A New Hybrid Cryptocurrency Returns Forecasting Method Based on Multiscale Decomposition and an Optimized Extreme Learning Machine Using the Sparrow Search Algorithm

Xiaoxu Du1, Zhenpeng Tang2, Junchuan Wu3, Kaijie Chen1 and Yi Cai 1
1School of Economics and Management, Fuzhou University, Fuzhou, CO 350108 China
2School of Economics and Management, Fujian Agriculture and Forestry University, Fuzhou, CO 350002 China
3 School of Economics and Management, Nanchang University, Nanchang, CO 330031 China

Corresponding author: Zhenpeng Tang (e-mail:zhenpt@fzu.edu.cn).

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 71973028 and Grant 71573042.

ABSTRACT The return series of cryptocurrencies, which are emerging digital assets, exhibit nonstationarity, nonlinearity, and volatility clustering compared to other traditional financial markets, making them exceptionally difficult to forecast. Therefore, accurate cryptocurrency price forecasting is important for both market participants and regulators. It has been demonstrated that improved data forecasting accuracy can be achieved through decomposition, but few researchers have performed information extraction on the residual series generated by data decomposition. Based on the construction of a "decomposition-optimization-integration" hybrid model framework, in this paper, we propose a multi-scale hybrid forecasting model that combines the residual components after primary decomposition for secondary decomposition and integration. This model uses the variational modal decomposition (VMD) method to decompose the original return series into a finite number of components and residual terms; then, the residual terms are decomposed and the features are extracted using the completed ensemble empirical mode decomposition with adaptive noise (CEEMDAN) method. The components are predicted by an extreme learning machine optimized by the sparrow search algorithm, and the final predictions are summed to obtain the final results. Forecasts for the returns of Bitcoin and Ethereum, which are major cryptocurrency assets, are compared with other benchmark models constructed based on different ideas, and we find that the proposed quadratic decomposition VMD-Res.-CEEMDAN-SSA-ELM hybrid model demonstrates the optimal and most stable forecasting performance in both one-step and multi-step ahead prediction of the cryptocurrency return series.

INDEX TERMS Cryptocurrency, model selection, decomposition-ensemble, extreme learning machine, sparrow search algorithm.

I. INTRODUCTION
In recent years, the rise of blockchain technology in the context of the increasing integration of finance and the internet has led to the rapid development of cryptocurrency, which is a new type of virtual asset. Cryptocurrency uses the principles of cryptography to secure transactions on a transaction-by-transaction basis using the encryption of virtual currencies and digital hashing combined with smart contracts. Bitcoin (BTC), the first cryptocurrency, was introduced in 2008[1]. Since then, cryptocurrencies such as Ethereum (ETH) and Ripple (XRP) have gradually emerged, and the variety of cryptocurrencies is growing by leaps and bounds. Currently, there are more than 9,000 cryptocurrencies and more than 400 related exchanges, and the terms “coin wave” and “chain wave” are buzzwords. An important issue regarding cryptocurrencies is price volatility.
As seen in Figure 1, the global cryptocurrency market experienced great volatility during the selected period, especially in 2017, when a major wave of growth occurred that represented a 3,175% year-on-year increase in market capitalization, and in 2018, when the market fell by 78%. Then, during the global COVID-19 epidemic, the market experienced a nearly two-year oscillation period when the cryptocurrency market was experiencing another major market[2]. Compared to traditional financial assets, whose valuation is based on fundamental information, cryptocurrencies have a decentralized and virtual existence without a physical backing; they are neither associated with any commodity nor with a company, and governments have no senior regulatory authority over them[3]. Due to the specificity of cryptocurrencies, the complexity of their price fluctuations can be attributed to the multiple factors and uncertainties that interact in the market, including the economic and political environment, as well as investor behavior that can lead to price instability. In addition, cryptocurrency trading rules are different from those of traditional financial markets. Information and events generated at any time can immediately affect the price of cryptocurrencies, rather than at specific market trading hours (e.g., the stock market), as decentralized cryptocurrencies can be traded 24 hours a day and 7 days a week[4]. In summary, one of the most challenging areas of time series is the accurate prediction of the trends in cryptocurrency market quotes[5, 6]. Although the cryptocurrency market itself is extremely complex and risky, it still represents an emerging alternative investment product with high returns and low correlation to other traditional financial assets[7]. It is these characteristics that make cryptocurrencies financial instruments that can be used to hedge against uncertainty[8-10]. Therefore, accurate cryptocurrency price forecasting models can deepen the grasp of cryptocurrency market price fluctuation patterns, provide investors with a reasonable basis for investment decisions in terms of optimal hedging, option pricing, and portfolio diversification, and provide the government with a reference for formulating relevant regulatory policies[11, 12].

According to existing research, asset price volatility in financial markets is dynamic and highly nonlinear[4]. The price forecasting problem of the cryptocurrency market is similar to that of traditional stock and foreign exchange markets and is also a financial time series forecasting problem. However, due to the special trading time system of the cryptocurrency market, its price volatility is more obvious and different from other financial markets[13]. Currently, the methods involved in financial time series forecasting mainly include the following traditional econometric models, artificial intelligence methods, and hybrid models.

Traditional econometric forecasting models include linear multiple regression models[14], error correction models (ECMs)[15], autoregressive integrated moving average models (ARIMA)[16, 17]and vector autoregressive models (VAR)[18]. However, econometric models have specific assumptions, such as that the time series are trending and repeatable and that the data are stable. For data that meet these assumptions, a good prediction can be achieved; however, for nonlinear, nonstationary, and volatile time series such as cryptocurrency prices, the prediction ability of these models is weak[19].

With the rapid development of computer technology, a series of artificial intelligence (AI) techniques have been used in research related to time series forecasting. Common representative models include random forests[20], back propagation neural networks (BPs)[21], artificial neural networks (ANNs)[22], bayesian neural networks (BNNs)[23], convolutional neural networks (CNNs)[24], and support vector regressions (SVRs)[25, 26]. extreme learning machines (ELMs)[27]have been developed as an emerging learning framework for feedforward neural networks that can overcome the training dilemma of backpropagation algorithms for single hidden layer feedforward networks (SLFN) and have been gradually applied to the classification and prediction of various complex sequences with a series of important results due to their advantages in learning convergence speed and parameter settings as well as noise immunity[28-30]. Artificial intelligence algorithms are data-driven, have addressed some limitations of traditional econometric models in forecasting to a certain extent and have significantly improve the forecasting accuracy. However, such methods are more sensitive to parameters and model settings and are prone to local optima and overfitting problems[31].

In the field of financial time series forecasting, hybrid models have become more popular forecasting methods, and their constructed frameworks have been widely used in many studies and proven to be an effective in improving forecasting ability. Many researchers have built hybrid models to achieve effective forecasting for time series,
including Bitcoin price forecasting[32], exchange rate forecasting[33, 34], and international crude oil price forecasting[35]. Generally, hybrid models are based on the idea of "decomposition-integration", which is divided into three steps: data decomposition, modal forecasting and integrated learning. Unlike traditional end-to-end price forecasting methods, hybrid algorithms are used to first decompose the original data through a decomposition algorithm that extracts time-domain features of the time series[36]. For common decomposition algorithms, such as empirical mode decomposition (EMD)[37], ensemble empirical mode decomposition (EEMD)[38], complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN)[39], With the advancement of decomposition algorithm research, VMD[40] algorithm was found to be effective in separating components with similar frequencies, thus enabling more efficient decomposition of the original sequence and demonstrating its superior performance in dealing with complex signals disturbed by noise. After decomposing the data, each decomposed component is predicted separately by the aforementioned prediction models such as econometric models and artificial intelligence models. Finally, the components are summed to obtain the overall prediction results.

