

de Araújo, Rafael Cavalcanti; Cajueiro, Daniel Oliveira

## Article

# Combining term structure of interest rate forecasts: The Brazilian case

EconomiA

## Provided in Cooperation with:

The Brazilian Association of Postgraduate Programs in Economics (ANPEC), Rio de Janeiro

*Suggested Citation:* de Araújo, Rafael Cavalcanti; Cajueiro, Daniel Oliveira (2013) : Combining term structure of interest rate forecasts: The Brazilian case, EconomiA, ISSN 1517-7580, Elsevier, Amsterdam, Vol. 14, Iss. 2, pp. 102-121, <https://doi.org/10.1016/j.econ.2013.08.007>

This Version is available at:

<https://hdl.handle.net/10419/179550>

## Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

## Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by-nc-nd/3.0/>

# Combining term structure of interest rate forecasts: The Brazilian case

Rafael Cavalcanti de Araújo<sup>a</sup>, Daniel Oliveira Cajueiro<sup>b,c,\*</sup>

<sup>a</sup> Ministry of Finance, 70048-900, Brazil

<sup>b</sup> Department of Economics and INCT-SC, University of Brasília, DF 70910-900, Brazil

<sup>c</sup> Department of Physics and Astronomy, Northwestern University, Evanston, IL 60208-3112, USA

## Abstract

Issues like structural breaks and misspecification biases make it difficult to find a term structure of interest rates forecast model that dominates all competitors. Focusing on Brazilian data, this paper aims to identify the existence of combining methods that provide superior performance than individual models. Empirical results confirm that it is not possible to determine an individual model that consistently produces superior forecasts. Furthermore, the relative performance of these models may vary over time. The problems of using individual models may be reduced by applying forecast combining schemes. The empirical results show consistent forecast gains of combining schemes over time. In particular, the longer the forecast horizon, the greater the contribution of forecast combination.

© 2013 National Association of Postgraduate Centers in Economics, ANPEC. Production and hosting by Elsevier B.V. All rights reserved.

*JEL classification:* C53; E43; E47; G17

*Keywords:* Interest rate; Forecast model; Combined forecast

## Resumo

Problemas como quebras estruturais e vieses causados por má-especificação dificultam achar um modelo de previsão de estrutura a termo de taxa de juros que domine todos os competidores. Esse artigo tem como objetivo identificar a existência de métodos de combinação que produzam resultados de previsão superiores a modelos individuais no caso Brasileiro. Resultados empíricos confirmam que não é possível determinar um modelo individual que consistentemente produza previsões superiores. Além disso, o desempenho desses modelos varia temporalmente. Os problemas encontrados nos modelos individuais podem ser reduzidos aplicando esquemas de combinação de previsão. Os resultados mostram consistentemente ganhos de previsão nos esquemas de combinação para o período considerado. Em particular, quanto maior o horizonte de previsão, maior a contribuição do esquema.

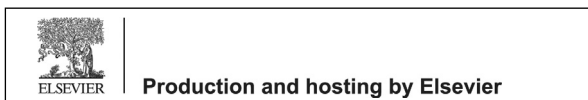
© 2013 National Association of Postgraduate Centers in Economics, ANPEC. Production and hosting by Elsevier B.V. All rights reserved.

*Palavras-chave:* Taxa de juros; modelo de previsão; previsão combinada

\* Corresponding author.

E-mail addresses: [rafael.araujo@fazenda.gov.br](mailto:rafael.araujo@fazenda.gov.br) (R.C. Araújo), [danielcajueiro@unb.br](mailto:danielcajueiro@unb.br) (D.O. Cajueiro).

Peer review under responsibility of National Association of Postgraduate Centers in Economics, ANPEC.



## 1. Introduction

The implementation of yield curve forecasting models for the Brazilian case is recent (Lima et al., 2006; Vereda et al., 2008; Vicente and Tabak, 2008; Almeida et al., 2009; Cajueiro et al., 2009). The results show that there is no single forecast model that dominates all competitors.<sup>1</sup> This is due to the fact that different models outperform the others, depending on time horizon ahead, maturity and forecast period.

Based on this fact, this study sets out to verify whether forecast combining methods can result in more accurate forecasts compared to individual models, focusing on the Brazilian market. In affirmative case, we analyze which combining schemes are adequate for different situations. For the forecasting of interest rates in Brazil there is no evidence of studies applying combination schemes. Hence, our paper contributes to the literature by implementing those schemes and evaluating their effectiveness in forecasting the yield curve.

The term structure of interest rates is of crucial importance for a variety of economic agents. For central banks, the yield curve helps to explain how monetary policy is implemented. For national treasuries, it indicates the yields for which public debt management should pay for issuing fixed rate bonds. For financial institutions, the term structure determines the allocation and exposure limits of different bonds maturities. For macroeconomists, it is an important leading indicator of economic activity and inflation. For firms, the term structure affects investments decisions. For households, the yield curve influences consuming and savings decisions.

In recent years studies in finance and in econometrics have tried to produce accurate forecasts of the term structure. Duffee (2002) uses the dynamic properties of affine models and concludes that his forecasts are dominated by the random walk model. The forecast underperformance of affine models stimulated Diebold and Li (2006) to adapt the static Nelson and Siegel (1987) factor model into a dynamic variant. Their results show that the model can beat the random walk, with the exception of the 1-month ahead horizon.<sup>2</sup>

After Diebold and Li (2006), many authors have developed different dynamic approaches of Nelson–Siegel model and have reported improvements in forecasting. Pooter (2007) adds new factors. Koopman et al. (2007) introduce time-varying parameters. Diebold et al. (2008) extend the model to a global context, modeling a large set of country yield curves in a framework that allows for both global and country-specific factors.

The interaction between macroeconomics and the yield curve has also generated promising results. Prominent examples are Evans and Marshall (1998), Ang and Piazzesi (2003), Hordahl et al. (2006), Diebold and Rudebush (2006), Moench (2008), Ludvigson and Ng (2009) and Pooter et al. (2010). It is important to point out that the use of macroeconomic information is beyond the scope of this study.

Despite good statistical properties, the models based on Nelson–Siegel framework do not impose the arbitrage-free restriction, which is a fundamental concept in the literature of bond pricing analysis. In order to solve this problem, Christensen et al. (2010) have developed an affine model that maintain the factor structure of Nelson and Siegel (1987), the so-called Arbitrage-Free Dynamic Nelson–Siegel. They conclude that improvements in predictive performance are achieved from the imposition of the absence of arbitrage. However, Duffee (2011) argues that this restriction is irrelevant for forecast gains and it only provides a mechanism to specify a functional form for risk compensation.

Bowsher and Meeks (2008) do not adopt the Nelson–Siegel framework. Instead, they introduce a factor model, known as Functional Signal Plus Noise with an Equilibrium Correction Model. Splines are used to model the yield curve. Besides, they argue that the knots of these splines act as factors. The results indicate that this model outperforms the random walk mainly at short-term horizons.

The concept of forecast combination was introduced in Bates and Granger (1969). They showed that combinations of individual forecasts often outperform individual forecasts. One possible reason is that the combination of forecasts adds value in the presence of structural breaks (Diebold and Pauly, 1987; Hendry and Clements, 2004). A number of papers (Diebold and Lopez, 1996; Stock and Watson, 2004; Timmermann, 2006) also defends that individual forecast models may be subject to misspecification biases of unknown form.

The main empirical findings of this paper are: (1) It is not possible to determine an individual model that consistently produces superior forecasts; (2) The relative performance of the individual models may vary over time; (3) The difficulties of using individual models can be reduced by applying forecast combining schemes; (4) Combining methods

<sup>1</sup> The same verification applies for U.S. data.

<sup>2</sup> In this work the shorter horizons is defined as 1 and 3-months ahead, whereas the longer horizons corresponds to 6 and 12-months ahead.

present consistent gains over time; (5) The longer the forecast horizon, the greater the contribution of the forecast combining scheme.

The remainder of the paper is structured as follows. Section 2 presents the forecast models used in this paper. Section 3 describes the alternative combining methods. Section 4 introduces the data. Section 5 presents the forecast procedures and the evaluating criteria. Section 6 shows the main results of this paper. It evaluates the performance of individual models and combining methods. Finally, Section 7 concludes the paper.

## 2. Individual forecast model

Let  $\tau = 1, 2, 3, 6, 9, 12, 15, 18, 21, 24, 30$  be the  $N = 11$  maturities in months of discount bonds with face value of \$1. Let  $y_t(\tau_i)$  denote their yields to maturity from time  $t$  to  $t + \tau_i$ , with  $i = 1, \dots, N$ . In addition, let  $h$  be the number of months. The  $h$ -month ahead forecast of each maturity  $\tau_i$  evaluated at time  $t$  is given by  $\hat{y}_{t+h}(\tau_i)$ .

In this work we consider models with different levels of complexity. The competing forecast models range from models with linear specifications (random walk, autoregressive, and vector autoregressive) to models with factor structures (Dynamic Nelson and Siegel and Functional Signal Plus Noise with an Equilibrium Correction Model). The random walk is used as a benchmark for all competing models.

### 2.1. Random walk

Consider a random walk (RW) without drift for the yield of each individual maturity  $\tau_i$  given by

$$y_t(\tau_i) = y_{t-1}(\tau_i) + \varepsilon_t(\tau_i), \quad (1)$$

where  $\varepsilon_t(\tau_i) \sim N(0, \sigma^2(\tau_i))$  is a white noise process.

The  $h$ -period ahead yield forecast is given by

$$\hat{y}_{t+h}(\tau_i) = y_t(\tau_i), \quad (2)$$

i.e., it is equal to the most recent observed value of  $y_t(\tau_i)$ .

### 2.2. Univariate autoregressive

In the forecast context, we consider two approaches of the first-order univariate autoregressive (AR) model. In the iterated approach, we estimate the model

$$y_t(\tau_i) = c(\tau_i) + \phi(\tau_i)y_{t-1}(\tau_i) + \varepsilon_t(\tau_i), \quad (3)$$

where  $\varepsilon_t(\tau_i) \sim N(0, \sigma^2(\tau_i))$  and  $c(\tau_i)$ ,  $\phi(\tau_i)$  and  $\sigma^2(\tau_i)$  are scalar parameters.

For each maturity, the  $h$ -months ahead forecast is produced by

$$\hat{y}_{t+h}(\tau_i) = \hat{c}(\tau_i) + \hat{\phi}(\tau_i)\hat{y}_{t+h-1}(\tau_i). \quad (4)$$

On the other hand, in the direct approach, the estimated coefficients summarize the  $h$ -months ahead effect. In this case, we estimate the model

$$y_t(\tau_i) = c_h(\tau_i) + \phi_h(\tau_i)y_{t-h}(\tau_i) + \varepsilon_t(\tau_i), \quad (5)$$

where  $\varepsilon_t(\tau_i) \sim N(0, \sigma^2(\tau_i))$ . To obtain the forecast, we regress  $y_t$  directly on its past value  $y_{t-h}$  as in

$$\hat{y}_{t+h}(\tau_i) = \hat{c}_h(\tau_i) + \hat{\phi}_h(\tau_i)\hat{y}_t(\tau_i). \quad (6)$$

Although we have also implemented the direct approach for the Brazilian case, we do not consider it in this work. Similar to the findings of [Carriero et al. \(2010\)](#) for U.S. data, the forecast results produced by this approach were inferior than those of the iterated approach.

