Fluctuation and Noise Letters
(2025) 2550008 (21 pages)

DOI: 10.1142/S0219477525500087

Volatility Information in High-Frequency Financial Interval-Valued Time Series: A Direct Modeling Pattern

Xu Hu [*], Jianwen Yu [*], Qin Xu [†] and Zhifu Tao [‡,§]

[*]*School of Big Data and Statistics, Anhui University*
Hefei, Anhui 230601, P. R. China

[†]*School of Computer Sciences, Anhui University*
Hefei, Anhui 230601, P. R. China

[‡]*Research Center for Data Fusion and Development Application*
Anhui University, Hefei, Anhui 230061, P. R. China
[§]*jeff.tao@ahu.edu.cn*

Received 10 April 2024
Revised 13 July 2024
Accepted 17 August 2024
Published

Communicated by Wei-Xing Zhou

The aim of this paper is to develop a forecasting method with global interval input and output for interval-valued financial time series by combining the VAR(p) process, the volatility information and neural network, namely VAR-NN. To reflect the volatility information, four types of interval-valued data volatility information from both the relative and absolute perspectives are constructed. Furthermore, the neural network is combined to produce the parameters. The developed forecasting model is finally applied to the highest and lowest hourly prices of the Shanghai Composite Index prediction. Numerical study shows the feasibility and validity of the developed improved VAR model.

Keywords: Interval-valued time series; high-frequency; interval-valued data volatility information; financial; VAR-NN.

1. Introduction

In the last two decades, theoretical and practical studies on interval-valued time series (IvTS) have attracted growing attention from both academia and practitioners. Different from classical real-valued time series, the observations at each time point are presented with a numerical interval but not a real number. Because of the ability to model uncertain with clear bounds and uniform distribution of observations, IvTS analysis have been widely applied to real-world decision scenarios include

[§]Corresponding author.

agriculture marketing service [1], predictions of financial or economic data include crude oil future prices [2], air quality index [3], consumer price index [4] and short-term load forecasting [5]. Since the financial industry is the industry with the deepest data accumulation, resulting in wide affection for human life, the interval-valued financial time series (or interval financial time series or financial IvTS Data) is an important branch in IvTS analysis.

To realize accurate forecasting on financial interval-valued time series, types of forecasting technologies from the perspective of statistics and artificial intelligence have been developed [6], in which the evidences of the predictability of the IvTS are found. Maia *et al.* [7] introduced the Autoregressive Integrated Moving Average ($ARIMA$ for short) Model[a] based IvTS forecasting method and its combination with artificial neural network (ANN). Xiong *et al.* [8] proposed a hybrid modeling framework combining interval Holt's exponential smoothing method (HoltI) and multi-output support vector regression (MSVR) for IvTS. Next, Maia and Carvalho [9] considered IvTS forecasting based on multilayer perceptron (MLP) neural networks, Holt's exponential smoothing methods and their combination. Hajek *et al.* [10] and Ouyang *et al.* [11] introduced different types of cognitive maps methods to forecasting IvTS, which allowed continuous model adaptation using the stream interval input data. Chien *et al.* [12] designed a heteroscedastic volatility auto-interval-regressive (HVAIR) model for the analysis of symbolic interval-valued data. Hamiye [13] presented two IvTS approaches to construct multivariate multi-step ahead joint forecast regions via two bootstrap algorithms. Guo *et al.* [14] proposed a novel rule-based granular IvTS model. Buansing *et al.* [15] developed an iterative and efficient information-theoretic estimator for forecasting interval-valued data. Maciel *et al.* [19] introduced an adaptive interval fuzzy modeling method using participatory learning and interval-valued stream data. It can be seen that current IvTS forecasting models can be mainly partitioned into three categories, i.e., the statistical theory-based, the artificial intelligence-based and the hybrid-based forecasting models. By considering the factors include interpretability and predictive ability, these forecasting models provide a wide range of applications under different decision scenarios.

With the development of information technologies, financial data can be recorded at microsecond level. Correspondingly, it is a common sense to face and handle IvTS with high frequency. To analyze the validity of interval forecasts of financial data, Taylor [37] considered a regression-based test, and an improved version of Christoffersen's Markov chain test for independence, and analyzed their properties when the financial time series exhibit periodic volatility. Geng *et al.* [16] proposed the wavelet multi-elements method to overcome the limitation of the wavelet finite element methods in low-frequency domain when capturing dynamic characteristics of

[a] Given that x_t is a stationary time series, $ARIMA$-based analysis on x_t can be finished according to $\nabla^d x_t = \phi_1 \nabla^d x_{t-1} + \phi_2 \nabla^d x_{t-2} + \cdots + \phi_p \nabla^d x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$, where the parameters ϕ_i, ε_j can be estimated by using the least squares method, moment estimation or maximum likelihood estimation, $i = 1, 2, \ldots, p, j = 1, 2, \ldots, q$, and ∇^d represents the d-order difference.

thin plate. Vella and Ng [17] investigated the ability of higher-order fuzzy systems to handle increased uncertainty, mostly induced by the market microstructure noise inherent in a high frequency trading (HFT) scenario. Kong [18] implemented the empirical likelihood approach to construct the confidence interval of the jump activity index of a pure jump model using high frequency data. Current studies mainly focused on interval output from the perspective of probability theory. However, it is worth noting that high-frequency IvTS has merely been reported.

By reviewing current literatures on IvTS, the following research gaps are worth to be further considered:

(1) Although forecasting on financial IvTS has been reported in theory and real-world applications, high-frequency IvTS has been merely studied. As an important branch of IvTS, it is meaningful to consider how to design a forecasting model of IvTS with high accuracy.
(2) Currently, most of the IvTS forecasting model are built based on the decomposition of initial observations, i.e., the lower/upper bound sequences or center/radius sequences. With the development of VAR(p) process-based forecasting method for IvTS analysis [25], the mode for global input and output of intervals was put forward. Up until now, it can be found that only Sun *et al.* [20] have reported the Generalized Autoregressive Conditional Heteroskedasticity[b] (*GARCH* for short) model for interval-valued data. On the other hand, how to model the volatility information in such models is an open problem.

2. Preliminaries

In this subsection, fundamental knowledge related to interval-valued high-frequency financial time series and vector autoregression are listed for further consideration.

2.1. *High-frequency financial interval-valued time series*

To describe practical uniformly varying datum with limited bounds, the following concept of interval-valued number was developed [21], i.e.,

Definition 2.1 ([21]). An interval number, denoted as $\tilde{a}$, can be presented as $\tilde{a} = \{[a_l, a_u] | a_l, a_u \in \mathbb{R}\}$, where a_l and a_u are, respectively, the lower and the upper bound, and $a_l \leq a_u$. $\tilde{a}$ is said to be an interval-valued number.

By Definition 2.1, it can be seen that $\tilde{a}$ is developed to describe the uncertainty with known bounds and distribution. Besides, an interval-valued number $\tilde{a}$ can also be presented as two tuple named as center and radius, i.e., $c = \frac{1}{2}(a_l + a_u)$ and $r = \frac{1}{2}(a_u - a_l)$. Correspondingly, $\tilde{a}$ can be equivalently expressed as $\tilde{a} = (c; r)$.

[b] Given an observed time series $\{x_t\}$, the fundamental GARCH model can be presented according to $\begin{cases} x_t = f(t, x_{t-1}, x_{t-2}, \ldots) + \varepsilon_t \\ \varepsilon_t = \sqrt{h_t} e_t \\ h_t = \omega + \sum_{i=1}^{p} \eta_i h_{t-i} + \sum_{j=1}^{q} \lambda_j \varepsilon_{t-j}^2 \end{cases}$,
where $f(t, x_{t-1}, x_{t-2}, \ldots)$ is the deterministic information fitting model of $\{x_t\}$ and $e_t \overset{i.i.d}{\sim} N(0, \sigma^2)$.

