J. COMBIN. MATH. COMBIN. COMPUT. 127a (2025) 5197--5213

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Research on financial market volatility prediction and risk response strategy based on LSTM network

Xinjie Chen[1, ✉]

[1] School of Economics, Jinan University, Guangzhou, Guangdong, 510630, China

ABSTRACT
In this paper, the autoregressive moving average model (ARMA) and LSTM deep neural network are first introduced, and the time series are decomposed into high volatility components and low volatility components by MA filtering method. Then the time series forecasting model ARMA and deep neural network LSTM are combined on the basis of MA filtering method to form ARMA-LSTM combination model based on MA filtering method, and the application effect of this model in financial market volatility forecasting and risk response is verified through empirical evidence. The results show that the ARMA-LSTM_t model will achieve relatively good results in predicting the GDP_IG of the current year using the data of the 12 months of the current year and the last month of the previous year, and the training and prediction sets of the ARMA-LSTM combination model proposed in this paper have the best results. In addition, there is a positive relationship between investment-related indicators and GDP_IG, and the addition of investment network search data improves the estimation accuracy of the model, obtains smaller prediction errors, and improves the prediction accuracy of the ARMA-LSTM model in the short and medium term.

Keywords: Autoregressive moving average model (ARMA); LSTM deep neural network; MA filtering method; volatility forecasting; financial market

1. Introduction

After the 1970s, many anomalies have emerged in the world financial market that cannot be explained by the efficient market hypothesis (EMH). For example, the return of real financial assets does not obey the normal distribution, but shows a biased "fat-tailed" pattern, and the price fluctuations in the financial market have a long time memory, etc. [1-3]. In addition, many important nonlinear features such as chaos and fractals are found in financial markets. All the signs indicate that the price

✉ Corresponding author.
E-mail address: zsrschscd77@163.com (X. Chen).
Received 15 January 2024; Revised 10 February 2024; Accepted 30 November 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-293

fluctuation of financial assets is not like the EMH assumption, completely obey the random Brownian motion, the complex fluctuation behavior of the financial market may be hidden behind some regular dynamics [4-6]. In recent years, many researches in econophysics have provided many strong supports to the abnormal volatility phenomenon and its predictability in financial markets [7-8].

Measurement of financial market volatility is of great significance both for risk measurement and management, and for asset and option pricing, and it has been a major topic of concern in finance, and academic research on financial volatility has made great progress and has become one of the relatively independent, quite distinctive, and most active research areas in financial econometrics [9-12]. Finance encompasses the circulation of money, the issuance of commodities, bonds, etc. Risk is a collective term for the uncertainty, secrecy, etc. that exists in the process of exchange, storage, recycling, utilization, collateralization, etc. of money and commodities. Financial risk is the likelihood of loss due to uncertainty for all transacting entities involved in financial activities such as companies and financial institutions [13-15]. Financial vulnerability is a financial state in which there is an imbalance in the structure of the financial system that leads to the accumulation of risk and the loss of some or all of the functions of the financial system. Financial vulnerability originated from Minsky's "financial instability hypothesis", which believes that the financial system has inherent vulnerability, and Bernanke's "financial gas pedal theory" also believes that there are cyclical changes in the financial system [16-18]. Therefore, the emergence of financial vulnerability is usually accompanied by the occurrence of financial crises, when the financial operating system is stimulated by a series of factors such as financial innovation, financial liberalization, financial globalization and so on, financial vulnerability will be presented, which is specifically reflected in the outbreak of financial crises. With the rapid development of the modern financial industry, financial crises constantly impact the entire financial system, and the financial system vulnerability has been increasingly aggravated [19-21]. To prevent systemic financial risks, the central bank should strengthen the supervision of the financial industry, improve the financial risk indicator system, optimize the financial organization and product structure, decompose the financial vulnerability, increase the supervision of network security, comprehensively carry out the construction of social credit, improve the risk control system of financial institutions, and reduce the systemic financial risks [22-24].

Literature [25] illustrated the machine learning theory in financial market prediction with good ability to capture market nonlinear dynamics. Literature [26] constructed a deep learning prediction model for stock market based on deep learning algorithm, which has better robustness and prediction accuracy than the benchmark model. Literature [27] designed an auxiliary prediction decision framework for financial risk in carbon trading market based on deep neural network algorithm, and confirmed the feasibility of the proposed model based on simulation experiments. Literature [28] conceptualized a time series forecasting model based on the fusion of long and short-term memory and wavelet analysis, which has good accuracy in predicting the dynamic trend of the financial market compared with machine learning models such as multilayer perceptron machine, support vector machine, and k-nearest neighbor. Literature [29] systematically reviews the research of machine learning in bank risk management, which is mainly applied to bank credit risk, market risk, operational risk and liquidity risk management, pointing out that there is still a large potential for machine learning to be explored in the field of risk management. Literature [30] categorizes financial risk management and explores the path of machine learning algorithm adaptation, filling the gap of machine learning algorithms in financial risk management. The above studies mainly explore financial risk prediction and management from the perspective of machine learning and other algorithmic technologies.

There are also scholars who analyze the response to financial risk from the perspective of futures hedging as well as insurance. Literature [31] reviews the correlation between the investment returns of U.S. stocks and gold and oil futures markets, and finds that the U.S. stocks are positively correlated with crude oil futures and negatively correlated with gold futures, proposing that it can be used for gold qi to cross-hedge in order to cope with financial risk. Literature [32] quantitatively analyzes the systematic risk of insurance financial market risk and proposes an asymptotic formula for risk and stochastic environment affecting insurer solvency.

In this paper, the exchange rate between RMB and USD is selected as the research dataset of financial market volatility, and descriptive statistics are analyzed to combine autoregressive model (AR) and moving average model (MA) with autoregressive moving average model (ARMA). After that, based on the characteristics of the time series components, the time series is decomposed into high volatility components and low volatility components of the MA filtering method, and based on the ARMA model, the high volatility components are predicted with a deep learning model, and the ARMA-LSTM combination model based on the MA filtering method is proposed to predict the volatility time series. Finally, an empirical study of ARMA-LSTM is conducted to examine the accuracy and timeliness of the model.

2. Data description and pre-processing

2.1. Data sources and description

The univariate forecasting model constructed in this paper is to forecast the median exchange rate of the US dollar against the Chinese yuan, and the input variable is the median exchange rate of the US dollar against the Chinese yuan, with data from the State Administration of Foreign Exchange and a date range of January 1, 2010, to September 30, 2023, respectively.

