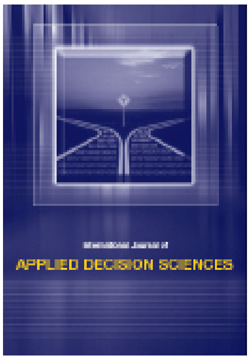


International Journal of Applied Decision Sciences

ISSN online: 1755-8085 - ISSN print: 1755-8077
https://www.inderscience.com/ijads

Does the optimal model always perform the best? A combined approach for interval forecasting

Zhe Zhang, Wei Chong Choo, Jayanthi Arasan

DOI: 10.1504/IJADS.2025.10058959

Article History:

Received:	28 May 2023
Last revised:	28 July 2023
Accepted:	28 July 2023
Published online:	03 December 2024

Does the optimal model always perform the best? A combined approach for interval forecasting

Zhe Zhang

School of Business and Economics,
Universiti Putra Malaysia,
Serdang, Selangor, Malaysia
Email: gs60278@student.upm.edu.my

Wei Chong Choo*

School of Business and Economics,
Institute for Mathematical Research,
Universiti Putra Malaysia,
Serdang, Selangor, Malaysia
Email: wcchoo@upm.edu.my
*Corresponding author

Jayanthi Arasan

Department of Mathematics and Statistics,
Faculty of Science,
Universiti Putra Malaysia,
Serdang, Selangor, Malaysia
Email: jayanthi@upm.edu.my

Abstract: Interval forecasting is widely applied by decision makers for it can provide more comprehensive information. In the literature, GARCH models under different distributional assumptions are applied and evaluated to find the optimal interval forecasting model for the experimental data. However, the optimal model selected based on sample data from a specific period may not always perform the best in future periods. Therefore, this study employs GARCH models based on different distributional assumptions for interval forecasting of the daily return data of the Nasdaq Composite Index. The results show that the forecasting performance of some models exhibits significant differences across different periods. To address this issue, this study proposes a Monte Carlo-based non-parametric interval forecasting combination method. The results demonstrate that this method can effectively avoid the risk of forecasting inaccuracies caused by relying on a single model.

Keywords: interval forecasting; optimal model; combined approach; GARCH model; distribution assumptions; Monte Carlo.

Reference to this paper should be made as follows: Zhang, Z., Choo, W.C. and Arasan, J. (2025) 'Does the optimal model always perform the best? A combined approach for interval forecasting', *Int. J. Applied Decision Sciences*, Vol. 18, No. 1, pp.64–83.

Biographical notes: Zhe Zhang received his Bachelor's degree in Economics from Beijing Technology and Business University. He completed his Master's degree from Macau University of Science and Technology in the field of Public Administration. Currently, he is pursuing his PhD at the Universiti Putra Malaysia. His research interests include times series forecasting and financial decision science.

Wei Chong Choo finished his BSc from Universiti Putra Malaysia. He received his MSc from Universiti Putra Malaysia. He completed his PhD in the University of Oxford. Currently, he is an Associate Professor and Director of Institute for Mathematical Research (INSPEM) at Universiti Putra Malaysia. His research field is time series analysis and forecasting.

Jayanthi Arasan finished her BSc (Hons) from Universiti Sains Malaysia. She received her MSc in Applied Statistics from Universiti Putra Malaysia. She completed her PhD in the University of Oxford. Currently, she is an Associate Professor at Universiti Putra Malaysia. Her research field is independent analysis.

1 Introduction

Interval forecasting refers to a method of forecasting that provides not only a single point estimate but also a range of possible values within which the true outcome is expected to fall. Interval forecasting provides decision makers with a more informative tool for managing uncertainty and making better-informed decisions. Interval forecasting has the advantage that it explicitly quantifies the uncertainty by providing a range of possible outcomes. Decision makers can have a better understanding of the potential variability in future events, allowing them to make more informed decisions. Moreover, interval forecasting also helps decision makers assess and manage risks. By considering the entire range of possible outcomes, decision makers can identify potential worst-case scenarios and take proactive measures to mitigate risks. Market participants are continually interested in increasing the accuracy of interval forecasting since forecasting accuracy significantly impacts the effectiveness of their decisions. Over the past few decades, numerous studies have focused on finding the optimal model to increase the accuracy of interval forecasting. However, it is unlikely that a single model will perform well on all data in all periods. In fact, for decision makers, relying on a single 'optimal' model can be risky. Decision makers who attempt to find the optimal model believe that there exists a perfect model. However, such a model may not actually exist. In practice, forecasting models are built based on certain assumptions, methodologies, and training data. Depending on the complexity of the problem and the available data, different models may yield different results. By relying on a single model, decision makers are exposed to the risk that the model may not accurately capture the true nature of the data, leading to forecasting failure. Real-world data is also subject to variability and uncertainty. A single model may perform well on certain subsets of data but may fail to generalise to new data. By relying on a single model, decision makers run the risk of making decisions based on a limited understanding of the data's variability. To mitigate these risks, one of the choices is to use ensemble methods, which combine forecasts from multiple models.

Combination refers to the use of multiple forecasting methods or models to reduce uncertainty, improve robustness, and provide a more comprehensive perspective for decision making. The idea is that different methods have their strengths and weaknesses, and by combining them, a more robust and accurate forecasting tool can be created. While the performance of the combined model may not be as good as a single model, the overall performance tends to be more robust and accurate. There are several benefits of using combination forecasting. Firstly, it reduces the reliance on any single model, which can be affected by specific data patterns and biases. Secondly, it can capture a wider range of potential future scenarios by incorporating the strengths and weaknesses of different models. Thirdly, it can improve the accuracy and reliability of the forecast by reducing the impact of outliers or extreme values in the data.

In interval forecasting, a straightforward method achieve combination is to combine the lower and upper bounds, respectively; however, this approach does not guarantee the combined interval with the correct probability (Timmermann, 2006). To address this problem, some studies combine the forecast intervals using the combined density (Wallis, 2005). To address this problem, some studies combine the forecast intervals using the combined density (Wallis, 2005). To implement the combination of the density function, the density functions of individual forecasts are obtained firstly. Then, these density functions can be combined to generate forecasting intervals for any required probability. However, the practical implementation of combining density functions from different distributions can be challenging. As a result, most studies in the literature assume a normal distribution for the individual models, as it is easier to combine the density functions of normal distributions.

In this study, the Nasdaq Composite daily data is used for the comparison of the accuracy of interval forecasting. The distributional assumption for individual models is assumed to follow different distribution, include normal (N), student-t (T), skewed t (ST), GED, and skewed GED (SGED). The variances are forecast using the GARCH model. This study firstly perform a comparison study of interval forecasting under different distributional assumption over the period, that revealing the challenges and risks involved in the pursuit of an optimal model for interval forecasting. And next, this study introduces a non-parametric combination method of different distributions that fills the gaps in related research. This study also discuss the potential directions for future research.

