IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 31 October 2023, accepted 18 November 2023, date of publication 21 November 2023,
date of current version 29 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3335609

RESEARCH ARTICLE

Dynamic Forecasting for Systemic Risk in China's Commercial Banking Industry Based on Sequence Decomposition and Reconstruction

ZHENGYONG LI[1], DEYIN FU[1,2], AND HAITING LI[3]
[1]School of Statistics and Data Science, Lanzhou University of Finance and Economics, Lanzhou 730020, China
[2]School of Labor Economics, China University of Labor Relations, Beijing 100048, China
[3]Office of Admission and Employment, Lanzhou University of Finance and Economics, Lanzhou 730020, China

Corresponding author: Zhengyong Li (lzy3303@outlook.com)

This work was supported in part by the 2023 Graduate 'Star of Innovation' Program under Grant 2023CXZX-700, in part by the Excellent Doctoral Project of Gansu Province under Grant 23JRRA1189, and in part by the Doctoral Research Innovation Project of Lanzhou University of Finance and Economics under Grant 2022D03.

ABSTRACT The Pressure Index of commercial banks is an effective measure of the systemic risk in the sector. This helps decision makers and market participants assess the potential levels of stress that commercial banks may face when confronted with impending risks. This study proposes a method for forecasting future trends in a Pressure Index for systemic risk prediction. The banking stress index is specifically constructed through an extreme value approach, followed by a non-stationary time series decomposition using variational mode decomposition (VMD). The number of decompositions was determined using the fuzzy entropy (FE) rule. These models were then used to construct autoregressive integrated moving average (ARIMA), artificial neural network (ANN), backpropagation neural network (BP), recurrent neural network (RNN), and long short-term memory (LSTM) models for independent prediction. The empirical results demonstrate the significant advantages of the VMD technique for forecasting non-linear and non-stationary complex time series. These findings highlight the substantial benefits of using VMD in forecasting intricate temporal patterns, especially in cases where traditional methods may face challenges in effectively capturing underlying dynamics. The VMD-ARIMA model showed superior prediction accuracy compared with the other models. Our study aims to model and forecast the data of the banking stress index, which is of utmost importance for the central bank in formulating macroeconomic policies and for commercial banks in managing credit risk.

INDEX TERMS China's commercial banking industry, systemic risk, forecasting, variational mode decomposition.

I. INTRODUCTION

With the progression of economic globalization and the flourishing development of financial markets, commercial banks have become increasingly interconnected with the entire financial system. This is primarily due to the expansion of their services and the growing trend of financial conglomerates. As a fundamental pillar of the financial system, commercial banks play a vital role in promoting a nation's economic growth. However, during financial crises, notable commercial banks, such as Lehman Brothers and Standard Chartered Bank, suffered severe blows, even triggering turbulence within the financial system. This finding highlights the importance of the commercial banking industry. On March 10, 2023, Silicon Valley Bank, the 16th largest bank in the United States, declared bankruptcy, with assets exceeding 200 billion dollars. This incident marked the largest bank failure in the United States since the financial crisis, leading regulatory authorities to prioritize financial risk prevention and control. The internationalization of commercial banks'

The associate editor coordinating the review of this manuscript and approving it for publication was Akin Tascikaraoglu.

development and the trend of financial innovation have increased the interdependence between commercial banks and other sectors of the economy. This has increased the likelihood of the systemic risk in the commercial banking industry. Accurately predicting and identifying systemic risks in the commercial banking sector in a timely manner are crucial. This proactive approach will aid in preventing and mitigating these risks, ultimately ensuring financial stability.

Current research on systemic risk warnings primarily focuses on the financial system, particularly on the construction and prediction of financial conditions or stress indices. However, limited attention has been paid to systemic risk warning in the commercial banking industry. Existing systemic risk-warning models can be broadly classified into two types: discrete and continuous. Early discrete warning models, such as the FR probability model [6], STV cross-sectional regression model [20], and KLR model [10], mainly focused on currency or international payment crises. These models utilized predetermined definitions of crises, employed virtual variables as dependent variables to determine crisis occurrence, and selected warning indicators as independent variables to construct regression equations and to examine the relationship between each warning indicator and a crisis. However, practical testing has revealed that these warning models do not effectively explain financial time series. First, a subjective setting of the crisis definitions and thresholds is necessary. Secondly, it is important to note that discretizing continuous financial stress variables may lead to a loss of information.

Recently, researchers have focused on developing continuous prediction frameworks to address the limitations of discrete warning models. One notable methodology is the Financial Stress Index (FSI) proposed [9]. They argue that external shocks and uncertainties, such as expected losses can impact economic entities and lead to financial stress. Consequently, stress indices were constructed for banks, stock markets, and foreign exchange markets. Reference [14] utilized the Financial Stress Index to track the trends and severity of overall risk pressure within the financial system over time. This index has the advantage of revealing and forecasting the potential systemic risk. Therefore, the Financial Stress Index is a valuable tool for assessing the level of the systemic risk in the banking industry. Thus, similar financial stress indices can be used to evaluate the systemic risk in the commercial banking industry. This allows for anticipation of an industry's systemic risk condition by predicting changes in the banking stress index. Conducting such research enhances our understanding of the risk landscape within the commercial banking industry and provides regulatory authorities and banking institutions with a scientific foundation for decision-making. In their study, Hao et al. [8] successfully constructed a stress index for the commercial banking industry using risk-free rate spreads (TED), non-performing loan rates (NPL), and interbank lending rates (RR). Building on this, our study adopts a stress index specifically for the Chinese commercial banking industry to assess its systemic risk condition and provide an early warning of the potential systemic risk by predicting changes in the banking stress index.