### 2.3. Vector Autoregressive

The Vector Autoregressive (VAR) model presents the advantage that each maturity also uses the historical data of the other maturities. In the present work, we use a VAR model with the following specification<sup>3</sup>

$$Y_t = c + \Phi Y_{t-1} + H\varepsilon_t(\tau), \quad (7)$$

where  $\varepsilon_t(\tau_i) \sim N(0, I)$ ,  $Y_t = (y_t(1), \dots, y_t(30))'$  is a vector  $(11 \times 1)$  that contains the yields of different maturities,  $c$  is a vector  $(11 \times 1)$ ,  $\Phi$  is a matrix  $(11 \times 11)$  and  $H$  is a matrix of residual variance containing  $(1/2)N(N+1) = 66$  free parameters.

Due to the high number of parameters, we replace the vector  $Y_{t-1}$  by a 3-dimensional principal components vector.<sup>4</sup> Hence, we consider the model given by

$$Y_t = c + \Phi F_{t-1} + \varepsilon_t(\tau), \quad (8)$$

where  $\varepsilon_t(\tau_i) \sim N(0, H)$  and  $\Phi$  becomes a matrix  $(11 \times 3)$ . Then, we obtain the forecast from

$$\hat{Y}_{t+h} = \hat{c} + \hat{\Phi} \hat{F}_{t+h-1}. \quad (9)$$

In order to calculate  $\hat{F}_{t+h-1}$ , we previously compute the principal component factor loadings using only data up to month  $t$ . Then, we multiply these loadings with the iterated forecasts.

### 2.4. Dynamic Nelson and Siegel

The Dynamic Nelson–Siegel (DNS) model has as starting point the static Nelson and Siegel (1987), in which the instantaneous forward rate is defined as a polynomial multiplied by the exponential decay

$$f_t(\tau) = \beta_1 + \beta_2 e^{-\lambda\tau} + \beta_3 \lambda \tau e^{-\lambda\tau}. \quad (10)$$

Nelson and Siegel (1987) affirm that the equation above can be interpreted as an approximation of a differential equation solution. Furthermore, the authors argue that the expectations theory provides a heuristic motivation for the utilization of this approach. More specifically, they defend that if the spot rate is generated by a differential equation, then forward rates, being forecasts, are the solutions of this equation. The correspondent term structure is given by<sup>5</sup>

$$y_t(\tau_i) = \beta_1 + (\beta_2 + \beta_3) \left( \frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} \right) + \beta_3 (-e^{-\lambda\tau_i}). \quad (11)$$

In order to adequate the Nelson and Siegel (1987) framework to a model of latent factors with dynamic structure, Diebold and Li (2006) adopt a modern interpretation of Eq. (11). They assume that  $\beta_1$ ,  $\beta_2$  and  $\beta_3$  are time dependent variables, respectively, interpreted as level, slope and curvature as presented in

$$y_t(\tau_i) = \beta_{1t} + \beta_{2t} \left( \frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} \right) + \beta_{3t} \left( \frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i} \right), \quad (12)$$

where each term that multiplies the factors are known as factor loadings.<sup>6</sup>

The level factor  $\beta_{1t}$  controls the long term maturities. The slope factor  $\beta_{2t}$  commands the short term maturities. The curvature factor  $\beta_{3t}$  governs the medium term maturities. The parameter  $\lambda$  controls not only the speed of exponential decay of the factor loadings on  $\beta_{2t}$  and  $\beta_{3t}$ , but also the position in which the factor loading  $\beta_{3t}$  achieves its maximum.

In this work,  $\lambda$  is fixed as in Diebold and Li (2006). Similar to Cajueiro et al. (2009), we consider the maturity of 12 months, which in turn corresponds to the middle point of the sector of the yield curve considered medium, with

<sup>3</sup> The model specification of the VAR model used in this paper is similar to Pooter et al. (2010).

<sup>4</sup> Similar to Litterman and Scheinkman (1991) and Pooter et al. (2010), the first three principal components explain almost 99.91% of yield curve variation and have the same interpretation: level, slope and curvature.

<sup>5</sup> The relation between the forward rate and the yield curve is obtained by  $y(\tau) = (1/\tau) \int_0^\tau f_t(u) du$ .

<sup>6</sup> Defined by Litterman and Scheinkman (1991) as the sensibility of the bond yield in relation to its respective factor.

extremes in 6 and 18 months.<sup>7</sup> Therefore, the value that maximizes the loading  $((1 - e^{-\lambda\tau_i})/\lambda\tau_i) - e^{-\lambda\tau_i}$  on  $\tau = 12$  is 0.14944.

Similar to [Diebold and Li \(2006\)](#), we assume that the factors follow a first-order autoregressive process given by

$$\hat{\beta}_{jt+h} = \hat{\alpha}_0 + \hat{\alpha}_1 \tilde{\beta}_{jt}, \quad (13)$$

where  $\tilde{\beta}_{jt}$ , for  $j = 1, 2, 3$ , are the Ordinary Least Squares (OLS) estimations for each date  $t$  using the data of all maturities and Eq. (12). Furthermore,  $\hat{\alpha}_0$  and  $\hat{\alpha}_1$  are the coefficients of the regression of  $\tilde{\beta}_{jt}$  on  $\tilde{\beta}_{j,t-h}$ .

Once the  $\hat{\beta}_j$  are available in  $t+h$ , the yields forecasts can be generated from

$$\hat{y}_{t+h}(\tau_i) = \hat{\beta}_{1t+h} + \hat{\beta}_{2t+h} \left( \frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} \right) + \hat{\beta}_{3t+h} \left( \frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i} \right). \quad (14)$$

## 2.5. Functional Signal Plus Noise with an Equilibrium Correction Model

[Bowsher and Meeks \(2008\)](#) have proposed the so-called Functional Signal Plus Noise with an Equilibrium Correction Model (FSN-ECM). This framework is based on an economic function (functional signal), in that case a latent yield curve, which is observed with a measurement error (noise). In order to model the yield curve, the authors establish a natural cubic spline (NCS), whose dynamic evolution is driven by a cointegrated VAR in the form of error correction model (ECM). They argue that the FSN-ECM can be interpreted as a special type of dynamic factor model in which the knots of the splines are the factors and the factor loadings are generated by the natural cubic spline.

Let  $S_{\gamma_t}(\tau) = (S_{\gamma_t}(\tau_1), \dots, S_{\gamma_t}(\tau_N))'$  be a dynamic NCS evaluated at the vector of observed maturities  $\tau = (\tau_1, \dots, \tau_N)$  with dimension  $N$ . Let  $m$  be the number of knots of the NCS positioned at the latent maturities  $k = (1, k_2, \dots, k_m)$ , which in turn are fixed over time. Consider  $\gamma_t = (\gamma_{1t}, \dots, \gamma_{mt})'$  the vector yields evaluated at the knots, i.e.,  $S_{\gamma_t}(k_{jt}) = \gamma_{jt}$  for  $j = 1, \dots, m$ , denominated the knot-yields of the spline. Therefore,  $S_{\gamma_t}(\tau)$  is interpolated from  $\gamma_t$ , being uniquely determined by  $\gamma_t$ . It is important to stress that the dimension of vector  $k$  is smaller than the dimension of vector  $\tau$ , that is,  $m < N$ .

The model for the observation vectors  $Y_t(\tau) = (y_t(\tau_1), \dots, y_t(\tau_N))'$  is given by

$$\begin{aligned} Y_t(\tau) &= S_{\gamma_t}(\tau) + \varepsilon_t(\tau) \\ &= W(k, \tau)\gamma_t + \varepsilon_t(\tau) \\ \Delta\gamma_{t+1} &= \alpha(\beta'\gamma_t - \mu) + \Psi\Delta\gamma_t + v_t, \end{aligned} \quad (15)$$

where  $W(k, \tau)$  corresponds to the interpolating matrix, depending only on  $\tau$  and  $k$ .  $S_{\gamma_t}(\tau) = W(k, \tau)\gamma_t$  is given by Lema B.1 of [Bowsher and Meeks \(2008\)](#). In addition,  $\alpha$  is a full rank matrix  $m \times (m-1)$  and  $\beta$  is a matrix determined uniquely by  $\beta'\gamma_t = (\gamma_{j+1,t} - \gamma_{j,t})_{j=1}^{m-1}$ .

The initial state  $(\gamma'_1, \gamma'_0)'$  has first and second moments given by  $\gamma^* e \omega^*$  respectively. The vector  $u_t = (\varepsilon_t(\tau)', v_t')'$  is a white noise process.

Eq. (15) describes the state equation of ECM using a cointegrated VAR. The process  $\gamma_t$  is integrated of first-order,  $I(1)$ . The difference among the latent yields  $\beta'\gamma_t$  are cointegrated and  $\beta'\gamma_t - \mu$  is a vector of stationary cointegrated relations with zero mean. Then  $E[\beta'\gamma_t] = \mu$  is the stationary mean of the cointegrating relations and  $E[\Delta\gamma_{t+1}] = 0$ , hence excluding deterministic trend.

Under the conditions defined above, [Bowsher and Meeks \(2008\)](#) argue that FSN-ECM model may be structured in a linear state-space form. This representation form allows the use of the Kalman filter to implement both quasi-maximum likelihood estimation and 1-step ahead forecast.

Therefore, it is possible to define a model specification through FSN( $m$ )-ECM( $p$ ), where  $m$  is the number of knots and  $p$  is the maximum lag of  $\gamma_{t+1}$  that enters in Eq. (15). In this work, we only consider the case in which  $m = 5$ . On one hand, smaller values of  $m$  may result in a very large mean residual sum of the squares ([Bowsher and Meeks, 2008](#)). On the other hand, larger values of  $m$  may cause overfitting of the data used in the estimation step. In addition, we

<sup>7</sup> In this study, the following definitions of maturities are considered: short maturities (up to 3 months), medium maturities (from 6 up to 12 months) and long maturities (above 18 months).

use  $p=2$ , since [Bowsher and Meeks \(2008\)](#) states that ECM(1) models provide forecasts with higher Mean Square Forecast Error (MSFE) than ECM(2) models do.

In order to proceed with the estimation, additional restrictions are imposed in the terms of the non-singular matrix  $Q$ , where  $\varphi = Q\gamma_t$  is the following transformed state vector, which contains the latent short rate and inter-knot latent yield spreads

$$\varphi_t = (\gamma_{1,t}, \gamma_{2,t} - \gamma_{1,t}, \dots, \gamma_{m,t} - \gamma_{m-1,t})' = \begin{bmatrix} 1 & 0_{1 \times (m-1)} \\ \beta' \end{bmatrix} \gamma_t = Q\gamma_t. \quad (16)$$

The state equation can be written as the VAR

$$\Delta\varphi_{t+1} = Q\alpha(\beta'Q^{-1}\varphi_t - \mu_s) + Q\Psi Q^{-1}\Delta\varphi_t + \eta_t, \quad (17)$$

where  $\eta_t = Qv_t \in \Omega_\eta = \text{var}[\eta_t] = Q\Omega_v Q'$ . Both covariance matrices  $\Omega_\eta$  and  $\Omega_\varepsilon = \sigma_\varepsilon^2 I_N$  are diagonal as in [Bowsher and Meeks \(2008\)](#).<sup>8</sup> The Kalman Filter is initialized by using  $(\gamma'_1, \gamma'_0) = (\mu^*, \Omega^*)$ , where  $\Omega^* = 0$  and  $\mu$  is set equal to  $(y_0(k)', y_{-1}(k)')'$ .