With the notion of interval-valued number, the concept of interval-valued time series was then considered [7], i.e.,

Definition 2.2 ([7]). Assume that T is a non-negative integer, an interval-valued time series (IvTS) with length T, denoted as $\tilde{y}_t, t \in \{1, 2, \dots, T\}$, is a sequence such that $\tilde{y}_t = [y_t^l, y_t^u], t \in \{1, 2, \dots, T\}$, i.e., the observation at each time unit is presented by an interval-valued number.

By Definition 2.2, it can be seen that the observation of $\tilde{y}_t$ at each time point is a numerical interval. Different from common real-valued time series, the observed value given by the form of numerical interval can reflect possible uncertainties for the object of study. A real-valued time series is a sequence of scalar data, while an interval-valued time series is formally composed by two sequences of scalar data, i.e., the lower and upper bound sequences or the center and radius sequences.

Currently, transaction data in the financial market can be comprehensively preserved. For instance, daily lowest, highest, and transaction prices. Since the strengths of providing more frequent and detailed market information, as well as more timely feedback, interval-valued financial data [22] can be used to obtain insights into market behavior and price changes by analyzing its patterns, trends, and price volatilities. On the other hand, the analysis of high-frequency financial data also suffers from practical problems including data quality and market noise. Therefore, the following high-frequency interval-valued financial time series can be defined:

Definition 2.3. High-frequency financial interval-valued time series (HFIvTS), also known as intra-day data, refers to trading data sampled between the opening and closing times. It is mainly a time series of interval-valued numbers arranged in chronological order with sampling frequencies of hours, minutes, and even seconds.

According to Definition 2.3, HFIvTS is an IvTS in essence, but the frequency of data collection is limited to within a day. Compared to common IvTS, HFIvTS owns the characteristics including more frequent and detailed market information and more timely feedback. On the other hand, it also has the limitation of higher requirements for data storage and processing and data quality issues, etc. Especially, how to model the volatility information has been widely studied in real-valued time series analysis.

Although traditional daily interval-valued data indeed reflect intraday price movements, we posit that HFIvTS data (at the hourly level) provide more granular information on market dynamics. This higher-frequency data capture the characteristics of price volatilities during different intraday periods, which may be averaged out or obscured in daily-level data.

The rationale for selecting hourly granularity can be summarized in the following:

(a) Information Density: Hourly-level data effectively preserve crucial market dynamic information while mitigating the noise inherent in ultra-high-frequency data (e.g., minute-level).

(b) Intraday Patterns: This frequency allows for the capture of trading characteristics specific to different periods within the day, such as the opening effect, midday trading lulls, and pre-closing trading surges.
(c) Market Microstructure: It facilitates the study of market microstructure features, including intraday variations in liquidity and trading volume patterns across different time segments.

2.2. *VAR process*

Generally, a p-order vector autoregression process with white noise, denoted as VAR(p), can be written as follows:

$$\mathbf{y}_t = \Phi_1\mathbf{y}_{t-1} + \Phi_2\mathbf{y}_{t-2} + \cdots + \epsilon_t = \sum_{j=1}^{p}\Phi_j\mathbf{y}_{t-j} + \epsilon_t, \tag{1}$$

where $\mathbf{y}_t$ is a vector composed by a series of variables, denoted as $y_{it}, i = 1, 2, \ldots, m$, m is an integer, and $\mathbf{y}_{t-j}, j = 1, 2, \ldots, p$ is the lags.

By using the lag operator, Eq. (1) can be rewritten as follows:

$$\Phi(L)\mathbf{y}_t = \epsilon_t, \tag{2}$$

where $\Phi(L) = I_k - \Phi_1 L - \cdots - \Phi_p L^p$ and ϵ_t is a vector white noise.

It is worth noting that the process $\{\mathbf{y}_t\}$ is required to be covariance-stationary, i.e., all the roots in the equation $|I_k - \Phi_1 z - \cdots - \Phi_p z^p| = 0$ lie outside the unit circle.

2.3. *Confidence intervals*

In the traditional example of calculating the confidence interval, it is customary to take a symmetrical quantile. In the normal distribution $N(\mu, \sigma^2)$, given the confidence level of $1 - \alpha$, if the population variance σ^2 is unknown, it is replaced by the sample variance S^2, thus obtaining the confidence interval of the confidence level of $1 - \alpha$ of the mean value μ as follows:

$$\left(\bar{Y} - Z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}, \bar{Y} + Z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right). \tag{3}$$

When applied to IvTSs, and to ensure that the interval covers the true value to the greatest extent, Eq. (4) can be predicted by the interval center and radius data, and it is summed and combined to obtain Eq. (5):

$$\begin{cases} [L_1, L_2] = (\bar{L} - Z_{\frac{\alpha}{2}}\frac{S_L}{\sqrt{n}}, \bar{L} + Z_{\frac{\alpha}{2}}\frac{S_L}{\sqrt{n}}), \\ [U_1, U_2] = (\bar{u} - Z_{\frac{\alpha}{2}}\frac{S_U}{\sqrt{n}}, \bar{U} + Z_{\frac{\alpha}{2}}\frac{S_U}{\sqrt{n}}), \end{cases} \tag{4}$$

$$\begin{aligned} \{[L_1, L_2] \ominus [U_1, U_2]\} \cup \{[U_1, U_2] \oplus [U_1, U_2]\} &= [L_1 - U_2, L_2 - U_1] \cup [L_1 + U_1, L_2 + U_2] \\ &= [L_1 - U_2, L_2 + U_2]. \end{aligned} \tag{5}$$

Therefore, $[L_1 - U_2, L_2 + U_2]$ is the confidence interval of the interval value. When the radius series data $r(\omega) = 0$, the confidence interval of IvTSs degenerates into the confidence interval of the point-value time series (i.e., $[L_1, L_2] = (\bar{L} - Z_{\frac{\alpha}{2}}\frac{S_L}{\sqrt{n}}, \bar{L} + Z_{\frac{\alpha}{2}}\frac{S_L}{\sqrt{n}}))$.

3. Improved VAR Process for HFIvTS

In this section, an integration of VAR(p) process with interval-valued data volatility information and neural network is proposed. Mechanism analysis and the construction process of the developed high-frequency financial interval-valued time series forecasting model are presented.

3.1. *Mechanism analysis*

In financial time series analysis, volatility information is commonly coupled with the forecasting model. Taking *GARCH* as an example, the usage of past volatility information is utilized to correct the bias of the mean equation, which can be shown in the following equation:

$$\left\{\begin{aligned} &y_t = \mu_t + \varepsilon_t, \\ &\mu_t = \alpha_0 + \sum_{i=1}^{p}\alpha_i y_{t-i} + \sum_{j=1}^{q}\beta_j\sigma_{t-j}, \\ &\sigma_t^2 = \omega + \sum_{i=1}^{p}\gamma_i\varepsilon_{t-i}^2 + \sum_{j=1}^{q}\delta_j\sigma_{t-j}^2. \end{aligned}\right\}. \tag{6}$$

By Eq. (6), volatility information contained in the error sequence after the extraction of the main trend is helpful to realize more accurate prediction on financial time series. Correspondingly, for multivariate financial time series analysis, whether the application of volatility information can be used to improve forecasting accuracy of VAR(p) process is a natural issue. Besides, if the answer is positive, how to model the volatility information should be further considered. Although Pantelidis and Pittis [23] discussed the estimation and forecasting in VAR(1) with near to unit roots and conditional heteroscedasticity, the VAR models fail to model the conditional heteroscedasticity [24]. Thus, it is interesting to study the affection of possible volatility information in multivariate financial time series analysis. Motivated by Eq. (6), the volatility information can be considered to be included in common VAR(p) process by adding the volatility variable(s).