Numerous achievements have been made regarding the prediction of time series data. These forecasting methods can be broadly categorized into three groups: traditional econometric models, machine learning models, and deep learning models. Traditional econometric models such as autoregressive moving average (ARMA) [2], autoregressive integrated moving average (ARIMA) [2], and generalized autoregressive conditional heteroscedasticity (GARCH) [1], have limitations in effectively handling time series modeling problems with non-linear and non-stationary characteristics. However, with the advancement of technologies, such as artificial intelligence and big data, computer-based approaches are increasingly being employed in the financial domain. In this context, machine learning methods, including artificial neural networks (ANN) and support vector machines (SVM), have demonstrated superior performance in time series prediction. These models can effectively handle non-linear, discontinuous, and high-frequency multi-dimensional data, and have gained wide acceptance in financial forecasting. Compared to traditional econometric models, machine learning models are not constrained by stringent data requirements and can provide more accurate prediction results. However, traditional machine learning methods often face challenges in effectively capturing the correlation between time series data, owing to overfitting. To address this issue, deep learning methods such as recurrent neural networks (RNN) have been introduced. RNN, with their self-feedback and cyclic structures, are capable of handling the autocorrelation characteristics of time series data and have shown promising results in time series prediction. However, traditional RNN using backpropagation algorithms may encounter problems such as gradient disappearance or explosion, making it difficult to handle long-term dependency issues. As an improvement over RNN, long short-term memory (LSTM) has been proposed. LSTM can automatically retain longer historical information, specifically addressing long-term dependency problems and overcoming gradient disappearance and explosion issues faced by ordinary RNN. Traditional econometric models are well suited for simple time series prediction problems, whereas machine learning and deep learning models such as ANN, RNN, and LSTM offer advantages in handling complex time series prediction problems. Time series data often display intricate trends, and decomposing and reconstructing these data can be helpful.

Modal decomposition is an approach used to process non-linear and non-stationary sequences. This allows the transformation of irregular frequency data into residual wave components with singular frequencies. Common algorithms for time series decomposition include the wavelet transform, empirical mode decomposition (EMD), and variational mode decomposition (VMD). Rua and Nunes [19] analyzed the major developed economies over the past four decades using wavelet square correlation. They then applied the wavelet transform to analyze the resonance of market movement in terms of time and frequency variations. However, wavelet transform utilizes a filter bank for signal decomposition, which does not capture the instantaneous features of the signal. Li et al. [12] applied an adaptive approach to decompose the original time series into multiple intrinsic mode functions (IMFs) using EMD. They then employed a hybrid deep learning model based on Convolutional Neural Networks (CNN) and Gated Recurrent Units (GRU) to reconstruct and predict time series. However, EMD has limitations as it relies heavily on local extrema points within the sequence, leading to mode-mixing issues among the decomposed sub-signals [4]. To address these issues, VMD introduces a mathematical optimization framework that provides more robust and accurate decomposition results in certain scenarios. VMD is an adaptive and fully non-recursive method for mode variation and signal processing, which allows the determination of the number of mode decompositions based on the actual situation. It effectively overcomes the endpoint effects and mode component overlap problems observed in EMD. The VMD model is known for its ability to effectively address noise issues in the input signals. The optimized approach is simple and fast, providing satisfactory quantitative results for pitch detection and separation even when harmonic frequencies are absent. It also demonstrated strong qualitative results for both synthesized and real test signals, showing excellent resilience to signal noise. The VMD model has been widely used in various fields, such as signal processing [22], time-frequency analysis in seismology [24], and fault diagnosis [15]. In the realm of economic and financial data forecasting, [11] introduced groundbreaking work by combining VMD, considering it an effective and promising technique for analyzing and predicting economic and financial time series. Moreover, with the growing trend in interdisciplinary learning, the application of the VMD model in economic and financial research has expanded. This includes prediction of futures prices [13], oil prices [5], [7], [18], and stock prices [16], [17]. However, it is essential to investigate the effectiveness of VMD in modeling and predicting the banking stress index. Modeling and predicting banking stress index data are crucial for the central bank to formulate macroeconomic policies and for commercial banks to manage credit risks.

Currently,the measurement of systemic risks in the commercial banking industry is primarily based on bilateral balance sheets data. This approach focuses mainly on large listed banks that disclose their balance sheets [21], [25]. However, this method is biased because of the presence of numerous small and non-listed commercial banks in China. To address this challenge, we identify three indicators reflecting the entire commercial banking system: TED, NPL, and RR. These indicators form the basis for constructing a Commercial Banking Stress Index, which provides an assessment of systemic risk in the commercial banking industry. To improve the accuracy of predicting the Commercial Banking Stress Index, we employed the concept of 'decomposition-reconstruction' to address the forecasting and modeling issues of non-stationary time series. Initially, an intricate system is decomposed, and subsequently, artificial intelligence techniques are employed to effectively handle the non-linear nature and high complexity of the system. The decomposed parts were subsequently recombined for overall analysis and modeling using the reconstruction approach. Our framework consists of the following steps: 1) Preprocess the original time series data using the VMD decomposition algorithm to obtain several distinct sub-sequences. 2) Individual prediction modeling was conducted for each sub-sequence, considering their different characteristics. 3) The predicted values of each sub-sequence are summed to obtain a more accurate overall prediction output. Specifically, the commercial bank stress index was used to evaluate the systemic risk of the commercial banking industry. It helps monitor systemic risks by predicting future stress indices. In this study, the VMD algorithm was applied to the stress index to generate sub-sequences. These sub-sequences were then reconstructed based on the fuzzy entropy of the decomposition term. To predict each reconstructed sub-sequence, various models, such as the conventional ARIMA, ANN, backpropagation neural network (BP), RNN, and LSTM models, have been independently constructed. The decomposed parts were consolidated for comprehensive analysis and modeling. A multiple linear regression model was established using the regression coefficients of each sub-sequence for reconstruction, and the final prediction results were obtained. By comparing the evaluation criteria of this combined model with those of single traditional statistical, machine learning, and deep learning models, it was found that the combined model based on the decomposition-reconstruction methodology can more accurately predict the commercial bank stress index and enhance the effectiveness of early warnings for systemic risk in the commercial banking industry.

Our main contributions can be summarized as follows:

(1) We formulated a Commercial Banking Stress Index, which closely aligns with the actual state of systemic risk in the Chinese commercial banking industry.

(2) We extensively utilized the VMD technique to handle the complex time series. By decomposing the original sequence into a lower-complexity IMF time series, we leveraged the power of the VMD. However, determining the appropriate number of variation modes using the VMD algorithm is not straightforward [11]. To overcome this challenge, we employed the rule of FE to determine the number of variation modes. The VMD signal decomposition method effectively reduces the complexity of the time series and significantly enhances the predictive accuracy of the overall time series. Comparative experiments demonstrated

that our proposed VMD-ARIMA model achieves the best predictive performance.

(3) Our research expands the field of predicting the Commercial Banking Stress Index and extends the theoretical framework of time series forecasting methods based on the 'decomposition-reconstruction' approach, making a valuable contribution to this domain.

The article is structured as follows: Section II introduces the relevant models, Section III applies the concept of decomposition-reconstruction to empirically analyze the commercial bank stress index, and Section IV provides a summary.