For computational purposes,<sup>9</sup> we write Eq. (17) as

$$\Delta\varphi_{t+1} = [I + Q\alpha\beta'Q^{-1} + Q\Psi Q^{-1}] \varphi_t + Q\Psi Q^{-1}\varphi_{t-1} + Q\alpha\mu_s + Qv_t. \quad (18)$$

We define vector  $X_t$  as

$$X_t = \begin{bmatrix} \varphi_t \\ \varphi_{t-1} \\ 1 \end{bmatrix}. \quad (19)$$

Therefore, the state-space representation becomes

$$X_{t+1} = \begin{bmatrix} A & B & \bar{C} \\ I & 0 & 0 \\ 0 & 0 & I \end{bmatrix} X_t + \begin{bmatrix} Q \\ 0 \\ 0 \end{bmatrix} v_t, \quad (20)$$

where  $A = [I + Q\alpha\beta'Q^{-1} + Q\Psi Q^{-1}]$ ,  $B = Q\Psi Q^{-1}$  and  $\bar{C} = -Q\alpha\mu_s$ . While matrix  $A$  is fully estimated, we assume that matrix  $B$  is a diagonal matrix<sup>10</sup> when  $m \leq 5$ , as in [Bowsher and Meeks \(2008\)](#) and [Cajueiro et al. \(2009\)](#).

### 3. Forecast combining methods

In this section we implement ten combining methods organized into four classes: simple (simple mean, median and trimmed mean), Inverted Mean Square Forecast Error (Inverted MSFE), weighted least squares (WLS) and via Model Confidence Set (MCS). In this last class, we apply all previous methods after excluding the worst performance forecast models.

In order to test the predictive capacity of the models, we split the data into in-sample (the first  $R$  observations) and out-of-sample (the last  $P$  observations). Furthermore, in some schemes, namely inverted MSFE, WLS and combining via MCS, we need an extra out-of-sample period (training period) for computing the weights of the individual forecast models. In this case, we use the first  $P_0$  out-of-sample observations. More specifically, we calculate the weights using the individual out-of-sample forecasts,  $\{\hat{y}_{t+h|t}(\tau_i)\}_{t=R+P_0}^{t-h}$ , and the observations,  $\{y_{t+h|t}(\tau_i)\}_{t=R+P_0}^{t-h}$ , both available from the beginning of the training out-of-sample period to time  $t$ . In turn, we use the  $P_1$  post-training out-of-sample observations to compare the forecast performance of the individual models and combining schemes.

<sup>8</sup> While the diagonality  $\Omega_\varepsilon$  may be justified by the assumption of the independence of the shocks, the diagonality of  $\Omega_\eta$  is clearly imposed as a simplification of the estimation framework.

<sup>9</sup> See in [Cajueiro et al. \(2009\)](#).

<sup>10</sup> This is a simplification meaning that only own lagged changes of the spreads matters for each equation of the system (17).



Let  $C$  be the number of forecast models to be combined. The combined  $h$ -months ahead forecast for the yield with maturity  $\tau_i$  is given by

$$\hat{y}_{t+h|t}(\tau_i) = f(\hat{\mathbf{y}}_{t+h|t}(\tau_i); \mathbf{w}_{t+h|t}(\tau_i)), \quad (21)$$

where  $\hat{\mathbf{y}}_{t+h|t}(\tau_i)$  indicates the  $C$ -dimensional vector of forecast models and  $\mathbf{w}_{t+h|t}(\tau_i)$  denotes the parameters or weights associated with the combination. In addition, the function  $f$  transforms the vector  $\hat{\mathbf{y}}_{t+h|t}(\tau_i)$  into the summary measure  $\hat{y}_{t+h|t}(\tau_i)$ .

### 3.1. Simple combinations

We consider three simple combining methods of individual forecasts: simple mean (SM), median (MED) and trimmed mean (TM).

**Simple mean:** It gives the same weights for all individual models, i.e.  $w_{t+h|t,c}(\tau_i) = 1/C$ . Then, the combined forecast is given by

$$\hat{y}_{t+h|t}(\tau_i) = \sum_{c=1}^C \frac{1}{C} \hat{y}_{t+h|t,c}(\tau_i). \quad (22)$$

**Median:** It uses the sample median of individual models. Thus, the combined forecast is

$$\hat{y}_{t+h|t}(\tau_i) = \text{median of } \left\{ \hat{y}_{t+h|t,c} \right\}_{c=1}^C. \quad (23)$$

**Trimmed mean:** It makes the weights of individual forecast models that produce the largest and smallest forecast values equal to zero ( $w_{t+h|t,c}(\tau_i) = 0$ ). For the other models, we assign  $w_{t+h|t,c}(\tau_i) = 1/(C-2)$  and evaluate the forecast as the simple mean that uses only these  $C-2$  remaining individual models as in

$$\hat{y}_{t+h|t}(\tau_i) = \sum_{c=1}^{C-2} \frac{1}{C-2} \hat{y}_{t+h|t,c}(\tau_i). \quad (24)$$

### 3.2. Inverted Mean Square Forecast Error

In this combining class, the weights in Eq. (21) are inversely proportional to the MSFE of each individual model, computed over a window of the 24 previous months<sup>11</sup> using

$$w_{t+h|t,c}(\tau_i) = \frac{1/MSFE_{h|t,c}^{(\tau_i)}}{\sum_{c=1}^C 1/MSFE_{h|t,c}^{(\tau_i)}},$$

where  $w_{t+h|t,c}(\tau_i)$  denotes the weight assigned to period  $t$  forecast of the  $c$ th model ( $\hat{y}_{t+h|t,c}$ ) and  $MSFE_{h|t,c}^{(\tau_i)} = (1/24) \sum_{p_0=1}^{24} (\hat{y}_{t-p_0+1|t-h-p_0+1,c}^{(\tau_i)} - y_{t-p_0+1}^{(\tau_i)})^2$ . Therefore, the combined forecast is given by

$$\hat{y}_{t+h|t}(\tau_i) = \sum_{c=1}^C w_{t+h|t,c}(\tau_i) \hat{y}_{t+h|t,c}(\tau_i). \quad (25)$$

<sup>11</sup> This is the same window that is considered in the training period used for the computation of the weights. This window is also used in the WLS and MCS combinations.



### 3.3. Weighted Least Squares

In order to compute the initial WLS combined forecast for the observation  $y_{R+P_0+h}(\tau_i)$ , we regress the previous observations  $\{y_{s+h}(\tau_i)\}_{s=R}^{R+P_0-h}$  on a constant and  $\{\hat{y}_{c,s+h|s}(\tau_i)\}_{s=R}^{R+P_0-h}$  for  $c = 1, \dots, C$ . The weight of each model is set equal to the WLS estimator

$$\hat{\beta}_{MQP} = (X'W^{-1}X)^{-1}X'W^{-1}y,$$

where  $W$  is the weighted mean matrix ( $T \times T$ ). To obtain the second forecast for the observation  $y_{R+P_0+h+1}^h(\tau_i)$ , we estimate the weights by regressing  $\{y_{s+h}^h(\tau_i)\}_{s=R+1}^{R+P_0-h+1}$  on a constant and  $\{\hat{y}_{c,s+h|s}^h(\tau_i)\}_{s=R+1}^{R+P_0-h+1}$ , for  $c = 1, \dots, C$ . This procedure is repeated until the end of the out-of-sample period. Then, we use the WLS coefficients as weights in the combined forecast equation

$$\hat{y}_{t+h|t}(\tau_i) = w_{0,t} + \sum_{c=1}^C w_{t+h|t,c}(\tau_i) \hat{y}_{t+h|t,c}(\tau_i). \quad (26)$$

Diebold and Pauly (1987) recommend the use of a diagonal matrix  $W = \text{diag}[\psi_{it}] = kt^\lambda$ , where  $t = 1, \dots, T$  and  $T$  is the number of observations used in the WLS regression. Hence, the weights may be declining at an increasing rate, at a decreasing rate or at a constant rate, depending on the signal of  $(\lambda - 1)$ . Similar to the case of the inverted MSFE method,  $T$  is fixed and equal to 24. Furthermore,  $\lambda$  and  $k$  used in this work are equal to 1, that is, the weights of older observations decrease linearly. Finally, it is worth emphasizing that the WLS is the only unrestricted combining method, i.e., the weights assigned to each individual model can assume any value. In other cases, the weights vary between 0 and 1, with the sum of them equal to 1.

### 3.4. Combining via Model Confidence Set

In this combining scheme, we remove the models with the worst forecasting performances as considered in Hansen et al. (2003) and Hansen et al. (2011). Then, we combine the forecasts of the remaining models by the simple combinations, the inverted MSFE and the WLS.

Given a forecast model set  $C_0$  (the set that contains all individual forecast models), the MCS procedure identifies

$$\hat{C}_\alpha^* \subset C_0$$

where  $\hat{C}_\alpha^*$  is the subset that contains the best forecast models given a confidence level  $(1 - \alpha)$ .

Starting with  $C = C_0$  and a forecast vector  $P_0$ ,<sup>12</sup> the MCS procedure repetitively tests the null hypothesis of forecast accuracy:

$$H_{0,C} : E[d_{ij,t}] = 0, \forall i, j \in C,$$

where  $d_{ij,t} = L_{i,t} - L_{j,t}$  is the differential loss between models  $i$  and  $j$  and  $L$  is the loss function. The MCS procedure sequentially eliminates the model with the worst performance, as long as the null hypothesis is rejected. We repeat the procedure until the null hypothesis is no longer rejected, in which case the remaining set is  $\hat{C}_\alpha^*$ . In order to proceed with the test, we use the semi-quadratic statistic

$$T_{SQ} \equiv \sum_{i,j \in C} t_{i,j}^2,$$

where  $t_{i,j} = (\bar{d}_{i,j} / \sqrt{\hat{v}ar(\bar{d}_{i,j})})$  for  $i, j \in C$  and  $\bar{d}_{i,j} = (1/P_0) \sum_{t=1}^{P_0} d_{ij,t}$ .

Therefore, to obtain a  $h$ -months ahead combined forecast, we use the statistic  $T_{SQ}$  to construct  $\hat{C}_{0.25}^*$ , with a confidence level of  $(1 - \alpha) = 75\%$ <sup>13</sup> as in Pooter et al. (2010). To generate the probability distribution under the null

<sup>12</sup> In this case  $P_0 = 24$ .

<sup>13</sup> The set of models are based on the significance level of 75%, since in many cases the forecast error differences are relatively small. Thus, traditional significance levels (such as 5%) are not very informative, since no model may be excluded of the MCS.