3.2. *VAR-NN*

The proposed model is Vector Autoregressive Neural Network (VAR-NN) derived from the Vector Autoregressive (VAR) model. The VAR-NN architecture is built similar to the AR-NN, with more than one variable (multivariate) in the output layers [27]. In this case, the input layer contains variables from the previous lag

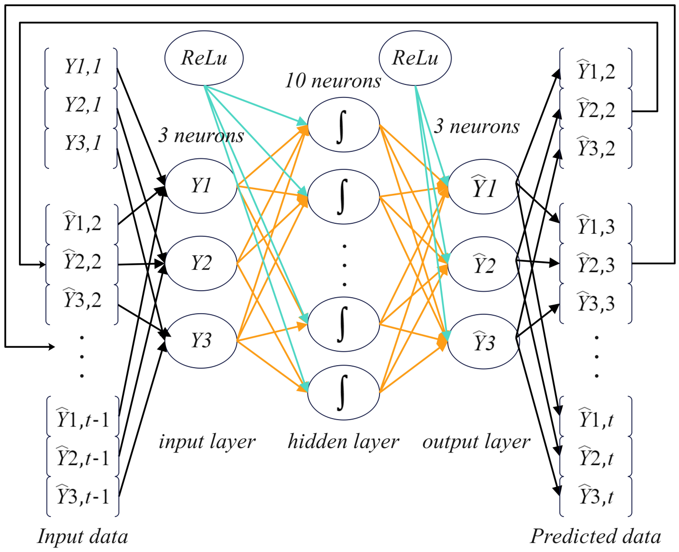

Fig. 1. The structure of VAR-NN.

observations of each output variable. Nonlinear estimates are processed in the hidden layer by the activation function. To handle the reconstructed $\mathbf{y}_t$ with triple cases, an architecture with one hidden layer is considered (see Fig. 1).

As can be seen in Fig. 1, the developed architecture of VAR-NN $(1, 1, 10)$ is composed of one hidden layer, 3 input neurons from 1 lag dependent variable, 10 neurons in hidden layer and 3 neurons in output layer from 3 dependent variables.

The integration of Vector Autoregression (VAR) and Neural Networks (NN) in VAR-NN models leverages their complementary strength [27]: VAR's ability to capture linear relationships and lagged effects, and NN's capacity to model nonlinear patterns. This hybrid approach enhances modeling flexibility, preserving VAR's interpretability while accommodating nonlinear relationships. VAR-NN models improve prediction accuracy by capturing residual nonlinear patterns, particularly effective for financial time series exhibiting both linear trends and nonlinear volatility. Furthermore, the non-parametric nature of NNs mitigates the parameter explosion problem in high-dimensional VAR models. Lastly, NNs enhance the model's robustness in handling outliers and extreme events, addressing a key limitation of traditional VAR models.

3.3. *Main results*

In common VAR(p) process-based interval-valued time series forecasting model, $\mathbf{y}_t$ is commonly composed of the lower and the upper bound sequences or the center and the radius sequences [25]. To add the volatility information contained in high-frequency IvTS, the calculation of interval-valued data volatility information is first put forward.

Since an IvTS can be decomposed into two crisp time series, i.e., the lower and the upper bound sequences or the center and the radius sequences, the interval-valued data volatility information is thus constructed from the decomposed subsequences. Let $\mathbf{y}_t = [\mathbf{y}_{1t}, \mathbf{y}_{2t}]^T$, where $\mathbf{y}_{1t}$ and $\mathbf{y}_{2t}$ are, respectively, the decomposed lower and upper bound sequences or center and radius sequences, we consider two types of interval-valued data volatility information, i.e.,

$$f_t = y_{1t} - E(y_{1t}) + y_{2t} - E(y_{2t}) \tag{7}$$

or

$$f_t = [y_{1t} - E(y_{1t})]^2 + [y_{2t} - E(y_{2t})]^2. \tag{8}$$

According to Eq. (7), f_t reflects the sum of relative bias between the interval observation at each time unit and the total mean of all observations. Actually, when the subtraction of interval-valued numbers is determined by Hukuhara [26], i.e., $\tilde{a} -_H \tilde{b} = [a_l - b_l, a_u - b_u]$, then Eq. (7) can be rewritten as $f_t = 2 * c_{[y_{1t}, y_{2t}] - E([y_{1t}, y_{2t}])}$, where $c_{\tilde{*}}$ is the center of an interval $\tilde{*}$. While Eq. (8) provides an absolute bias between the interval observation at each time unit and the total mean of all observations (including possible difference(s)). Since there are two decomposition modes of a high-frequency IvTS, 4 types of variables representing interval-valued data volatility information can be then derived. For convenience, the notation $FlInf \in \{\text{LUR, CRR, LUA, CRA}\}$ is used to mark 4 kinds of variables representing interval-valued data volatility information, we have

$$\begin{aligned} f_t^{\text{LUR}} &= y_t^L - E(y_t^L) + y_t^U - E(y_t^U), \\ f_t^{\text{CRR}} &= c_t - E(c_t) + r_t - E(r_t), \\ f_t^{\text{LUA}} &= [y_t^L - E(y_t^L)]^2 + [y_t^U - E(y_t^U)]^2, \\ f_t^{\text{CRA}} &= [c_t - E(c_t)]^2 + [r_t - E(r_t)]^2. \end{aligned} \tag{9}$$

Common range volatility mainly represents the variance of the observed time series. By Eq. (9), it can be seen that there are four types of information derived from the observed interval-valued time series. All sequences measure the difference between real observations and the means, which can be divided into absolute indicators and relative indicators, i.e., the interval-valued volatilities listed in Eq. (9) are special cases of Eqs. (7) and (8). With the interval-valued data volatility information defined above, $\mathbf{y}_t$ can be renewed as $\mathbf{y}_t = [\mathbf{y}_{1t}, \mathbf{y}_{2t}, \mathbf{y}_{3t}]^T$, where $\mathbf{y}_{1t}$ and $\mathbf{y}_{2t}$ are, respectively, the decomposed lower and upper bound sequences or center and radius sequences, and $\mathbf{y}_{3t} \in \{f_t^{\text{LUR}}, f_t^{\text{CRR}}, f_t^{\text{LUA}}, f_t^{\text{CRA}}\}$ (including possible difference). By the reconstructed $\mathbf{y}_t$ with interval-valued data volatility information, the interval-valued data volatility information will be set as an analyzed object in core VAR(p) model. Correspondingly, a direct mode for forecasting HFIvTS with interval-valued data volatility information can be considered.

Different from classical statistical theory-based parameter solution strategies, the neural network is utilized for the solution of VAR(p), which is denoted as VAR-NN. Table 1 shows the solution algorithm.

By Table 1, the developed neural network-based VAR(p) process with interval-valued data volatility information for high-frequency IvTS can be mainly summarized according to:

Step 1. Data. Real-world HFIvTS collected from Tonghuashun[c] is considered in this study. For the reason that the Shanghai Composite Index occupies a leading position in trading volume, trading value and liquidity in China's stock market, which has received extensive attention, the highest and lowest hourly prices of the Shanghai Composite Index from 6 June 2022 to 7 April 2023 (a total of 828 data points) are collected.

With the collected HFIvTS, the stationarity test or the cointegration test is first discussed. After testing for stationarity or cointegration, the data preprocessing is then prepared for further consideration.

Step 2. Model implementation. To realize the solution and further evaluation of the developed forecasting model, the collected real-world observations are divided into training set and testing set with respectively the ratio of 70% and 30%. For the sake of obtaining high-frequency financial predicted IvTS, the neural network model with standard three-layer feedforward network is combined in this study. The ReLU activation function is used between the input layer and the hidden layer, as well as between the hidden layer and the output layer. By using the optimizer Adam (Adaptive Moment Estimation), the training process is implemented by minimizing the mean squared error (MSE) between the predictions and the real-world observations. The usage of neural network can capture nonlinear relationships contained in the data that traditional VAR(p) models cannot do.

Aiming at illustrating the efficiency of the developed VAR-NN forecasting model, classical VAR(p) process and types of neural network-based forecasting models will also be selected and realized.

