ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.9.1.P9

Prediction of Weibo Retweet Volume for the China Internet Joint Rumor-Refutation Platform Based on Uncertain Time Series Analysis

Jianing Yu

School of Management, University Sains Malaysia, Minden Penang 11800, Malaysia

yxd20040402@163.com

Abstract: With the widespread use of social media, the speed and reach of information dissemination have reached unprecedented levels. However, this has also accelerated the spread of misinformation and rumors. The China Internet Joint Rumor-Refutation Platform plays a crucial role in combating rumors and disseminating verified information. This study analyzes retweet data from the platform's official Weibo account to assess and forecast uncertainty in retweet volume. Using uncertain time series analysis, data from January to June 2024 are modeled to determine the optimal model order and estimate model parameters. Residual analysis and uncertain hypothesis testing show that the selected model reasonably fits the observed data and that the residuals conform to the assumed uncertainty distribution. The forecasting results indicate that the model can provide reliable predictions of future retweet volumes together with prediction intervals, thereby providing a quantitative basis for assessing forecast uncertainty.

Keywords: uncertain time series analysis; social media; information dissemination; prediction model

1. Introduction

In the era of information explosion, social media platforms have become one of the primary channels for information dissemination. However, this has also led to the rapid spread of misinformation and rumors. As an important channel for combating rumors and disseminating verified information, the official Weibo account of the China Internet Joint Rumor-Refutation Platform plays a crucial role in maintaining a healthy online information environment through its influence and dissemination effectiveness. This study analyzes the retweet data of this Weibo account and forecasts the uncertainty in its retweet volume.

Uncertainty theory, proposed and developed by Liu and others, is a branch of mathematics for studying uncertain phenomena and is distinct from probability theory, which is used to analyze random phenomena[1] [2]. In practice, when observational data are used to study quantities such as stock prices and temperatures, a distribution function must first be determined. If we believe that this distribution function is sufficiently close to the true frequency distribution, probability theory should be used; otherwise, uncertainty theory should be adopted. Although many people believe that a sufficiently large amount of observational data can yield a distribution function close to the frequency distribution, Liu pointed out that such observations are often neither drawn from the same population nor random sequences in the probabilistic sense [3]. If they are nevertheless used to construct a distribution function, the resulting distribution may deviate from future frequencies.

As a branch of uncertainty theory, uncertain statistics was proposed by Liu in 2010 as a system of statistical tools for collecting, analyzing, and interpreting data [4]. After years of development, uncertain statistics has mainly covered five areas: uncertain statistical model estimation, uncertain hypothesis testing, uncertain regression analysis, parameter estimation for uncertain differential equations, and uncertain time series analysis. Uncertain time series analysis uses uncertainty theory to predict future values from previously observed data. Yang and Liu first proposed that the disturbance terms in time-series models should be treated as uncertain variables rather than random variables and therefore referred to them as uncertain disturbance terms [5]. This work marked the beginning of uncertain time series analysis. To estimate uncertain disturbance terms, Yang and Liu introduced the method of moments [5]; Chen and Yang proposed maximum likelihood estimation [6];

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.9.1.P9
Liu and Liu subsequently modified the method [7]; and Ye and Yang applied uncertain time series analysis to epidemic-spread research [8].

The retweet volume of rumor-refutation information is an important factor in suppressing rumors. This study uses uncertain time series analysis to model and forecast the retweet volume of the official Weibo account of the China Internet Joint Rumor-Refutation Platform. The remainder of this paper is organized as follows. Section 2 briefly introduces the basic concepts of uncertain time series analysis. Section 3 presents the retweet-volume data from the platform's official Weibo account. Section 4 applies uncertain time series analysis to model and forecast these data. Section 5 concludes the paper.

2. Uncertain Time Series Analysis

This section introduces the basic concepts and methods of uncertain time series analysis (UTSA), including the uncertain autoregressive model, parameter estimation, residual analysis, uncertain hypothesis testing, and forecasting.

2.1 Uncertain Autoregressive Model

The uncertain autoregressive model is an application of uncertainty theory to time-series analysis and is suitable for data that do not satisfy conventional probability-distribution assumptions. In a classical autoregressive model, the error term is usually assumed to follow a normal distribution. In real-world data, however, uncertainty often exists that cannot be fully described by traditional probability theory. By introducing uncertainty theory, the uncertain autoregressive model provides a new tool for analyzing such time-series data.