II. COMPONENT MODELS

This section outlines the methodology used to construct our proposed Commercial Banking Stress Index and presents the systemic design principles of the decomposition-reconstruction model. However, detailed explanations of individual models such as autoregressive integrated moving average (ARIMA), artificial neural network (ANN), multilayer feedforward neural network (BP), recurrent neural network (RNN), and long short-term memory network (LSTM), are not provided in this paper.

A. COMMERCIAL BANKING STRESS INDEX

The modeling and prediction of systemic risks in the commercial banking industry are of utmost importance for central banks. This helps them to formulate and implement macroeconomic regulations, financial supervision, and financial stability. An accurate assessment of the systemic risk status of the commercial banking industry is crucial to provide risk warnings. A literature review of systemic risk warnings in the commercial banking industry highlights the effectiveness of the commercial bank stress index. It has proven to be an effective tool for revealing the systemic pressure situation in China's commercial banking industry and offers significant advantages for evaluating systemic risks.

In this study, we construct a commercial bank stress index for China's commercial banking industry, drawing inspiration from extreme value theory [8], [9], [14]. The instability of the commercial banking system is often accompanied by significant fluctuations in the risk-free interest rate spread (TED), non-performing loan ratio (NPL), and interbank lending rate (RR). Therefore, we used the rate of change of these three indicators as the basis for constructing a commercial bank stress index. We reflected on the systemic risk status of the commercial banking industry. The synthesized commercial bank stress index is expressed as follows:

$$
\begin{aligned}
BSI = {} & \frac{1}{\frac{1}{\sigma_{TED}} + \frac{1}{\sigma_{NPL}} + \frac{1}{\sigma_{RR}}} \\
& \times \left(\frac{1}{\sigma_{TED}} \frac{TED_t - \min(TED_t)}{\max(TED_t) - \min(TED_t)} \right. \\
& + \frac{1}{\sigma_{NPL}} \frac{NPL_t - \min(NPL_t)}{\max(NPL_t) - \min(NPL_t)} \\
& \left. + \frac{1}{\sigma_{RR}} \frac{RR_t - \min(RR_t)}{\max(RR_t) - \min(RR_t)} \right) \quad (1)
\end{aligned}
$$

where, σ_{TED},σ_{NPL},and σ_{RR} represent the standard deviations of commercial banks' risk-free interest rate spreads, non-performing loan ratios, and interbank lending rates, respectively. BSI is the banking stress index. TED_t represents the interest rate spread of risk-free rates for commercial banks over a period of t, which is obtained by subtracting the risk-free rate from the interbank lending rate with a three-month maturity. An increase in the interest rate spread signifies insufficient liquidity and a rapidly expanding bank risk. NPL_t represents the non-performing loan ratio for commercial banks over period of t. A higher ratio indicates a higher level of bad debt risk in the commercial banking system. RR_t represents the weighted average interest rate for a 7-day pledged repurchase agreement in the interbank market. A higher repo rate suggests tighter market funding and a higher crisis level.

B. VMD MODEL

Dragomiretskiy and Zosso [4] proposed a variational mode decomposition technique, which is essentially an iterative variational framework that automatically searches for the optimal solution of the variational problem by minimizing the total bandwidth of all modes. The banking stress index can be considered as a signal that reflects market information of commercial banks. Mode decomposition treats the signal as a superposition of sub-signals from different 'modes', while VMD views the signal as a superposition of sub-signals at various frequencies. The purpose of VMD is to decompose a signal into a series of sparse component signals at different frequencies. The main idea of VMD is to decompose the input signal into k finite bandwidths, with each mode identified as mode u_k, striving to minimize the total estimation of the bandwidths for all modes. The signal decomposition process is directly related to the resolution of the variational problems. The constrained variational problem is expressed as:

$$
\begin{cases}
\min\limits_{\{u_k\},\{w_k\}} \left\{ \sum_{k=1}^{K} \left\| \partial_t \left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) e^{-jwk^t} \right\|_2^2 \right\} \\
\text{s.t. } \sum_{k=1}^{K} u_k = f
\end{cases} \quad (2)
$$

Among these variables, $\{u_k\} = \{u_1, u_2, \ldots, u_k\}$ represents a modal component, $\{w_k\} = \{w_1, w_2, \ldots, w_k\}$ denotes the corresponding central frequency for each modal component, K signifies the number of decomposed modes, ∂_t corresponds to the partial derivative of t, $\delta(t)$ is identified as an impulse function, u_k is the k-th sub-sequence of f, and f denotes the original input signal.

To obtain the optimal solution for the constrained variational mode, we introduce the penalizing function α and Lagrangian operator λ, thereby transforming the constrained variational problem into an under-constrained variational problem:

$$L(\{u_k\},\{w_k\},\lambda)=\alpha\sum_k\left\|\partial_t\left(\delta(t)+\frac{j}{\pi t}\right)u_k(t)e^{-jwk^t}\right\|_2^2+\left\|f(t)-\sum_k u_k(t)\right\|_2^2+\left\langle\lambda(t),f(t)-\sum_k u_k(t)\right\rangle \quad (3)$$

The optimum solution was obtained using the alternating direction multiplier algorithm, which identifies the saddle point of the Lagrangian function. The specific steps are as follows:

Step 1. Initialize each modal component $\{u_k^1\}$, center frequency $\{w_k^1\}$, λ^1, $n = 0$, transforming them into the frequency domain.

Step 2. Update the modal component within the non-negative frequency range:

$$\frac{\hat{f}(w)-\sum_{i<k}\hat{u}_i^{n+1}(w)-\sum_{i>k}\hat{u}_i^{n}(w)+\frac{\hat{\lambda}^n(w)}{2}}{1+2\alpha\left(w-w_k^n\right)^2}\rightarrow\hat{u}_k^{n+1}(w) \quad (4)$$

Step 3. Update the center frequency w_k within the non-negative frequency range:

$$\frac{\int_w^\infty w\left|\hat{u}_k^{n+1}(w)\right|^2dw}{\int_w^\infty\left|\hat{u}_k^{n+1}(w)\right|^2dw}\rightarrow w_k^{n+1} \quad (5)$$

Step 4. Update the Lagrangian operator λ within the non-negative frequency range:

$$\hat{\lambda}^n+\tau\left(\hat{f}(w)-\sum_k\hat{u}_k^{n+1}(w)\right)\rightarrow\hat{\lambda}^{n+1} \quad (6)$$

Step 5. The iteration stops when the following formula is satisfied, given a specific decision precision $\varepsilon > 0$:

$$\frac{\sum_k\left\|\hat{u}_k^{n+1}-\hat{u}_k^n\right\|_2^2}{\left\|\hat{u}_k^n\right\|_2^2}<\varepsilon \quad (7)$$

where $\hat{f}(w)$, $\hat{u}_k^{n+1}(w)$, and $\hat{\lambda}^{n+1}$ are the respective Fourier transforms of $f(t)$, u_k^{n+1}, and λ^{n+1}.

