Journal of Systems Science and Information
Jun., 2025, Vol. 13, No. 3, pp. 325–344
DOI: 10.12012/JSSI-2024-0047

Can a Hierarchical Approach Using Interval Information Improve Gasoline Volatility Forecast?

Yao YUE

Post-Doctoral Research Center, Agricultural Bank of China, Beijing 100005*, China; School of Finance, Renmin University of China, Beijing* 100190*, China*

Qi ZHANG

Business School, Beijing Technology and Business University, Beijing 100048*, China; School of Economics and Management, and MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation, University of Chinese Academy of Sciences, Beijing* 100190*, China*
*E-mail: zhangqi*190*@mails.ucas.ac.cn*

Yuying SUN*

State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190*, China; School of Economics and Management, and MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation, University of Chinese Academy of Sciences, Beijing* 100190*, China*

Shouyang WANG

State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190*, China; School of Economics and Management, and MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation, University of Chinese Academy of Sciences, Beijing* 100190*, China*

Abstract Accurately forecasting gasoline volatility is significant for risk management, economic analysis, and option pricing formulas for future contracts. This study proposes a novel interval-valued hierarchical decomposition and ensemble (IHDE) approach to investigate gasoline price volatility. Our interval-based IHDE method can decompose the complex price process into different components to capture the distinct features of each component, which is helpful for forecasting and analyzing complex price processes. By using interval-valued data, the dynamics of gasoline prices in terms of levels and variations can be fully utilized in this study. Fully utilizing the informational gain of interval-valued

Received April 14, 2024, accepted July 29, 2024
Supported by National Natural Science Foundation of China (72322016, 72073126, 71988101); Beijing Natural Science Foundation (9254024)
*Corresponding author

data improves forecasting performance. In forecasting weekly gasoline volatility, we document that the proposed IHDE approach outperforms the GARCH, EGARCH, CARR, and ACI models, indicating the importance of capturing features of different frequency components and utilizing the informational gain of interval-valued data for gasoline volatility forecasts.

Keywords volatility forecast; interval-valued time series; variational mode decomposition; ACI model; interval neutral network; interval Holt's model

1 Introduction

Gasoline is an indispensable strategic commodity in modern economies, and its price fluctuations have profound impacts on countries' economic performance[1,2]. Gasoline price surges will weaken consumption and investment, leading to slower economic growth[2,3]. Gasoline prices are affected by international crude oil markets, policy changes, and other major events. For example, the lockdown policy in many countries during the COVID-19 outbreak period decreases traveling needs, leading to a drop in demand for gasoline and a falling price. The geopolitical conflicts such as the Russia-Ukraine War have also led to great fluctuations in gasoline price[4]. Historically, gasoline prices have shown a tendency to be nonlinear and uncertain in their behavior. Therefore, developing effective forecasting models for gasoline price volatility is an important research agenda.

Nowadays, one class of popular volatility forecasting models is the generalized autoregressive conditional heteroskedasticity (GARCH) model[5] and its extensions. The GARCH model can capture volatility clustering and time-vary heteroskedasticity. To capture the nonlinear characteristics, Nelson[6] developed the Exponential GARCH (EGARCH) model to forecast volatility. GARCH model and its extensions are applied for volatility forecasting in the fields of equities[7–9], foreign exchange[10,11], interest rates[12], cryptocurrencies[13,14] and commodities markets[15,16]. On the other hand, GARCH-type models only use closing prices, which may neglect the information on high and low prices and not reflect variability or uncertainty in a given period. Models harnessing the information content of range price (high price-low price) can deliver superior volatility forecasts[17–19]. This is because range price delivers fluctuation information in a given period while close price does not. Chou[19] proposed the conditional autoregressive range (CARR) model and found that the CARR model can produce more accurate volatility forecasts than GARCH-type models. However, the range data only reflects price fluctuations and neglects the price level information[20]. For example, the last week of Dec.2013 and the second week of Dec.2019 have the same range ($ 0.705 per gal), while the midpoint prices are $ 2.81 per gal and $ 1.65 per gal, respectively.

Instead, some literature attempts to model the price process by using interval-valued time series (ITS) and associated tools to study the price characteristics[21–31]. Modeling ITS can encompass information on price trends and price fluctuations simultaneously[20,32]. One class of interval-valued models simplifies the ITS into a traditional point-valued process (i.e., low price and high price, midpoint and range price)[21,23,33], which neglects the inner information

of intervals. The other streams consider the interval as an inseparable extended random set that contains more information about intervals to generate more precise predictions[20,29,32]. One popular method is the autoregressive conditional interval (ACI) model proposed by He et al.[32]. The ACI model with two-stage D_K-distance estimation can capture the interval-valued gasoline price dynamics and is more parsimonious than prior interval-valued models. In addition, the two-stage D_K-distance estimation can efficiently and simultaneously use the information contained in intervals to estimate model parameters by a kernel function (see more details in He et al.[32]). Thus, the ACI model can alleviate the overparameterization issue, generate accurate forecasts, and enhance model parameter efficiency[20].

However, the gasoline price process is complex, and the aforementioned econometric models cannot fully capture the complex features such as nonlinear characteristics. Thus, based on the "divide and conquer" principle, some interval-based hybrid ensemble approaches have been applied to energy markets[34–37] and have better forecasting performances. Nevertheless, these hybrid approaches are employed for price forecasting, and no such interval-based hybrid approach is proposed for gasoline volatility forecast to our best known.

To address these limitations, we follow the "divide and conquer" principle to propose a novel interval-based hybrid multivariate decomposition and ensemble (IHDE) approach for gasoline price volatility forecasting based on different frequency components. Our IHDE approach is the first method to forecast gasoline volatility according to different frequency components by using interval-valued data to our best known. We forecast gasoline volatility by using weekly interval-valued prices based on the IHDE approach. First, to capture the volatility characteristics from different frequency components, we employ multivariate variational mode decomposition (MVMD)[38] to decompose the original interval-valued price into low- and high-frequency components. The low-frequency component is the fundamental factor of gasoline volatility. The high-frequency component refers to the fluctuations caused by short-term supply-demand disequilibrium or market noise. Although bivariate empirical mode decomposition (BEMD) is popularly used in the interval-valued decomposition[4,34], it suffers from end effects, mode mixing, and lack of robustness to noise[38], which may affect the forecast performance. Second, to capture the dynamics of the fundamental factors of gasoline volatility, we apply the ACI model to forecast the low-frequency component of gasoline volatility. Third, we apply the MLPI-HoltI approach proposed by Maia and de Carvalho[39] to distinguish the short-term supply-demand disequilibrium and market noise of high-frequency component. The HoltI model is applied to capture the short-term supply-demand disequilibrium characteristics of the high-frequency component, that is, forecasting the linear structure of the high-frequency component by the HoltI model. The remaining high-frequency component is often referred to as market noise, which is typically a nonlinear structure of the high-frequency component. The MLPI method is used to forecast the nonlinear structure. Furthermore, we ensemble the forecasted low- and high-frequency components to generate the forecasted volatility. Our IHDE approach outperforms all benchmarks, including GARCH, EGARCH, CARR, and ACI models.