Step 3. Evaluation. To show the feasibility and validity of the developed model, the following error indicators are used to realize the comparisons, i.e.,

$$\mathrm{MASE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{\frac{1}{n-1}\sum_{i=2}^{n}|y_i - y_{i-1}|}\right|, \tag{10}$$

$$\mathrm{MAPE}^B = \frac{100}{N}\sum_{n=1}^{N}\frac{|X_n^B - \hat{X}_n^B|}{X_n^B}, \tag{11}$$

$$\mathrm{RMSSE}^B = \sqrt{\mathrm{mean}\left(\frac{(X_n^B - \hat{X}_n^B)^2}{\frac{1}{N-1}\sum_{i=2}^{N}(X_i^B - X_{i-1}^B)^2}\right)}, \tag{12}$$

[c]http://stockpage.10jqka.com.cn/1A0001/.

Table 1. VAR-NN model algorithm.

Function VAR-NN_Model($X \in \mathbb{R}^{T\times 3}$, N_{iter}, N_{train}, N_{hidden})**:**

For $i = 1$ to N_{iter}:

$X_{diff} = \nabla X$

$X_{std} = (X_{diff} - \mu)/\sigma$

$X_{train} = X_{std}[1 : N_{train}]$, $X_{test} = X_{std}[N_{train} + 1 :]$

\# Neural Network Structure

$NN(x) = W_2 \cdot ReLU(W_1 \cdot x + b_1) + b_2$

where:

$W_1 \in \mathbb{R}^{N_{hidden}\times 3}$, $b_1 \in \mathbb{R}^{N_{hidden}}$

$W_2 \in \mathbb{R}^{3\times N_{hidden}}$, $b_2 \in \mathbb{R}^3$

$ReLU(z) = \max(0, z)$

\# Model parameters $\theta = \{W_1, W_2, b_1, b_2\}$

For $epoch = 1$ to N_{epochs}:

For $t = 1$ to $N_{train} - 1$:

$y_t = NN(X_{train}[t])$

$L = ||y_t - X_{train}[t + 1]||^2$

$\nabla\theta = \partial L/\partial\theta$

\# Update all parameters $\theta- = \eta \cdot \nabla\theta$

$\hat{Y}_{train} = [NN(x)\ for\ x \in\ X_{train}]$

$\hat{Y}_{test} = [NN(x)\ for\ x \in\ X_{test}]$

For $S \in \{train, test\}$:

$RMSSE_S = \sqrt{(MSE(\hat{Y}_S, Y_S))}/\sigma_{naive}$

$MASE_S = MAE(\hat{Y}_S, Y_S)/MAE_{naive}$

$MAPE_S = mean(|Y_S - \hat{Y}_S|/(Y_S + \varepsilon))/2$

$ARV_S = \sum(\hat{Y}_S - Y_S)^2/\sum(Y_S - \mu_Y)^2$

$CI_{S,\alpha} = [\mu_{\hat{Y}} \pm z_\alpha \cdot \sigma_{\hat{Y}}/\sqrt{n}]\ for\ alpha\ in \{0.1, 0.05, 0.01\}$

Store_Results ($RMSSE_S$, $MASE_S$, $MAPE_S$, ARV_S, $CI_{S,\alpha}$)

$R_{mean} = mean(Results)$, $R_{std} = std(Results)$

Return R_{mean}, R_{std}

Main:

$X = Load\_Data()$

$R = VAR - NN\_Model(X, N_{iter} = 50, N_{train} = 600, N_{hidden} = 10)$

Output(R)

$$\text{ARV}^{\text{LU}} = \frac{\sum_{t=1}^{n}(X_{t+1}^{U} - \hat{X}_{t+1}^{U})^2 + \sum_{t=1}^{n}(X_{t+1}^{L} - \hat{X}_{t+1}^{L})^2}{\sum_{t=1}^{n}(X_{t+1}^{U} - \bar{X}^{U})^2 + \sum_{t=1}^{n}(X_{t+1}^{L} - \bar{X}^{L})^2}, \tag{13}$$

where the superscript B denotes either the lower bound L, or the upper bound U of the interval prices (the same applies to center and radius). Four classical accuracy measures, i.e., interval average relative variance (ARV^{LU}), mean absolute percentage error (MAPE), mean absolute scaled error (MASE) and root mean squared scaled error (RMSSE) are adopted in this study to compare the forecasting performance.

4. Empirical Findings

Following Step 1 in Sec. 3, the highest and lowest hourly prices of the Shanghai Composite Index from 6 June 2022 to 7 April 2023 (a total of 828 data points) are selected to verify the feasibility and validity of the developed model.

By using Eq. (6), a 828×3 input matrix composed of the lower bound, the upper bound and the volatility information (or corresponding sequences with respect to the center and the radius sequences) is constructed. Next, by partitioning the inputs into training set and testing set, the following procedures are finished.

4.1. *Data preprocessing*

As the precondition of VAR(p) process, the stationarity of each input or the cointegration relationship among all input variables is necessary. Since practical observed time series are commonly not stationary, the difference is often used as a tool for stationarity. Figure 1 shows the initial observed high-frequency financial IvTS and related differential sequences.

The ADF test results before and after differencing of the data are given in Tables 2 and 3.

According to Tables 2 and 3, it can be seen that the initial lower bound sequence y_t^L, the upper bound sequence y_t^U, the center sequence c_t and the interval-valued data volatility information $f_t^{\text{LUA}}, f_t^{\text{CRA}}$ are not stationary. While the differential sequences

Table 2. ADF test values-original value.

Sequence	y_t^L	y_t^U	c_t	r_t
ADF value	0.4524	0.3846	0.4972	0.0067
Sequence	f_t^{LUR}	f_t^{CRR}	f_t^{LUA}	f_t^{CRA}
ADF value	0.4972	0.1297	0.0327	0.0327

Table 3. ADF test values-differenced value.

Sequence	∇y_t^L	∇y_t^U	∇c_t	∇f_t^{LUR}	∇f_t^{CRR}
ADF value	0.0000	0.0000	0.0000	0.0000	0.0010

Table 4. White noise testing results.

Variable	∇y_t^U	∇y_t^L	r_t	∇c_t	$\nabla f_t^{\mathrm{LUR}}$	$\nabla f_t^{\mathrm{CRR}}$	f_t^{LUA}	f_t^{CRA}
Lag: 5	0.013	0.015	0.000	0.000	0.000	0.013	0.000	0.000
Lag: 10	0.040	0.096	0.000	0.000	0.000	0.040	0.000	0.000

of these non-stationary sequences are stationary. As a result, the data used in the subsequent modeling process are all original stationary or differentially stationary sequences. The white noise testing results shown in Table 4 show that those variables are not white noise, which can be further used for modeling.

4.2. *Model implementation*

By setting $p = t = 3$ and $s = 10$, i.e., in the first fully connected layer, the weights corresponding to the input layer and the hidden layer are set as 3 and 10, while the bias in the hidden layer is set as 10. Besides, in the second fully connected layer, the weights corresponding to the hidden layer and the output layer are set as 10 and 3, the bias in the output layer is set as 3. $I_{\max} = 1000$. After 50 times iterations, Figs. 2 to 5 show the mean of forecasting results with respect to four types of interval-valued data volatility information.

According to Figs. 2 to 6, it can be seen that the forecasting results produced by four data groups can well describe the trend of changes in raw observations.

4.3. *Comparisons and analysis*

To illustrate the feasibility and validity of the developed model, the following baselines are selected for further comparisons.

- VAR(p) ([25]). The processed stationary lower bound sequence and upper bound sequence (or center sequence and radius sequence) are set as the inputs of Eq. (1).
- VAR-NN without interval-valued data volatility information. Different from classical VAR(p) process, the parameters in Eq. (1) are produced by using the neural network similarly as the developed model. Similarly, a standard three-layer neural network is applied.
- BEMD-MLP ([33]). The Bidirectional Empirical Mode Decomposition (BEMD) method, proposed by Norden Huang [28], is a powerful tool for handling nonlinear and non-stationary data. This method decomposes the data, isolating different frequency components, which allows for a more detailed and nuanced analysis. Furthermore, the use of Multilayer Perceptrons (MLPs) for Intrinsic Mode Function (IMF) forecasting, as demonstrated by Sun *et al.* [29], effectively models potentially nonlinear IMFs. This divide and conquer approach, combining BEMD and MLPs, provides a robust method for handling the complexities of nonlinearity and non-stationarity in data.