The optimal solution for the constrained variational model was obtained using the alternating direction method of multipliers (ADMM). This method allows decomposition of the original signal into K narrowband modal components.

C. FUZZY ENTROPY

Fuzzy entropy (FE) utilizes an exponential function to fuzzify the formula for measuring similarity. As the time series of a signal becomes more complex, its entropy value increases [23]. For a time series $\{X(i), i = 1, 2, \ldots, N\}$ of length N, the specific steps of FE are as follows:

Step 1. To analyze the time series data, we first divided each sequence into subsequences of length m. Each subsequence is represented as follows:

$$X_i^m(t)=\{x_i(t),x_{i+1}(t),\ldots,x_{i+m-1}(t),i=1,2,\ldots,N-m+1\} \quad (8)$$

Step 2. We calculated the distance between each subsequence and all k subsequences. The distance was determined by calculating the maximum absolute difference between the corresponding elements of the two vectors.

$$d_{ij}^m=\max\left|\left[x_{i+k}(t)-x_{j+k}(t)\right]\right|,k=0,1,\ldots,m-1\mid \quad (9)$$

Step 3. We define a fuzzy function $\mu\left(d_{ij}^m,n,r\right)$ to measure the similarity between the entities. Parameters n and r represent the gradient and width of the boundaries of the fuzzy function, respectively.

$$D_{ij}^m=\mu\left(d_{ij}^m,n,r\right)=e^{-\left(d_{ij}^m/r\right)^n} \quad (10)$$

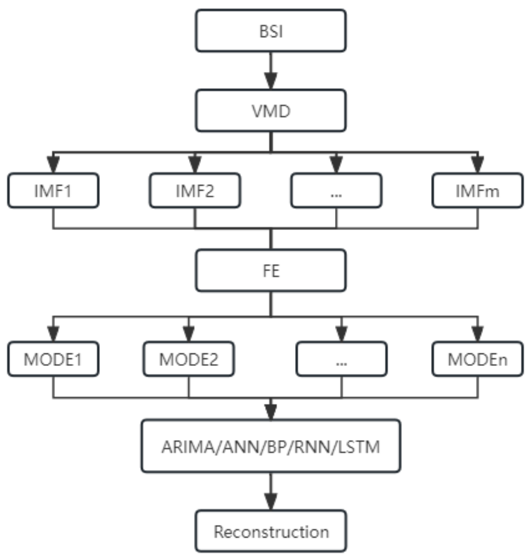

FIGURE 1. Decomposition and reconstruction modeling.

Step 4. We then average the similarity over all subsequences except itself.

$$\phi^m(n,r)=\frac{1}{N-m}\sum_{j=1}^{N-m}\left(\frac{1}{N-m-1}\sum_{j=1,j\neq i}^{N-m}D_{ij}^m\right) \quad (11)$$

Step 5. We increment the dimensionality from m to $m+1$ and repeat Steps 1-4.

$$\phi^{m+1}(n,r)=\frac{1}{N-m}\sum_{i=1}^{N-m}\left(\frac{1}{N-m-1}\sum_{N-m}D_{ij}^{m+1}\right) \quad (12)$$

Step 6. We can calculate the fuzzy entropy $FuzzyEn(m, n, r, N)$.

$$FuzzyEn(m, n, r, N) = \ln \phi^{m}(n, r) - \ln \phi^{m+1}(n, r) \quad (13)$$

D. DECOMPOSITION AND RECONSTRUCTION PREDICTIVE MODEL

The business bank stress index falls under the category of a typical non-linear and non-stationary time series. However, relying solely on conventional econometric models, machine learning, or deep learning for modeling has a limited predictive efficacy. To overcome this limitation, we utilized the VMD technique, which has significant advantages in handling non-stationary time series. We construct a composite model by employing decomposition and reconstruction methods to forecast the systemic risk in the commercial banking industry. The detailed modeling process is shown in Figure 1.

Step 1. The BSI was disassembled into various IMF components using VMD.

Step 2. Each IMF component, $IMF_1, IMF_2, \ldots, IMF_m$, is then reconstructed into new subsequences $MODE_1, MODE_2, \ldots, MODE_n$, where $n < m$, based on their varying FE ranges.

Step 3. Separate predictions were conducted on these new subsequences using models such as ARIMA, ANN, BP, RNN, and LSTM.

Step 4. A multivariate linear regression model was constructed using the decomposed sequences, and the coefficients for this regression were obtained. The sequence is then reconstructed based on this regression to obtain the final prediction results. The reconstructed predictions of each MODE subsequence were aggregated to reconstruct the BSI. Finally, the reconstructed BSI is subjected to backtesting.

E. EVALUATION CRITERIA

Six evaluation metrics were utilized to assess the performance of each predictive model. Among them, MSE, MAE, RMSE, MAPE, and SMAPE serve as measures of accuracy, whereas *D_stat* represents a vertical precision indicator. The expressions for these metrics are as follows. Mean Square Error (MSE):

$$MSE = \frac{1}{n} \sum \left(\hat{y}_i - y_i\right)^2 \quad (14)$$

Mean Absolute Error (MAE):

$$MAE = \frac{1}{n} \sum \left|\hat{y}_i - y_i\right| \quad (15)$$

Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum \left(\hat{y}_i - y_i\right)^2} \quad (16)$$

Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left|\frac{\hat{y}_i - y_i}{y_i}\right| \quad (17)$$

Symmetric Mean Absolute Percentage Error (SMAPE):

$$SMAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{\left|\hat{y}_i - y_i\right|}{\left(\hat{y}_i + y_i\right)/2} \quad (18)$$

D Statistical measures(D_stat):

$$D_{\text{stat}} = \frac{1}{M} \sum_{i=1}^{M} a_i \times 100\%, \qquad a_i = I\left(\left(f_{i+1} - y_i\right)\left(y_{i+1} - y_i\right) \geq 0\right) \quad (19)$$

where, $\hat{y}_i$ represents the predicted BSI index values and y_i denotes the actual BSI index values. n represents the sample size, and i represents the specific ranking number of the sample. The smaller the values of MSE, MAE, RMSE, and SMAPE, the higher is the corresponding level of predictive accuracy. $I(\bullet)$ serves as an indicator function, taking a value of 1 when the condition is true and 0 otherwise. A larger *D_stat* value indicated a higher degree of directional accuracy in the model.