Currently, hybrid models have achieved many results in forecasting research in related fields. Sun et al.[41] proposed a new carbon price prediction model based on EEMD-IBA-ELM, and tested the validity and stability of the model by examining the historical carbon prices, then concluded that the proposed model can significantly improve the prediction accuracy. CAO et al.[42] constructed CEEMDAN to decompose the stock price series and then predicted them by LSTM, empirically obtained that CEEMDAN was more thorough in decomposition than EMD, and this hybrid prediction model showed superior performance in predicting the stock price series. Zhu et al.[43] decomposed the carbon price into multiple modes using the VMD model, further reconstructed the modes according to the evolutionary clustering algorithm proposed by CCI, and made predictions to obtain the carbon price prediction results, which demonstrated that the "decomposition-clustering-prediction" approach could predict the carbon price. Jiang et al.[44] constructed new two-stage ensemble models by combining EMD (or VMD), ELM and improved harmony search (IHS) algorithm for stock price prediction, and the results show that the proposed model has superior performance in terms of accuracy and stability compared with other models.

The advantages of each model can be maximized through model integration. As a result, the advantages of each model can be used to better overcome the shortcomings of a single model that obtains large differences in prediction results in different situations and has different prediction evaluation criteria[45, 46]. Therefore, more prediction information is used, which improves the prediction performance. However, the existing hybrid model construction methods still have the following shortcomings. Most of the previous integrated hybrid models decompose the original sequence into a finite number of modal components and residual terms through one decomposition, and then the resulting modalities are predicted through the prediction model[47, 48]. The residual terms, which are discarded as general components, will lead to the accumulation of the prediction errors generated by the decomposition, thus causing a certain degree of data distortion[49]. In fact, the residual term in its complex and nonlinear form may still carry valid predictive information, and further dissection of the residual term is necessary[50]. As a single forecasting model, the ELM model has advantages in terms of convergence speed and parameter settings, but the results obtained by forecasting using an unoptimized single learning machine algorithm are unstable and the forecasting accuracy is not high. Therefore, advanced related algorithms are needed to optimize the single forecasting model to improve the accuracy and stability of the forecasting part. Many existing time series forecasts only consider one-step ahead forecasting, but in the real investment process, investors are also very concerned about the short-term market, especially for the cryptocurrency market, which is characterized by more significant volatility clusters. Therefore, multi-day-ahead forecasting is more important, and multi-step ahead forecasting can help investors provide a more comprehensive and effective reference basis.

In view of the shortcomings of the existing research and based on the inheritance of the abovementioned model construction idea and overcoming its limitations, a hybrid model is constructed in this paper that is composed of a data decomposition algorithm, an optimization algorithm, and a forecasting model, namely, the proposed VMD-Res.-CEEMDAN-SSA-ELM hybrid model, where Res. represents the residual term after VMD. The innovation of the VMD-Res.-CEEMDAN-SSA-ELM model lies in the following points.

1. The quadratic decomposition technique combining VMD and CEEMDAN is applied to the nonlinear, nonstationary, volatility-clustered cryptocurrency price series. First, the VMD algorithm, which is the most effective for processing complex signals[51], is adopted. The original series is decomposed, and then the residual terms obtained after primary decomposition are taken into account. Then, the CEEMDAN algorithm is used to further decompose the residual terms obtained after the VMD decomposition of the original series to extract the complex, nonlinear information contained in the residual terms, and the overall data characteristics of the original time series can be better understood through the secondary decomposition technique, which is more accurate and complete for the decomposition of the original data. (2) In the VMD-Res.-CEEMDAN-SSA-ELM model, the SSA-ELM model is used as the prediction model. The ELM model is advantageous in the field of classification and prediction research, but it depends on the
input parameters; therefore, the hidden layer neuron parameters of ELM are optimized by introducing the cluster intelligence optimization SSA algorithm[52], which has advantages in search accuracy, convergence rate, stability, and avoidance of local optima, so that the decomposed subsequence is input to the prediction model as part of the input features in the prediction module to achieve better stability and improved prediction accuracy. (3) The VMD-Res-CEEMDAN-SSA-ELM model is applied to one-step ahead and multistep ahead forecasting of cryptocurrency returns series and compared with the benchmark model to validate the accuracy and robustness of the model in forecasting complex, nonlinear, volatility-clustered time series such as cryptocurrency returns. The proposed method is completely data-driven and does not require excessive assumptions or consideration of exogenous influences that cause fluctuations in market sentiment, which makes it more consistent with reality.

The rest of this paper is presented as follows. In Section 2, the individual components and details of the VMD-Res.-CEEMDAN-SSA-ELM model are proposed. In Section 3, the daily closing price data of Bitcoin and Ethereum, which are representative among cryptocurrencies, obtained from the CoinMarketCap website as empirical samples are compared. The traditional single benchmark model, the integrated benchmark approach and the VMD-Res.-CEEMDAN-SSA-ELM model constructed in this paper are compared based on the one-step-ahead and the five-step-ahead forecasting using evaluation metrics to test the performance of the proposed model. This paper is concluded in Section 4 and a plan for future work is presented.

II. INTRODUCTION TO THE METHODOLOGY

A combined model (VMD-Res.-CEEMDAN-SSA-ELM) was developed based on the idea of ”decomposition-ensemble” and the combination of secondary decomposition techniques with machine learning methods, aiming to better predict cryptocurrency returns. Because the prediction model proposed in this paper consists of several models, the components of the model and the overall model are described below: the VMD algorithm, the CEEMDAN algorithm, the extreme learning machine, the sparrow search algorithm, and the whole new hybrid model built in this study.

A. Variational Mode Decomposition (VMD)

VMD is an adaptive, quasi-orthogonal, and completely non-recursive decomposition method proposed by Dragomiretskiy and Zosso (2013). In the process of signal decomposition, by searching and solving, the optimal center frequency and limited bandwidth of each mode can be adaptively matched, and the effective separation and frequency domain division of the eigenmode components of the signal can be realized, so that the effective decomposition components of the given signal can be obtained, and finally the optimal solution of the problem can be obtained. The detailed VMD steps are shown as follows.

Step 1: A Hilbert transform is implemented on every modal signal to obtain a unilateral spectrum. The exponential term of the modal function corresponding to the center frequency is mixed and multiplied with the phase to adjust the spectrum of each component signal to the fundamental frequency band. Then, the bandwidth is determined by estimating each component using the Gaussian smoothing method. The corresponding constrained variational model can be described as

\[
\min_{\{u_k(t)\}} \left\{ \sum_k \left| \partial_t \left[ \left( \delta(t) + \frac{i}{\pi t} \right) u_k(t) \right] e^{-\gamma t^2} \right|_2 \right\} \quad \text{s.t. } \sum_k u_k = f
\]

(1)

In this equation, \(\{u_k\} = \{u_1, \ldots, u_k\}\) is the mode component obtained after the VMD decomposition, and \(\{w_k\} = \{w_1, \ldots, w_k\}\) is the center frequency of the mode components, VMF. Where \(\partial_t\) denotes the partial derivative, \(\delta(t)\) refers to the shock function, and \(*\) denotes the convolution sign. \(f\) is the original input signal.

Step 2: To make the signal reconstruction accurate, it must be constrained by introducing an incremental Lagrange function to convert the original equation into a variational problem that is unconstrained. As a result, the optimal solution can be derived as follows:

\[
\ell([\hat{u}_k], [\Lambda]) = \alpha \sum_{i=1}^{K} \left[ \left| \left[ \hat{u}_k(t) + \frac{i}{\pi t} \right] e^{i \omega_k t} \right|_2^2 \right] + \lambda \left[ f(t) - \sum_{i=1}^{K} \hat{u}_k(t) \right]^2 + \lambda \left[ f(t) - \sum_{i=1}^{K} \hat{u}_k(t) \right]^2
\]

(2)

where \(\alpha\) is the quadratic penalty factor introduced to guarantee the accuracy of signal reconstruction when it occurs, and \(\lambda\) is the Lagrange multiplier used to control the strictness of the constraints.