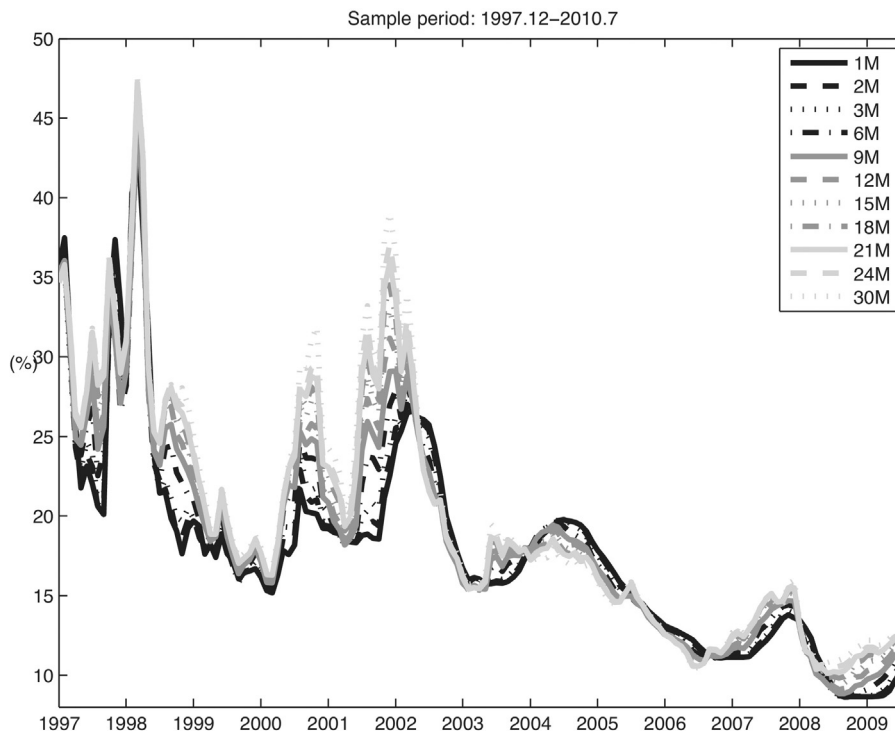


Fig. 1. Evolution of interest rates.

Source: BM&FBovespa.

hypothesis, we use a bootstrap implementation through stationary blocks,<sup>14</sup> with 1000 resamples, average block length of 20<sup>15</sup> and the MSFE as the loss function.

#### 4. Yield data

We have obtained the data employed for the yield curve construction from the interest rates swaps traded at BM&FBovespa futures and stocks exchange.<sup>16</sup> In these contracts, a party pays a fixed rate and in exchange receives a floating rate. The reverse occurs with the counterparty of the contract. The resulting cash flow exchange is computed over the same notional value. In turn, by acting as a central counterpart, BM&FBovespa guarantees these contracts. Therefore, those interest rates may be seen as proxies for default-free interest rates.<sup>17</sup>

We have considered the following maturities of the swaps contracts: 1, 2, 3, 6, 9, 12, 15, 18, 21, 24 and 30 months. The data begins in December 1997 and ends in July 2010. Each swap is daily observed (3112 samples), with the monthly data constructed from the daily averages (152 samples).

Fig. 1 presents the dynamic evolution of the data. After a period of many crises that affected Brazil (from 1997 to 2002), we observe a long run tendency of level reduction of the yield curve. This phenomenon may be justified by the improvement of macroeconomic conditions verified in Brazil from 2003 until the end of the sample period.

<sup>14</sup> Developed by Politis and Romano (1994).

<sup>15</sup> The choice of block size 20 can be justified by a trade off between the preservation of the dependence structure of the data and the variation of the bootstrap sample (small variations may result in poor approximations).

<sup>16</sup> On the contrary to what occurs in the U.S., where the prices of the bonds are observed, in the Brazilian market the interest rates are the ones that are directly traded. Hence, it is not necessary to estimate the Brazilian term structure from the bonds prices.

<sup>17</sup> Although international literature uses government bond yields as reference, the use of interest rates swaps is justified for some reasons: (1) there are not enough old data for government bonds; (2) interest rates swaps are more liquid and are highly correlated with government bonds yields.

Table 1  
Descriptive statistics of interest rates.

Maturity	Mean	StdDev	Min	Max	$\rho(1)$	$\rho(12)$	$\rho(24)$
1m	17.81	6.67	8.61	43.95	0.92	0.40	0.14
2m	17.84	6.57	8.62	43.59	0.92	0.41	0.16
3m	17.93	6.56	8.62	44.86	0.92	0.42	0.17
6m	18.27	6.66	8.69	44.88	0.93	0.46	0.20
9m	18.59	6.87	8.84	45.01	0.93	0.47	0.20
12m	18.83	7.02	9.10	45.64	0.93	0.47	0.20
15m	19.14	7.26	9.47	46.19	0.93	0.47	0.20
18m	19.39	7.45	9.82	46.92	0.94	0.47	0.19
21m	19.58	7.60	10.10	47.23	0.94	0.47	0.18
24m	19.75	7.73	10.37	47.37	0.94	0.47	0.18
30m	20.04	7.95	10.43	47.43	0.94	0.48	0.18

Note: The last three columns present the sample autocorrelations for 1, 12 and 24 months.

Table 2  
(Trace) Relative Root Mean Squared Forecast Error, 1-month horizon.

Models	(T)RMSFE	1m	2m	3m	6m	9m	12m	15m	18m	21m	24m	30m
RW	0.54	0.39 (0.00)	0.39 (0.00)	0.40 (0.07)	0.43 (0.10)	0.48 (0.05)	0.53 (0.05)	0.57 (0.07)	0.60 (0.07)	0.63 (0.12)	0.66 (0.22)	0.70 (0.20)
<b>Panel A: Individual models</b>												
AR	0.75	0.78 (0.00)	0.71 (0.02)	0.69 (0.34)	0.74* (0.34)	0.75* (0.56)	0.76* (0.66)	0.76* (0.66)	0.76* (0.68)	0.76* (0.66)	0.76* (0.66)	0.76* (0.76)
VAR	0.90	0.50 (0.56)	0.46* (0.63)	0.51* (0.54)	0.80 (0.02)	0.97 (0.02)	1.02 (0.07)	1.02 (0.15)	0.99 (0.17)	0.95 (0.17)	0.92 (0.27)	0.91 (0.07)
DNS	0.85	1.00 (0.24)	0.91 (0.44)	0.87 (0.56)	0.84 (0.61)	0.86 (0.56)	0.88 (0.51)	0.88 (0.59)	0.86 (0.59)	0.82 (0.68)	0.79 (0.71)	0.79 (0.83)
FSN-ECM	0.82	0.44* (0.73)	0.48 (0.37)	0.54* (0.27)	0.74* (0.17)	0.86 (0.12)	0.92 (0.07)	0.91 (0.07)	0.89 (0.07)	0.86 (0.12)	0.85 (0.12)	0.87 (0.12)
<b>Panel B: Combining methods</b>												
SM	0.79	0.67	0.63	0.64	0.75	0.81	0.84	0.85	0.84	0.81	0.80	0.80
MED	0.76	0.66	0.62	0.62	0.74	0.78	0.80	0.79	0.79	0.77	0.77	0.78
TM	0.77	0.65	0.61	0.62	0.73	0.78	0.82	0.82	0.81	0.79	0.78	0.79
MSFE	0.77	0.55	0.54	0.58	0.73	0.80	0.83	0.83	0.83	0.80	0.79	0.79
WLS	0.91	0.53	0.60	0.55	0.73	0.85	0.98	0.94	0.98	0.95	0.97	1.08
<b>Panel C: Combining via MCS</b>												
MCS-SM	0.75	0.49	0.47	0.56	0.69	0.76	0.80	0.84	0.81	0.78	0.79	0.79
MCS-MED	0.75	0.48	0.47	0.57	0.70	0.76	0.78	0.82	0.81	0.77	0.78	0.79
MCS-TM	0.75	0.48	0.47	0.57	0.70	0.76	0.78	0.82	0.81	0.77	0.78	0.79
MCS-MSFE	0.75	0.49	0.47	0.56	0.69	0.76	0.80	0.83	0.81	0.78	0.79	0.79
MCS-WLS	0.77	0.59	0.62	0.68	0.71	0.92	0.84	0.65	0.73	0.75	0.79	0.87

Note: The out-of-sample period for 1-month horizon used in the table is 2006.4–2009.8. The training period used in the combining methods contains a 24-months rolling window, beginning in 2004.4–2006.3 and ending in 2007.8–2009.7. The individual models parameters are estimated using a 76-months rolling window, beginning in 1997.12–2004.3 and ending in 2001.4–2007.7. We use the following abbreviations: RW stands for the Random Walk, AR for Autoregressive, VAR for Vector Autoregressive, DNS for Dynamic Nelson Siegel, FSN-ECM for Functional Signal Plus Noise with an Equilibrium Correction Model, SM for Simple Mean, MED for Median, TM for Trimmed Mean, MSFE for Mean Squared Forecast Error, WLS for Weighted Least Squares. The prefix “MCS” in Panel C indicates that the combining methods of Panel B are applied after excluding the worst performing forecast models according to  $\hat{C}_{0.25}^*$ .

Table 3  
(Trace) Relative Root Mean Squared Forecast Error, 3-months horizon.

Models	(T)RMSFE	1m	2m	3m	6m	9m	12m	15m	18m	21m	24m	30m
RW	1.29	1.09 (0.63)	1.10* (0.44)	1.12* (0.49)	1.18* (0.76)	1.25* (0.83)	1.31* (0.85)	1.37* (0.83)	1.39* (0.83)	1.41* (0.83)	1.42* (0.83)	1.46* (0.83)
<b>Panel A: Individual models</b>												
AR	1.06	1.15 (0.00)	1.09 (0.02)	1.07* (0.34)	1.10* (0.46)	1.08* (0.61)	1.07* (0.66)	1.05* (0.66)	1.04* (0.68)	1.04* (0.66)	1.03* (0.66)	1.03* (0.85)
VAR	1.48	0.81 (0.59)	0.91* (0.68)	1.02* (0.61)	1.34 (0.10)	1.49 (0.02)	1.56 (0.07)	1.59 (0.15)	1.63 (0.17)	1.65 (0.17)	1.67 (0.27)	1.70 (0.07)
DNS	1.17	1.34 (0.78)	1.31 (0.80)	1.28* (0.83)	1.24* (0.88)	1.20* (0.80)	1.17 (0.73)	1.14 (0.78)	1.12 (0.71)	1.10 (0.73)	1.08 (0.78)	1.07* (0.83)
FSN-ECM	1.44	0.78* (0.76)	0.91* (0.37)	1.03* (0.27)	1.30* (0.17)	1.45 (0.12)	1.53 (0.07)	1.55 (0.07)	1.57 (0.07)	1.58 (0.12)	1.61 (0.12)	1.64 (0.12)
<b>Panel B: Combining methods</b>												
SM	1.12	0.91	0.94	0.98	1.09	1.14	1.16	1.16	1.17	1.17	1.17	1.18
MED	1.05	0.97	1.04	1.06	1.11	1.09	1.08	1.06	1.05	1.04	1.04	1.03
TM	1.11	0.89	0.94	0.98	1.09	1.13	1.15	1.15	1.15	1.15	1.15	1.16
MSFE	1.03	0.85	0.91	0.96	1.05	1.07	1.08	1.07	1.07	1.06	1.06	1.05
WLS	1.25	0.89	0.80	0.74	0.94	1.31	1.39	1.41	1.40	1.33	1.34	1.42
<b>Panel C: Combining via MCS</b>												
MCS-SM	0.97	0.83	0.83	0.92	1.04	0.99	0.99	1.00	0.99	0.99	0.99	0.99
MCS-MED	0.96	0.83	0.84	0.89	1.03	0.99	0.98	1.00	0.99	0.99	0.98	0.98
MCS-TM	0.96	0.83	0.84	0.89	1.03	0.99	0.98	1.00	0.99	0.99	0.98	0.98
MCS-MSFE	0.97	0.83	0.83	0.91	1.03	0.99	0.99	1.00	0.99	0.99	0.99	0.99
MCS-WLS	1.10	0.92	0.73	0.82	1.01	1.25	1.24	1.19	1.13	1.10	1.12	1.19

*Note:* The out-of-sample period for 3-months horizon used in the table is 2006.6–2009.10. The training period used in the combining methods contains a 24-months rolling window, beginning in 2004.6–2006.5 and ending in 2007.10–2009.9. The individual models parameters are estimated using a 76-months rolling window, beginning in 1997.12–2004.3 and ending in 2001.4–2007.7.

