Exchange Rate Volatility Forecasting by Hybrid Neural Network Markov Switching Beta-t-EGARCH

RUOFAN LIAO[1], WORAPHON YAMAKA[ID][2], AND SONGSAK SRIBOONCHITTA[2]

[1]Faculty of Economics, Chiang Mai University, Chiang Mai 50200, Thailand
[2]Centre of Excellence in Econometrics, Faculty of Economics, Chiang Mai University, Chiang Mai 52000, Thailand

Corresponding author: Woraphon Yamaka (woraphon.econ@gmail.com)

ABSTRACT The motivation of this study is built from the previous research to find a way to enhance the forecast of advanced and emerging market currency volatilities. Given the exchange rate's nonlinear and time-varying characteristics, we introduce the neural networks (NN) approach to enhance the Markov Switching Beta-Exponential Generalized Autoregressive Conditional Heteroscedasticity (MS-Beta-t-EGARCH) model. Our hybrid model synthesizes these two approaches' advantages to predict exchange rate volatility. We validate the performance of our proposed model by comparing it with various traditional volatility forecasting models. In-sample and out-of-sample volatility forecasts are considered to achieve our comparison. The empirical results suggest that our hybrid NN-MS Beta-t-EGARCH outperforms the other models for both emerging and advanced market currencies.

INDEX TERMS Exchange rate volatility, neural networks, Markov-switching Beta-t-EGARCH.

I. INTRODUCTION

Volatility represents the degree to which variable changes over time, and it is an essential facet in risk evaluation of many essential economic tasks such as value at risk, financial asset pricing, and exchange rate [52]. Volatility has three significant characteristics in the financial area, namely, volatility clustering property [37], asymmetric property [10], and nonlinearity property [19]. These properties lead to uncertainty in financial time series. In the context of economic globalization, international transactions, and capital flow across the borders have increased. There is an indisputable fact that the foreign exchange market is a crucial factor affecting the transactions and capital flows. Thus, policymakers must understand the exchange rate volatility before making fiscal and monetary policy decisions, especially in those import-led and export-led countries.

In this respect, our study aims to forecast the exchange rate volatility of the three emerging markets' currencies, namely the Brazilian Real, Chinese Yuan, and Indian Rupee, and three advanced markets' currencies, namely, Euro,

The associate editor coordinating the review of this manuscript and approving it for publication was Francesco Benedetto[ID].

Japanese Yen, and Pound Sterling. These currencies are selected as there are the top traded currencies in recent years. We use six currencies data for the investigation on the premise that they can function as the accurate predictors of exchange rate risk on a worldwide scale, allowing investors and exporting-importing firms to capture the volatility and prepare themselves against the risks.

Advanced and emerging countries have different development stages and have different positions in the world industrial chain; therefore, their innate characteristics make them need to make a suitable exchange rate system and exchange rate policy according to the economic development level and the opening degree. Advanced economies tend to adopt flexible systems while Emerging economies are characterized by their trade openness, economic development, foreign-currency liabilities, foreign exchange reserve holdings, economic size, export concentration ratios, and financial development [1]. Therefore, there is still a difference between exchange rate behaviors in both advanced and emerging countries. Although countries have similar macroeconomic fundamentals and adopted the freely floating exchange rates, the volatility of the exchange rate may vary significantly from one country to another [33], and scholars have attempted

to find the best method for forecasting the emerging and advanced currencies; however, inclusive results have been obtained [7], [16], [17], [20], [26], [29], [31], [32].

In the past few decades, there are several periods of high market volatility, such as the Asian financial crisis in 1997, the United States' financial crisis, the European debt crisis, and Japan's dotcom bubble. Turner *et al.* [5] noted that these crises are the structural change sources in the volatility persistence, thereby leading to regime-switching in the parameters in the variance equation. Thus, they suggested using the Markov Switching GARCH (MS-GARCH) model to cope with the exchange rate's low- and high-volatility regime. The application of this model can be found in the work of la Torre-Torres *et al.* [34]. Although this model has shown good performance in volatility forecasting, it seems not to take into account the asymmetric leverage effects for the conditional volatility. Thus, in this study, we consider Markov Switching (MS) Beta-t-EGARCH, a relatively recent volatility model introduced by Blazsek and Ho [39].

Furthermore, we consider adding the neural network's method in our volatility forecasting endeavor to recognize the similarity and close relationship between this method and many statistical techniques and its other usefulness. The neural networks' superior ability is to learn and generalize the pattern or characteristics of data [53]. More specifically, neural networks can learn the nonlinear and complex relationship among variables without model specification and the hidden relationship in the data without imposing any fixed relationships. Besides, after learning from historical data, various NN models can generalize and predict unobservable or future data. Therefore, neural network as a form of artificial intelligence inspires the present researchers with fresh consideration and cognition on volatility forecasting.

Consequently, the conventional regime-switching volatility models, which do not take into consideration the neural network's method, may become obsolete in modeling volatility in the exchange rate and predicting the complicated phenomena. To deal with this problem, the exchange rate volatility can be modeled by combining both the conditional mean and the conditional variance with nonlinear time series and a neural network model to achieve possible gains in forecasting and modeling capabilities. In other words, in this study, the MS-Beta-t-EGARCH will be incorporated with the multi-layer perceptron (MLP) type of neural networks model. As suggested by Bildirici and Ersin [23], [25], augmenting the regime-switching GARCH model with multi-layer perceptron (MLP) type of neural networks would enhance forecasting capabilities. Specifically, this study proposes a hybrid forecasting model that combined MLP and Markov Switching Beta-t-EGARCH.

This paper contributes to the literature in two aspects. First, to the best of our knowledge, this is the first attempt to extend the MS Beta-t-EGARCH to the neural networks model. Second, this paper's results provide a way to improve the prediction of exchange rate volatility, and these findings are useful for the related issues of currencies, such as international trade, foreign income and expenses, and foreign exchange rate management.

This paper will be organized as follows. We review the previous research in Section 2. In Section 3, the regime-switching Beta-t-EGARCH with multi-layer perceptron (MLP) type of neural network models will be explained in detail. Section 4 is the data description. Section 5 explains different models' empirical results and compares these models' performance based on the loss functions. Section 6 makes a summary.

II. RELATED WORKS

Scholars have studied the characteristics and behavior of exchange rate volatility extensively. Among the econometric methods, the standard GARCH (1,1) (SGARCH) of Bollerslev [44] is most widely used for volatility forecasting in the foreign exchange market (see [20], [21], [43]). However, it is unable to capture asymmetric effects like the leverage effect, which is commonly found in asset return series. The Exponential GARCH (EGARCH) model of Nelson [9] was introduced to account for the asymmetric volatility. The advantage of the EGARCH model over the basic GARCH (1,1) is that it can capture the asymmetric effect or leverage effect in financial time-series. Recently, the Beta-t-EGARCH model of Harvey and Sucarrat [2] was introduced. This model broke ARCH and GARCH terms' positive parameter assumptions, thereby improving the long-term forecasting performance.

However, the Beta-t-EGARCH still faces some limitations, and the conditional mean and conditional variance equations are based on the linear structure. This model specification is not consistent with recent financial data as the financial data usually present low- and high-volatility periods. Under this consideration of different volatility levels or regimes, Gray [40] introduced Markov Switching GARCH to cope with the financial data's structural change. Caporale and Zekokh [14] revealed that the forecasting performance of the MS-GARCH model is superior to that of the one-regime GARCH model. Recently, Blazsek and Ho [39] considered asymmetric leverage effects for conditional volatility and proposed MS-Beta-t-EGARCH. They compared the in-sample statistical performance of the MS-Beta-t-EGARCH model with that of the single-regime Beta-t-EGARCH model. The result showed that the two-regime model is better than the one-regime alternative.

Alternatively, many studies have pursued nonlinear modeling using neural network (NN) models. Galeshchuk [41] revealed that the Markov Switching models do not improve exchange rate forecasts. She also mentioned that NN could be an alternative approach that can produce a nonlinear model without prior knowledge about the functional forms. In addition, NN does not require assumptions regarding the distribution of data while it can learn from the data. The NN models have recently been applied to predict volatility series in many fields [11], [38], [45], [52]. In the case of the exchange rate volatility forecasting, Panda and Narasimhan [6] found that Neural Network (NN) has a better exchange rate

forecasting performance for not only in-sample but also out-of-sample period, compared to the linear regression and random walk models. Other researchers also find a similar result (see, [3], [13], [22]). Although the NN technique has several advantages that distinguish it from the other existing prediction methods, it is a black box learning approach because it cannot interpret the relationship between input and output or deal with uncertainties. Another major disadvantage of neural networks is that there is no formal systematic model-building approach. Fortunately, there is a new idea considering combining NN with the GARCH-type models.