Step 3: Solve the variational problem by searching for the optimal solution of Equation (2) using (alternating direction method of multipliers). Equations (3)–(4) are iterated several times to obtain \(\hat{u}_{k+1}^m(w)\), \(\hat{w}_{k+1}^m\), and \(\Lambda^m\). The optimal solution of the constrained variational model is obtained until the iterative condition (6) is satisfied.

\[
\hat{u}_{k+1}^m(w) = \frac{f(w) - \sum_{i \neq k} \hat{u}_i^m(w) + \Lambda^m}{1 + 2\alpha (w - u_k)^2}
\]

(3)

\[
\hat{w}_{k+1}^m = \frac{\int_0^\infty w \hat{u}_k^m(w)^2 dw}{\int_0^\infty \left| \hat{u}_k^m(w) \right|^2 dw}
\]

(4)

\[
\Lambda_{\Lambda+1} = \Lambda^m + \tau f(w) - \sum_k \hat{u}_{k+1}^m(w)
\]

(5)

In this equation, \(\hat{u}_k^m(w)\), \(\hat{f}(w)\) and \(\lambda^m\) are the fourier transforms of \(\hat{u}_k\), \(\hat{f}(t)\) and \(\lambda^m\) respectively.

\[
\sum_k \left\| \hat{u}_{k+1}^m - \hat{u}_k^m \right\|_2^2 < \varepsilon
\]

(6)

B. Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN)

Torres et al. (2011) proposed a CEEMDAN algorithm based on EMD and EEMD. The algorithm effectively suppresses the mode mixing of EMD by adding finite times of adaptive white
noise at each stage. It can achieve a more thorough
decomposition of the signal data with a smaller reconstruction
error by removing noise residuals with fewer averaging times.
The decomposition steps of CEEMDAN are as follows:
Step 1: Let \( x(n) \) be the original signal sequence, \( \varepsilon \) be the
adaptive coefficient, \( w_i(n) \) be the noise sequence added for
each decomposition, and \( x_i(n) \) be the signal sequence after \( i \)
times of noise is added. The average value of \( N \) subexperiments of EMD decomposition is the first intrinsic
mode component \( IMF_1 \),
\[
IMF_1(n) = \frac{1}{N} \sum_{i=1}^{N} IMF_1(x_i(n))
\] (7)
Step 2: Calculate the residual sequence \( r_1(n) \) of the first stage
and obtain a new \( r_1(n) \) for the \( N \) subexperiments until the
EMD decomposition finishes its work on the IMF component.
\[
r_1(n) = x(n) - IMF_1(n)
\] (9)
Step 3: Based on Step 2, calculate the second intrinsic mode
component \( IMF_2 \) of Phase 2.
\[
IMF_2(n) = \frac{1}{N} \sum_{i=1}^{N} IMF_2\{r_1(n) + \varepsilon_i IMF_1[w_i(n)]\}
\] (10)
Step 4: Repeat the calculation to stage \( k+1 \) and obtain the new \( r_k(n) \) for the \( N \) subexperiments until the
EMD decomposition finishes its work on the IMF component.
\[
r_k(n) = r_{k-1}(n) - IMF_k(n)
\] (11)
\[
IMF_{k+1}(n) = \frac{1}{N} \sum_{i=1}^{N} IMF\{r_k(n) + \varepsilon_i IMF_k[w_i(n)]\}
\] (12)
Step 5: The above steps are repeated. If the number of extreme
points of the residual sequence is \( \leq 2 \), the EMD is stopped, and
the final residual sequence \( R(n) \) and intrinsic mode
component \( IMF_k \). Then, the initial signal sequence \( x(n) \) is
finally decomposed as
\[
x(n) = \sum_{i=1}^{K} IMF_i(n) + R(n)
\] (13)

C. Extreme learning machine optimized by Sparrow
Search algorithm(SSA-ELM)
1) Extreme learning machine(ELM)
Huang et al. (2004) proposed an ELM algorithm to solve
single hidden-layer feedforward neural networks. It mainly
uses the generalized inverse theory of matrices. Compared
with the traditional neural network learning algorithm, only
a unique optimal solution needs to be generated in ELM. This
is achieved by setting the number of hidden layer nodes of
the network. The input weights and biases do not need to be
adjusted during execution; therefore, the advantages of ELM
are fast learning and good generalization performance. The
ELM network model is illustrated in Figure 2.
In Figure 2, \( x_1 \sim x_n \) are the nodes of the input neuron, \( \omega_{11} \sim \omega_{nk} \) is the
weight between the input layer and the hidden layer,
g(\( x \)) is the activation function, \( b_1 \sim b_n \) are the hidden layer
node thresholds, \( \beta_{11} \sim \beta_{nk} \) are the weights between the hidden
layer and the output network model layer, and \( y_1 \sim y_n \) are the
outputs of the model.
Suppose there are \( N \) training samples \( \{(x_i,y_i)\}_{i=1}^{N} \), and
\( x_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T \in \mathbb{R}^n \) refers to the \( n \) dimensional
input data of \( n \) dimensional of the training set.
\( t_i = [t_{i1}, t_{i2}, \ldots, t_{im}]^T \in \mathbb{R}^m \) is the \( m \) dimensional ideal
output value of \( m \) dimension of the expression of the training
set, the ELM network model expression of the hidden layer
nodes \( K \) and assuming an activation function of \( g_i(x) \) can
be expressed as
\[
y_j = \sum_{i=1}^{K} \beta_i g_i(\omega_{ij}x_j + b_i), \ j = 1, 2, \ldots, K
\] (14)
In this equation, \( \omega_i \) is the input weight vector connected
between the input layer nodes and the first hidden layer node
\( i ; \beta_i \) is the weight matrix between the first hidden layer
and the output layer; \( b_i \) is the threshold value of the \( i \)th hidden
layer node; \( y_j \) is the actual output of the network model and
\( g_i(\omega_{ij}x_j + b_i) \) is the activation function.
If the single-feeder neuron network in the hidden layer can
approach any training sample with zero error, that is,
\[
\sum_{i=1}^{N} \| y_i - t_i \| = 0
\] (15)
Then we obtained
\[
t_j = \sum_{i=1}^{K} \beta_i g_i(\omega_{ij}x_j + b_i), \ j = 1, 2, \ldots, N
\] (16)
Which can be matrix represented as
\[
H\beta = T
\] (17)
In this equation,
\[
H = \begin{bmatrix}
g_1(\omega_1 x_1 + b_1) & \cdots & g_k(\omega_k x_1 + b_k) \\
\vdots & \ddots & \vdots \\
g_1(\omega_1 x_N + b_1) & \cdots & g_k(\omega_k x_N + b_k)
\end{bmatrix}_{N \times K}
\] (18)
\( H \) stands for the hidden-layer output matrix and \( T \) for the
ideal output vector.
Therefore, the optimal solution of \( H\beta = T \) is obtained,
which is given by
\[
\hat{\beta} = H^+ T
\] (19)
In this equation, \( H^+ \) refers to the augmentation matrix of the
matrix \( H \).
The entire training process needs to be run only once to obtain the optimal solution, which makes the ELM’s generalization ability very strong.

2) Sparrow Search Algorithm (SSA)

The sparrow search algorithm (SSA) is a new intelligent optimization algorithm proposed by Xue et al. in 2020 that idealizes and formulates the corresponding rules for the predatory behavior of sparrow groups. This algorithm assumes that there are two types of sparrow: discoverers and foragers. Discoverers actively search for food, and foragers obtain food from discoverers. In addition, there are predators that can grab food.

The role of the discoverer is to guide the entire sparrow population in search and predation, and its position can be expressed by equation.

\[
X_{i,d}^{t+1} = \begin{cases} 
X_{i,d}^t \cdot \exp \left( \frac{-i}{\alpha \cdot \text{iter}_{\text{max}}} \right) & R_2 < ST \\
X_{i,d}^t + Q \cdot L & R_2 \geq ST 
\end{cases} \tag{20}
\]

In this equation, \( t \) is the current number of iterations; \( X_{i,d}^{t+1} \) is the location of the first sparrow in the \( t + 1 \) th iteration of the \( d \) th dimension; \( T \) is the maximum number of iterations; \( \alpha \in [0,1] \) is a random number; \( Q \) is a random number that obeys a normal distribution; \( R_2 \) and \( ST \) is a matrix whose warning and safety values are of size \( 1 \times d \) and each element is 1, respectively.

