Evolving Participatory Learning Fuzzy Modeling for Yield Curve Forecasting with Time-Varying Volatility

Leandro Maciel  Fernando Gomide  Rosangela Ballini
Department of Computer Engineering and Automation  School of Electrical and Computing Engineering  Department of Economic Theory
University of Campinas  Institute of Economics  University of Campinas
Albert Einstein Ave. 400, 13083-852, Campinas, SP  Pitágoras St. 353, 13083-857, Campinas, SP
E-mails: {maciel,gomide}@dca.fee.unicamp.br  E-mail: ballini@eco.unicamp.br

Abstract. This paper suggests a dynamic approach for the term structure of interest rates forecasting using evolving participatory learning fuzzy modeling (ePL). The model includes a time-varying volatility structure in order to predict the yield curve factors. Thus, this framework both comprises an adaptive framework for term structure parameters behavior and deal with the uncertainty related to these factors by describing its variability patterns. Results based on US Treasury market data indicate that the ePL model outperformed a common approach in the literature, based on autoregressive processes, for short- and long-term horizons considering the fitness accuracy.

Keywords: Fuzzy Systems, Adaptive Forecasting, Interest Rate, GARCH Models.
1 Introduction

The term structure of interest rates depicts the relationship between the nominal rates of default-free zero-coupon bonds and time to maturity. Its fitting and forecasting is an extremely useful tool for finance and macroeconomics. The yield curve may be employed for the pricing of bonds, interest rate derivatives, asset allocation, portfolio and risk management and corporate financial decisions. In macroeconomics, the interest rate gives relevant informations on the economic status and it is also used for the implementation and evaluation of the monetary policy.

The major models devoted for the yield curve modeling can be categorized into three main classes: no-arbitrage, equilibrium, and statistical models. No-arbitrage models are constructed by the imposition of consistency conditions between the interest rates of several maturities so as to prevent the systematic existence of arbitrage opportunities [12]. On the other hand, equilibrium models are characterized by the imposition of equilibrium conditions between the yields of several maturities for the interest rate. Using the structure of affine models, this class is obtained on the modeling of the instantaneous forward rate, in which the rates of other maturities can be derived by assuming that the risk premium is given by an affine function [5]. The third class of models, statistical models, is obtained as statistical representations of the evolution of the term structure of interest rates without impositions of no-arbitrage or equilibrium conditions.

Statistical models comprise principal components, factor or latent variable and interpolation models. Among these approaches, latent variable models, especially Nelson-Siegel model [18] and its variations, are the most popular and used by fixed income managers and central banks. The attractiveness of factor models in Nelson-Siegel class is due to their parsimony, good empirical performance and capability to capture most of the evolution of the term structure of interest rates with the use of three factors [3].

Despite several works have been devoted to term structure models, out-of-sample forecasting performance is mainly not focused [16]. [9] investigated the relationship between forward and future spot rates as the first study regarding predictability questions. More recently, [7] stated that affine models produce poor US yields forecasts. Otherwise, Diebold and Li [6] (henceforth DL) proposed a dynamic variant of the NS components framework to model the yield curve and provide a consistent way of generating forecasts of the yield curve. Considering the US fixed income market, the authors show that their model outperforms traditional benchmarks such as the random walk model. However, this model does not outperform random walk for short-term horizon forecasts (one-month ahead).

Equilibrium correction models were also suggested by [2] and [13], which outperforms the DL model and random walk at the short-term. [11] and [16] are examples which used Bayesian Markov Chain Monte Carlo to the estimation of the DL model. On the other hand, [22] applied state space models to perform term structure forecasting. However, these frameworks outperform other competitors on the out-of-sample forecast accuracy especially for short-term horizons.

Recently, [15] suggested the use of evolving fuzzy models for term structure of interest rates forecasting. Using a Nelson-Siegel framework, the authors made predictions of yield curve latent factors using an evolving fuzzy functional approach, based on the concept of density/potential as unsupervised clustering algorithm. The results indicate that the adaptive approach outperforms the DL model for both short- and long-term horizons in terms of curve fitting.