Table 1 presents descriptive statistics of the full sample. The typical yield curve is upward sloping. The volatility of the yields increases with the maturity. Finally, the amplitude of the yields indicates the huge change of level of yield curve during the period under analysis.

## 5. Methodology

This section presents the forecast procedures and evaluating criteria.

### 5.1. Forecast procedure

Similar to [Cajueiro et al. \(2009\)](#), in order to proceed with out-of-sample forecasts, we estimate the models using the first half of the sample (in-sample), that is, a rolling estimation window of 76 months. The initial estimation window is 1997.12–2004.03. Then, we reestimate the models including the next month and excluding the first month of the previous sample. The parameters are updated until the final estimation window, 2001.04–2007.07. We obtain the forecasts for the following horizons ahead: 1, 3, 6 and 12 months. We use the models to generate daily forecasts and we convert these forecasts to the monthly average.

Similar to [Almeida et al. \(2008\)](#), every month ahead corresponds to 21 working days. Inside the out-of-sample period, we use the first 24 months to obtain the weights of the individual models that are used in the forecast combinations (training period  $P_0$  described in Section 3). We compare the individual models and the combining schemes with the RW, which is the model used as the benchmark in the yield curve forecast.

Table 4  
(Trace) Relative Root Mean Squared Forecast Error, 6-months horizon.

Models	(T)RMSFE	1m	2m	3m	6m	9m	12m	15m	18m	21m	24m	30m
RW	2.16	1.96* (0.76)	1.97* (0.83)	1.99* (0.88)	2.09* (0.85)	2.18* (0.83)	2.24 * (0.85)	2.29* (0.83)	2.29* (0.83)	2.27* (0.83)	2.24* (0.83)	2.20* (0.83)
<b>Panel A: Individual models</b>												
AR	1.15	1.24 (0.00)	1.20* (0.12)	1.18* (0.49)	1.20* (0.51)	1.17* (0.61)	1.15* (0.66)	1.12* (0.66)	1.12* (0.68)	1.10* (0.66)	1.10* (0.66)	1.08* (0.85)
VAR	1.77	1.15 (0.59)	1.25* (0.78)	1.35* (0.61)	1.58 (0.12)	1.70 (0.12)	1.77 (0.07)	1.83 (0.15)	1.91 (0.17)	1.98 (0.17)	2.06 (0.27)	2.20 (0.07)
DNS	1.15	1.32 (0.83)	1.30* (0.93)	1.27* (0.93)	1.21* (0.90)	1.16* (0.88)	1.11* (0.85)	1.08* (0.95)	1.07* (0.93)	1.06* (0.93)	1.06* (0.95)	1.07* (0.90)
FSN-ECM	1.69	1.11* (0.93)	1.23* (0.41)	1.33* (0.29)	1.53 (0.17)	1.65 (0.12)	1.72 (0.07)	1.76 (0.07)	1.82 (0.07)	1.88 (0.12)	1.96 (0.12)	2.08 (0.12)
<b>Panel B: Combining methods</b>												
SM	1.24	1.08	1.11	1.14	1.21	1.24	1.25	1.25	1.27	1.29	1.31	1.35
MED	1.15	1.14	1.16	1.17	1.21	1.18	1.17	1.14	1.13	1.12	1.12	1.11
TM	1.24	1.08	1.12	1.15	1.22	1.24	1.25	1.25	1.27	1.28	1.30	1.33
MSFE	1.11	1.07	1.10	1.13	1.16	1.15	1.13	1.11	1.10	1.09	1.08	1.07
WLS	0.85	0.81	0.60	0.56	0.71	0.83	0.85	0.89	0.91	0.95	0.97	1.00
<b>Panel C: Combining via MCS</b>												
MCS-SM	1.05	1.12	1.15	1.13	1.15	1.11	1.03	1.00	0.98	0.97	0.97	0.97
MCS-MED	1.04	1.13	1.14	1.12	1.13	1.09	1.03	1.00	0.99	0.97	0.98	0.97
MCS-TM	1.04	1.13	1.14	1.12	1.13	1.09	1.03	1.00	0.99	0.97	0.98	0.97
MCS-MSFE	1.04	1.11	1.14	1.13	1.15	1.10	1.02	1.00	0.98	0.97	0.97	0.97
MCS-WLS	0.92	0.88	0.85	0.76	0.82	0.87	0.96	0.95	0.95	1.06	0.96	0.92

Note: The out-of-sample period for 6-months horizon used in the table is 2006.9–2010.1. The training period used in the combining methods contains a 24-months rolling window, beginning in 2004.9–2006.8 and ending in 2008.1–2009.12. The individual models parameters are estimated using a 76-months rolling window, beginning in 1997.12–2004.3 and ending in 2001.4–2007.7.

## 5.2. Evaluation result procedures

In order to evaluate the full out-of-sample period, we compute the Root Mean Squared Forecast Error (RMSFE) for each forecast horizon and maturity

$$RMSFE_c^{(\tau_i)} = \sqrt{\frac{1}{P_1} \sum_{t=1}^{P_1} (\hat{y}_{t+h|t,c}^{(\tau_i)} - y_{t+h}^{(\tau_i)})^2}. \quad (27)$$

The accuracy of forecasts is appraised by the Relative Root Mean Squared Forecast Error (ReRMSFE) given by

$$ReRMSFE_c^{(\tau_i)} = \frac{RMSFE_c^{(\tau_i)}}{RMSFE_{RW}^{(\tau_i)}}. \quad (28)$$

Similar to [Hordahl et al. \(2006\)](#) and [Pooter et al. \(2010\)](#), we summarize the performances of each model and combining method across all maturities by evaluating the Trace Root Mean Squared Forecast Error ([Christoffersen and Diebold, 1998](#)) given by

$$TRMSFE_c = \sqrt{\frac{1}{N} \frac{1}{P_1} \sum_{i=1}^N \sum_{t=1}^{P_1} (\hat{y}_{t+h|t,c}^{(\tau_i)} - y_{t+h}^{(\tau_i)})^2}. \quad (29)$$

Furthermore, in order to compare to the RW, we use

$$TRReRMSFE_c = \frac{TRMSFE_c}{TRMSFE_{RW}}. \quad (30)$$

Table 5  
(Trace) Relative Root Mean Squared Forecast Error, 12-months horizon.

Models	(T)RMSFE	1m	2m	3m	6m	9m	12m	15m	18m	21m	24m	30m
RW	2.97	2.83* (0.98)	2.84* (0.93)	2.86* (0.98)	2.96* (0.95)	3.04* (0.93)	3.09* (0.88)	3.11* (0.83)	3.08* (0.83)	3.02* (0.83)	2.96* (0.83)	2.89* (0.83)
<b>Panel A: Individual models</b>												
AR	1.21	1.3 (0.12)	1.27 (0.17)	1.26 (0.49)	1.27 (0.51)	1.24 (0.61)	1.21 (0.66)	1.18 (0.66)	1.17 (0.68)	1.16 (0.66)	1.16 (0.66)	1.14* (0.85)
VAR	2.16	1.67 (0.59)	1.75 (0.78)	1.82 (0.61)	1.97 (0.12)	2.05 (0.12)	2.11 (0.07)	2.19 (0.15)	2.28 (0.17)	2.39 (0.17)	2.50 (0.27)	2.70 (0.07)
DNS	1.15	1.29 (0.83)	1.27 (0.93)	1.24 (0.93)	1.17 (0.95)	1.12 (0.95)	1.09* (0.88)	1.07* (0.95)	1.07* (0.93)	1.08* (0.93)	1.10* (0.95)	1.13* (0.90)
FSN-ECM	1.95	1.56 (0.98)	1.64 (0.46)	1.70 (0.29)	1.82 (0.17)	1.88 (0.12)	1.93 (0.07)	1.97 (0.07)	2.04 (0.07)	2.13 (0.12)	2.22 (0.12)	2.38 (0.12)
<b>Panel B: Combining methods</b>												
SM	1.38	1.28	1.30	1.32	1.36	1.36	1.36	1.36	1.39	1.42	1.45	1.51
MED	1.23	1.30	1.30	1.30	1.30	1.26	1.23	1.19	1.18	1.17	1.17	1.16
TM	1.37	1.29	1.32	1.33	1.36	1.36	1.36	1.35	1.37	1.39	1.42	1.47
MSFE	1.17	1.24	1.25	1.25	1.23	1.19	1.16	1.14	1.12	1.11	1.11	1.10
WLS	0.57	0.59	0.55	0.52	0.53	0.55	0.57	0.59	0.59	0.59	0.58	0.56
<b>Panel C: Combining via MCS</b>												
MCS-SM	1.04	1.25	1.21	1.19	1.02	0.99	0.97	0.97	0.96	0.97	0.97	0.99
MCS-MED	1.04	1.26	1.20	1.15	1.01	0.99	0.97	0.97	0.96	0.97	0.97	0.99
MCS-TM	1.04	1.26	1.20	1.15	1.01	0.99	0.97	0.97	0.96	0.97	0.97	0.99
MCS-MSFE	1.04	1.24	1.20	1.18	1.02	0.99	0.97	0.97	0.96	0.97	0.97	0.99
MCS-WLS	0.88	0.85	0.79	0.74	1.13	1.04	0.91	0.84	0.83	0.82	0.82	0.82

Note: The out-of-sample period for 12-months horizon used in the table is 2007.3–2010.7. The training period used in the combining methods contains a 24-months rolling window, beginning in 2005.3–2007.2 and ending in 2008.7–2010.6. The individual models parameters are estimated using a 76-months rolling window, beginning in 1997.12–2004.3 and ending in 2001.4–2007.7.

In order to test if the forecast error differentials of the models are statistically different, we apply the MCS approach described in Section 3.4. The relevant change here is the number of observations considered,  $P_1$  instead of  $P_0$ . This is due to the fact that  $P_1$  is the number of out-of-sample observations used to compare the performances of the models and combining schemes.

To analyze in which periods of the sample the models or methods underperform or outperform the RW, we graphically display the Cumulative Squared Forecast Error (CSFE) for the accumulated period  $T$  given by

$$CSFE_{c,T}^{(\tau_i)} = \sum_{t=1}^T [(\hat{y}_{t+h|t,RW}^{(\tau_i)} - y_{t+h}^{(\tau_i)})^2 - (\hat{y}_{t+h|t,c}^{(\tau_i)} - y_{t+h}^{(\tau_i)})^2]. \quad (31)$$

Finally, to summarize this measure for all maturities, we use the TCSFE, which is the simple average of  $CSFE_{c,T}^{(\tau_i)}$ , for  $i = 1, \dots, N$ .<sup>18</sup>

It is worth mentioning that this approach was applied for stocks by Welch and Goyal (2008) and adapted to interest rates by Pooter et al. (2010).