If \( R_2 < ST \), this means that predators are not nearby, so the discoverer can perform a wide-area search. If \( R_2 \geq ST \), that is, the predator has been found, and the rest of the sparrows need to leave their present position.

Foragers should observe the finder during this process, and when a good food source is noted, the foragers will leave their location to compete for food. If the scramble is successful, they will receive food from the finder; thus, the foragers' positions are updated as

\[
X_{i,d}^{t+1} = \begin{cases} 
Q \cdot \exp \left( \frac{X_{\text{worst}} - X_{i,d}^t}{\varepsilon_i^2} \right) & i > n \\
X_{i,d}^{t+1} + |X_{i,d}^t - X_{p,d}^{t+1}| \cdot A \cdot L & \text{other situation}
\end{cases} \tag{21}
\]

In this equation, \( X_{p,d}^{t+1} \) is the worst position of the entire process in the \( t \) th iteration, \( X_{p,d}^{t+1} \) is the best position of the finder in the \( t + 1 \) th iteration, and \( A \) is an \( 1 \times d \) dimensional matrix with the same dimensions as the input. Each element is randomly assigned 1 or -1, and \( A^+ = A \cdot (A^T \cdot A)^{-1} \); \( n \) is the number of sparrows.

Usually, some foragers will act as warning sparrows to help the discoverer forage, and when in danger, they will counterattack or withdraw close to other sparrows.

\[
X_{i,d}^{t+1} = \begin{cases} 
X_{\text{best}}^t + \lambda \cdot |X_{i,d}^t - X_{\text{best}}| & f_i > f_g \\
X_{i,d}^t + J \left( \frac{X_{i,d}^t - X_{\text{worst}}^t}{(f_i - f_g) + \varepsilon} \right) & f_i = f_g 
\end{cases} \tag{22}
\]

In this equation, \( X_{\text{best}}^t \) is the best position for the \( t \) th iteration; as a step control parameter, \( \lambda \) is a random number that obeys a normal distribution with an average of 0 and variance of 1; \( J \in [-1,1] \) is a random number; \( f_i \) is the current sparrow’s fitness value; \( f_g \) and \( f_w \) are the best fitness value and the worst fitness value, respectively; \( \varepsilon \) is defined as a particularly small constant, which is mainly used to prevent the case of \( f_i - f_w = 0 \). \( f_i \), \( f_g \), and \( f_w \) are the current sparrow's fitness value, best global fitness value, and worst fitness value, respectively; \( \varepsilon \) is defined as a very small constant, mainly to prevent the case of \( f_i - f_w = 0 \).

The main steps of the sparrow algorithm are as follows.

Step 1: Initialize the population, set the total number of sparrow population \( n \), number of discoverers, number of warners, maximum number of iterations \( T \), and alarm threshold \( R_2 \).

Step 2: Use Mean Squared Error (MSE) as the fitness function and then calculate the fitness value of each sparrow. Find and define the best and worst fitness values as \( f_g \) and \( f_w \), respectively.

Step 3: The new positions of the discoverer, forager, and predator. The corresponding fitness value \( P_i \) of each sparrow is calculated separately, and the best fitness is defined along with the position of the corresponding sparrow as \( X_{\text{best}} \).

Step 4: Perform iterations and repeat step 3 to continuously update the positions of sparrows, stopping when the number of iterations is performed \( T \). So the position of the sparrow with the lowest fitness value in all iterations is the optimal solution.

3) The Extreme Learning Machine for Sparrow Search Algorithm Optimization (SSA-ELM)

ELM can be used for nonlinear function fitting and prediction problems with small-sample learning, but the stability of model training can be affected by its input weights and implied layer thresholds. SSA algorithm has the advantages of high search accuracy, fast convergence, and good stability, so that the ELM input parameters and weights can be optimized by SSA algorithm to obtain more stable prediction results and prediction efficiency, and the detailed operation flow of SSA-ELM model is as follows:

Step 1: The sparrow population is initialized as the discoverer, forager, and predator. The corresponding fitness value \( P_i \) of each sparrow is calculated separately, and the best fitness is defined along with the position of the corresponding sparrow as \( X_{\text{best}} \).

Step 2: Iterations are performed to determine the optimal initial weights and thresholds. These data can be obtained by comparing the fitness MSE because the mean square error is the fitness function. More specifically, when exercising the second iteration, the minimum MSE of the current sparrow generations should be compared with the optimal adaptive value of the previous generations \( P_i \). If it is less than \( P_i \), the optimal adaptation value \( P_i \) must be updated to the minimum MSE of the current generation of sparrows, and the position of this sparrow should be updated to the optimal position \( X_{\text{best}} \). Otherwise, the optimal adaptation value \( P_i \) and the optimal
position $X_{\text{best}}$ need not be updated, and the next iteration can be performed.

Step 3: The iteration should be stopped until it reaches the set value $iter_{\text{max}}$. Finally, the optimal weights and thresholds obtained from the model optimized by the sparrow algorithm are used to construct a new ELM model for prediction. The specific steps of SSA-ELM are shown in figure 3.

**D. VMD-Res.-CEEMDAN-SSA-ELM model**

As previously mentioned, cryptocurrency return series have complex characteristics, such as typical nonstationary, nonlinear, and volatility clustering, which have limited accuracy when using a single forecasting method. Because the VMD decomposition technique can decompose the complex signal into several mode components with much lower complexity, the prediction accuracy is substantially improved when each modal component obtained from the VMD decomposition is modeled separately through common forecasting methods. However, previous studies have only modeled the modal components from the decomposition of VMD, discarding the complex information contained in the residual directly after modal decomposition. However, the decomposed modal components have received sufficient attention from scholars, while the remaining terms have not been investigated; therefore, in this study, based on the primary decomposition, a secondary decomposition of the residual terms by CEEMDAN is considered, which in turn enables a more thorough decomposition of the original sequence. Finally, ELM is used to optimize the SSA. To a certain extent, the advanced SSA-ELM model can improve the prediction accuracy, convergence speed, and stability of cryptocurrency returns. The specific modelling steps are shown in Figure 4.

Step 1: Decompose the original return series into each mode component $\text{VMF}_i$ through the VMD decomposition technique and subtract each mode component from the original return series to obtain the residual series ($\text{Res.}$).

Step 2: Secondary Decomposition. CEEMDAN was applied to further decompose the residual series to obtain another set of sub-series $\text{IMF}_i$. Normalize decomposed $\text{VMF}_i$ and $\text{IMF}_i$.

Step 3: The modal components obtained from the decomposition of the original series and the residual term decomposition are predicted by the SSA-ELM.

Step 4: Overall forecast cryptocurrency returns. The predictions of the residuals are superimposed with the

---

**FIGURE 3. SSA-ELM process**

**FIGURE 4. Structure of the VMD-Res.-CEEMDAN-SSA-ELM model**
predictions of each VMF, to forecast the final cryptocurrency return.

### III. Empirical Analysis

#### A. Data description and evaluation criteria

In the cryptocurrency market, Bitcoin and Ethereum account for nearly 66% of the market capitalization and enjoy the majority of the daily trading volume, even reaching more than 70% of the whole market on June 30, 2021. Therefore, in this study, the log returns of the daily closing prices of Bitcoin and Ethereum were selected as the prediction objects. The log returns of the \( t \) th trading day, \( r_t = (\ln (p_t) - \ln (p_{t-1})) \times 100 \), and the returns of the above two virtual currencies are predicted using the proposed VMD-Res.-CEEMDAN-SSA-ELM to verify the effectiveness of the model, and the daily closing prices of BTC and ETH were obtained from the web (http://www.CoinMarketCap.com/). People were not highly aware of or interested in cryptocurrencies such as Bitcoin before the upsurge of cryptocurrency, and the period after 2017 is what truly raises the attention of investors and academics. Trading volume data also confirmed this trend. Therefore, this study selects the return rate of BTC and ETH from January 1, 2017, to June 30, 2021, with 1,642 returns data each, were selected in this study. In the respective returns datasets of Bitcoin and Ethereum, the training and test sets were divided; the first 1,492 returns data were used as the train set, and the remaining 150 data were used as the test set. Table I lists the descriptive information related to Bitcoin and Ethereum returns data. The empirical operation in this study was completed by MATLAB 2019b.

