Forecasting EUR/RON Exchange Rate using a Hybrid Model

Alexandru TINDECHE
The Bucharest University of Economic Studies
Doctoral School of Economic Cybernetics and Statistics
Romania

Prof. Stelian STANCU, Ph.D
The Bucharest University of Economic Studies
Informatics and Economic Cybernetics Department
Romania
Centre for Industrial and Services Economics, Romanian Academy

Andreea-Cristina PETRICĂ
The Bucharest University of Economic Studies
Doctoral School of Economic Cybernetics and Statistics
Romania

Abstract

The paper illustrates the importance of Delay Embedding Technique (DET) - a chaotic model applied to the daily EUR/RON exchange rate. It assumes that the time series is a projection of a dynamical system that belongs to a larger space. Thus, we reconstruct the phase space and choose some random dimension and time delay. The results of applying DET instead of selecting random input dimension and time delay indicates that a better forecast of EUR/RON exchange rate is obtained and the dynamic of the system is preserved. Artificial Neuro-Fuzzy Inference Systems ($ANFIS$) receives as inputs the delayed vectors resulted from the chaotic model, these two mechanisms outline a hybrid model. $ANFIS$ proved to be the most efficient tool that can handle large data sets in order to provide qualitative predictions, so the combination between DET, auto-mutual information function and $ANFIS$ could represent an outstanding prediction tool for complex time series.

Keywords: Artificial Neural Network; Forecasting; Delay Embedding Technique; Artificial Neuro-Fuzzy Inference Systems; Chaos Theory; Mutual Information.

1. Literature Review- Brief History of Chaos

James Clerk Maxwell studied the sensitivity to initial conditions of dynamical systems and assumed that this could only happen in systems with a large number of variables. On the other hand, Henri Poincaré paid attention to the analysis regarding the sensitivity to initial conditions and unpredictability of chaotic systems with a small number of variables. In 1963 Lorenz, E. [15] came with the seminal paper on sensitivity to initial conditions, later named the “butterfly effect”, but it did not receive much interest until his talk at the American Association for the Advancement of Science meeting in 1972. The first application field of the chaos theory was in meteorology when Lorenz simulated the atmosphere with a set of 12 differential equations. Later, as the neural networks field emerged, they represented a good support to further develop of the chaos theory, also building powerful tools for predicting time series. Some of the most representative papers regarding the time series prediction using hybrid models (chaotic tools and neural networks) are presented below:

Atmaca, H., Cetişli, B. and Yavuz, H. S. (2001) [3] realized a comparative analysis between $ANFIS$ and ANN using fuel consumption (Mile per Gallon) support data based on which the main conclusions were outlined. $ANFIS$ Proved to be more efficient in terms of computational complexity problems, but it presents lack of qualitative prediction when it comes to small amounts of data. Also, regarding the training data, $ANFIS$ provides the result with minimum total error compared toANN. On the other hand, when the trained parameters are applied to the checking data, the total error of the ANN is smaller than the total error of the $ANFIS$.

Finnsson, I. (2005) [10] applied $ANFIS$ to a small sample (952 exchange rate points available), some time delays were chosen after trying different combinations, but when it comes to prediction - in this case $ANFIS$ encounters instability problems. The author claims that the $GARCH$ model is recommended in this case, providing very good prediction for the return series although it is difficult to map it back into the price series. Zhang, J., Dai, W., Fan, M., Chung, H. and Wei, Z. (2005) [26] proved that if we decide to use only Fuzzy Inference System to predict chaotic time series, the result will be weak but the prediction accuracy can increase significantly if we preprocess the time series data with delay embedding technique. Ma, H. and Han, C. (2006) [16] described a new algorithm for calculating the embedding dimension and the time delay of the reconstructed phase space which is based on the nonbias multiple autocorrelation and Gamma test, this combination also reduces the effect of noise which leads to a better highlight of the strange attractor.

In 2009 Ragulskis, M. and Lukoseviciute, K. (2009) [22] approached a new method of delay embedding by finding a set of non-uniform time lags. The set of non-uniform time lags results from maximizing a predefined objective function which depends on the magnitude of the attractor's spreading in the phase space. The identification of the optimal time lags also helps finding the embedding dimension of the reconstructed phase space. Some experiments using classical chaotic time series (Rossler) were conducted and used $ANFIS$ for prediction purposes. The input for the fuzzy inference system is represented by delayed vectors of non-uniform time lags previously optimized. The results of these experiments reveal significantly improved predictions compared to uniform delay embedding technique. Ciobanu, D. and Bar M. V. (2013) [7] used the Chaos Theory to predict the USD/EUR exchange rate: starting with the calculation of the embedding dimension and time delay, then using these values for k-Nearest Trajectories Algorithm, he concluded that if we have enough observations in order to have a good coverage of the attractor, the chaotic model holds. The Chaos Theory model is presented as a local model which is more appropriate for financial time series compared to Neural Networks models which are global function approximators that can lead to weaker predictions.

Behmanesh, M., Mohammadi, M. and Naeini, V. S. (2014) [4] proposed an improved version of $ANFIS$ used in chaotic time series prediction. This version consists of a new learning algorithm which is a combination between Least Squares Method (to update the consequent parameters) and Imperialist Competitive Algorithm (to update the previous parameters) in the iterative learning process. This new hybrid algorithm doesn't depend on the derivative of the error's surface while adapting the synaptic strengths, concluding that the network will not be trapped in local optima while trying to update the previous parameters. Several time series derived from classical Dynamical Systems are investigated: Mackey-Glass time series, Lorenz model and Rossler model and the results reveal an outstanding performance of the new hybrid algorithm. In 2015 Allen, D. E., McAleer, M., Peiris, S. and Singh, A. [2] modeled the return series for four parities including US Dollar.

