Received: 9 January 2018 | Revised: 22 November 2018 | Accepted: 23 November 2018

DOI: 10.1002/isaf.1440

RESEARCH ARTICLE

WILEY

Using Artificial Neural Networks to forecast Exchange Rate, including VAR-VECM residual analysis and prediction linear combination

Alejandro Parot | Kevin Michell | Werner D. Kristjanpoller

Departamento de Industrias, Universidad Tecnica Federico Santa Maria, Valparaiso, Chile

Correspondence
Werner D. Kristjanpoller, Universidad Tecnica Federico Santa Maria, Departamento de Industrias, Avenida España 1680, Valparaiso, Chile.
Email: werner.kristjanpoller@usm.cl

Summary

The Euro US Dollar rate is one of the most important exchange rates in the world, making the analysis of its behavior fundamental for the global economy and for different decision-makers at both the public and private level. Furthermore, given the market efficiency of the EUR/USD exchange rate, being able to predict the rate's future short-term variation represents a great challenge. This study proposes a new framework to improve the forecasting accuracy of EUR/USD exchange rate returns through the use of an Artificial Neural Network (ANN) together with a Vector Auto Regressive (VAR) model, Vector Error Corrective model (VECM), and post-processing. The motivation lies in the integration of different approaches, which should improve the ability to forecast regarding each separate model. This is especially true given that Artificial Neural Networks are capable of capturing the short and long-term non-linear components of a time series, which VECM and VAR models are unable to do. Post-processing seeks to combine the best forecasts to make one that is better than its components. Model predictive capacity is compared according to the Root Mean Square Error (RMSE) as a loss function and its significance is analyzed using the Model Confidence Set. The results obtained show that the proposed framework outperforms the benchmark models, decreasing the RMSE of the best econometric model by 32.5% and by 19.3% the best hybrid. Thus, it is determined that forecast post-processing increases forecasting accuracy.

KEYWORDS

Embedded Models, Artificial Neural Network, Exchange rate return, VAR, VECM, Cointegration

1 | INTRODUCTION

Parities or values between currencies are fundamental to the exchange of goods and services. This is even more true when there is strong international trade as the exchange rate can generate incentives for imports or exports. In this context, one of the main global exchange rates is the Euro US Dollar (EUR/USD). This rate is essential for the decision-making of governments, businesses, and people. For this reason, having more accurate predictions has become a challenge; thus, accurate exchange rate forecasting, both in terms of return and volatility, is a problem of global importance, especially in the financial sector, where a poor forecast can cause big losses. Due to financial market efficiency, it is difficult to make good short- and long-term forecasts (Kilian & Taylor, 2003). In the case of the EUR/USD exchange rate, this rate has long been the most traded exchange rate, making it a very efficient market where abnormal returns cannot be frequently obtained. In this study, we aim to forecast next day returns, which is more complicated as basic techniques are often ineffective, in order to achieve better results that consider changes in interest rates. In addition, exchange rate series are generally non-linear, chaotic, non-parametric, and dynamic (Zhang & Wu, 2009). Therefore, different techniques and methodologies have been developed to predict the exchange rate in a better way than the random walk model. Econometric models have been widely applied for this task; however, they have also been heavily criticized as the majority are linear and work under assumptions that restrict them. Thus, the use of artificial intelligence techniques has spread, such as Artificial Neural Networks (ANN), which are able to model the non-linear behavior of time series through their learning, adaptability, and training properties, without needing to know the data structure in advance, favoring models focused on forecasting ((Chen & Leung, 2004); (Davis, Episcopos, & Wettimuny, 2001); (Kodogiannis & Lolis, 2002); (Yao & Tan, 2000)).Despite the discrepancies between different studies on exchange rate linearity, the literature shows a greater interest in non-linear models to analyze the behavior of this time series ((Clements & Lan, 2010); (Yu, Wang, & Lai, 2005); (Leung, Chen, & Daouk, 2000)). The goal of this study is to verify that there is an improvement in the forecasts of exchange rate returns when using embedded models as opposed to using only econometric or artificial intelligence models, all in a rolling windows frame. To achieve this goal, we first implement a Vector Autoregressive (VAR) model using a daily rolling window in order to determine the long-term relationship between the variables. Second, the presence of cointegration vectors is verified in each window, and the Vector Error Correction Model (VECM) is implemented to obtain the econometric forecasts part of the framework. Thus, the forecast in each window is determined by the VECM or VAR (Called VAR-VECM) depending on whether or not there are cointegration vectors. Then, the error generated by it with respect to the original series is obtained, and these are combined with the past information of the original series in another daily rolling window to feed different configurations of the ANN to obtain the prediction. Finally, a ranking of the best embedded models is performed to determine the best configuration, and their predictions are weighted to obtain the final forecast.This paper's originality lies in the proposition of a multi-step framework for the EUR/USD daily return forecast, which seeks to improve forecasting accuracy in order to support decision-making. This framework is composed by VAR and VECM econometric models and the artificial intelligence ANN model. ANN is fed by the errors of econometric models. Therefore, we seek to capture the residual non-linearity to improve forecasting accuracy. To achieve this, cointegration analysis is performed on a rolling window basis for three daily exchange rates: EUR/USD, GBP/USD, and JPY/USD. The results of the analysis are assessed to make a selection between two models: VECM or VAR (VAR-VECM). Subsequently, the residuals are fed to the ANN along with different amounts of past data. Finally, a ranking is done of the best ANN models to perform predictions, which are weighted according to a supervised linear adjustment. This framework uses several techniques that have been tested in the literature, and it combines them to create a methodology that allows us to obtain better forecasts than the individual techniques. This new proposed framework has the advantage of being simple and rather accurate on forecasting time series. These improvements are of great interest to the financial field, shareholders, and international companies. This is due to the fact that the success of a model utilized for exchange rate return prediction helps to offer additional information on market behavior in order to avoid losses and to decrease exchange exposure. On the other hand, these improvements are also important for researchers because they offer additional analytical tools that may be used to improve research processes and to enrich the literature regarding forecasts. The main difference of this study's approach compared to those of previous studies is the steps that make up the proposed framework. Specifically, the multi-step models mentioned previously have been used in several studies in the literature. However, in this study, we demonstrate that their combination can generate better forecasts for a variable as sensitive and important as EUR/USD. In particular, this study integrates the contribution of linear analysis which considers cointegration, residual analysis by ANN, and the weighting of the best forecasts to generate more accurate predictions. The framework presented as a whole represents the biggest difference between this study and previous studies. In addition, the results obtained in this work substantiate the use of the framework's different steps. In a practical sense, it is possible to apply the methodology sequentially to acquired future information in order to perform investments when it would be convenient, given that the EUR/USD is a highly tradable asset in the financial market. The remainder of the paper is ordered as follows: Section 2 presents an overview of the literature on the topics covered. Section 3 explains the data, different models, and statistical metrics in detail. Section 4 shows and discusses the results obtained in the study. Section 4 discusses the conclusions of the paper.