Compared with existing literature for gasoline volatility forecasting, our proposed IHDE approach has some appealing features. First, our IHDE approach is the first interval-valued hybrid decomposition and ensemble approach for gasoline volatility forecasting to our best known. This approach captures the different volatility characteristics from different frequency components, revealing that volatility comprises both persistent long-term and transient short-term dynamics, consistent with Engle et al.[40]. Second, superior to the traditional point-based and range-based models, the interval-valued model can better capture both the trend and volatility dynamics of the gasoline simultaneously due to the fully utilizing informational gains of the ITS data. Third, MVMD is applied in our approach to deal with the complex interval-valued gasoline price, which can benefit from the "divide and conquer" principle and overcome the shortcomings of the EMD-based method. Our IHDE approach also outperforms the ACI model, verifying the importance of forecasting gasoline volatility according to the different frequency components by using ITS data.

The rest of this paper is organized as follows. Section 2 proposes the IHDE approach. Section 3 introduces the benchmark models for forecasting gasoline price volatility. Section 4 describes the sample data and provides some basic analysis. Section 5 discusses the empirical results of forecasting comparison. Finally, Section 6 concludes.

2 Methodology

2.1 A Novel IHDE Framework for Gasoline's Volatility Forecasting

In this part, we establish a novel interval-valued hybrid multivariate decomposition and ensemble (IHDE) framework to forecast the volatility of gasoline prices.

Given an interval-valued gasoline price series $y_t = [y_t^L, y_t^U]$, $t = 1, 2, \cdots, n$, y_t^L and y_t^U denote the low bound and high bound of the interval-valued gasoline price at time t.

Figure 1 shows that the IHDE learning approach is generally composed of the following 4 steps. First, we employ MVMD to decompose the input signal $y_t = [y_t^L, y_t^U]$ $(t = 1, 2, \cdots, n)$

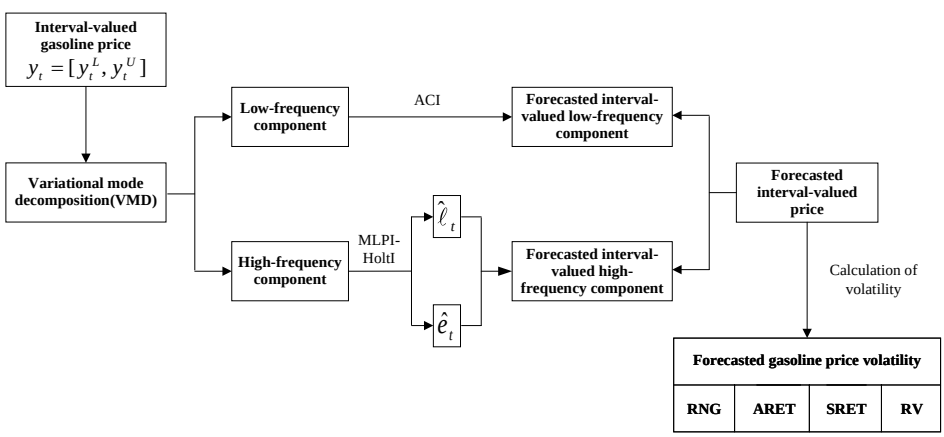

Figure 1 The framework of IHDE approach

into 2 VMF components g_t^i, $(i = 1, 2)$. The VMF with a higher center frequency is the high-frequency component, and the other is the low-frequency component. Second, we forecast high- and low-frequency components separately. We employ MLPI-HoltI developed by Maia and de Carvalho [39] to forecast the low and high boundaries of the high-frequency component simultaneously, and the ACI model proposed by Han et al. [41] to model the low-frequency component. In addition, we ensemble the forecasted high- and low-frequency components to generate the interval-valued gasoline price forecasting results. Finally, we calculate the volatility of gasoline prices based on the forecasting results. See Section 4.1 for more calculation details.

2.2 MVMD for Interval-Valued Time Series

MVMD is a multivariate extension of the variational mode decomposition (VMD) technique, which overcomes the shortcomings of empirical mode decomposition (EMD) extensions, such as bivariate EMD (BEMD) and multivariate EMD (MEMD). These EMD extensions also suffer from the drawbacks of the original EMD algorithm, including sensitivity to sampling rate, lack of robustness to noise, mode aliasing, ending effects, and other issues related to the algorithmic nature of EMD [38]. In this study, MVMD is employed as an effective method for decomposing interval-valued gasoline prices, enabling the original interval-valued gasoline price data to be decomposed into high- and low-frequency components based on their frequency patterns and oscillatory characteristics, thereby overcoming the deficiencies of the EMD algorithm and its extensions. The MVMD method is as follows:

1) For the input data y_t containing C channels, i.e., $y_t = [y_1(t), y_2(t), \cdots, y_C(t)]$

$$y_t = \sum_{k=1}^{K} u_k(t), \tag{1}$$

where $u_k(t) = [u_1(t), u_2(t), \cdots, u_C(t)]$.

2) Obtain an analytical modulated signal $u(t)$ by using the Hilbert transform operator, denoted as $u_+^k(t)$, and then obtain the single frequency component ω_k, which is used in the harmonic mixing of the whole vector $u_+^k(t)$. In addition, the bandwidths of $u_k(t)$ are estimated by the L_2 norm of the gradient function of harmonically transferred $u_+^k(t)$. To ensure the extracted multivariate oscillation signal can reconstruct the original signal fully, the sum of bandwidths of all modes should be minimized. The constrained optimization problem for MVMD is:

$$\begin{aligned} &\min_{\{u_+^{k,c}\},\{\omega_k\}} \left\{ \sum_k \sum_c \|\partial_t[u^{k,c}(t)\mathrm{e}^{-\mathrm{j}\omega_k \mathrm{t}}]\|_2^2 \right\} \\ &\text{s.t.} \quad \sum_k u_{k,c}(t) = y_c(t), c = 1, 2, \cdots, C. \end{aligned} \tag{2}$$

3) Using ADMM approach to solve the multi-constrained optimization problem (see details in ur Rehman and Aftab [38]), and the augmented Lagrangian function is as:

$$\mathscr{L}(\{u_{k,c}\},\{\omega_k\},\lambda_c)=\alpha\sum_k\sum_c\|\partial_t[u_+^{k,c}(t)e^{-\mathrm{j}\omega_k t}]\|_2^2+\sum_c\left\|y_c(t)-\sum_k u_{k,c}(t)\right\|_2^2+\sum_c\left\langle\lambda_c(t),y_c(t)-\sum_k\boldsymbol{u}_{k,c}(t)\right\rangle. \tag{3}$$

4) The optimization problem in (3) is solved by the mode update using the multiplier alternating direction method and by the central frequency updates. Then, the mode update is as follows:

$$\hat{u}_{k,c}^{n+1}(\omega)=\frac{\hat{y}_c(\omega)-\sum_{i\neq k}\hat{u}_{i,c}(\omega)+\frac{\hat{\lambda}_c(\omega)}{2}}{1+2\alpha(\omega-\omega_k)^2}, \tag{4}$$

$$\omega_k^{n+1}=\frac{\sum_c\int_0^\infty\omega|\hat{u}_{k,c}(\omega)|^2\mathrm{d}\omega}{\sum_c\int_0^\infty|\hat{u}_{k,c}(\omega)|^2\mathrm{d}\omega}. \tag{5}$$

Finally, the K VMFs are obtained by decomposing the original signals according to the abovementioned steps.

