Forecasting Exchange Rates with Fuzzy Granular Evolving Modeling for Trading Strategies

Leandro Maciel[1] Fernando Gomide[1] Rosangela Ballini[2]

[1]School of Electrical and Computer Engineering, University of Campinas, Brazil
Emails: {maciel,gomide}@dca.fee.unicamp.br
[2]Institute of Economics, University of Campinas, Brazil
Email: ballini@eco.unicamp.br

Abstract

This paper addresses a fuzzy set based evolving modeling (FBeM) approach and the task of forecasting exchange rates in order to perform trading strategies. FBeM is a granular computing technique that uses fuzzy information granules to model nonstationary functions providing functional and linguistic approximations. As an application, this work considers the BRL/USD exchange rate market data for the period from January 2000 to October 2012. Comparisons in terms of goodness of fit and based on trading performance indicators includes the granular model against a Multi-Layer Perceptron (MLP), an autoregressive moving average (ARMA), a naïve strategy and some state of the art evolving fuzzy systems. Computational results suggest that the FBeM model statistically outperforms the alternative approaches.

Keywords: Granular Computing, Evolving Systems, Exchange Rates, Trading, Forecasting.

1. Introduction

Forecasting exchange rate trend is a challenging task due to its high nonlinear, time-varying and noisy environment. Several factors such as economic activity, trader's expectations, political events, and inflation influence the dynamic of currency markets. Therefore, new tools and techniques are needed in dealing with exchange rate prediction, since the traditional econometric approach based on autoregressive moving average (ARMA) models has been criticized by their limitations to reply the dynamics inherent to exchange rate time series [1, 2]. Moreover, the development of a timely and accurate trading decision-making tool is the key for traders to make profits [3].

Due to their success in financial forecasting, artificial neural networks (ANNs) have been adopted as an alternative method in the prediction of exchange rates [4, 5, 6]. Recently, the literature focused the application of different kinds of structures of ANNs such as Multi-layer Perceptron (MLP), Recurrent, Radial Basis and Psi Sigma Networks to the problem of foreign exchange rate forecasting and trading [7, 8, 9, 10]. In the trade strategy based on forecasts, the ANN models achieve higher returns (profits) when compared to traditional ARMA models.

However, the main drawback with ANNs is their black-box nature, revealing difficulty in interpreting the results by not providing an insight into the dynamic of the interactions between the technical indicators and the currency market fluctuations [11]. Another challenge in ANN approaches is the learning process for models with memory like recurrent networks, essential for time-dependent environments, which requires intensive training and suffers low convergence, revealing a key problem in dynamic markets [12].

To overcome these limitations, the use of fuzzy-based models provides both predictive accuracy and interpretation to deal with complex real-world problems. Unlike traditional sets, fuzzy sets allow for the concept of partial membership. This enables discrimination between elements that are relevant to the phenomenon of interest and those of borderline importance that involve imprecision and uncertainty. Moreover, information granules can be processed using fuzzy logic whereby each linguistic term describes a fuzzy set, providing an interpretable system [13]. Recent studies have been revealing the high potential of fuzzy models to deal with financial time series, including exchange rate forecasting [14, 3, 15, 16]. In general, they considered evolving fuzzy systems, constructed by fuzzy rules in a form of "If-Then" statements, comprising an adaptive structure based on streaming data [17, 18].

In this paper, we investigate the financial forecasting performance of fuzzy granular evolving modeling by benchmarking their trading results over the Brazilian BRL/USD daily exchange rate with a MLP, an ARMA model, a naïve trading strategy and some state of the art evolving fuzzy modeling approaches. Evolving granular systems lay emphasis on granulated views of detailed data and computing with granules coarser than the data in order to simplify complex real-world problems and provide low cost solutions [19].

Evolving granular modeling comes as an approach to capture the essence of streaming data and as a framework to extrapolate spatio-temporal correlations from lower-level raw data, providing a more abstract human-like representation of them [20, 21, 22]. Here, we analyze a fuzzy set based evolving modeling (FBeM) approach, proposed by Leite et al. [19], which uses fuzzy-type information granules to construct granular maps. FBeM models combine functional and linguistic fuzzy systems to provide singular and granular approximation of nonstationary functions. The linguistic component provides model interpretation and encloses possible model outputs. The functional component is derived from input data and real-valued local functions, producing more accurate approximators.

These advantages overcome the limitations derived from ANNs models and also comprise an autonomous mechanism without requiring expert knowledge or even specific assumptions about the data, which is essential for dynamic environments such as exchange rates. Hence, due to these motivations, the contribution of this paper concerns the application of the FBeM approach to the task of exchange rate forecasting in order to perform trading strategies.

After this introduction, the paper is organized as follows. In Section 2, the FBeM approach is presented. Section 3 describes the data used for this research, the competing forecasting models and the metrics for comparisons in terms of goodness of fit and trading performance. The empirical results are given in Section 4 while Section 5 provides some concluding remarks and suggests issues for further investigation.