### Table I

**Descriptive statistics for Bitcoin and Ethereum return data**

| Cryptocurrency | Mean | Maximum | Minimum | Median | Standard deviation | Sample Length |
|----------------|------|---------|---------|--------|-------------------|---------------|
| **BTC**        | 0.2371 | 22.5119 | -46.4730 | 0.2476 | 4.2630            | 1492          |
| **Test set**   | 0.0377 | 17.1820 | -14.8107 | -0.0143 | 4.7594           | 150           |
| **ETH**        | 0.3422 | 29.0130 | -55.0714 | 0.1408 | 5.7034            | 1492          |
| **Test set**   | 0.3653 | 22.5649 | -31.7459 | 0.6860 | 6.4284            | 150           |

In this study, four evaluation metrics, Mean Absolute Error (MAE), Normalized Root Mean Square Error (NRMSE), and Symmetric Mean Absolute Percentage Error (SMAPE), coefficient of determination (\( R^2 \)) were selected to test the prediction effectiveness of the models. In addition, in order to more concisely see the differences in evaluation metrics between different benchmark models and the proposed model, we define the following three evaluation metrics relative to the proposed model.

Table II presents the definitions and formulas for the relevant evaluation indicators.

### Table II

**Relevant Evaluation Indicators**

| Evaluation Indicator | Definition | Equation |
|----------------------|------------|----------|
| MAE                  | Mean Absolute Error | \( \frac{1}{n} \sum_{i=1}^{n} |x_i - \bar{x}| \) |
| NRMSE                | Normalized Root Mean Squared Error | \( \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2} \) |
| SMAPE                | Symmetric Mean Absolute Percentage Error | \( \frac{1}{n} \sum_{i=1}^{n} \frac{|x_i - \bar{x}|}{\bar{x} + x_i} \) |
| \( R^2 \)            | Coefficient of Determination | \( R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y_i})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \) |
| \( R_{\text{MAE}} \) | Reducing percentages of Mean Absolute Error | \( R_{\text{MAE}} = \frac{\text{MAE}_p - \text{MAE}_r}{\text{MAE}_p} \) |
| \( R_{\text{NRMSE}} \) | Reducing percentages of Normalized Root Mean Squared Error | \( R_{\text{NRMSE}} = \frac{\text{NRMSE}_p - \text{NRMSE}_r}{\text{NRMSE}_p} \) |
| \( R_{\text{SMAPE}} \) | Reducing percentages of Symmetric Mean Absolute Percentage Error | \( R_{\text{SMAPE}} = \frac{\text{SMAPE}_p - \text{SMAPE}_r}{\text{SMAPE}_p} \) |

#### B. Data Processing

In practical applications, the optimal number of modal components cannot be directly determined when decomposing the original time series through VMD decomposition because of the admixture of noise in the original time series. Therefore, the average instantaneous frequency observation method was used in this study to determine the optimal \( K \) value. For Bitcoin returns series data, the average instantaneous frequency decreased less at the end when the value of \( K \) is 11, that is, overdecomposition exists. Therefore, the optimal number of VMD decomposition modes for the Bitcoin returns series was 10. Similarly, the optimal number of components for VMD decomposition of the Ethereum returns series was calculated to be 12. After determining parameter \( K \), time series data decomposition was performed. Then, after subtracting the original series from each component of the mode decomposed by VMD, the residual series was obtained. Owing to its complexity, the residual series using a predictive algorithm is difficult to accurately predict. Hence, previous studies typically neglected this series. To a certain extent, this operation tends to lead to a loss of information. Thus, to extract more available information, the
secondary decomposition technique was adopted in this study and the complex residual series was further decomposed using CEEMDAN technology. Taking the decomposition of the Bitcoin returns series as an example, the decomposition process is shown in Figure 5. In addition, owing to the large span of values of individual features, the differences in their units and magnitudes led to features that are not comparable with each other; therefore, the data needed to be normalized for each decomposed subseries of modal component data before being predicted using machine learning methods.

In this study, the data were linearly altered using the minimum-maximum deviation normalization method with the following expressions:

\[ \tilde{x} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \]  

(23)

where \( x \) denotes the original feature data, \( \tilde{x} \) denotes the standardized subseries data, \( x_{\text{max}} \) represents the maximum value in the original sequence, and \( x_{\text{min}} \) denotes the minimum value.

C. Forecast results

In order to verify the validity of the proposed model, we compare the predictive validity and stability of the proposed model with the basic models.

First, in order to test the differences in forecasting performance between the hybrid model and the single forecasting model, as well as for the purpose of selecting the optimal basic model for the forecasting module. We introduce traditional econometric models and artificial intelligence models as basic models to obtain the differences in forecasting performance between different single forecasting models. The intelligent cluster SSA algorithm is also selected to optimize the above better basic prediction models, so that the model with the best prediction performance can be selected as the main component of the hybrid model.

Second, in order to examine the difference in prediction performance between the proposed quadratic decomposition model considering residual terms compared with the commonly used methods in the general decomposition-integration framework, we empirically modeled the commonly used methods in each possible combination. Specifically, in the modal decomposition stage, EMD, CEEMDAN or VMD can be used to decompose the return series. At the same time, due to different modeling ideas, scholars have been able to improve the forecasting accuracy by decomposing the high-frequency components of the primary decomposition quadratically and integrating the forecasts, and we correspondingly incorporate such models. Finally, based on the residual terms generated after considering the primary decomposition proposed in this paper,
we construct an integrated forecasting model combining the residual terms. Based on the above considerations, ten additional different benchmark models are constructed in this paper for comparison with the proposed models. Specifically: (1) single prediction models: ARIMA, BP, SVR, ELM, SSA-ELM; (2) hybrid prediction models: EMD-SSA-ELM, CEEMDAN-SSA-ELM, CEEMDAN-VMD-SSA-ELM, VMD-SSA-ELM, VMD-Res.-SSA-ELM, VMD-Res.-CEEMDAN-SSA-ELM.

In addition, one-step and multistep ahead forecasting was also performed in this paper by one-step-ahead and five-step-ahead, i.e., the data of the first six trading days allows forecasting the