2.3 ACI Model for Low-Frequency Component

He et al.[32] developed the ACI model to capture the conditional mean dynamics of interval time series processes. An ACI model with order (p,q) is as follows:

$$\Delta Y_t=\alpha_0+\beta_0 I_0+\sum_{i=1}^{j}\beta_i\Delta Y_{t-i}+\sum_{i=1}^{q}\gamma_i\Delta u_{t-i}+u_t, \tag{6}$$

where Y_t is the interval-valued time series (ITS), Δ denotes the Hukuhara difference of an ITS (i.e., $\Delta Y_t=[Y_t^L-Y_{t-1}^L,Y_t^H-Y_{t-1}^H]=[\Delta Y_t^L,\Delta Y_t^H]$). ΔY_t is a stationary ITS though Y_t is not. $\alpha_0,\beta_i\ (i=0,1,2,\cdots,p),\gamma_i\ (i=1,2,\cdots,q)$ are scalar-valued unknown parameters, $I_0=[-\frac{1}{2},\frac{1}{2}]$ is the constant unit interval. u_t is an interval martingale difference sequence with respect to the information set I_{t-1}, satisfying $E(u_t|I_{t-1})=[0,0]$ almost everywhere. $\alpha_0+\beta_0 I_0=[\alpha_0-\frac{1}{2}\beta_0,\alpha_0+\frac{1}{2}\beta_0]$ is the interval intercept. It is observed that ACI models can be used to capture some well-known empirical stylized facts, including mean-reversion. For example, $\beta_1>0$ indicates the mean-reversion that historical returns eventually will revert to the long-run mean.

To be noteworthy, we follow Sun et al.[25] and Yang et al.[20] to allow the low bound can be larger than the high bound in the ITS which is called extended random interval time series (see details in Sun et al.[25]). Specifically, the ACI model exhibits superior performance in capturing the dynamics of ITS by employing the estimation of the D_K distance.

2.4 MLPI-HoltI Models for High-Frequency Component

The high-frequency component contains linear and nonlinear structures. The linear structure is the short-term supply-demand disequilibrium, while the nonlinear structure denotes the market noise. An interest-based hybrid MLPI-HoltI method achieves better results than the individual forecasting of each structure of high-frequency component[39]. We follow Maia and de Carvalho[39] to employ the two-step hybrid system. The linear structure $\boldsymbol{\ell}_t$ is modeled by

the HoltI method, and the nonlinear structure is captured by the MLPI method. The residuals of the HoltI model,

$$\boldsymbol{e}_t = \boldsymbol{I}_t - \widehat{\boldsymbol{\ell}}_t, \tag{7}$$

contain the nonlinear structure of the ITS. After adjusting HoltI model, we model residuals through an MLPI with $2p$ input nodes.

$$\boldsymbol{e}_t = f(\boldsymbol{e}_{t-1}, \boldsymbol{e}_{t-2}, \ldots, \boldsymbol{e}_{t-p}) + \boldsymbol{\epsilon}_t, \tag{8}$$

where f is a nonlinear function determined by the neutral network; $\boldsymbol{\epsilon}_t$ denotes a random vector of errors. After using the MLPI to get the predictions of the nonlinear structure $\widehat{\boldsymbol{e}}_t$, the forecasted high-frequency component provided by the hybrid system is:

$$\widehat{\boldsymbol{I}}_t = \widehat{\boldsymbol{\ell}}_t + \widehat{\boldsymbol{e}}_t. \tag{9}$$

The MLPI-HoltI method does not presume any specific correlation structure in the time series. The forecasting performance could be improved by separately modeling the linear and nonlinear structures in the high-frequency component using tailored models and then integrating the individual forecasts (see more details in Zhang[42]).

3 Benchmark Models

The benchmark models in this study are point-based (GARCH, EAGRCH), range (CARR), and interval-valued (ACI) models. The GARCH model can characterize the long-term memory component. Let y_t^C be the closing price at time t, $Y_t^C = \ln y_t^C$ is the log closing price of the asset at time t, and the log return of the asset from $t-1$ to t is $\Delta Y_t^C = Y_t^C - Y_{t-1}^C$, the models are as follows:

$$\begin{aligned}
&\Delta Y_t^C = \mu_t + \epsilon_t, \\
&\epsilon_t = \sqrt{h_t} z_t, z_t \sim IIDf(\cdot), \\
&h_t = w + \sum_{i=1}^{p} \beta_i h_{t-i} + \sum_{j=1}^{q} \gamma_j \varepsilon_{t-j}^2,
\end{aligned} \tag{10}$$

where $\mu_t = \mu(I_{t-1}, \theta)$ is the conditional mean and $h_t = h(I_{t-1}, \theta)$ is the model for variance $\mathrm{var}(\Delta Y_t | I_{t-1})$ with I_{t-1} being the information set available at time $t-1$. Z_t is residual fluctuations, which is a white noise and IID sequence of standardized innovations with $E(Z_t) = 0, E(Z_t^2) = 1$. z_t is independently identically distributed to the normal distribution, t-distribution, skewed t-distribution (SSTD), generalized error distribution (GED) and others according to data characteristics.

However, the non-negative variance requirement of the GARCH model leads to parameters being non-negative and unconditional variance being smooth. Additionally, the symmetric effect of the disturbance term on the variance limits the GARCH model to capture the asymmetric

characteristics of gasoline markets[43]. Thus, we employ the EGARCH model to mitigate these issues. The EGARCH model is as follows:

$$\begin{aligned}
&\Delta Y_t^C = \mu_t + \epsilon_t,\\
&\epsilon_t = \sqrt{(h_t)} z_t, z_t \sim IIDf(\cdot),\\
&\ln h_t = w + \sum_{i=1}^{p} \beta_i \ln h_{t-i} + \sum_{j=1}^{p} \gamma_j g(z_{t-j}),
\end{aligned} \tag{11}$$

where $g(z_t) = \theta z_t + \lambda[|z_t - E(z_t)]$ is the perturbation function introduced in the EGARCH model. In EGARCH, the range of $\ln h_t$ is the real number set $\mathcal{R}$, which relaxed the parameter constraints of the GARCH model. The positivity and negativity of parameters w, β_i, γ_j is no longer constrained. $g(z_t)$ satisfies $E_{t-1}(g(z_t)) = 0$. To deal with positive and negative effects, the perturbation function $g(z_t)$ can be written as the following subsection function:

$$g(z_t) = \begin{cases} (\theta + \lambda) z_t - \lambda E(z_t), & z_t \geq 0, \\ (\theta - \lambda) z_t - \lambda E(z_t), & z_t < 0, \end{cases} \tag{12}$$

where $g(z_t)$ satisfies $E_{t-1}(g(z_t)) = 0$, $E(z_t) = \sqrt{\pi/2}$. When $\lambda \neq 0$, the influence of disturbance information on price is asymmetric. When $\lambda > 0$, time series data are more sensitive to positive than negative external shocks.