## 6. Results

This section presents the results of the paper. For the sake of clearness, we divide this section in two parts. While in the first subsection we present the results for the individual models, in the second subsection we analyze the results for the combining methods.

<sup>18</sup> In fact, although the original trace MSFE statistic proposed by Christoffersen and Diebold (1998) is a sum, in this work also we divide it by  $N$  to normalize the measures.

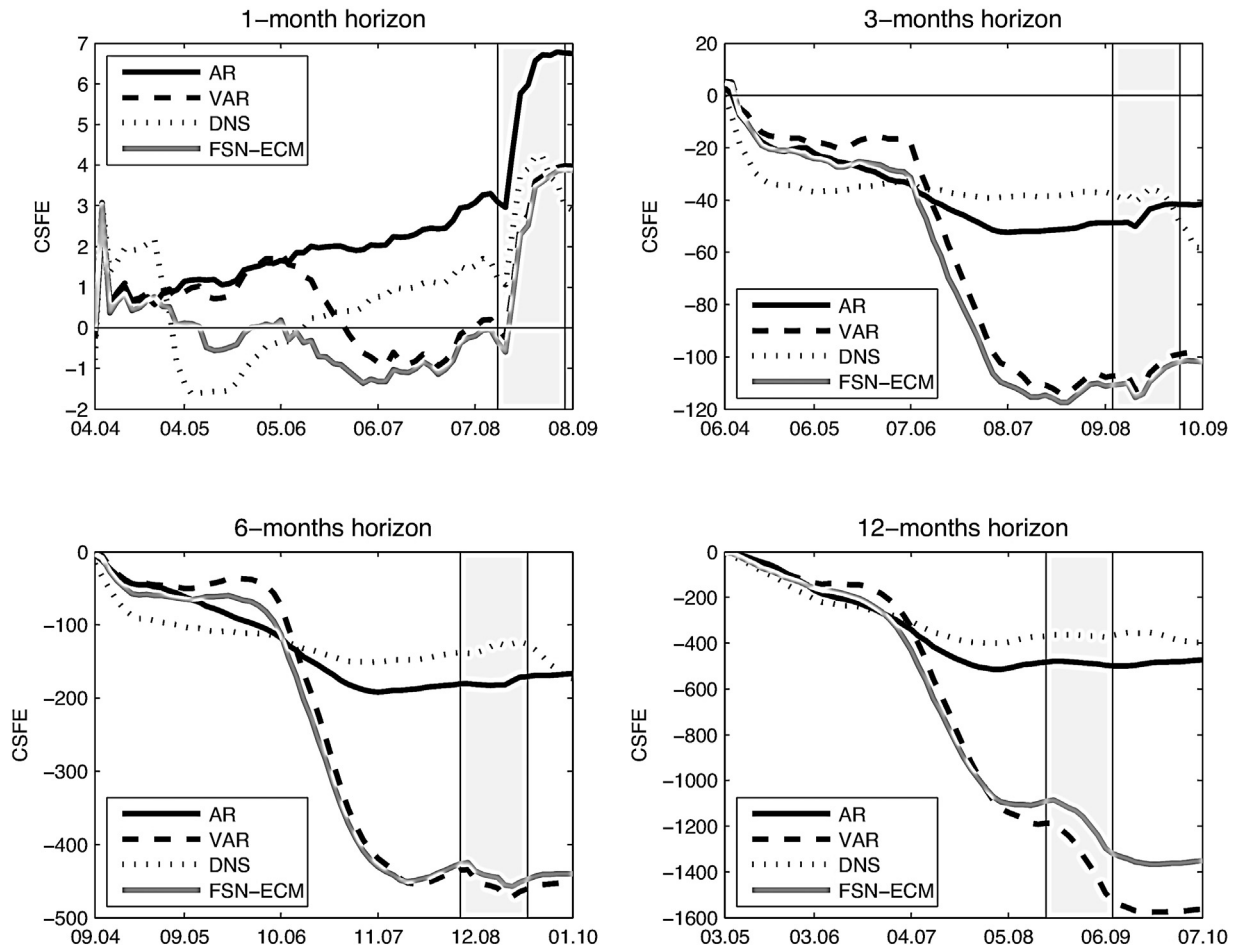


Fig. 2. Trace Cumulative Squared Forecast Error (including training period). TCSFE of individual models. 1-month horizon out-of-sample period: 2004.4–2009.8. 3-months horizon out-of-sample period: 2004.06–2009.10. 6-months horizon out-of-sample period: 2004.9–2010.1. 6-months horizon out-of-sample period: 2005.03–2010.07.

### 6.1. Individual models results

We start the discussion of the individual models performance considering the results in terms of the (Trace) Relative Root Mean Squared Forecast Error, (T)ReRMSFE, displayed for each forecast horizon in Tables 2–5. The first row of each table contains the (T)RMSFE for the RW. The remaining rows in Panels A, B and C<sup>19</sup> report the (T)ReRMSFE for the other models relative to the RW, in accordance with Eq. (29). While the first column of each row corresponds to the TRReRMSFE, the remaining columns present the ReRMSFE.

Thus, a number smaller than 1 indicates that the model outperforms the RW and a number larger than 1 suggests the model underperforms. The stars on the upper right of the ReRMSFE of the models denote that the model belongs to the set  $\hat{C}_{0.25}^*$ , whereas the data in parentheses corresponds to the percentage of times that each model is included in  $\hat{C}_{0.25}^*$ .

In terms of 1-month ahead forecasts, in general, the individual models beat the RW for almost all maturities. The only exceptions are the DNS and the VAR models for the few maturities. Additionally, the AR model seems to be the one with the best general performance. However, in the other forecast horizons ( $h=3$ ,  $h=6$  and  $h=12$ ) no model

<sup>19</sup> For the forecast evaluation of the individual models only matters Panel A.



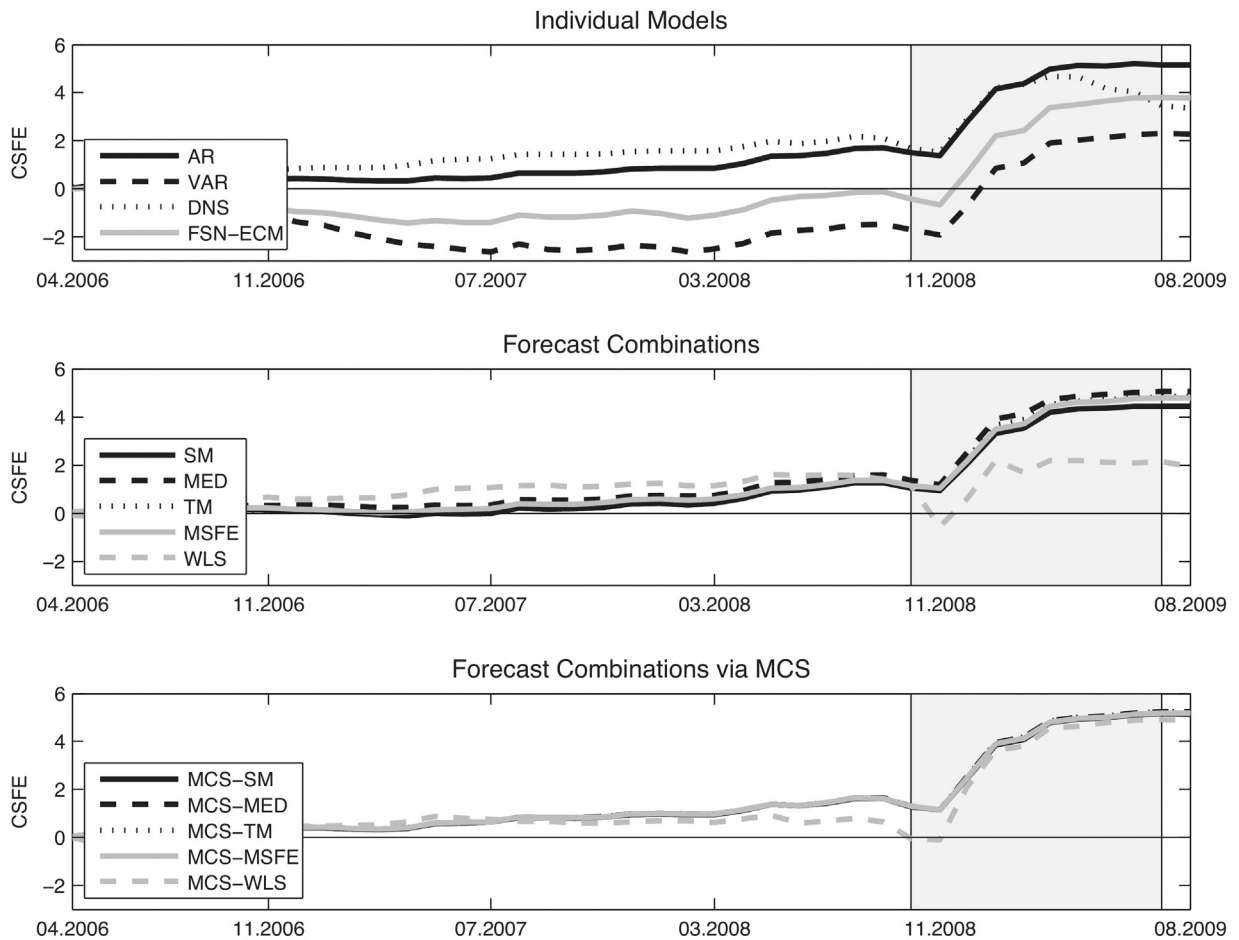


Fig. 3. Trace Cumulative Squared Forecast Error, 1-month horizon. TCSFE of individual models and combining methods: 2006.6–2009.8.

systematically outperforms the RW. In these horizons, the AR and the DNS models present regular performances, whereas the VAR and the FSN-ECM underperform the RW.

Regarding the significance test of the MCS approach, it is worth pointing out that although the AR and the DNS present ReRMSFE above 1 for horizons larger than 1 month, the results of the tests suggest not being possible to assume the superiority of the RW in many maturities, specially in the longest ones. However, this result differs in some sense from [Cajueiro et al. \(2009\)](#),<sup>20</sup> once we also cannot affirm the superiority of the DNS. This fact demonstrates that the forecast performances of the models can vary over time, a result defended theoretically by [Timmermann \(2006\)](#) and verified empirically for the US yield curve<sup>21</sup>.

Analyzing over the specific maturities, the FSN-ECM and the VAR models present superior performance at the shorter maturities and horizons. On the other hand, the AR demonstrates good performance at longer maturities until the 6-month horizons. Finally, the DNS has better performance at the longer maturities and horizons.

We consider now the CSFE graphical analysis generated for the entire out-of-sample period.<sup>22</sup> Fig. 2 shows the graphs for each forecast horizon. Every line in the graph represents a model and displays the forecast performance relative to the RW. An upward slope of CSFE signals the model is outperforming, whereas the opposite indicates the

<sup>20</sup> In their work, considering a shorter sample period (1997.12–2008.03), they find that the DNS outperformed the RW for all maturities and horizons. It is worth mentioning that we also find this result, if we consider this smaller period.

<sup>21</sup> As in [Bowsher and Meeks \(2008\)](#), and [Pooter et al. \(2010\)](#).