TABLE III

| Cryptocurrency | Model                              | One-step ahead prediction | Five-step ahead prediction |
|----------------|------------------------------------|---------------------------|---------------------------|
|                | MAE (R<sub>M</sub>) | NRMSE (R<sub>NRMSE</sub>) | SMAPE (R<sub>SMAPE</sub>) | R<sup>2</sup> | MAE (R<sub>M</sub>) | NRMSE (R<sub>NRMSE</sub>) | SMAPE (R<sub>SMAPE</sub>) | R<sup>2</sup> |
| BTC            | ARIMA                             | 3.5206 (82.49%)           | 0.1491 (84.05%)           | 0.0260 (66.02%) | -0.0045 | 3.5303 (68.44%)           | 0.1493 (71.06%) | 0.0651 (60.22%) | -0.0070 |
|                | BP                                | 3.5040 (82.41%)           | 0.1486 (84.00%)           | 0.0320 (66.23%) | 0.0922 | 3.5079 (68.24%)           | 0.1487 (79.92%) | 0.0948 (59.51%) | 0.0099 |
|                | SVR                               | 3.4602 (82.24%)           | 0.1481 (83.94%)           | 0.0242 (65.95%) | 0.0095 | 3.4883 (68.06%)           | 0.1485 (70.89%) | 0.0940 (59.33%) | 0.0112 |
|                | ELM                               | 3.4581 (82.18%)           | 0.1478 (83.90%)           | 0.0234 (65.92%) | 0.0134 | 3.4851 (68.03%)           | 0.1483 (70.85%) | 0.0939 (59.11%) | 0.0099 |
|                | SSA-ELM                           | 3.4420 (82.09%)           | 0.1475 (83.87%)           | 0.0205 (65.17%) | 0.0176 | 3.4761 (67.95%)           | 0.1482 (70.83%) | 0.0921 (58.36%) | 0.0076 |
|                | EMD-SSA-ELM                       | 2.6954 (77.14%)           | 0.1065 (77.67%)           | 0.0797 (60.54%) | 0.4873 | 3.6302 (69.31%)           | 0.1470 (79.59%) | 0.0972 (58.14%) | 0.0233 |
|                | CEEMDAN-SSA-ELM                   | 2.5622 (75.95%)           | 0.1011 (76.48%)           | 0.0781 (60.17%) | 0.5379 | 3.4664 (67.86%)           | 0.1416 (69.46%) | 0.0875 (55.95%) | 0.0943 |
|                | CEEMDAN-VMD-SSA-ELM               | 1.0199 (39.57%)           | 0.0431 (44.78%)           | 0.0207 (25.20%) | 0.9162 | 2.2919 (51.39%)           | 0.0994 (54.45%) | 0.0620 (38.67%) | 0.5930 |
|                | VMD-SSA-ELM                       | 0.9922 (37.89%)           | 0.0313 (37.99%)           | 0.0405 (21.62%) | 0.9336 | 1.1985 (7.03%)            | 0.0561 (13.62%) | 0.4121 (6.84%)  | 0.8868 |
|                | VMD-Res.-SSA-ELM                  | 0.8844 (30.31%)           | 0.0346 (31.32%)           | 0.0543 (11.18%) | 0.9458 | 1.1161 (4.08%)            | 0.0473 (8.58%)  | 0.4083 (5.96%)  | 0.8989 |
|                | VMD-Res.-CEEMDAN-SSA-ELM          | 0.6163                     | 0.0238                     | 0.3147                     | 0.9744 | 1.1142                     | 0.0432                     | 0.3839                     | 0.9185 |
| ETH            | ARIMA                             | 4.78143 (83.10%)          | 0.1190 (84.43%)           | 0.0394 (70.96%)           | -0.0103 | 4.7817 (69.45%)           | 0.1189 (70.60%) | 0.0413 (58.66%) | -0.0092 |
|                | BP                                | 4.7438 (82.97%)           | 0.1184 (84.37%)           | 0.0391 (70.95%)           | -0.0055 | 4.7522 (69.26%)           | 0.1187 (70.54%) | 0.0391 (58.21%) | -0.0049 |
|                | SVR                               | 4.7376 (82.94%)           | 0.1171 (84.20%)           | 0.0343 (70.80%)           | 0.0220 | 4.7453 (69.26%)           | 0.1183 (70.45%) | 0.0554 (59.27%) | 0.0009 |
|                | ELM                               | 4.7032 (82.82%)           | 0.1169 (84.17%)           | 0.0257 (70.53%)           | 0.0242 | 4.7185 (69.05%)           | 0.1179 (70.35%) | 0.0921 (57.77%) | 0.0074 |
|                | SSA-ELM                           | 4.6413 (82.59%)           | 0.1158 (84.02%)           | 0.0201 (69.69%)           | 0.0420 | 4.6547 (68.62%)           | 0.1174 (70.22%) | 0.0913 (57.26%) | 0.0154 |
|                | EMD-SSA-ELM                       | 3.2330 (75.01%)           | 0.0773 (76.07%)           | 0.0716 (59.38%)           | 0.5738 | 4.4833 (67.42%)           | 0.1116 (68.68%) | 0.0857 (54.61%) | 0.1104 |
|                | CEEMDAN-SSA-ELM                   | 3.0361 (73.39%)           | 0.0721 (74.34%)           | 0.0649 (57.86%)           | 0.6287 | 4.4525 (66.99%)           | 0.1115 (68.65%) | 0.0845 (53.93%) | 0.1127 |
|                | CEEMDAN-VMD-SSA-ELM               | 1.2655 (36.15%)           | 0.0338 (45.27%)           | 0.0362 (24.81%)           | 0.9184 | 2.8797 (49.28%)           | 0.0727 (51.91%) | 0.0618 (36.30%) | 0.6228 |
|                | VMD-SSA-ELM                       | 1.2072 (32.82%)           | 0.0281 (34.16%)           | 0.0321 (17.36%)           | 0.9437 | 1.5710 (7.03%)            | 0.0386 (9.44%)  | 0.4014 (3.06%)  | 0.8919 |
|                | VMD-Res.-SSA-ELM                  | 1.1557 (30.09%)           | 0.0271 (31.73%)           | 0.0323 (17.91%)           | 0.9477 | 1.5529 (5.94%)            | 0.0380 (8.01%)  | 0.3962 (1.79%)  | 0.8984 |

**Figure 6.** True values, one-step and five-step ahead prediction results of the different models for BTC and ETH
return of the next 1 and 5 trading days, respectively. The evaluation metrics of the prediction results are shown in Tables III. Figures 6 (a)-(d) show the true values and prediction results of the different models for Bitcoin and Ethereum in one-step and five-step ahead.

### D. Model Comparison and Analysis

1) Non-combined Models

First, the common econometric and artificial intelligence models ARIMA, BP, SVR, and ELM are used as benchmark prediction models to predict the return data of BTC and ETH as shown in Table III. From the overall performance, under the comparison of the four single models, it can be seen that ELM performs the best in all evaluation metrics in the prediction of Bitcoin and Ethereum 1 day ahead and 5 days ahead. ARIMA, on the other hand, has the worst prediction performance. The possible reason is that ARIMA as a classical linear model is difficult to capture the pattern due to the non-linear and high volatility characteristics of the cryptocurrency return data. Further, we introduce the SSA optimization algorithm combined with ELM, and the four evaluation indicators of SSA-ELM improve to a certain extent compared with other single forecasting models when forecasting complex financial time series with high volatility like cryptocurrency yields. Therefore, we adopt SSA-ELM as the main model for the forecasting module in the next construction of the hybrid model.

However, in a comprehensive view, the $R^2$ indicators of the overall fit of the single forecasting model starting from data-driven do not perform well and cannot effectively capture the complex cryptocurrency return data characteristics.

2) Combined model without considering residual term decomposition

The evaluation indices of CEEMDAN-SSA-ELM were generally better than those of EMD-SSA-ELM, which shows that the CEEMDAN decomposition is more complete than the EMD decomposition, and thus the data features were more fully extracted. The best prediction results at this stage were achieved when using VMD-SSA-ELM. In this regard, it can be shown that the VMD technique has a stronger decomposition ability for time series with a high degree of complexity and volatility, such as cryptocurrency prices, and can be used to better extract series data features and deal with complex signals. By comparing the forecasting results of the common hybrid models EMD-SSA-ELM, CEEMDAN-SSA-ELM, CEEMDAN-VMD-SSA-ELM and VMD-SSA-ELM constructed under the decomposition-based integration framework. It is easy to find that the combined model has significantly improved in four evaluation metrics, MAE, NRMSE, SMAPE and $R^2$, by comparing with the single prediction model. The prediction accuracy and stability of the combined model with the decomposition technique are better than those of the single prediction model.

Further, these combined models are viewed separately. The evaluation indices of the prediction results of CEEMDAN-SSA-ELM are generally better than those of EMD-SSA-ELM, which shows that the CEEMDAN decomposition is more complete than the EMD decomposition for data decomposition, and thus makes the data features extraction more adequate. The VMD-SSA-ELM achieves the best prediction results in this stage. In this regard, it can be shown that the VMD decomposition technique has stronger decomposition ability for time series with high complexity and high volatility like cryptocurrency prices, and can better extract serial data features and handle complex signals.