2 | LITERATURE REVIEW

The existing literature shows the complexity of forecasting the exchange rate, especially after Meese and Rogoff (1983) demonstrated that most linear models are unable to surpass the random walk forecast for the exchange rate. Years after, Alvarez-Diaz et al. (2008) noted that the work of Meese & Rogoff starts from an assumption of linearity, which is not shown empirically. Thus, it is worth exploring new forecasting tools that consider a non-linear structure of the exchange rate. Therefore, machine learning, through ANN, begins to look promising, mainly for its potential as a great tool for estimating and analyzing time series. This can also be seen in the work of Galeshchuk and Mukherjee (2017) where they used Deep Learning for the change direction prediction of foreign exchange rates by employing convolutional neural networks (CNN) over three different exchange indexes. Although CNNs are especially good at image recognition tasks, the study showed that this kind of architecture can also successfully detect directional changes. The work of Rivas, Parras-Gutíerrez, Merelo, Arenas, and García-Fernández (2017) uses ANN together with Genetic Algorithm (GA) to implement a kind of environment that can perform time series predictions in a browser environment by following the jsEvRBF methodology, an approach that was also applied on currency exchange. With a more economical impact, the work of León, Machado, and Murcia (2016) used the Fuzzy Logic Inference System to create an index that can capture the importance of the financial institution size, connectedness and non-substitutability in the Colombian financial market. Also, by using a Principal Component Analysis (PCA) to construct another index, both approaches were used to find relevant information to the study of the systemic importance of financial institutions in Colombia. Empirically, ANNs have been able to accurately model continuous and non-linear functions, without being aware of the data distribution with which they are fed (Galeshchuk, 2016). Furthermore, financial time series are numerical data and do not require prior manipulation to be used nominally as inputs to the ANN. This ignores problems such as a loss of information or using a conversion method that alters the results, among others. Therefore, ANNs are especially well suited to the use of this kind of series, as reported by Lam (2004). ANNs have been used as alternatives to classical prediction techniques for financial series in several studies, such as: Kuan and White (1994) and Swanson and White (1997), who showed the implications and usefulness of applying neural networks to financial market forecasts; Khashei and Bijari (2011) proposed a hybrid model that uses ARIMA to identify the linear structure in data and then uses ANN to determine a model that captures the underlying data generating the process in order to predict its future behavior. They concluded that the proposed model improves the forecasting accuracy achieved by traditional hybrid models and also both of the techniques used separately. Leigh, Purvis, and Ragusa (2002) and Zhang and Wu (2009) demonstrated the successful application of ANN in stock markets to improve forecasts; Dhamija and Bhalla (2011) studied a comparison of various ANN architectures for exchange rate forecasting, finding that ANNs can be effectively used in this task; Khashei, Bijari, and Ardali (2009) showed improvements in predictive results when using ANNs embedded with autoregressive models, concluding that the hybrid model obtained better results than an ARIMA model; Dunis and Huang (2002) applied the ANN to forecast the EUR/USD one step ahead by using autoregressive terms as inputs, showing that the ANN performed better out-of-sample than in the traditional model for annualized returns; Galeshchuk (2016) compared the ANN's predictive power in 3 currencies (EUR/USD, GBP/USD, and JPY/USD) in daily, monthly, and quarterly terms; Yao and Tan (2000) showed the advantages of using ANN, specifically in capturing the nonlinear component of the exchange rate; Laboissiere, Fernandes, and Lage (2015) proposed the stock price forecasting of Brazilian power distribution companies based on ANNs, finding that the methodology provides very good performance on MAE, MAPE, and RMSE terms. Furthermore, Muzhou et al. (2017) showed that the application of a theoretical model called hybrid constructive neural network method (HCNNM) can correctly forecast the price of tungsten and they were also able to validate it statistically as a good forecasting tool. Model composition is an idea that has been growing since the publication of Bates and Granger (1969) and, later, with Clemen (1989), who performs an analysis and bibliographic compilation on the subject. The motivation lies in the possibility of picking up from each model the ability to recognize different elements and patterns of a series in order to combine these capacities and to improve the predictive performance of each individual model, especially when these models have very different characteristics (Baxt, 1992; Zhang, 2007). Hybrid models can be homogeneous, i.e., composed of a succession of models with similar characteristics, or heterogeneous, i.e., composed of linear and nonlinear models, as an example (Taskaya-Temizel & Casey, 2005). Clear examples of the effectiveness of using hybrid models are shown in various studies. For example, Wedding and Cios (1996) described a combined methodology between the Radial Basis Function Network and the ARIMA model; Tsaih, Hsu, and Lai (1998) presented a hybrid model integrating ANNs with rules-based systems techniques to predict the S&P 500 index; Bildirici, Alp, and Ersin (2010) proposed a family of ANNs in combination with a TAR-VEC (Threshold Autoregressive Vector Error Correction) model to obtain improvements in the modeling of non-linear cointegration; Fatima and Hussain (2008) drew a comparison between ARIMA, ARCH/GARCH, and ANN models, separately and together, concluding that the hybrid model had the best performance; Babu and Reddy (2014) proposed an ARIMA-ANN model that provides better prediction accuracy on sunspots, electricity price, and Indian stock data; Adhikari (2015) showed an improvement in forecast accuracy when using Box-Jenkins ARIMA combined with FANN and EANN models; Eswaran and Logeswaran (2012) proposed the use of linear models (linear regression, exponential smoothing, ARIMA) together with ANN to improve time series forecasts, obtaining very good results in terms of MAE, MAPE, and RMSE; Li, Luo, Zhu, Liu, and Le (2008) proposed the application of an AR-GRNN model for financial time series forecasting, showing that the combined model provides an effective way to improve forecasting performance, which can be achieved by either of the models used separately; Khashei and Bijari (2010) introduced a hybrid model, composed of an ARIMA model and an ANN, which had better results in predicting 3 sets of time series with respect to the ANN on its own; Pradeepkumar and Ravi (2017) presented a Particle Swam Optimization (PSO) regression with a ANN to forecast volatility from financial time series and outperform other models; and Abraham, Nath, and Mahanti (2001) proposed using ANNs to feed a neuro-fuzzy system to analyze and forecast the trend of the Nasdaq-100, which obtained favorable results. Therefore, one of the objectives of this paper is to continue contributing to the literature and to the development of new methodologies based on what has previously been developed and presented.

3 | METHODOLOGY AND DATA

3.1 | Data

The exchange rate data used in this study corresponded to the EUR/USD, GBP/USD, and JPY/USD parities in daily terms for the period between 4 January 1999 and 30 December 2015 (4,242 days for the entire process), which are shown in Figures 1, 2, and 3. It is possible to see the relation of the two later exchanges with the former in a VEC-M environment of cointegration.

These figures show that the trend of the three exchange rates was in decline until approximately mid-2008, due to the consequences of the Sub-prime crisis which caused a rise in exchange rates, except for in the case of JPY/USD which showed no large increase in variation and continued to decline until about 2010. In Table 1, it can be seen that the EUR/USD was on average 0.7948 EUR/USD and the average daily variation was -0.0078% during the period to be forecast. The minimum value reached during the period was 0.6252 EUR/USD while the highest value reached was 1.1643 EUR/USD. The high standard deviation of the daily variations compared to the average was observed and shows a high volatility in the series to forecast. The kurtosis of the price and its variation is similar while the skewedness in the case of the exchange rate is positive but negative in the case of variation.

For modeling, the series are used in terms of natural logarithm at time t, defined as Y_{kt}, where $k = \{1, 2, 3\}$ represents the series Euro