Range data is a more efficient volatility indicator proved by Parkinson[17] and Chou[19]. Range data contains more information than the closing price. For example, Brandt and Jones[44] found that the standard deviation of the range is merely about 25% of the standard deviation of absolute return. Chou[19] also showed that the CARR model has better forecast performance than GARCH-type models. The CARR model is employed as another benchmark model in our study.

Let $Y_t^r = Y_t^H - Y_t^L$ be the observed range data in time period t, where Y_t^L, Y_t^H denote the lowest and highest log-price in time period t respectively. CARR model is described as:

$$\begin{aligned}
&Y_t^r = \lambda_t \varepsilon_t,\\
&\lambda_t = \omega + \sum_{i=1}^{q} \alpha_i Y_{r,t-i} + \sum_{j=1}^{p} \beta_j \lambda_{t-j},\\
&\varepsilon_t | I_{t-1} \sim f(\cdot),
\end{aligned} \tag{13}$$

where λ_t is the conditional mean $E(Y_t^H | I_{t-1})$ of Y_t^r based on information set I_{t-1}, ε_t is an IID standardized nonnegative disturbance with $E(\varepsilon_t) = 1$. The parameters $(\omega, \alpha_i, \beta_j)$ of the conditional mean equation are all positive to ensure the conditional mean is positive. The distribution of innovation ε_t is usually set to be the standard exponential distribution.

The last benchmark is the ACI model, that is, we employ the ACI model to forecast interval range and then generate volatility forecasts; See more discussions in Section 2.

4 Preliminary Analysis

4.1 Calculations of Volatility

We employ four volatility proxies from the literature to estimate weekly volatility: Range (RNG), absolute return (ARET), squared return (SRET), and realized variation (RV). The calculations are followings:

$$\mathrm{RNG}_t = Y_t^H - Y_t^L, \tag{14}$$

where Y_t^H, Y_t^L are the logarithms of weekly high and low gasoline prices.

$$\mathrm{ARET}_t = |Y_t^C - Y_{t-1}^C|, \tag{15}$$

where Y_t^C is the logarithms of the weekly closing price.

$$\mathrm{SRET}_t = (Y_t^C - Y_{t-1}^C)^2, \tag{16}$$

$$\mathrm{RV}_t = \sum_{i=0}^{ti} (Y_{ti}^C - Y_{t(i-1)}^C)^2, \tag{17}$$

where ti is the ith day within t week. Due to the large missing gasoline minute frequency data in the original data, we follow Chou [19] to use daily data to calculate the weekly realized variation instead. Y_{ti}^C is the ith daily closing price with each day t.

In this paper, FV denotes forecasted volatility and MV denotes actual volatility (calculated from Equations (14)∼(17)). To convert forecasting results in range form to volatility in quadratic form, Chou [19] used the conversion relationship established by Parkinson [17],

$$h_t = \frac{\lambda_t^2}{4 \ln 2}. \tag{18}$$

Equation (18) describes the relationship between a range forecast and a variance forecast when innovations z_t of the GARCH model is IID $N(0,1)$. This range volatility estimator is theoretically shown by Parkinson [17] to be 2.5 to 5 times more efficient than the classical estimator based on closing prices. When z_t is IID but not $N(0,1)$, we follow Zhu et al. [45] to forecast volatility as following:

$$\begin{cases} \rho = \dfrac{Sd^2}{(\varepsilon_{\max} - \varepsilon_{\min})^2}, \\ \mathrm{FV}_t = \rho \times (\hat{Y}_t^r)^2, \end{cases} \tag{19}$$

where Sd, $\varepsilon_{\max}, \varepsilon_{\min}$ are standard deviation, the maximum and minimum from the innovation of a CARR model, $\hat{Y}_t^r$ is the out-of-sample forecast results for range.

To compare the forecasting accuracy of different models, we follow Zhu et al. [45] to convert forecasting results to volatility. For GARCH family models, the predicted volatility σ is compared with RNG and ARET, and the conditional variance σ^2 is compared with SRET and RV. The interval ranges forecasted by CARR, ACI, and IHDE are also compared with RNG and

ARET. For forecasting volatility in the quadratic power form, we use the prediction results of the CARR model to calculate the transformation coefficient ρ of the formula, and the volatility in the quadratic power form is calculated according to the transformation formula, which is shown in Equation (19).

4.2 Data Discriptions

The data analyzed in this paper are weekly interval-valued Reformulated Blendstock for Oxygenate Blending (RBOB) gasoline futures prices Y_t. For each week t, the lower and upper bounds of the interval-valued gasoline price observation Y_t are formed using the low and high daily gasoline future prices within that week. The samples span from January 1, 2007, to December 31, 2023, and are converted to the logarithmic scale.

Figure 2 shows a series of charts showing the logarithm of high price Y_t^H, low price Y_t^L, midpoint Y_t^M, and range RNG_t. The Y_t^H, Y_t^L and Y_t^M measure the gasoline price trends, while the RNG_t quantifies the gasoline price fluctuations. Figure 2 shows that the gasoline market prices fluctuate violently, especially when the range rises while the gasoline prices fall, indicating a negative correlation between the trends and volatility contained in the range gasoline price data. Meanwhile, studies have shown that interval-valued gasoline prices may have the same range but different price levels in different periods, and similar price levels that occur in different periods may be accompanied by completely different ranges, indicating that interval-valued prices contain more information gain than point value price data.

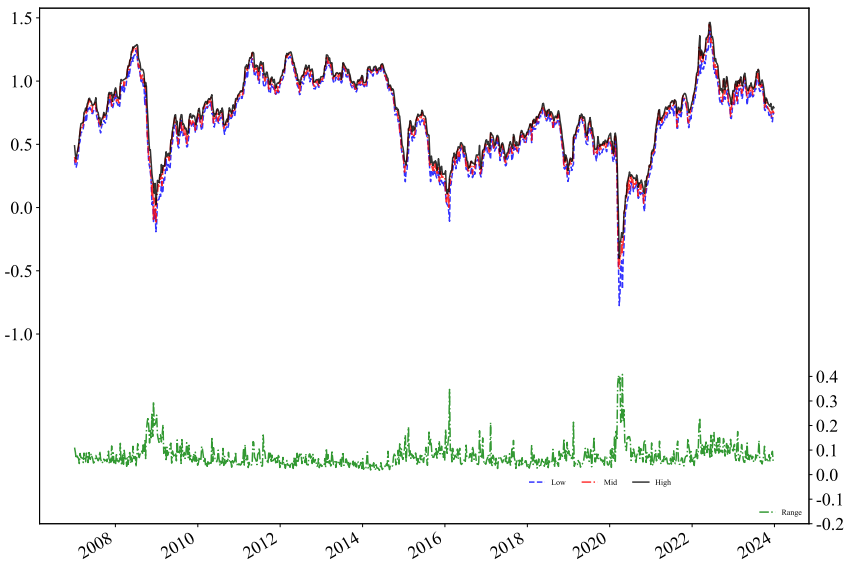

Figure 2 The weekly interval-valued gasoline price from Jan. 1, 2007, to Dec. 31, 2023