Due to the existence of complex nonlinear correlation structure among variables and more massive data sets, the conventional GARCH-type models' prediction results may not be reliable [51]. Many papers discussed and proposed extending the NN approach to both the one-regime and two-regime GARCH-type models in the literature. Examples are NN-GARCH of Donaldson and Kamstra [38], Carvalho Griebeler [26], Kristjanpoller and Minutolo [47]; NN-EGARCH of Tseng *et al.* [8] and Lahmiri and Boukadoum [42]; NN-GARCH typed models of Bildirici and Ersin [23], NN-Integrating GARCH of Kristjanpoller and Minutolo [47]; NN-two-regime Markov Switching (MS)-ARMA-GARCH models of Bildirici and Ersin [24]; and NN-two-regime Smooth Transition(ST)-GARCH models of Bildirici and Ersin [25]. These studies combine GARCH-type models and the NN model by adding the NN structure to the existing GARCH-type equations. They confirmed that the hybrid neural network and GARCH-type models provide greater predictability than the traditional GARCH-type models alone. Kristjanpoller and Minutolo [47] mentioned that the NN model could capture volatile properties that could not be captured by the GARCH-type models alone. Besides, the NN model introduces a potential enhancement to earlier GARCH models since it can learn and manipulate erroneous, incomplete, or fuzzy data inputs [46]. This is to say, and it is not easy to have an accurate volatility prediction when using traditional GARCH models alone. The time-series are often influenced by different characteristics, such as high persistence in the conditional variance, nonstationary, and nonlinear structure of the volatility process.

Combining regime-switching GARCH-type models with neural network methods to deal with the nonlinear dynamics of volatility has been pursued limitedly in recent years. Bildirici and Ersin [24] used a neural network augmented MS-GARCH models to predict stock volatility. They concluded that neural network models are promising. Bildirici and Ersin [25] applied hybrid Smooth Transition-GARCH models and MLP-type neural networks to forecast petrol price volatility. Their results indicated significant improvement of this hybrid model over the traditional ST-GARCH models.

In a nutshell, there is evidence that its integration could improve the regime-switching GARCH-type models' forecasting performance with the NN model. However, there are still opportunities to improve the existing models and their

forecasting performance further. In this study, we proposed to combine the MS-Beta-t-EGARCH model with the NN model based on Multilayer Perceptron (MLP) to achieve improvement in terms of forecasting accuracy.

Nevertheless, to the best of our knowledge, no paper has tried to combine the NN model with the MS-Beta-EGARCH model yet. Thus, we aim to fill the research gap and enhance the MS Beta-t-EGARCH performance by combining it with the NN model. Specifically, we take advantage of the NN to improve the forecasting ability of the MS-Beta-EGARCH approach. To do this, we extend the MLP typed neural networks to the MS-Beta-t-EGARCH (1,1) model. Thus, we have NN-MS-Beta-t-EGARCH (1,1) to depict the essential stylized facts about asymmetries, volatility clustering, or mean reversion in different financial regimes. We expect that our model would be likely to bring about an increase in forecasting capabilities.

To illustrate that our model may work in exchange rate volatility forecasting since deep learning requires using high-performance computing access, we show it on the simplified example of a shallow NN (single hidden layer). We show that with a shallow NN, combining it with MS-Beta-t-EGARCH leads to better predictions than using NN by itself (and that traditional MS-Beta-t-EGARCH).

III. METHODOLOGY
In this section, we present the approaches related to our proposed NN-MS-Beta-t-EGARCH model.

A. BETA-t-GARCH MODEL
The Beta-t-EGARCH model of Harvey and Chakravarty [2] is a particular type of the Dynamic conditional score model. It is an extended version of the EGARCH model that uses the conditional score of beta-t-distribution to estimate the conditional variance. Thus, the model becomes more robust against outliers. The formulation also follows Nelson's conventional EGARCH model [8], where the conditional volatility is log-transformed. However, this model does not restrict the ARCH and GARCH parameters to be positive. The Beta-t-EGARCH model is constructed by two components, the conditional mean and the conditional volatility, and the model is driven by the conditional scale parameter λ_t. The Beta-t-EGARCH (1,1) model considered in this study is formulated as:

$$y_t = u_t + \varepsilon_t = u_t + \exp(\lambda_t)z_t, \qquad (1)$$

where μ_t is the constant term. The conditional volatility with leverage effects is expressed as:

$$\lambda_t = \omega + \alpha e_{t-1} + \beta \lambda_{t-1}, \qquad (2)$$

where $e_t = ((v+1)b_t) - 1$ and $b_t = (\varepsilon_t^2/v \cdot \exp(2\lambda_t))/(1 + \varepsilon_t^2/v \cdot \exp(2\lambda_t))$. α and β are ARCH effect and GARCH effect coefficients, respectively. z_t is the standardized residual which is assumed to follow the student-t distribution with mean zero, variance 1 and degree of freedom $v > 2$. To satisfy the covariance stationarity of Beta-t-EGARCH (1,1),

we restrict $|\beta| < 1$. Then the conditional volatility is computed as

$$\sigma_t = \exp(\lambda_t)\sqrt{v/(v-2)}. \tag{3}$$

B. MARKOV SWITCHING BETA-t-EGARCH MODEL

One of the Markov Switching model's key characteristics is that it allows all or some parameters to switch across different regimes. Thus, the stylized facts in financial series with two volatility regimes (i.e., high and low) can be captured under a Markov chain process hidden in the regime-switching. The MS-Beta-t-EGARCH of Blazsek and Ho [39] is formulated as:

$$y_t = u(S_t) + \varepsilon_t(S_t) = u(S_t) + \exp\left(\lambda_t(S_t)\right)z_t(S_t) \tag{4}$$

The logarithm of regime dependent conditional scale of volatility $\lambda_t(s_t)$ is expressed as

$$\lambda_t(S_t) = \omega(S_t) + \alpha(S_t)e_{t-1}(S_t) + \beta(S_t)\lambda_{t-1}(S_t), \tag{5}$$

$$e_t(S_t) = ((v(S_t) + 1)b_t(S_t)) - 1, \tag{6}$$

$$b_t(S_t) = \left(\varepsilon_t^2(S_t)/v(S_t) \cdot \exp(2\lambda_t(S_t))\right) / \left(1 + \varepsilon_t^2(S_t)\right. \\ \left. /v(S_t) \cdot \exp(2\lambda_t(S_t))\right). \tag{7}$$

Thus, the estimated parameters varying across regimes in this model are $\omega(S_t)$, $\alpha(S_t)$, $\beta(S_t)$ and $v(S_t)$. $S_t \in (0, 1)$ indicates the realization of the two-state Markov Chain, which is the probabilistic structure of the switching regime indicator and is governed by the first-order Markov process with a constant transition probability matrix P. P is determined by two parameters: p_{00} and p_{11}, taking values between 0 and 1. Thus,

$$P = \begin{bmatrix} \pi(S_t = 0 \,|S_{t-1} = 0) & \pi(S_t = 1 \,|S_{t-1} = 0) \\ \pi(S_t = 0 \,|S_{t-1} = 1) & \pi(S_t = 1 \,|S_{t-1} = 1) \end{bmatrix}, \tag{8}$$

$$P = \begin{bmatrix} p_{00} & 1 - p_{00} \\ 1 - p_{11} & p_{11} \end{bmatrix}, \tag{9}$$

where p_{00} and p_{11} are the transition probabilities from regime 0 to regime 0, and regime 1 to regime 1, respectively. To find the optimal parameters in this model, the maximum likelihood estimation (MLE) is used. Therefore, the full conditional likelihood function is specified as

$$L(\Theta(y_t \,|S_t = j)) = f(y_t \,|\Theta(S_t = 0))\Pi(S_t = 0 \,|y_{t-1}) \\ + f(y_t \,|\Theta(S_t=1))\Pi(S_t=1|y_{t-1}), \tag{10}$$

where the $f(y_t \,|\Theta(S_t = j))$ and $\Pi(S_t = j \,|y_{t-1})$ are, respectively, the regime-dependent conditional density of y_t and the filtered probability of the regimes $j = 0, 1$. $\Theta(S_t)$ is the set of the regime-dependent parameters.

C. ARTIFICIAL NEURAL NETWORK

ANN is a network of artificial neurons that can receive inputs, change their internal states according to the inputs, and then compute outputs based on the inputs and internal states. These artificial neurons have weights that can be modified by a process called learning [49]. The ANN model can be presented as

$$y_t = \tilde{g}(\widehat{g}(x_iw^I + b^I)w^O + b^O), \tag{11}$$

where $\tilde{g}$ and $\widehat{g}$ are the output and input activation functions, respectively. y_t and x_i are the output and input, respectively. b^I and b^O are the bias term of input and output layers, respectively. w^I is the weight vector between the hidden layer and the input layer, while w^O is the weight vector between the hidden layer and the output layer. Fig. 1 illustrates the architecture of the ANN.
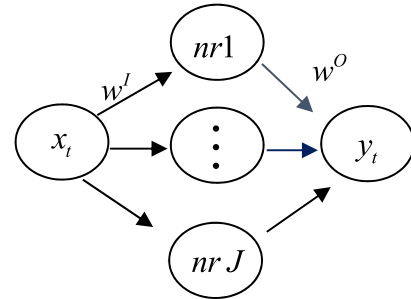

FIGURE 1. The ANN architecture. Note that *nr* is a neuron.