<sup>22</sup> In particular, [Pooter et al. \(2010\)](#) asserts that this option helps to show explicitly the performance of the individual models during the training period.

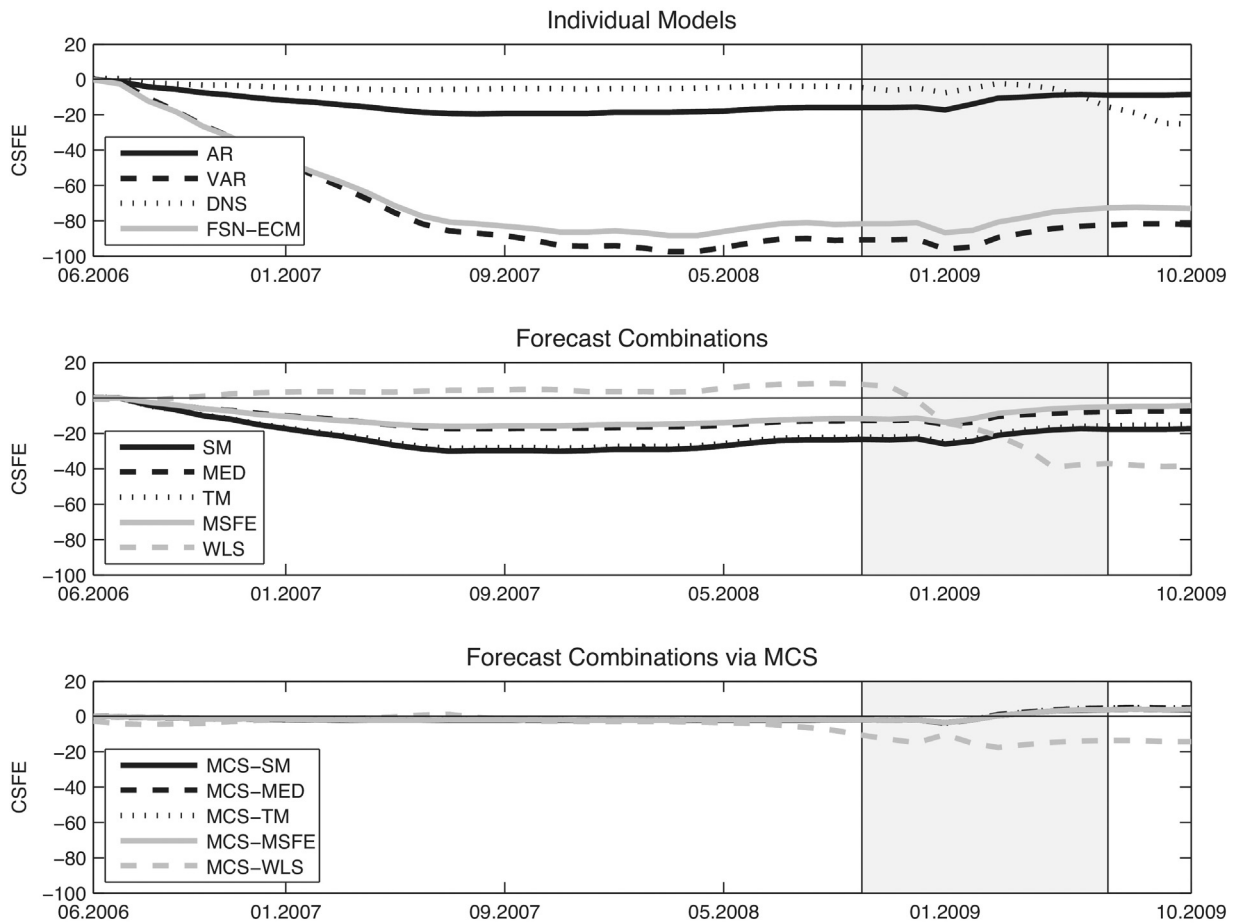


Fig. 4. Trace Cumulative Squared Forecast Error, 3-months horizon. TCSFE of individual models and combining methods: 2006.6–2009.10.

model is underperforming. In order to reduce the number of graphs displayed, we use the Trace Cumulative Squared Error (TCSFE).

In general terms, the training period analysis ( $P_0$ ) suggests a reduced volatility in the performances of the models, except for the DNS at 1-month ahead horizon. On the other hand, in the subsequent periods ( $P_1$ ), we have found that: (1) the VAR and the FSN-ECM present poor performance from the end of 2006 to the end of the sample for the longer horizons; (2) all models outperform the RW in the period of international financial crisis (from the end of 2008 to the middle of 2009) for the 1-month ahead horizon; (3) the DNS underperforms the RW after the peak of international crisis (middle 2009) for the 3 and 6-months ahead horizons; (4) the AR and the DNS, on one side, and the VAR and the FSN-ECM, on the other side, present a similar performance.

## 6.2. Combining methods results

In order to analyze the performance of the combining schemes, we consider Panels B and C in Tables 2–5. While Panel B refers to the combinations of all models, Panel C considers only the ones that are not eliminated by the MCS approach.

First, at each forecast horizon, it is always possible to find at least one combining scheme that has presented equal or superior performance than the best individual model. In addition, the longer the horizon, the greater the contribution of the combinations. At the shorter horizons, the simple combinations and the inverted MSFE trimmed by MCS presented performances similar to the AR in 1-month ahead and superior performance than the RW in 3-months ahead. At the

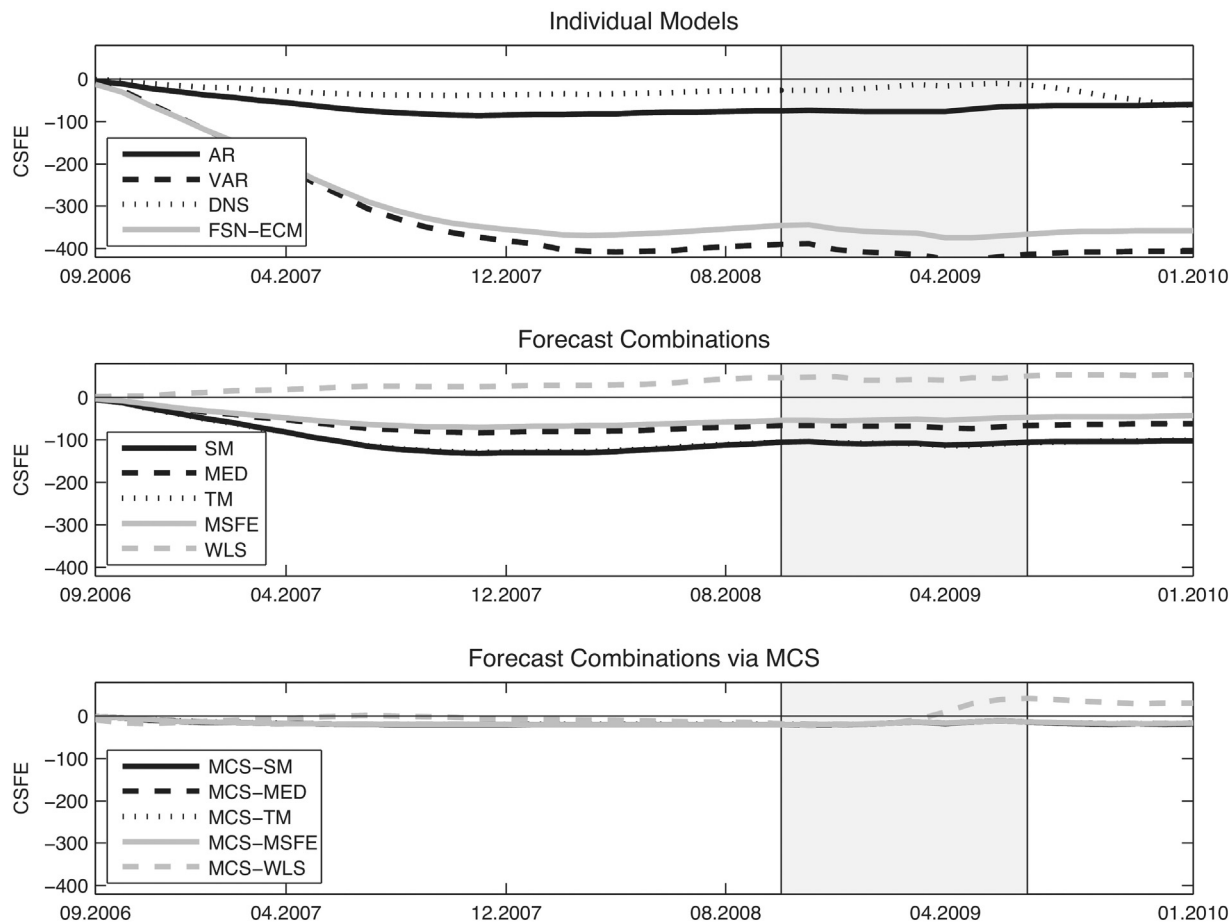


Fig. 5. Trace Cumulative Squared Forecast Error, 6-months horizon. TCSFE of individual models and combining methods: 2006.9–2010.1.

longer horizons, the combinations by WLS (trimmed or not by MCS approach) produce smaller forecast errors than the RW.

Second, comparing Panels B and C, we can verify the good performance of the combinations after applying the MCS approach. Except for the WLS combinations for the longer horizons, we have found for each scheme superior forecast results after the exclusion of worst individual models. The percentages in the tables (in parentheses) indicate that most of the time the methods via MCS select the best performance models. Other interest point is that, in the shorter horizons, in which the combinations via MCS present the best results, the differences among them are negligible. Therefore, we can conclude that in these cases the trimming procedure of individual models via MCS is more beneficial for the forecast than the way how the weights are computed. This is in line with [Timmermann \(2006\)](#) that affirms that the trimming of the worst models often improves performance.

Third, the WLS combining method (without MCS), which is the unique unrestricted method, presents quite different performances depending on the forecast horizon. While, in the shorter horizons, this scheme presents large forecast errors, in the longer horizons, this scheme is the one that presents the best performance, with even more emphasis in the 12-month ahead horizon. In this horizon, the outperformance relative to the RW was above 40% in all maturities. This feature of best performance of an unrestricted method is similar to the empirical findings of [Bolder and Romanyuk \(2008\)](#) with the application of the simple OLS for the Canadian interest rates. As [Timmermann \(2006\)](#) points out, the schemes that use regressions can adjust model bias through the intercept term. It is also important to note that although the WLS combining methods are effective in the longer horizons, it has the disadvantage of being more sensitive to the training period used in the weights determination.