The mathematical expression of an uncertain autoregressive model can be written as follows:

$$X_t = a_0 + \sum_{i=1}^{k} a_i X_{t-i} + \varepsilon, \tag{1}$$

where X_t is the observation at time t, a_0 is the constant term of the model, a_i denotes the model parameters, k is the order of the model, and ε is the uncertain disturbance term. Unlike a random variable in the conventional sense, it is an uncertain variable used to describe fluctuations that cannot be characterized by traditional probability distributions.

2.2 Parameter Estimation

Uncertainty theory uses least-squares estimation (LSE) to determine the model parameters. Specifically, the parameters are obtained by minimizing the sum of squared residuals:

$$\min_{a_0,a_1,\ldots,a_k} \sum_{t=k+1}^{n} \left(X_t - a_0 - \sum_{i=1}^{k} a_i X_{t-i} \right)^2, \tag{2}$$

n is the sample size, while X_t and X_{t-i} are the observations at time points t and $t-i$, respectively. If the minimizing solution is $a_0, a_1, \ldots, a_k$, then the fitted autoregressive model is

$$X_t = \hat{a}_0 + \sum_{i=1}^{k} \hat{a}_i X_{t-i}. \tag{3}$$

2.3 Residual Analysis

Residual analysis is an important step in evaluating the goodness of fit of an uncertain autoregressive model. The residual ε_t is defined as the difference between the actual observation X_t and the model-predicted value $\widehat{X}_t$:

$$\varepsilon_t = X_t - \widehat{X}_t. \tag{4}$$

In uncertainty theory, the residual ε_t is assumed to follow an uncertainty distribution. Its expected value $E[\varepsilon_t]$ and uncertainty measure can be used to evaluate model forecasting performance.

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.9.1.P9
Ideally, if the model fits well, the expected value should be close to zero, indicating that the model has no systematic bias. The predicted value $\widehat{X}_{t+1}$ can be expressed as:

$$X_{t+1} = \hat{a}_0 + \sum_{i=1}^{k} \hat{a}_i X_{t+1-i}\,. \tag{5}$$

The residuals $\varepsilon_{k+1}, \varepsilon_{k+2}, \ldots, \varepsilon_n$, treated as samples of the uncertain disturbance term ε in Equation (1), can be used to estimate the expected value and variance of the uncertain disturbance term as follows:

$\hat{e} = \frac{1}{n-k}\sum_{t=k+1}^{n} \varepsilon_t$,

The standard deviation is

$$\widehat{\sigma}^2 = \frac{1}{n-k}\sum_{t=k+1}^{630} (\varepsilon_t - \hat{e})^2\,.$$

Therefore, if ε is further assumed to follow a normal uncertainty distribution, an uncertain autoregressive model is obtained as follows:

$$X_t = \hat{a}_0 + \sum_{\mathrm{i}=1}^{\mathrm{k}} \hat{\mathrm{a}}_{\mathrm{i}} \mathrm{X}_{\mathrm{t}-\mathrm{i}} + N(\hat{e}, \widehat{\sigma})\,. \tag{6}$$

2.4 Uncertain Hypothesis Testing

By analyzing the residual sequence, we can examine whether the model satisfies the basic assumptions of uncertainty theory, such as whether the residuals behave as non-probabilistic uncertain variables. Uncertain hypothesis testing is used to evaluate whether the model is suitable for the observed data. To determine whether the uncertain autoregressive model fits the observed data well, we need to test whether the normal uncertainty distribution $N(\hat{e}, \widehat{\sigma})$ fits the $N-k$ residuals. Uncertain time series analysis generally uses testing methods based on uncertainty theory to evaluate goodness of fit, that is, to determine whether the residuals follow a specified uncertainty distribution.