3) Combined model considering residual term decomposition

The VMD-Res.-SSA-ELM and the proposed model take into account the residual terms after first decomposition and incorporate the prediction of the resulting residual series as well. The difference is that the model proposed in this paper will further decompose the residual terms generated from the primary decomposition to obtain the corresponding components, so that the components obtained from the decomposition of the original sequence and the components obtained from the decomposition of the residual terms can be predicted. Then integrated to obtain the final prediction results. The prediction results of the VMD-Res.-SSA-ELM and VMD-Res.-CEEMDAN-SSA-ELM are shown in the Table III. The predictive evaluation metrics obtained from the model proposed in this paper are further improved compared to VMD-Res.-SSA-ELM. It can be seen that the residual series generated after the original returns series decomposed by VMD also contains important and complex information, and the direct use of the SSA-ELM model to forecast the residual terms directly has limited effect. Specifically, for example, for the forward one-step prediction of bitcoin returns, the proposed model achieves an overall fit evaluation index $R^2$ of 0.9744.

In general, the proposed model performs best in all evaluation metrics compared to all other benchmark models. Therefore, the quadratic decomposition technique proposed in this paper for the residual term appears to be necessary. The original sequence is decomposed twice by VMD and CEEMDAN, which can effectively combine the advantages of the two algorithms to better grasp the characteristics of the original sequence, and thus can be combined with the SSA-ELM prediction module to further obtain more accurate prediction results.

4) Analysis of forward multiscale forecasting

A longitudinal comparison of multiscale forward forecasting shows that the forecasting performance of both the benchmark model and the proposed multiscale hybrid model gradually decreased as the forecasting scale increased. This means that all models have the best forecasting ability when forecasting 1 day in advance, and the overall accuracy of forecasting 5 days in advance is not as good as that of forecasting 1 day in advance. This is mainly because as the forecasting scale increases, the complexity and high volatility of the cryptocurrency return series increases. Therefore, the
accuracy of forecasting 5 days in advance is lower than that of forecasting 1 day in advance. In addition, some data information is not trained by the model in actual Bitcoin and Ethereum return forecasting, which leads to the gradual weakening of the model’s forecasting ability and explains the increased difficulty of forecasting cryptocurrency return series multiple steps ahead compared to one day ahead.

5) Forecast Error Analysis

To further understand and compare the forecast error distribution between the models, we analyze the forecast errors generated by the benchmark models and the proposed model in the empirical study of one-step and five-step ahead prediction scenarios for major cryptocurrency returns. Figures 7 (a)-(d) show the forecast error distributions and the corresponding fitted error distribution curves of the benchmark and proposed models, then depict Taylor plots with the error data. It is worth mentioning that Taylor diagrams are used to facilitate the analysis of correlation measures between different models by presenting statistical information such as standard deviation, correlation coefficient, and root mean square deviation (RMSD) generated by each different forecasting models in a single plot during the actual forecasting process [53]. The Taylor plots plotted in the paper are done by normalizing the standard deviation and root mean square deviation in order to make the comparison between models more intuitive.

We combine the results of the forward one-step and five-step forecasts of the two cryptocurrency asset returns for error analysis. First, the error distribution curves fitting the prediction errors of the different models show that the hybrid model has a smaller error than the single model. In addition, the latter three hybrid models using VMD decomposition techniques have a smaller range of error point fluctuations and run more smoothly. The errors of the proposed models are more distributed around zero, with the smallest corresponding standard deviations. Taylor plots show that among all empirical results, the proposed model is closest to the actual values compared to the other models and shows that the proposed model has optimal prediction accuracy and stability by the positions of the three main statistical indicators.

6) Short summary

Overall, the VMD-Res.-CEEMDAN-SSA-ELM multiscale hybrid model proposed in this paper can be used to forecast cryptocurrency returns more accurately. Compared to the benchmark model, the hybrid model proposed in this paper performs best in the four evaluation metrics shown in the forward one-step and five-step forecasting process for two cryptocurrency asset returns. Moreover, compared with the single forecasting model, the hybrid model without considering the residual term improves the evaluation indicators significantly. Specifically, compared with the ARIMA model, the MAE of the proposed model decreases by 82.49%, the NRMSE by 84.05%, and the SMAPE by 66.02% in the forward one-step forecasting of bitcoin return, and the $R^2$ index reaches 0.9744. In comparing the prediction errors,
we found that the proposed model had a more concentrated prediction error distribution and showed excellent and stable prediction performance.

IV. Conclusion and Future Work

In this paper, we proposed a combined model of quadratic decomposition, optimization, and prediction considering residual terms for Bitcoin and Ethereum daily returns for forecasting research, drawing on the decomposition-integration technique, and the empirical analysis led to the following conclusions.

1) Based on complex systems methodology, the decomposition and integration techniques were used to decompose the cryptocurrency return series into subseries with different frequencies, predict each subseries individually, and finally integrate and reconstruct the prediction results of each subseries to form the overall prediction results. This process could improve the accuracy of forecasting more effectively than a single model.

2) The VMD technique performed better when dealing with highly complex time series data such as nonstationary and nonlinear data. Adapting VMD decomposition combined with the SSA-ELM algorithm could be used to substantially improve the prediction results compared with CEEMDAN and EMD.

3) The combined VMD-Res.-CEEMDAN-SSA-ELM quadratic decomposition model had a significantly stronger forecasting ability than the combined single decomposition model. Considering the residual terms, the VMD and CEEMDAN quadratic decomposition techniques could be used to effectively decompose the nonstationary, nonlinear and highly complex financial time series with clustered fluctuations into a number of more regular smooth subseries. The combined model had significant advantages over the single model and other combined models without considering the residual terms. Information and the optimal prediction results achieved in both the 1-step-ahead and 5-step-ahead prediction studies proved the robustness of the model.

The proposed multiscale hybrid model, VMD-Res.-CEEMDAN-SSA-ELM, conducted an empirical study based on the returns of Bitcoin and Ethereum, and found that the proposed model could be used to effectively improve the accuracy of cryptocurrency return forecasting. These results can help participants in the cryptocurrency market, especially short-term investors, to more accurately understand the price trends of the cryptocurrency market. However, the empirical evidence in this paper for cryptocurrency return series also proves that the trends of price series can be affected by multidimensional complex factors, the data are still...
characterized by sharp fluctuations, and the performance of the model is weakened in multistep forecasting. Therefore, in future research, we plan to consider combining other multidimensional complex influence modeling to further upgrade this proposed model by considering higher frequencies (e.g., hours) and longer time horizons (e.g., months or years) to better capture financial time series characteristics and apply them to actual portfolio strategy design to provide a reference for investors and regulators.

REFERENCES

[1] S. Nakamoto, “A Peer-to-Peer Electronic Cash System,” Decentralized Business Review, p. 21260, 2008.

[2] S. Aras, "On improving GARCH volatility forecasts for Bitcoin via a meta-learning approach," Knowledge-Based Systems, vol. 230, Oct 27 2021.

[3] D. H. Shen, A. Urquhart, and P. F. Wang, "Does twitter predict Bitcoin?," Economics Letters, vol. 174, pp. 118-122, Jan 2019.

[4] H. Z. Guo, D. Zhang, S. Y. Liu, L. Wang, and Y. Ding, "Bitcoin price forecasting: A perspective of underlying blockchain mining," Decision Support Systems, vol. 151, Dec 2021.

[5] R. Chowdhury, M. A. Rahman, M. S. Rahman, and M. R. C. Mahdy, "An approach to predict and forecast the price of constituents and index of cryptocurrency using machine learning," Physica a-Statistical Mechanics And Its Applications, vol. 551, Aug 1 2020.

[6] Z. R. Zhang, H. N. Dai, J. H. Zhou, S. K. Mondal, M. M. Garcia, and H. Wang, "Forecasting cryptocurrency price using convolutional neural networks with weighted and attentive memory channels," Expert Systems with Applications, vol. 183, Nov 30 2021.

[7] S. F. Cheng, G. De Franco, H. B. Jiang, and P. K. Lin, "Riding the Blockchain Mania: Public Firms’ Speculative R-K Disclosures," Management Science, vol. 65, pp. 5901-5913, Dec 2019.

[8] Y. C. Ma, F. Ahmad, M. Liu, and Z. L. Wang, "Portfolio optimization in the era of digital financialization using cryptocurrencies," Technological Forecasting And Social Change, vol. 161, Dec 2020.