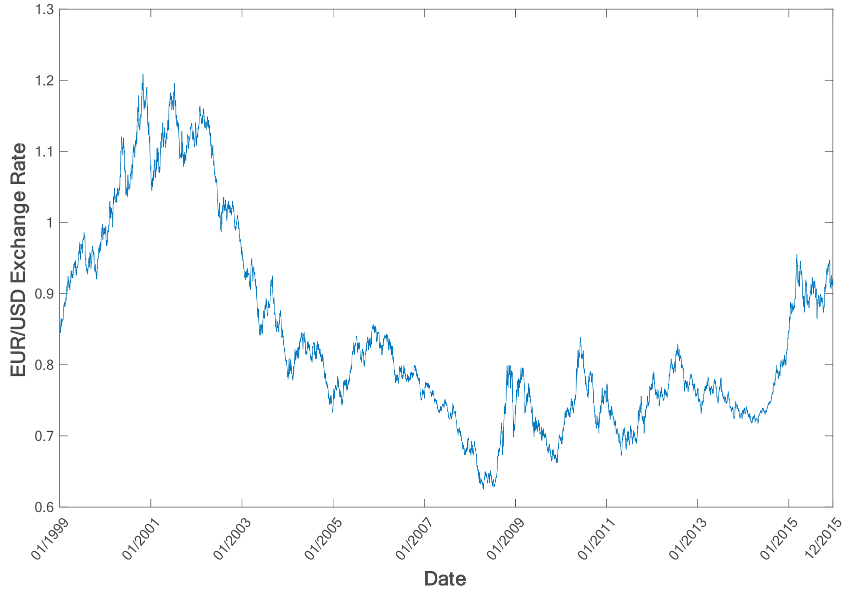

FIGURE 1 EUR/USD exchange rate on days

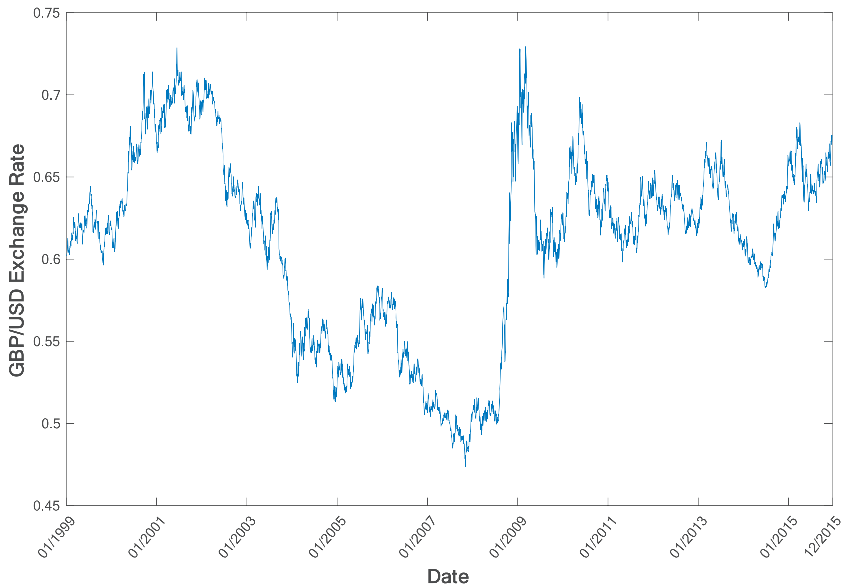

FIGURE 2 GBP/USD exchange rate on days

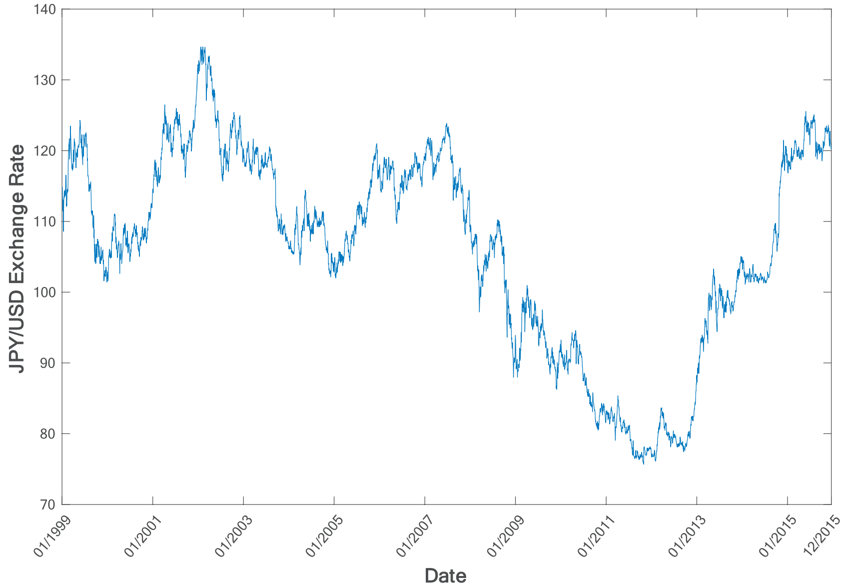

FIGURE 3 JPY/USD exchange rate on days

TABLE 1 Descriptive statistics of EUR/USD

	EUR/USD	EUR/USD % Var.
Average	0.7948	-0.0078%
St. deviation	0.0978	0.6433%
Min	0.6252	−3.7152%
Max	1.1643	3.2356%
Kurtosis	2.3035	1.9867
Skewness	1.3686	−0.0060

(EUR/USD), Pound Sterling (GBP/USD), and Yen (JPY/USD) exchange rates, respectively. The exchange rate return, or variation, is defined as the first difference of the series, and this corresponds to the variation of Y_{kt}, defined as $\Delta Y_{kt} = Y_{kt} - Y_{kt-1} = y_{kt}$. These transformations can be seen in Figures 4, 5, and 6.

These figures demonstrate the movement of the exchange rate return. Here, the values are kept within a stable range but, as in the previous exchange rate figures, the values close to 2008 stand out because they exceed these margins due to the anomalies resulting from the 2008 Subprime crisis.

3.2 | Vector autoregressive and vector error correction models

The Vector Autoregressive (VAR) model is used to show the simultaneous interactions among a group of variables through a system of equations that considers that the set of explanatory variables of each equation is formed by the lags of each of the model's endogenous variable in addition to its deterministic or exogenous variables and their respective lags. In general, a VAR model of order n, with Gaussian error, can be expressed in matrix form as:

$$y_t = \sum_{s=1}^{n} A_s y_{t-s} + Bx_t + e_t \qquad t = 1, \ldots, T \tag{1}$$

where y_t is a $k \times 1$ column vector of k endogenous variables in time t; A_s are $k \times 1$ matrices of the coefficients of each endogenous variable