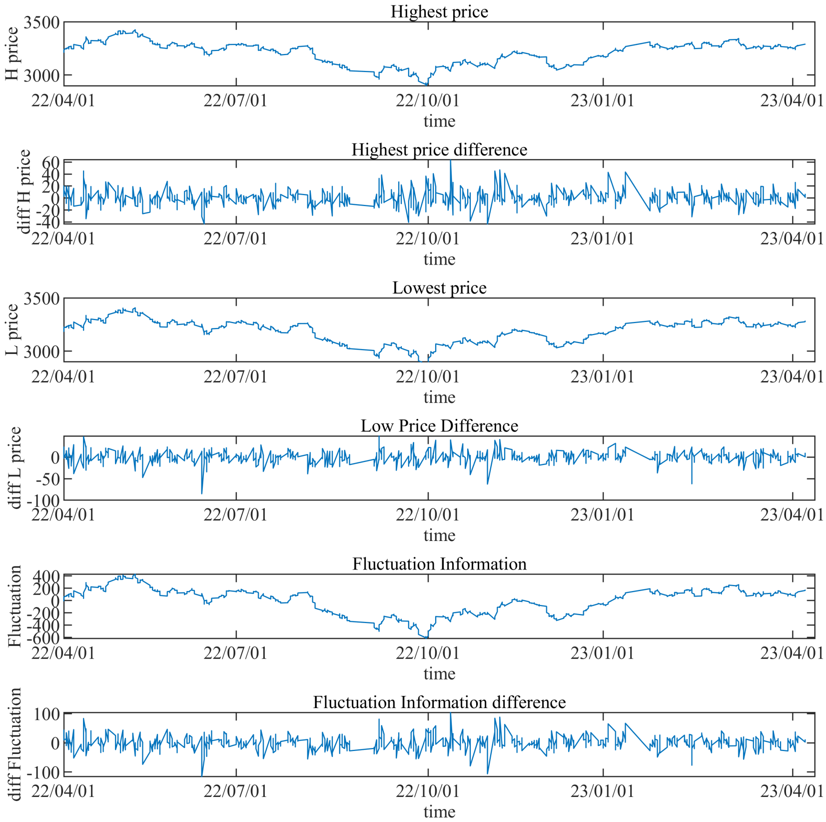

Fig. 2. Time series diagram of lower bound, the upper bound and the interval-valued data volatility information (type-1).

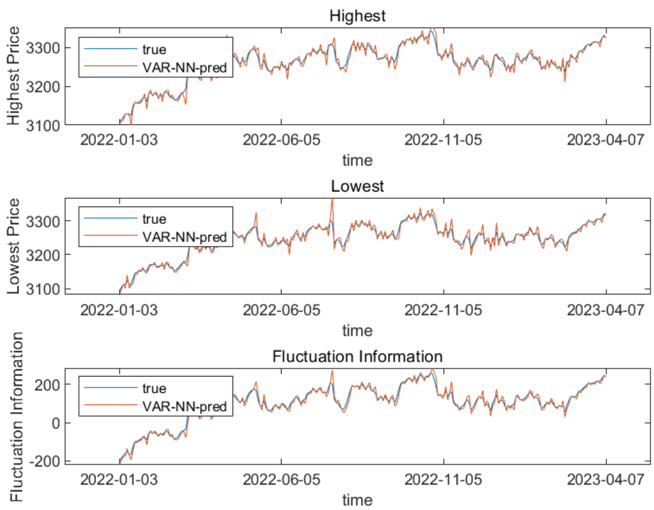

Fig. 3. VAR-NN prediction with y_t^L, y_t^U and f_t^{LUR}.

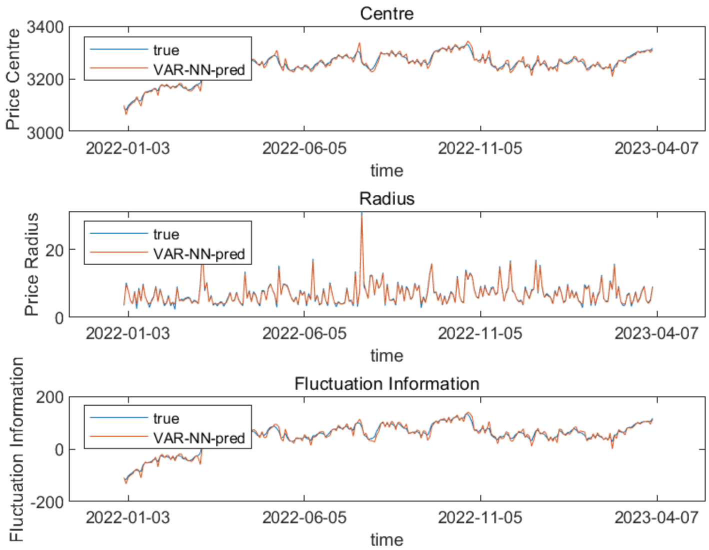

Fig. 4. VAR-NN prediction with c_t, r_t and f_t^{CRR}.

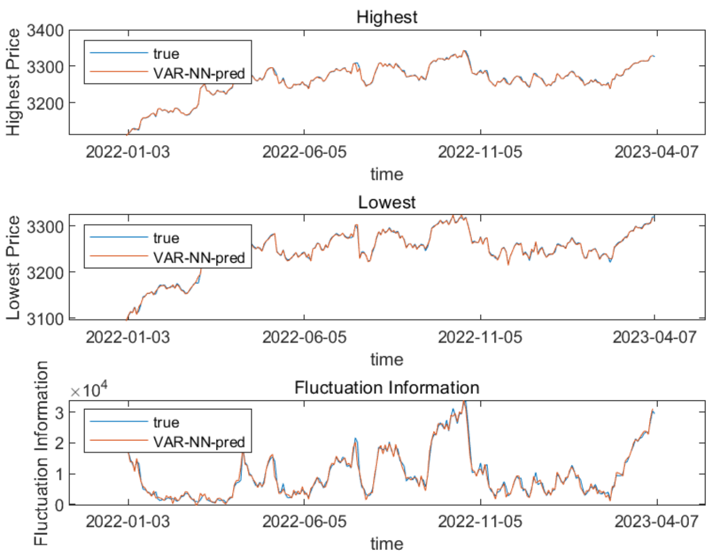

Fig. 5. VAR-NN prediction with y_t^L, y_t^U and f_t^{LUA}.

- ARIMA-NN ([34]). The Autoregressive Integrated Moving Average (ARIMA) models, pioneered by Box and Jenkins in their book "Time Series Analysis: Forecasting and Control" [30], are a cornerstone in time series analysis. These models are capable of capturing trends, seasonality and autocorrelation in time series data. A key feature of ARIMA models is their ability to transform

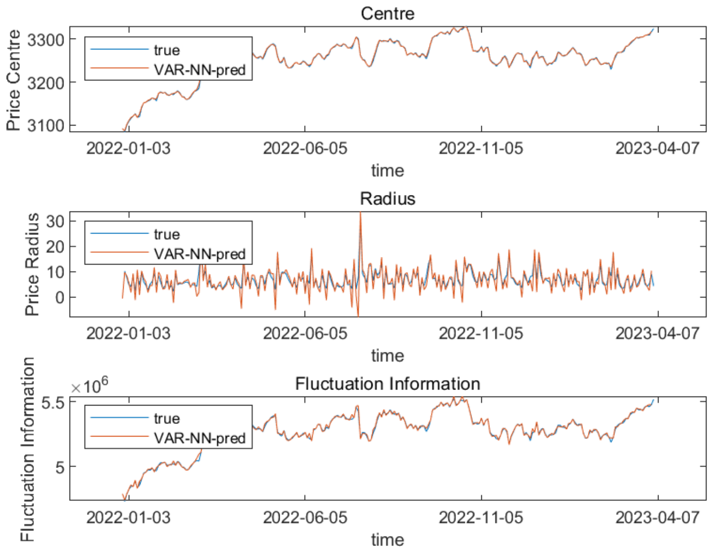

Fig. 6. VAR-NN prediction with c_t, r_t and f_t^{CRA}.

non-stationary series into stationary ones through differencing, thereby facilitating more robust analysis. As linear models, ARIMA models offer interpretable and parsimonious parameters, making them a popular choice for univariate time series forecasting. The systematic Box–Jenkins modeling methodology further enhances the utility of ARIMA models in practical applications.

