Chapter 2 INTERVAL TIME SERIES ANALYSIS AND FORECASTING

2.1 INTRODUCTION

Researchers have been exploring the subject of how to forecast the trends from excessive information. The analysis and forecasting of time series are extensively utilized in a variety of applications. The traditional time series analysis selects the best suitable model from *a priori* models, such as *ARIMA* model, *ARCH* model or the threshold model *etc* [5]. Considering the uncertainty of the predicted points, interval data are used to estimate the prediction values. Montgomery and Johnson [22], Abraham and Ledolter [1], Chatfield [7] proposed the interval prediction methods by using a traditional time series to perform the prediction. Because of the diversity of research backgrounds and purposes, there are various methods of interval forecasting. Nevertheless, the collection of data typically is on the basic form of single numerical values.

In a practical case, uncertain or incomplete factors might interfere with the data collection so that the observed single-valued (real number) samples cannot fully describe the true situations of the sample from the population. Consequently, this chapter attempts to utilize interval data to perform the interval forecasting. In the studies of time series forecasting and analysis, the data of interval form are paid more and more attention, such as daily temperature changes, the fluctuation of the exchange rates, the level prices of petroleum *etc*. However, it is quite difficult to calculate the traditional forecasting by interval operations. Nguyen and Wu [23] have indicated that intervals have the fuzzy characteristic due to their uncertainty. It is a challenging work to define a forecasting criterion from the concept of fuzzy numbers. For example, Kubo *et al* [18] proposed an integer ambiguity estimation and validation method in carrier phase GPS positioning, and Wu and Tseng [33] applied the fuzzy regression models with application to business cycle analysis.

This chapter applies fuzzy set theory (Zimmermann [37]) to perform the analysis and the forecasting of interval data. As the forecasting methods are being improved, it is noteworthy that bi-directional computing architecture has been applied to the time series forecasting (Wakuya [28]). We proposes three interval forecasting models all employing the bi-directional computation style. The proposed interval forecasting methods are the interval moving average (*IMA*) of order *k*, the weighted interval moving average (*WIMA*) of order *k*, and the *ARIMA* interval forecasting, respectively. The *IMA* approach uses the traditional method to forecast the interval data; the *WIMA* approach regards interval lengths as fuzzy weights so as to amend the *IMA* forecasting model to more coincide with actual situations; the *ARIMA* interval forecasting constructs an *ARIMA* model by interval variables, which is a breakthrough in the development of forecasting methods.

The proposed interval time series is different from the fuzzy time series proposed by Wu and Hsu [32]. Since the fuzzy time series is a method combining linguistic variables with the analysis process which applies fuzzy logic into time series to solve the fuzziness of data, the predicted values are not real values but the linguistic numbers. On the other hand, the interval time series analyzes interval data by the concept of fuzzy number into the mean values, the right lengths, and the left lengths of intervals. Therefore, it constructs a multi-dimensional forecasting model and its forecast result is still in the form of interval. For example, the forecast results of the fuzzy time series in the stock market are merely outputting linguistic variables such as "*plunge*", "*drop*", "*draw*", "*soar*" and "*surge*". Nevertheless, the interval time series could tell us how much the stock would fluctuate. Consequently, the interval time series provides more precise and more objective forecast results than the traditional fuzzy time series does.

Since the forecasting methods are designed, it is necessary to analyze the validity of the forecasting methods by means of the estimated errors between the forecast data and the actual data. Chatfield [9] declared that the error made by an inappropriate interval prediction method is more severe than the error made by a simple point prediction. Therefore, in order to assess the efficiency of an interval forecasting, this chapter also defines several criteria, referring to two kinds of the fuzzy distances which were proposed by Yang and Ko [34], and Hébert *et al* [16] respectively, to evaluate the efficiency of forecasting. An integrated analysis of the forecasting efficiency is formulated by incorporating the position and length of the interval, which are the mean squared error of interval and the mean relative interval error respectively.

In order to demonstrate the proposed forecasting methods, four sets of stable and unstable interval time series are simulated by the methods of *AR*(1) and *ARCH*(1) for the use of analyzing the efficiency of the proposed forecasting methods. Besides, the monthly highest and lowest prices of stocks are also used as a practical case study in the end of this chapter. Regardless of the simulated interval time series or the practical data, using *ARIMA* to achieve the interval forecasting is more appropriate than using traditional technologies.