2.5 Forecasting

In uncertain time series analysis, forecasting uses an established model to estimate future observations. Given an uncertain autoregressive model, the next step is to forecast the value at the next time point and quantify the associated uncertainty. This process involves not only calculating the forecast value but also evaluating forecast uncertainty and determining a confidence interval. Based on the estimated uncertain autoregressive model, Yang and Liu proposed the following predictive uncertain variable for the prediction variable X_{t+1} with respect to the time series $X_1, X_2, \ldots, X_t$:

$$\widehat{X}_{t+1} = \hat{a}_0 + \sum_{\mathrm{i}=1}^{\mathrm{k}} \hat{\mathrm{a}}_{\mathrm{i}} \mathrm{X}_{\mathrm{t}+1-\mathrm{i}} + N(\hat{e}, \widehat{\sigma})\,. \tag{7}$$

Based on the predictive uncertain variable, Yang and Liu suggested defining the forecast value as the expected value of X_{t+1}, namely,

$$\hat{\mu} = \hat{a}_0 + \sum_{\mathrm{i}=1}^{\mathrm{k}} \hat{\mathrm{a}}_{\mathrm{i}} \mathrm{X}_{\mathrm{t}+1-\mathrm{i}} + \hat{e}. \tag{8}$$

When forecasting future values, calculating a confidence interval is also important because it provides the uncertainty range of the forecast. In uncertain time series analysis, the calculation of the confidence interval takes the characteristics of uncertain variables into account. According to Equations (7) and (8) and the operational laws of uncertain variables, the predictive uncertain variable X_{t+1} has the normal uncertainty distribution $N(\hat{e}, \widehat{\sigma})$. Yang and Liu proposed the following confidence interval for X_{t+1} at confidence level α:

$$\hat{\mu} \pm \frac{\widehat{\sigma}\sqrt{3}}{\pi} \ln \frac{1+\alpha}{1-\alpha}. \tag{9}$$

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.9.1.P9

In uncertain time series analysis, future uncertain variables are forecast using the estimated model and parameters. Given an uncertain autoregressive model, we can forecast the number of retweets at a future time point and quantify the uncertainty associated with the forecast.

3. Data Analysis

The China Internet Joint Rumor-Refutation Platform is a national-level public service initiative established to address the challenges posed by rumor propagation in the information age. By integrating the efforts of government agencies, media organizations, professional institutions, and the public, the platform identifies, clarifies, and corrects online rumors in a timely manner, thereby helping to maintain a clean online environment and the authenticity of public information. It also employs advanced technologies such as artificial intelligence and big-data analytics to monitor and analyze the propagation paths of online rumors, effectively curb the spread of false information, safeguard the public's right to know, and enhance cyberspace security.

Each retweet of a post from the platform's official Weibo account creates another opportunity for information diffusion, allowing rumor-refutation information to reach a broader audience. Retweet volume is therefore one of the key indicators for measuring the influence and dissemination effectiveness of rumor refutation. A high retweet volume not only indicates that users recognize the quality and credibility of the Weibo content, but also reflects the platform's role in guiding online public opinion and improving the public's ability to identify rumors. In addition, analysis of Weibo retweet data can provide feedback to platform operators by showing which types of rumor-refutation information are more popular, thereby helping them optimize content strategies and improve the targeting and effectiveness of rumor-refutation efforts. The dataset used in this study was collected from the official Weibo account of the China Internet Joint Rumor-Refutation Platform and covers the daily average retweet volume over 182 days from January to June 2024.

Table 1. Daily Average Retweet Volume of the China Internet Joint Rumor-Refutation Platform Weibo Account (2024.1.1–2024.6.30)

21	102	149	117	73	112	0	82	246	306
144	157	13	14	167	157	121	126	109	115
90	208	129	81	209	158	14	13	151	468
231	118	116	46	159	96	256	125	217	0
0	0	13	0	0	13	0	0	165	228
135	122	128	158	10	16	180	136	177	147
112	13	25	183	202	144	131	166	8	31
138	110	221	246	184	13	16	154	115	410
190	121	12	19	142	267	133	99	170	17
21	221	266	125	0	0	0	150	189	291
161	106	114	15	22	255	172	117	150	83
12	27	140	338	289	261	273	21	195	284
160	37	150	55	18	0	210	161	468	335
159	128	20	250	171	331	152	114	14	56
243	125	147	107	105	11	18	302	183	129
126	189	14	17	187	152	475	410	198	0
130	17	277	267	161	153	18	23	210	157
157	162	152	13	16	196	170	112	202	149
18	23								