D. MARKOV SWITCHING BETA-t-EGARCH WITH NEURAL NETWORK MODEL AND ESTIMATION

In this section, we extend the MLP models that belong to the ANN family to the Markov Switching Beta-t-EGARCH. In this respect, the conditional mean and conditional variance processes augmented with MLP typed neural network are regime dependent. By combining the neural networks with the Markov Switching model, Bildirici and Ersin [23] confirmed the model's better performance in terms of in-sample and out-of-sample volatility forecasting accuracy. The formulation of this model is similar to the formal Markov switching Beta-t-EGARCH in Eqs. (4-6); however, the estimation of the logarithm of the conditional scale $\lambda_t(S_t)$ is different, such that

$$\lambda_t(S_t) = \omega(S_t) + \alpha(S_t)e_{t-1} + \beta(S_t)\lambda_{t-1}(S_t) \\ + \tilde{g}\left(\sum_{h=1}^{h}\xi_h(S_t)\widehat{g}\left(z_t(S_t)\delta_h(S_t)\right)\right), \tag{12}$$

where $\xi_h(S_t)$ is the additional regime-dependent output weight in the neural network. $\tilde{g}(\cdot)$ is the linear activation function of the output layer. $\widehat{g}(z_t(S_t)\delta_h(S_t))$ is the regime-dependent multi-layer perceptron neural network (MLP) that possesses h hidden neurons in each regime, which can be written as

$$\widehat{g}(z_t(S_t)\delta_h(S_t)) \\ = \left(1 + \exp\left(\delta_{h,d}(S_t) + \sum_{d=1}^{1}\left[\sum_{h=1}^{h}\delta_{h,d}(S_t)z_{t-d,h}(S_t)\right]\right)\right)^{-1} \tag{13}$$

$$\delta_{h,d}(S_t) \sim U(-1, +1) \tag{14}$$

where $z_{t-d} = (\varepsilon_{t-d}(S_t) - E(\varepsilon_t(S_t)))/\sqrt{E(\varepsilon_t(S_t)^2)}$ is the standardized residual at lag d(input) in the log-sigmoid activation function $\widehat{g}(\cdot)$ and $\delta_{h,d}(S_t)$ is the additional regime-

dependent input weight. We note that three layers, namely one input layer, one hidden layer, and one output layer, are considered in the multi-layer perceptron neural network. To have a simple estimation, one neuron is assumed ($h = 1$).

In the estimation point of view, the maximum likelihood estimation is also used to estimate all unknown parameters in MS Beta-t-EGARCH. Thus, the log of the regime-dependent conditional likelihood of y_t is written as

$$
\begin{aligned}
&L(y_t \mid \Phi(S_t = j)) \\
&= \log f(y_t \mid \Phi(S_t = 0))\Pi(S_t = 0 \mid y_{t-1}) \\
&\quad + \log f(y_t \mid \Phi(S_t = 1))\Pi(S_t = 1 \mid y_{t-1}),
\end{aligned}
\tag{15}
$$

where $\Phi(S_t) = \{\Theta(S_t), \xi_h(S_t), \delta_{h,d}(S_t)\}$ is all unknown parameter and $f(y_t \mid \Theta(S_t = j))$ is the conditional density of y_t, which can be defined as

$$
\begin{aligned}
&\ln f(y_t \mid y_1, \dots, y_{t-1}, S_t) \\
&= \ln \Gamma[\frac{v(S_t + 1)}{2}] - \ln \Gamma[\frac{v(S_t)}{2}] - \lambda_t(s_t) - \frac{\ln[\pi v(S_t)]}{2} \\
&\quad - \frac{v(S_t) + 1}{2} \ln \left\{ 1 + \frac{[y_t - \mu(S_t)]^2}{v(S_t)\exp[2\lambda_t(S_t)]} \right\},
\end{aligned}
\tag{16}
$$

where $\Gamma(\cdot)$ is the gamma distribution. $\Pi(S_t = j \mid y_{t-1})$ is the filtered probability, which can be filtered by Hamilton's filter [18] as written below.

$$
\Pi(S_t = j \mid y_{t-1}) = \frac{f(\Phi(y_t \mid S_t = j))\Pi(S_t = j \mid y_{t-1})p_{jj}}{\sum\limits_{j=1}^{2} f(\Phi(y_t \mid S_t = j)\Pi(S_t = j \mid y_{t-1})p_{jj}}.
\tag{17}
$$

Therefore, the maximum likelihood estimator for the Markov Switching Beta-t-EGARCH Model with NN model is given by

$$
\widehat{\Phi} = \arg \max_{\Phi} \sum_{t=1}^{T} L(\Phi(S_t = j) \mid y_{t-1}).
\tag{18}
$$

We note that the output weight parameter is estimated by minimizing the sum of the square's residuals of the volatilities, $\frac{1}{T}\sum_{i=1}^{T}\left(\sigma_t - \widehat{\sigma}_t\right)^2$, whereas $\widehat{\sigma}_t$ is the actual volatility.

E. MODEL SELECTION
Although AIC and BIC are goodness-of-fit measures of the models, they are different for some reasons. AIC gives prediction upon the assumption of maximum likelihood, while BIC consists of a likelihood term and a penalty term. The penalty term depends on the number of free parameters in the trained model for different states. However, it is hard to justify the best model when the competing models' AIC values are not much different; hence, the raw AIC and BIC may lead to ambiguous interpretation. AIC weights (AICw) and BIC weights (BICw) of Wagenmakers and Farrell [12] are introduced as the alternative model selection methods in this study. Note that these statistics are referred to as

conditional probabilities for each model. The formulas of AICw and BICw are as follows:

$$
(\text{AICw}) = \frac{\exp\left\{-\frac{1}{2}\Delta_i(\text{AIC})\right\}}{\sum\limits_{k=1}^{K} \exp\left\{-\frac{1}{2}\Delta_k(\text{AIC})\right\}},
\tag{19}
$$

$$
(\text{BICw}) = \frac{\exp\left\{-\frac{1}{2}\Delta_i(\text{BIC})\right\}}{\sum\limits_{k=1}^{K} \exp\left\{-\frac{1}{2}\Delta_k(\text{BIC})\right\}},
\tag{20}
$$

where $\Delta_i(\text{AIC}) = \text{AIC}_i - \min \text{AIC}$ and $\Delta_i(\text{BIC}) = \text{BIC}_i - \min \text{BIC}$. The minAIC andminBICare the minimum AIC and BIC of the best model. We note that these two weight statistics can be viewed as the probability that model i (M_i) is the best.

F. FORECASTING EVALUATION
The Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are also used to evaluate the prediction error rates and model performance in volatility forecasting analysis. These measures can be computed by

$$
MAE = \frac{1}{N}\sum_{i=1}^{N}|\sigma_i - \widehat{\sigma}_i|
\tag{21}
$$

$$
RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\sigma_i - \widehat{\sigma}_i\right)^2}
\tag{22}
$$

where $\widehat{\sigma}_i$ is the estimated forecasting volatility and σ_i is the realized volatility. Note that the realized volatility is measured by $\sigma_i = |y_i - E(y_i).|$

IV. DATA DESCRIPTION
The data sets analyzed in this paper are the financial time-series of the daily US Dollar to Chinese Yuan Renminbi (CNY), US Dollar to Japanese Yen (JPY), US Dollar to Euro (EUR), US Dollar to Indian Rupee (INR), US Dollar to the British Pound (GBP), and US Dollar to Brazilian Real (BRL) covering 3rd January 2007 to 31st December 2019, consisting of consecutive 3390 observations. The data is obtained from Thomson Reuters DataStream database. We should mention that the data set for each currency is subdivided into two subsets, say in-sample and out-of-sample. In-sample subsets are used for model training with data from 3rd January 2007 to 10th December 2019, whereas out-of-sample subsets are used for testing the model prediction ability with data from 11th December 2019 to 31st December 2019. We also note that two different prediction horizons, namely T=1, and 21, are considered.

These exchange rate series are chosen because the corresponding advanced and emerging currencies are those highly traded in the world. Note that the US dollar is not included in our analysis as it is used as the benchmark currency for those six currencies.