They used 5 types of nonlinear regression models (Logistic Smooth Transition Model, Threshold Autoregressive Model, Additive Nonlinear Model, Smooth Transition Model and Nonlinear Autoregressive Model) and two types of Neural Networks: with linear and nonlinear activation function (depending on the case). The Neural Network with nonlinear activation function has proven to be the most suitable model for prediction exchange rate in terms of errors. Pedram, M. and Ebrahimi, M. (2015) [21] investigated the prediction of exchange rate data using artificial neural networks and concluded that the ANN performance indicates the advantage of estimating complex models even if we use a small set of data as an input, but we should take into account that data limitation often leads to a hard reach of a global minimum error, recommending Genetic Algorithms in order to fix this problem. The author also analyzed the sensitivity of the ANN to the input variables concluding that the Iran Consumer Price Index has the biggest impact in the trend of USD/IRR exchange rate.

2. *Preliminaries*

2.1. Takens Theorem

The deterministic part of a dynamical system relies on the concept of phase space, the set of all possible states of the system which is mathematically described by using the equations of motion or a collection of coordinates that give a complete description of the system. For chaotic dynamical systems, which are more common in nature, the easiest way to reconstruct the phase space is through attractors.

An attractor represents 'a set of states (points in the phase space), invariant under the dynamics, towards which neighboring states in a given basin of attraction asymptotically approach in the course of dynamic evolution'[1]. Usually, attractors are simple fixed points, but they can also take different geometrical shapes, moreover if they cannot be easily described as intersection of fundamental geometrical manifolds, then the attractors are called strange attractors.

In reality, a dynamical system issues multiple signals at different moments in time, generating a time series. Thus, the challenge is to reconstruct the phase space using the time series as a proxy.

Suppose we have a dynamical system represented by the states $x(t) \in \mathbb{R}^N$ that evolves through the following differential equation$\dot{x} = \psi(x)$, where $\psi: \mathbb{R}^N \to \mathbb{R}^N$ represents the vector field of the dynamical system.
Let $\mathcal{M}$ be a submanifold of $\mathbb{R}^N$. If we want to describe the system states, $x(t)$, at any given point in time, t, we define the following flow function:

$$\boldsymbol{G}: \boldsymbol{\mathcal{M}} \times \mathbb{R} \to \boldsymbol{\mathcal{M}}, \boldsymbol{G}(\boldsymbol{x}(\boldsymbol{t_0}), \boldsymbol{T}) = \boldsymbol{x}(\boldsymbol{t_0} + \boldsymbol{T}) \quad (2.1.1)$$

For the moment, we are interested in dynamical systems that are uniformly sampled in time (sampling time is denoted by T_s) and define the time T-map of the manifold $\mathcal{M}$ by:

$$\boldsymbol{G_T}: \boldsymbol{M} \to \boldsymbol{M}, \boldsymbol{G_T}\big(\boldsymbol{x}(\boldsymbol{t_0})\big) = \boldsymbol{x}(\boldsymbol{t_0} + \boldsymbol{T}) \quad (2.1.2)$$

and

$$\boldsymbol{G_T} \circ \dots \circ \boldsymbol{G_T}\big(\boldsymbol{x}(\boldsymbol{t_0})\big) = \boldsymbol{G_T^k}\big(\boldsymbol{x}(\boldsymbol{t_0})\big) \quad (2.1.3)$$

Due to the technological limitations we only get to see one-dimensional time series $y(t) = \phi\big(x(t)\big)$, where ϕ is a measurement function. The main question is: Can information about $x(t)$ be retained in this time series data? Takens confirmed this theory through the following immersion:

$$\begin{aligned}\boldsymbol{F}\big(\boldsymbol{x}(\boldsymbol{t})\big) &= \boldsymbol{F}_{(\boldsymbol{\phi}, \boldsymbol{G}_{-\boldsymbol{T_s}})}\big(\boldsymbol{x}(\boldsymbol{t})\big) = \big[\boldsymbol{\phi}\big(\boldsymbol{x}(\boldsymbol{t})\big), \boldsymbol{\phi} \circ \boldsymbol{G}_{-\boldsymbol{T_s}}\big(\boldsymbol{x}(\boldsymbol{t})\big), \dots, \boldsymbol{\phi} \circ \boldsymbol{G}_{-\boldsymbol{T_s}}^{\boldsymbol{M}-\boldsymbol{1}}\big(\boldsymbol{x}(\boldsymbol{t})\big)\big] \\ &= \big[\,\boldsymbol{y}(\boldsymbol{t}), \boldsymbol{y}(\boldsymbol{t} - \boldsymbol{T_s}), \dots, \boldsymbol{y}(\boldsymbol{t} - (\boldsymbol{M} - \boldsymbol{1})\boldsymbol{T_s})\big)\big]\end{aligned} \quad (2.1.4)$$