When selecting the multivariate predictor variables, considering that the USD-RMB exchange rate prediction is usually made by inputting the exchange rate closing price, opening price, high price and low price to predict the closing price, or inputting the exchange rate data and the economic and political factors that cause the exchange rate changes to predict the exchange rate closing price, this paper selects the exchange rate data and the stock trading data as the input variables to predict the closing price of the USD-RMB exchange rate. Among them, the stock trading data selected NASDAQ index, Dow Jones Industrial Index, SSE index and Hang Seng index, which NASDAQ index and Dow Jones Industrial Index through the influence of the dollar in the international exchange rate market fluctuations caused by the dollar exchange rate changes against the yuan, SSE index and Hang Seng index through the influence of the fluctuations of the renminbi in the international exchange rate market caused by the change of the dollar exchange rate against the renminbi. The exchange rate and stock transaction data required for the multivariate forecasts are obtained from the Wind database, with a date range from January 1, 2010 to September 30, 2023.

The data field descriptions are shown in Table 1. Specific field descriptions of all variables: Date denotes the date; the daily mid-rate of the USD-RMB exchange rate (Mid), which is the most important reference indicator of the spot interbank foreign exchange market and the banks' quoted exchange rates; the daily closing price of the USD-RMB exchange rate (Close), which is the exchange rate of the foreign exchange banks at the end of the foreign exchange transactions of a trading day; the daily opening price of the USD-RMB exchange rate (Open), which is the exchange rate of the foreign exchange banks at the end of each trading day; the daily opening price of the USD-RMB exchange rate (Open), which is the exchange rate of the foreign exchange banks in each trading day. , is the foreign exchange bank in each business day, the initial reported exchange rate; dollar to yuan exchange rate daily maximum price (High), is the foreign exchange bank in a trading day the highest exchange rate; dollar to yuan exchange rate daily minimum price (Low), is the foreign exchange bank in a trading day the lowest exchange rate. Nasdaq index closing price (Ixic), is reflected in the Nasdaq stock market sentiment changes in the average index of stock prices; Dow Jones Industrial Average closing price (Djia), is a measure of the U.S. stock market, the development of the industrial composition of the index; SSE closing price (Sh), is reflected in the Shanghai Stock Exchange, listed stocks price changes in the index; Hang Seng Index closing price (His), is to The Hang Seng Index (His) is a weighted average stock price index calculated by taking all Chinese H-share companies listed on the Stock Exchange as constituent stocks.

Table 1. Data field specification

Field	Field specification	Example
Date	Time	2023/3/15
Mid	Intermediate price	682.5344
Close	Closing price	6.9652
Open	Opening price	6.6663
High	Maximum price	6.767
Low	Lowest price	6.7156
Ixic	Nasdaq	2308.6144
Djia	Dow Jones industrial index	10583.9811
Sh	Shanghai index	3243.4564
His	Hang seng index	21823.4095

2.2. Data pre-processing

Because of the problem of missing data and date after the data are downloaded from the database, the data are summarized with the same date and cleaned. The sample covers the period from January 1, 2010 to September 30, 2022. Table 2 shows the original data of the median exchange rate of USD-RMB, the value indicates the median exchange rate of USD-RMB per 100 USD, and the missing values have been automatically deleted from the downloaded data, and a total of 3,097 pieces of data are left.

Table 2. The original data of the middle price of the dollar against the RMB

Date	Mid
2023/3/4	682.8071
2023/3/5	682.8302
2023/3/6	682.8898
2023/3/7	682.8474
2023/3/8	682.8707
2023/3/11	682.7016

After downloading and integrating the multivariate data, it is found that the exchange rate data has missing data on weekends and holidays, so it needs to be deleted. After deleting the corresponding date data, there is no missing value in the exchange rate data, so it is only necessary to fill in the data of the stock exchange data, and the way to fill in is the average of the values of the two rows above and below the missing value, and the total number of processed data is 3,542, including 8 variables except time, and the processed data of the exchange rate and the stock index part are shown in Table 3.

Table 3. The exchange rate and the stock index are partially processed

Date	Close	Open	High	Low	Ixic	Djia	Sh	His
2023/3/4	7.0948	6.6773	6.5881	6.7906	2311.8573	10564.886	3251.3089	21847.6963
2023/3/5	7.0012	6.7115	6.9617	6.9012	2302.8534	10560.7119	3272.5389	21816.4441
2023/3/6	6.9442	6.3887	6.8557	6.7165	2307.6459	10591.828	3229.2624	21753.3544
2023/3/7	6.9944	6.8309	6.6913	6.5746	2314.9411	10580.0337	3259.7654	21734.5663
2023/3/8	6.8607	6.4408	6.6788	6.5495	2308.8178	10549.7666	3237.5285	21794.6889
2023/3/11	7.0486	6.8637	6.7334	6.6807	2313.7775	10610.1484	3240.0317	21853.79
2023/3/12	6.8992	6.5211	6.689	6.567	2301.5558	10520.4825	3221.0774	21802.0244
2023/3/13	6.8045	6.6576	7.0466	6.4863	2306.3432	10560.8906	3234.3822	21808.4301
2023/3/14	7.0157	6.7683	6.8603	6.6301	2304.3987	10585.0627	3253.1295	21818.672

2.3. Descriptive statistical analysis

Figure 1 shows a chart of the daily median exchange rate of the dollar against the renminbi, and Figure 2 shows a chart of the data of the exchange rate of the dollar against the renminbi. As a result,it is found that the closing price,opening price,maximum price and minimum price of USD-RMB exchange rate chart,the data trend is the same. As can be seen from the two charts, the USD-RMB exchange rate data showed a downward trend from the beginning of 2010 to the beginning of 2014, indicating that the dollar continued to depreciate during this time period, which was due to the fact that the United States had just experienced a financial crisis and was in the context of a real estate rupture, while China had implemented the policy of “exchange rate reform” in 2010, which had led to an upturn in the economy. From 2015 to 2017, the exchange rate showed an upward trend, this period of time, the United States economy continued to recover, which contributed to the appreciation of the dollar. 2017, the dollar against the RMB exchange rate with international political, economic and other factors showing short-term up and down fluctuations.