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.9.1.P9

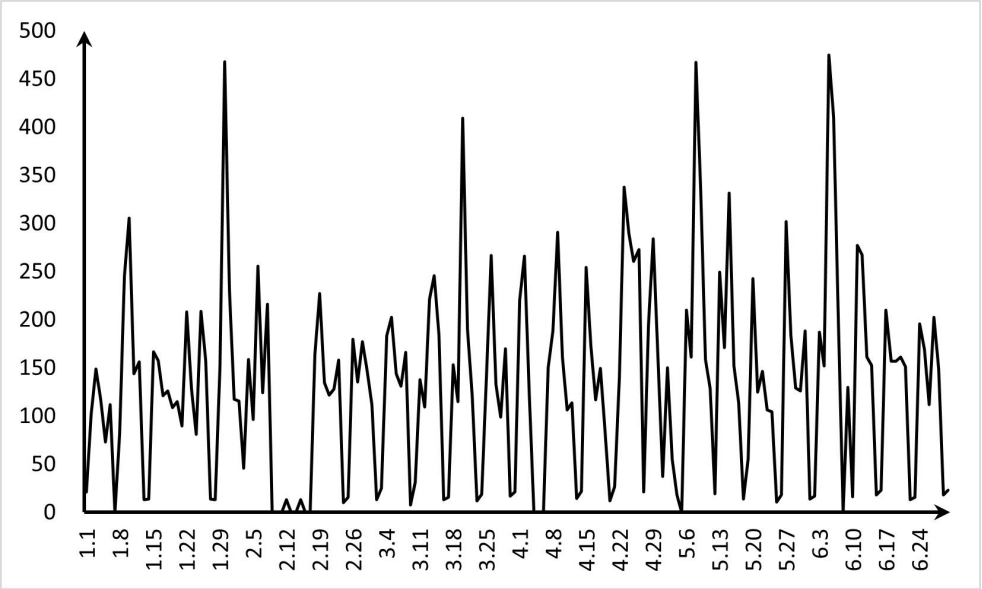


Fig.1 Daily Retweet Volume of the China Internet Joint Rumor-Refutation Platform Weibo Account (2024.1.1–2024.6.30)

We manually collected daily retweet data from the platform's official Weibo account. Because of the public-service nature of the account, some posts were unrelated to rumor refutation. Examples include "Starting today, these new regulations will affect your life" (2024.1.1), "Stay warm and protect yourself from the cold during Minor Cold" (2024.1.6), and "In summer, which foods should not be left overnight?" (2024.6.30), which are science-popularization posts or seasonal and holiday greetings. Each post was reviewed individually to determine whether it contained rumor-refutation information. Posts that did not were excluded from the retweet calculation to ensure the integrity and accuracy of the data. The retweet volumes are shown in Tab.1 and Fig.1. It should be noted that the dataset contains several consecutive days with values of 0, such as February 9–17 and April 4–6, 2024. Further inspection showed that these periods coincided with the Spring Festival and Qingming Festival holidays, during which the platform did not publish rumor-refutation posts for special reasons. We also examined the other dates with a daily average retweet volume of 0 (2024.1.7, 2024.5.5, and 2024.6.8) and found that the platform likewise did not publish any rumor-refutation posts on those days; the zero values therefore do not mean that rumor-refutation posts received zero retweets.

During the 182 days in the first half of 2024, the China Internet Joint Rumor-Refutation Platform published 525 rumor-refutation Weibo posts, which received a total of 80,440 retweets, averaging 153 retweets per post. Combined with the daily average retweet volume, these results indicate that the platform's overall retweet volume was relatively low. Although the China Internet Joint Rumor-Refutation Platform is one of China's official national rumor-refutation platforms, it did not receive broad attention from internet users. The maximum daily average retweet volume was 475. Examination of the corresponding Weibo posts showed that a major social event occurred on that day, and the views of one rumor-refutation post increased sharply, leading to a higher retweet volume.