Table 1 and Fig. 2 show the basic data information and a graph of the six currencies' returns, respectively. We can

TABLE 1. Data description.

|	CNY	JPY	EUR	INR	GBP	BRL
Min	-1.150	-4.610	-3.056	-3.064	-4.474	-11.033
Max	1.810	3.710	3.725	3.251	8.311	10.951
Mean	-0.003	-0.003	-0.005	0.014	0.011	-0.018
Median	0.000	0.000	0.000	0.000	0.000	-0.014
Skewness	0.600	-0.253	0.021	0.250	0.695	0.233
Kurtosis	13.564	4.985	2.757	4.897	13.179	93.424
Std (%)	15.828	63.709	59.706	44.505	62.484	76.635
Jarque-Bera	25399[a]	3600[a]	1098[a]	3466[a]	24895[a]	34073[a]
ARCH-LM	3.597[a]	5.158[a]	3.211[a]	7.288[a]	5.245[a]	4.413
ADF	·12.705[a]	14.595[a]	14.429[a]	13.374[a]	14.389[a]	-15.091[a]
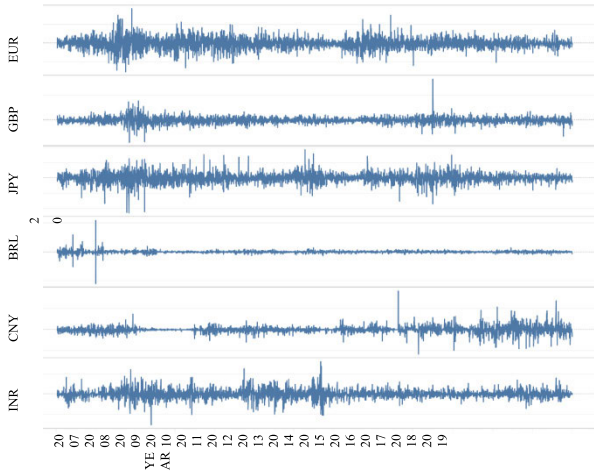
FIGURE 2. Returns for six currencies during the period from 3rd January 2007 to 31st December 2019.

observe that the INR shows the highest average return (0.0014), followed by GBP's average return (0.0012). BRL return has the highest standard deviation (76.635%), followed by JPY (63.709%) and GBP (62.484%). This evidence suggests that the British Pound is characterized by high return and high risk. Positive skewness is observed in all series, except for JPY. High kurtosis is also observed in all series. These results indicate that our currency returns are leptokurtic, with a longer distribution and fatter tails. In other words, the kurtosis and skewness statistics show that there had been a deviation from normality in the currency returns. We thus conduct the Jarque–Bera test to confirm the non-normal distribution. Jarque–Bera test strongly suggests that the null hypothesis of the normal distribution for daily returns can be rejected. Furthermore, the ARCH-LM heteroscedasticity test is used to verify the ARCH effects in our currency returns. The result of heteroscedasticity cannot be rejected for the daily returns series. Then, we also employ the Augmented Dickey-Fuller test to examine our series's stationarity, and the result shows strong evidence supporting the stationarity of our series.

As shown in Fig 2, there are low-volatility and high-volatility periods along 2007-2019. This indicates that the variance of the return of currencies is not stationary and presents the volatility clustering, confirming the GARCH-type models' use. We can observe that most of the currencies

experience high volatility around 2008-2009, corresponding to the US's global financial crisis. This figure suggests that all currencies are generated by a sequence of random variables with regime-switching probability distributions. Therefore, it might be appropriate to forecast these six currencies' volatility by using the Markov Switching Beta-t-EGARCH model.

V. EMPIRICAL RESULTS

The parameter estimation, in-sample volatility forecast, and out-of-sample forecast of the NN-MS-Beta-t-EGARCH and other baseline models (Beta-t-EGARCH, NN-Beta-t-EGARCH, and MS-Beta-t-EGARCH) were undertaken. We also considered the previous hybrid deep learning models, namely the Neural Network (NN)-MS-GARCH of Bildirici and Ersin [23], and the Neural Network (NN)-GARCH of Kristjanpoller and Minutolo [48]. In addition, the neural network without GARCH (NN) of Malliaris and Salchenberger [28] was also considered as other competing models. By doing this, we can investigate the performance of the hybrid models and determine whether improvements of MS-Beta-E-GARCH can be achieved in the forecasting of exchange rate volatility by combining a neural network method with MS-Beta-t-EGARCH. Then, several criteria were used to evaluate the performance of these models.

A. ESTIMATION RESULTS FOR DIFFERENT VOLATILITY MODELS

At the first stage, we present the parameter estimates of our proposed NN-MS-Beta-EGARCH model and baseline models (MS-Beta-EGARCH and Beta-EGARCH). The estimation of NN-MS-Beta-EGARCH model for six currencies is provided in Table 2, and two baseline models for six currencies are reported in Table 3. The last two rows show the results of the diagnostic test on the standardized residuals. According to the results, it can be summarized as in the following: First, the results reported in Tables 2 and 3 show that the GARCH effect of the first and second regimes, $\beta(0)$ and $\beta(1)$, are close to 1 for all currencies, except for the first regime of INR. This indicates that these exchange rate markets experience a high degree of volatility persistence in both regimes. Second, the measures of the unconditional variance $\alpha(0) + \beta(0)$ for regime 1 and $\alpha(1) + \beta(1)$ for regime 2 show high persistence for all the returns in both regimes. Besides, the summation of these ARCH and GARCH effects are less than one satisfying the stationary covariance condition. Third, the transition matrix of the Markov Switching model is estimated; and we observe that $p_{00} > 0.9$ and $p_{11} > 0.9$ are observed in all currencies, except for INR, indicating that the probability of staying in the same regime is larger than 90% and there is only a 10% chance of switching from one to the other regime. Finally, the diagnostic tests are presented in the last two rows: Ljung–Box autocorrelation test and ARCH effects test. The result shows that the Ljung–Box test strongly accepts the null hypothesis of no serial correlation of the standardized residual. Furthermore, the ARCH effects reveal that we cannot

TABLE 2. Parameter estimates results of NN-MS-Beta-t-EGARCH model.

Par.	GBP	JPY	EUR	INR	CNY	BRL
$\mu_t(0)$	-0.00001	0.0001	0.0001	-0.0001	-0.0001	-0.0001
	(0.00001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
$\mu_t(1)$	0.00001	0.0002	0.0001	-0.0001	0.0001	0.0001
	(0.00001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
$\omega(0)$	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
	(0.00001)	(0.0001)	(0.0001)	(0.0001)	(0.0001	(0.0001
$\omega(1)$	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002
	(0.00001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
$\alpha(0)$	0.0128	0.0227	0.0311	0.1213	0.0301	0.0209
	(0.0035)	(0.0125)	(0.0138)	(0.0435)	(0.0021)	(0.0028)
$\alpha(1)$	0.0485	0.0884	0.0358	0.0354	0.0453	0.0501
	(0.0215)	(0.0221)	(0.0112)	(0.0121)	(0.0019)	(0.0012)
$\beta(0)$	0.9541	0.9054	0.8548	0.7113	0.9854	0.9671
	(0.0125)	(0.0112)	(0.0033)	(0.0855)	(0.0259)	(0.0200)
$\beta(1)$	0.9036	0.9033	0.9188	0.9158	0.8764	0.8385
	(0.0289)	(0.0413)	(0.0159)	(0.0032)	(0.0659)	(0.0152)
$v(0)$	6.4154	3.6544	6.6858	7.6641	2.3394	5.6588
	(0.5841)	(1.0225)	(0.9564)	(1.3554)	(0.589)	(1.3259)
$v(1)$	8.1259	4.6585	7.6698	5.3658	3.4482	5.1051
	(1.3551)	(1.8424)	(1.0315)	(1.5126)	(0.9988)	(1.2551)
p_{11}	0.9789	0.9425	0.9368	0.5038	0.9984	0.9549
	(0.0022)	(0.0015)	(0.0113)	(0.0014)	(0.0324)	(0.213)
p_{22}	0.9598	0.9536	0.9874	0.7689	0.9360	0.9569
	(0.0033)	(0.0050)	(0.0053)	(0.0038)	(0.0101)	(0.1001)
$\xi_1(S_t)$	0.1558	0.2588	0.9917	0.4789	0.3221	0.4479
	(0.0018)	(0.0124)	(0.0010)	(0.0112)	(0.1158)	(0.1598)
$\delta_{1,1}(S_t)$	0.6554	0.1245	0.9870	-0.4987	0.6987	-0.1225
	(0.1158)	(0.0010)	(0.0052)	(0.0118)	(0.1003)	(0.2155)
ARCH(1) (p-value)	0.5847	0.9254	0.7845	0.3984	0.2884	0.8755
Ljung–Box (p-value)	0.6884	0.8887	0.3984	0.6844	0.8744	0.99847

Note: the parenthesis indicates standard error. $\xi_1(S_t)$ and $\delta_{1,1}(S_t)$ are output and input weights of the neural network part.

reject the null hypothesis of no serial correlation and the ARCH effects in the squared residual series.

One of the advantages of the MS-GARCH typed model is that it can be used to characterize and forecast the probability of being in a high- or low-volatility regime [34]. Therefore, the regime probabilities obtained from the MS-Beta-t-EGARCH are illustrated in Fig 3. According to Fig. 3, the high exchange rate volatility regime's smoothed probabilities are presented for our sample period from 2007 to 2019. We observe the high-volatility regime in some periods. It is observed that the probability of staying in the high-volatility regime of INR is close to one during the entire sample period, indicating the high volatility of INR during 2007-2019; however, the probability of staying in this regime is not persistent. We also notice that the probability that CNY stays in the high volatility is low during 2009-2010. This finding

is consistent with the Chinese government's intervention toward CNY appreciation because of the effects of the global economic crisis on China's exporters. For GBP, JPY, and BRL, the smoothed probabilities of all currencies display that upswings are abrupt and much shorter while downswings are more gradual and highly persistent. Among these three currencies, the most severely fluctuating one is JPY. Finally, three episodes of high volatility in the EUR are formally identified in the period 2007-2013, 2015, and 2017-2018, respectively. Specifically, the first period coincides with the global financial crisis in 2007-2009 and the European debt crisis during 2010-2012, the second period coincides with the negative interest rate announced by the European central bank and the bond purchased. Lastly, the third period corresponds to the unexpected robust economic growth in Europe that strengthened the EURO currency.