III. EMPIRICAL ANALYSIS

In this study, we present a concrete empirical design, comparative analysis of model fitting effectiveness, and comparative analysis of model prediction efficacy.

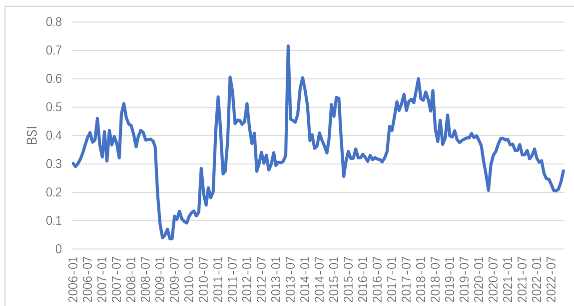

FIGURE 2. Commercial bank stress index.

A. DATA DESCRIPTION AND BSI INDEX CONSTRUCTION

To assess the systemic risk in China's banking sector, we used the Bank Stress Index. Our analysis is based on monthly data from January 2006 to December 2022, including the risk-free interest rate spread (TED), non-performing loan rate (NPL), and interbank lending rate (RR) of commercial banks. The data were obtained from the WIND database. To ensure consistency, we converted quarterly NPL data into monthly data using the cubic spline interpolation method. The overall sample consisted of 204 months, with the first 180 months used as training samples and the remaining 24 months as testing samples for model building. Figure 2 shows the synthesized commercial banking stress index.

The figure shows that the index has reached elevated levels over several periods. In 2007, this was the primary reason for the economic downturn caused by the credit crisis, which intensified the systemic risk in the commercial banking

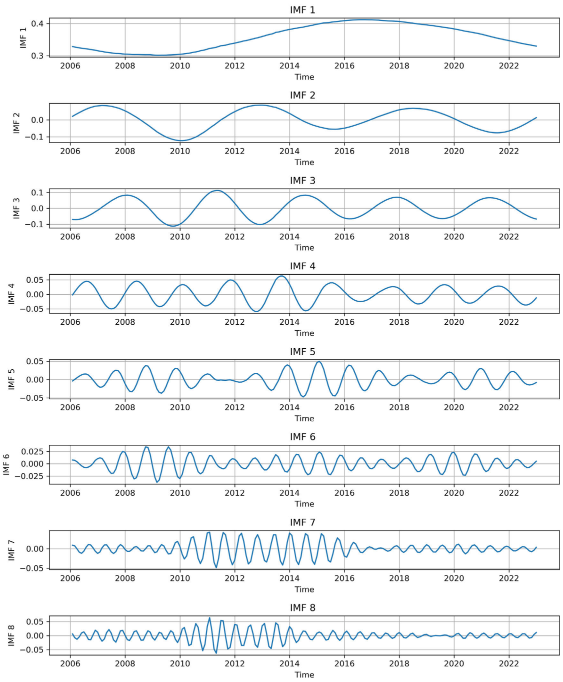

FIGURE 3. Decomposition results of BSI.

sector. In 2008, the index was influenced by the economic crises. In 2011, it was affected by the European debt crisis, which posed a significant risk to China's commercial banking industry. In 2013, the 'liquidity crunch' emerged in China. In 2015, the stock market crash occurred in China, and in 2018, the implementation of the financial deleveraging policy exposed the commercial banking sector to considerable systemic risk. This index closely reflects the systemic risk situation of China's commercial banking industry and aligns with the industry's actual conditions.

B. VMD AND FE

Before applying the decomposition technique to the commercial banking stress index series, it is necessary to select an appropriate number of components denoted as n. To determine this, we follow an empirical rule and choose n as an approximation to the sample data size, where n is a power of 2. In this study, we found that the optimal number of components n, is 8. Using the VMD method, we decomposed BSI into eight distinct modes. The decomposition results are presented in Figure 3, where modes 1 to 8 are arranged from top to bottom in ascending order of frequency from low to high.

However, an excessive number of modes can increase the computational burden and lead to prediction errors [7]. Therefore, a method to reduce the computational burden and minimize the prediction errors is to reconstruct similar modes. FE is a metric used to quantify the complexity of a time series, and is an improved method of approximate entropy and sample entropy [3]. In this study, we employed FE to estimate the complexity of all sub-modes obtained from VMD. The FE values for each mode are listed in Table 1.

TABLE 1. FE value of IMFs.

IMF	Value
IMF_1	0.00232
IMF_2	0.01224
IMF_3	0.01904
IMF_4	0.01154
IMF_5	0.00823
IMF_6	0.00487
IMF_7	0.00775
IMF_8	0.00682

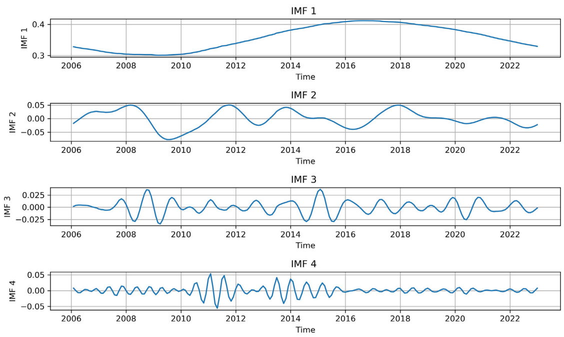

FIGURE 4. Reconstruction modes.

The complexity range of Mode 1 and Mode 6 was between 0.002 and 0.005, whereas that of Mode 2 and Mode 4 was between 0.010 and 0.015. Modes 5, 7, and 8 had complexity range of 0.006-0.010. Based on these complexity ranges, we considered Mode 3 to be an independent mode. We combined Mode 1 with Mode 6, Mode 2 with Mode 4, and Modes 5, 7, and 8 to form new modes. Finally, we reconstruct these eight modes into four new modes. Figure 4 illustrate all the reconstructed modes.

Subsequently, four new modes were used to construct a multivariate linear regression model for the reconstruction. These coefficients can be obtained through regression, and was used as the foundation for sequence reconstruction, which ultimately yielded the final predictive outcome. The primary manifestations of the multivariate linear regression model are as follows:

$$\hat{y_i} = \hat{\alpha_1} x_{i1} + \hat{\alpha}_2 x_{i2} + \hat{\alpha}_3 x_{i3} + \hat{\alpha}_4 x_{i4} + \varepsilon \tag{20}$$

where x represents the four new modes; y represents the predicted value; $\hat{\alpha_1}$, $\hat{\alpha_2}$, $\hat{\alpha_3}$, and $\hat{\alpha_4}$ represent the estimated regression coefficients,; and ε represents the random disturbance term.