4. Uncertain Time Series Analysis

This section applies uncertain time series analysis to the retweet-volume data from the official Weibo account of the China Internet Joint Rumor-Refutation Platform. Through this analysis, we aim to evaluate the influence of the Weibo account and forecast the uncertainty in its retweet volume. Let $t = 1,2, ..., 182$ represent the 182 days from January 1 to June 30, 2024. The observations in Table 1 are denoted by X_t, representing the Weibo retweet volume on day t. To model the daily average retweet volume of the rumor-refutation platform, we adopt the following uncertain autoregressive model:

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.9.1.P9

$$X_t = a_0 + \sum_{i=1}^{k} a_i X_{t-i} + \varepsilon, \tag{10}$$

k is the order to be determined, $a_0, a_1, ..., a_k$ are the unknown parameters to be estimated, and ε is the uncertain disturbance term.

4.1 Determining the Order of the Uncertain Autoregressive Model

To determine the order k in the uncertain autoregressive model (10), we use the fixed-origin cross-validation method proposed by Liu and Yang. Setting the maximum order to 10 and the training-set length to 22 yields the average testing errors (ATEs) corresponding to different orders, as shown in Tab.2.

Since ATE(1) is the smallest, we select order k = 1. Therefore, we use the following uncertain autoregressive model:

$$X_t = a_0 + a_1 X_{t-1} + \varepsilon, \tag{11}$$

where a_0 and a_1 are the unknown parameters to be estimated, and ε is the uncertain disturbance term.

Table 2. ATE Cross-Validation

k	ATE(k)	k	ATE(k)	k	ATE(k)	k	ATE(k)	k	ATE(k)
1	13598	2	21256	3	533748	4	223547	5	8825955
6	8672024	7	9599829	8	25015510	9	38981273	10	47162743

4.2 Parameter Estimation

After selecting an appropriate order for the uncertain autoregressive model, the next step is to estimate the model parameters. The objective of parameter estimation is to identify a set of parameter values that minimizes the difference between the retweet volume predicted by the model and the observed retweet volume. Based on the observed data, we use Python to solve the following minimization problem:

$$\min_{a_0,a_1} \sum_{t=1}^{182} \left(X_t - a_0 - \sum_{i=1}^{k} a_i X_{t-i} \right)^2,$$

The least-squares estimates of a_0 and a_1 are

$$\hat{a}_0 = 103.2, \hat{a}_1 = 0.428,$$

Therefore, the fitted autoregressive model is

$$X_t = 103.2 + 0.428 X_{t-1}. \tag{12}$$

4.3 Residual Analysis

According to the fitted autoregressive model (12), we have

$$\varepsilon_t = X_t - 103.2 - 0.428 X_{t-1},$$

The 175 residuals obtained are shown in Figure 2. If the disturbance term ε in model (11) is regarded as a random variable, then, in the probabilistic sense, the residuals shown in **Fig.2** should come from the same population. As can be seen from the figure, the residuals range from -200 to 400, indicating substantial deviations between the model-predicted values and the actual retweet volumes at different time points. Although the absolute residuals vary considerably, no obvious overall trend is observed, suggesting that the model has no systematic bias and does not consistently overestimate or underestimate the retweet volume.

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.9.1.P9

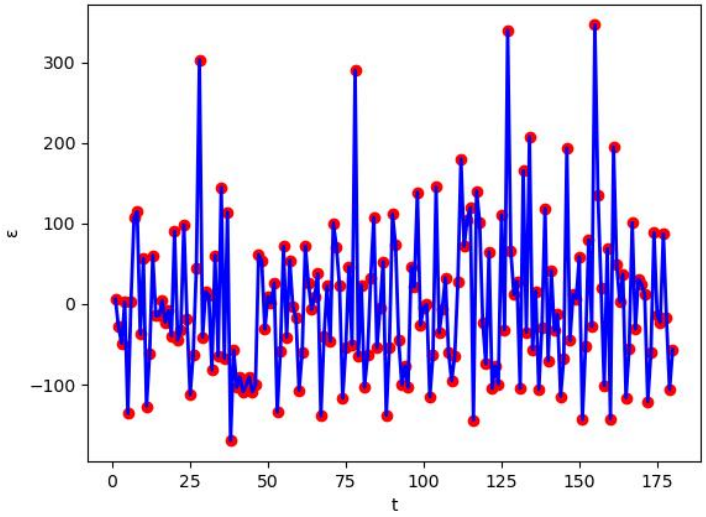


Fig.2 Residuals of the Uncertain Autoregressive Model (13)