TABLE 3. Parameter estimates results of baseline models.

Par.	GBP Beta-t-EGARCH	GBP MS-Beta-t-EGARCH	JPY Beta-t EGARCH	JPY MS-Beta-t-EGARCH	EUR Beta-t-EGARCH	EUR MS-Beta-t-EGARCH	INR Beta-t-EGARCH	INR MS-Beta-t-EGARCH	CNY Beta-t-EGARCH	CNY MS-Beta-t-EGARCH	BRL Beta-t-EGARCH	BRL MS-Beta-t-EGARCH
$\mu_t(0)$	-0.0001 (0.000006)	-0.00001 (0.0001)	0.0001 (0.0001)	0.0001 (0.0001)	0.00001 (0.00001)	0.0001 (0.0001)	-0.0001 (0.0001)	-0.0001 (0.0001)	-0.0001 (0.0001)	-0.0001 (0.0001)	-0.0001 (0.0001)	-0.0001 (0.0001)
$\mu_t(1)$		0.00001 (0.00001)		0.0002 (0.0001)		0.0001 (0.0001)		-0.0001 (0.0001)		0.0001 (0.0001)		0.0001 (0.0001)
$\omega(0)$	0.00001 (0.00001)	0.0001 (0.0001)	0.0001 (0.0001)	0.0001 (0.0001)	0.0001 (0.0001)	0.0001 (0.0001)	0.00001 (0.00001)	0.0001 (0.0001)	0.00001 (0.00001)	0.0001 (0.0001)	0.00001 (0.00001)	0.0001 (0.0001
$\omega(1)$		0.0001 (0.00001)		0.0001 (0.0001)		0.0001 (0.0001)		0.0001 (0.0001)		0.0002 (0.0001)		0.0002 (0.0001)
$\alpha(0)$	0.1001 (0.0367)	0.0154 (0.0064)	0.0374 (0.0043)	0.0370 (0.0241)	0.0376 (0.0028)	0.0307 (0.0125)	0.1004 (0.0133)	0.1308 (0.0756)	0.0378 (0.0046)	0.0259 (0.0031)	0.0235 (0.0023)	0.0125 (0.0020)
$\alpha(1)$		0.0412 (0.0231)		0.0521 (0.0641)		0.0341 (0.0114)		0.0882 (0.0598)		0.0488 (0.0010)		0.0594 (0.0009)
$\beta(0)$	0.8972 (0.0324)	0.9150 (0.0033)	0.9595 (0.0029)	0.9560 (0.0040)	0.9609 (0.0022)	0.9652 (0.0011)	0.8962 (0.0119)	0.6042 (0.0962)	0.9564 (0.0040)	0.9708 (0.0314)	0.9784 (0.0035)	0.9877 (0.0258)
$\beta(1)$		0.8644 (0.0321)		0.8830 (0.0542)		0.9492 (0.0213)		0.9085 (0.0024)		0.8024 (0.0254)		0.8791 (0.0365)
$v(0)$	4.2654 (0.3749)	5.1581 (0.4792)	5.2496 (0.7042)	4.9506 (0.4745)	7.7849 (1.1992)	7.7198 (1.2315)	6.1675 (0.5039)	4.5315 (0.9621)	7.7577 (1.2135)	12.5202 (2.2516)	3.6545 (0.5231)	8.6410 (2.6551)
$v(1)$		10.3254 (1.5974)		6.3490 (1.4387)		11.6652 (1.6655)		8.6746 (1.4017)		6.2146 (1.3211)		4.6565 (2.1151)
p_{11}		0.9687 (0.0018)		0.9995 (0.0026)		0.9917 (0.0010)		0.5581 (0.0036)		0.9937 (0.0125)		0.9578 (0.0100)
p_{22}		0.9447 (0.0054)		0.9511 (0.0050)		0.9870 (0.0052)		0.7544 (0.0088)		0.9265 (0.0115)		0.9894 (0.0154)
ARCH(1) (p-value)	0.9785	0.9514	0.7588	0.9011	0.4247	0.6411		0.3218	0.7555	0.2158	0.8874	0.2158
Ljung–Box (p-value)	0.3800	0.5469	0.1075	0.8511	0.3741	0.3265		0.2551	0.4437	0.3321	0.6974	0.3321

Note: the parenthesis indicates standard error.

B. IN-SAMPLE STATISTICAL PERFORMANCE

In this section, the first forecasts that are performed are those from the conventional neural network and Beta-E-GARCH model to use them as a benchmark to compare with our proposed hybrid model. Specifically, six conventional volatility models, consisting of the artificial neural network (Model 1), the hybrid artificial neural network and GARCH model (Model 2), the hybrid artificial neural network and Markov Switching GARCH model (Model 3), Beta-t-EGARCH model (Model 4), the hybrid artificial neural network and Beta-t-EGARCH (Model 5), and Markov Switching Beta-t-GARCH model (Model 6), are compared with our proposed hybrid artificial neural network and Markov Switching Beta-t-EGARCH (Model 7). These models are compared in terms of AIC, BIC, AICw, BICw, MAE, and RMSE criteria for in-sample forecast performance.

In the second stage, to obtain the most accurate models, the model is further selected using the AIC, BIC, AICw, and BICw. For AIC and BIC, a lower value indicates a more parsimonious model. For AICw and BICw, a higher value indicates a better model fit. The model selection criteria results have arrived at the same conclusions for the in-sample results. We find that the regime-switching Beta-t-EGARCH model with a multi-layer perceptron (MLP) achieves a more reasonable degree of fit than the competing models for all six currencies. Besides, it is conspicuous that the regime-switching has improved the conditional volatility modeling performance as the Markov Switching model is superior to the single-regime alternative. To obtain robustness results, we need to select the model for in-sample performance and the forecasting ability. Hence, we should further compare the model performance by the out-of-sample volatility forecast.

C. IN-SAMPLE STATISTICAL PERFORMANCE

In the third stage, the prediction power of each model was compared by the out-of-sample performance. The world-famous economist Friedman [27] points out that ''the only relevant test of the validity of a hypothesis is a comparison of its predictions with experience.'' We need to compare the prediction result with the real data to validate the forecasting performance.

As we mentioned before, in this paper, we employ MAE and RMSE as loss functions for making the comparison, and a lower value of MAE and RMSE indicates a better prediction while a higher value might lead to more significant uncertainty in prediction.

Table 5 presents the out-of-sample forecast performance of four competing models. The forecast horizon is selected as 1 and 21 days ahead to evaluate the models' performances in short and long horizons. The results show a significant improvement of the Markov Switching Beta-t-EGARCH with NN models in both the short-term (1-day) and long-term (21-day) volatility forecasts for CNY, JPY, EUR, INR returns. The results in this table demonstrate the NN-MS-Beta-t-EGARCH for six currency volatilities have reduced MAE and RMSE. One crucial point is that the NN-based models' MAE and RMSE values are comparatively lower than expected. Starting from T+1 day to T + 21 days, the NN based models provide lower MAE and RMSE values. We also observe that the MAE and RMSE values for the NN-MS-Beta-t-EGARCH model obtained for long horizons are higher than short horizons.

Next, we compare the performance of the single-regime and two-regime models. The two-regime switching models, NN-MS-Beta-t-EGARCH and MS-Beta-t-EGARCH, outperform the single-regime counterparts as the lower MAE and
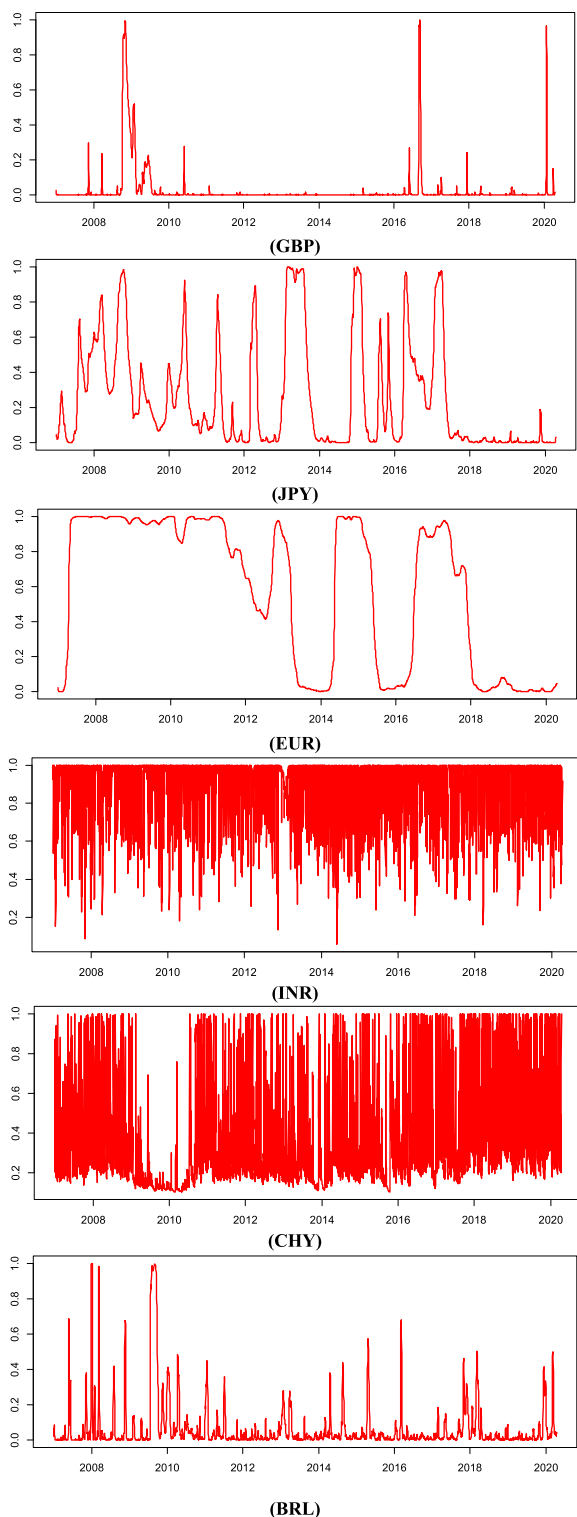
FIGURE 3. Smoothed probability of the high-volatility regime.