2. Fuzzy set based evolving modeling

The fuzzy set based evolving modeling (FBeM) produces higher level information granules in an evolving modeling structure based on streaming data and recursive learning algorithm. A FBeM model is comprised by a set of If-Then fuzzy rules extracted from the data, managing information granules and gradually evolved over time. Each granule correspond to a rule. Rules R^i governing information granules γ^i are of the type:

IF $(x_1$ is $A_1^i)$ AND ... AND $(x_j$ is $A_j^i)$ AND ... AND $(x_n$ is $A_n^i)$ THEN $(y$ is $B^i)$ AND $y = p^i(x_j \forall j)$

where x_j and y are variables of the data stream $(x, y)^{[t]}$, $j = 1, 2, \ldots, n$, $t = 1, \ldots$; A_j^i and B^i are membership functions; p^i are approximation polynomials. One must note that the consequent part $(y$ is $B^i)$ comprises the output linguistic part, providing interpretability of the results, and the term $y = p^i(x_j \forall j)$ is the functional output, offering precision[1]. The collection of rules R^i, $i = 1, 2, \ldots, c$,

forms the rule base. FBeM takes advantage of both, linguistic and functional systems, within a single modeling framework [21].

In this paper, scattering-type mechanism was used for granulation of data into fuzzy objects[2]. Specifically, FBeM uses Gaussian fuzzy subsets $A_j^i = \mathcal{G}(\mu_j^i, \sigma_j^i)$, where μ_j^i is the modal value and σ_j^i the spread. The Gaussian representation are useful since their necessary parameters are set straightforward from a data stream, the infinite support does not ignore the data and comprises a smoothness and continuously differentiable surface.

Rule's consequent combines functional and linguistic fuzzy information. The functional part of the consequent, p^i, concerns singular local functions whereas the linguistic part describes information granules B^i along the domain of output variable. This paper assumes affine local functions for functional part of the consequent:

$$p^i = a_0^i + \sum_{j=1}^{n} a_j^i x_i \qquad (1)$$

where a_0^i and a_j^i are the corresponding coefficients.

Since Gaussian representation allows all granules to overlap, each rule in FBeM contributes to the system output. The model singular output is obtained as the weighted mean value:

$$p = \frac{\sum_{i=1}^{c} \min(A_1^i, \ldots, A_n^i) p^i}{\sum_{i=1}^{c} \min(A_1^i, \ldots, A_n^i)} \qquad (2)$$

Consequents of rules, B^i, also assume Gaussian fuzzy subsets $B^i = \mathcal{G}(\mu_y^i, \sigma_y^i)$ to assemble granular objects in the output space. The granular output provides useful information than the specific numerical output p, enriching decision making. Moreover, information granules tend to reflect the essence of the structure of the underling data stream and emphasize the interpretability of the result [19].

There are two main sub-tasks related to FBeM identification: learning rules antecedents and consequents. These sub-tasks are described as follows.

2.1. Learning antecedents

FBeM learns online from a stream of instances $(x, y)^{[t]}$, where $y^{[t]}$ is known given $x^{[t]}$ or will be known at some latter step. The recursive algorithm associated decide when and how to proceed structural and parametric adaptation of models. The learning procedure to evolve fuzzy granular systems FBeM decides on how to accommodate new information given an instance $(x, y)^{[t]}$. When a new instance does not fit current knowledge, the model creates a new information granule and a rule governing the granule. Otherwise, if a new instance

---

[1] This paper considers the FBeM model on its single out-

put form, the multivariate structure extension is straightforward. See [19].

[2] Granulation of data into fuzzy objects can be based on grid, tree or scatter partitioning.

fits current knowledge, the procedure adapts existing granules and rules. Furthermore, the quotient structure of fuzzy rules may be optimized, coarsed or refined, according to inter-granules relationships.

FBeM model can start from scratch. Rules are constructed and evolve when data are input. When a new granule γ^c is created, a rule R^{c+1} is added to the current rule base $R = \{R^1, \ldots, R^i, \ldots, R^c\}$. If a instance $x^{[t]}$ does not activate the current collection of rules, a new granule is created. It is assumed that $x^{[t]}$ brings new information about the system.

As a new instance $x^{[t]}$ arrives, a new granule is created if the following condition holds:

$$\min(A_1^i, \ldots, A_n^i) \leq \rho \ \forall \ i, \qquad (3)$$

where $\rho \in [0, 1]$ is a threshold value determining the granularity of FBeM models. If ρ is set to zero, the system is structurally stable and unable to capture eventual concept shift. On the other hand, if ρ equals 1, FBeM creates a rule for each new instance, which is not practical [21].