We then calculate the expected value of the uncertain disturbance term ε as

$$\hat{e} = \frac{1}{176}\sum_{1}^{182} \varepsilon_t = 0.0000,$$

The standard deviation is

$$\hat{\sigma}^2 = \frac{1}{176}\sum_{1}^{182} (\varepsilon_t - \hat{e})^2 = 92.795^2.$$

Therefore, we obtain the following uncertain autoregressive model:

$$X_t = 103.2 + 0.428X_{t-1} + N(0,92.795). \quad (13)$$

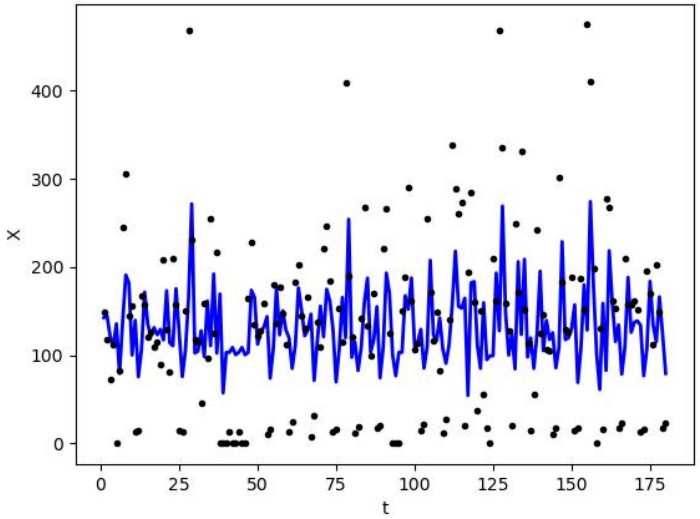


Fig.3 Fitted Uncertain Autoregressive Model (13)

The fitted-model plot of the uncertain autoregressive model (Fig.3) provides an intuitive view of the results obtained by applying the model to the rumor-refutation platform's retweet data. The X-axis represents the time points in the series, while the Y-axis represents the observed retweet volume at those time points. The scatter points show the actual retweet data, whereas the continuous line depicts the retweet volume predicted by the model using the estimated parameters. As shown in the figure, the fitted curve generally follows the fluctuations in the data points, although deviations occur on several days. In addition, the variability of the data points suggests that retweet volume may exhibit uncertainty or periodic patterns, which may be related to user behavior, the dynamics of information dissemination, or other external factors.

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.9.1.P9

4.4 Uncertain Hypothesis Testing

After parameter estimation and residual analysis, we conduct a hypothesis test to evaluate whether the residuals conform to a normal uncertainty distribution; the results are shown in Fig.4. In the hypothesis test, the significance level is set to $\alpha = 0.05$ to determine whether the residuals fall within an acceptable uncertainty range.

By calculation, we obtain the mean and standard deviation of the residuals and then determine the test interval. The results show that most residuals fall within the calculated test interval [-187.4298, 187.4298], with only 7 data points outside this interval. This result is consistent with the specified significance level, indicating that the model passes the hypothesis test. In addition, the standard deviation of the disturbance term ε is 92.795. Although some fluctuation exists, this value is acceptable given the inherent uncertainty of social media data.

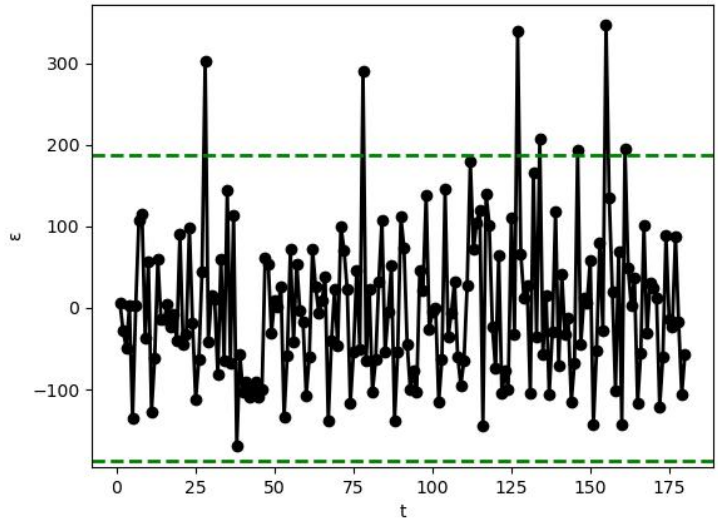


Fig.4 Hypothesis Test for the Uncertain Autoregressive Model (13)