Table 1 shows the descriptive statistical results of each indicator. Among them, Y_t^L is the lowest price, Y_t^H is the highest price, and Y_t^M is the midpoint ($Y_t^M = (Y_t^L + Y_t^H)/2$). RNG, ARET, SRET, and RV are the volatility indicators. On the one hand, compared to the Y_t^H and Y_t^M, the standard deviation and range of Y_t^L are larger, indicating a greater

Table 1 Descriptive stastics

	Y^L	Y^H	Y^M	RNG	ARET	SRET	RV
Mean	0.679	0.755	0.717	0.076	0.040	0.003	0.003
Sd	0.324	0.304	0.314	0.046	0.040	0.010	0.007
Median	0.703	0.767	0.731	0.065	0.029	0.001	0.002
Min	−0.775	−0.402	−0.589	0.017	0.000	0.000	0.000
Max	1.399	1.465	1.432	0.408	0.435	0.189	0.098
Skew	−0.683	−0.402	−0.540	2.944	3.216	11.781	8.910
Kurtosis	0.693	−0.168	0.205	14.031	19.439	186.103	97.493

Note: The table presents the descriptive statistics for a sample of weekly gasoline price and volatility from Jan. 1, 2007, to Dec. 31, 2023. Y^L, Y^H, Y^M, and RNG are the low, high, midpoint, range of gasoline log prices with a week. Sd is the standard deviation.

degree of dispersion. Meanwhile, Y_t^L, Y_t^H, and Y_t^M all exhibit a negatively skewed distribution, with Y_t^L and Y_t^M exhibiting a peak distribution and Y_t^H exhibiting a low peak distribution, indicating that gasoline prices are concentrated at higher levels within the sample interval, but the distribution of the highest prices is flatter. On the other hand, RNG and ARET have higher mean values and greater dispersion compared to SRET and RV. All four indicators exhibit a positively skewed distribution and a peak distribution, indicating that the volatility of gasoline prices is concentrated at a relatively low level within the sample interval.

5 Empirical Results

5.1 Forecast Criteria

We use root mean square error (RMSE) and mean absolute error (MAE) to evaluate the prediction accuracy of different prediction methods:

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\text{MV}_t - \text{FV}_t)^2}, \tag{20}$$

$$\text{MAE} = \frac{1}{T}\sum_{t=1}^{T}|\text{MV}_t - \text{FV}_t|, \tag{21}$$

where MV_t and FV_t denote actual volatility and forecasted volatility.

5.2 Out-of-Sample Forecasting Results

This subsection demonstrates the interval forecasting performance of our IHDE learning approach by using interval-valued gasoline prices. We compare the forecasting performance of the proposed IHDE approach with the benchmarks (i.e. GARCH, EGARCH, CARR, ACI). We select RBOB Gasoline Continuous Contract price as the experimental datasets of ITS from NYEX markets, downloaded from tradingview.com. Our sample datasets consist of 886 interval-valued data of spot price spinning from Jan. 1, 2007, to Dec. 31, 2023. We choose 70% of the sample as the training set and 20% as the test sets.

We decompose gasoline prices into high- and low-frequency components based on MVMD. Figure 3 shows the MVMD results of gasoline prices with different frequencies from Jan. 1, 2007, to Dec. 31, 2023. The low-frequency component shows the fundamental factors of gasoline price fluctuations. For example, during the initial outbreak period of COVID-19 in early 2020, many countries implemented lockdown policies, leading to decreased travel demand and consequently diminished gasoline consumption. Thus, the low-frequency component of gasoline price reached its lowest on April 16, 2020. In the first half of 2022, the oil production increase plan formulated by OPEC+ failed to materialize, leading to a tightening of the oil supply. At the same time, the Russian-Ukrainian conflict has increased the geopolitical risks in the crude oil market, fuelling supply constraints. As a result, the low-frequency component reached its highest in the last week of June 2022. The high-frequency component is more volatile than the low-frequency component, presenting the market noise and short-term supply-demand disequilibrium.

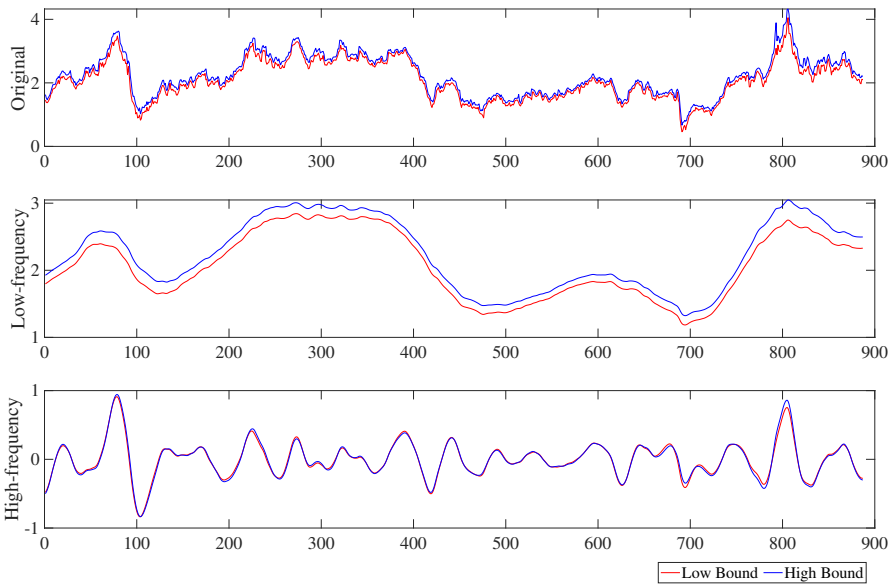

Figure 3 The MVMD results of weekly interval-valued gasoline price from Jan. 1, 2007, to Dec. 31, 2023

Then, we employ a hybrid MLPI-HoltI to forecast interval-valued high-frequency components. The HoltI model separates the linear structure of the high-frequency component, capturing the short-term supply-demand disequilibrium. The MLPI models the nonlinear structure, such as unobserved market behaviors. We optimize smoothing parameter matrices A and B within the range of (0, 1) by minimizing the interval sum of squared errors of one-step ahead forecast and solve the optimization problem by the limited memory Broyden-Fletcher-Goldfarb-Shanno algorithm method for bound constrained optimization (L-BFGS-B) which is provided in the R package 'optimx'. The best number of inputs and hidden nodes selected by the simulation approach is used for the MLPI method. A set of neural networks with different numbers of inputs and hidden nodes are trained by trial and error testing, and the best combination of

the two values is determined by the sum of the squared errors of the fifty iterations. We use fifty iterations of the conjugate gradient algorithm, where the initial weights are applied randomly to mitigate the issue of local minima[39]. Third, we employ ACI models to forecast the low-frequency component and ensemble the forecasted interval-valued high- and low-frequency components to produce the forecasted volatility of gasoline. Finally, we compare the forecasting performance through RMSE and MAE measures by Equation (20) and Equation (21).