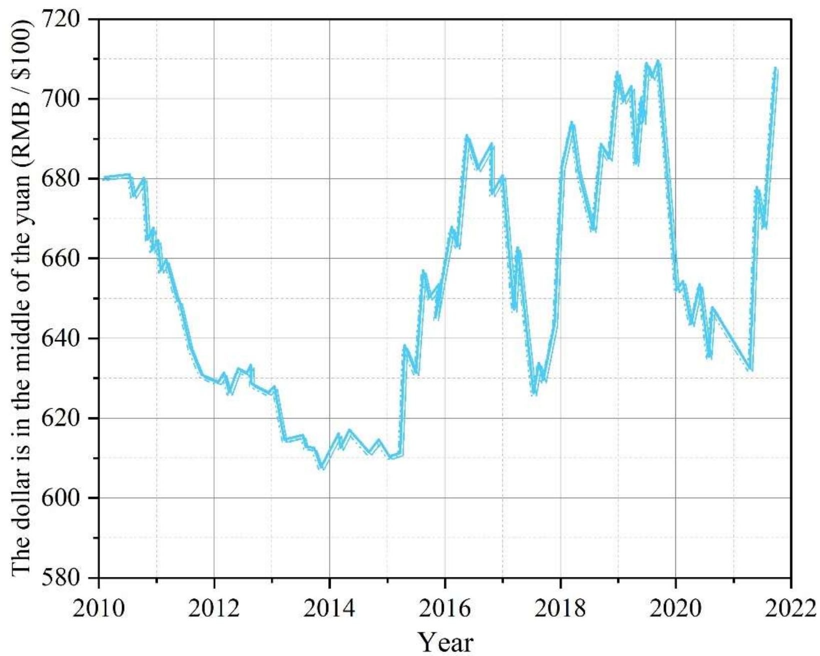


Fig. 1. The dollar is on the middle of the daily exchange rate

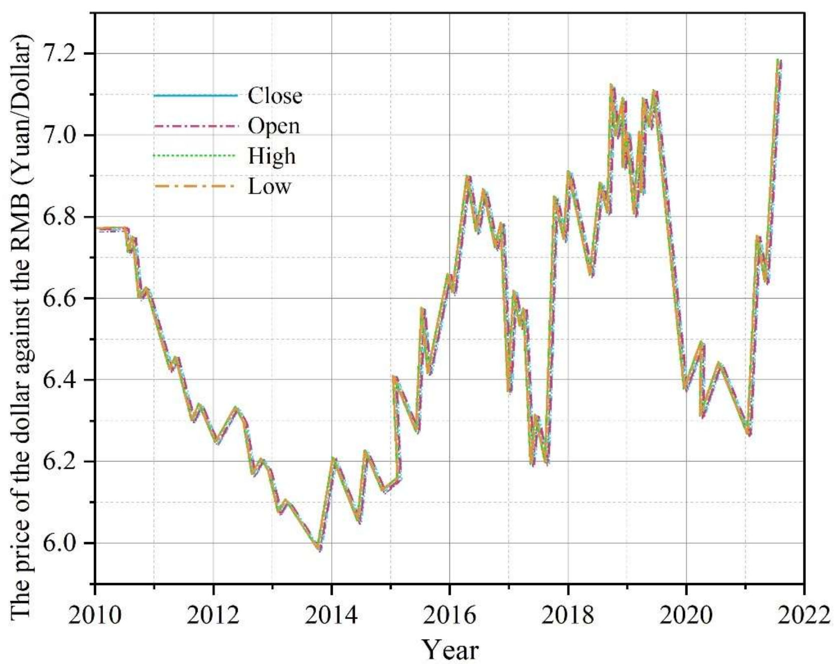


Fig. 2. The dollar is a sign of the renminbi's exchange rate data

Figure 3 shows the trend of the stock exchange sample data over time. It can be seen that all four stock indices show a strong or weak upward trend in the early period and a downward trend after 2021 due to the international economic situation. Table 4 shows the descriptive statistical indicators of stock trading data. The table gives the mean, median, maximum, minimum and standard deviation of the closing prices of the four stock indices, which can show the magnitude of the stock data and the degree of dispersion of the stock data in the time range. Meanwhile, the p-value of J-B test is 0, which indicates that the distribution of stock data rejects the original hypothesis of normal distribution at 1% significant level, which is in line with the distribution characteristics of financial time series data.

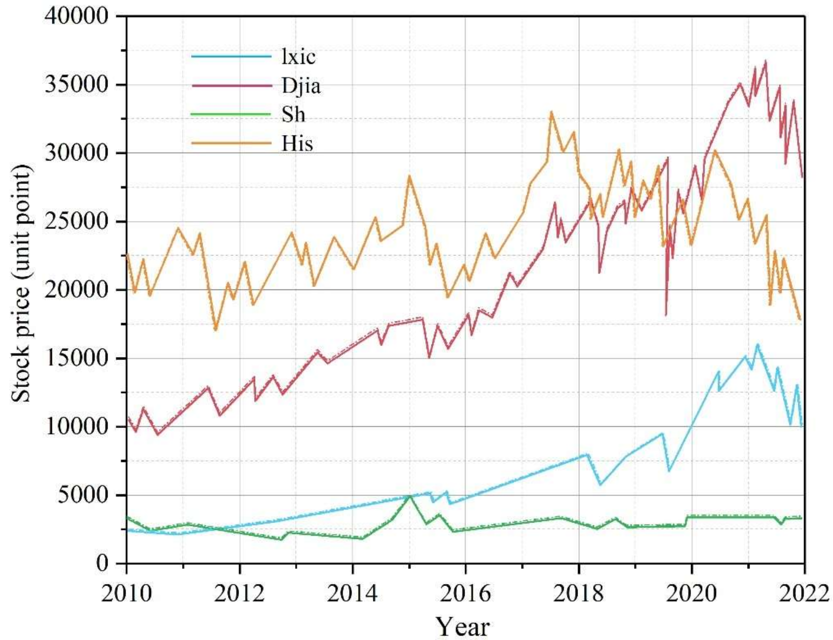


Fig. 3. The change of the stock trading sample number according to the change

Table 4. Descriptive statistics on stock trading data

Statistic	Nasdaq	Dow Jones industrial index	Shanghai index	Hang seng index
Mean value	6405.3623	20554.4582	2896.9839	23999.1941
Median	5098.4609	18056.888	2985.5973	23470.3735
Maximum value	16058.1304	36786.032	5172.6216	33139.953
Minimum value	2092.9756	9705.167	1925.2942	16273.4278
Standard deviation	3692.6338	7427.753	506.3587	3178.0628
Test value	523.3162	239.6371	119.04553	150.5385
Test p	0.0000	0.0000	0.0000	0.0000