In this paper, we suggest the use of evolving participatory learning fuzzy modeling to the issue of term structure forecasting using US Treasury market data from January 1985 through December 2000 and the results are compared with the DL framework. The concept of evolving fuzzy systems introduces the idea of gradual self-organization and parameter learning in fuzzy rule-based models. Evolving fuzzy systems use data streams to continuously adapt the structure and functionality of fuzzy rule-based models.

The evolving fuzzy participatory learning (ePL) modeling was suggested in [14]. The approach joins the concept of participatory learning (PL) [21] with the evolving fuzzy modeling idea. In evolving systems the PL concept is viewed as an unsupervised clustering algorithm [20] and is a natural candidate to find rule base structures in dynamic environments.

Moreover, other contribution of this paper consists in the inclusion of a time-varying volatility scheme in the modeling of yield curve latent factors. This issue is particularly important because the assumption of constant interest rate volatility in these models often has remarkable practical impli-

---

1 A review on interest rate modeling can be found in [10].
2 Using a broad class of nonlinear methods, [17] suggested that error-correction models are able to capture the dynamics of short-term interest rates, outperforming linear and nonlinear alternatives.
cations for risk management policies, which can be too simple, neglecting the risk of a time-varying volatility structure [3]. Latent factors volatility were performed using a generalized autoregressive conditional heteroscedasticity (GARCH) model [8], which is a convenient framework for dealing with time-dependent volatility in financial data. Finally, the volatility was included in the ePL model in order to provide latent factors forecasts.

This paper is organized as follows. Section 2 describes the basic structure of Nelson-Siegel models including the presence of time-varying volatility. Section 3 presents the evolving participatory learning fuzzy model applied. Section 4 provides the computational results and the comparison between models. Finally, Section 5 concludes the paper and suggests issues for future works.

2 Yield Curve Modeling and Forecasting

Motivated by the fact that a yield curve is essentially humped, Nelson-Siegel [18] expressed the interest rate curve at time \( t \), denoted by \( y_t(\tau) \), as a function of maturities \( \tau \) according to the following expression:

\[
y_t(\tau) = \beta_{1,t} + \beta_{2,t} \left( \frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} \right) + \beta_{3,t} \left( \frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau} \right)
\]

where \( \beta_{1,t}, \beta_{2,t}, \beta_{3,t} \) and \( \lambda_t \) are the parameters.

The parameters and its associate weights determine the shape of the yield curve. Parameter \( \lambda_t \), regarded as fixed in DL [6], governs the exponential decay rate, and small (large) values of \( \lambda_t \) are associated with a smooth (fast) decay, and provide longer (shorter) maturities with a better fit [3]. The weight related to \( \beta_{1,t} \), factor is 1 (constant) for all maturities, thus the parameter \( \beta_{1,t} \) represents the long-term factor. On the other hand, the weight of the second component, \( \frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} \), begins at 1 and converges to zero monotonically, and \( \beta_{2,t} \) is considered as the short-term factor, since influences strongly short-term interest rates. Finally, the weight of the third component, \( \frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau} \), is a concave function, which begins at zero for maturity zero, increases, and then converges monotonically to zero in longer maturities. The component \( \beta_{3,t} \) is associated with medium-term interest rates, called also as the medium-term factor. The three factor loadings are shown in Figure 1.

![Image of Nelson-Siegel factors loadings](image)

**Fig. 1.** Factors loadings of Nelson-Siegel curve with fix \( \lambda_t = 0.0609 \).

Nelson-Siegel term structure representation is able to provide a good fit to the cross section of yields at a given point in time, and this is a key reason for its popularity and empirical appeal with financial market practitioners. At each time period, there are interest rates \( y_t(\tau) \) for several maturities \( \tau \). Therefore, the equation can be estimated at each time period, obtaining time series for parameters


\[ \beta_{i,t} = \zeta_i \beta_{i,t-1} + \varepsilon_{i,t}, \quad i = 1, 2, 3 \]  

(2)

where \( \zeta \) and \( \varepsilon \) are the parameters and \( \varepsilon \) is a white noise i.i.d. process.