Panel A of Table 2 reports the two forecasting criteria of one-step-ahead forecasts for the weekly volatility of gasoline. Meanwhile, we employ the DM test to assess the performance difference between different models. Table 3 reports the results. The IHDE performs most accurately in all cases, indicating the gain of using information contained in ITS data and nonlinearity for volatility forecasting.

Table 2 Forecast performance

Volatility	RNG		ARET		SRET		RV	
Criteria	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Panel A: Forecast performance criteria during 01/01/2007–31/12/2023. Out-of-sample size: 20% of the sample								
GARCH	89.3561	82.5442	55.6904	42.4364	6.4376	3.2725	4.3690	3.0284
EGARCH	88.4501	81.5626	54.9467	41.5200	6.4351	3.2679	4.3656	3.0235
CARR	81.5666	74.0065	49.5286	35.2443	6.3845	3.1837	4.2959	2.9211
ACI	79.9486	72.8677	48.5781	34.3542	6.3609	3.1547	4.2421	2.8807
IHDE	**78.6719**	**71.5115**	**47.3018**	**33.3396**	**6.3356**	**3.1324**	**4.2262**	**2.8510**
Panel B: Forecast performance criteria during 01/01/2007–31/12/2023. Out-of-sample size: 30% of the sample								
GARCH	106.4098	89.7434	67.4898	46.3903	15.7373	4.6994	12.3716	4.6679
EGARCH	102.5607	85.1442	64.4159	42.2855	15.7267	4.6682	12.3582	4.6324
CARR	99.8069	81.8016	62.3028	39.5892	15.7113	4.6261	12.3387	4.5802
ACI	94.6805	78.7590	59.1790	37.4119	15.6061	4.5576	12.1143	4.4636
IHDE	**91.9999**	**75.6916**	**57.2408**	**35.7953**	**15.5329**	**4.4969**	**12.0024**	**4.3478**
Panel C: Forecast performance criteria during 01/01/1998–31/12/2023. Out-of-sample size: 20% of the sample								
GARCH	104.0476	87.2412	65.6006	44.1998	15.5917	4.6400	12.2561	4.6039
EGARCH	102.6132	85.5253	64.4615	42.7148	15.5870	4.6261	12.2501	4.5878
CARR	99.6001	81.9036	62.1342	39.7496	15.5712	4.5826	12.2301	4.5347
ACI	95.1044	79.3749	59.3934	37.8942	15.4806	4.5254	12.0365	4.4393
IHDE	**92.1221**	**76.2943**	**57.1609**	**36.0632**	**15.3867**	**4.4554**	**11.8936**	**4.3223**

Note: Unit: 10^{-3}.

First, comparing the forecasting accuracy between the traditional GARCH, EGARCH, range models and the interval-valued forecasting methods (ACI, IHDE), we can see that, whatever accuracy measure is employed, the interval-valued models outperform the traditional forecasting

Table 3 DM test results for out-of-sample forecasting using a 20% sample from Jan. 1, 2007, to Dec. 31, 2023

Model	EGARCH	CARR	ACI	IHDE
Panel A: RNG				
GARCH	9.4422e+12 (0.000)	144.01 (0.000)	32.60 (0.000)	42.58 (0.000)
EGARCH		127.46 (0.000)	29.29 (0.000)	38.79 (0.000)
CARR			3.70 (0.000)	8.93 (0.000)
ACI				5.20 (0.000)
Panel B: ARET				
GARCH	37.02 (0.000)	23.07 (0.000)	18.75 (0.000)	19.77 (0.000)
EGARCH		21.35 (0.000)	17.13 (0.000)	18.34 (0.000)
CARR			2.92 (0.004)	6.48 (0.000)
ACI				3.87 (0.000)
Panel C: SRET				
GARCH	36.53 (0.000)	20.21 (0.000)	11.88 (0.000)	14.93 (0.000)
EGARCH		19.54 (0.000)	11.47 (0.000)	14.52 (0.000)
CARR			3.31 (0.001)	6.72 (0.000)
ACI				3.12 (0.002)
Panel D: RV				
GARCH	1.1682e+13 (0.000)	72.25 (0.000)	17.38 (0.000)	27.23 (0.000)
EGARCH		68.94 (0.000)	16.80 (0.000)	26.47 (0.000)
CARR			4.65 (0.000)	10.07 (0.000)
ACI				4.20 (0.000)

Note: The corresponding p-value is given in parentheses.

methods. Compared with benchmark models, the HIDE approach has a maximum prediction accuracy improvement rate of 27.34%. This is because of the information gained from interval-valued data and the ability to capture the characteristics of different frequency components. The interval-valued gasoline price contains more information (i.e., price trends, and price variations) than the point-valued series. For example, the last week of Dec. 2013 and the second week of Dec. 2019 have the same range ($ 0.705 per gal), while the midpoint prices are $ 2.81 per gal and $ 1.65 per gal, respectively. Thus, the interval-valued models perform better in volatility forecasting with the information gained from the intervals, which is consistent with Sun et al.[25].

Second, the HIDE approach has a better forecast performance than the ACI model. The HIDE approach captures the features of different frequency components of gasoline volatility by using interval-valued prices. Thus, the HIDE model has improved prediction accuracy by 0.39%∼6.20% compared with the ACI model, indicating the importance of the "divide and conquer" principle in gasoline volatility forecasting.

Third, it is worth highlighting that nonlinear models consistently outperform linear counterparts, such as the EGARCH model outperforms the GARCH model, and the IHDE model outperforms the ACI model. This happens because the gasoline market has nonlinear characteristics. In particular, the MVMD method employed in the proposed IHDE approach takes advantage of "divide and conquer" and provides a new way to forecast volatility according to different frequency components. The MLPI method employed in this study captures the nonlinear characteristics of ITS data.

Furthermore, interval-valued models such as the ACI and IHDE outperform the CARR model. The range data modeled by the CARR loses the price trends' information contained in intervals due to the range calculations. While the interval-valued models can fully utilize the price trends and variations within an interval. It supports the idea that combining price trends and uncertainty information is important for volatility forecasting.