TABLE 4. Forecast evaluation: In-sample forecast performance.

Currency	CNY			
Model	AIC	BIC	AICw	BICw
Model 1	-20139.58	-20274.35	0.0000	0.0000
Model 2	-20441.85	-20464.33	0.0000	0.0000
Model 3	-24548.01	-24567.39	0.0000	0.0000
Model 4	-21111.02	-21135.92	0.0000	0.0000
Model 5	-22154.02	-22173.18	0.0000	0.0000
Model 6	-28254.26	-28191.99	0.0069	0.0038
Model 7	**-28264.21**	**-28203.14**	**0.9931**	**0.9962**
Currency	JPY			
Model	AIC	BIC	AICw	BICw
Model 1	-28101.65	-28144.63	0.0000	0.0000
Model 2	-28231.56	-28255.17	0.0000	0.0000
Model 3	-28330.97	-28364.19	0.0000	0.0000
Model 4	-28240.59	-28265.50	0.0000	0.0000
Model 5	-28260.31	-28285.39	0.0000	0.0000
Model 6	-28392.51	-28754.26	0.0138	0.0724
Model 7	**-28401.05**	**-28759.36**	**0.9862**	**0.9276**
Currency	EUR			
Model	AIC	BIC	AICw	BICw
Model 1	-27025.94	-27054.51	0.0000	0.0000
Model 2	-27441.39	-27466.55	0.0000	0.0000
Model 3	-28814.36	-28835.69	0.0000	0.0000
Model 4	-28783.07	-28807.97	0.0000	0.0000
Model 5	-28790.51	-28821.35	0.0000	0.0000
Model 6	-28917.21	-28979.49	0.0000	0.0005
Model 7	**-28969.31**	**-28994.64**	**1.0000**	**0.9995**
Currency	INR			
Model	AIC	BIC	AICw	BICw
Model 1	-25043.65	-25078.54	0.0000	0.0000
Model 2	-25314.55	-25336.47	0.0000	0.0000
Model 3	-25554.69	-25575.66	0.0000	0.0000
Model 4	-25461.64	-25486.55	0.0000	0.0000
Model 5	-25485.11	-25414.61	0.0000	0.0000
Model 6	-25489.04	-25491.31	0.0000	0.0000
Model 7	**-25814.31**	**-25814.10**	**1.0000**	**1.0000**
Currency	GBP			
Model	AIC	BIC	AICw	BICw
Model 1	-28001.54	-28040.51	0.0000	0.0000
Model 2	-28200.55	-28239.48	0.0000	0.0000
Model 3	-28414.87	-28465.18	0.0000	0.0000
Model 4	-28392.66	-28417.56	0.0000	0.0000
Model 5	-28401.36	-28433.14	0.0000	0.0000
Model 6	-28422.03	-28484.31	0.0000	0.0002
Model 7	**-28452.36**	**-28501.35**	**1.0000**	**0.9998**
	BRL			
Currency	AIC	BIC	AICw	BICw
Model 1	-31441.15	-31480.33	0.0000	0.0000
Model 2	-31364.32	-31403.98	0.0000	0.0000
Model 3	-32125.41	-32169.73	0.0000	0.0000
Model 4	-31413.78	-31457.41	0.0000	0.0000
Model 5	-32438.96	-32478.10	0.0000	0.0000
Model 6	-32199.14	-32241.33	0.0000	0.0000
Model 7	**-32847.58**	**-32886.54**	**1.0000**	**1.0000**

Note: Bold numbers suggest superior in-sample forecast performance. The AIC and BIC of the neural networks without GARCH are computed using the method of Qi and Zhang [30].

RMSE are obtained. This confirms the presence of the structural change in six currencies' volatilities.

According to the above results, our NN-MS-Beta-t-EGARCH improves the capability to model and forecast exchange rate returns and volatility over the baseline models, not only for the short term but also for the long-term

volatility forecasting. Hence, we conclude that a neural network approach can enhance the MS-Beta-t-EGARCH model. This study proves the generalization and forecasting power of the NN-MS-Beta-t-EGARCH model.

Finally, to validate our proposed model's forecasting performance, we compare our NN-MS-Beta-t-EGARCH model with two hybrid models, namely the Neural network-MS-GARCH and the Neural network GARCH. It is observed that our proposed model provided superior results relative to these hybrid models.

TABLE 5. Forecast evaluation: Out-of-sample forecast performance.

Currency		CNY		
	1 day		21 days	
Model	MAE	RMSE	MAE	RMSE
Model 1	0.00398	0.00561	0.00491	0.00887
Model 2	0.00378	0.00549	0.00458	0.00843
Model 3	0.00341	0.00454	0.00423	0.00815
Model 4	0.00353	0.00542	0.00431	0.00812
Model 5	0.00351	0.00540	0.00423	0.00815
Model 6	0.00341	0.00454	0.00424	0.00811
Model 7	**0.00340**	**0.00452**	**0.00415**	**0.00803**
Currency		JPY		
	1 day		21 days	
Model	MAE	RMSE	MAE	RMSE
Model 1	0.00501	0.00697	0.00601	0.08144
Model 2	0.00493	0.00671	0.00531	0.00783
Model 3	0.00443	0.00640	0.00509	0.00705
Model 4	0.00453	0.00652	0.00512	0.00734
Model 5	0.00453	0.00652	0.00512	0.00734
Model 6	0.00451	0.00650	0.00503	0.00699
Model 7	**0.00450**	**0.00649**	**0.00502**	**0.00698**
Currency		EUR		
	1 day		21 days	
Model 1	MAE	RMSE	MAE	RMSE
Model 2	0.00501	0.06123	0.00597	0.00701
Model 3	0.00443	0.00561	0.00565	0.00650
Model 4	0.00344	0.00460	0.00571	0.00688
Model 5	0.00351	0.00465	0.00598	0.00710
Model 6	0.00344	0.00464	0.00587	0.00699
Model 7	0.00344	0.00460	0.00566	0.00651
Model 1	**0.00344**	**0.00459**	**0.00518**	**0.00510**
Currency		INR		
	1 day		21 days	
Model	MAE	RMSE	MAE	RMSE
Model 1	0.40215	0.83554	0.60369	1.10255
Model 2	0.30100	0.63151	0.57810	0.82001
Model 3	0.30001	0.63058	0.55319	0.81132
Model 4	0.31910	0.64161	0.55810	0.81313
Model 5	0.29841	0.62775	0.55801	0.81307
Model 6	0.30558	0.63690	0.56325	0.81946
Model 7	**0.20560**	**0.40123**	**0.31245**	**0.63954**
Currency		GBP		
	1 day		21 days	
Model	MAE	RMSE	MAE	RMSE
Model 1	0.00398	0.00561	0.02511	0.04159
Model 2	0.00352	0.00458	0.01225	0.02101
Model 3	0.00366	0.00418	0.00925	0.02001
Model 4	0.00352	0.00458	0.01240	0.02126
Model 5	0.00352	0.00458	0.01238	0.02122
Model 6	0.00365	0.00418	0.01035	0.02024
Model 7	**0.00364**	**0.00417**	**0.09014**	**0.01997**
Currency		BRL		
	1 day		21 days	
Model	MAE	RMSE	MAE	RMSE
Model 1	0.31281	0.61759	0.43348	0.85458
Model 2	0.23311	0.44125	0.24688	0.51381
Model 3	0.21457	0.43458	0.23468	0.50122
Model 4	0.29454	0.54121	0.31594	0.64031
Model 5	0.22364	0.44036	0.24321	0.51294
Model 6	0.23548	0.44201	0.24703	0.51401
Model 7	**0.11239**	**0.29543**	**0.15549**	**0.21591**

Note: Bold numbers suggest superior predictive performance.