C. COMPARISON OF MODEL FITTING RESULTS

To validate the remarkable advantages of the VMD technique, we compared the fitting performances of the various models (ARIMA, ANN, BP, RNN, and LSTM) without incorporating the VMD algorithm, with the corresponding models (VMD-ARIMA, VMD-ANN, VMD-BP, VMD-RNN, and VMD-LSTM) incorporating the VMD algorithm. Figure 5 shows that when dealing with multimodal data, mode decomposition often leads to a decline in the performance of the RNN model. This decline can be attributed to issues such

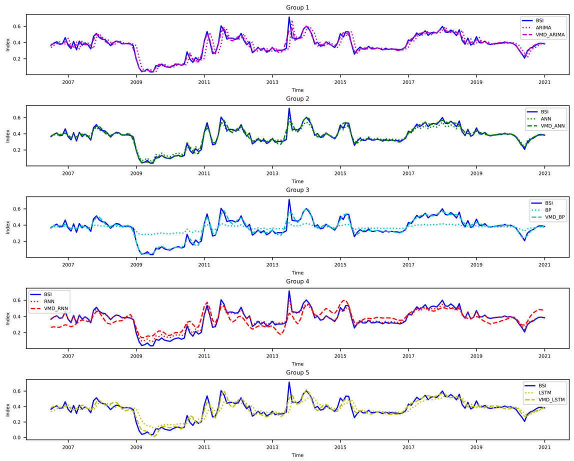

FIGURE 5. Fitting results of BSI.

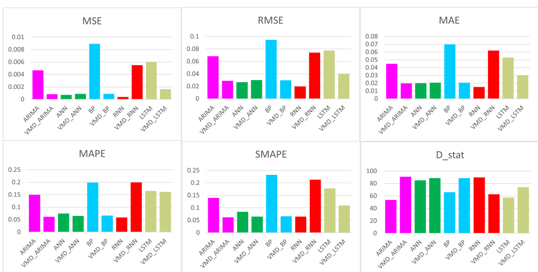

FIGURE 6. Evaluating the fitting performance of different models.

TABLE 2. Comparison of fitting accuracy among different models.

Index	ARIMA	ANN	BP	RNN	LSTM	ARIMA	VMD_ANN	VMD_BP	VMD_RNN	VMD_LSTM
MSE	0.00467	0.00071	0.00892	0.00038	0.00599	0.00083	0.00089	0.00087	0.00550	0.00162
RMSE	0.06835	0.02666	0.09444	0.01958	0.07739	0.02888	0.02985	0.02957	0.07414	0.04020
MAE	0.04485	0.02000	0.06989	0.01504	0.05277	0.01967	0.02069	0.02070	0.06195	0.03007
MAPE	0.14964	0.07450	0.19910	0.05854	0.16497	0.06158	0.06516	0.06620	0.19918	0.16165
SMAPE	0.13973	0.08418	0.23252	0.06463	0.17792	0.06156	0.06469	0.06559	0.21262	0.10966
D_stat	53.4483	85.0575	66.0920	89.6552	57.4713	90.8046	88.5056	88.5058	62.6437	74.1379

as data loss, inadequate intermodal information exchange, and insufficient modeling of modal correlations. The impact of VMD on the fitting performance was insignificant for the ANN, BP, and RNN models. However, the ARIMA and LSTM models exhibited less sensitivity to intermodal information exchange, making them less affected by mode decomposition. The models combined with the VMD algorithm showed closer proximity between the fitted values and actual values of the test samples, thereby significantly improving the fitting performance of the trends.

Table 2 presents a visual representation of the evaluation metrics used to compare the fitting performances of different models in the future. These metrics included MSE, RMSE, MAE, MAPE, SMAPE, and *D_stat* for each model. To facilitate an easier comparison between the models, we have provided Figure 6.

Based on the data presented in Table 2 and Figure 6, it is clear that the models incorporating VMD technology, with the exception of RNN, consistently show lower values of MSE, RMSE, MAE, and SMAPE than the models that did not incorporate VMD. In addition, these models improved the accuracy of direction prediction.

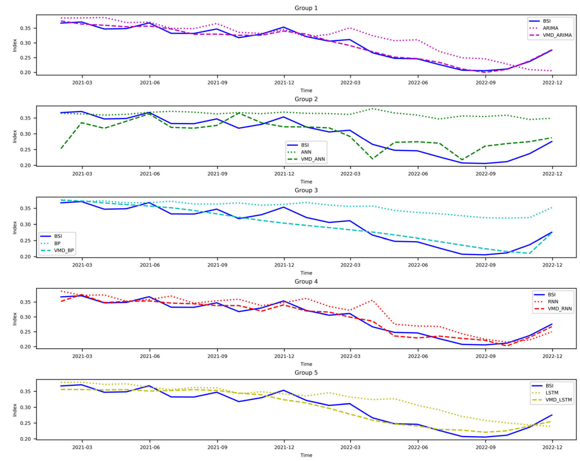

FIGURE 7. Prediction results of BSI.

TABLE 3. Comparison of prediction accuracy among different models.

Model	MSE (Rank)	RMSE (Rank)	MAE (Rank)	MAPE (Rank)	SMAPE (Rank)	D_stat (Rank)	Average (Rank)
ARIMA	0.001230 (6)	0.035064 (6)	0.028737 (6)	0.099305 (6)	0.102631 (6)	50.000000 (7)	6
ANN	0.006645 (10)	0.081517 (10)	0.063986 (10)	0.178005 (10)	0.209261 (10)	45.454545 (8)	10
BP	0.004488 (9)	0.066991 (9)	0.054789 (9)	0.160344 (9)	0.183670 (9)	45.454545 (8)	9
RNN	0.000908 (5)	0.030140 (5)	0.023307 (5)	0.074945 (5)	0.079091 (5)	45.454545 (8)	5
LSTM	0.001457 (8)	0.038175 (8)	0.031803 (8)	0.103400 (7)	0.111343 (8)	59.090909 (4)	8
VMD_ARIMA	0.000079 (1)	0.008901 (1)	0.007136 (1)	0.023118 (1)	0.023030 (1)	86.363636 (1)	1
VMD_ANN	0.001452 (7)	0.038110 (7)	0.029258 (7)	0.104575 (8)	0.103448 (7)	54.545455 (5)	7
VMD_BP	0.000381 (4)	0.019512 (4)	0.016443 (4)	0.058066 (4)	0.057704 (4)	54.545455 (5)	4
VMD_RNN	0.000154 (2)	0.012421 (2)	0.011173 (2)	0.039536 (2)	0.039703 (2)	77.27273 (2)	2
VMD_LSTM	0.000259 (3)	0.016100 (3)	0.013540 (3)	0.045697 (3)	0.045670 (3)	72.72727 (3)	3