3. Time series forecasting models

Time series forecasting methods [33-34] include descriptive analysis and statistical analysis. Descriptive analysis is the earliest time series forecasting method, which plots the time series data according to the time by tracing the points and connecting the lines, and forecasts according to the trend of the graph. Descriptive analysis is a method explored by human beings in actual production, through the discovery of its laws to make plans and arrangements in advance, so as to improve production efficiency. The descriptive analysis method is simple and efficient, and can visualize the fluctuation trend of time series. However, the pure descriptive analysis method has great limitations in the case of strong randomness of data. With the continuous broadening of the field of human understanding, statistical analysis methods have gradually developed and matured. Starting from the last century, Markov chain model was proposed, which predicts the future state based on the current state and trend.AR, MA, and ARMA models were successively proposed to solve the problem of forecasting smooth time series. The modeling principles of several time series forecasting models are described below.

3.1. AR model

Autoregressive models utilize the prior period data as well as a white noise disturbance term to create regression equations that lead to forecasts. A time series is said to be a p st order autoregressive model if it satisfies the following structure:

$$x_t = \varphi_0 + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + +\varphi_p x_{t-p} + \varepsilon_t \tag{1}$$

where φ_0 is the constant term, $\varphi_1, \cdots, \varphi_p$ is the regression coefficient, and ε_t is the white noise. In order to ensure that the model is valid, the above equation needs to satisfy three constraints: first, because the establishment of the autoregressive model of order p, so to ensure that $\varphi_p \neq 0$; second, to ensure that the random perturbation term ε_t to meet all the requirements of the white noise, that is, it needs to satisfy the $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \sigma_e^2$, $E(\varepsilon_t \varepsilon_s) = 0$, $s \neq t$; and, finally, in the sequence needs to ensure that the disturbing term is independent of the past value, that is, at $s < t$ time $E(x_s s_t) = 0$.

Equation (1) is a general p st order autoregressive process, i.e., an uncentered $AR(p)$ model. When $\varphi_0 = 0$ or by subtracting φ_0 from both the left and right sides of the equation, Eq. (1) becomes the centered $AR(p)$ model. When $Bx_t = x_{t-1}$, the operator B is called a one-step delay operator; similarly when $B^k x_t = x_{t-k+1}$, the operator B^k is called a k-step delay operator. At this point the centered autoregressive model $AR(p)$ can be expressed in the form of operator as follows:

$$\varphi(B) x_t = \varepsilon_t \tag{2}$$

Of these, $\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p$.

3.2. MA model

The moving average model $MA(q)$ of order q is an important model for describing time series data, which uses a random perturbation term of the past q order to represent the present value x_t of the time series with the following model structure:

$$x_t = \theta_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \tag{3}$$

where θ_0 is a constant and $\theta_1, \theta_2, \cdots, \theta_q$ is the coefficient of moving average. In order to ensure that the $MA(q)$ model holds, the above equation needs to satisfy two constraints: firstly, because the model is built as a qth order MA model, it is necessary to ensure that $\theta_q \neq 0$; secondly, it is necessary to ensure that the random perturbation term ε_t satisfies all the requirements of white noise, i.e., it is necessary to satisfy $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \sigma_\varepsilon^2$, $E(\varepsilon_t, \varepsilon_s) = 0$, and $s \neq t$. The centered sliding average model [35] $MA(q)$ is expressed in the form of an operator that can be expressed as follows:

$$x_t = \theta(B) \varepsilon_t \tag{4}$$

Of these, $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$.

3.3. ARMA modeling

The autoregressive moving average model $ARMA(p, q)$, which defines the present value of the time series based on historical observations and random perturbation terms in the present as well as in the past x_t, is structured as follows:

$$x_t = \mu + \varphi_1 x_{t-1} + \cdots + \varphi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_t - \cdots - \theta_q \varepsilon_{t-q} \tag{5}$$

The constraints of this model are the same as those of the first two models. The centered

$ARMA(p,q)$ model in the form of an operator can be expressed as:

$$\varphi(B)x_t = \theta(B)\varepsilon_t \tag{6}$$

Among them:

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p \tag{7}$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q \tag{8}$$

It can be seen from equation (5) that model $ARMA(p,q)$ is a combination of $AR(p)$ and $MA(q)$, and model $ARMA(p,q)$ will degenerate to $MA(q)$ when p in model $ARMA(p,q)$ is 0, and model $ARMA(p,q)$ will degenerate to $AR(p)$ when q in model $ARMA(p,q)$ is 0. Therefore, the restriction of model $ARMA$ is a combination of models AR and MA, which are all used for forecasting analysis of smooth time series.

4. Combined ARMA-LSTM model based on the MA filtering method

4.1. LSTM deep neural network

Long Short-Term Memory (LSTM) network is a more accurate network structure combined with gradient learning algorithm. It is not essentially different from the general RNN network structure, but only uses a different function to calculate the state of the hidden layer, i.e., the neurons in the hidden layer of the RNN are all replaced by a special block, in order to make up for the shortcomings of the native RNN such as gradient disappearance and short-term memory.

The LSTM network structure is shown in Fig. 4. In the figure, f_i is a part of the information passed from the previous time point to the current time point, $\tilde{I}_t$ measures the flow of information passed from the previous step to the current step, and c_i is a measure of the importance of the current information. In the actual prediction of volatility, at each point in time i, the input to the implicit layer consists of three parts: the volatility at the previous five points in time, which is represented by a vector H_i; the predicted value of volatility at the previous step $\hat{h}_t$; and the flow of information at the previous step I_{i-1}. The function of the control gate is either a sig function or a $\tanh$ function, with × denoting scalar multiplication and + denoting addition. The following equation 1 represents the update of the linear memory of information; I_i and the predicted value of volatility at the next time point $\hat{h}_{i+1}$ are carried forward to the next iteration step.