To estimate the model according to this framework, a two-step procedure could be applied as in [6]. In the first step, fixing the parameter \( \lambda_t = \lambda \) to a pre-specified value, one estimates the parameters \( \beta_t \), for all \( t \), by solving the following optimization problem:

\[
\min_{\beta_{i,t}} \sum_{t=1}^{N} \left( y_t(\tau) - \hat{y}_t(\tau) \right)^2
\]

(3)

subject to:

\[
\beta_{1,t} > 0 \\
\beta_{1,t} + \beta_{2,t} > 0
\]

(4)

where \( y_t(\tau) \) and \( \hat{y}_t(\tau) \) are the actual and fitted spot interest rates, respectively, and \( N \) is the sample size.

Estimated \( \beta_t \) parameters are obtained by ordinary least squares (OLS). Thus, in the second step, three time series of \( \beta_{1,t}, \beta_{2,t}, \beta_{3,t} \) are modeled by a first-order autoregressive process as in Equation (2) in order to provide its forecasts.

One must be note that this approach assumes a homoskedastic structure, i.e. constant volatility, for all latent factors. In order to overcome this limitation, this paper includes time-varying volatility for yield curve factors forecasting. Modeling volatility directly in terms of factors is a more flexible approach and allows capturing the uncertainty over the level, slope, and curvature of the interest rate. Thus, the so-called level volatility represents the volatility of the general interest rate level, while slope volatility captures the uncertainty over the spread between long- and short-term rates. In turn, curvature volatility is associated with the risk of changes in the curvature of the term structure [3]. Therefore, in this paper, the yield curve factors assume the following system of equations:

\[
\beta_{i,t} = \zeta_i \beta_{i,t-1} + \varepsilon_{i,t} \sigma_{i,t-1}, \quad i = 1, 2, 3
\]

(5)

\[
\sigma_{i,t-1}^2 = \phi_{i,0} + \phi_{i,1} \kappa_{i,t-1} + \phi_{i,2} \sigma_{i,t-2}^2
\]

(6)

where \( \phi_{i,j} \), for \( i = 1, 2, 3 \) and \( j = 0, 1, 2 \), are the volatility process parameters, \( \sigma_t \) is the factors volatility at \( t \) and \( \kappa_t \) corresponds to factors returns.

Equation (6) indicates that the yield curve factors follow a GARCH(1,1) process [8]. The choice of this model is justified by several factors. There is empirical evidence indicating that these models are better at capturing stylized facts about financial series and at forecasting than other volatility models, including term structure forecasting [13].

Therefore, in this paper, the two-step procedure was applied as in DL [6]. However, in the second step, instead of model the factors as a first-order autoregressive process, they were estimated with the evolving participatory learning model (ePL), described in the next section. Moreover, as inputs in ePL, factors volatility were also included, obtained by a GARCH(1,1) process.

---

3 Diebold-Li [6] turned their attention to Nelson-Siegel model and reinterpreted it as a three-factor statistical model to describe the interest rate curve over time. The \( \beta_{1,t}, \beta_{2,t}, \) and \( \beta_{3,t} \) factors are interpreted as level, slope, and curvature, respectively.