Herein, by decomposing the initial high-frequency financial IvTS into two crisp time series, the following ARIMA(p, d, q) model can be used for prediction, where d is the order of difference and p, q are, respectively, the lags of autoregression and moving average.

$$\phi(\mathcal{L})\nabla^d Y_t = \theta(\mathcal{L})\epsilon_t,$$

where $\phi(\mathcal{L}) = 1 - \phi_1\mathcal{L} - \cdots - \phi_p\mathcal{L}^p$ and $\theta(\mathcal{L}) = 1 + \theta_1\mathcal{L} + \cdots + \theta_p\mathcal{L}^q$ are the autoregressive coefficient polynomial and moving average coefficient polynomial, $\mathcal{L}$ is the difference operator.

- BPNN ([35]). Backpropagation neural network is a classic artificial neural network model. It is a feed-forward neural network with supervised learning capabilities and is often used to solve classification and regression problems, which is composed of multiple neurons, usually including an input layer, a hidden layer and an output layer. Each neuron is connected to all neurons in the previous layer, forming the forward propagation path of the network. The input layer accepts external input, the hidden layer and the output layer transmit signals layer by layer, and finally produce the output result.
- LSTM ([35]). Long Short-Term Memory is a special type of Recurrent Neural Network (RNN), which is widely used in processing and modeling time series data.

Compared with traditional RNN, LSTM has better performance in dealing with long-term dependency problems.

- GARCH ([20]). The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model was introduced, which is a generalization of the ARCH model [31]. The GARCH model is used to estimate volatility in financial time series. It assumes that the variance of the current error term is a function of previous squared error terms and previous variance terms. The GARCH model allows the conditional variance to be dependent on previous own lags, which makes it more flexible than the ARCH model. The main application of GARCH models is in modeling and forecasting volatility in financial markets. It is commonly used to model and predict asset price volatility in stocks, bonds, currencies and commodities [32].

Tables 5 and 6 show the comparisons among different methods with the training set and the testing set.

According to Tables 5 and 6, the following conclusions are valid:

(1) By considering the developed types of interval-valued data volatility information, the developed high-frequency IvTS forecasting model with VAR(p) process, neural network and interval-valued data volatility information performs the best when comparing with the other baselines. In the training stage, the usage of interval-valued data volatility information calculated by f_t^{CRA} can improve the forecasting accuracy the best. While in the testing stage, the type of interval-valued data volatility information given by f_t^{LUR} shows the best performance.

Table 5. Comparisons of forecasting results on training set.

Method	MASE	MAPE	RMSSE	ARV
VAR-NN with f_t^{LUR}	0.7724 ± 0.0020	0.9701 ± 0.0351	0.8774 ± 0.0030	$\mathbf{1.01979 \pm 0.0062}$
VAR-NN with f_t^{CRR}	0.9035 ± 0.0085	1.3129 ± 0.0525	1.3106 ± 0.0235	1.1462 ± 0.0779
VAR-NN with f_t^{LUA}	0.7823 ± 0.0164	$\mathbf{0.8646 \pm 0.1794}$	0.9067 ± 0.0084	1.0364 ± 0.0176
VAR-NN with f_t^{CRA}	$\mathbf{0.7419 \pm 0.0009}$	0.9229 ± 0.0095	$\mathbf{0.8768 \pm 0.0014}$	1.0681 ± 0.0040
VAR-NN$^{\mathrm{LH}}$	1.1231 ± 0.0115	1.3782 ± 0.0233	1.1894 ± 0.0123	1.2438 ± 0.0132
VAR-NN$^{\mathrm{CR}}$	1.0138 ± 0.1132	1.4233 ± 0.0089	1.0295 ± 0.0034	1.3321 ± 0.0325
BEMD-MLP$^{\mathrm{LU}}$	1.1238 ± 0.3960	1.3328 ± 0.0811	1.0950 ± 0.0481	1.0869 ± 0.0026
BEMD-MLP$^{\mathrm{CR}}$	1.1240 ± 0.0602	1.2328 ± 0.0182	1.1907 ± 0.0722	1.1869 ± 0.0119
LSTM$^{\mathrm{LU}}$	1.0329 ± 0.0015	1.3625 ± 0.0112	1.0012 ± 0.0028	1.3038 ± 0.1304
LSTM$^{\mathrm{CR}}$	0.8072 ± 0.0011	1.2948 ± 0.0991	1.0966 ± 0.0015	1.1985 ± 0.0027
ARIMA-NN$^{\mathrm{LU}}$	1.4590 ± 0.1099	1.5291 ± 0.0324	1.5512 ± 0.0127	1.4588 ± 0.0562
ARIMA-NN$^{\mathrm{CR}}$	1.3475 ± 0.2099	1.4291 ± 0.0722	1.3116 ± 0.0220	1.2539 ± 0.0780
BPNN$^{\mathrm{LU}}$	1.2590 ± 0.0785	1.3581 ± 0.0352	1.2532 ± 0.0284	1.5588 ± 0.0331
BPNN$^{\mathrm{CR}}$	1.1475 ± 0.0116	1.5291 ± 0.0235	1.2336 ± 0.0106	1.3539 ± 0.0254
GARCH$^{\mathrm{LU}}$	1.4381	1.5213	1.2262	1.2431
GARCH$^{\mathrm{CR}}$	1.3182	1.3342	1.1168	1.1345
VAR$^{\mathrm{LU}}$	1.8955	1.6468	1.3981	1.1494
VAR$^{\mathrm{CR}}$	1.6955	1.5468	1.4343	1.4651

Table 6. Comparisons of forecasting results on testing set.

Method	MASE	MAPE	RMSSE	ARV
VAR-NN with f_t^{LUR}	$\mathbf{0.8038 \pm 0.0019}$	1.0368 ± 0.0594	$\mathbf{0.9357 \pm 0.0018}$	$\mathbf{1.0469 \pm 0.0033}$
VAR-NN with f_t^{CRR}	1.0864 ± 0.0177	1.2331 ± 0.0412	1.0922 ± 0.0078	1.1398 ± 0.0147
VAR-NN with f_t^{LUA}	0.8168 ± 0.0124	1.0483 ± 0.0718	0.9695 ± 0.0041	1.0691 ± 0.0191
VAR-NN with f_t^{CRA}	0.8768 ± 0.0014	$\mathbf{1.0172 \pm 0.0271}$	0.8646 ± 0.0011	1.1011 ± 0.0027
VAR-NN$^{\text{LU}}$	1.3135 ± 0.0259	1.4625 ± 0.0855	1.2394 ± 0.0320	1.3438 ± 0.1000
VAR-NN$^{\text{CR}}$	1.1284 ± 0.0132	1.4357 ± 0.0387	1.1874 ± 0.0053	1.2355 ± 0.0238
BEMD-MLP$^{\text{LU}}$	1.2256 ± 0.2619	1.2385 ± 0.2731	1.1933 ± 0.1191	1.1777 ± 0.0161
BEMD-MLP$^{\text{CR}}$	1.1465 ± 0.0215	1.2458 ± 0.0083	1.2207 ± 0.0072	1.1908 ± 0.0225
LSTM$^{\text{LU}}$	1.0165 ± 0.0012	1.2764 ± 0.0110	1.0022 ± 0.0024	1.1758 ± 0.0815
LSTM$^{\text{CR}}$	0.8288 ± 0.0013	1.3471 ± 0.0572	1.1846 ± 0.0015	1.2156 ± 0.0028
ARIMA-NN$^{\text{LU}}$	1.6590 ± 0.0785	1.9342 ± 0.2138	1.4282 ± 0.1214	1.7527 ± 0.2991
ARIMA-NN$^{\text{CR}}$	1.3482 ± 0.3179	1.9291 ± 0.5232	1.3230 ± 0.2177	1.5439 ± 0.7805
BPNN$^{\text{LU}}$	1.3950 ± 0.0673	1.4215 ± 0.0234	1.1552 ± 0.0227	1.4538 ± 0.0831
BPNN$^{\text{CR}}$	1.1242 ± 0.0169	1.2345 ± 0.0753	1.4664 ± 0.0882	1.3559 ± 0.0665
GARCH$^{\text{LU}}$	1.3381	1.4796	1.2384	1.2135
GARCH$^{\text{CR}}$	1.2562	1.2534	1.2413	1.1643
VAR$^{\text{LU}}$	1.7623	1.6552	1.4450	1.3321
VAR$^{\text{CR}}$	1.6675	1.5822	1.4768	1.3554