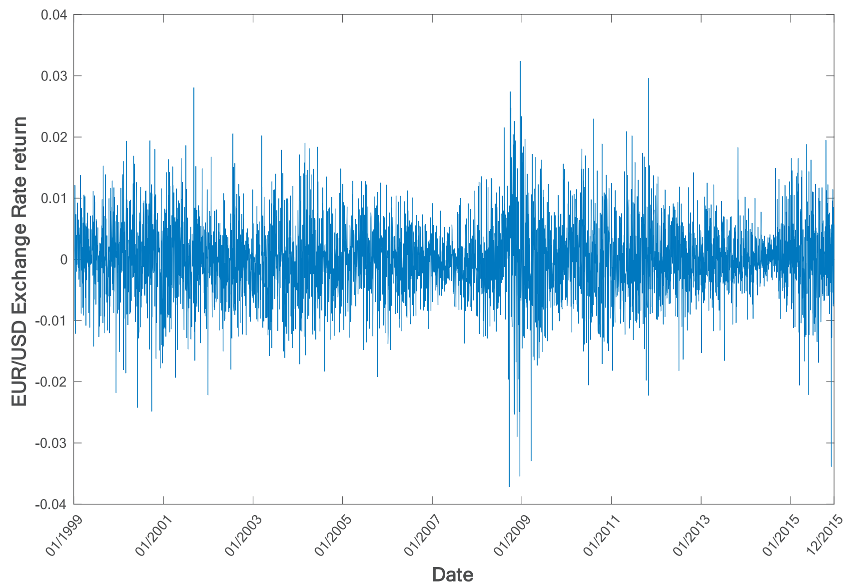

FIGURE 4 EUR/USD exchange rate on days

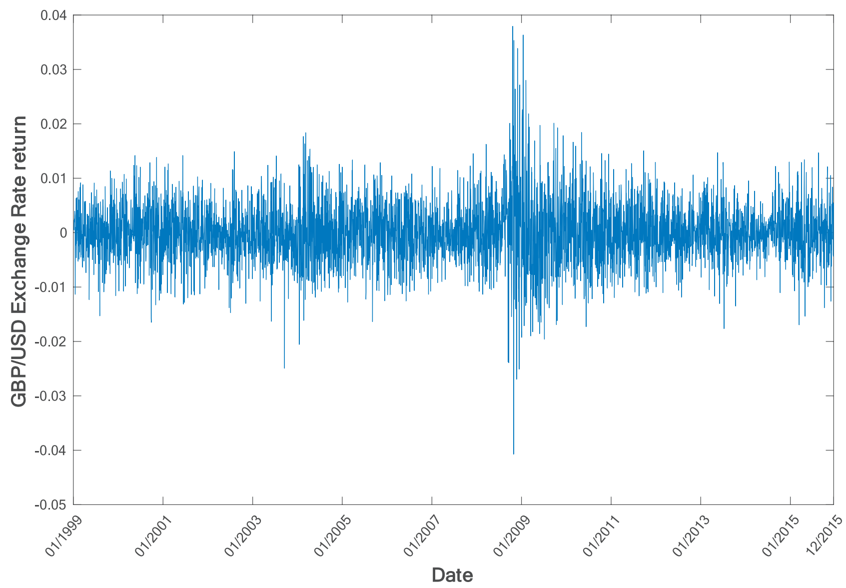

FIGURE 5 GBP/USD exchange rate on days

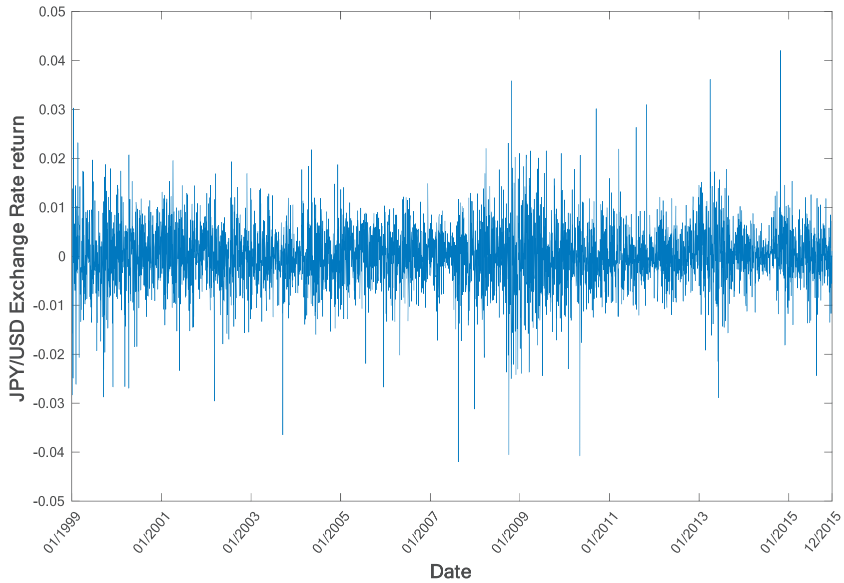

FIGURE 6 JPY/USD exchange rate on days

with lag s; y_{t-s} is a $k \times 1$ column vector of the k endogenous variables with lag s; x_t is a $p \times 1$ column vector of the p exogenous variables in time t; B is a $k \times p$ matrix of the coefficients of each exogenous variable; e_t is a $k \times 1$ column vector of the random errors of each endogenous variable in time t, distributed Gaussian, without autocorrelation, with mean equal to zero and a constant variance. This study uses cointegration analysis, introduced by Granger (1981) and developed later by Engle and Granger (1987), Johansen (1988), and Johansen and Juselius (1990). Granger's theorem establishes the existence of an error-correction model of x_t, assuming Δx_t and $\beta' x_t$ have stationary and invertible VARMA representations for the cointegration vectors β' (Engle & Granger, 1987). The cointegration vectors β have the property of making $\beta' x$ stationary, although x is non-stationary (Johansen & Juselius, 1990). Considering the VAR model of equation (1), the error correction representation (VECM) applied according to Johansen and Juselius (1990) corresponds to:

$$y_t = \sum_{s=1}^{n-1} \Gamma_s y_{t-s} + \Pi Y_{t-n} + Bx_t + e_t \qquad t = 1, \ldots, T \tag{2}$$

where the terms and dimensions of the VAR model are maintained, and we define: the matrix $\Gamma_s = -I + \sum_{i=1}^{s} A_i$, where I is the identity matrix; and $\Pi = -\mathrm{I} + \sum_{i=1}^{n} A_i$ that score the long-term information between the variables. Cointegration verifies the existence of a stochastic tendency in common within a set of variables. The Johansen & Juselius test (1990) is used to evaluate the existence of cointegration through 2 different tests: The Trace test and the Maximum Eigenvalue test. In the Trace test, the null hypothesis assumes that there are at most r cointegration vectors, which is tested against the alternative hypothesis that there is exactly $r \pm 1$. In the Maximum Eigenvalue test, the null hypothesis corresponding to exactly r cointegration vectors is contrasted against the existence of $r \pm 1$ cointegration vectors.

3.3 | Artificial neural network (ANN)

ANNs are non parametric processing systems that have the great advantage of recognizing non-linear patterns in a time series as well as making inferences even though the data contain noise. The parallel processing of the data causes this type of model to be strongly affected by it. For this reason, it does not require previous assumptions. In addition, the ANN can approximate a large number of functions with a high degree of accuracy (Hornik, Stinchcombe, & White, 1989). This makes the ANN especially efficient for working with series such as the exchange rate return. The learning method that was used in this study is supervised, i.e., adjusts the weights by comparing the outputs obtained from the model and the available real outputs. A widely used ANN model for modeling and forecasting time series is the Perceptron, which consists of one single hidden layer and one node in the output layer (Zhang, Patuwo, & Hu, 1998). An extended version of this structure is characterized by a l-layer network containing forward-connected processing units, i.e., each of the nodes of the input layer $(z_1, z_2, \ldots, z_P)$ are connected to each node of the first hidden layer $\left(h_1^{(1)}, h_2^{(1)}, \ldots, h_{Q_1}^{(1)}\right)$; then, these are connected to the nodes of the next hidden layer $\left(h_1^{(2)}, h_2^{(2)}, \ldots, h_{Q_2}^{(2)}\right)$ and so on until the nodes of the last hidden layer $\left(h_1^{(L-1)}, h_2^{(L-1)}, \ldots, h_{Q_{L-1}}^{(L-1)}\right)$ are connected to the node of the output layer $(\hat{y}_t)$, where $l = 1, \ldots, L$. The structure is expressed mathematically in the equations (3), (4), (5):

$$h_j^{(1)} = g^{(1)}\left(\sum_{i=1}^{P} w_{i,j}^{(1)} z_i + w_{0,j}^{(1)}\right) \tag{3}$$

$$h_j^{(l)} = g^{(l)}\left(\sum_{k=1}^{Q_{l-1}} w_{k,j}^{(l)} h_k^{(l-1)} + w_{0,j}^{(l)}\right) \tag{4}$$

$$\hat{y}_t = h_j^{(L)} = g^{(L)}\left(\sum_{k=1}^{Q_{L-1}} w_{k,j}^{(L)} h_k^{(L-1)} + w_{0,j}^{(L)}\right) \tag{5}$$

where the hat in y indicates that this is the final forecast of the proposed model; the input z_i can respond to exogenous variables in time t or lagged, endogenous variables, residual in time t or lagged, among others (for $i = 1, \ldots, P$); $w_{k,j}^{(l)}$ represent the weights from node k to node j in layer l (for $l = 1, \ldots, L$, $k = 1, \ldots, P$ and $j = 1, \ldots, Q_l$, with Q_l as the amount of neurons in the layer l); $h_j^{(l)}$ is the output from neuron j in layer l (for l = 1, … , L and j = 1, … , Q_l, with Q_l as the amount of neurons in the layer l); $w_{0,j}^{(l)}$ represents the weights from the bias unit to the node j in the layer l (for l = 1, … , L and j = 1, … , Q_l); $g^{(l)}$ represents the transfer function in the layer l. The most commonly used learning algorithm is backpropagation. This optimization algorithm adjusts the weights by an iterative process based on the gradient descent method, in order to minimize a cost function (Rumelhart, Hinton, Williams, et al., 1988). The choices of the parameters P and Q_l, as well as the number of variables that are included as inputs, were not made according to a rule. This is due to the fact that there is no method that guarantees finding these values in an optimal way, because they depend on the characteristics of the data. Therefore, it is common to test different ANN configurations until identifying the one that produces the smallest generalized error (Hosseini, Luo, & Reynolds, 2006). This is done following a build-up logic by which the method is computed with the most basic configurations for the ANN (this is, few layers and few neurons) and then they are incremented to achieve better performance/complexity ratio. It is important to note that a more complex ANN is more complicated to train, and it needs more data to learn its parameters. The specific configurations of these hyper parameters are presented in the following section.

3.4 | Proposed model

The proposed model corresponds to an embedded approach for predicting one step ahead, based on the idea that combining different models unifies their characteristics, allowing more information to be rescued from a time series. This leads to increased forecasting accuracy with respect to the use of a single model. Specifically, we propose combining a VAR model with an Error Correction Model (VECM) (linear model) and Artificial Neural Network model (ANN) (non-linear model), using backpropagation as a learning algorithm, forming a VAR-VECM-ANN model. After the predictions, a weighting of these is done as a post-processing to generate better forecasts. Both in the choice of the best embedded model as well as in the number of models to weigh, the determination goes through the RMSE loss function. The MCS test is applied for the robustness of the results. The general view of the framework can be seen in Figure 7.