To test the robustness of forecast evaluation, we change the out-of-sample and whole sample sizes, respectively. In the first robustness test, we choose 30% samples as test sets. In the second

Table 4 DM test results for out-of-sample forecasting using a 30% sample from Jan. 1, 2007, to Dec. 31, 2023

Model	EGARCH	CARR	ACI	IHDE
Panel A: RNG				
GARCH	7.3071e+08 (0.000)	1286.63 (0.000)	25.08 (0.000)	24.80 (0.000)
EGARCH		541.52 (0.000)	14.58 (0.000)	16.68 (0.000)
CARR			6.94 (0.000)	10.77 (0.000)
ACI				7.50 (0.000)
Panel B: ARET				
GARCH	36.89 (0.000)	30.83 (0.000)	18.76 (0.000)	15.97 (0.000)
EGARCH		22.93 (0.000)	11.32 (0.000)	10.60 (0.000)
CARR			5.46 (0.000)	6.72 (0.000)
ACI				3.98 (0.000)
Panel C: SRET				
GARCH	32.84 (0.000)	26.91 (0.000)	6.81 (0.000)	6.27 (0.000)
EGARCH		22.69 (0.000)	5.36 (0.000)	5.34 (0.000)
CARR			3.36 (0.001)	4.08 (0.000)
ACI				3.03 (0.003)
Panel D: RV				
GARCH	1.8092e+09 (0.000)	404.10 (0.000)	9.18 (0.000)	9.97 (0.000)
EGARCH		240.51 (0.000)	7.59 (0.000)	8.86 (0.000)
CARR			5.24 (0.000)	7.24 (0.000)
ACI				6.02 (0.000)

Note: The corresponding p-value is given in parentheses.

robustness test, we expand the sample period from Jan. 1, 1998, to Dec. 31, 2023, and choose 20% of the sample as test sets. The results in Panel B and Panel C of Table 2, Table 4, and Table 5 are also robust, supporting that IHDE outperforms all benchmarks.

Table 5 DM test results for out-of-sample forecasting using a 20% sample from Jan. 1, 1998, to Dec. 31, 2023

Model	EGARCH	CARR	ACI	IHDE
Panel A: RNG				
GARCH	1.4507e+15 (0.000)	234.84 (0.000)	19.54 (0.000)	20.29 (0.000)
EGARCH		159.35 (0.000)	15.28 (0.000)	17.11 (0.000)
CARR			6.30 (0.000)	10.38 (0.000)
ACI				8.34 (0.000)
Panel B: ARET				
GARCH	28.89 (0.000)	26.06 (0.000)	15.21 (0.000)	13.70 (0.000)
EGARCH		23.97 (0.000)	12.26 (0.000)	11.65 (0.000)
CARR			5.15 (0.000)	7.03 (0.000)
ACI				4.95 (0.000)
Panel C: SRET				
GARCH	28.80 (0.000)	25.17 (0.000)	6.53 (0.000)	6.12 (0.000)
EGARCH		23.54 (0.000)	5.77 (0.000)	5.68 (0.000)
CARR			3.33 (0.001)	4.28 (0.000)
ACI				3.77 (0.000)
Panel D: RV				
GARCH	2.3073e+14 (0.000)	163.52 (0.000)	8.62 (0.000)	9.07 (0.000)
EGARCH		125.64 (0.000)	7.78 (0.000)	8.55 (0.000)
CARR			5.00 (0.000)	6.84 (0.000)
ACI				6.32 (0.000)

Note: The corresponding p-value is given in parentheses.

In summary, the above empirical results have yielded several key findings. Firstly, the proposed IHDE approach significantly outperforms all benchmarks regarding forecasting accuracy and the priority of models, indicating that the "divide and conquer" can effectively capture the volatility characteristics of different frequency components. Secondly, the nonlinear models outperform corresponding linear models. Specifically, the nonlinearity in the high-frequency component is more important. Thirdly, the interval-valued model outperforms the point-valued model, supporting that the informational advantage improves the volatility forecasting ability.

6 Conclusion

This study proposes a novel IHDE volatility forecasting approach for gasoline volatility using ITS data. In this approach, we employ MVMD to decompose and ensemble the complex gasoline price into low- and high-frequency components. To capture the volatility's fundamental

dynamics of long-term trends, the ACI model is employed to forecast low-frequency component. To capture the market noise and short-term demand-supply disequilibrium impacts on gasoline volatility, the MLPI-HoltI method is applied to forecast high-frequency component. The forecasted low- and high-frequency components are ensemble to generate the forecasted volatility. The empirical analysis demonstrated that our IHDE significantly outperforms all benchmarks with different sample sizes and forecasting periods. Our study emphasizes the importance of different components' characteristics and informational gain of interval-valued prices for volatility forecasting.

The empirical results of this study are critical for the government to implement policies to manage inflation and to help investors manage risks. Our approach can also be extended to stock markets, crude oil markets, exchange rate markets, etc. In future work, we can integrate related factors in our approach, such as structural breaks in energy markets, economic indexes, and policy uncertainties. The asymmetric characteristic of gasoline volatility also remains to be studied.

References

[1] Kilian L, Zhou X. The impact of rising oil prices on US inflation and inflation expectations in 2020–2023. Energy Economics, 2022, 113: 106228.

[2] Baumeister C, Kilian L. Forty years of oil price fluctuations: Why the price of oil may still surprise us. Journal of Economic Perspectives, 2016, 30(1): 139–160.

[3] Kilian L, Zhou X. Heterogeneity in the pass-through from oil to gasoline prices: A new instrument for estimating the price elasticity of gasoline demand. Journal of Public Economics, 2024, 232: 105099.

[4] Yang H, Tian X, Jin X, et al. A novel hybrid model for gasoline prices forecasting based on Lasso and CNN. Journal of Social Computing, 2022, 3(3): 206–218.

[5] Bollerslev T. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 1986, 31(3): 307–327.

[6] Nelson D B. Conditional heteroskedasticity in asset returns: A new approach. Econometrica: Journal of the Econometric Society, 1991: 347–370.

[7] Segnon M, Gupta R, Wilfling B. Forecasting stock market volatility with regime-switching GARCH-MIDAS: The role of geopolitical risks. International Journal of Forecasting, 2024, 40(1): 29–43.

[8] Shi Y. Modeling and forecasting volatilities of financial assets with an asymmetric zero-drift GARCH model. Journal of Financial Econometrics, 2023, 21(4): 1308–1345.

[9] Song Y, Tang X, Wang H, et al. Volatility forecasting for stock market incorporating macroeconomic variables based on GARCH-MIDAS and deep learning models. Journal of Forecasting, 2023, 42(1): 51–59.

[10] Afuecheta E, Okorie I E, Nadarajah S, et al. Forecasting value at risk and expected shortfall of foreign exchange rate volatility of major african currencies via GARCH and dynamic conditional correlation analysis. Computational Economics, 2024, 63(1): 271–304.

[11] You Y, Liu X. Forecasting short-run exchange rate volatility with monetary fundamentals: A GARCH-MIDAS approach. Journal of Banking & Finance, 2020, 116: 105849.

[12] Tian S, Hamori S. Modeling interest rate volatility: A realized GARCH approach. Journal of Banking & Finance, 2015, 61: 158–171.

[13] Aras S. Stacking hybrid GARCH models for forecasting Bitcoin volatility. Expert Systems with Applications, 2021, 174: 114747.

[14] Zhang C, Ma H, Arkorful G B, et al. The impacts of futures trading on volatility and volatility asymmetry of Bitcoin returns. International Review of Financial Analysis, 2023, 86: 102497.

[15] Zhang Y J, Zhang H. Volatility forecasting of crude oil market: Which structural change based GARCH models have better performance? The Energy Journal, 2023, 44(1): 175–194.

[16] Lin Y, Xiao Y, Li F. Forecasting crude oil price volatility via a HM-EGARCH model. Energy Economics, 2020, 87: 104693.