4 The order of GARCH model was selected according to the Bayesian Information Criteria (BIC) [19].
3 Evolving Participatory Learning Fuzzy Models

Evolving functional fuzzy participatory learning (ePL) modeling adopts the same philosophy as the classical evolving Takagi-Sugeno (eTS) methodology [1]. After the initialization phase, data processing is performed at each step to verify if a new cluster must be created, if an old cluster should be modified to account for the new data, or if redundant clusters must be eliminated, in an on-line mode. Cluster centers are the focal point of the rules and each rule corresponds to a cluster. Parameters of the consequents are computed using the local recursive least squares algorithm. In this paper we assume, without loss of generality, linear consequent functions as in [14].

ePL models assume fuzzy rule-based systems whose rules are endowed with local models forming their consequents, commonly referred to as fuzzy functional models. Here we consider Takagi-Sugeno (TS) type of fuzzy model with linear functions as rule consequents. These models is a set of fuzzy rules of the following form:

\[ R^i : \text{IF } x_1 \text{ is } \Gamma^i_1 \text{ AND } x_2 \text{ is } \Gamma^i_2 \text{ AND } \ldots \text{ AND } x_n \text{ is } \Gamma^i_n \text{ THEN } y^i = \gamma^i_0 + \sum_{j=1}^{n} \gamma^i_j x_j \]  

where \( R^i \) is the \( i^{th} \) fuzzy rule (\( i = 1, 2, \ldots, R \)), \( x_j \in \mathbb{R} \) is the input data (\( j = 1, 2, \ldots, n \)), \( \Gamma^i_j \) is the fuzzy set associated with the \( j^{th} \) input variable of the \( i^{th} \) fuzzy rule, \( y^i \in \mathbb{R}^n \) is the linear output of the \( i^{th} \) rule, and \( \gamma^i_j \) are the parameters of the consequent.

In a TS model, the fuzzy regions are parameterized and each region is associated with a linear subsystem. Hence, the nonlinear system forms a collection of loosely (fuzzily) coupled (blended) multiple linear models. The contribution of a local linear model to the overall output is proportional to the degree of firing of each rule. Antecedent fuzzy sets are described with Gaussian membership functions, which ensures greatest possible generalization of the description [1]:

\[ \mu^i(x_j) = e^{-\frac{(x_j - v^i_j)^2}{2r^2}} \]

where \( \mu^i(x_j) \) is the membership degree of the current input \( x_j \) in \( \Gamma^i_j \), \( v^i_j \) is the respective cluster center or focal point, and \( r \) defines the radius of the antecedent, the zone of influence of the \( i^{th} \) model.

The activation degree of particular fuzzy rule is, choosing the algebraic product t-norm:

\[ \omega^i(x_k) = \prod_{j=1}^{n} \mu^i_j(x_k) = \mu^i_1(x_k) \times \ldots \times \mu^i_n(x_k) \]

The TS model output at step \( k \) is the weighted average of the individual rule contributions:

\[ y = \sum_{i=1}^{R} \pi_i y_i, \quad \pi_i = \frac{\omega_i}{\sum_{l=1}^{R} \omega_l} \]

where \( \pi_i \) is the normalized firing level of the \( i^{th} \) rule, and \( R \) is the number of fuzzy rules.

As stated in [1], identification of a TS model requires two sub-tasks: i) learning the antecedent part of the model using a fuzzy clustering algorithm, and ii) learning the parameters of the linear consequents. To the problem of this paper we focus on the evolving fuzzy participatory learning algorithm for antecedent learning, and on the recursive least squares algorithm to estimate the consequent parameters.

3.1 Participatory Learning Clustering

Evolving fuzzy participatory learning modeling adopts the same philosophy as eTS. After the initialization phase, data processing is performed at each step to verify if a new cluster must be created, if an old cluster should be modified to account for the new data, or if redundant clusters must be eliminated. Cluster centers are the focal point of the rules. Each rule corresponds to a cluster. The main difference between ePL and eTS concerns the procedure to update the rule base structure. Differently from eTS, ePL uses a fuzzy similarity measure to determine the proximity between new data and the existing rule base structure. The rule base structure is isomorphic to the cluster structure because each rule is associated with a cluster. Participatory learning assumes that model learning depends on what the system already knows about the model. Therefore, in ePL, the current model...
is part of the evolving process itself and influences the way in which new observations are used for self-organization. An essential property of participatory learning is that the impact of new data in causing self-organization or model revision depends on its compatibility with the current rule base structure, or equivalently, on its compatibility with the current cluster structure [14].