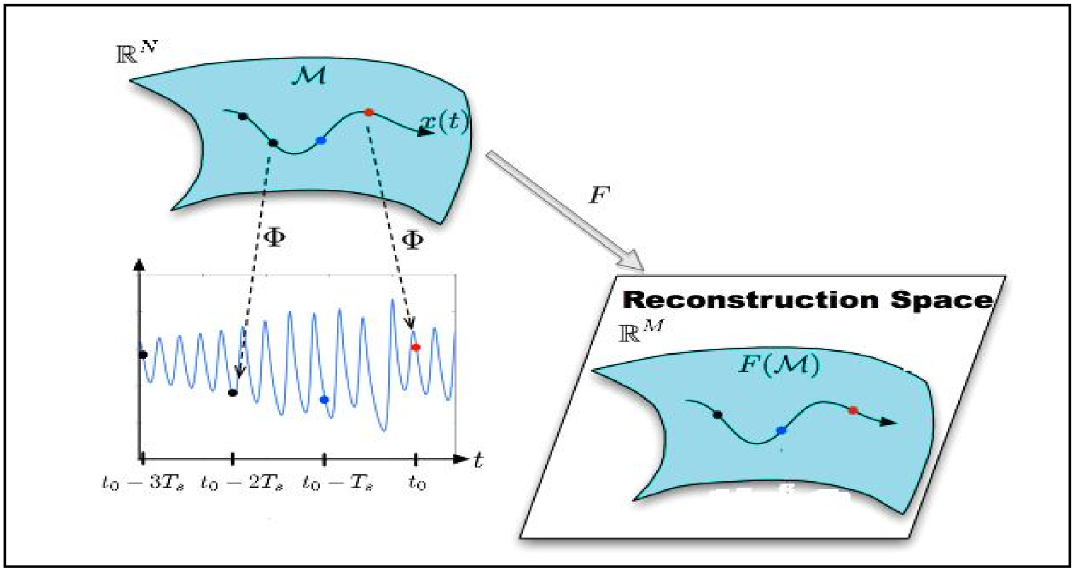
Where F is a mapping from the manifold $\mathcal{M} \subset \mathbb{R}^N$ to a reconstructed space $\mathbb{R}^M$ formed with time series measurements.

The idea of using time delay coordinates in order to represent a system state comes from the theory of ordinary differential equations where existence theorem says that there is a unique solution for $g(t, y(t), \dot{y}(t), \dots) = 0$, given all the initial conditions (current knowledge of the position and momentum suffices to uniquely determine the future dynamics). The same approach could be applied for delay embedding technique, approximating the derivatives by delay-coordinate terms as follows:

$$\boldsymbol{F}\big(\boldsymbol{x}(\boldsymbol{t})\big) = \left[\,\boldsymbol{y}(\boldsymbol{t}), \frac{\boldsymbol{y}(\boldsymbol{t}) - \boldsymbol{y}(\boldsymbol{t} - \boldsymbol{T_s})}{\boldsymbol{T_s}}, \frac{\boldsymbol{y}(\boldsymbol{t}) - \boldsymbol{2y}(\boldsymbol{t} - \boldsymbol{T_s}) + \boldsymbol{y}(\boldsymbol{t} - \boldsymbol{2T_s})}{\boldsymbol{T_s^2}}, \dots\right] \quad (2.1.5)$$

[1]http://mathworld.wolfram.com/Attractor.html

Figure 1: Takens theorem illustration



Source: https://cnx.org/contents/k57_M8Tw@2/Takens-Embedding-Theorem

2.2. Phase space reconstruction of a dynamical system

Considering Takens theorem, we can assume that a chaotic time series is actually a projection of the realizations of a dynamic system in a larger space. For the reconstruction of the phase space, we must determine the time delay (T_s) and the embedding dimension (M), meaning that from the initial time series $\{y(1), y(2), \ldots, y(N)\}$ we have to extract the following vectors (that belong to the phase space):

$$\boldsymbol{y_M(t) = [\,y(t), y(t-T_s), \ldots, y(t-(M-1)T_s)\,]\,, t = (M-1)T_s + 1, \ldots, N} \qquad \textbf{(2.2.1)}$$

It is important to select a suitable pair (T_s, M) when performing the delay embedding, since the embedding dimension and time delay are directly related to the characteristics of the strange attractors included in the phase space.

According to **[**16**]** there are two methods of choosing the pair (T_s, M):

1) The first one assumes that T_s and M are not correlated which means that they have to be selected independently. Takens proved that this method is ideal if the chaotic time series is infinite and has no noise which, in this case, we can estimate the embedding dimension using False Nearest Neighbors algorithm and the time delay using autocorrelation function, mutual information function, wavering product, average displacement (AD algorithm), etc.

2) The second method assumes that T_s and M are closely related due to the fact that real time series are not infinitely long and we can't avoid the noise. The dependence between the two variables is given by the time span, as follows:

$$\boldsymbol{t_w = (M-1)\,T_s} \qquad \textbf{(2.2.2)}$$

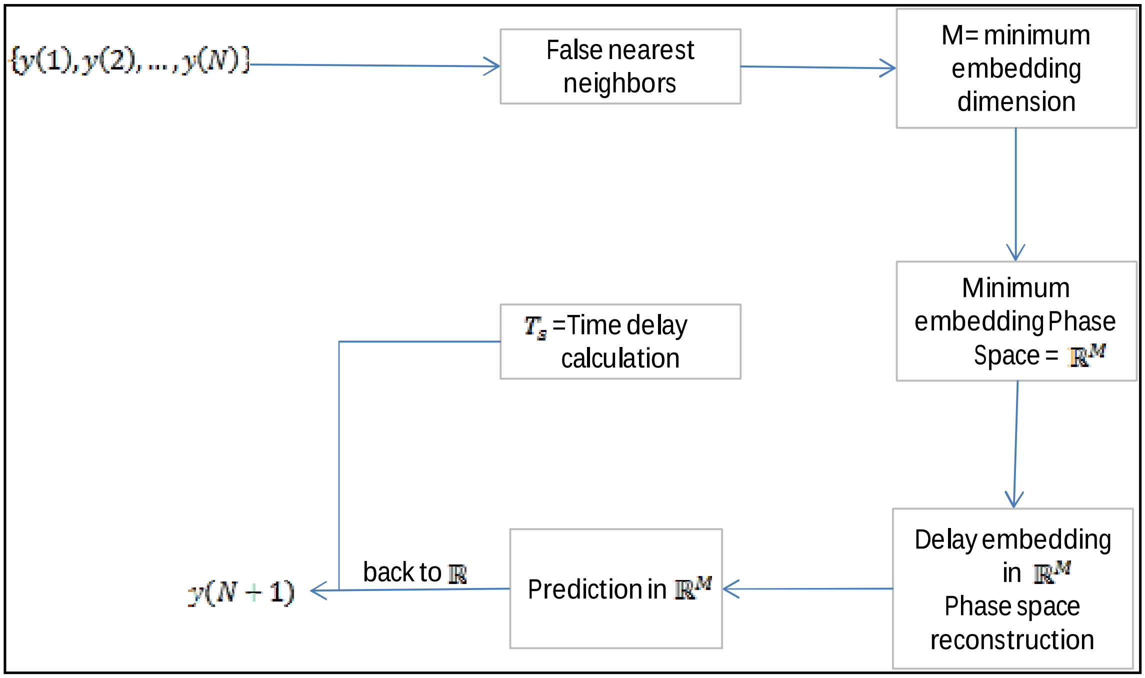
Note that experiments on chaotic time series proved that t_w is almost constant but an irrelevant relation between T_s and M will degenerate the equivalence between the original dynamic system and the reconstructed phase space. The second method is more practical than the first one due to the noise and data limitation. An idealistic approach (as the 1st method) could be applied if we have a consistent historical data, and eventually reduce the noise by applying spectral decomposition.

Theoretically, an embedding of the original space can be obtained if we take a sufficiently large M and any value of T_s. In reality, if M is too large - the noise can reduce the visibility of the attractor and it will harden the computational efficiency, but on the other hand if M is too small - the attractor will be folded. The choice of T_s is also important in order to establish data correlation in the delayed vectors.

If T_s is too large the elements of the delayed vectors will behave like uniformly random data (this is not the purpose of the embedding) and no information can be gained from the plot, on the other hand if T_s is too small, when plotted, all of the data stays near the line $y(t) = y(t + T_s)$. Therefore, optimal values of T_s and M have to be calculated in order to mimic the behavior of the original dynamic system.

In this paper, optimal values of T_s and M are selected using Average Mutual Information function and the False Nearest Neighbors, respectively (see Figure 2).

Figure 2: Prediction scheme in a chaotic system



Source: **Authors' calculations.**

3. Data And Methodology

3.1. EUR/RON exchange rate data

The paper aims to investigate the prediction efficiency and dynamics preserving of Chaotic Models in combination with $ANFIS$ using the EUR/RON exchange rate, daily quotations for the time period from January 03, 2011 to February 26, 2016, which involves investigating a time series of 1305 observations. The data were collected from the official website of the National Bank of Romania[2].

3.2. Selection of the embedding dimension

One of the most important features of the attractors is the neighbors acquired on their orbits. These neighbors provide information about how phase space neighborhoods evolve, in order to use them in the prediction of new points near or on the attractor. In an embedding dimension that is too small to unfold the attractor, not all the points that seem to be close to one another are actually neighbors due to the dynamics. If we are in M dimensions and $y_M^{(r)}(t)$ is the rth nearest neighbor of $y_M(t)$, then the square Euclidian distance between $y_M^r(t)$ and $y_M(t)$ is given by the following equation:

$$R_M^2(t, r) = \sum_{k=0}^{M-1} \left[y(t + k\,T_s) - y^{(r)}(t + k\,T_s) \right]^2 \qquad \textbf{(3.2.1)}$$

[2]http://bnr.ro/Cursul-de-schimb-524.aspx

Adding an extra dimension to the reconstructed phase space is equivalent to adding a new coordinate to each vector of $y_M(t)$ which is $y(t + M\,T_s)$, therefore the new square Euclidian distance becomes:

$$\boldsymbol{R^2_{M+1}(t,r) = R^2_M(t,r) + \left[y(t + M\,T_s) - y^{(r)}(t + M\,T_s)\right]^2} \quad \textbf{(3.2.2)}$$

Using this measure, we start to designate as a false neighbor any neighbor for which the following inequation holds:

$$\boldsymbol{\frac{R^2_{M+1}(t,r) - R^2_M(t,r)}{R^2_M(t,r)} = \frac{\left|y(t + M\,T_s) - y^{(r)}(t + M\,T_s)\right|}{R^2_M(t,r)} > R_{tol}} \quad \textbf{(3.2.3)}$$

Where R_{tol} is some predefined threshold.