[17] Parkinson M. The extreme value method for estimating the variance of the rate of return. Journal of Business, 1980: 61–65.

[18] Alizadeh S, Brandt M W, Diebold F X. Range-based estimation of stochastic volatility models. The Journal of Finance, 2002, 57(3): 1047–1091.

[19] Chou R Y. Forecasting financial volatilities with extreme values: The conditional autoregressive range (CARR) model. Journal of Money, Credit and Banking, 2005: 561–582.

[20] Yang W, Han A, Hong Y, et al. Analysis of crisis impact on crude oil prices: A new approach with interval time series modelling. Quantitative Finance, 2016, 16(12): 1917–1928.

[21] Neto E A L, De Carvalho F D A T. Centre and range method for fitting a linear regression model to symbolic interval data. Computational Statistics & Data Analysis, 2008, 52(3): 1500–1515.

[22] Han A, Hong Y, Lai K K, et al. Interval time series analysis with an application to the sterling-dollar exchange rate. Journal of Systems Science and Complexity, 2008, 21(4): 558–573.

[23] Teles P, Brito P. Modeling interval time series with space-time processes. Communications in Statistics-Theory and Methods, 2015, 44(17): 3599–3627.

[24] Lin W, González-Rivera G. Interval-valued time series models: Estimation based on order statistics exploring the agriculture marketing service data. Computational Statistics & Data Analysis, 2016, 100: 694–711.

[25] Sun Y, Han A, Hong Y, et al. Threshold autoregressive models for interval-valued time series data. Journal of Econometrics, 2018, 206(2): 414–446.

[26] Qiao K, Liu Z, Huang B, et al. Brexit and its impact on the US stock market. Journal of Systems Science and Complexity, 2021, 34(3): 1044–1062.

[27] Mate C, Jimeńez L. Forecasting exchange rates with the iMLP: New empirical insight on one multilayer perceptron for interval time series (ITS). Engineering Applications of Artificial Intelligence, 2021, 104: 104358.

[28] Lu Q, Sun Y, Hong Y, et al. Forecasting interval-valued crude oil prices using asymmetric interval models. Quantitative Finance, 2022, 22(11): 2047–2061.

[29] Sun Y, Zhang X, Wan A T K, et al. Model averaging for interval-valued data. European Journal of Operational Research, 2022, 301(2): 772–784.

[30] Zhong W, Qian C, Liu W, et al. Feature screening for interval-valued response with application to study association between posted salary and required skills. Journal of the American Statistical Association, 2023, 118(542): 805–817.

[31] Shen T, Tao Z, Chen H. Exploring long-memory process in the prediction of interval-valued financial time series and its application. Journal of Systems Science and Complexity, 2024, 37(2): 759–775.

[32] He Y, Han A, Hong Y, et al. Forecasting crude oil price intervals and return volatility via autoregressive conditional interval models. Econometric Reviews, 2021, 40(6): 584–606.

[33] González-Rivera G, Lin W. Constrained regression for interval-valued data. Journal of Business & Economic Statistics, 2013, 31(4): 473–490.

[34] Sun S, Sun Y, Wang S, et al. Interval decomposition ensemble approach for crude oil price forecasting. Energy Economics, 2018, 76: 274–287.

[35] Yang D, Guo J, Sun S, et al. An interval decomposition-ensemble approach with data-characteristic-driven reconstruction for short-term load forecasting. Applied Energy, 2022, 306: 117992.

[36] Zhu B, Wan C, Wang P. Interval forecasting of carbon price: A novel multiscale ensemble forecasting approach. Energy Economics, 2022, 115: 106361.

[37] Zheng L, Sun Y, Wang S. A novel interval-based hybrid framework for crude oil price forecasting and trading. Energy Economics, 2024, 130: 107266.

[38] ur Rehman N, Aftab H. Multivariate variational mode decomposition. IEEE Transactions on Signal Processing, 2019, 67(23): 6039–6052.

[39] Maia A L S, de Carvalho F A T. Holt's exponential smoothing and neural network models for forecasting interval-valued time series. International Journal of Forecasting, 2011, 27(3): 740–759.

[40] Engle R F, Ghysels E, Sohn B. Stock market volatility and macroeconomic fundamentals. Review of Economics and Statistics, 2013, 95(3): 776–797.

[41] Han A, Hong Y, Wang S, et al. A vector autoregressive moving average model for interval-valued time series data. Essays in honor of Aman Ullah. Emerald Group Publishing Limited, 2016: 417–460.

[42] Zhang G P. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 2003, 50: 159–175.

[43] Kang S H, Yoon S M. Modeling and forecasting the volatility of petroleum futures prices. Energy Economics, 2013, 36: 354–362.

[44] Brandt M W, Jones C S. Volatility forecasting with range-based EGARCH models. Journal of Business & Economic Statistics, 2006, 24(4): 470–486.

[45] Zhu M, Hong Y, Wang S. Can interval data improve volatility forecasts? Evidence from foreign exchange markets. Evidence From Foreign Exchange Markets, 2024. Available at SSRN: https://ssrn.com/abstract=4785170.

A Appendix

A.1 Estimation for ACI Models

Han et al.[41] proposed the estimation methods for ITS by minimizing squared D_K distance. The D_K distance method can be used to estimate unknown parameters $\phi = (\alpha_0, \beta_i, \gamma_i)'$ in Equation (6) for minimizing squared D_K distance between the fitted interval value and observed interval value:

$$\hat{Q}_T(\phi) = \sum_{t=1}^{T} D_K^2[\Delta Y_t, \Delta \hat{Y}_t(\phi)], \tag{A1}$$

where $\Delta \hat{Y}_t(\phi)$ is the fitted value of $\Delta Y_t(\phi)$ based on Equation (6). The squared D_K distance between ΔY_t and $\Delta \hat{Y}_t$ is defined as:

$$D_K^2[\Delta Y_t, \Delta \hat{Y}_t] = \begin{pmatrix} \Delta Y_t^U - \Delta \hat{Y}_t^U \\ -(\Delta Y_t^L - \Delta \hat{Y}_t^L) \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} K_{LL} & K_{LU} \\ K_{UL} & K_{UU} \end{pmatrix} \begin{pmatrix} \Delta Y_t^U - \Delta \hat{Y}_t^U \\ -(\Delta Y_t^L - \Delta \hat{Y}_t^L) \end{pmatrix}, \tag{A2}$$

where $K = \begin{pmatrix} K_{LL} & K_{LU} \\ K_{UL} & K_{UU} \end{pmatrix}$ is a positive definite matrix. The parameters $\phi = (\alpha_0, \beta_i, \gamma_i)'$ can be estimated as follows:

$$\hat{\phi} = \arg\min_{\phi \in \Phi} T^{-1} \sum_{t=1}^{T} D_K^2[\Delta Y_t, \Delta \hat{Y}_t(\phi)]. \tag{A3}$$

The D_K metric considers the set of absolute differences between all possible pairs of points (range and interior points) in random extended intervals A and B with a proper weighting function implied by matrix K. Thus, the ACI model utilizes the interval information more efficiently than conventional point-valued time series models[32].