2 Literature review

Wang et al. (2017) propose that the stock market is an important part of the capital market, which plays a significant role in optimising capital allocation, financing and other areas. The stock price is also of great concern to investment decision makers. It is possible to make estimated guesses and informed forecasts of the stock price based on the past and present information (Chhajer et al., 2022). Li et al. (2019) propose that compare with point forecasting, interval forecasting offers decision makers more comprehensive information. Zhu et al. (2022) propose that interval forecasting can help decision makers and market participants make more informed decisions. Probabilistic forecasting is a class of forecasting in which the method provides intervals or probability distributions as outcomes of its forecasting (de Lima Silva et al., 2019). In the literature, calculating quantiles of the distribution is a commonly used parametric approach for interval forecasting. Granger et al. (1989) use this method by modelling the conditional mean and variance processes to calculate the quantiles. Chatfield (1993) makes a review study of the history and methods for interval forecasting. He also uses this parametric method based on a fitted probability model to construct forecast intervals. Li et al. (2019) also construct forecast intervals by calculate the quantiles with distributional assumptions and forecast variances. Tian and Hao (2020) obtained the interval forecasting results based on the results of point forecasting and the distribution function. They also propose that interval forecasting provide more reliable information for decision making. Ma and Dong (2020) propose an interval forecasting approach based on feature selection, the optimised machine learning method, and correction of the Gaussian distribution.

Compared to modelling the means, modelling the variance is relatively complex. It has been proved that financial market data commonly exhibit volatility clustering and highly responsive to the arrival of new information. Das et al. (2023) believe that this happens due to the human nature of overweighing new information as opposed to the prior data available. Black (1976) proposes the leverage effect to explain the relationship between stock returns and volatility. In stock market, the leverage effect refers to a phenomenon where stock volatility tends to increase when stock prices drop. The generalised autoregressive conditional heteroskedasticity (GARCH) model introduced by Bollerslev (1986) is a model that commonly used for modelling the variance. Khosravi et al. (2013) propose that the primary goal of the GARCH model is to model changes in variance. In their study, they use the GARCH model for the estimation of time-varying variance. Chong et al. (1999) performed stock market volatility forecasting with different types of GARCH models and compared their forecasting performance.

In the practical application of GARCH models, the choice of distribution assumption has a significant impact (Sun and Zhou, 2014). Adubisi et al. (2022) propose that most financial time series have non-normal features such as heavy tails, excess kurtosis and skewness. Cerqueti et al. (2020) performed volatility forecasting with GARCH-type models under skewed and non-skewed distributions in cryptocurrencies' return data. They selected normal, student-t, GED, and their skewed extensions for a comparison study. They propose that abnormal and asymmetric distributions should be taken into account in the selection of GARCH models. Horvath and Šopov (2016) used eight GARCH-type models under normal and student-t distributions to perform a simulation study. They find that the GARCH-type models under student-t distribution outperform those under normal distribution because the student-t distribution captures the tail risk better. Liu and Hung (2010) forecast S&P-100 stock index volatility using GARCH models under four distribution assumptions, include N, T, HT, and SGT. Shi and Feng (2016) believe that financial data in practice generally follow fat-tail distributions instead of a normal distribution. So they try the student-t, GED, and tempered stable distributions in their study. To account for the excess kurtosis and fat tails that are present in the residuals of the returns series, Gyamerah (2019) models GARCH models with student-t distribution, generalised error distribution (GED), and normal inverse Gaussian (NIG) types of distributions. To model the conditional variance, Yelamanchili (2020) applies four conventional GARCH models under Gaussian distribution, t-distribution, and generalised error distribution. Guo (2019) introduces a new type of heavy-tailed distribution, and compare its empirical performance with two popular types of heavy-tailed distribution, the student's t distribution and the normal inverse Gaussian distribution (NIG). The results illustrate that there is no overwhelmingly dominant distribution in fitting the data under the GARCH framework. Vacca et al. (2022) agree that Gaussian distribution fails to account for skewness and leptokurtosis. Horvath and Šopov (2016) used eight GARCH-type models under normal and student-t distributions to perform a simulation study. They find that the GARCH-type models under student-t distribution outperform those under normal distribution because the student-t distribution captures the tail risk better. Cerqueti et al. (2020) performed volatility forecasting with GARCH-type models under skewed and non-skewed distributions in cryptocurrencies' return data. They selected normal, student-t, GED, and their skewed extensions for a comparison study. They propose that abnormal and asymmetric distributions should be taken into account in the selection of GARCH models. Chen et al. (2023) also used the multivariate realised GARCH under a multivariate skew-t distribution.

All concepts of decision making are forward looking, and the explanation-induction-prediction-innovation sequence forms the basis of any method of decision making (Thomakos and Xidonas, 2023). Combination methods have gained increasing popularity in recent years to improve forecast accuracy by integrating information from multiple models. Similarly, the combination of interval forecasts has emerged as a valuable approach to enhance forecast performance. However, it is important to consider different heuristics and methodologies to optimise the combination process. One important aspect of forecast combination is the combination of experts' probability distributions in risk analysis (Clemen and Winkler, 1999). Experts play a vital role in providing valuable information, especially when 'hard data' is limited for uncertainties in risk analysis. By combining experts' probability distributions, a more comprehensive and reliable assessment of risks can be achieved. Gaba et al. (2017) further extend this notion by comparing different heuristics for combining interval forecasts and evaluating their performance against the simple average method. Their findings highlight the need to consider factors such as overconfidence and dependence among individual forecasts in the combination process.

In the context of interval forecasting competitions, the study by Grushka-Cockayne and Jose (2020) on the M4 forecasting competition emphasises the importance of interval calibration and accuracy. They investigate the benefits of interval combination using six recently-proposed heuristics, suggesting that interval aggregation offers improvements in terms of both calibration and accuracy. Their results support the notion that combining interval forecasts can lead to more reliable and informative predictions. Lv and Qi (2022) examined the effectiveness of combination forecast models in predicting stock market returns. They employed combination forecasts, which took the form of a simple average of 14 individual forecasts. They find that the mean combination forecast model outperformed other forecasting models in terms of out-of-sample performance. Moreover, specific domains such as tourism and carbon price forecasting have demonstrated the advantages of combining interval forecasts. Li et al. (2019) discuss the use of combination as an effective way to improve tourism forecasting accuracy. Their empirical results show that combination improves the accuracy of tourism interval forecasting for different horizons. In the context of carbon price forecasting, Liu et al. (2022) propose a combination forecasting model based on a hybrid interval multi-scale decomposition method. This model outperforms benchmark models in terms of accuracy and stability, showcasing the potential of combining interval forecasts in this domain. Wang et al. (2022a) provide a comprehensive review of forecast combinations, discussing various methods and highlighting crucial issues concerning their utility. They emphasise the need for flexibility and customisation in selecting combination methods to maximise forecast accuracy.