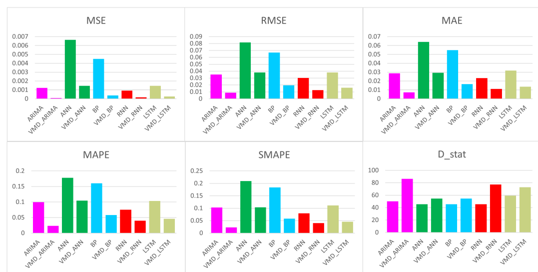

FIGURE 8. Evaluating the predictive performance of different models.

D. COMPARISON OF MODEL PREDICTION RESULTS

In our study, we utilized a real-time prediction strategy called 'one-step-ahead rolling forecast'. This strategy involves a continuous update of the training set by adding the actual data that follows each prediction. To evaluate the performance of various models, including ARIMA, ANN, BP, RNN, and LSTM, as well as their combinations with the VMD algorithm (VMD-ARIMA, VMD-ANN, VMD-BP, VMD-RNN, and VMD-LSTM), graphical representations of the prediction results were presented.

The results in Figure 7 show that the model combinations using the VMD algorithm have predictive values that are significantly closer to the actual values of the test samples. Although VMD algorithm-based models may not have a satisfactory fitting performance for local extrema, they can still provide early indications of upward or downward pressure trends, making them better warning signals. To further compare the predictive performances of the different models, Table 3 displays a visual representation of the evaluation metrics, including MSE, RMSE, MAE, MAPE, SMAPE, and

D_{stat}, for each model. To facilitate an easier comparison between the models, we have provided Figure 8.

Table 3 and Figure 8 present the evaluation criteria for the predictive results of the models. The table and figure show that the model combinations incorporating the VMD algorithm and reconstruction technique had significantly lower metrics (MSE, RMSE, MAE, MAPE, and SMAPE) than those without incorporation. Furthermore, the accuracy of the directional prediction (*D_stat*) was significantly improved. This indicates that the models incorporating the VMD algorithm and reconstruction technique demonstrate superior predictive performance compared with the non-incorporated models, effectively highlighting the superiority of the decomposition and reconstruction techniques. In contrast, traditional statistical, machine learning, and deep learning models struggle to predict sequence trends accurately without incorporating the VMD algorithm or reconstruction techniques. However, models incorporating the VMD algorithm and reconstruction techniques have significantly enhanced the predictive capability. In particular, the *VMD_ARIMA* model outperformed the machine learning and deep learning-based reconstruction models in terms of the reconstruction effect. This suggests that the *VMD_ARIMA* model accurately predicts the changing trend of price time series and exhibits excellent stability.

The complexity of the time series can be reduced using the VMD signal decomposition method. This method breaks down the original sequence into IMF time series with lower complexity, making it easier to apply machine learning algorithms and improve the predictive accuracy of each IMF sequence. The complexity of all VMD sub-modes was estimated using FE, and the approximate sub-sequences were reconstructed as new modes for individual prediction and reconstruction. Consequently, the overall time series prediction accuracy was significantly improved. Therefore, the decomposition and reconstruction methods effectively enhance the predictive performance of the entire sequence.

IV. CONCLUSION

This study constructs a stress index using data on the risk-free interest rate spread (TED), non-performing loan rate (NPL), and interbank lending rate (RR) in the Chinese commercial banking industry. The purpose of this index is to measure the systemic risk level of the industry and serve as an early warning system for risk assessment by predicting future trends. Drawing upon decomposition and reconstruction techniques as well as research achievements in econometrics, artificial intelligence, and machine learning, a combined decomposition and reconstruction model is proposed for predicting systemic risks in the commercial banking industry. The following conclusions were drawn from the empirical analysis.

(1) We employed the extreme value method to determine the stress index for China's commercial banking industry. The movement of this index aligns closely with the risk situation encountered by China's commercial banking sector. The prompt compilation and release of China's real-time commercial bank stress index can effectively serve as a vital indicator of systemic financial risk, accurately assessing the magnitude of systemic risk within China's commercial banking industry.

(2) Using the VMD signal decomposition method, this study decomposes the non-stationary stress index of the commercial banking industry into signals of different frequencies. Subsequently, sub-sequences with distinct frequency and amplitude characteristics were extracted. Fuzzy entropy (FE) was employed to estimate the complexity of all VMD decomposed sub-modes. Similar entropy values were then utilized to reconstruct new modes with the aim of reducing cumulative prediction errors. Independent predictions were conducted for all the new modes, and a multivariate regression model was established to obtain the final predicted output. The findings reveal that the decomposition and reconstruction methods significantly enhance forecasting accuracy for systemic risks in the commercial banking industry.

(3) Through comparison, the VMD-ARIMA model leverages the VMD decomposition and reconstruction technique to handle complex time series, while also incorporating the performance advantages of the ARIMA model for stationary time series. Compared to other models, the VMD-ARIMA model exhibits clear superiority in predicting systemic risks in the commercial banking industry.

Although our approach yields positive results, it is important to acknowledge its limitations. First, the stress index for the commercial banking industry comprises only three variables: the risk-free interest rate spread (TED), non-performing loan ratio (NPL), and interbank lending rate (RR). This limited set of variables imposes certain constraints. Future indicators should incorporate forward-looking information to enhance the effectiveness of the bank stress index as an early warning system. Secondly, our commercial bank stress index is limited to a monthly indicator, which hinders its ability to provide timely reference information to regulatory authorities. Finally, our prediction of the stress index did not consider the impact of other significant factors or exogenous variables.

In future research, we propose to explore higher-frequency data, such as weekly or daily data, to construct a hybrid commercial banking stress index. This improves the timeliness of our predictions. Additionally, to enhance the accuracy of our forecasts, future studies can investigate the integration of other variables, such as GDP and M1, which are macroeconomic indicators, as well as non-structural indicators, such as news media and search index data. We plan to delve into these meaningful inquiries in future research.