Since a new granule γ^{c+1} is created, membership functions A_j^{c+1} and B^{c+1} are initiated as follows:

$$
\begin{aligned}
\mu_j^{c+1} &= x_j^{[t]} \\
\mu_y^{c+1} &= y^{[t]} \\
\sigma_j^{c+1} &= \sigma_y^{c+1} = \frac{1}{2\pi}
\end{aligned}
\qquad (4)
$$

The coefficients of local-valued polynomials p^{c+1} are set to:

$$a_0^{c+1} = y^{[t]}, \quad a_j^{c+1} = 0, \quad j \neq 0 \qquad (5)$$

As stated by [19], this initial parametrization gives preference to design of granules balanced along all dimensions rather than granules with unbalanced geometry, following the principle of balanced information granularity [23].

If a new instance $x^{[t]}$ does not satisfy the condition in (3), the rule base need adaptation, i.e.: i) expand or contract A_j^i and B^i to accommodate new data; ii) move granules γ^i toward denser regions of data over the input and output domains; iii) adjust coefficients of local approximation function p^i.

A rule R^i is adapted if a data instance $x^{[t]}$ holds:

$$\min(A_1^i, \ldots, A_n^i) > \rho \qquad (6)$$

The most active rule for $x^{[t]}$ is then chosen for adaptation by updating the model value and the spread of membership functions A_j^i recursively:

$$\mu_j^i(\text{new}) = \frac{(\omega^i - 1)\mu_j^i(\text{old}) + x_j}{\omega^i} \qquad (7)$$

$$
\begin{aligned}
\sigma_j^i(\text{new}) &= \frac{(\omega^i - 1)}{\omega^i}\sigma_j^i(\text{old}) \\
&\quad + \frac{1}{(\omega^i - 1)}(x_j - \mu_j^i(\text{new}))^2 \quad (8)
\end{aligned}
$$

where ω^i is the number of times that the granule γ^i has been activated by the data stream.

Adaptation of fuzzy sets of rule consequents B^i uses output data $y^{[t]}$. Moreover, polynomial coefficients a_0^i and a_j^i are updated according to recursive weighted least squares (wRLS).

As mentioned before, the granularity threshold ρ impacts model accuracy and transparency. To alleviate guesses on how fast and how often the structure of the data changes, the parameter ρ takes values in the interval according to prediction errors comprising a recursive adaptation framework. Let E be the squared error between predictions $p(x^{[t]})$ and actual value $y^{[t]}$, then

$$E = (y^{[t]} - p(x^{[t]}))^2 \qquad (9)$$

Hence, ρ learning values for itself from

$$\rho(\text{new}) = \rho(\text{old}) + \eta(E_D - E) \qquad (10)$$

where η is a learning rate and E_D the desired prediction error. We also assume $\rho^{[0]} = 0.5$ as default initial value.

As the expansion of granules occurs, relationships among pairs of granules may be strong enough to justify a granule structure. In order to avoid redundant granules and to construct a more concise structure, the quotient structure was coarsened based on a distance measure between granules. Considering Gaussian membership functions as internal representatives of granules, the distance between two different granules, γ^{i_1} and γ^{i_2}, is computed as follows:

$$
\begin{aligned}
D(\gamma^{i_1}, \gamma^{i_2}) &= \frac{1}{n}\sum_{j=1}^{n} ||\mu_j^{i_1} - \mu_j^{i_2}||^2 + \sigma_j^{i_1} \\
&\quad + \sigma_j^{i_2} - 2\sqrt{\sigma_j^{i_1}\sigma_j^{i_2}} \qquad (11)
\end{aligned}
$$

Choosing the pair of granules with the lowest entry of $D(\cdot)$, if this entry is less or equal than a threshold Δ, a new granule γ^i, coarsening of γ^{i_1} and γ^{i_2}, is represented by Gaussian membership function with modal value

$$\mu_j^i = \frac{\frac{\sigma_j^{i_1}}{\sigma_j^{i_2}}\mu_j^{i_1} + \frac{\sigma_j^{i_2}}{\sigma_j^{i_1}}\mu_j^{i_2}}{\frac{\sigma_j^{i_1}}{\sigma_j^{i_2}} + \frac{\sigma_j^{i_2}}{\sigma_j^{i_1}}}, \quad j = 1, \ldots, n \qquad (12)$$

and spread

$$\sigma_j^i = \sigma_j^{i_1} + \sigma_j^{i_2}, \quad j = 1, \ldots, n \qquad (13)$$

Therefore, the coefficients of the new local polynomial are:

$$a_j^i = \frac{1}{2}(a_j^{i_1} + a_j^{i_2}), \quad j = 0, 1, \ldots, n \qquad (14)$$

The mechanism of coarsening granules reduces the number of rules and redundancy. Moreover,

the FBeM deletes granules when they become inactive during a number of processing steps, h_r. This mechanism ensures to keep the rule base updated. Once the structure of the FBeM is defined and established, the problem of parameter identification becomes important, as described below.