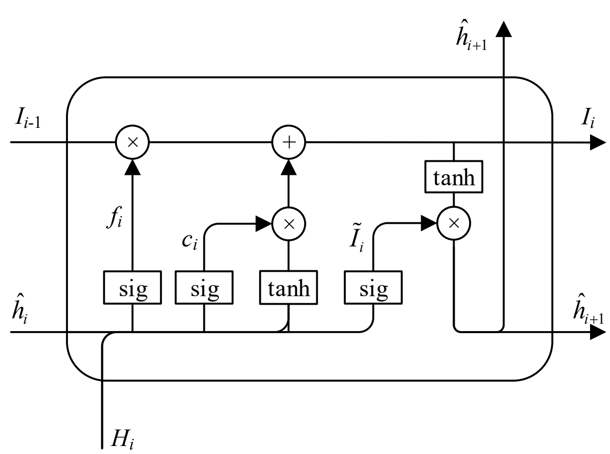


Fig. 4. LSTM network structure diagram

A step-by-step approach to understanding LSTM:

Step 1: Decide what information we will discard from the cell state. This decision is made through a layer called the "forget gate". This gate reads the volatility prediction from the previous moment and the inputs from that moment, and outputs a value between 0 and 1 for each number in cell state c_{i-1}. 1 means "keep completely" and 0 means "discard completely". Then:

$$f_i = sigmoid\left[\left(\hat{h}_i, H_i\right) \cdot W_f + b_f\right] \tag{9}$$

Step 2: Determine what new information is stored in the cell state, which consists of two parts: first, a sigmoid layer, called the "input gate layer", decides what values we are going to update; then, a tanh layer creates a new vector of candidate values, which are added to the state. Then:

$$c_i = sigmoid\left[\left(\hat{h}_v, H_i\right) \cdot W_c + b_c\right] \tag{10}$$

$$\tilde{I}_t = \tanh\left[\left(\hat{h}_v, H_i\right) \cdot W_{\tilde{I}} + b_{\tilde{I}}\right] \tag{11}$$

Step 3: Determine the new information for the update. It is now time to update the old cell state, c_{i-1} to c_i. The previous steps have determined what will be done, now it is time to actually get it done. Multiply the old state I_{i-1} with f_i and discard the information we determined needed to be discarded. Then add $c_i \cdot \tilde{I}_t$. Which is the new candidate value. After which we decide how much to update each state for the change. Then:

$$I_i = I_{i-1} \cdot f_i + c_i \cdot \tilde{I}_t \tag{12}$$

Step 4: Determine what value to output. This output will be based on the cell state, but will be a filtered version. First a sigmoid layer is run to determine what part of the cell state will be output. Next, run the cell state through tanh (to get a value between -1 and 1) and multiply it by the output of the sigmoid gate, which will ultimately output just the portion we determined to be output. Then:

$$\hat{h}_{i+1} = \alpha + \beta \cdot o_i \tanh\left[I_i\right] \tag{13}$$

$$o_i = sigmoid\left[\left(\hat{h}_l, H_i\right) \cdot W_0 + b_0\right] \tag{14}$$

4.2. Combined ARMA-LSTM model based on MA filtering method

In practice, time series are usually characterized by composite with high volatility components, which may not be well portrayed by a single ARMA model [36] or RNN model. After obtaining the stock trading data, the volatility time series is first calculated and denoted as $\{y_t\}$, and then $\{y_t\}$ is decomposed into a low volatility series $\{l_t\}$ and a high volatility series $\{h_t\}$ based on the MA filtering method described above. Since the ARMA model is a typical linear model, which can well characterize linear sequences, i.e., low volatility sequences, we use the ARMA model for the prediction of $\{l_t\}$; while the LSTM model happens to be able to better characterize nonlinear sequences, so we use the LSTM deep neural network model to do the prediction of $\{h_t\}$.

The steps of ARMA-RNN combined model building are as follows:

1) Decompose the volatility time series $\{y_t\}$ into two parts using MA filtering:

$$y_t = l_t + h_t \tag{15}$$

where l_t denotes the low volatility component and h_t denotes the high volatility component.

2) Test the stability of the low volatility series $\{l_t\}$ and build an ARMA model for forecasting:

$$l_t = f\left(l_{t-1}, l_{t-2}, \cdots, l_{t-p}, \varepsilon_t, \varepsilon_{t-1}, \cdots, \varepsilon_{t-q}\right) \tag{16}$$

where $\varepsilon_t, \varepsilon_{t-1}, \cdots, \varepsilon_{t-q}$ is a sequence of random errors and f is a linear ARMA equation, viz:

$$\begin{cases} x_t = \phi_1 l_{t-1} + \cdots + \phi_p l_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \\ \phi_p \neq 0, \theta_q \neq 0 \\ E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_s^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(\varepsilon_t \varepsilon_s) = 0, \forall s < t \end{cases} \tag{17}$$

3) High volatility sequence $\{h_t\}$ with 3-step forward forecasting using information from $h_{t-1}, h_{t-2}, \cdots, h_{t-N}$:

$$h_t = g\left(h_{t-1}, h_{t-2}, \cdots, h_{t-N}\right) + \varepsilon_t \tag{18}$$

where $g(\cdot)$ is an LSTM neural network [37].

4) Calculate the final predicted value of the time series $\{y_t\}$ predicted in the first 3 steps:

$$\hat{y}_t = \hat{l}_t + \hat{h}_t \tag{19}$$

Where $\hat{l}_t = \hat{f}\left(l_{t-1}, l_{t-2}, \cdots, l_{t-p}, \varepsilon_t, \varepsilon_{t-1}, \cdots, \varepsilon_{t-q}\right)$, $\hat{h}_t = \hat{g}\left(h_{t-1}, h_{t-2}, \cdots, h_{t-N}\right)$.

5. Empirical studies

The same in-sample and out-of-sample intervals are set up as in the previous model, with data from 2011-2020 as the training set and data from 2021-2022 as the test set.

5.1. ARMA-LSTM modeling empirical evidence

In order to conduct a comparative study with the MIDAS model above, the models were modeled separately using ISI, using INSI, and using both ISI and INSI data. The three models are denoted as ARMA-LSTM_t, ARMA-LSTM_n, and ARMA-LSTM_tn, respectively. n_test denotes the number of months used in the test set, epochs and batch_size represent the number of iterations and the number of samples taken in a training session, respectively, nodes represents the number of hidden neurons, month_t and month_p represent the number of months in the current quarter and the previous quarter used for training, respectively, and gdp_p represents the number of lags of GDP_IG data. The ARMA-LSTM_t model is an example, and the optimal parameter settings are shown in Table 5. The results show that the ARMA-LSTM_t model will achieve relatively good results in predicting the GDP_IG of the current year using the data of the 12 months of the current year and the last month of the previous year.