4.5 Forecasting

Based on the uncertain autoregressive model (13), the uncertain variable for forecasting the retweet volume of the rumor-refutation platform on July 1, 2024 is

$$\widehat{X}_t = 103.2 + 0.428 \times 18 + N(0,92.795),$$

$\widehat{X}_t \sim N(109,92.795)$. The retweet volume of the rumor-refutation platform is 327, and the 95% confidence interval is

$$109 \pm \frac{92.795\sqrt{3}}{\pi} \ln \frac{1+0.95}{1-0.95},$$
$$109 \pm 270 = [-161,379].$$

Because the predicted quantity is retweet volume, the final 95% confidence interval is $[0,379]$. We then retrieved the retweet volume of the rumor-refutation Weibo posts published on July 1 from the China Internet Joint Rumor-Refutation Platform and calculated a daily average retweet volume of 159, which falls within the 95% confidence interval.

5. Conclusion

This study conducted detailed modeling and analysis of retweet-volume data from the official Weibo account of the China Internet Joint Rumor-Refutation Platform using uncertain time series analysis. Based on retweet data from January to June 2024, the study not only revealed the influence of the Weibo account in rumor refutation, but also forecast the time-series characteristics of its retweet

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.9.1.P9
volume, thereby providing data support and a decision-making basis for the platform's information-dissemination strategy.

First, this study constructed an uncertain autoregressive model to capture the dynamic characteristics of the time-series data. The parameter-estimation results show that the selected model can reasonably fit the observed data, and the residual analysis further verifies the effectiveness of the model. The optimal model order determined through cross-validation provides a robust parameter basis for the uncertain time-series model. In the forecasting stage, the uncertain time-series model was used to forecast the retweet volume of the Weibo account and to provide a prediction interval. The accuracy and reliability of the forecasting results were supported by the hypothesis test: most residuals fell within the expected uncertainty range, indicating that the model can explain the uncertainty in the data well.

Despite the positive results, this study has several limitations. For example, the model may not fully account for all factors that affect Weibo retweet volume. Future research could incorporate additional explanatory variables, such as user behavior and content characteristics. In addition, an in-depth investigation of the outliers identified in the residual analysis could help reveal potential abnormal patterns or the effects of special events, thereby supporting further model adjustment and improvement. Because user interests and behavioral patterns may change over time, model parameters should be updated regularly to adapt to new data trends. Developing a real-time forecasting system for dynamic monitoring and immediate feedback on Weibo retweet volume would provide more flexible and responsive decision support for platform operations. At the same time, evaluating the effects of different content-publishing strategies based on forecasting results could provide guidance for practical information-dissemination activities.

References

[1] B. Liu, Uncertainty Theory, 2nd edn. (Springer, Berlin, 2007).

[2] B. Liu, Some research problems in uncertainty theory, J. Uncertain Syst. 3(1) (2009)3–10.

[3] B. Liu, Uncertain urn problems and Ellsberg experiment, Soft Comput. 23(15) (2019) 6579–6584.

[4] B. Liu, Uncertainty Theory: A Branch of Mathematics for Modeling Human Uncertainty (Springer, Berlin, 2010).

[5] X. F. Yang and B. Liu, Uncertain time series analysis with imprecise observations, Fuzzy Optim. Decis. Mak. 18(3) (2019) 263–278.

[6] D. Chen and X. F. Yang, Maximum likelihood estimation for uncertain autoregressive model with application to carbon dioxide emissions, J. Intell. Fuzzy Syst. 40(1) (2021)1391–1399.

[7] Y. Liu and B. Liu, A modified uncertain maximum likelihood estimation with applications in uncertain statistics, Technical Report (2022).

[8] T. Q. Ye and X. F. Yang, Analysis and prediction of confirmed COVID-19 cases in China with uncertain time series, Fuzzy Optim. Decis. Mak. 20(2) (2021) 209–228.