2.2. Learning consequents

Estimation of the parameters of the consequent linear models can be formulated as a least squared problem [24]. The output can be represented by:

$$y = \Lambda^T \Phi \qquad (15)$$

where $\Lambda = \left[\lambda_1 x_e^T, \lambda_2 x_e^T, \ldots, \lambda_n x_e^T \right]^T$ denotes the fuzzily weighted extended inputs vector, $x_e = \left[1 \ x^T \right]^T$ is the expanded input data vector, $\lambda_i = \frac{\min(A_1^i, \ldots, A_n^i)}{\sum_{j=1}^{c} \min(A_1^j, \ldots, A_n^j)}$ is the normalized firing level of the rule i, $\Phi = \left[\Psi_1^T, \Psi_2^T, \ldots, \Psi_c^T \right]^T$ represents the vector of parameters of the rule base, and $\Psi_i = [a_0^i, a_1^i, \ldots, a_n^i]$ is the vector of consequent parameters.

Since the actual target output is provided at each step, the parameters of the consequents can be updated using weighted recursive least squares algorithm wRLS [24] considering locally or globally optimization. This paper uses the locally optimal error criterion that minimizes the following error criterion:

$$E_L^i = \sum_{h=1}^{T} \lambda^i(x^{[t]}) \left(y^{[t]} - (x_e^{[t]})^T \Psi^i \right)^2 \qquad (16)$$

where T is the sample size.

There are not only fuzzily coupled linear subsystems and streaming data, but also structure evolution, therefore the optimal update of the parameters of the local subsystem i is given by [24]:

$$\Sigma^i(\text{new}) = \Sigma^i(\text{old}) - \frac{\lambda^i \Sigma^i(\text{old}) x_e^{[t]} (x_e^{[t]})^T \Sigma^i(\text{old})}{1 + \lambda^i (x_e^{[t]})^T \Sigma^i(\text{old}) x_e^{[t]}} \qquad (17)$$

$$\begin{aligned} \Psi^i(\text{new}) &= \Psi^i(\text{old}) + \Sigma^i(\text{new}) x_e^{[t]} \lambda^i \\ &\cdot \left(y^{[t]} - (x_e^{[t]})^T \Psi^i(\text{old}) \right) \end{aligned} \qquad (18)$$

where $\Sigma_1^i = \Omega I$, I is a $(n+1) \times (n+1)$ identity matrix, Ω denotes a large number, usually $\Omega = 1000$, and Σ a dispersion matrix.

2.3. FBeM algorithm

The detailed steps of the FBeM model are as follows. All the steps of the algorithm are non-iterative. The model can develop/evolve an existing model when the data pattern changes, and by being recursive it means that it is computationally efficient.

BEGIN
Set parameters E_D, Δ, h_r, η, $c = 0$;
Read $(x, h)^{[t]}$, $t = 1$;
Create granule γ^{c+1};
For $t = 2, \ldots$ do
    Read $(x, h)^{[t]}$;
    If $\min(A_1^i, \ldots, A_n^i) \leq \rho, \ \ \forall \ i$
      Create granule γ^{c+1};
    Else
      Adapt the most active granule γ^i;
    If $\exists \ D(\gamma^{i_1}, \gamma^{i_2}) \leq \Delta$
      Form γ^i by coarsening γ^{i_1} and γ^{i_2};
    Remove inactive granules;
    Update granularity;
    Adapt parameters a_j^i using wRLS;
END

---

3. Methodology

3.1. Data

In this paper, we considered the daily closing BRL/USD (Brazilian Real/Dollar) exchange rate from 3 January 2000 trough 26 October 2012. The data period is partitioned into two sub-periods namely training and validation data sets. The training data set covers the period from 3 January 2000 to 31 December 2009. The remaining data comprises the validation set (out-of-sample). This mechanism is only necessary for ARMA and MLP models, since evolving fuzzy systems can start from scratch avoiding training steps.

The BRL/USD data was transformed into stationary daily series of returns using:

$$y_t = \left(\frac{P_t}{P_{t-1}} \right) - 1 \qquad (19)$$

where y_t is the rate of return and P_t is the price level at time t.

3.2. Forecasting models

We compared the performance of the FBeM model with MLP, ARMA models, a naïve strategy and some state of the art evolving fuzzy models for forecasting and trading the BRL/USD exchange rate.

The naïve strategy simply takes the most recent period change as the best prediction of the future change, i.e., a simple random walk model. The model is defined by:

$$y_{t+1} = y_t + \epsilon \qquad (20)$$

where y_t is the actual rate of return at period t and $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$ the error term with variance σ_ϵ^2.