1. The first step is to apply the unit root test to check stationarity, as proposed by Dickey and Fuller (1979). If the series are stationary, the VAR model is applied considering the return of EUR/USD, GBP/USD, and JPY/USD exchange series. The cointegration tests are performed considering the existence of a relationship between the 3 series, in order to obtain the error correction model (VECM). The VECM model is used when there exists cointegration; otherwise, the window considers a VAR model. Thus, it is analyzed whether to VAR or VECM model is better at each step. Later on, this fixed window size is sensitized to 42 (two months), 63 (three months), 84 (four months), 252 (a year), and 378 (one year and three months) days. These parameters were selected because of their economical interpretation, as this information is generally useful in making seasonal predictions. The periods selected are used in different studies (Antonakakis & Badinger, 2016; Antonakakis, Dragouni, & Filis, 2015a; Antonakakis, Dragouni, & Filis, 2015b) and these periods resemble short run (two to three months), mid run (four to six months) and long run (year to year and three months). Thus, the aim is to identify the best VAR-VECM in terms of RMSE. The model that is selected as the best will be used as a benchmark against the proposed model and its residuals will be used to feed the ANN. For example, in the 126-day rolling window, we analyze cointegration to the data from day 1 to 126 to forecast day 127 with the model VAR or VECM according to the existence of vectors of cointegration. Then, the data are analyzed from day 2 to 127 to forecast day 128, and so on until the end of the data. The VECM model works with the optimal lags, indicating the lowest AIC in each window, with a maximum of 8.
2. With the residual of the best VAR-VECM model, the ANN is applied to identify the relationships between data and residuals extracted. Specifically, it is fed with the return of the EUR/USD series in time t and its lags, and also with residuals of the best VAR-VECM model, previously chosen, in the time t (e_t) and its lags. To be consistent with the maximum imposed in the VAR-VECM model, it was determined to feed the ANN with 8 lags of each variable. The ANN forecast is also done in a rolling windows approach with 126 days (six months) as training data for windows size. Similar to what was done in the econometric part, the fixed windows size is sensitized to 42 (two months), 63 (three months),

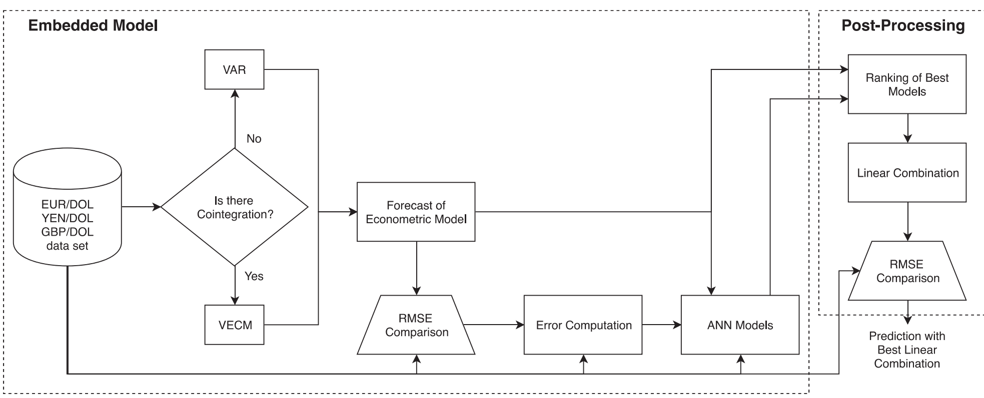

FIGURE 7 Flow chart of the proposed framework

84 (four months), 252 (a year), and 378 (one year and three months) days.

3. A sensitization of hidden layers and neurons is performed, i.e., the number of hidden layers can be: 1, 2, 3, or 4; while the number of neurons per hidden layer can be: 5, 10, 15, or 20. These configurations were selected because they are widely used in the literature and fit well with the data analyzed.

The equations (3), (4) and (5) are rewritten, for this case, as equations (6), (7) and (8):

$$h_j^{(1)} = g^{(1)}\left(\sum_{i=1}^{a} w_{i,j}^{(1)} y_{1,t-i} + \sum_{i=a+1}^{a+b} w_{i,j}^{(1)} e_{1,t+a-i} + w_{0,j}^{(1)}\right) \tag{6}$$

$$h_j^{(l)} = g^{(l)}\left(\sum_{i=1}^{Q_{l-1}} w_{k,j}^{(l)} h_k^{(l-1)} + w_{0,j}^{(l)}\right) \tag{7}$$

$$\hat{y}_t = h_t^{(L)} = w_0^{(L)} + g^{(L)}\left(\sum_{i=1}^{Q_{L-1}} w_k^{(L)} h_k^{(L-1)}\right) + u_t \tag{8}$$

where the previous variables are maintained and in addition: the input $y_{1,\,t\,-\,i}$ corresponding to the return of the EUR/USD series at time $t-i$ (for $i = 1, \ldots, a$); the input $e_{t\,+\,a\,-\,i}$ which responds to the residuals of the VAR-VECM model over time $t + a - i$ (for $i = a + 1, \ldots, a + b$); the values a and b are integers that correspond to the number of lags in the series and the residuals, respectively, with $a + b = P$ .The reason for considering residuals (e) in this analysis is that they contain the nonlinear relationship that the linear model is unable to recognize (Khashei & Bijari, 2010), helping to obtain a more accurate forecast than the one delivered by the benchmark.

4. The results obtained for the VAR-VECM-ANN models are assessed using the RMSE and this RMSE is then compared with that of the best fit VAR-VECM model to verify that there is indeed an improvement in the predictive capacity of the proposed model.
5. Finally, a ranking is performed with the best embedded VAR-VECM-ANN models based on their performance with respect to the RMSE loss function. These models are combined, weighing them according to a simple linear regression, allowing 126 days for adjustment of the weightings, taking the best 5 models. Similar to the other steps of the framework, a sensitization is performed of the adjustment windows, considering 42 (two months), 63 (three months), 84 (four months), 126 (six months), 252 (a year), and 378 (one year and three months) days and also the number of models to weigh, considering 2, 10, 15, 20, and 25 as a maximum. With these weightings, the final forecast of the framework is presented, which is then compared to the benchmark models and by the Model Confidence Test (MCS)

3.5 | Loss function

In order to understand whether or not the forecasts provided by a model are better than the ones delivered by another, it is necessary to compare the predictive capacity of each one. As an indicator to measure performance in forecasts, the RMSE is specified in equation (9)

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(y_{1t} - \hat{y}_{1t}\right)^2} \tag{9}$$

This indicator shows the value of the sum in t of the errors between the forecast $\hat{y}_{1t}$ and the target y_{1t} for $k = 1$, that is EUR/USD, all under a squared root. Once we have the RMSE of each model, it is possible to compare the predictive capacity, with the best model delivering the lowest RMSE.

3.6 | Model confidence set (MCS)

The Model Confidence Set (MCS), proposed by Hansen, Lunde, and Nason (2011), is an improvement of the Superior Predictive Ability (SPA) model proposed by Hansen (2005) to show robustness in the results. The SPA is designed to address whether a particular benchmark is significantly outperformed by any of the alternatives used in the comparison in terms of expected loss. In particular, the null hypothesis is that the benchmark is not inferior to any of the $k = 1, \ldots, m$ alternatives. A rejection of the SPA test only identifies one or more models as significantly better than the benchmark. Thus, the SPA test offers little guidance about which models reside in M^*, i.e., the set of superior objects that contains a finite number of objects where the null is true, but does not specify which model is best in that set. The MCS test does not require a benchmark to be specified, because all models serve as the benchmark in the series of comparison. It also has the advantage that it can be employed for model selection, unlike the SPA. So, a sequence of SPA tests defines the MCS to be the set of benchmark models that are found not to be significantly inferior to the alternatives, excluding the inferior models. Here, the null hypothesis is $H_0\colon \mathbb{E}(d_{ij}) = 0$ where $d_{ij} = L_i - L_j$ for the models i and j in the set, and the loss L. The objective of the MCS procedure is to determine the set M^* where a model i is preferred in terms of expected loss to an alternative j if $\mathbb{E}(d_{ij}) < 0$ for all j in the set M^0 that contains all the models to analyze. This is done through a sequence of significance tests, where objects that are found to be significantly inferior to other elements of M^0 are eliminated to determine M^*.