D. THE OUT-OF-SAMPLE FORECASTING PERFORMANCE

Finally, we investigate whether or not our proposed hybrid models statistically significantly outperform other competing models. To achieve this analysis, out-of-sample twenty days-ahead forecasts of conditional volatilities are generated based

TABLE 6. The out-of-sample evaluation of forecasts of volatility for six currency returns (21 days).

Model	MAE		RMSE	
	Range statistic	Semi-quadratic statistic	Range statistic	Semi-quadratic statistic
Currency		CNY		
Model 1	0	0	0	0
Model 2	0	0	0	0
Model 3	0.5431	0.8242	0.8314	0.8441
Model 4	0.1251	0.8974	0.9044	0.9134
Model 5	0.5431	0.9998	0.9984	0.9999
Model 6	0.5432	0.9047	0.9138	0.9212
Model 7	1.0000*	1.0000*	1.0000*	1.0000*
Currency		JPY		
Model 1	0	0	0	0
Model 2	0.5124	0.5328	0	0
Model 3	0.8433	0.8641	0	0
Model 4	0.9901	0.9987	0	0
Model 5	0.9901	0.9987	0	0
Model 6	0.9943	0.9999	0.9936	0.9999
Model 7	1.0000*	1.0000*	1.0000*	1.0000*
Currency		EUR		
Model 1	0	0	0	0
Model 2	0	0	0	0
Model 3	0	1	0	0
Model 4	0	0	0	0
Model 5	0	0	0	0
Model 6	0	0	0	0
Model 7	1	1.0000*	1.0000*	1.0000*
Currency		INR		
Model 1	0	0	0	0
Model 2	0	0	0	0
Model 3	0	0	0	0
Model 4	0	0	0	0
Model 5	0	0	0	0
Model 6	0	0	0	0
Model 7	1.0000*	1.0000*	1.0000*	1.0000*
Currency		GBP		
Model 1	0	0	0	0
Model 2	0	0	0	0
Model 3	0.6941	0.7103	0.7119	0.7354
Model 4	0	0	0	0
Model 5	0	0	0	0
Model 6	0	0	0.5385	0.5787
Model 7	1.0000*	1.0000*	1.0000*	1.0000*
Currency		BRL		
Model 1	0	0	0	0
Model 2	0	0	0	0
Model 3	0	0	0	0
Model 4	0	0	0	0
Model 5	0	0	0	0
Model 6	0	0	0	0
Model 7	1.0000*	1.0000*	1.0000*	1.0000*

Note: (*) indicates the best forecasting model.

on seven models. Then, the model confidence set (MCS) of Hansen, Lunde, and Nason [36] is used to compare these seven models' forecasting performance. We would like to note that the MCS test has the ability to find the best forecasting models with a certain probability in the set of competing models [50]. Hansen, Lunde, and Nason [36] proposed two test statistics, which are the semi-quadratic statistic and the range statistic; thus, we consider both tests in our comparison. These tests are constructed from two loss functions, MAE and RMSE, and we use 5,000 bootstrap simulations to obtain the p-values. Note that the higher p-value indicates higher forecasting performance.

According to the MCS test results shown in Table 6, both of the loss functions indicate that the volatility forecast based on our proposed NN-MS-Beta-t-EGARCH shows the best performance as the p-value is equal to 1.000 for all currencies. This result confirms the robustness of our proposed forecasting model.

VI. CONCLUSION

In this work, we combine the merits of two-regime Markov Switching (MS) Beta-t-EGARCH and multi-layer perceptron (MLP) type of neural networks (NN) models for predicting the emerging and advanced market currencies which are most traded nowadays. The proposed model aims at modeling the nonlinear conditional mean processes, the conditional variance, and neural network architecture simultaneously through the hidden Markov process. The model is suggested to forecast the volatility of the six widely traded currencies in advanced and emerging market currencies: Brazilian Real, China Yuan, Indian Rupee, Japanese Yen, Euro, and Pound Sterling. Several criteria are conducted to assess the accuracy of our model. Both in-sample and out-of-sample forecasts of NN-MS-Beta-t-EGARCH are introduced to investigate the accuracy and performance of our model.

Based on this study's findings, both in-sample and out-of-sample volatility forecasts suggest that the NN-MS-Beta-t-EGARCH is superior to the conventional MS-Beta-t-EGARCH as well as the single-regime alternative. The results show that the MS-Beta-t-EGARCH based neural networks augmented models provide the lower AIC, BIC, MAE, and RMSE compared with the competing models' results. We then conclude that the conditional variance models augmented with artificial neural networks enable improved modeling as long as the proposed model captures the volatility more efficiently. Moreover, we compare our proposed model with NN-MS-GARCH and the NN GARCH and found that our proposed model still outperforms these two models under the major currencies data. We also confirm that augmenting MS-Beta-t-EGARCH with MLP leads to better predictions than using NN or MS-Beta-t-EGARCH alone. The MCS test also investigates the robustness forecast, and the result confirms the superiority of our proposed model.

The exchange rate is a key factor driving a national monetary policy and international trade balance for policy implications. According to the present findings, the foreign exchange markets experienced a high degree of volatility persistence over the past decades, and only an extreme event might switch the market returns to the other regime. Hence, from the perspective of policymakers, to create a favorable environment for economic development and avoid more significant fluctuations in the exchange rate, the government does not need to interfere with the foreign exchange market when the market stays in the high volatility regime. For further research, as the exchange rate dynamics are featured by nonlinearity with high and low volatility regimes, it is worth considering the neural networks approach when undertaking some research regarding the financial time-series volatility forecasting. It will enhance the forecasting capabilities.

REFERENCES

[1] A. Ghosh, "A comparison of exchange rate regime choice in emerging markets with advanced and low income nations for 1999–2011," *Int. Rev. Econ. Finance*, vol. 33, pp. 358–370, Sep. 2014.

[2] A. Harvey and G. Sucarrat, "EGARCH models with fat tails, skewness and leverage," *Comput. Statist. Data Anal.*, vol. 76, pp. 320–338, Aug. 2014.

[3] A. K. Dhamija and V. K. Bhalla, "Exchange rate forecasting: Comparison of various architectures of neural networks," *Neural Comput. Appl.*, vol. 20, no. 3, pp. 355–363, May 2010.

[4] A. C. Harvey and T. Chakravarty, "Beta-t-(E)GARCH," Dept. Econ., Univ. Cambridge, Cambridge, U.K., Cambridge Working Papers Econ. 0840, 2008.

[5] C. M. Turner, R. Startz, and C. R. Nelson, "A Markov model of heteroskedasticity, risk, and learning in the stock market," *J. Financial Econ.*, vol. 25, no. 1, pp. 3–22, Nov. 1989.

[6] C. Panda and V. Narasimhan, "Forecasting exchange rate better with artificial neural network," *J. Policy Model.*, vol. 29, no. 2, pp. 227–236, Mar. 2007.

[7] C. Pierdzioch, J.-C. Rülke, and G. Stadtmann, "A note on forecasting emerging market exchange rates: Evidence of anti-herding: Anti-herding in exchange rate forecasts," *Rev. Int. Econ.*, vol. 20, no. 5, pp. 974–984, Oct. 2012.

[8] C.-H. Tseng, S.-T. Cheng, Y.-H. Wang, and J.-T. Peng, "Artificial neural network model of the hybrid EGARCH volatility of the taiwan stock index option prices," *Phys. A, Stat. Mech. Appl.*, vol. 387, no. 13, pp. 3192–3200, May 2008.

[9] D. B. Nelson, "Conditional heteroskedasticity in asset returns: A new approach," *Econometrica*, vol. 59, no. 2, p. 347, Mar. 1991.

[10] D. B. Nelson, "Filtering and forecasting with misspecified ARCH models i," *J. Econometrics*, vol. 52, nos. 1–2, pp. 61–90, Apr. 1992.

[11] E. Chong, C. Han, and F. C. Park, "Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies," *Expert Syst. Appl.*, vol. 83, pp. 187–205, Oct. 2017.

[12] E.-J. Wagenmakers and S. Farrell, "AIC model selection using akaike weights," *Psychonomic Bull. Rev.*, vol. 11, no. 1, pp. 192–196, Feb. 2004.

[13] F. Charef and F. Ayachi, "A comparison between neural networks and GARCH models in exchange rate forecasting," *Int. J. Academic Res. Accounting, Finance Manage. Sci.*, vol. 6, no. 1, pp. 94–99, Feb. 2016.

[14] G. M. Caporale and T. Zekokh, "Modelling volatility of cryptocurrencies using Markov-switching GARCH models," *Res. Int. Bus. Finance*, vol. 48, pp. 143–155, Apr. 2019.

[15] J. Arnerić, T. Poklepović, and Z. Aljinović, "GARCH based artificial neural networks in forecasting conditional variance of stock returns," *Croatian Oper. Res. Rev.*, vol. 5, no. 2, pp. 329–343, Dec. 2014.

[16] I. A. Moosa and J. J. Vaz, "Cointegration, error correction and exchange rate forecasting," *J. Int. Financial Markets, Institutions Money*, vol. 44, pp. 21–34, Sep. 2016.