Autoregressive moving average models (ARMA) assume that a time series depends on its previous

values and on previous residual values, taking the form:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + \varepsilon_t$$
$$-\beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \ldots - \beta_q \varepsilon_{t-q} \qquad (21)$$

where α_i and β_j, for $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, q$, are regression coefficients and the residual weights respectively, ε_t the residual at period t such that $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, and σ_ε^2 the model variance.

The MLP model presents nonlinear neurons for hidden layers with tangent hyperbolic transfer functions and a linear output. It is a feed forward network trained with back-propagation algorithm using descent gradient. Finally the empirical analysis include some state of the art evolving fuzzy models such as eTS [25], xTS [17], ePL [26] and eTS+ [24].

3.3. Performance measurement

Comparison of BRL/USD forecasts was done assuming one-step ahead forecast and using as error criterion the mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean squared error (RMSE):

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^{T} |y_t - \hat{y}_t| \qquad (22)$$

$$\text{MAPE} = \frac{1}{T} \sum_{t=1}^{T} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \qquad (23)$$

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{y}_t)^2} \qquad (24)$$

where y_t and $\hat{y}_t$ are the actual and predicted returns at time t, and T is the sample size.

Moreover, we applied the Diebold-Mariano [27] statistic test which evaluates the null hypothesis of equal predictive accuracy. The alternative hypothesis states that one model can be considered as a better predictor than some other model in terms of goodness of fit. Moreover, it is a widely approach used for forecasting model validation concerning applications related to economic and finance. If T is the sample size and e_i^1, e_i^2 $(i = 1, 2, \ldots, T)$ are the forecast errors of the two competing forecasts, then the loss functions are estimated as:

$$L_1^{\text{MSE}}(e_i^1) = (e_i^1)^2, \; L_2^{\text{MSE}}(e_i^2) = (e_i^2)^2 \qquad (25)$$

$$L_1^{\text{MAE}}(e_i^1) = |e_i^1|, \; L_2^{\text{MAE}}(e_i^2) = |e_i^2| \qquad (26)$$

The Diebold-Mariano statistic is based on the loss differentials:

$$d_i^{\text{MSE}} = L_1^{\text{MSE}}(e_i^1) - L_2^{\text{MSE}}(e_i^2) \qquad (27)$$

$$d_i^{\text{MAE}} = L_1^{\text{MAE}}(e_i^1) - L_2^{\text{MAE}}(e_i^2) \qquad (28)$$

The null hypotheses testes based on the d_i^{MSE} and d_i^{MAE} are: i) H_0: $E(d_i^{\text{MSE}}) = 0$ against the alternative H_1: $E(d_i^{\text{MSE}}) \neq 0$; ii) H_0: $E(d_i^{\text{MAE}}) = 0$ against the alternative H_1: $E(d_i^{\text{MAE}}) \neq 0$. The Diebold-Mariano test statistic s is estimated as:

$$s = \frac{\bar{d}_i}{\sqrt{\hat{V}(\bar{d}_i)}} \sim N(0, 1) \qquad (29)$$

where $\hat{V}(\bar{d}_i) = T^{-1} \left[\hat{\kappa}_0 + 2 \sum_{k=1}^{T-1} \hat{\kappa}_k \right]$ and $\kappa_k = T^{-1} \sum_{i=k+1}^{T} (d_i - \bar{d}_i)(d_{i-k} - \bar{d}_i)$.

Besides the model's fitness one also evaluates their trading performance according to annualized returns (R^A), cumulative returns (R^C), annualized volatility (σ^A) and maximum drawdown (MD), described as follows:

$$R^A = 252 \cdot \frac{1}{T} \sum_{t=1}^{T} y_t \qquad (30)$$

$$R^C = \sum_{t=1}^{T} y_t \qquad (31)$$

$$\sigma^A = \sqrt{252} \cdot \sqrt{\frac{1}{T-1} \sum_{t=1}^{T} (y_t - \bar{y})^2} \qquad (32)$$

$$\text{MD} = \min_{i=1,\ldots,t; t=1,\ldots,T} \left(\sum_{j=i}^{t} y_j \right) \qquad (33)$$

where $\bar{y}_t$ is the mean value of returns.

4. Empirical results

We examine the BRL/USD exchange rate over the period from January 2000 to October 2012. The BRL/USD series was transformed into a stationary daily series of rate returns. Figure 1 shows the dataset in level and in terms of rate returns.
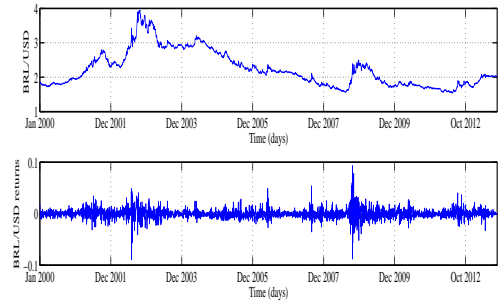

Figure 1: BRL/USD daily closing prices (top panel) and BRL/USD daily returns (bottom panel).