Furthermore, studies by Hill and Miller (2011), Xiong et al. (2015) and Wallis (2005) offer additional perspectives on combining datasets, conflation techniques, and the statistical framework for combined density and interval forecasts. Hill and Miller (2011) present the mathematical method of conflation, which allows for the consolidation of data from independent experiments measuring the same physical quantity. Xiong et al. (2015) propose a combination method for interval forecasting of agricultural commodity futures prices using vector error correction model (VECM) and multi-output support vector regression (MSVR). Wallis (2005) introduces the finite mixture distribution as a statistical model for a combined density forecast, providing insights into measures of uncertainty, disagreement, and the decomposition of the variance of the aggregate distribution.

Genest and Zidek (1986) reviewed several approaches for combining probability distributions. Gneiting and Ranjan (2013) studied combination formulas and aggregation methods for predictive cumulative distribution functions, and both linear and nonlinear aggregation methods are investigated. Liu and Chang (2021) presented a combination of Lorentz stable distribution and transition probability distribution description for the stock index changes. Alfaro and Inzunza (2023) propose GARCH models are a powerful tool for characterising the distribution of asset prices, which oftentimes requires Monte Carlo simulations to obtain relevant statistics, such as tail-risk measures. Wang et al. (2022b) propose Markov chain Monte Carlo (MCMC) is frequently used methods to provide an excellent approximation to the parameters of combination. Chen and Maung (2023) use Monte Carlo simulations to assess the finite-sample performance of the proposed estimators for both low-dimensional and high-dimensional data. They focus on developing a new non-parametric estimator of time-varying forecast combination weights that can deal with both model uncertainty and structural changes in forecasting.

The result of interval forecasting is a range of values, bounded by upper and lower forecast bounds, instead of a single value as in point forecasting. Hence, the coverage rate and the interval width are commonly used measurement to asses the accuracy of interval forecasts (Athanasopoulos et al., 2011). The coverage rate measures the proportion of actual values falling within the forecast interval. In practice, a coverage probability closer to the nominal coverage rate suggests better model performance. Hansen (2006) assesses the accuracy of forecast intervals with the coverage rate. Another important measurement of forecast interval is the average width. The forecast interval with a narrower interval width is considered to have better forecasting performance (Kim et al., 2011). Zhou et al. (2021) mention the use of prediction interval normalised average width to measure forecast intervals. Wang et al. (2022a) and Tsao et al. (2021) also apply the average width to asses the forecast intervals. Although the coverage rate and the interval width are good measurements of interval forecasting, selecting any of these measurements alone may lead to bias in model selection. Li et al. (2019) propose that the trade-off between coverage rates and interval widths is an unavoidable issue in interval forecasting research. They use the Winkler score (Winkler, 1972) to asses the forecast intervals, which penalises true values outside the forecast interval and rewards for narrower interval widths. Wang et al. (2022b) also use the Winkler score for the assessment of forecast intervals. Winkler score has the advantage that it takes into account both the accuracy of the forecasting and the width of the interval, providing a comprehensive assessment of the quality of interval forecasting. Winkler score is more appropriate for evaluating the overall performance of interval forecasts, so this study uses the Winkler score for the assessment of forecast intervals.

3 Interval forecasting under different distributional assumptions

3.1 GARCH models

One of the basic assumptions of the ordinary regression model is that the residual term has homoscedasticity, which means the residuals have consistent variances. However, in the practice, a large number of financial time series data show heteroscedasticity. To address this issue, the autoregressive conditional heteroskedasticity (ARCH) model was introduced by Engle (1982). And then an improved generalised autoregressive conditional heteroskedasticity (GARCH) model was introduced by Bollerslev (1986). The GARCH(1, 1) model is usually sufficient for most financial time series applications (Andersen and Bollerslev, 1998; Hansen and Lunde, 2005). In this study, the GARCH(1, 1) model can be expressed as:

$$r_t = \mu + \varepsilon_t \tag{1}$$

$$\varepsilon_t = \sigma_t z_t, \qquad z_t \; iid \; F(0,1) \tag{2}$$

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{3}$$

where r_t is the returns series, μ is the conditional mean and σ_t^2 is the conditional variance. ε_t is the innovation term, and $F(0, 1)$ is the assumed distribution of the innovation term. Parameters ω, α and β are strictly limited to be non-negative with the restriction of $\alpha + \beta < 1$. The GARCH (1, 1) model is usually sufficient for most financial time series applications (Andersen and Bollerslev, 1998; Hansen and Lunde, 2005).

3.2 Distributional assumptions of GARCH model

According to the relevant literature, the most commonly used distributional assumptions of GARCH models include symmetric distributions include: normal distribution (N), t distribution (T), and generalised error distribution (GED) and skewed distributions include: skewed t (ST), and skewed GED (SGED). This study also chose these distributional assumptions for the comparison of interval forecasting.

The normal distribution, also known as the Gaussian distribution, is a bell-shaped distribution that is widely used in GARCH modelling. It assumes that the errors are normally distributed, which means that the data is symmetrically distributed around the mean. The probability density function of the normal distribution can be expressed by:

$$f(x) = \frac{e^{\frac{-0.5(x-\mu)^2}{\sigma^2}}}{\sigma\sqrt{2\pi}} \tag{4}$$

The normal distribution is characterised by two parameters: the mean (μ) and the variance (σ^2).

The t-distribution is a more flexible distribution than the normal distribution and is commonly used in GARCH modelling when the data exhibits heavy tails. The t-distribution allows for more extreme values in the tails of the distribution compared to the normal distribution. The probability density function of the t distribution can be expressed by:

$$f(x)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\beta\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{(x-\alpha)^2}{\beta\nu}\right)^{-\left(\frac{\nu+1}{2}\right)} \tag{5}$$

where α, β, and ν are the location, scale, and shape parameters respectively, and Γ is the gamma function. The location parameter α is the mean of the distribution and the variance of the distribution can be calculated by:

$$Var(x)=\frac{\beta\nu}{(\nu-2)} \tag{6}$$

The generalised error distribution (GED) distribution is a flexible distribution that can capture both heavy-tailed and skewed distributions. It is particularly useful when the data is not normally distributed and exhibits non-constant variance. The probability density function of the generalised error distribution can be expressed by:

$$f(x)=\frac{ke^{-0.5\left|\frac{x-\alpha}{\beta}\right|^k}}{2^{1+k^{-1}}\beta\Gamma\left(k^{-1}\right)} \tag{7}$$

where α, β and k representing the location, scale and shape parameters. The location parameter α is the mean of the distribution and the variance of the distribution can be calculated by:

$$Var(x)=\beta^2 2^{2/k}\frac{\Gamma\left(3k^{-1}\right)}{\Gamma\left(k^{-1}\right)} \tag{8}$$