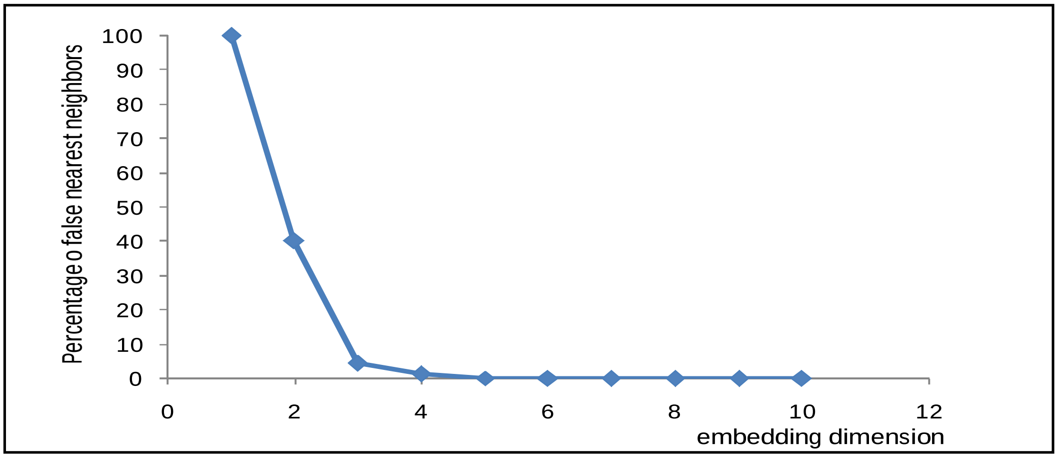
Kennel, M., et al. [13] proved that for $R_{tol} \geq 10$ the false neighbors are clearly identified, but also that $R_{tol} \geq 10$ is not a sufficient condition for determining the embedded dimension.Analyzing a white noise signal, if we are adding new data points to the signal, the embedding dimension (that dimension where the number of false nearest neighbors drops to zero) systematically increases, diverging to infinity for very large noise time series.The conclusion is that we need an extra measure (we can assimilate it to a convergence criterion) in the algorithm of false nearest neighbors:

$$\boldsymbol{\frac{R_{M+1}(t,r)}{R_A} > A_{tol}} \quad \textbf{(3.2.4)}$$

whereR_A is chosen to be $\sqrt{\frac{1}{M}\sum_{k=1}^{M}(y(k) - \bar{y})^2}$.

Using experimental data, Kennel, M., et al. [13] stated that $A_{tol} = 2$ is a good convergence criterion. Thus, it is recommended to select the embedding dimension that sets to zero the number of false nearest neighbors.
As can be seen from Figure 3, for our EUR/RON exchange rate data the number of false nearest neighbors drops to 0 when $M = 6$.

Figure 3: Number of false nearest neighbors depending on the embedding dimension



Source: Authors' calculations.

3.3. Selection of the time delay

The main difference between the mutual information function and the correlation function is that the first one measures the mutual dependence between two variables i.e. the amount of information about one random variable through another random variable, while the second function measures the linear dependence. Another important difference between these two functions is given by the applicability of mutual information also to symbolic sequences, while the correlation function is specific only for numerical sequences. Formally, the mutual information function can be defined as follows:

$$I(X,Y) = \sum_{i=1}^{N_x}\sum_{j=1}^{N_y} P_{XY}(i,j)\log_2\frac{P_{XY}(i,j)}{P_X(i)P_Y(j)} \qquad (3.3.1)$$

Where:

X, Y – discrete random variables;
P_X – The probability mass function of X;
P_Y – The probability mass function of Y;
P_{XY} – The probability mass function of the joint distribution;
N_x – The number of elements that are chosen for the histogram of X;
N_y– The number of elements that are chosen for the histogram of Y.

If we refer to the two variables as delayed version of one from the other, we can view the mutual information as auto-mutual information function (specifically for stochastic processes) which is a function depending on the time delay (T) and it is given by the following equation:

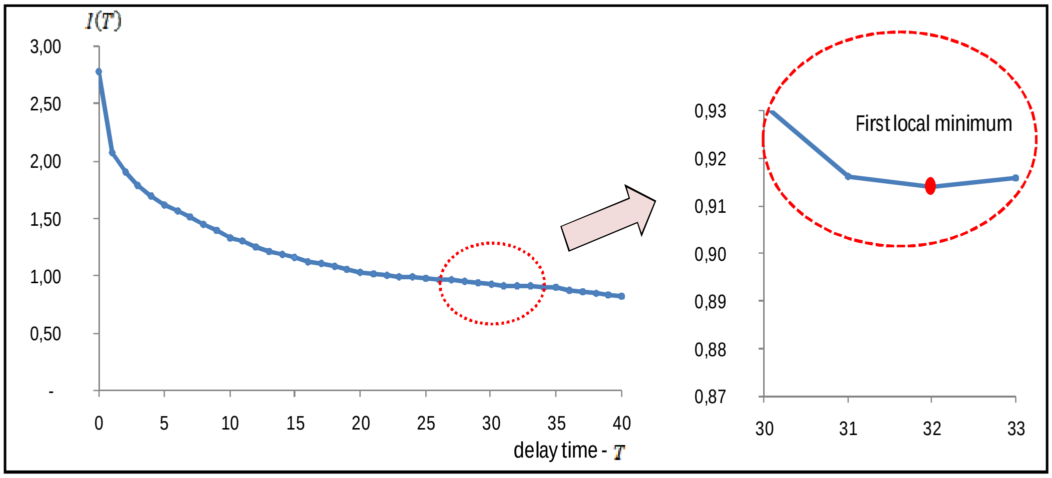
$$I(Y_t, Y_{t-T}) = I(T) = \sum_{i=1}^{N-T}\sum_{j=T+1}^{N} P(Y_i = y(i), Y_j = y(j))\log_2\frac{P(Y_i = y(i), Y_j = y(j))}{P(Y_i = y(i))P(Y_j = y(j))} \qquad (3.3.2)$$