The summary statistics of the BRL/USD returns, shown in Table 1, reveal positive skewness and high kurtosis. The Jarque-Bera [28] statistic confirms

that the BRL/USD return series is non-normal at the 95% level.

Statistic	Returns
Mean	0.00008
Median	-0.00025
Maximum	0.09330
Minimum	-0.08937
Std. Dev.	0.00984
Skewness	0.28204
Kurtosis	11.61198
Jarque-Bera	18077.14
Probability	0.0010

Table 1: Descriptive statistics for BRL/USD daily returns.

The inputs selection for FBeM, MLP and the evolving models was conducted based on the return series partial autocorrelation function (PACF) analysis. Simulations considering the significant lags in the PACF and models' performance indicated as inputs a set of autoregressive terms of the BRL/USD returns: y_t and y_{t-2}. Therefore, predictions were performed for one-step-ahead horizon, i.e., y_{t+1}.

To estimate the ARMA model, the autoregressive and moving average parameters range in the following intervals, respectively, $1 \leq p \leq 15$ and $1 \leq q \leq 15$. An ARMA(2,1) model was chosen as the best for the in-sample estimation and can be specified as follows:

$$y_t = -.1838y_{t-1} + .0762y_{t-2} + .2473\varepsilon_{t-1} \quad (34)$$

After an extensive experimentation, the MLP model with best performance presents the characteristics shown in Table 2.

Parameters	MLP
Learning algorithm	Gradient descent
Learning rate	0.001
Momentum	0.002
Iteration steps	1500
Weights initialization	$\mathcal{N}(0,0.01)$
Input nodes	4
Hidden layers	1
Hidden nodes	5
Output node	1
Transfer function	Sigmoid

Table 2: MLP characteristics.

FBeM control parameters were chosen according simulation experiments to produce better performance in terms of fitness. The model adopted the following values: $\eta = 0.01$, $\Delta = 0.08$, $h_r = 70$ and $E_D = 0.007$. Finally, Table 3 indicates the evolving fuzzy models control parameters setting[3].

To evaluate models forecasts, the MAE, the MAPE and the RMSE statistics were computed for

---

[3]The xTS model does not require control parameters.

Parameters	eTS	ePL	eTS+
Clusters spread	0.05	0.04	-
Learning rate	-	0.01	-
Utility threshold	-	-	0.10
Spread update	-	-	0.50
Arousal update	-	0.16	-
Redundancy threshold	-	0.84	-
Arousal threshold	-	0.16	-

Table 3: Evolving fuzzy models control parameters.

the out-of-sample subset. Table 4 presents the error measures for the evaluated models, the lower the output, the better the forecasting accuracy of the model concerned. The FBeM model outperformed all remaining models in terms MAE, MAPE and RMSE measures. Naïve, ARMA and MPL strategies showed the worst results, and the evolving models, eTS, xTS, ePL and eTS+, performed similar.

Models	MAE	MAPE	RMSE
naïve	0.0231	7,19%	0.0513
ARMA	0.0214	6.96%	0.0496
MLP	0.0208	7.00%	0.0351
eTS	0.0192	6.52%	0.0322
xTS	0.0200	7.02%	0.0341
ePL	0.0196	6.77%	0.0331
eTS+	0.0192	6.41%	0.0283
FBeM	0.0134	4.39%	0.0231

Table 4: Summary of out-of-sample statistical performance.

Besides the models goodness of fit, this paper also compares them in statistical terms. Hence, the Diebold-Mariano [27] statistic for predictive accuracy is also computed for both MSE and MAE loss functions, as described in Section 3. The results of the Diebold-Mariano statistic, comparing the FBeM model with each other method are summarized in Table 5. The Diebold-Mariano test were performed to couples of forecasts, i.e., FBeM vs. another forecasting model. The null hypothesis of equal predictive accuracy is rejected for all comparisons and for both loss functions at 5% confidence level, since $|s_{\text{MSE}}| > 1.96$ and $|s_{\text{MAE}}| > 1.96$. Therefore, the statistical superiority of the FBeM model forecasts is confirmed as for both loss functions the realizations of the statistic are negative.

In order to evaluate the results in a real-world application, trading strategies were applied for all forecasting methods. The trading strategy is to go or stay "long" when the forecasting return is above zero and go or stay "short" when the forecast return is below zero. The "long" and "short" BRL/USD positions are defined as buying and selling Brazilian Reals at the current price respectively. Table 6 reports the out-of-sample trading performance of all models. The FBeM model performs better than all other models in terms of trading performance. It

Models	s_{MSE}	s_{MAE}
naïve	-3.6588	-7.9584
ARMA	-3.6333	-7.7268
MLP	-3.2199	-7.8831
eTS	-2.5466	-7.2442
xTS	-2.8825	-7.4257
ePL	-2.6541	-7.1430
eTS+	-2.1209	-7.2755

Table 5: Diebold-Mariano statistic for MSE and MAE loss functions.

presents a slightly higher annualized (R^A) and cumulative (R^C) returns than the other models. On the other hand, naïve and ARMA strategies provide lower results. The annualized volatility (σ^A) and maximum drawdown (MD) were similar for all models.