4 | RESULTS AND DISCUSSION

First, the ADF unit root test, proposed by Dickey and Fuller (1979), showed that the EUR/USD, GBP/USD, and JPY/USD exchange series are stationary when applying the natural logarithm difference transformation, i.e., they are $I(1)$ and we used the return of the series. This prior analysis, and all of the following, was done through rolling windows for different lengths: 42, 63, 84, 126, 252 and 378 days. Once the series were stationary, the VAR and VECM econometric models were applied over the set to obtain forecasts and the residuals of this model with respect to the EUR/USD series. Different ANN configurations were then used to model the nonlinear and linear components that exist in the data and in the residuals. Subsequently, the best 5 models were combined linearly, as a post-processing in order to improve the forecast. Finally, RMSE and two statistical tests were applied to compare the predictive capacity of the proposed framework with respect to the benchmark defined below. The first models analyzed were the traditional models used to predict the exchange rate, ARIMA, VAR-VECM and ANN, as well as the Random Walk model. The performance of these models is presented in Table 2. The Random Walk model with the premise that the best prediction for the next exchange rate is today's price, has a RMSE of 0.6432 %. The best ARIMA model used to predict the EUR/USD exchange rate has one autoregressive term and one moving average term and the accuracy is a RMSE of 0.6439%, which was slightly higher than the Random Walk. The prediction made through ANN also generated a slightly worse forecast than the Random Walk. On the other hand, the best VAR-VECM model outperformed the Random Walk, showing that the incorporation of other exchange rates and their variations as explanatory variables provides predictive information. The RMSE of the best VAR-VECM is 0.5814 % and it will be the RMSE benchmark. These first results show how efficient the EUR/USD market is, which makes it challenging to try to predict its values. As an input, the ANN model uses the EUR/USD variation and the residual of the best econometric model (VAR-VECM W:63). Both series are input to the ANN as well as their lags. The first ANN configurations are with a 126-day rolling window, with one to four hidden layers, and with 5, 10, 15, and 20 neurons per layer. RMSE is applied to each to find the combination that yields the best forecast. In Table 3, it can be observed that by performing the second step of the embedded model, the RMSE was reduced by 7.64% with the configuration 2 hidden layers and 15 neurons per layer, thus increasing the forecast accuracy and demonstrating how the econometric and neural network models complement one another to make better predictions.In search of the best ANN configuration, we performed all the forecasts with the same configurations of layers and neurons, but reduced the window length to 42, 63, and 84 days. In Table 4, it can be observed that by reducing the longitude of the window, although in some configurations the econometric model improves, better results are not achieved than the result obtained with the 126-day window. By increasing the size of the window that feeds the ANN to 252 and 378 days, it can be observed (Table 5) that there was an improvement in the RMSE in various configurations. It can be noted that all the combinations with 4 hidden layers have worse results than the best econometric model. This once again confirms that the configuration of 2 layers is the best. In particular, 378 days and 20 neurons per layer obtains the best performance, reducing the RMSE of the best econometric model by 16.36%. We ran the model with 25 neurons per layer and the RMSE did not improve. After analyzing all the embedded model configurations, the accuracy results of the models were sorted to define the ranking of the top 5 models. The ranking can be seen in Table 6. With these 5 models, we proceeded to perform different linear combinations for the different window lengths. We first optimize the 2 best models, then the best 3, and so on until obtaining the combination of the best 5. In each window, the best model was determined according to OLS, and then forecast. Once the forecast was made, we moved forward a day and again took the window, optimized, and predicted the following day. It can be seen in Table 7 that, by performing this post-processing, the results improved significantly. The best model is the linear combination of the top 5 VAR-VECM-ANN models with a 63-day rolling window. This model improves the RMSE of the best econometric model by 27.55% while it decreases the RMSE of the best VAR-VECM-ANN model by 13.38%. This result is very important because, with simple post-processing, the error in EUR/USD variation forecast was reduced by almost a quarter. Due to the fact that the optimal linear combination was reached with the maximum of models included, we analyzed the top 10 VAR-VECM-ANN models. In this case, the best model incorporated the 10 models with a 378-day window, reducing the RMSE by 28.03% and 13.95 % with respect to the best econometric and VAR-VECM-ANN model, respectively. Given these results, the number of models was further extended, taking up to 20 top VAR-VECM-ANN models. With the top 20 VAR-VECM-ANN as a maximum, we determined that the best model is made up by a combination of the top 17, reducing the RMSE of the best econometric model by 32.52% and the RMSE of the best VAR-VECM-ANN by 19.32%. Once we had a first impression of the results in terms of RMSE, statistical tests were performed. This is due to the fact that, even though a model delivers the lowest RMSE, there are other models that approach it. Therefore, based solely on this indicator, we cannot be assured that a given model is best.The MCS test was then applied to the 216

TABLE 2 RMSE for the top 5 ARIMA, VAR-VECM and ANN models

ARIMA	AR/MA/W	VAR-VECM	W	ANN	L/N/AR/W
0.6439%	1/1/376	0.5814%	63	0.6448%	2/20/376/1
0.6445%	1/1/252	0.5828%	126	0.6449%	3/5/376/1
0.6447%	1/1/126	0.5829%	84	0.6450%	1/20/42/10
0.6454%	1/2/376	0.6108%	252	0.6453%	3/10/376/10
0.6456%	2/1/376	0.6128%	378	0.6454%	3/20/376/1

AR: Number of autoregressive terms; MA: Number of moving average terms; W: Window length; L: Hidden layers; N: Neurons per layer.

TABLE 3 RMSE for VAR-VECM-ANN with 126-day window length

	Hidden layers							
	1		2		3		4	
Neurons per layer	RMSE	% Var.	RMSE	% Var.	RMSE	% Var.	RMSE	% Var.
5	0.5704%	−1.90%	0.5515%	−5.14%	0.5796%	−0.32%	0.6041%	3.90%
10	0.5726%	−1.52%	0.5413%	−6.89%	0.5746%	−1.17%	0.6061%	4.24%
15	0.6390%	9.90%	0.5370%	−7.64%	0.5756%	−0.99%	0.6142%	5.65%
20	0.7210%	24.01%	0.5409%	−6.97%	0.5847%	0.56%	0.6087%	4.69%

Window is the length of the rolling windows.

% Var. means the variation percentage of RMSE compared to the best econometric model.

TABLE 4 RMSE for VAR-VECM-ANN with 42, 63, and 84-day window length

Window neurons per layer	RMSE	% Var.	RMSE	% Var.	RMSE	% Var.	RMSE	% Var.
42 days								
5	0.5718%	−1.65%	0.6076%	4.50%	0.6271%	7.86%	0.6434%	10.66%
10	0.6227%	7.09%	0.6142%	5.63%	0.6386%	9.84%	0.6452%	10.98%
15	0.7590%	30.55%	0.6153%	5.84%	0.6394%	9.98%	0.6492%	11.65%
20	0.7975%	37.17%	0.6042%	3.92%	0.6474%	11.35%	0.6588%	13.3%
63 days								
5	0.5472%	−5.89%	0.5841%	0.46%	0.6112%	5.12%	0.6255%	7.57%
10	0.6194%	6.54%	0.5857%	0.74%	0.6106%	5.03%	0.6282%	8.04%
15	0.6856%	17.92%	0.5902%	1.51%	0.6088%	4.72%	0.6369%	9.55%
20	0.7692%	32.29%	0.5756%	−1.00%	0.6075%	4.49%	0.6397%	10.02%
84 days								
5	0.5635%	−3.07%	0.5734%	−1.38%	0.5947%	2.28%	0.6117%	5.21%
10	0.5834%	0.35%	0.5613%	−3.46%	0.5878%	1.1%	0.6042%	3.92%
15	0.6848%	17.78%	0.555%	−4.54%	0.5757%	−0.98%	0.5972%	2.72%
20	0.773%	32.96%	0.5428%	−6.63%	0.5707%	−1.84%	0.5924%	1.89%

Window is the length of the rolling windows.

% Var. means the variation percentage of RMSE compared to the best econometric model.