The skewed t-distribution assumes that the errors are normally distributed with skewness. This distribution is particularly useful in GARCH modelling when the data exhibits asymmetry. For example, when modelling stock returns, negative returns may have a different distribution than positive returns due to market asymmetry. The skewed t-distribution assumes that the errors are t-distributed with skewness. This distribution is similar to the skewed normal distribution but allows for more flexibility in modelling the tails of the distribution. The skewed t-distribution is particularly useful when the data exhibits both asymmetry and heavy tails. The skewed GED distribution can capture both heavy-tailed and skewed distributions. Lee and McLachlan (2022) make an overview of skew distributions and propose that most literature handle skewness by introducing latent skewing variable. The 'rugarch' package in R introducing skewness into symmetric distributions by adding a skew parameter ξ. The specific method can refer to Fernandez and Steel (1998). Overall, the parameters of five distributions are shown in Table 1.

Table 1 Parameters of distributional assumptions

	Location	*Scale*	*Skew*	*Shape*
N	√	√	NA	NA
T	√	√	NA	√
GED	√	√	NA	√
ST	√	√	√	√
SGED	√	√	√	√

3.3 Experimental data

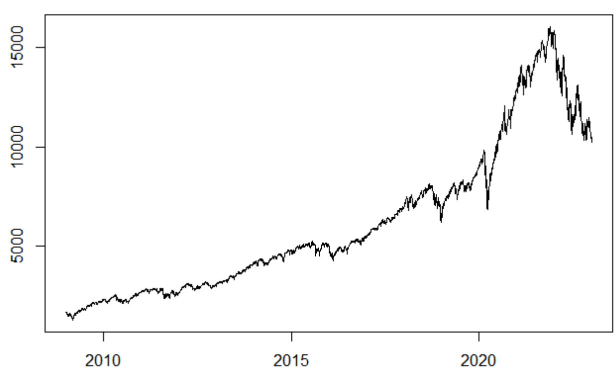
This study selects the Nasdaq Composite Index for the comparison of interval forecasting based on five distributions. The data used in this study is obtained from Yahoo Finance website (http://finance.yahoo.com/) covering the period between 15 January 2009 and 30 December 2022 (14 years). The original data is transformed into return data. The return (r_t) is defined as the logarithm of the daily closing price differences times 100, which can be expressed by:

$$r_t = 100 \times \ln\left(\frac{p_t}{p_{t-1}}\right) \tag{9}$$

Figure 1 shows the pattern of the Nasdaq Composite Index. Figure 2 shows the pattern of the returns series of the Nasdaq Composite Index.

Table 2 shows the descriptive statistics of five returns series over the whole sample period. The statistics include mean, standard deviation, skewness, kurtosis, ADF, Jarque-Bera, and the number of observations.

Figure 1 The Nasdaq Composite Index



From Table 2, the experimental data has 3,523 observation. The means of the data is 0.0528, and the standard deviation is 1.3407. Skewness is a measure of the asymmetry of a distribution. It indicates whether the distribution is symmetric or skewed to one side. A positive skewness value indicates that the tail of the distribution is longer on the positive side, while a negative skewness value indicates that the tail is longer on the negative side. In this study, the data has a negative skewness value, which means the distribution of the data is skewed to the left. Kurtosis is a measure of the 'peakedness' of a distribution. It indicates whether the distribution is more or less peaked than a normal distribution. A positive kurtosis value indicates that the distribution has more extreme values than a normal distribution, while a negative kurtosis value indicates that the distribution has fewer extreme values. In this study, the data has positive skewness values, which means the distribution of the data has more extreme values than a normal distribution. The augmented Dickey-Fuller (ADF) test is a statistical test used to determine whether a time series is stationary. The test's p-values in this study is less than the chosen significance level 0.01, indicate that the return data are stationary. The Jarque-Bera test is a statistical test that examines the normality assumption of data by assessing its skewness and kurtosis. If the test's p-value is less than the chosen significance level (in this study, 0.01), the null hypothesis of normal distribution is rejected. Based on this test, all of the data are rejected to follow normal distribution.

Figure 2 Return series of the Nasdaq Composite Index

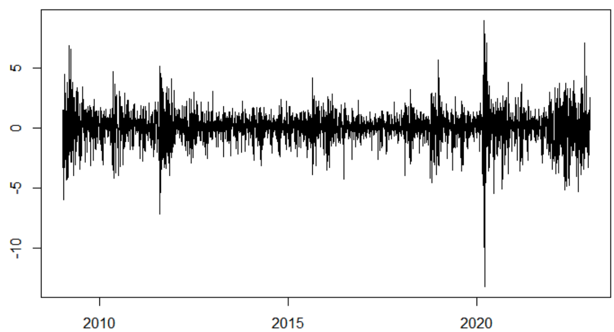


Table 2 Descriptive statistics

Mean	*Standard deviation*	*Skewness*	*Kurtosis*	*ADF*	*Jarque-Bera*	*Observations*
0.0528	1.3407	–0.5169	9.9777	–15.509*	7,303.9*	3,523

3.4 Forecast intervals

This study employs a one-day-ahead rolling forecasting method for the variances series using the GARCH model. The 'rugarch' package in R is used to implement this approach. First, the experimental data is divided by year (for the purpose of comparing the forecasting performance of models within different time periods). The data within 2009 is used as the in-sample data, as well as the initial moving window. Subsequently, the window moves forward by removing the earliest data and adding the latest data. The parameters are re-estimated regularly to account for the changes of data over time, and the continuous variances series can be obtained.

The forecast intervals in this study are obtained by calculating the quantiles based on given distributional assumptions and confidence level. The quantile function can be expressed as:

$$Q_t(\theta) = F_t^{-1}(\theta) \tag{10}$$

where $F_t(x)$ is the assumed distribution function. The forecast interval on confidence level p is bounded by Q_t $(q/2)$ and Q_t $(1 - q/2)$, where $q = 1 - p$. In this study, the confidence level p is set at 0.90, so the forecast intervals under N, T, ST, GED, and SGED are bounded by Q_t (0.05) and Q_t (0.95), respectively.