In online mode, the training data are collected continuously, rather than being a fixed set. Let $v^i_k$ be a variable that encodes the $i^{th}$ ($i = 1, \ldots, R_k$) cluster center at the $k^{th}$ step. The aim of the participatory mechanism is to learn the value of $v^i_k$, using a stream of data $x_k$. In other words, each $x_k$, $k = 1, 2, \ldots$, is used as a vehicle to learn about $v^i_k$. We say that the learning process is participatory if the contribution of each data $x_k$ to the learning process depends upon its acceptance by the current estimate of $v^i_k$ as being a valid. Implicit in this idea is that, to be useful and to contribute to the learning of $v^i_k$, observations $x_k$ must somehow be compatible with current estimates of $v^i_k$.

In ePL, the object of learning are cluster structures. Cluster structures are defined by cluster centers (or prototypes). More formally, given an initial cluster structure, a set of vectors $v^i_k$, $i = 1, \ldots, R_k$, is updated using a compatibility measure, $\rho^i_k \in [0, 1]$ and an arousal index, $a^i_k \in [0, 1]$. While $\rho^i_k$ measures how much a data point is compatible with the current cluster structure, the arousal index $a^i_k$ acts as a critic to remind when current cluster structure should be revised in front of new information contained in data.

Due to its unsupervised, self-organizing nature, the PL clustering procedure may create a new cluster or modify the existing ones at each step. If the arousal index is greater than a threshold value $\vartheta \in [0, 1]$, then a new cluster is created. Otherwise, the $i^{th}$ cluster center, the one most compatible with $x_k$, is adjusted as follows:

$$v^i_{k+1} = v^i_k + G^i_k(x_k - v^i_k)$$

(11)

where

$$G^i_k = \alpha \rho^i_k$$

(12)

$\alpha \in [0, 1]$ is the learning rate, and

$$\rho^i_k = 1 - \frac{||x_k - v^i_k||}{n}$$

(13)

with $|| \cdot ||$ a norm, $n$ the dimension of input space, and

$$i = \arg \max_j \{\rho^j_k\}$$

(14)

Notice that the $i^{th}$ cluster center is a convex combination of the new data sample $x^k$ and the closest cluster center.

Similarly as (11), the arousal index $a^i_k$ is updated as follows:

$$a_{k+1}^i = a_k^i + \eta(1 - \rho_{k+1}^i - a_k^i)$$

(15)

The value of $\eta \in [0, 1]$ controls the rate of change of arousal: the closer $\eta$ is to one, the faster the system is to sense compatibility variations.

The way in which ePL considers the arousal mechanism is to incorporate the arousal index (15) into (12), that is, we assume

$$G^i_k = \alpha(\rho^i_k)^{1-a_k^i}$$

(16)

When $a_k^i = 0$, we have $G^i_k = \alpha \rho^i_k$ which is the PL procedure with no arousal. If the arousal index increases, the similarity measure has a reduced effect. The arousal index can be interpreted as the complement of the confidence we have in the truth of the current belief, the rule base structure. The arousal mechanism monitors the performance of the system by observing the compatibility of the current model with the observations. Therefore learning is dynamic in the sense that (11) can be viewed as a belief revision strategy whose effective learning rate (16) depends on the compatibility between new data, the current cluster structure, and on model confidence as well.

Notice that the learning rate is modulated by compatibility. In conventional learning models, there are no participatory considerations and the learning rate is usually set small to avoid undesirable oscillations due to spurious values of data that are far from cluster centers. Small values of learning rate while protecting against the influence of noisy data, slow down learning. Participatory learning allows the use of higher values of the learning rate and the compatibility index acts to lower the effective learning rate when large deviations occur. On the contrary, when the compatibility is large, it increases the effective rate, which means speeding up the learning process.
Whenever a cluster center is updated or a new cluster added, the PL fuzzy clustering procedure should verify if redundant clusters are created. This is because updating a cluster center using (11) may push a given center closer to another one and a redundant cluster may be formed. Thus, a mechanism to exclude redundancy is needed. One mechanism is to verify if similar outputs due to distinct rules are produced. In PL clustering, a cluster center is declared redundant whenever its similarity with another center is greater than or equal to a threshold value \( \delta \). If this is the case, then we can either maintain the original cluster center or replace it by the average between the new data and the current cluster center. Similarly as in (13), the compatibility index among cluster centers is computed as follows:

\[
\rho_{v_i,k} = 1 - \frac{1}{n} \sum_{j=1}^{n} |v_i^k - v_j^k| \tag{17}
\]

Therefore, if

\[
\rho_{v_i,k} \geq \delta \tag{18}
\]

then the cluster \( i \) is declared redundant.

Detailed guidelines to choose appropriate values of \( \alpha, \eta, \delta \) and \( \vartheta \) are given in [14], where it is shown that they should be chosen such that:

\[
0 < \vartheta \leq \eta \leq 1 - \delta \leq 1
\]

where

\[
\vartheta \leq \eta \text{ and } \delta \leq 1 - \vartheta
\]

### 3.2 Parameter Identification

Estimation of the parameters of the consequent linear models can be formulated as a least squared problem [1]. Equation (10) can be transformed into a vector form as follows:

\[
y = \Lambda^T \Phi \tag{19}
\]

where \( y \) is the output of the ePL, \( \Lambda = [\pi_1 x_e^T, \pi_2 x_e^T, \ldots, \pi_n x_e^T]^T \) denotes the fuzzily weighted extended inputs vector, \( x_e = [1 \ x_e^T]^T \) is the expanded data vector, \( \Phi = [\Psi_1^T, \Psi_2^T, \ldots, \Psi_R^T]^T \) represents the vector of parameters of the rule base, and \( \Psi_i = [\gamma_{i0} \gamma_{i1} \ldots \gamma_{in}]^T \) is the vector of consequent part parameters (parameters of the \( i \)th linear local subsystem).

Since the actual target output is provided at each step, the parameters of the consequents can be updated using recursive least squares algorithm RLS [4] considering locally or globally optimization. In this paper we use the locally optimal error criterion:

\[
\min E_k^i = \min \sum_{l=1}^{k} \pi_i(x_l) \left(y_l - \pi_i x_e^T \Psi_i^l\right)^2 \tag{20}
\]

There are not only fuzzily coupled linear subsystems and streaming data, but also structure evolution, therefore the optimal update of the parameters of the \( i \)th local subsystems is given by:

\[
\Psi_{k+1}^i = \Psi_k^i + \Sigma_k^i x_e^T \pi_i (y_k - (x_e^T \Psi_k^i), \Psi_1^i = 0 \tag{21}
\]

\[
\Sigma_{k+1}^i = \Sigma_k^i - \pi_i^i \Sigma_k^i x_e^T (x_e^T \Psi_k^i) \Sigma_k^i \Sigma_k^i = \Omega I_{(n+1) \times (n+1)} \tag{22}
\]

where \( I \) is a \((n+1) \times (n+1)\) identity matrix, \( \Omega \) denotes a large number, usually \( \Omega = 1000 \), and \( \Sigma \) a dispersion matrix.

When a new fuzzy rule is added, a new dispersion matrix is initiated \( \Sigma_k^{R+1} = I \Omega \). Parameters of the new rules are approximated from the parameters of the existing \( R \) fuzzy rules as:

\[
\Psi_k^{R+1} = \sum_{i=1}^{R} \pi_i \Psi_k^{i} \tag{23}
\]
Otherwise, parameters of all other rules are inherited from the previous time step, while the dispersion matrices are updated independently. Finally, with the consequent parameters, the prediction of the output is obtained by Equation (10).