It can be seen that the introduction of interval-valued data volatility information in traditional VAR(p) process is helpful.

(2) Different from the way to handle the volatility information in the mean equation given by GARCH, the developed VAR-NN with volatility information introduces a direct mode. Compared with classical statistical theory-based forecasting models include ARIMA and VAR, the developed forecasting model can effectively improve the forecasting accuracy.
(3) The combination of neural network has shown better performance when comparing with the methods without such combination. Besides, the developed hybrid high-frequency financial IvTS forecasting model can provide forecasting results with both interpretability and high accuracy. Furthermore, such method puts forward a novel interval-valued time series prediction technology from a holistic perspective, i.e., the inputs and outputs of the forecasting model are both interval-valued time series.

4.4. *Confidence intervals*

Table 7 compares the performance of various models at three different confidence levels ($\alpha = 0.01, 0.05, 0.10$). The study focuses on the coverage of confidence intervals and appropriately widens the interval width according to Eqs. (7) and (8), employing the union method. In the test set data, regardless of whether α is 0.01, 0.05, or 0.10, the confidence intervals of all models can effectively predict the original intervals.

Figure 7 shows that the VAR-NN with f_t^{LUA} model can fully predict the mean interval of the test set, achieving the highest coverage. However, Table 7 indicates

Table 7. Average internal width of the confidence interval of VAR.

Method	AIW[a] ($\alpha = 0.01$)	AIW ($\alpha = 0.05$)	AIW ($\alpha = 0.10$)
VAR-NN with f_t^{LUR}	**281.5205**	**180.8052**	**79.4067**
VAR-NN with f_t^{CRR}	439.0052	399.7024	345.6282
VAR-NN with f_t^{LUA}	330.9625	234.9024	111.0647
VAR-NN with f_t^{CRA}	418.5107	290.2751	150.8450
VAR-NN$^{\mathrm{LU}}$	574.0028	427.9762	379.7758
VAR-NN$^{\mathrm{CR}}$	576.6380	261.0442	54.5496
BEMD-MLP$^{\mathrm{LU}}$	589.4098	332.8486	179.3827
BEMD-MLP$^{\mathrm{CR}}$	637.7822	342.0099	173.9077
LSTM$^{\mathrm{LU}}$	386.1545	331.1712	168.4328
LSTM$^{\mathrm{CR}}$	434.5268	310.3324	162.9578
ARIMA-NN$^{\mathrm{LU}}$	782.8992	369.4937	157.4829
ARIMA-NN$^{\mathrm{CR}}$	831.2715	378.6549	152.0079
BPNN$^{\mathrm{LU}}$	579.6439	387.8162	216.5330
BPNN$^{\mathrm{CR}}$	628.0162	396.9774	241.0580
GARCH$^{\mathrm{LU}}$	576.3885	406.1387	235.5831
GARCH$^{\mathrm{CR}}$	624.7609	315.2999	130.1081
VAR$^{\mathrm{LU}}$	873.1332	524.4612	124.6332
VAR$^{\mathrm{CR}}$	821.5056	333.6224	119.1582

Note: [a]AIW represents the average interval width.

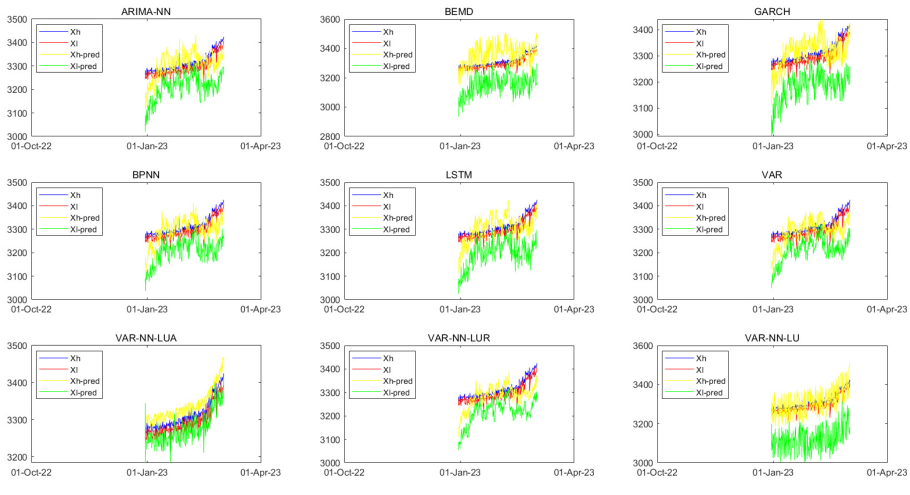

Fig. 7. 90% confidence intervals of predicted interval-valued time series.

that the VAR-NN with f_t^{LUR} model predicts the narrowest confidence intervals ($\alpha = 0.01$: 281.5202, $\alpha = 0.05$: 180.8052, $\alpha = 0.10$: 79.4067), but its prediction performance is poor.

Overall, from the perspective of confidence intervals, regardless of whether α is 0.01, 0.05, or 0.10, the VAR-NN with f_t^{LUA} model outperforms other comparative models, demonstrating the effectiveness of this model.

Figure 7 presents the 90% confidence interval prediction results for different models, visually illustrating the performance of each model.

5. Concluding Remarks

The hybrid interval-valued financial time series forecasting model developed in this paper provide a novel way to handle the possible interval-valued data volatility information contained in the initial observations. Since the volatility information contained in the IvTS has merely been reported, the developed forecasting method extends the application of VAR(p) process in IvTS analysis. The results produced by the application of developed model to the highest and lowest hourly prices of the Shanghai Composite Index show the superiority, feasibility and validity. The advantages of the developed forecasting model include global interval-valued input and output and processing of volatility information in an integrated model. The introduction of neural network also improves the accuracy.

The constructed model tries to provide a direct way for the forecasting of interval-valued time series by reconsidering classical model like GARCH(p, q). Although the interval-valued data is a measure of volatility in essence, as a type of numerical data, the volatility in interval-valued time series forecasting is also worth to be considered. By Eq. (9), it can be seen that there are four types of information derived from the observed interval-valued time series. All sequences measure the difference between real observations and the means, which can be divided into absolute indicators and relative indicators. It can be seen that different forms of volatility have diverse efficiency in forecasting process. Similar to Residual gray prediction model [38], we try to construct such a direct mode for the forecasting of interval-valued time series, i.e., the volatility can be seen as a supplementary information in interval-valued time series analysis.

AQ: Please approve. We have deleted the Data Availability, Ethics, Conflict sections.