3.5 Average Winkler score

This study uses the average Winkler score (AWS) to assess the accuracy of interval forecasting. The AWS is calculated by assigning a score to each forecast interval based on its coverage of the true value, and the width of the interval. The Winkler score gives higher weights to intervals that have better coverage of the true value and are narrower and lower weights to intervals that have poor coverage or are wider. The AWS in this study can be expressed as:

$$W_t = \hat{U}_t - \hat{L}_t \tag{11}$$

$$WS_t = \begin{cases} W_t & L_t \le y_t \le U_t \\ W_t + 2\left(L_t - y_t\right)/(1-p) & L_t > y_t \\ W_t + 2\left(y_t - U_t\right)/(1-p) & U_t < y_t \end{cases} \tag{12}$$

W_t refers to the width of the forecast interval at time t, $\hat{U}_t$ and $\hat{L}_t$ refer to the predicted upper and lower bounds at time t. WS_t is the Winkler score of the forecast interval at time t, and p refers to the confidence level. In this study, interval forecasting is a continuous process. For continuous rolling interval forecasting, we introduce the AWS to evaluate the overall forecasting performance, which can be calculated as follow:

$$AWS = \frac{1}{m}\sum\nolimits_{t=1}^{m} WS_t \tag{13}$$

The m is the number of the forecast intervals. Under a confidence level p, the model with lower AWS is considered to have better interval forecasting performance.

3.6 Results

The forecast intervals of the returns data in different years under different distributional assumptions are assess by the AWS, and the results are shown in Table 3. From Table 3, each model exhibits significant differences in AWS values across different years, while the differences in AWS values among different models within the same year are small. This indicates that these models have adapted well to the changes in the data and provide reliable interval forecasting results.

To demonstrate the performance of different models in different years, Table 4 provides the ranking of models in 2010–2022.

From Table 4, significant changes in the ranking of model performance across different years can be observed. Figure 3 depicts the ranking of the forecasting performance of five individual models in different years using line graphs. The numbers in the figures represent the ranking of the model in the year. For example, the number 1 indicates that the model ranked first in the year, which means it has the best forecasting performance.

Table 3 AWS of forecast intervals

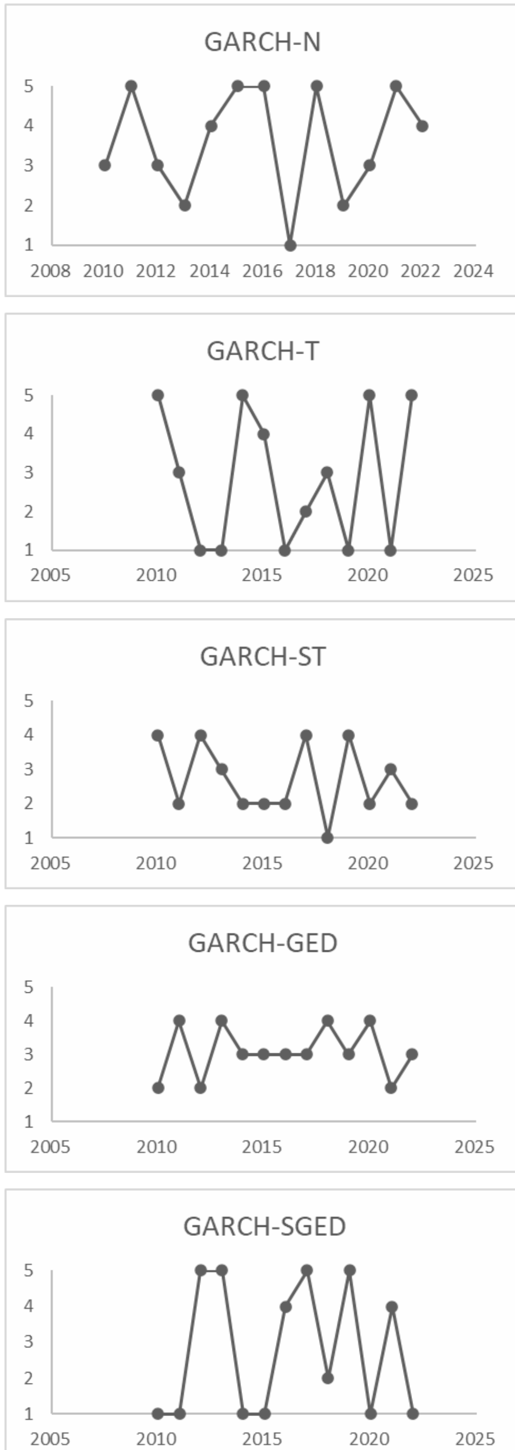
Year	*GARCH-N*	*GARCH-T*	*GARCH-ST*	*GARCH-GED*	*GARCH-SGED*
2010	5.7094	5.7443	5.7203	5.7084	5.6719
2011	6.8044	6.6601	6.6249	6.6749	6.5928
2012	4.2553	4.2472	4.2690	4.2546	4.2756
2013	3.4929	3.4805	3.4934	3.5034	3.5154
2014	3.9033	3.9386	3.8621	3.8925	3.8264
2015	4.7574	4.7571	4.4678	4.7410	4.4601
2016	4.3776	4.3230	4.3416	4.3580	4.3598
2017	2.9044	2.9209	2.9901	2.9431	3.0092
2018	5.8364	5.7570	5.6165	5.7929	5.6912
2019	4.2867	4.2545	4.2964	4.2938	4.3123
2020	8.4254	8.4553	8.3978	8.4289	8.3637
2021	4.8621	4.7795	4.8270	4.8190	4.8406
2022	8.3528	8.3565	8.2094	8.3461	8.1958

Table 4 Rankings of single models

Ranking	*2010*	*2011*	*2012*	*2013*
1	GARCH-SGED	GARCH-SGED	GARCH-T	GARCH-T
2	GARCH-GED	GARCH-ST	GARCH-GED	GARCH-N
3	GARCH-N	GARCH-T	GARCH-N	GARCH-ST
4	GARCH-ST	GARCH-GED	GARCH-ST	GARCH-GED
5	GARCH-T	GARCH-N	GARCH-SGED	GARCH-SGED

Ranking	*2014*	*2015*	*2016*
1	GARCH-SGED	GARCH-SGED	GARCH-T
2	GARCH-ST	GARCH-ST	GARCH-ST
3	GARCH-GED	GARCH-GED	GARCH-GED
4	GARCH-N	GARCH-T	GARCH-SGED
5	GARCH-T	GARCH-N	GARCH-N
Ranking	*2017*	*2018*	*2019*
1	GARCH-N	GARCH-ST	GARCH-T
2	GARCH-T	GARCH-SGED	GARCH-N
3	GARCH-GED	GARCH-T	GARCH-GED
4	GARCH-ST	GARCH-GED	GARCH-ST
5	GARCH-SGED	GARCH-N	GARCH-SGED
Ranking	*2020*	*2021*	*2022*
1	GARCH-SGED	GARCH-T	GARCH-SGED
2	GARCH-ST	GARCH-GED	GARCH-ST
3	GARCH-N	GARCH-ST	GARCH-GED
4	GARCH-GED	GARCH-SGED	GARCH-N
5	GARCH-T	GARCH-N	GARCH-T

Figure 3 The ranking of five models



From Figure 3, the GARCH-N, GARCH-T, and GARCH-SGED are ranked first in some years, but last in other years. Such results imply that if only use the data from a single year as the sample data for model comparison to select the optimal model for future forecasting, the forecasting performance will significantly decline. So it is unwise to select the optimal model based on data from a single year. The approach to improving overall forecasting performance is to find a method that has a more stable expected forecasting performance. Therefore, in the proposal of combined interval forecasting method, the significance of the combination lies in obtaining relatively stable forecasting performance by combining individual models, rather than attempting to achieve optimal forecasting performance.