TABLE 5 RMSE for VAR-VECM-ANN with 252 and 378-day window length

Window neurons per layer	RMSE	% Var.	RMSE	% Var.	RMSE	% Var.	RMSE	% Var.
252 days								
5	0.8071%	38.82%	0.543%	−6.6%	0.5716%	−1.69%	0.5937%	2.12%
10	0.5534%	−4.82%	0.5244%	−9.8%	0.5673%	−2.42%	0.5926%	1.92%
15	0.5592%	−3.82%	0.5154%	−11.35%	0.5671%	−2.46%	0.5943%	2.21%
20	0.602%	3.55%	0.4981%	−14.33%	0.5601%	−3.67%	0.5945%	2.26%
378 days								
5	0.7152%	23.01%	0.5339%	−8.17%	0.5781%	−0.57%	0.5916%	1.75%
10	0.5824%	0.17%	0.5063%	−12.92%	0.565%	−2.82%	0.5924%	1.89%
15	0.5648%	−2.85%	0.4986%	−14.25%	0.5606%	−3.59%	0.5908%	1.61%
20	0.6158%	5.92%	0.4863%	−16.36%	0.5606%	−3.58%	0.5975%	2.76%

Window is the length of the rolling windows.

% Var. means the variation percentage of RMSE compared to the best econometric model.

TABLE 6 Ranking of the best VAR-VECM-ANN models

Position	Hidden layers	Neurons	Window	RMSE	% Var.
1	2	20	378	0.4863%	−16.36%
2	2	20	252	0.4981%	−14.33%
3	2	15	378	0.4986%	−14.25%
4	2	10	378	0.5063%	−12.92%
5	2	15	252	0.5154%	−11.35%

Window is the length of the rolling windows. Neurons is the number of neurons per layer. % Var. means the variation percentage of RMSE compared to the best econometric model.

models run for 3,100 predictions each. The results of this test are consistent with the RMSE seen in Table 8, showing that the best forecasts are obtained when post-processing was included. Table 9 only shows the 10 models with the greatest p-value. It can be observed that the last 5 models shown possess a p-value of 0.032, which means that 211 are not considered and are immediately discarded by the test.

TABLE 7 RMSE of linear combination models

Models	Windows	RMSE	% Var-VECM	% Var ANN
5	63	0.4212%	−27.55%	−13.38%
5	84	0.4221%	−27.40%	−13.20%
4	63	0.4243%	−27.02%	−12.74%
4	84	0.4252%	−26.86%	−12.55%
5	126	0.4268%	−26.60%	−12.24%

Window is the length of the rolling windows.

% Var-VECM means the variation percentage of RMSE compared to the best econometric model.

% Var-ANN means the variation percentage of RMSE compared to the best embedded model.

The best 5 models according to the MCS test corresponded to models with post-processing, implying that this approach generates an improvement in forecasting accuracy that is statistically significant, but only the first two surpass a p-value of 0.500. For this test, a model

TABLE 8 RMSE of linear combination models for top 10 and top 20 VAR-VECM-ANN models

Top 10 models	Models	Windows	RMSE	% Var-VECM	% Var ANN
	10	378	0.4148%	−28.03%	−13.95%
	9	378	0.4198%	−27.80%	−13.68%
	10	252	0.4204%	−27.70%	−13.56%
	9	252	0.4216%	−27.48%	−13.30%
	8	378	0.4222%	−27.38%	−13.18%
Top 20 models	**Models**	**Windows**	**RMSE**	**% Var-VECM**	**% Var ANN**
	17	378	0.3923%	−32.52%	−19.32%
	18	378	0.3927%	−32.46%	−19.25%
	19	378	0.3934%	−32.33%	−19.10%
	20	378	0.3944%	−32.17%	−18.90%
	17	252	0.3967%	−31.76%	−18.41%

Window is the length of the rolling windows.

% Var-VECM means the variation percentage of RMSE compared to the best econometric model.

% Var-ANN means the variation percentage of RMSE compared to the best embedded model.

TABLE 9 MCS Test p-values of the best models

Models	Windows	MCS p-value
17	378	1.000
18	378	0.583
19	378	0.425
20	378	0.163
17	252	0.134
15	378	0.032
18	252	0.032
16	378	0.032
19	252	0.032
15	252	0.032

MCS is the p-value of the Model Confidence Set, Hansen, Lunde, and Nason (2003) and Hansen et al. (2011).

that yields a p-value greater than 0.95, at a significance level of 5%, rejects the hypothesis that there is another model in the set that equals or improves it. Thus, the model that delivers a p-value of 1 means that it is the last model in the set and therefore, is as good as itself. Analyzing the best five, clearly there is only one that is the best, since with an M95, the only linear combination model that qualifies for the top 17 VAR-VECM-ANN models is with the 378-day window.

5 | CONCLUSION

Modern financial markets present a high degree of competition among participants. Therefore, having as much information as possible is fundamental. In this scenario, exchange rate forecasts play a special role in anticipating market movements. This paper applies an embedded model using a rolling window analysis to predict the EUR/USD exchange rate return, itself the most important and traded exchange rate in the world. This makes it a very efficient market and thus, there is greater difficulty in predicting its behavior. There will never be a perfect prediction. However, through more advanced models, we can obtain better accuracy. The motivation to use embedded models rests on the fact that, theoretically and empirically, the integration of different models improves the ability to forecast with respect to each individual model. Thus, the proposed model that considers an ANN fed by residuals from an econometric VAR-VECM model should provide better forecasts than classical models. Following the same idea, the linear combination of several forecasts provides a better prediction because it takes the information of several characteristics of the best models, therefore enhancing the results. We use a rolling window analysis to obtain a dynamic model that can better adjust changes to the data with respect to a model that only considers a general framework. The results show that the proposed model yields better forecasts than the econometric one, especially using a 378-day rolling window with a configuration of 2 hidden layers and 20 neurons in the case of VAR-VECM-ANN, and a 17 model linear combination in the last part of the framework. These results are evaluated in terms of RMSE and are then tested statistically through MCS. The RMSE shows that the VAR-VECM-ANN model improves the forecast by 16.36% in comparison to the best econometric model. Meanwhile, the model with post-processing improves the best econometric model by 32.35% and the best VAR-VECM-ANN model by 19.32%. Finally, the MCS test shows that the 378-day rolling window with a linear combination of 17 models outperforms all other approaches considered in this study, including VAR-VECM-ANN with a configuration of 2 hidden layers and 20 neurons, concluding that there is no better model in predictive terms. Without a doubt, these results are excellent for those who must be vigilant of daily variations in the EUR/USD exchange rate since we obtained greater forecasting accuracy with a hybrid model with linear combination post-processing. It is also important to note that the post-processing performed is simple. Nonetheless, its results are very important regarding the increase in forecast accuracy. Finally, it can be concluded that the search for the best forecasting model is sometimes incomplete without post-processing. In this case, the combination of the best models generated more accurate forecasts than the best model on its own.

ACKNOWLEDGMENTS

We are grateful for the financial support from the Dirección General de Investigación, Innovación y Postgrado (DGIIP), Universidad Técnica Federico Santa María, Valparaíso, Chile, through the “Programa de Incentivos a la Iniciación Científica (PIIC)” initiative.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

ORCID

Werner D. Kristjanpoller https://orcid.org/0000-0002-5878-072X

BIBLIOGRAPHY

Abraham, A., Nath, B., & Mahanti, P. K. (2001). Hybrid intelligent systems for stock market analysis. In *International Conference on Computational Science* (pp. 337–345). Berlin, Heidelberg: Springer.

Adhikari, R. (2015). A mutual association based nonlinear ensemble mechanism for time series forecasting. *Applied Intelligence*, *43*(2), 233–250.

Alvarez-Diaz, M., Bissoondeeal, R., Binner, J., Bhuruth, M., Gazel, A., & Mootanah, V. (2008). Forecasting exchange rates with linear and nonlinear models. *Applied Economics*, *40*, 1969–1984.

Antonakakis, N., & Badinger, H. (2016). Economic growth, volatility, and cross-country spillovers: New evidence for the g7 countries. *Economic Modelling*, *52*, 352–365.

Antonakakis, N., Dragouni, M., & Filis, G. (2015a). How strong is the linkage between tourism and economic growth in europe? *Economic Modelling*, *44*, 142–155.