Models	R^A	R^C	σ^A	MD
naïve	12.86%	32.14%	12.54%	-5.32%
ARMA	25.34%	55.87%	11.13%	-4.57%
MLP	44.50%	112.84%	10.94%	-4.13%
eTS	60.85%	130.69%	10.50%	-4.23%
xTS	66.47%	144.95%	9.25%	-4.12%
ePL	63.12%	127.84%	9.66%	-4.96%
eTS+	86.09%	163.04%	10.01%	-3.72%
FBeM	94.74%	171.41%	9.43%	-3.57%

Table 6: Summary of trading performance results.

The models were also compared in terms of computational complexity concerning the final number of rules/nodes and processing (cpu) time, which results are shown in Table 7. One may see that the models show similar performance, except the MLP model which presents a high processing time.

Models	# rules/nodes	Time (in sec.)
ARMA	-	4.510
MLP	5	110.9
eTS	4	5.247
xTS	4	5.112
ePL	3	6.714
eTS+	2	5.322
FBeM	2	6.217

Table 7: Models computational complexity.

Besides the high potential of the FBeM in the problem of trading and forecasting the BRL/USD exchange rate, the model recursively adapts its structure to match current knowledge about the system, as in all evolving techniques. Figure 2 shows the number of rules evolution of the FBeM model. We can note that the number of rules increases, reaching the maximum value equal 6, in the periods of BRL/USD returns high volatility (Figure 2).

One must note that the FBeM model provides functional and linguistic output terms, which may be employed to improve decision making process by
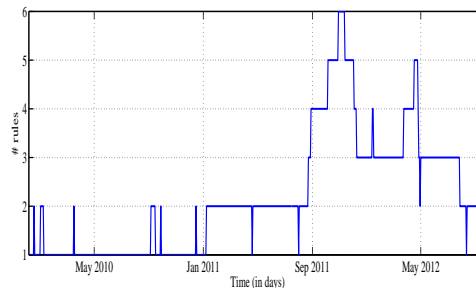

Figure 2: Evolution of the number of rules for BRL/USD using the FBeM model.

labeling the output granules in order to give some insight about the system or even about the trading rules. Rules of particular interest can be displayed at any time. The rule base at the final process is:

R^1: IF y_t is $\mathcal{G}(.5835, .2486)$ AND y_{t-2} is $\mathcal{G}(.5892, .2537)$ THEN y_{t+1} is $\mathcal{G}(.6011, .2549)$ AND $p^1 = .0154 - .1037y_t + 1.0864y_{t-2}$.

R^2: IF y_t is $\mathcal{G}(.9299, .1592)$ AND y_{t-2} is $\mathcal{G}(.9045, .1344)$ THEN y_{t+1} is $\mathcal{G}(.8978, .1106)$ AND $p^2 = .5264 + .1031y_t + .3048y_{t-2}$.

From this rule base, the consequents output membership functions may be interpreted as "low returns" and "high returns", for example, improving the model interpretability.

5. Conclusion

This paper applies a fuzzy set based evolving model (FBeM) to a one-day-ahead forecasting and trading task of the Brazilian Real/Dollar (BRL/USD) exchange rate. The FBeM model recursively granulates data instances to output singular and linguistic granular approximations. The trading performance of the granular method was compared with a Multi-Layer Perceptron (MLP) neural network, an autoregressive moving average model (ARMA), a naïve strategy and some state of the art evolving fuzzy models like eTS, xTS, ePL and eTS+. Using the BRL/USD exchange rate series from January 2000 to October 2012, models comparisons were conducted based on error measures, statistical tests and trading performance indicators. Results have shown the effectiveness of the FBeM approach by its best trading measures in terms of annualized and cumulative returns. Besides the statistical superiority and the high potential to describe the BRL/USD nonlinear behaviour, FBeM allows the construction of decision making rules according to its linguistic outputs, improving the results in terms of interpretability. Future studies shall concern the application of a time-varying leverage in the trading strategies as well as the inclusion of transaction costs in the model.

References

[1] A. Timmermann and C. Granger. Efficient market hypothesis and forecasting. *International Journal of Forecasting*, 20:15–27, 2004.

[2] H. Ni and H. Yin. Exchange rate prediction using hybrid neural networks and trading indicators. *Neurocomputing*, 72:2815–2823, 2009.