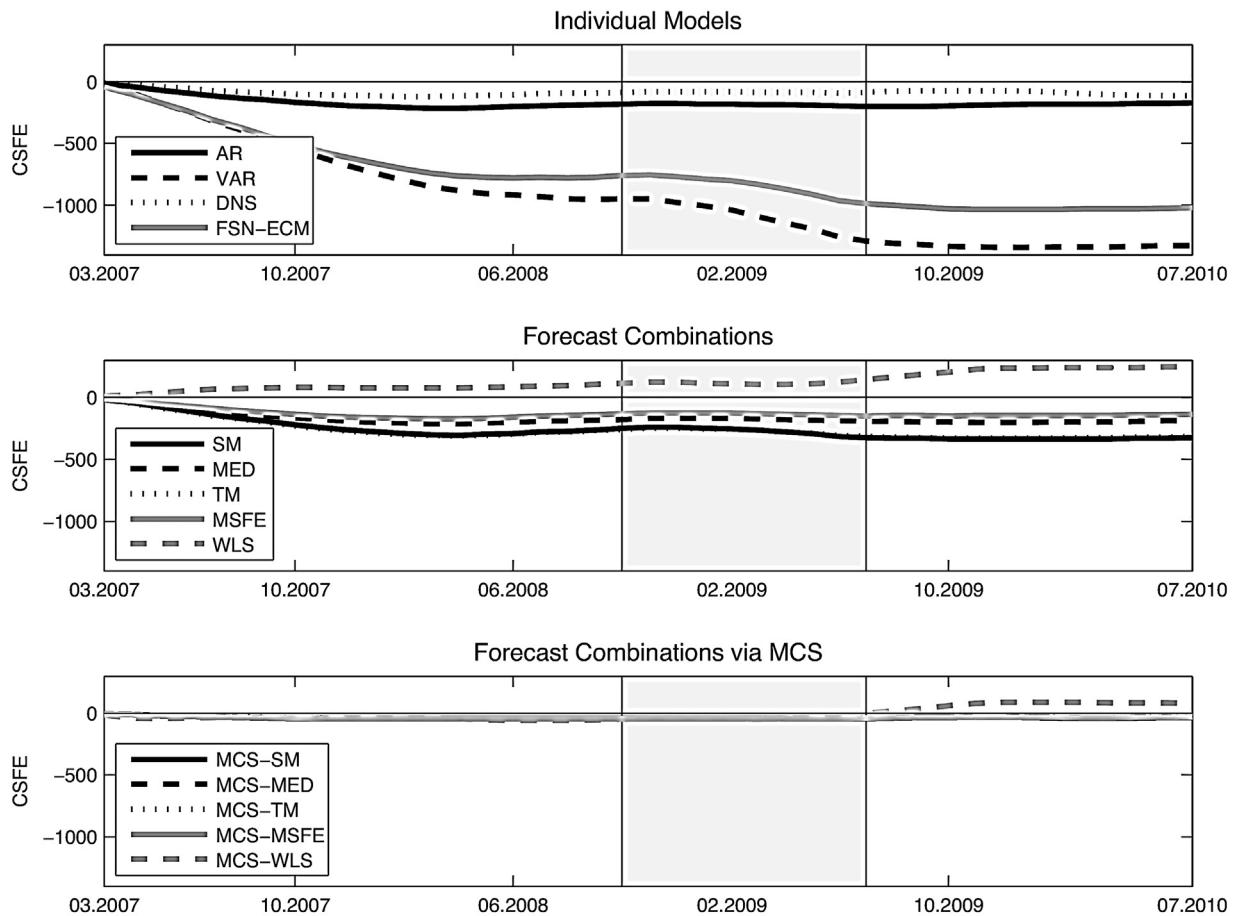


Fig. 6. Trace Cumulative Squared Forecast Error, 12-months horizon. TCSFE of individual models and combining methods: 2007.3–2010.7.

Figs. 3–6 can also be used to value the quality of the combining methods. In order to confront them in a more effective way, these figures contain the same period of Tables 2–5. The forecast errors of the combinations are certainly less volatile than the ones of individual models. Among the combining methods, emphasis is given to the methods via MCS, whose forecast errors are even less volatile. In terms of eventual benefits of the use of time-varying parameters in the case of structural breaks, we find empirical evidence of forecast improvement in the shorter horizons of the inverted MSFE combination and in the longer horizons of the WLS combination<sup>23</sup> in the period of international financial crisis (2008–2009). Another important point is that the poor forecast performance of the VAR and the FSN-ECM above 1-month ahead is mitigated with the introduction of the combining methods.<sup>24</sup> This verification corroborates with Timmermann (2006) argument that the combining schemes may be interpreted as a way of turning the forecasts more robust against the specification bias of individual models.

## 7. Conclusions

In this paper we have assessed the performance of individual models and combining methods using Brazilian data.

In the case of individual models, we conclude that it is not possible to identify an individual model that consistently produces smaller forecast errors. The performance of each model is contingent to the forecast horizon, maturity and period of the sample. This fact confirms that models with different specifications contribute in a complementary way to

<sup>23</sup> In both cases, with or without trimming via MCS.

<sup>24</sup> In particular, in the VAR and the FSN-ECM with trimming via MCS, the forecast errors are in most of the time strongly reduced.

approximate the data generating process of the term structure, which is a strong argument for the use of the combining schemes.

The results of the combining methods are in consonance with this analysis. For each forecast horizon it is possible to find at least one combining scheme that presents performance equal or superior to the best individual model. In addition, the longer the horizon, the greater the contribution of the combinations.

The forecast errors of the combining schemes are certainly less volatile than the individual models, specially in the combinations via MCS. These specific combinations also present superior results in the shorter horizons, suggesting that they can be used in these situations.

For 1-month ahead forecasts, the simple combinations via MCS and the MSFE via MCS present a performance similar to the best individual model, which is the AR. On the other hand, for 3-month ahead forecasts, they present a performance superior than the best individual model, which is the RW. Another relevant point is that the difference among the results of these combinations is negligible. Thus, we can conclude that in these cases the trimming procedures by MCS are more beneficial for the forecast combination than the way how the weights are computed.

In the longer horizons, the WLS combination presents the best performance. This is a significant result since it is well known that in the longer ahead horizons it is very difficult to find a yield curve forecast model that beats the simple RW.

## Acknowledgments

We thank Professor José Guilherme de Lara Resende (Department of Economics – University of Brasília), Rodrigo Cabral (World Bank), the referees and the editors for their comments on a preliminar version of this work. The remaining errors are ours. The second author also thanks the Brazilian agencies CAPES (9207-11-2) and CNPQ (300974/2009-4) for partial financial support in different steps of this work. The opinions expressed in this paper are those of the authors and do not necessarily reflect those of the Ministry of Finance.

## References

- Almeida, C., Gomes, R., Leite, A., Simonsen, A., Vicente, J., 2009. Does curvature enhance forecasting? *International Journal of Theoretical and Applied Finance* 12, 1171–1196.
- Almeida, C., Gomes, R., Leite, A., Vicente, J., 2008. Movimentos da estrutura a termo e critérios de minimização do erro de previsão em um modelo paramétrico exponencial. *Revista Brasileira de Economia* 62 (4), 497–510.
- Ang, A., Piazzesi, M., 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics* 50, 745–787.
- Bates, J.M., Granger, C.W.J., 1969. The combination of forecasts. *Operational Research Quarterly* 29, 451–468.
- Bolder, D.J., Romanyuk, Y., 2008. Combining Canadian interest-rate forecasts, Working Paper Bank of Canada 34.
- Bowsher, C., Meeks, R., 2008. The dynamics of economic functions: modeling and forecasting the yield curve. *Journal of American Statistical Association* 103, 1419–1437.
- Cajueiro, D., Divino, J.A., Tabak, B., 2009. Forecasting the yield curve for Brazil, Central Bank of Brazil Working Paper Series 197.
- Carriero, A., Kapetanios, G., Marcellino, M., 2010. Forecasting government bond yields with large Bayesian vars, University of London, Queen Mary Working Paper 662.
- Christensen, J., Diebold, F., Rudebusch, G., 2010. The affine arbitrage-free class of Nelson–Siegel term structure models. *Journal of Econometrics* 164, 4–20.
- Christoffersen, P.F., Diebold, F.X., 1998. Cointegration and long-horizon forecasting. *Journal of Business & Economic Statistics* 16, 450–458.
- Diebold, F., Li, C., 2006. Forecasting the term structure of government bond yields. *Journal of Econometrics* 130, 337–364.
- Diebold, F., Li, C., Yue, V., 2008. Global yield curve dynamics and interactions: a generalized Nelson–Siegel approach. *Journal of Econometrics* 146, 351–363.
- Diebold, F., Rudebusch, G.B.A., 2006. The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics* 131, 309–338.
- Diebold, F.X., Lopez, J.A., 1996. Forecast Evaluation and Combination in *Handbook of Statistics*. North-Holland, pp. 241–268, chapter 14.
- Diebold, F.X., Pauly, P., 1987. Structural change and the combination of forecasts. *Journal of Forecasting* 6, 21–40.
- Duffee, G., 2002. Term premia and interest rate forecasts in affine models. *The Journal of Finance* 57, 405–443.
- Duffee, G., 2011. Forecasting with the term structure: the role of no-arbitrage. The Johns Hopkins University, Department of Economics Working Paper 576.
- Evans, C., Marshall, D., 1998. Monetary policy and the term structure of nominal interest rates: evidence and theory. *Carnegie-Rochester Conference Series on Public Policy* 49, 53–111.

- Hansen, P.R., Lunde, A., Nason, J., 2003. Choosing the best volatility model: the model confidence set approach. *Oxford Bulletin of Economics and Statistics* 65, 839–861.
- Hansen, P.R., Lunde, A., Nason, J., 2011. The model confidence set. *Econometrica* 79 (2), 453–497.
- Hendry, D., Clements, M., 2004. Pooling of forecasts. *Econometrics Journal* 7, 1–31.
- Hordahl, P., Tristani, O., Vestin, D., 2006. A joint econometric model of macroeconomic and term-structure dynamics. *Journal of Econometrics* 131, 405–444.
- Koopman, S., Malle, M., Well, M.V.D., 2007. Analyzing the term structure of interest rates using the dynamic Nelson–Siegel model with time-varying parameters. *Tinbergen Institute Discussion Papers*, Tinbergen Institute 095/4.
- Lima, E., Ludvigson, F., Tabak, B., 2006. Forecasting interest rates: an application for Brazil, Central Bank of Brazil Working Paper Series 120.
- Litterman, R., Scheinkman, J., 1991. Common factors affecting bond returns. *Journal of Fixed Income* 1, 51–61.
- Ludvigson, S.C., Ng, S., 2009. Macro factors in bond risk premia. *Review of Financial Studies* 22 (12), 5027–5067.
- Moench, E., 2008. Forecasting the yield curve in a data-rich environment: a no-arbitrage factor-augmented var approach. *Journal of Econometrics* 146 (1), 26–43.
- Nelson, C., Siegel, A., 1987. Parsimonious modeling of yield curves. *The Journal of Business* 60, 473–489.
- Politis, D., Romano, J.P., 1994. The stationary bootstrap. *Journal of the American Statistical Association* 89, 1303–1313.
- Pooter, M.D., 2007. Examining the Nelson–Siegel class of term structure models, 07-043/4.
- Pooter, M.D., Ravazzolo, F., Dijk, D.V., 2010. Term structure forecasting using macro factors and forecast combination. Board of Governors of the Federal Reserve System International Finance Discussion Papers 993.
- Stock, J., Watson, M., 2004. Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting* 23, 405–430.
- Timmermann, A., 2006. Forecast Combinations in *Handbook of Economic Forecasting*. North-Holland, pp. 135–196, chapter 4.
- Vereda, L., Lopes, H., Fukuda, R., 2008. Estimating var models for the term structure of interest rates. *Insurance: Mathematics and Economics* 42, 548–559.
- Vicente, J., Tabak, B.M., 2008. Forecasting bonds yields in the Brazilian fixed income market. *International Journal of Forecasting* 24 (3), 490–497.
- Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies* 21, 1455–1508.