The training set and test set errors of the model built using ISI data and the model built using INSI data are relatively small, which is consistent with the results of the MIDAS model. The reason may be that ISI data is synthesized from various macroeconomic indicators, which are directly related to economic fluctuations. Whereas INSI data is based on the linkage established from the search data of investment-related indicators, which has a certain degree of uncertainty. It can be seen that both the training set and prediction set errors are minimized using the combined ARMA-LSTM model based on the MA filtering method.

Table 5. Optimal parameter setting

Optimal parameter	ARMA-LSTM_n	ARMA-LSTM_t	ARMA-LSTM_tn
n_test	9	9	9
epochs	1000	2000	3000
batch_size	30	28	26
nodes	6	6	6
month_t	4	4	4
month_p	3	3	3
gdp_p	2	2	2
Training set RMSE	2.7519	1.0997	0.8443
Test set RMSE	5.5865	1.2815	0.8317

In order to verify whether the model has the function of predicting the GDP_IG data in advance, the model parameters month_t are set to 3, 4, and 5, respectively.The training and testing errors of the model are obtained by adjusting the other parameters in Table 5, and a comparison of the training and prediction errors for the ARMA-LSTM_tn model is shown in Table 6.20+1 means that the INSI data for the first 30 days of each year and the 1-month ISI data to predict the GDP_IG data for that year. The comparison reveals that using only the first 30 days and 1 month of the year's data can predict that GDP_IG data about 50 days earlier than INSI. In contrast, the prediction using the first 50 days and 1 month of the year's data is comparable in accuracy to the AR-MIDAS model, but it can predict the GDP_IG data 30 days in advance. Therefore, this forecasting advance can be utilized to provide a decision-making basis for government departments to formulate economic policies.

Table 6. ARMA-LSTM_tn model training and prediction error comparison

Mixing frequency	Training set error	Test set error
20+1	0.7754	3.0677
40+1	0.7326	1.1439
50+1	0.6662	0.8615

5.2. Evaluation of the prediction effect of the combined ARMA-LSTM model

5.2.1. Comparison of Predictive Performance of Models. Using the estimation results of each model in the empirical analysis section, the in-sample fitting error and out-of-sample prediction error of each model are calculated, and the in-sample and out-of-sample prediction performance of each model is shown in Table 7. There is a positive relationship between the INSI and ISI data synthesized from investment and the GDP_IG series, indicating that the higher investment rate is an important factor causing economic fluctuations since 2010. From the estimation results of all models, the MIDAS model and the model of this paper (ARMA-LSTM) have improved the prediction effect than the AR model, which indicates that the synthetic INSI and ISI of investment web search in this paper have optimized the prediction of China's economic fluctuations. In addition, this paper argues that web search data, as an immediate indicator reflecting changes in user concerns, provides an effective complement to traditional monitoring methods and can be used to predict economic fluctuations. In summary, the effectiveness of the ARMA-LSTM model in the prediction of China's macroeconomic fluctuations is confirmed from the prediction results of the ARMA-LSTM model and the comparison model.

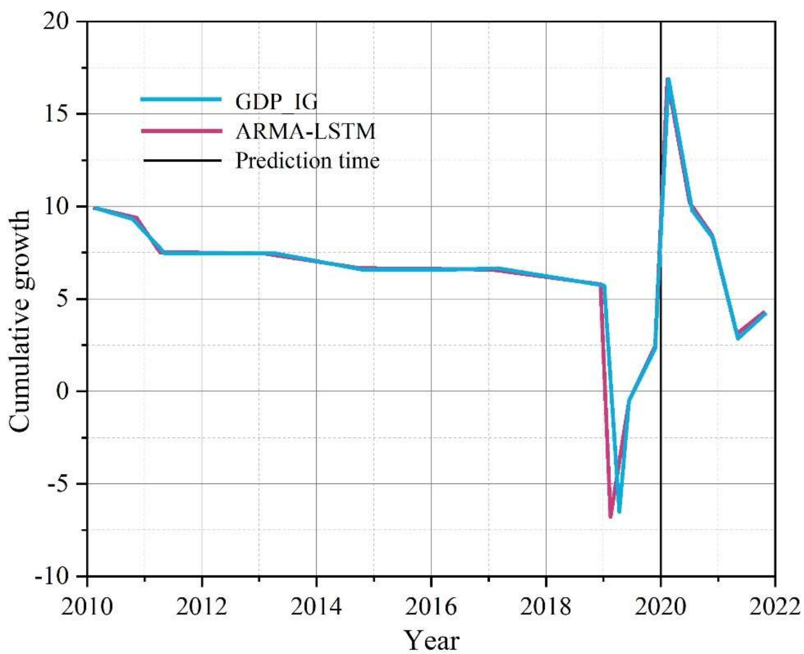
Table 7. The predictive ability of each model is compared

Model	Sample mismatch error	Sample external prediction error
AR(1)	2.0239	5.8128
AR(1)-MIDAS(4,2)	0.9219	1.1372
AR(1)-MIDAS(3,0)	2.442	5.3302

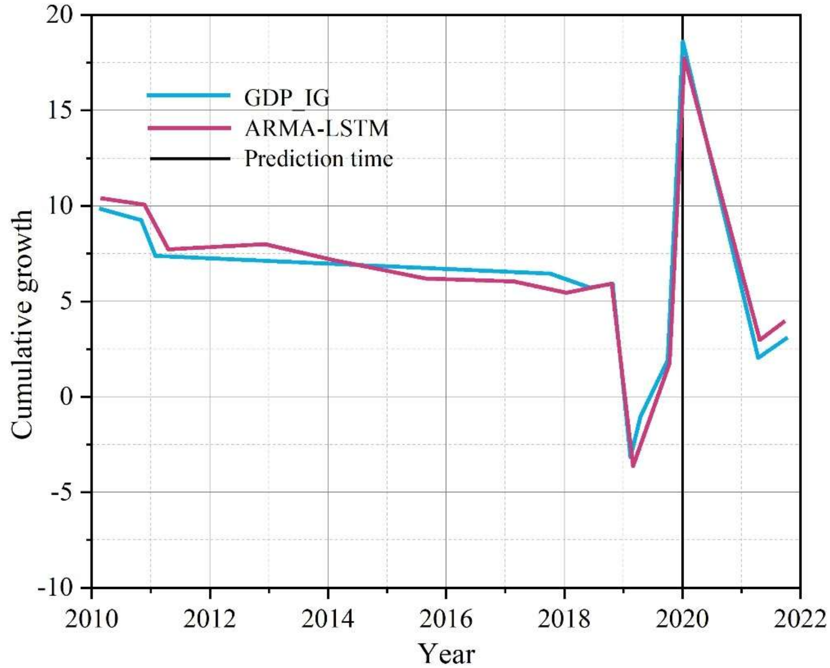
AR(1)-MIDAS_tn	0.6438	1.1705
ARMA-LSTM_t	1.1816	1.4121
ARMA-LSTM_n	2.919	5.6292
ARMA-LSTM_tn	0.6554	0.8768