The method of mutual information function for finding the best delay (T_s) was proposed by Fraser, A. M. and Swinney, H. L. [11]. According to this paper, the mutual information is the answer to the following question: Given a measurement of $y(t)$, how many bits (of information) on the average can be predicted about $y(t+T)$? Idealistic will be to have $I(T)$ as small as possible. Actually, $I(T)$ starts from a very high value; given a measurement of $y(t)$ we know as many bits as possible about $(t+0) = y(t)$. As T grows, $I(T)$ decreases, then usually rises again. The authors also suggested using the first local minimum of $I(T)$ in order to select T_s .

As it is, the numerical approximation of the joint probability distribution $P(Y_i = y(i), Y_j = y(j))$ is the most demanding element of the computation. Fraser, A. M. and Swinney, H. L [11] also proposed an algorithm for estimating this function by constructing a locally adaptive partition of the XY plane. Therefore, a Matlab routine was created in order to calculate the auto-mutual information function.

For our data series the first local minimum is reached at $T = T_s = 32$ (see Figure 4) which is not very pronounced, but it indicates a monthly seasonality of the exchange rate.

Figure 4: Auto-mutual information function and its first local minimum



Source: Authors' calculations.

3.4. Forecast EUR/RON exchange rate using ANFIS

*ANFIS*combines two approaches: fuzzy systems and neural networks (a hybrid platform which offers the combinationof learning and nonlinear-adaptability from neural networks plus the approximate reasoning provided by the fuzzy set theory).

It is a five layer network (see Figure 5) that allows dealing with complex issues of uncertainty, improving also the robustness of the control systems. An input signal in the network has the following path:
Step 1. The first layer of the network performs the fuzzification process;

Step 2. The second layer executes the fuzzy AND operator of the antecedent part:

$$w_i = \mu_{A_i}(x) \times \mu_{B_i}(y), i = 1,2 \quad \textbf{(3.4.1)}$$

Step 3. The third layer normalizes the membership functions:

$$\bar{w}_i = \frac{w_i}{w_1 + w_2}, i = 1,2 \quad \textbf{(3.4.2)}$$

Step 4. The fourth layer evaluates the conclusion part of the fuzzy rules. It uses Sugeno fuzzy rules, 'if x is A and y is B then $z = f(x,y)$', which is equivalent to:

$$f_i = f_i(x,y) = a_i x + b_i y + c_i, i = 1,2 \quad \textbf{(3.4.3)}$$

where$a$, b and $c$are consequent parameters.

Step 5. The fifth layer sums up the outputs from the layer four and gives the final output of the fuzzy system:

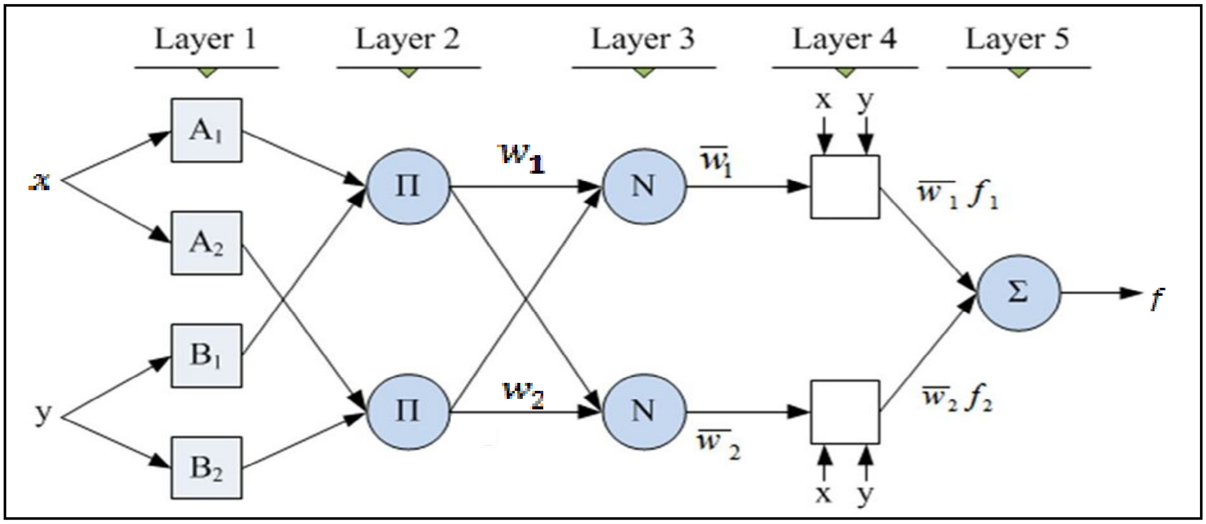
$$f = \bar{w}_1 f_1 + \bar{w}_2 f_2 \quad \textbf{(3.4.4)}$$

The membership function is $Gaussian(m,\sigma)$ and has the following form:

$$\mu_{A_i}(x) = e^{-\frac{(x-m_i)^2}{2\sigma_i}} \quad \textbf{(3.4.5)}$$

where:
m – the center;
σ – the width. This parameters are commonly called the premise parameters.