4 Combination of interval forecasting

4.1 A non-parametric method to combine forecast intervals

One of the most straightforward and practical methods for combining prediction intervals is to directly compute the average of the upper and lower bounds of individual prediction intervals to obtain the combined prediction interval.

Consider N forecast intervals from individual models at time t, $L_{i,t}$ and $U_{i,t}$ are the lower and upper bounds of the forecast interval obtained based on model i. The lower and upper bounds of equal weights combined interval $L_{c,t}$ and $U_{c,t}$ can be express as:

$$L_{c,i} = \sum_{i=1}^{N} \frac{1}{N} L_{i,t}, \tag{14}$$

$$U_{c,i} = \sum_{i=1}^{N} \frac{1}{N} U_{i,t}. \tag{15}$$

However, the combined interval obtained through this method lacks the corresponding probability, making it difficult to provide an effective interpretation for it. To address this problem, an axiomatic approach is introduced by combining the density of individual distributions. For example, for the value x_i at time t, the forecast probability density of model i is $g_{i,t}(x_i)$, then the equal weights combined probability density function $f_{i,t}(x_i)$ is:

$$f_{i,t}(x_i) = \sum_{i=1}^{N} \frac{1}{N} g_{i,t}(x_i). \tag{16}$$

Subsequently, the lower and upper bounds of forecast interval can be obtained by calculating the quantiles with the combined density. In practice, combining probability density functions and calculating quantiles can be computationally demanding, especially when dealing with a large number of complex distributions. Therefore, this study draws inspiration from the Monte Carlo method and employs a non-parametric sampling approach to achieve the combinations of different distributions. Let x_j be a random variable obtained by sampling from distribution n. Let n represent the equal sample sizes from distributions 1, 2, …, j, respectively. The process of independently sampling from the j distributions and combining them to form a combined sample S can be expressed mathematically as:

$$S = [x_{1,1}, x_{1,2}, \ldots, x_{1,n}, x_{2,1}, x_{2,2}, \ldots, x_{2,n}, \ldots, x_{j,1}, x_{j,2}, \ldots, x_{j,n}], \quad (17)$$

where $x_{1,1}$, $x_{1,2}$, …, $x_{1,n}$ are n independent observations sampled from distribution 1, $x_{2,2}$, …, $x_{2,n}$ are n independent observations sampled from distribution 2, …, $x_{j,1}$, $x_{j,2}$, …, $x_{j,n}$ are n independent observations sampled from distribution j. Once the combined sample is formed, the upper and lower bounds of the combined forecast interval can be obtained by calculating the quantiles of the combined sample. By repeating the process of randomly sampling and combining the samples, multiple composite samples can be obtained, thereby obtaining different forecast intervals.

Based on the five distribution assumptions mentioned above, we conducted five repeated experiments using this non-parametric combination method. The equal sampling size are set at 100,000.

4.2 Results

The results of five combinations are shown in Table 5. From Table 5, the results of the five combinations are very close, indicating that the results obtained using this non-parametric combination method are relatively stable. The random sampling process does not have a significant impact on the results, which also implies that this non-parametric combination method is an effective approach for interval prediction combination.

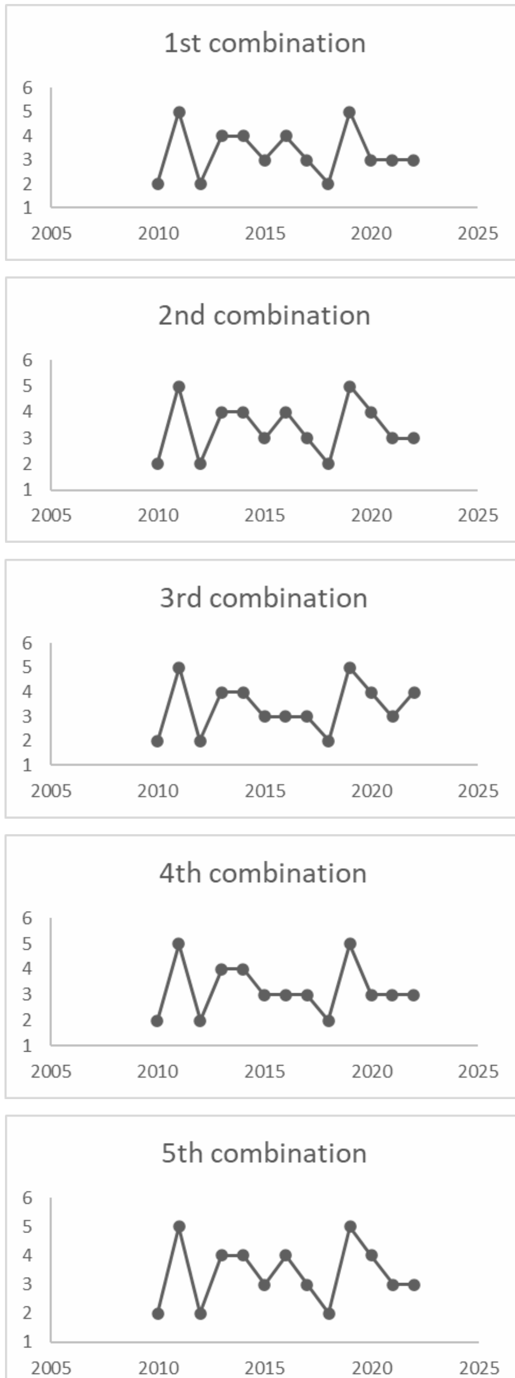
Table 5 AWS of five combinations

Year	*1st combined*	*2nd combined*	*3rd combined*	*4th combined*	*5th combined*	*Average*
2010	5.6873	5.6892	5.6890	5.6904	5.6887	5.6889
2011	6.6819	6.6824	6.6821	6.6794	6.6847	6.6821
2012	4.2525	4.2511	4.2504	4.2519	4.2506	4.2513
2013	3.5011	3.5001	3.5004	3.4992	3.5007	3.5003
2014	3.8957	3.8964	3.8973	3.8955	3.8954	3.8960
2015	4.6638	4.6637	4.6671	4.6625	4.6628	4.6640
2016	4.3591	4.3583	4.3571	4.3569	4.3594	4.3582
2017	2.9376	2.9369	2.9364	2.9352	2.9394	2.9371
2018	5.6732	5.6652	5.6708	5.6732	5.6691	5.6703
2019	4.3066	4.3081	4.3079	4.3041	4.3110	4.3075
2020	8.4248	8.4269	8.4288	8.4248	8.4256	8.4262
2021	4.8229	4.8242	4.8237	4.8220	4.8260	4.8237
2022	8.3448	8.3407	8.3483	8.3425	8.3417	8.3436

To compare the performance of forecast intervals obtained from this combination method with those obtained from five individual models, the ranking of five combined models with five individual models are shown in Figure 4, respectively.