The use of the recursive least squares algorithm depends on the initial values of the parameters $\Psi_0$, and of the initial values of the entries of the dispersion matrix $\Sigma_0$. These initial values are chosen based on: i) existence of previous knowledge about the system, exploring a database to find an initial rule base and, consequently, $\Psi_0$ and $\Sigma_0$; ii) a useful technique when no previous information is available is to choose large values for the entries of matrix as described above. In this paper, we use the first option, that is, we use a database to choose the initial rule base and its parameters.

4 Computational Analysis

4.1 Data

Each database entry is the end-of-month price quotes (bid-ask average) of the US Treasuries, from January 1985 through December 2000, and maturities of 1, 3, 6, 12, 24, 36, 60, 84 and 120 months. Figure 2 shows the 3-dimensional plot of the yield curve data. Notice the large temporal variation of the yield values, and the smoother, low variations of the slope and curvature.

4.2 Methodology

In this work, yield curve forecasts were obtained by the two-step procedure using autoregressive process and the ePL model. The first step was the same for both approaches, i.e. time series of Nelson-Siegel parameters ($\beta_t$) were obtained by solving the optimization problem in (3) via ordinary least squares, by fixing $\lambda_t = \lambda = 0.0609$ as suggested by Diebold-Li [6]. For this purpose, data base was split into training and validation sets. The training sample covers the period from January 1985 through December 1999. The remaining data were used for forecasting evaluation. Thus, $\beta_t$ time series were constructed using the in-sample (training) data.

Therefore, as the benchmark, the second step was first performed by modeling the term structure factors using a first-order autoregressive process as in Equation (2). However, as suggested in this paper, factors $\beta_t$ were also modeled with the ePL model suggested, described in the previous section. Moreover, one also includes a time-varying volatility structure, described by a GARCH (1,1) process as in Equation (6). Thus, ePL inputs are: first previous factors values and its volatility; on the other hand, the output is the level, slope and curvature parameters at $h$-steps ahead.

The database is available in http://www.federalreserve.gov/econresdata/releases/statisticsdata.htm.

To estimate the GARCH process for each yield curve factor the MatLab function garchfit was employed.
One defines forecast errors at $t+h$ as $\hat{y}_{t+h}(\tau) - y_{t+h}(\tau)$ where $y_{t+h}$ and $\hat{y}_{t+h}$ correspond to the actual and predicted future yields and $\tau$ is the maturity. Out-of-sample forecasts were compared in terms of mean squared forecasts errors (MSFE):

$$\text{MSFE}(\tau) = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_{t+h,i}(\tau) - y_{t+h,i}(\tau))^2$$

(24)

where $N$ is the sample size.

In this paper, for comparison purposes, predictions were performed for 1 (short-term), 3 (medium-term) and 6 (long-term) months forecasting horizons.

### 4.3 Results and Discussion

The procedure related to the issue of term structure of interest rates forecasting comprises two steps. In the first one, factors time-series were constructed by solving the optimization problem in (3) with OLS procedure, by fixing $\lambda$ at 0.0609 as suggested by Diebold-Li.

Figure 3 presents the in-sample fit using the Nelson-Siegel model. The residual plot indicates a good fit. As pointed out by [6], regardless of the term structure estimation method used, there is a persistent discrepancy between actual bond prices and prices estimated from term structure models. Presumably these discrepancies arise from persistent tax and/or liquidity effects.

Therefore, the second step procedure consists in modeling $\beta_t$ time-series in order to perform forecasts. As the benchmark, a first-order autoregressive process was applied for each yield curve factor. Alternatively, this work also suggests the modeling of these factors using the evolving participatory learning fuzzy model (ePL), described in Section 3, including a volatility term obtained by a GARCH(1,1) process. The ePL model adopts the following control parameters: $r = 0.05$, $\eta = \vartheta = 0.16$, $\alpha = 0.01$, $\delta = 0.84$ and $\Omega = 1000$.