[17] J. Barunik, T. Krehlik, and L. Vacha, "Modeling and forecasting exchange rate volatility in time-frequency domain," *Eur. J. Oper. Res.*, vol. 251, no. 1, pp. 329–340, May 2016.

[18] J. D. Hamilton, "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, vol. 57, no. 2, p. 357, Mar. 1989.

[19] J. M. Maheu and T. H. McCurdy, "Nonlinear features of realized FX volatility," *Rev. Econ. Statist.*, vol. 84, no. 4, pp. 668–681, Nov. 2002.

[20] J. Vilasuso, "Forecasting exchange rate volatility," *Econ. Lett.*, vol. 76, no. 1, pp. 59–64, Jun. 2002.

[21] K. Pilbeam and K. N. Langeland, "Forecasting exchange rate volatility: GARCH models versus implied volatility forecasts," *Int. Econ. Econ. Policy*, vol. 12, no. 1, pp. 127–142, May 2014.

[22] L. Falat, Z. Stanikova, M. Durisova, B. Holkova, and T. Potkanova, "Application of neural network models in modelling economic time series with non-constant volatility," *Procedia Econ. Finance*, vol. 34, pp. 600–607, Jan. 2015.

[23] M. Bildirici and Ö. Ö. Ersin, "Improving forecasts of GARCH family models with the artificial neural networks: An application to the daily returns in istanbul stock exchange," *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7355–7362, May 2009.

[24] M. Bildirici and Ö. Ersin, "Modeling Markov switching ARMA-GARCH neural networks models and an application to forecasting stock returns," *Sci. World J.*, vol. 2014, pp. 1–21, Jan. 2014.

[25] M. Bildirici and Ö. Ersin, "Forecasting volatility in oil prices with a class of nonlinear volatility models: Smooth transition RBF and MLP neural networks augmented GARCH approach," *Petroleum Sci.*, vol. 12, no. 3, pp. 534–552, Jul. 2015.

[26] M. Griebeler, "Models for forecasting exchange rate volatility: A comparison between developed and emerging countries," *Econ. Bull., AccessEcon*, vol. 34, no. 3, pp. 1618–1630, 2014.

[27] M. Friedman, "The methodology of positive economics," *Philosophy Econ.*, vol. 3, no. 3, pp. 145–178, 1953.

[28] M. Malliaris and L. Salchenberger, "Using neural networks to forecast the S&P 100 implied volatility," *Neurocomputing*, vol. 10, no. 2, pp. 183–195, Mar. 1996.

[29] M. McCrae, Y.-X. Lin, D. Pavlik, and C. M. Gulati, "Can cointegration-based forecasting outperform univariate models? An application to asian exchange rates," *J. Forecasting*, vol. 21, no. 5, pp. 355–380, Aug. 2002.

[30] M. Qi and G. P. Zhang, "An investigation of model selection criteria for neural network time series forecasting," *Eur. J. Oper. Res.*, vol. 132, no. 3, pp. 666–680, Aug. 2001.

[31] N. L. Joseph, "Model specification and forecasting foreign exchange rates with vector autoregressions," *J. Forecasting*, vol. 20, no. 7, pp. 451–484, 2001.

[32] O. Ince, "Forecasting exchange rates out-of-sample with panel methods and real-time data," *J. Int. Money Finance*, vol. 43, pp. 1–18, May 2014.

[33] O. Jeanne and A. K. Rose, "Noise trading and exchange rate regimes," *Quart. J. Econ.*, vol. 117, no. 2, pp. 537–569, May 2002.

[34] O. V. De la Torre-Torres, D. Aguilasocho-Montoya, and M. D. L. C. del Río-Rama, "A two-regime Markov-switching GARCH active trading algorithm for coffee, cocoa, and sugar futures," *Mathematics*, vol. 8, no. 6, p. 1001, Jun. 2020.

[35] P. Maneejuk and W. Yamaka, "Significance test for linear regression: How to test without P-values?" *J. Appl. Statist.*, pp. 1–19, Mar. 2020.

[36] P. R. Hansen, A. Lunde, and J. M. Nason, "Choosing the best volatility models: The model confidence set approach*," *Oxford Bull. Econ. Statist.*, vol. 65, no. s1, pp. 839–861, Dec. 2003.

[37] R. Cont, "Long memory in economics," in *Volatility Clustering in Financial Markets: Empirical Facts and Agent-Based Models*. Berlin, Germany: Springer, 2007, pp. 289–309.

[38] R. G. Donaldson and M. Kamstra, "An artificial neural network-GARCH model for international stock return volatility," *J. Empirical Finance*, vol. 4, no. 1, pp. 17–46, Jan. 1997.

[39] S. Blazsek and H. V. Ho, "Markov regime-switching beta-t-EGARCH," *Appl. Econ.*, vol. 49, no. 47, pp. 4793–4805, 2017.

[40] S. F. Gray, "Modeling the conditional distribution of interest rates as a regime-switching process," *J. Financial Econ.*, vol. 42, no. 1, pp. 27–62, Sep. 1996.

[41] S. Galeshchuk, "Neural networks performance in exchange rate prediction," *Neurocomputing*, vol. 172, pp. 446–452, Jan. 2016.

[42] S. Lahmiri and M. Boukadoum, "An ensemble system based on hybrid EGARCH-ANN with different distributional assumptions to predict S&P 500 intraday volatility," *Fluctuation Noise Lett.*, vol. 14, no. 1, Dec. 2014, Art. no. 1550001.

[43] S. Z. S. Abdalla, "Modelling exchange rate volatility using GARCH models: Empirical evidence from Arab countries," *Int. J. Econ. Finance*, vol. 4, no. 3, pp. 216–229, Feb. 2012.

[44] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *J. Econometrics*, vol. 31, no. 3, pp. 307–327, Apr. 1986.

[45] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *Eur. J. Oper. Res.*, vol. 270, no. 2, pp. 654–669, Oct. 2018.

[46] W. Kristjanpoller, A. Fadic, and M. C. Minutolo, "Volatility forecast using hybrid neural network models," *Expert Syst. Appl.*, vol. 41, no. 5, pp. 2437–2442, Apr. 2014.

[47] W. Kristjanpoller and M. C. Minutolo, "A hybrid volatility forecasting framework integrating GARCH, artificial neural network, technical analysis and principal components analysis," *Expert Syst. Appl.*, vol. 109, pp. 1–11, Nov. 2018.

[48] W. Kristjanpoller and M. C. Minutolo, "Gold price volatility: A forecasting approach using the artificial neural Network–GARCH model," *Expert Syst. Appl.*, vol. 42, no. 20, pp. 7245–7251, Nov. 2015.

[49] W. S. Sarle, "Neural networks and statistical models," in *Proc. 19th Annu. SAS Users Group Int. Conf.* Cary, NC, USA: SAS Institute, 1994, pp. 1538–1550.

[50] W. Yamaka and P. Maneejuk, "Analyzing the causality and dependence between gold shocks and asian emerging stock markets: A smooth transition copula approach," *Mathematics*, vol. 8, no. 1, p. 120, 2020.

[51] Y. Hu, J. Ni, and L. Wen, "A hybrid deep learning approach by integrating LSTM-ANN networks with GARCH model for copper price volatility prediction," *Phys. A, Stat. Mech. Appl.*, vol. 557, Nov. 2020, Art. no. 124907.

[52] Y. Wang, H. Liu, Q. Guo, S. Xie, and X. Zhang, "Stock volatility prediction by hybrid neural network," *IEEE Access*, vol. 7, pp. 154524–154534, 2019.

[53] Y. Zhao, J. Li, and L. Yu, "A deep learning ensemble approach for crude oil price forecasting," *Energy Econ.*, vol. 66, pp. 9–16, Aug. 2017.

RUOFAN LIAO received the bachelor's degree in English and economics from Guangxi Normal University, in 2015, the master's degree in finance from the University of Newcastle upon Tyne, in 2016, and the Ph.D. degree in economics from Chiang Mai University, in 2020. Since 2019, she has published five articles that are indexed in Scopus and ISI. Her research interests include machine learning, finance, economics, and econometrics.

WORAPHON YAMAKA received the bachelor's, master's, and Ph.D. degrees in economics from Chiang Mai University, Chiang Mai, Thailand, in 2011, 2014, and 2017, respectively. He has been a Lecturer with the Faculty of Economics, Chiang Mai University, since 2018, where he is currently the Vice-Director of the Centre of Excellence in Econometrics. Since 2015, he has published more than 120 articles that are indexed in Scopus and ISI. His research interests include economics and econometrics.

SONGSAK SRIBOONCHITTA received the bachelor's and master's degrees in economics from Thammasat University, Bangkok, Thailand, in 1972 and 1975, respectively, and the Ph.D. degree in agricultural economics from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 1983. He is currently a Professor of Economics and has been with the Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand, since 1976, where he was the Dean, from 2004 to 2008. He is currently the President of Thailand Econometric Society and also the Director of the Centre of Excellence in Econometrics, Chiang Mai University. He has published one book and five edited books, and over 300 articles. His current research interests include econometrics, data mining, and decision analysis.

• • •