Figure 4 indicates that the combined interval forecasting does not have the worst forecasting performance in any period and, when compared to individual models, generally falls around the intermediate level. This result also indicates that the combined method is influenced by the individual model that has the poorest performing in certain years, resulting in a decrease in forecasting performance. However, overall, the use of combination method reduces the loss in forecasting performance of selecting the individual model that has the poorest performing in the year.

Figure 4 The ranking of five combinations



5 Summary

5.1 Conclusions and contributions

This study uses daily return data of the Nasdaq Composite Index in the periods from 2009 to 2022 for interval forecasting. The GARCH model under five distributional assumptions is employed to generate forecast intervals, and the forecasting accuracy is evaluated using the average Winkler score. By dividing the forecast intervals by the year, the performance of each model within each years are assessed respectively. The results demonstrated significant variations in the performance rankings of individual models across different years. Particularly, a model that has the highest predictive accuracy in one year could have the lowest accuracy in other years. These findings indicate the risk involved in relying on data from a specific time period to compare and select the optimal model for future predictions. One possible reason for these results is that the patterns of stock price index changes across different periods. This study proposes a combination approach to address the issue of insufficient robustness in the forecasting performance of individual models. A non-parametric combination method based on the Monte Carlo sampling is presented. The random sampling combination is repeated five times. The results revealed that although the interval forecasting performance of the combined model is not optimal in any period, they are never the worst either. The results from the five repeated experiments are highly stable and indicate that the interval forecasting performance of the combined model is consistently close to the average level of the individual models.

In conclusion, this study highlights the risks associated with relying on specific time periods to select the optimal predictive model. The proposed combination method offers decision makers a more robust interval forecasting approach, mitigating potential risks associated with model selection in forecasting.

5.2 Implications and future research directions

Although the results of this study have demonstrated the practical value of this combination method, the theoretical research of this method still needs to be conducted. This will require further integration of literature and research in the mathematical field. Additionally, this study utilised an equal-weight combination approach. However, some studies have pointed out that the weight allocation of different models in the combination process significantly affects the performance of the combination. This aspect will also be one of our future research directions.

References

Adubisi, O.D., Abdulkadir, A., Farouk, U.A. and Chiroma, H. (2022) 'The exponentiated half logistic skew-t distribution with GARCH-type volatility models', *Scientific African*, Vol. 16, p.e01253.

Alfaro, R. and Inzunza, A. (2023) 'Modeling S&P500 returns with GARCH models', *Latin American Journal of Central Banking*, Vol. 4, No. 3, p.100096.

Andersen, T.G. and Bollerslev, T. (1998) 'Answering the skeptics: yes, standard volatility models do provide accurate forecasts', *International Economic Review*, pp.885–905.

Athanasopoulos, G., Hyndman, R.J., Song, H. and Wu, D.C. (2011) 'The tourism forecasting competition', *International Journal of Forecasting*, Vol. 27, No. 3, pp.822–844.

Black, F. (1976) 'Studies of stock price volatility changes', *Proceedings of the 1976 Meeting of the American Statistical Association, Business and Economic Statistics Section*, American Statistical Association, Washington DC, pp.177–181.

Bollerslev, T. (1986) 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics*, Vol. 31, No. 3, pp.307–327.

Cerqueti, R., Giacalone, M. and Mattera, R. (2020) 'Skewed non-Gaussian GARCH models for cryptocurrencies volatility modelling', *Information Sciences*, Vol. 527, pp.1–26.

Chatfield, C. (1993) 'Calculating interval forecasts', *Journal of Business & Economic Statistics*, Vol. 11, No. 2, pp.121–135.

Chen, B. and Maung, K. (2023) 'Time-varying forecast combination for high-dimensional data', *Journal of Econometrics*, p.105418.

Chen, Q., Huang, Z. and Liang, F. (2023) 'Measuring systemic risk with high-frequency data: a realized GARCH approach', *Finance Research Letters*, Vol. 54, p.103753.

Chhajer, P., Shah, M. and Kshirsagar, A. (2022) 'The applications of artificial neural networks, support vector machines, and long-short term memory for stock market prediction', *Decision Analytics Journal*, Vol. 2, p.100015.

Chong, C.W., Ahmad, M.I. and Abdullah, M.Y. (1999) 'Performance of GARCH models in forecasting stock market volatility', *Journal of Forecasting*, Vol. 18, No. 5, pp.333–343.

Clemen, R.T. and Winkler, R.L., (1999) 'Combining probability distributions from experts in risk analysis', *Risk analysis*, Vol. 19, pp.187–203.

Das, D., Kayal, P. and Maiti, M. (2023) 'A K-means clustering model for analyzing the Bitcoin extreme value returns', *Decision Analytics Journal*, Vol. 6, p.100152.

de Lima Silva, P.C., Sadaei, H.J., Ballini, R. and Guimarães, F.G. (2019) 'Probabilistic forecasting with fuzzy time series', *IEEE Transactions on Fuzzy Systems*, Vol. 28, No. 8, pp.1771–1784.

Engle, R.F. (1982) 'Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation', *Econometrica*, Vol. 50, No. 4, pp.987–1007.

Fernandez, C. and Steel, M.F. (1998) 'On Bayesian modeling of fat tails and skewness', *Journal of the American Statistical Association*, Vol. 93, No. 441, pp.359–371.

Gaba, A., Tsetlin, I. and Winkler, R.L., (2017) 'Combining interval forecasts', *Decision Analysis*, Vol. 14, No. 1, pp.1–20.

Genest, C. and Zidek, J.V. (1986) 'Combining probability distributions: a critique and an annotated bibliography', *Statistical Science*, Vol. 1, No. 1, pp.114–135.