REFERENCES

[1] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *J. Econometrics*, vol. 31, no. 3, pp. 307–327, Apr. 1986.
[2] T. G. Wilson, *Time Series Analysis: Forecasting and Control*, G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, 5th ed. Hoboken, NJ, USA: Wiley, 2016, p. 712.

[3] W. Chen, J. Zhuang, W. Yu, and Z. Wang, "Measuring complexity using FuzzyEn, ApEn, and SampEn," *Med. Eng. Phys.*, vol. 31, no. 1, pp. 61–68, Jan. 2009.

[4] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.

[5] E. Jianwei, Y. Bao, and J. Ye, "Crude oil price analysis and forecasting based on variational mode decomposition and independent component analysis," *Phys. A, Stat. Mech. Appl.*, vol. 484, pp. 412–427, Oct. 2017.

[6] J. A. Frankel and A. K. Rose, "Currency crashes in emerging markets: An empirical treatment," *J. Int. Econ.*, vol. 41, nos. 3–4, pp. 351–366, Nov. 1996.

[7] J. Guo, Z. Zhao, J. Sun, and S. Sun, "Multi-perspective crude oil price forecasting with a new decomposition-ensemble framework," *Resour. Policy*, vol. 77, Aug. 2022, Art. no. 102737.

[8] Q. Hao, J. H. Shen, and C.-C. Lee, "Risk contagion of bank-firm loan network: Evidence from China," *Eurasian Bus. Rev.*, vol. 13, no. 2, pp. 341–361, Jun. 2023.

[9] M. Illing and Y. Liu, "An index of financial stress for Canada," Staff Working Papers 03-14, 2003, vol. 29.

[10] G. Kaminsky, S. Lizondo, and C. M. Reinhart, "Leading indicators of currency crises," *Staff Papers*, vol. 45, no. 1, pp. 1–48, Mar. 1998.

[11] S. Lahmiri, "A variational mode decompoisition approach for analysis and forecasting of economic and financial time series," *Exp. Syst. Appl.*, vol. 55, pp. 268–273, Aug. 2016.

[12] M.-W. Li, D.-Y. Xu, J. Geng, and W.-C. Hong, "A ship motion forecasting approach based on empirical mode decomposition method hybrid deep learning network and quantum butterfly optimization algorithm," *Nonlinear Dyn.*, vol. 107, no. 3, pp. 2447–2467, Feb. 2022.

[13] W. Liu, C. Wang, Y. Li, Y. Liu, and K. Huang, "Ensemble forecasting for product futures prices using variational mode decomposition and artificial neural networks," *Chaos, Solitons Fractals*, vol. 146, May 2021, Art. no. 110822.

[14] X. Liu and F. Lei, "Construction and effectiveness test of financial stress index—Empirical analysis based on China data," *J. Ind. Eng. Eng. Manag.*, vol. 26, pp. 1–6, Jan. 2012.

[15] Y. Miao, M. Zhao, and J. Lin, "Identification of mechanical compound-fault based on the improved parameter-adaptive variational mode decomposition," *ISA Trans.*, vol. 84, pp. 82–95, Jan. 2019.

[16] H. Nasiri and M. M. Ebadzadeh, "Multi-step-ahead stock price prediction using recurrent fuzzy neural network and variational mode decomposition," 2022, *arXiv:2212.14687*.

[17] H. Niu, K. Xu, and W. Wang, "A hybrid stock price index forecasting model based on variational mode decomposition and LSTM network," *Int. J. Speech Technol.*, vol. 50, no. 12, pp. 4296–4309, Dec. 2020.

[18] Z.-J. Peng, C. Zhang, and Y.-X. Tian, "Crude oil price time series forecasting: A novel approach based on variational mode decomposition, time-series imaging, and deep learning," *IEEE Access*, vol. 11, pp. 82216–82231, 2023.

[19] A. Rua and L. C. Nunes, "International comovement of stock market returns: A wavelet analysis," *J. Empirical Finance*, vol. 16, no. 4, pp. 632–639, Sep. 2009.

[20] J. D. Sachs, A. Tornell, A. Velasco, and C. R. N. Cooper, "Financial crises in emerging markets: The lessons from 1995," *Brookings Papers Econ. Activity*, vol. 1996, no. 1, pp. 147–215, 1996.

[21] G.-J. Wang, Z.-Q. Jiang, M. Lin, C. Xie, and H. E. Stanley, "Interconnectedness and systemic risk of China's financial institutions," *Emerg. Markets Rev.*, vol. 35, pp. 1–18, Jun. 2018.

[22] Y. Wang and R. Markert, "Filter bank property of variational mode decomposition and its applications," *Signal Process.*, vol. 120, pp. 509–521, Mar. 2016.

[23] J. Xiang, C. Li, H. Li, R. Cao, B. Wang, X. Han, and J. Chen, "The detection of epileptic seizure signals based on fuzzy entropy," *J. Neurosci. Methods*, vol. 243, pp. 18–25, Mar. 2015.

[24] Y.-J. Xue, J.-X. Cao, D.-X. Wang, H.-K. Du, and Y. Yao, "Application of the variational-mode decomposition for seismic time-frequency analysis," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 8, pp. 3821–3831, Aug. 2016.

[25] X. Zhang, X. Zhang, C.-C. Lee, and Y. Zhao, "Measurement and prediction of systemic risk in China's banking industry," *Res. Int. Bus. Finance*, vol. 64, Jan. 2023, Art. no. 101874.

ZHENGYONG LI received the M.S. degree in finance from the Lanzhou University of Finance and Economics, China, in 2021, where he is currently pursuing the Ph.D. degree in statistics. He has published several original articles in some academic journals. His research interests include China's commercial banking systemic risk and risk management.

DEYIN FU received the Ph.D. degree in economics from the Renmin University of China, China, in 1997. He is currently a Professor with the School of Statistics and Data Science, Lanzhou University of Finance and Economics. He has successfully completed 12 national and provincial social science projects. He has authored eight books and published over 60 articles in professional journals, including the *Journal of Cleaner Production*, *Statistical Research*, and *Journal of Statistics Information*. His research interests include multivariate statistical analysis, economic statistics, and applied statistics. Furthermore, he holds esteemed positions, including the Vice Chairperson of the Gansu Federation of Social Sciences, a member of the Education Executive Committee of the Statistics Undergraduate Major of the Ministry of Education, the Vice President of the China Statistical Society, the Vice President of the Gansu Provincial Statistical Society, and the Vice President of the Gansu Provincial Finance Society.

HAITING LI received the M.S. degree in economics from the Lanzhou University of Finance and Economics, China, in 2022. Her research interests include risk management and forecasting.