[3] R. K. Lai, C. Fan, W. Huang, and P. Chang. Evolving and clustering fuzzy decision tree for financial time series data forecasting. *Expert Systems with Applications*, 36:3761–3773, 2009.

[4] J. Kaashoek and D. Van Dijk. Neural network pruning applied to real exchange rate analysis. *Journal of Forecasting*, 21:559–577, 2002.

[5] A. Chen and M. T. Leung. Regression neural network for error correction in foreign exchange forecasting and trading. *Computers & Operations Research*, 31:1049–1068, 2004.

[6] B. Majhi, M. Rout, R. Majhi, G. Panda, and P. J. Fleming. New robust forecasting models for echange rates prediction. *Expert Systems with Applications*, 39:12658–12670, 2012.

[7] C. L. Dunis, J. Laws, and G. Sermpinis. Higher order and recurrent neural architectures for trading the EUR/USD exchange rate. *Quantitative Finance*, 11:615–629, 2011.

[8] W. L. Tung and C. Quek. Financial volatility trading using a self-organising neural-fuzzy semantic network and option straddle-based approach. *Expert Systems with Applications*, 38:4668–4688, 2011.

[9] G. Sermpinis, J. Laws, A. Karathanasopoulos, and C. L. Dunis. Forecasting and trading the EUR/USD exchange rate with gene expression and psi sigma neural networks. *Expert Systems with Applications*, 39:8865–8877, 2012.

[10] G. Sermpinis, C. Dunis, J. Laws, and C. Stasinakis. Forecasting and trading the EUR/USD exchange rate with stochastic neural network combination and time-varying leverage. *Decision Support Systems*, 54:316–329, 2012.

[11] G. P. Zhang. Avoiding pitfalls in neural network research. *IEEE Transactions on Systems Man and Cybernetics, Part C*, 37:3–16, 2007.

[12] M. Lukoševičius and H. Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3:127–149, 2009.

[13] N. Gradojevic and R. Gençay. Fuzzy logic, trading uncertainty and technical trading. *Journal of Banking & Finance*, 37:578–586, 2013.

[14] Y. Zhang and X. Wan. Statistical fuzzy interval neural networks for currency exchange rate time series prediction. *Applied Soft Computing*, 7:1149–1156, 2007.

[15] Y. Leu, C. Lee, and Y. Jou. A distance-based fuzzy time series model for exchange rates forecasting. *Expert Systems with Applications*, 36:8107–8114, 2009.

[16] L. Maciel, A. Lemos, F. Gomide, and R. Ballini. Evolving fuzzy systems for pricing fixed income options. *Evolving Systems*, 3:5–18, 2012.

[17] P. Angelov and X. Zhou. Evolving fuzzy-rule-based classifiers from data streams. *IEEE Transactions on Fuzzy Systems*, 16:1462–1475, 2008.

[18] I. Luna and R. Ballini. Adaptive fuzzy systems for forecast financial time series volatility. *Journal of Intelligent & Fuzzy Systems*, 23:27–38, 2012.

[19] D. Leite, F. Gomide, R. Ballini, and P. Costa. Fuzzy granular evolving modeling for time series prediction. In *IEEE International Conference on Fuzzy Systems*, Taipei, Taiwan, June 27–30, 2011.

[20] W. Pedrycz. Granular computing – The emerging paradigm. *Journal of Uncertainty Systems*, 1:38–61, 2007.

[21] D. Leite, R. Ballini, P. Costa, and F. Gomide. Evolving fuzzy granular modeling from nonstationary fuzzy data streams. *Evolving Systems*, 3:65–79, 2012.

[22] D. Leite, P. Costa, and F. Gomide. Evolving granular neural networks from fuzzy data streams. *Neural Networks*, 38:1–16, 2013.

[23] A. Bargiela and W. Pedrycz. *Granular Computing: An Introduction*. Kluwer Academic Publishers- Boston, Dordrecht, London, 2003.

[24] P. Angelov. *Evolving Systems: Methodology and Applications*, chapter Evolving Takagi-Sugeno fuzzy systems from streaming data (eTS+), pages 21–50. John Wiley & Sons, Inc., Hoboken, NJ, USA, 2010.

[25] P. Angelov and D. Filev. An approach to online identification of Takagi-Sugeno fuzzy models. *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, 4:484–498, 2004.

[26] E. Lima, M. Hell, R. Ballini, and F. Gomide. *Evolving Systems: Methodology and Applications*, chapter Evolving Fuzzy Modeling Using Participatory Learning, pages 67–86. John Wiley & Sons, Inc., Hoboken, NJ, USA, 2010.

[27] F. X. Diebold and R. S. Mariano. Comparing predictive accuracy. *Journal of Business and Economics Statistics*, 13:253–263, 1995.

[28] C. M. Jarque and A. K. Bera. A test for normality of observations and regression residuals. *International Statistical Review*, 55:163–172, 1987.