In Figure 4, the estimated parameters $\{\hat{\beta}_{1,t}, \hat{\beta}_{2,t}, \hat{\beta}_{3,t}\}$ were plotted, respectively, along with the empirical level, slope and curvature, defined by the ones provided in the optimization procedure (first step forecasting procedure). The figure illustrates the high performance of the ePL model to describe the term structure factors. Moreover, one may note that ePL provides a more accurate fit than the benchmark, mainly for the level ($\beta_1$) and curvature ($\beta_3$) parameters. This fact can be explained by the inclusion of a volatility structure in the ePL approach.

Figure 5 shows term structure parameters volatility, obtained by the GARCH(1,1) model, which evidences a time-varying volatility behavior for level, slope and curvature yield curve factors. Thus, this inclusion translates in a more robust model, providing a better fit for the parameters evolution.
Fig. 4. Actual and estimated yield curve factors.
In Figure 6 one presents actual yields and the fitted yield curves for some selected dates considering the ePL model suggested. Clearly the ePL model is capable of replicating a variety of yield curve shapes by the three factors modeling: upward sloping, downward sloping, humped, and inverted humped. However, one could see some difficulties at some dates, especially when yields are dispersed, with multiple interior minima and maxima.
Therefore, forecasts were performed for 1, 3 and 6 months ahead. Table 1 indicates the MSFE for some selected maturities by taking into account the whole redemption curve for out-of-sample set. The ePL model achieves a better performance than the benchmark (DL model) for both sort- and long-term maturities, which confirms the high adherence of the evolving fuzzy model to deal with the Nelson-Siegel factors forecasting.

Table 1. Out-of-sample MSFE for $h$-month ahead horizons.

<table>
<thead>
<tr>
<th>Maturity (in months)</th>
<th>1-month ahead</th>
<th>3-month ahead</th>
<th>6-month ahead</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Models 1 3 6 12 60 120</td>
<td>Models 1 3 6 12 60 120</td>
<td>Models 1 3 6 12 60 120</td>
</tr>
<tr>
<td>DL</td>
<td>0.189 0.133 0.117 0.123 0.110 0.094</td>
<td>0.190 0.135 0.120 0.126 0.120 0.103</td>
<td>0.096 0.211 0.119 0.112 0.136 0.289</td>
</tr>
<tr>
<td>ePL</td>
<td>0.118 0.119 0.097 0.030 0.044 0.016</td>
<td>0.144 0.111 0.108 0.110 0.110 0.089</td>
<td>0.066 0.182 0.090 0.079 0.101 0.177</td>
</tr>
</tbody>
</table>
Finally, the ePL adaptive dynamic were evaluated by the evolution of the number of fuzzy rules, as shown in Figure 7. Variation of the number of rules is similar for the three yield curve parameters, and it shows the continuous model structure adaptation through changes in the rule base structure. It is interesting to note that the number of rules varies according to the respective behavior of level, slope and curvature Nelson-Siegel parameters, evidencing the adaptability mechanism in the ePL model.

![Figure 7. Evolution of the number of rules for ePL model.](image-url)
5 Conclusion

The construction of yield curve forecasts is essential for risk and portfolio managers, financial institutions, policy making and for all market practitioners in general. This work proposed the performance evaluation of evolving participatory learning fuzzy modeling for the term structure of government bond yields forecasting with a time-varying volatility scheme using data from the US fixed income market. In this approach, the model includes mechanisms to deal with the yield curve factors adaptive dynamic and to describe the uncertainty related to these parameters by taking into account its volatility structure. Results indicate that the ePL model with time-varying volatility provides more accurate forecasts than autoregressive models for both short- and long-term maturities. Future researches shall include the evaluation of the ePL model for a more flexible yield curve structure by considering no-arbitrage conditions. Moreover, some other evolving fuzzy approaches with non-constant volatility could be a promising tool for interest rate curve prediction.
Acknowledgments

The authors would like to thank the Brazilian Ministry of Education (CAPES), the Brazilian National Research Council (CNPq) grant 30543/2009-4 and 306343/2011-8, and the Foundation for Research Support of the State of São Paulo (FAPESP) for their support.

References