[9] I. Shaikh, "Policy uncertainty and Bitcoin returns," Borsa Istanbul Review, vol. 20, pp. 257-268, Sep 2020.

[10] D. Aggarwal, "Do bitcoins follow a random walk model?," Research In Economics, vol. 73, pp. 15-22, Mar 2019.

[11] J. J. Li, Y. Yuan, and F. Y. Wang, "A novel GSP auction mechanism for ranking Bitcoin transactions in blockchain mining," Decision Support Systems, vol. 124, Sep 2019.

[12] R. Matkovskyy and A. Jalal, "From financial markets to Bitcoin markets: A fresh look at the contagion effect," Finance Research Letters, vol. 31, pp. 93-97, Dec 2019.

[13] M. Muramal, Q. Qu, and B. Nasir, "Renovating blockchain with distributed databases: An open source system," Future Generation Computer Systems-the International Journal Of Es Science, vol. 90, pp. 105-117, Jan 2019.

[14] N. Urs, L. Marchesi, M. Marchesi, and R. Tonelli, "Forecasting Bitcoin closing price series using linear regression and neural networks models," Peerj Computer Science, Jul 6 2020.

[15] B. Kapar and J. Olmo, "Analysis of Bitcoin prices using market and sentiment variables," World Economy, vol. 44, pp. 45-63, Jan 2021.

[16] Z. H. Munim, M. H. Shakil, and I. Alon, "Next-Day Bitcoin Price Forecast," Journal Of Risk And Financial Management, vol. 12, Jun 2019.

[17] D. T. Nguyen and H. V. Le, "Predicting the Price of Bitcoin Using Hybrid ARIMA and Machine Learning," Future Data And Security Engineering (Fdsse 2019), vol. 11814, pp. 696-704, 2019.

[18] R. Bohle and L. Rossini, "Comparing the Forecasting of Cryptocurrencies by Bayesian Time-Varying Volatility Models," Journal Of Risk And Financial Management, vol. 12, Sep 2019.

[19] P. Jiang, Z. K. Liu, J. Z. Wang, and L. F. Zhang, "Decomposition-selection-ensemble forecasting system for energy futures price forecasting based on multi-objective version of chaos game optimization algorithm," Resources Policy, vol. 73, Oct 2021.

[20] J. Yoon, "Forecasting of Real GDP Growth Using Machine Learning Models: Gradient Boosting and Random Forest Approach," Computational Economics, vol. 57, pp. 247-265, Jan 2021.

[21] S. Q. Qi, K. J. Jin, B. S. Li, and Y. F. Qian, "The exploration of internet finance by using neural network," Journal Of Computational And Applied Mathematics, vol. 369, May 1 2020.

[22] O. Sohaib, W. Hussain, M. Asif, M. Ahmad, and M. Mazzara, "A PLS-SEM Neural Network Approach for Evaluating Cryptocurrency Adoption," IEEE Access, vol. 8, pp. 13138-13150, 2020.

[23] L. Cocco, R. Tonelli, and M. Marchesi, "Predictions of bitcoin prices through machine learning based frameworks," Peerj Computer Science, Mar 29 2021.

[24] S. Alonso-Monsalve, A. L. Suarez-Cetruo, A. Cervantes, and D. Quintana, "Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators," Expert Systems with Applications, vol. 149, Jul 1 2020.

[25] R. Gupta, S. Tanwar, S. Tyagi, and N. Kumar, "Machine Learning Models for Secure Data Analytics: A taxonomy and threat model," Computer Communications, vol. 153, pp. 406-440, Mar 2020.

[26] M. Poongodi, A. Sharma, V. Vijayakumar, V. Bhardwaj, A. P. Sharma, R. Iqbal, et al., "Prediction of the price of Ethereum blockchain cryptocurrency in an industrial finance system," Computers & Electrical Engineering, vol. 81, Jan 2020.

[27] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: A new learning scheme of feedforward neural networks," 2004 IEEE International Joint Conference on Neural Networks, Vols 1-4, Proceedings, pp. 985-990, 2004.

[28] G. B. Huang, D. H. Wang, and Y. Lan, "Extreme learning machines: a survey," International Journal Of Machine Learning And Cybernetics, vol. 2, pp. 107-122, Jun 2011.

[29] M. Shariati, M. S. Mafiopour, P. Mehrabi, Y. Zandi, D. Dehghani, A. Bahadori, et al., "Application of Extreme Learning Machine (ELM) and Genetic Programming (GP) to design steel-concrete composite floor systems at elevated temperatures," Steel And Composite Structures, vol. 33, pp. 319-332, Nov 10 2019.

[30] I. Ahmad, M. Basheri, M. J. Iqbal, and A. Rahim, "Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection," IEEE Access, vol. 6, pp. 33789-33795, 2018.

[31] S. L. Sun, S. Y. Wang, and Y. J. Wei, "A new ensemble deep learning approach for exchange rates forecasting and trading," Advanced Engineering Informatics, vol. 46, Oct 2020.

[32] D. Aggarwal, S. Chandrasekaran, and B. Annamalai, "A complete empirical ensemble mode decomposition and support vector machine-based approach to predict Bitcoin prices," Journal Of Behavioral And Experimental Finance, vol. 27, Sep 2020.

[33] H. L. Lin, Q. B. Sun, and S. Q. Chen, "Reducing Exchange Rate Risks in International Trade: A Hybrid Forecasting Approach of CEEMDAN and Multilayer LSTM," Sustainability, vol. 12, Mar 2 2020.

[34] S. L. Sun, S. Y. Wang, Y. J. Wei, and G. W. Zhang, "A Clustering-Based Nonlinear Ensemble Approach for Exchange Rates Forecasting," IEEE Transactions On Systems Man Cybernetics-Systems, vol. 50, pp. 2284-2292, Jun 2020.

[35] W. Liu, W. D. Liu, and J. W. Gu, "Forecasting oil production using ensemble empirical model decomposition based Long Short-Term Memory neural network," Journal Of Petroleum Science And Engineering, vol. 189, Jun 2020.

[36] Y. Z. Li, S. Y. Wang, Y. J. Wei, and Q. Zhu, "A New Hybrid VMD-ICSS-BiGRU Approach for Gold Futures Price
Forecasting and Algorithmic Trading,” IEEE Transactions on Computational Social Systems, vol. 8, pp. 1357-1368, Dec 2021.

Z. W. a. N. E. HUANG, “ENSEMBLE EMPIRICAL MODE DECOMPOSITION: A NOISE-ASSISTED DATA ANALYSIS METHOD,” Advances in Adaptive Data Analysis, vol. 01, p. 41, 2009.

W. Sun and J. J. Zhang, “Carbon Price Prediction Based on Ensemble Empirical Mode Decomposition and Extreme Learning Machine Optimized by Improved Bat Algorithm Considering Energy Price Factors,” Energies, vol. 13, Jul 2020.

Q. D. Qin, K. Q. Xie, H. D. He, L. Li, X. H. Chu, Y. M. Wei, et al., “An effective and robust decomposition-ensemble energy price forecasting paradigm with local linear prediction,” Energy Economics, vol. 83, pp. 402-414, Sep 2019.

J. K. Xue and B. Shen, “A novel swarm intelligence optimization approach: sparrow search algorithm,” Systems Science & Control Engineering, vol. 8, pp. 22-34, Jan 1 2020.

H. C. Huang, J. Y. Chen, X. T. Huo, Y. F. Qiao, and L. Ma, “Effect of Multi-Scale Decomposition on Performance of Neural Networks in Short-Term Traffic Flow Prediction,” IEEE Access, vol. 9, pp. 50994-51004, 2021.

J. K. Xue and B. Shen, “A novel swarm intelligence optimization approach: sparrow search algorithm,” Systems Science & Control Engineering, vol. 8, pp. 22-34, Jan 1 2020.

K. E. Taylor, “Summarizing multiple aspects of model performance in a single diagram,” Journal of Geophysical Research-Atmospheres, vol. 106, pp. 7183-7192, Apr 16 2001.