5.2.2. Comparison of model predictions. The in-sample fitting and out-of-sample prediction results of the ARMA-LSTM model are shown in Figure 5, with (a) to (c) representing ARMA-LSTM, AR and ARMA-LSTM_tn.

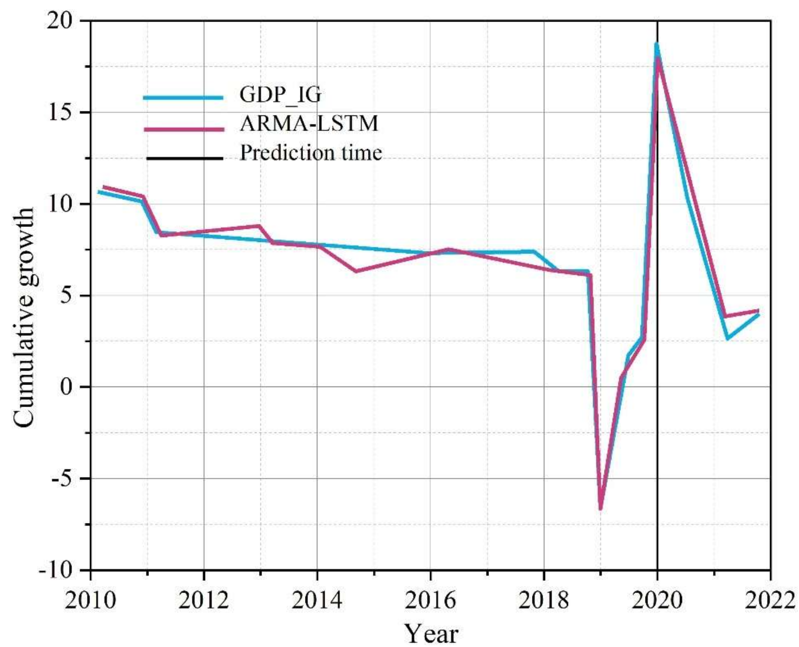
The estimation results of the ARMA-LSTM model for GDP_IG are similar to shifting the original value backward by 1 period, which cannot be predicted in time when there are large fluctuations in GDP_IG. For example, the prediction results in the first quarter of 2019, the first quarter of 2018, and the second quarter of 2020 show a large deviation. Compared with the AR model, after adding investment statistics and web search index data, the AR(1)-MIDAS-tn model accurately predicts the fluctuations in the first quarter of 2020, the first quarter of 2021, and the second quarter of 2022, while the out-of-sample short-term prediction is better, but the fluctuation prediction of the second quarter of 2022 has a large error.The short-term prediction of ARMA-LSTM_tn model is more accurate than the MI-DAS model, and the medium-term forecasting error is also relatively small, which indicates that the ARMA-LSTM model proposed in this paper improves the out-of-sample short-term and medium-term forecasting accuracy.



(a) ARMA-LSTM



(b) AR



(c) ARMA-LSTM_tn

Fig. 5. The model sample is the result of the sample and the sample

6. Conclusion

In order to explore the prediction of the volatility of the financial market and to cope with the risk it brings, this paper introduces the LSTM deep neural network on the basis of the autoregressive moving average model (ARMA) to construct the ARMA-LSTM combination model. Finally, the application effect of the model is analyzed through empirical evidence. The main conclusions are as follows:

1) In this paper, the exchange rate data between RMB and USD between 2010.1.1-2023.9.30 is selected as the dataset, and descriptive statistical analysis is used to draw the time-series graph of each variable data and characterize the data by data mean, maximum and minimum values, and standard deviation.

Through the J-B test, the distribution of stock data is at 1% significant level, which is in line with the distribution characteristics of financial time series data.
2) There is a positive relationship between investment-related indicators and GDP_IG. Adding investment web search data improves the estimation accuracy of the model. The out-of-sample prediction error decreases after adding the investment web search index, indicating that the web search data has a positive impact on economic forecasting.The ARMA-LSTM model has a smaller prediction error and improves the prediction accuracy in the short and medium term.

References

[1] Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European journal of operational research, 270(2), 654-669.

[2] Khan, W., Ghazanfar, M. A., Azam, M. A., Karami, A., Alyoubi, K. H., & Alfakeeh, A. S. (2022). Stock market prediction using machine learning classifiers and social media, news. Journal of Ambient Intelligence and Humanized Computing, 1-24.

[3] jun Gu, W., hao Zhong, Y., zun Li, S., song Wei, C., ting Dong, L., yue Wang, Z., & Yan, C. (2024, August). Predicting stock prices with finbert-lstm: Integrating news sentiment analysis. In Proceedings of the 2024 8th International Conference on Cloud and Big Data Computing (pp. 67-72).

[4] Cao, J., Li, Z., & Li, J. (2019). Financial time series forecasting model based on CEEMDAN and LSTM. Physica A: Statistical mechanics and its applications, 519, 127-139.

[5] Zhang, X., Liang, X., Zhiyuli, A., Zhang, S., Xu, R., & Wu, B. (2019, July). At-lstm: An attention-based lstm model for financial time series prediction. In IOP Conference Series: Materials Science and Engineering (Vol. 569, No. 5, p. 052037). IOP Publishing.

[6] Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. Expert Systems with Applications, 124, 226-251.

[7] Rezaei, H., Faaljou, H., & Mansourfar, G. (2021). Stock price prediction using deep learning and frequency decomposition. Expert Systems with Applications, 169, 114332.