In the future, it is direct to generalize the novel mode to the other types of time series, for instance, the interval gray time series [36].

Acknowledgments

The study was supported in part by the Humanities and Social Sciences Research Youth Project of the Ministry of Education of China (No. 21YJCZH148), the Natural Science Foundation of Anhui Province (No. 2108085MG239) and the National Natural Science Foundation of China (No. U22A20366).

ORCID

Xu Hu https://orcid.org/0009-0003-4434-3937
Jianwen Yu https://orcid.org/0009-0003-2579-5209
Qin Xu https://orcid.org/0000-0002-9020-2006
Zhifu Tao https://orcid.org/0000-0003-4039-9178

References

[1] W. G. Gloria, Interval-valued time series models: Estimation based on order statistics exploring the Agriculture Marketing Service data, *Comput. Stat. Data Anal.* **100** (2016) 694–711.

[2] Y. Sun, X. Y. Zhang, A. T. K. Wan *et al.*, Model averaging for interval-valued data, *Eur. J. Oper. Res.* **301**(2) (2022) 772–784.

[3] A. B. Ji, J. J. Zhang, X. He and Y. H. Zhang, Fixed effects panel interval-valued data models and applications, *Knowl.-Based Syst.* **237** (2022) 107798.1–107798.12.

[4] L. Sun, K. Wang, L. Xu, C. Zhang and T. Balezentis, A time-varying distance based interval-valued functional principal component analysis method - A case study of consumer price index, *Inf. Sci.* **589** (2022) 94–116.

[5] D. Yang, J. E. Guo, S. Sun, J. Han and S. Wang, An interval decomposition-ensemble approach with data-characteristic-driven reconstruction for short-term load forecasting, *Appl. Energy* **306** (2022) 117992.

[6] J. Arroyo, R. Espínola and C. Maté, Different approaches to forecast interval time series: A comparison in finance, *Comput. Econ.* **27** (2011) 169–191.

[7] A. L. S. Maia, F. D. A. T. D. Carvalho and T. B. Ludermir, Forecasting models for interval-valued time series, *Neurocomputing* **71**(16–18) (2008) 3344–3352.

[8] T. Xiong, C. Li and Y. Bao, Interval-valued time series forecasting using a novel hybrid Holt and MSVR model, *Econ. Model.* **60** (2017) 11–23.

[9] A. L. S. Maia and F. D. A. T. D. Carvalho, Holt's exponential smoothing and neural network models for forecasting interval-valued time series, *Int. J. Forecast.* **27**(3) (2011) 740–759.

[10] P. Hajek, W. Froelich and O. Prochazka, Intuitionistic fuzzy grey cognitive maps for forecasting interval-valued time series, *Neurocomputing* **400** (2020) 173–185.

[11] C. Ouyang, F. Yu, Y. Hao, Y. Tang and Y. Jiang, Build interval-valued time series forecasting model with interval cognitive map trained by principle of justifiable granularity, *Inf. Sci.* **652** (2024) 119756.

[12] H. L. Chien, S. Lee and L. C. Lin, Symbolic interval-valued data analysis for time series based on auto-interval-regressive models, *Stat. Methods Appl.* **30** (2021) 295–315.

[13] B. Hamiye, Bootstrap based multi-step ahead joint forecast densities for financial interval-valued time series, *Communications* **70**(1) (2021) 156–179.

[14] J. Guo, W. Lu, J. Yang and X. Liu, A rule-based granular model development for interval-valued time series, *Int. J. Approx. Reason.* **136** (2021) 201–222.

[15] T. T. Buansing, A. Golan and A. Ullah, An information-theoretic approach for forecasting interval-valued SP500 daily returns, *Int. J. Forecast.* **36** (2020) 800–813.

[16] J. Geng, X. W. Zhang, X. F. Chen and X. F. Xue, High-frequency vibrations prediction of kirchhoff plate based on b-spline wavelet on interval finite element method, *Sci. China Technol. Sci.* **60**(5) (2017) 792–806.

[17] V. Vella and W. L. Ng, Improving risk-adjusted performance in high frequency trading using interval type-2 fuzzy logic, *Expert Syst. Appl.* **55** (2016) 70–86.

[18] X. B. Kong, Confidence interval of the jump activity index based on empirical likelihood using high frequency data, *J. Stat. Plan. Inference* **142**(6) (2012) 1378–1387.

[19] L. Maciel, R. Ballini and F. Gomide, Adaptive fuzzy modeling of interval-valued stream data and application in cryptocurrencies prediction, *Neural Comput. Appl.* **35** (2023) 7149–7159.

[20] Y. Sun, G. Lian, Z. Lu, J. Loveland and I. Blackhurst, An interval-valued GARCH model for range-measured return processes (2019), arXiv:1901.02947.

[21] R. E. Moore, *Methods and Applications of Interval Analysis* (SIAM, 1979).

[22] C. Cappelli, C. Cerqueti, P. DUrso and F. D. Iorio, Multiple breaks detection in financial interval-valued time series, *Expert Syst. Appl.* **164** (2021) 113775.

[23] T. Pantelidis and N. Pittis, Estimation and forecasting in first-order vector autoregressions with near to unit roots and conditional heteroscedasticity, *J. Forecast.* **28**(7) (2009) 612–630.

[24] A. Agboluaje, S. Ismail, Y. C. Yin and A. Rahman, Research article comparing vector autoregressive (VAR) estimation with combine white noise (CWN) estimation, *Res. J. Appl. Sci., Eng. Technol.* **12**(5) (2016) 544–549.

[25] C. Garcia-Ascanio and C. Mate, Electric power demand forecasting using interval time series: A comparison between VAR and iMLP, *Energy Policy* **38**(2) (2010) 715–725.

[26] M. Hukuhara, Intégration des applications measurable dont la valeur est un compact convexe, *Funkcial. Ekvac.* **10** (1967) 205–223 (in French).

[27] D. U. Wutsqa, The Var-NN model for multivariate time series forecasting, *MatStat* **8**(1) (2008) 35–43.

[28] Y. X. Huang and F. G. Schmitt, Time dependent intrinsic correlation analysis of temperature and dissolved oxygen time series using empirical mode decomposition, *J. Mar. Syst.* **130** (2014) 90–100.

[29] S. Sun, Y. Sun, S. Wang and Y. Wei, Interval decomposition ensemble approach for crude oil price forecasting, *Energy Econ.* **76** (2018) 274–287.

[30] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control* (Holden-day, San Francisco, CA, 1970).

[31] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, *J. Econom.* **31**(3) (1986) 307–327.

[32] R. F. Engle, Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* **50**(4) (1982) 987–1007.

[33] Z. C. Wang, L. R. Chen, Z. N. Ding and H. Y. Chen, An enhanced interval pm_(2.5) concentration forecasting model based on bemd and MLPp 1 with influencing factors, *Atmos. Environ.* **223** (2020) 117200.1–117200.16.

[34] D. O. Faruk, A hybrid neural network and ARIMA model for water quality time series prediction, *Eng. Appl. Artif. Intell.* **23**(4) (2010) 586–594.

[35] J. P. Liu, P. Wang, H. Y. Chen and J. M. Zhu, A combination forecasting model based on hybrid interval multi-scale decomposition: Application to interval-valued carbon price forecasting, *Expert Syst. Appl.* **191** (2022) 116267.

[36] H. L. Huang, Z. F. Tao, J. P. Liu, J. H. Cheng and H. Y. Chen, Exploiting fractional accumulation and background value optimization in multivariate interval grey prediction model and its application, *Eng. Appl. Artif. Intell.* **104** (2021) 104360.

[37] C. N. Taylor, Evaluating interval forecasts of high-frequency financial data, *J. Appl. Econom.* **18**(4) (2010) 445–456.

[38] S. Prakash, A. Agrawal, R. Singh, R. K. Singh and D. Zindani, A decade of grey systems: Theory and application — Bibliometric overview and future research directions, *Grey Syst.: Theory Appl.* **13**(1) (2023) 14–33.