Figure 5: The Artificial Neuro-Fuzzy Inference SystemsArchitecture



Source: http://www.intechopen.com/books/artificial-neural-networks-architectures-and-applications

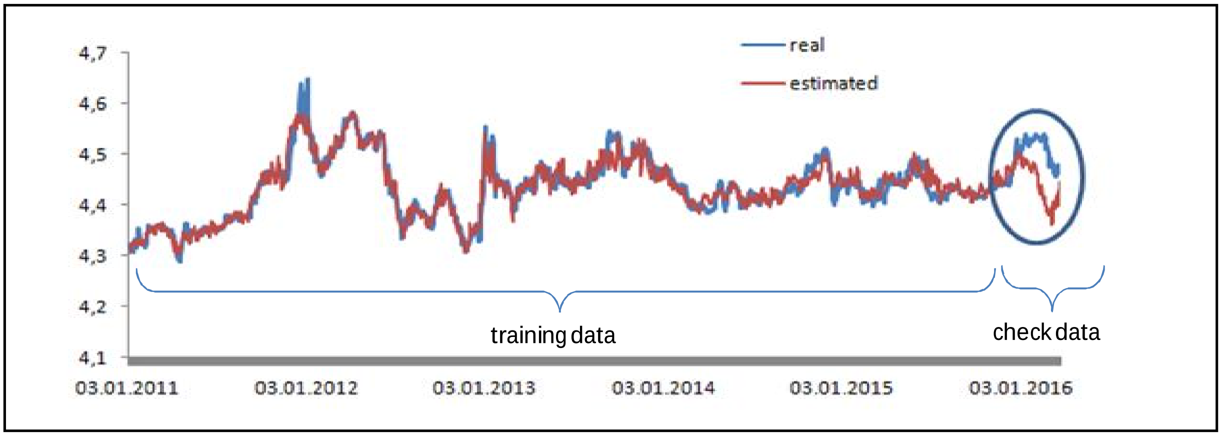
Using Takens theorem and auto-mutual average information function, the input vectors of the network are 6-dimensional vectors with 32 delays between components:

$$\left(y(t), y(t-32), y(t-64), y(t-96), y(t-128), y(t-160)\right) \quad (3.4.4)$$

where $t = 161, \ldots, 1273$.

Next, the first 1080 vectors are used as training data and the last 33 as checking data (for validation purposes). The results show a very good fit of the training data and the dynamic of the series is preserved for the validation data.

Figure 6: The *ANFIS* Output for EUR/RON Exchange Rate



Source: Authors' calculations.

Afterwards, another three simulations (see Table 1) of the EUR/RON exchange rate have been conducted, using the following parameters that were chosen on empirical basis:

1) $T_s = 10$ and $M = 4$
2) $T_s = 5$ and $M = 5$
3) $T_s = 20$ and $M = 3$

Table 1: RMSE for simulations with random pairs of (T_s, M)

	All	Training data	Check data
Ideal ($T_s = 32, M = 6$)	**0,023221**	**0,017803**	**0,087132**
($T_s = 10, M = 4$)	0,027641	0,027568	0,030233
($T_s = 5, M = 5$)	0,020369	0,020467	0,016211
($T_s = 20, M = 3$)	0,037489	0,037533	0,035833

Source: Authors' calculations.

We denoted by 'ideal' the base case where T_s is calculated with *AMIF* and M is calculated using delay embedding. Withal, we used Root-Mean-Square-Error ($RMSE$) as a comparison measure between simulations. The results reveal the lowest $RMSE$ for training data in the 'ideal' case ($RMSE = 0.017803$) which is the most comprehensive data set.

On the other hand, we have the highest $RMSE$ for the check data ($RMSE = 0.087132$), but is not that significant since we have only 33 validation vectors in this set. Extending the validation set, it is recommended to use the network with the parameters that are most fitted to the training data in order to get qualitative prediction.

4. Concluding Remarks

For that matter, it is quite difficult to 'guesses a right pair (T_s, M) for a certain time series since we have to deal with chaos. Besides this, we also have to deal with noise which is structurless and unpredictable.

In practice, in most of the cases there are used random pairs (T_s, M) in order to build input vectors for different neural networks that will predict complex time series, or chose a certain pair that will minimize the error of a certain sample of the time series.

The method presented in this paper ensures that computing T_s with *AMIF* and M using delay embedding, we will also comprise features of the underlying dynamical system that generates the time series. This property is useful in case we want to predict a larger time frame. The analysis performed in this paper reveals that we can have lower prediction errors even if we choose (T_s, M) randomly, but there are two main disadvantages such as:

1). The prediction quality depends on the sample, an increase of the time horizon for prediction or the number of observations will lead to larger prediction errors.
2). The dynamics of the series is not preserved, even if we obtain minimum prediction error of the validation set, we lose the trend of the series, this effect is also specific when we are using least squares method.

As can be seen in most of the research papers, prediction using chaos needs large amounts of observations, the underlying dynamical system that generates an exchange rate time series is usually high dimensional and a consistent training set in this case is mandatory, also *ANFIS* proved to be the most efficient tool that can handle large data sets in order to provide qualitative predictions so the combination between delay embedding technique, auto-mutual information function and *ANFIS* could represent an outstanding prediction tool for complex time series.