[8] Reboredo, J. C., & Ugolini, A. (2020). Price connectedness between green bond and financial markets. Economic Modelling, 88, 25-38.

[9] Barra, S., Carta, S. M., Corriga, A., Podda, A. S., & Recupero, D. R. (2020). Deep learning and time series-to-image encoding for financial forecasting. IEEE/CAA Journal of Automatica Sinica, 7(3), 683-692.

[10] Shah, D., Isah, H., & Zulkernine, F. (2019). Stock market analysis: A review and taxonomy of prediction techniques. International Journal of Financial Studies, 7(2), 26.

[11] Valle-Cruz, D., Fernandez-Cortez, V., López-Chau, A., & Sandoval-Almazán, R. (2022). Does twitter affect stock market decisions? financial sentiment analysis during pandemics: A comparative study of the h1n1 and the covid-19 periods. Cognitive computation, 14(1), 372-387.

[12] Yu, P., & Yan, X. (2020). Stock price prediction based on deep neural networks. Neural Computing and Applications, 32(6), 1609-1628.

[13] Ng, S. L., Chin, W. C., & Chong, L. L. (2017). Multivariate market risk evaluation between Malaysian Islamic stock index and sectoral indices. Borsa Istanbul Review, 17(1), 49-61.

[14] Oláh, J., Kovács, S., Virglerova, Z., Lakner, Z., Kovacova, M., & Popp, J. (2019). Analysis and comparison of economic and financial risk sources in SMEs of the Visegrad group and Serbia. Sustainability, 11(7), 1853.

[15] Chang, C. P., & Zhang, L. W. (2020). Do natural disasters increase financial risks? An empirical analysis. Bulletin of Monetary Economics and Banking, 23, 61-86.

[16] Chang, C. L., McAleer, M., & Wong, W. K. (2020). Risk and financial management of COVID-19 in business, economics and finance. Journal of Risk and Financial Management, 13(5), 102.

[17] van Riet, A. (2023). European financial law and the state-finance nexus: Sovereign privileges or market discipline for safe public debt?. Finance and Society, 9(2), 39-57.

[18] Yang, T., Li, A., Xu, J., Su, G., & Wang, J. (2024). Deep learning model-driven financial risk prediction and analysis. Applied and Computational Engineering, 77, 196-202.

[19] So, M. K., Chan, L. S., & Chu, A. M. (2021). Financial network connectedness and systemic risk during the COVID-19 pandemic. Asia-Pacific Financial Markets, 28(4), 649-665.

[20] Blajer-Gołębiewska, A., Wach, D., & Kos, M. (2018). Financial risk information avoidance. Economic research-Ekonomska istraživanja, 31(1), 521-536.

[21] Ihyak, M., Segaf, S., & Suprayitno, E. (2023). Risk management in Islamic financial institutions (literature review). Enrichment: Journal of Management, 13(2), 1560-1567.

[22] Luo, C., Liu, L., & Wang, D. (2021). Multiscale financial risk contagion between international stock markets: Evidence from EMD-Copula-CoVaR analysis. The North American Journal of Economics and Finance, 58, 101512.

[23] Campiglio, E., Dafermos, Y., Monnin, P., Ryan-Collins, J., Schotten, G., & Tanaka, M. (2018). Climate change challenges for central banks and financial regulators. Nature climate change, 8(6), 462-468.

[24] Heo, W., Rabbani, A., & Grable, J. E. (2021). An evaluation of the effect of the COVID-19 pandemic on the risk tolerance of financial decision makers. Finance Research Letters, 41, 101842.

[25] Zheng, H., Wu, J., Song, R., Guo, L., & Xu, Z. (2024). Predicting financial enterprise stocks and economic data trends using machine learning time series analysis. Applied and Computational Engineering, 87, 26-32.

[26] Yan, B., & Aasma, M. (2020). A novel deep learning framework: Prediction and analysis of financial time series using CEEMD and LSTM. Expert systems with applications, 159, 113609.

[27] Luo, J., Zhuo, W., & Xu, B. (2024). A Deep Neural Network-Based Assistive Decision Method for Financial Risk Prediction in Carbon Trading Market. Journal of Circuits, Systems & Computers, 33(8).

[28] Yan, H., & Ouyang, H. (2018). Financial time series prediction based on deep learning. Wireless Personal Communications, 102, 683-700.

[29] Leo, M., Sharma, S., & Maddulety, K. (2019). Machine learning in banking risk management: A literature review. Risks, 7(1), 29.

[30] Mashrur, A., Luo, W., Zaidi, N. A., & Robles-Kelly, A. (2020). Machine learning for financial risk management: a survey. Ieee Access, 8, 203203-203223.

[31] Junttila, J., Pesonen, J., & Raatikainen, J. (2018). Commodity market based hedging against stock market risk in times of financial crisis: The case of crude oil and gold. Journal of International Financial Markets, Institutions and Money, 56, 255-280.

[32] Tang, Q., & Yang, Y. (2019). Interplay of insurance and financial risks in a stochastic environment. Scandinavian Actuarial Journal, 2019(5), 432-451.

[33] Lingli Chen,Xin Gao,Jing Liu,Yunkai Zhang,Xinping Diao,Taizhi Wang... & Zhihang Meng. (2025). A multivariate time series anomaly detection method with Multi-Grain Dynamic Receptive Field. Knowledge-Based Systems112768-112768.

[34] XiwenQin,HongyuShi,XiaogangDong & SiqiZhang. (2024). A new method based on generative adversarial networks for multivariate time series prediction. Expert Systems(12),e13700-e13700.

[35] Kleber H.Santos & FranciscoCribari Neto. (2024). A Varying Precision Beta Prime Autoregressive Moving Average Model With Application to Water Flow Data. Environmetrics(8),e2886-e2886.

[36] Zahia Nabi,Mohammed Assam Ouali,Mohamed Ladjal & Hamza Bennacer. (2024). A Novel ANN-ARMA Scheme Enhanced by Metaheuristic Algorithms for Dynamical Systems and Time Series Modeling and Identification. Revue d'Intelligence Artificielle(3),

[37] Zhiganov S. V.,Ivanov Y. S. & Grabar D. M.. (2024). Investigation of Neural Network Algorithms for Human Movement Prediction Based on LSTM and Transformers. Doklady Mathematics(Suppl 2),S484-S493.