Gneiting, T. and Ranjan, R. (2013) 'Combining predictive distributions', *Electronic Journal of Statistics*, Vol. 7, pp.1747–1782.

Granger, C.W.J., White, H. and Kamstra, M. (1989) 'Interval forecasting: an analysis based upon ARCH-quantile estimators', *Journal of Econometrics*, Vol. 40, No. 1, pp.87–96.

Grushka-Cockayne, Y. and Jose, V.R.R. (2020) 'Combining prediction intervals in the M4 competition', *International Journal of Forecasting*, Vol. 36, No. 1, pp.178–185.

Guo, Z.Y. (2019) 'Empirical performance of GARCH models with heavy-tailed innovations', *Bulletin of Economic Research*, Vol. 71, No. 3, pp.359–387.

Gyamerah, S.A. (2019) 'Modelling the volatility of Bitcoin returns using GARCH models', *Quantitative Finance and Economics*, Vol. 3, No. 4, pp.739–753.

Hansen, B.E. (2006) 'Interval forecasts and parameter uncertainty', *Journal of Econometrics*, Vol. 135, Nos. 1–2, pp.377–398.

Hansen, P.R. and Lunde, A. (2005) 'A forecast comparison of volatility models: does anything beat a GARCH (1, 1)?', *Journal of Applied Econometrics*, Vol. 20, No. 7, pp.873–889.

Hill, T.P. and Miller, J. (2011) 'How to combine independent data sets for the same quantity', *Chaos: An Interdisciplinary Journal of Nonlinear Science*, Vol. 21, No. 3, p.033102.

Horvath, R. and Šopov, B. (2016) 'GARCH models, tail indexes and error distributions: an empirical investigation', *The North American Journal of Economics and Finance*, Vol. 37, pp.1–15.

Khosravi, A., Nahavandi, S., and Creighton, D. (2013) 'A neural network-GARCH-based method for construction of prediction intervals', *Electric Power Systems Research*, Vol. 96, pp.185–193.

Kim, J.H., Wong, K., Athanasopoulos, G. and Liu, S. (2011) 'Beyond point forecasting: evaluation of alternative prediction intervals for tourist arrivals', *International Journal of Forecasting*, Vol. 27, No. 3, pp.887–901.

Lee, S.X. and McLachlan, G.J. (2022) 'An overview of skew distributions in model-based clustering', *Journal of Multivariate Analysis*, Vol. 188, p.104853.

Li, G., Wu, D.C., Zhou, M. and Liu, A. (2019) 'The combination of interval forecasts in tourism', *Annals of Tourism Research*, Vol. 75, pp.363–378.

Liu, C., and Chang, C. (2021) 'Combination of transition probability distribution and stable Lorentz distribution in stock markets', *Physica A: Statistical Mechanics and its Applications*, Vol. 565, p.125554.

Liu, H.C. and Hung, J.C. (2010) 'Forecasting S&P-100 stock index volatility: the role of volatility asymmetry and distributional assumption in GARCH models', *Expert Systems with Applications*, Vol. 37, No. 7, pp.4928–4934.

Liu, J., Wang, P., Chen, H. and Zhu, J. (2022) 'A combination forecasting model based on hybrid interval multi-scale decomposition: application to interval-valued carbon price forecasting', *Expert Systems with Applications*, Vol. 191, p.116267.

Lv, W. and Qi, J. (2022) 'Stock market return predictability: a combination forecast perspective', *International Review of Financial Analysis*, Vol. 84, p.102376.

Ma, X. and Dong, Y. (2020) 'An estimating combination method for interval forecasting of electrical load time series', *Expert Systems with Applications*, Vol. 158, p.113498.

Shi, Y. and Feng, L. (2016) 'A discussion on the innovation distribution of the Markov regime-switching GARCH model', *Economic Modelling*, Vol. 53, pp.278–288.

Sun, P. and Zhou, C. (2014) 'Diagnosing the distribution of GARCH innovations', *Journal of Empirical Finance*, Vol. 29, pp.287–303.

Thomakos, D. and Xidonas, P. (2023) 'The origins of forward-looking decision making: cybernetics, operational research, and the foundations of forecasting', *Decision Analytics Journal*, Vol. 8, p.100284.

Tian, C. and Hao, Y. (2020) 'Point and interval forecasting for carbon price based on an improved analysis-forecast system', *Applied Mathematical Modelling*, Vol. 79, pp.126–144.

Timmermann, A. (2006) 'Forecast combinations', in Elliott, G., Granger, C.W.J. and Timmermann, A. (Eds.): *Handbook of Economic Forecasting*, Vol. 1, pp.135–196, Elsevier.

Tsao, H.H., Leu, Y.G. and Chou, L.F. (2021) 'A center-of-concentrated-based prediction interval for wind power forecasting', *Energy*, Vol. 237, p.121467.

Vacca, G., Zoia, M.G. and Bagnato, L. (2022) 'Forecasting in GARCH models with polynomially modified innovations', *International Journal of Forecasting*, Vol. 38, No. 1, pp.117–141.

Wallis, K.F. (2005) 'Combining density and interval forecasts: a modest proposal', *Oxford Bulletin of Economics and Statistics*, Vol. 67, pp.983–994.

Wang, J., Gao, J. and Wei, D. (2022a) 'Electric load prediction based on a novel combined interval forecasting system', *Applied Energy*, Vol. 322, p.119420.

Wang, S., Wang, L., Gao, S. and Bai, Z. (2017) 'Stock price prediction based on chaotic hybrid particle swarm optimisation-RBF neural network', *International Journal of Applied Decision Sciences*, Vol. 10, No. 2, pp.89–100.

Wang, X., Hyndman, R.J., Li, F. and Kang, Y. (2022b) 'Forecast combinations: an over 50-year review', *International Journal of Forecasting*.

Winkler, R.L. (1972) 'A decision-theoretic approach to interval estimation', *Journal of the American Statistical Association*, Vol. 67, No. 337, pp.187–191.

Xiong, T., Li, C., Bao, Y., Hu, Z. and Zhang, L. (2015) 'A combination method for interval forecasting of agricultural commodity futures prices', *Knowledge-Based Systems*, Vol. 77, pp.92–102.

Yelamanchili, R.K. (2020) 'Modeling stock market monthly returns volatility using GARCH models under different distributions', *International Journal of Accounting & Finance Review*, Vol. 5, No. 1, pp.42–50.

Zhou, M., Wang, B., Guo, S. and Watada, J. (2021) 'Multi-objective prediction intervals for wind power forecast based on deep neural networks', *Information Sciences*, Vol. 550, pp.207–220.

Zhu, B., Wan, C. and Wang, P. (2022) 'Interval forecasting of carbon price: a novel multiscale ensemble forecasting approach', *Energy Economics*, Vol. 115, p.106361.