References

Alipour, Z., Akhund, A. M., Radmanesh, F., Joorabyan, M. (2014). Comparison of three methods of ANN, ANFIS and Time Series Models to predict ground water level: (Case study: North Mahyar plain). Bulletin of Environment, Pharmacology and Life Sciences, 3 [Special Issue V], 128-134.

Allen, D. E., McAleer, M., Peiris, S., Singh, A. (2015). Nonlinear Time Series and Neural-Network Models of Exchange Rates between the US Dollar and Major Currencies. Tinbergen Institute.

Atmaca, H., Cetişli, B., Yavuz, H. S. (2001). The Comparison of Fuzzy Inference Systems and Neural Network Approaches with ANFIS method for Fuel Consumption Data.Osmangazi University, Electrical and Electronic Engineering Department.

Behmanesh, M., Mohammadi, M., Naeini, V. S. (2014). Chaotic Time Series Prediction using Improved ANFIS with Imperialist Competitive Learning Algorithm. International Journal of Soft Computing and Ennineering (IJSCE), 4(4), pp. 25-33.

Cellucci, C.J., Albano, A. M., Rapp P. E. (2004). Statistical Validation of Mutual Information Calculations: Comparisons of Alternative Numerical Algorithms. Naval Medical Research Institute.

Cellucci, C.J., Albano, A. M., Rapp P. E. (2005). Statistical Validation of Mutual Information Calculations: Comparisons of Alternative Numerical Algorithms. Bryn Mawr College.Physics Faculty Research and Scholarship.

Ciobanu, D., Bar M. V. (2013). On the prediction of the exchange rate USD/EURO by chaos theory. Annals of the "Constantin Brâncuși" University of Târgu-Jiu, Economy Series, Issue 6, pp. 239-252.

Ciorba, A. (2014). Reconstrucțiaatractoruluisursei de semnal.Curs de Procesareasemnalelor, Politehnica University, Bucharest.

Cruz, A. O. (2015). ANFIS: Adaptive Neuro-Fuzzy Inference systems. PPGI, IM-NCE, UFRJ.

Finnsson, I. (2005). Prediction of exchange rate using ANFIS, comparative method study. Soft computing 2005. SigillumUniversitatisIslandiae.

Fraser, A. M., Swinney, H. L. (1986). Independent coordinates for strange attractors from mutual information. Physical Review A, 33(2).

Hunt, B. R., Yorke, J. A. (1993). Maxwell on chaos.Springer-Verlag, 3(1).

Kennel, M. B., Brown, R. ,Abarbanel, H. D. I. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. Physical Review A, 45(6).

Li, W. (1990). Mutual information functions versus correlation functions. Journal of Statistical Physics, 60, Issue 5, pp. 823-837.

Lorenz, E. N. (1963). Deterministic Nonperiodic Flow. Journal of the Atmospheric Sciences, 20, pp. 130-141.

Ma, H. & Han, C.(2006). Selection of Embedding Dimension and Delay Time in Phase Space Reconstruction. Journal of Xi'an Jiaotong University, pp. 335-338.
McNames, J. (1998). A Nearest Trajectory Strategy for Time Series Prediction.Proceedings of the International Workshop on Advanced Black-Box Techniques for Nonlinear Modeling, Katholieke Universiteit Leuven, Belgium, pp. 112-128.
Navarro, R. I. (2013). Introduction to neural networks and Adaptive Neuro-Fuzzy Inference Systems (ANFIS).UniversitatPolitecnica de Catalunya, Department of Aeronautical Engineering.
Otani, M., Jones, A. J. (1997).Automated embedding and the creep phenomenon in chaotic time series.Cardiff School of Computer Science.
Paplinski, A. P. (2005).Fuzzy Logic Toolbox. User's Guide. Neuro-Fuzzy Computing, 12.1-12.19.
Pedram, M., Ebrahimi, M. (2015). The Effects of Economic Variables on Exchange Rate, Modeling and Forecasting: Case of Iran. MacrothinkInstitute™. Business and Management Horizons, 3(1), pp. 13-23.
Ragulskis, M., Lukoseviciute, K. (2009).Non-uniform attractor embedding for time series forecasting by fuzzy inference systems. Elsevier, Neurocomputing 72, pp. 2618-2626.
Robinson, J. C. (2005). A topological delay embedding theorem for infinite-dimensional dynamical systems. Institute of Physics Publishing, Nonlinearity, pp. 2135-2143.
Rohrbacker, N. (2009). Analysis of Electroencephologram Data Using Time-Delay Embeddings to Reconstruct Phase Space.Dynamics at the Horsetooth, 1.Colorado State University, Department of Mathematics.
Rontani, D., Citrin, D. (2013). Introduction to chaos, Georgia Institute of Technology.
Zhang, J., Dai, W., Fan, M., Chung, H., Wei, Z. (2005).An investigation into the use of delay coordinate embedding technique with MIMO ANFIS for nonlinear prediction of chaotic signals. Springer Link, Fuzzy systems and knowledge discovery, 3614,pp. 677-688.
http://www.scholarpedia.org/article/Attractor_reconstruction
https://en.wikipedia.org/wiki/Attractor
https://cnx.org/contents/k57_M8Tw@2/Takens-Embedding-Theorem
http://www.physics.emory.edu/faculty/weeks//research/tseries3.html
http://www.bindichen.co.uk/post/AI/adaptive-neuro-fuzzy-inference-systems.html