Antonakakis, N., Dragouni, M., & Filis, G. (2015b). Tourism and growth: The times they are a-changing. *Annals of Tourism Research*, *50*, 165–169.

Babu, C. N., & Reddy, B. E. (2014). A moving-average filter-based hybrid Arima–ann model for forecasting time series data. *Applied Soft Computing*, *23*, 27–38.

Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. *Or*, *20*(4), 451–468.

Baxt, W. G. (1992). Improving the accuracy of an artificial neural network using multiple differently trained networks. *Neural Computation*, *4*(5), 772–780.

Bildirici, M., Alp, E. A., & Ersin, Ö. Ö. (2010). Tar-cointegration neural network model: An empirical analysis of exchange rates and stock returns. *Expert Systems with Applications*, *37*(1), 2–11.

Chen, A.-S., & Leung, M. T. (2004). Regression neural network for error correction in foreign exchange forecasting and trading. *Computers & Operations Research*, *31*(7), 1049–1068.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*(4), 559–583.

Clements, K. W., & Lan, Y. (2010). A new approach to forecasting exchange rates. *Journal of International Money and Finance*, *29*(7), 1424–1437.

Davis, J. T., Episcopos, A., & Wettimuny, S. (2001). Predicting direction shifts on Canadian–us exchange rates with artificial neural networks. *Intelligent Systems in Accounting, Finance and Management*, *10*(2), 83–96.

Dhamija, A., & Bhalla, V. (2011). Exchange rate forecasting: Comparison of various architectures of neural networks. *Neural Computing and Applications*, *20*(3), 355–363.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*(366a), 427–431.

Dunis, C. L., & Huang, X. (2002). Forecasting and trading currency volatility: An application of recurrent neural regression and model combination. *Journal of Forecasting*, *21*(5), 317–354.

Engle, R. F., & Granger, C. W. (1987). Co-integration and error correction: Representation, estimation, And Testing. *Econometrica: journal of the Econometric Society*, *55*(2), 251–276.

Eswaran, C., & Logeswaran, R. (2012). An enhanced hybrid method for time series prediction using linear and neural network models. *Applied Intelligence*, *37*(4), 511–519.

Fatima, S., & Hussain, G. (2008). Statistical models of kse100 index using hybrid financial systems. *Neurocomputing*, *71*(13), 2742–2746.

Galeshchuk, S. (2016). Neural networks performance in exchange rate prediction. *Neurocomputing*, *172*, 446–452.

Galeshchuk, S., & Mukherjee, S. (2017). Deep networks for predicting direction of change in foreign exchange rates. *Intelligent Systems in Accounting, Finance and Management*, *24*(4), 100–110.

Granger, C. W. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, *16*(1), 121–130.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics*, *23*(4), 365–380.

Hansen, P. R., Lunde, A., & Nason, J. M. (2003). Choosing the best volatility models: The model confidence set approach. *Oxford Bulletin of Economics and Statistics*, *65*, 839–861.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366.

Hosseini, H. G., Luo, D., & Reynolds, K. J. (2006). The comparison of different feed forward neural network architectures for ecg signal diagnosis. *Medical Engineering & Physics*, *28*(4), 372–378.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, *12*(2–3), 231–254.

Johansen, S., & Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration—With applications to the demand for money. *Oxford Bulletin of Economics and Statistics*, *52*(2), 169–210.

Khashei, M., & Bijari, M. (2010). An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with Applications*, *37*(1), 479–489.

Khashei, M., & Bijari, M. (2011). A novel hybridization of artificial neural networks and Arima models for time series forecasting. *Applied Soft Computing*, *11*(2), 2664–2675.

Khashei, M., Bijari, M., & Ardali, G. A. R. (2009). Improvement of autoregressive integrated moving average models using fuzzy logic and artificial neural networks (anns). *Neurocomputing*, *72*(4), 956–967.

Kilian, L., & Taylor, M. P. (2003). Why is it so difficult to beat the random walk forecast of exchange rates? *Journal of International Economics*, *60*(1), 85–107.

Kodogiannis, V., & Lolis, A. (2002). Forecasting financial time series using neural network and fuzzy system-based techniques. *Neural Computing & applications*, *11*(2), 90–102.

Kuan, C.-M., & White, H. (1994). Artificial neural networks: An econometric perspective*. *Econometric Reviews*, *13*(1), 1–91.

Laboissiere, L. A., Fernandes, R. A., & Lage, G. G. (2015). Maximum and minimum stock price forecasting of Brazilian power distribution companies based on artificial neural networks. *Applied Soft Computing*, *35*, 66–74.

Lam, M. (2004). Neural network techniques for financial performance prediction: Integrating fundamental and technical analysis. *Decision Support Systems*, *37*(4), 567–581.

Leigh, W., Purvis, R., & Ragusa, J. M. (2002). Forecasting the nyse composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: A case study in romantic decision support. *Decision Support Systems*, *32*(4), 361–377.

León, C., Machado, C., & Murcia, A. (2016). Assessing systemic importance with a fuzzy logic inference system. *Intelligent Systems in Accounting, Finance and Management*, *23*(1–2), 121–153.

Leung, M. T., Chen, A.-S., & Daouk, H. (2000). Forecasting exchange rates using general regression neural networks. *Computers & Operations Research*, *27*(11), 1093–1110.

Li, W., Luo, Y., Zhu, Q., Liu, J., & Le, J. (2008). Applications of ar*-grnn model for financial time series forecasting. *Neural Computing and Applications*, *17*(5–6), 441–448.

Meese, R., & Rogoff, K. (1983). Empirical exchange rate models of the seven- ties: Do they fit out of sample? *Journal of International Economics*, *14*, 3–24.

Muzhou, H., Taohua, L., Yunlei, Y., Hao, Z., Hongjuan, L., Xiugui, Y., & Xinge, L. (2017). A new hybrid constructive neural network method for impacting and its application on tungsten price prediction. *Applied Intelligence*, *47*(1), 28–43.

Pradeepkumar, D., & Ravi, V. (2017). Forecasting financial time series volatility using particle swarm optimization trained quantile regression neural network. *Applied Soft Computing*, *58*, 35–52.

Rivas, V., Parras-Gutíerrez, E., Merelo, J., Arenas, M., & García-Fernández, P. (2017). Time series forecasting using evolutionary neural nets implemented in a volunteer computing system. *Intelligent Systems in Ac- Counting, Finance and Management*, *24*(2–3), 87–95.

Rumelhart, D. E., Hinton, G. E., Williams, R. J., et al. (1988). Learning representations by back-propagating errors. *Cognitive Modeling*, *5*(3), 1.

Swanson, N. R., & White, H. (1997). Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. *International Journal of Forecasting*, *13*(4), 439–461.

Taskaya-Temizel, T., & Casey, M. C. (2005). A comparative study of autoregressive neural network hybrids. *Neural Networks*, *18*(5), 781–789.

Tsaih, R., Hsu, Y., & Lai, C. C. (1998). Forecasting s&p 500 stock index futures with a hybrid ai system. *Decision Support Systems*, *23*(2), 161–174.

Wedding, D. K., & Cios, K. J. (1996). Time series forecasting by combining rbf networks, certainty factors, and the box-Jenkins model. *Neurocomputing*, *10*(2), 149–168.

Yao, J., & Tan, C. L. (2000). A case study on using neural networks to perform technical forecasting of forex. *Neurocomputing*, *34*(1), 79–98.

Yu, L., Wang, S., & Lai, K. K. (2005). A novel nonlinear ensemble forecasting model incorporating glar and ann for foreign exchange rates. *Computers & Operations Research*, *32*(10), 2523–2541.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, *14*(1), 35–62.

Zhang, G. P. (2007). A neural network ensemble method with jittered training data for time series forecasting. *Information Sciences*, *177*(23), 5329–5346.

Zhang, Y., & Wu, L. (2009). Stock market prediction of s&p 500 via combination of improved bco approach and bp neural network. *Expert Systems with Applications*, *36*(5), 8849–8854.

How to cite this article: Parot A, Michell K, Kristjanpoller WD. Using Artificial Neural Networks to forecast Exchange Rate, including VAR-VECM residual analysis and prediction linear combination. *Intell Sys Acc Fin Mgmt*. 2019;1–13. https://doi.org/10.1002/isaf.1440