2.2 THE CHARACTERISTICS OF INTERVAL DATA

2.2.1 *Why Using Interval?*

The traditional social and economic studies have brought in various analyses of interactive relationships and models which are related to human. In the traditional model construction, it often confronts the uncertainty problem of data. For instance, should it count the number of yearly enrolled students at the beginning, the middle, or the end of a year? The obtained numbers are often different at different time points. Wu and Chen [31] have given an extensive review of literature on this issue. In the social science study or economic research, obviously the answer to these questions is not just true or false. There are lots of uncertain and incomplete information or events so that we can not apply the conventional real number system to process it. Due to the influences of various factors in many practical cases, the observed data usually appear not only in the type of single numerical value but also in a "*range*" of numerical values. As well as a value of interval type is capable of representing dynamic incidents. By use of the continuity characteristic of interval value, it can make analysts capable of dealing with the uncertainty of factors. Consequently, an interval is indeed a better measure tallying with the actual situation in practical applications.

2.2.2 *An Interval as a Fuzzy Number*

Since the data of interval type are considered in this chapter, it must encounter various problems of interval operations as well as the realistic meanings. Besides, it is unable to give the standard rules of interval operations on computer hardware. Hayes [15] pointed out that the rules of interval operations seem simple, but there often appears a trap of miscalculation in the practical calculations. However, each interval can be explained as a set of possible values for the actual unknown number. This feature of interval coincides to the fuzzy theory. Hence, the viewpoint of fuzzy can be applied to describe intervals and thus a basic definition of a fuzzy number is described as follows [15].

Definition 2.1 *Fuzzy numbers*

A fuzzy number $\widetilde{M}$ *is a fuzzy subset M of real number* $\Re$ *such that:*

(1) $0 \le \mu_{\widetilde{M}}(x) \le 1$*;*

(2) The support $\left\{ x : \mu_{\widetilde{M}}(x) > 0 \right\}$ *of* $\widetilde{M}$ *is bounded;*

(3) The α*–level set* M_α *of M are closed interval.*

where $\mu_{\widetilde{M}}(x)$ *is a membership function.*

Fuzzy numbers are special fuzzy quantities. An interval is a special case of fuzzy numbers. By given a membership function, an interval can be considered as a fuzzy number.

Then by using the fuzzy expression for interval data, appropriate operations of intervals could be defined in accordance with the fuzzy theorem. With the assistance of computer programs, even more complicated calculation can be easily solved. As a result, before constructing and forecasting a model of interval time series, several definitions relevant to interval are given first. In order to consider interval data as the *LR*-representation of fuzzy numbers, we use the definition of a fuzzy number given by Zimmermann [37] to define the *LR*-type interval data.

Definition 2.2 *LR-type interval data,* $\mathbf{X} = (m,l,u)_{LR}$

An interval data $\mathbf{X}$ *with lower boundary a and upper boundary b which is denoted as* $\mathbf{X} = [a,b]$ *is of LR-type if there exist two decreasing shape functions* L:$[0,\infty]\to[0,1]$ *and* R:$[0,\infty]\to[0,1]$ *with the membership function*

$$\mu_X(x) = \begin{cases} L\left(\dfrac{m-x}{l}\right) & \text{for } x \le m \\ R\left(\dfrac{x-m}{u}\right) & \text{for } x \ge m \end{cases}$$

where the real number m is the mean value of $\mathbf{X}$*,* $l = m - a$*, and* $u = b - m$*. Therefore,* $\mathbf{X}$ *is denoted by* $(m,l,u)_{LR}$.

Example 2.1 Let $\mathbf{X} = (2, 1, 4)_{LR}$, $L(x) = 1 - x$, and $R(x) = exp(-x)$, then

$$\mu_X(x) = \begin{cases} 1-\dfrac{2-x}{1} & \text{for } 1 \le x \le 2. \\ \exp\left(-\dfrac{x-2}{4}\right) & \text{for } 2 \le x \le 6. \end{cases}$$

is the membership function which is shown in Figure 2.1. ♦

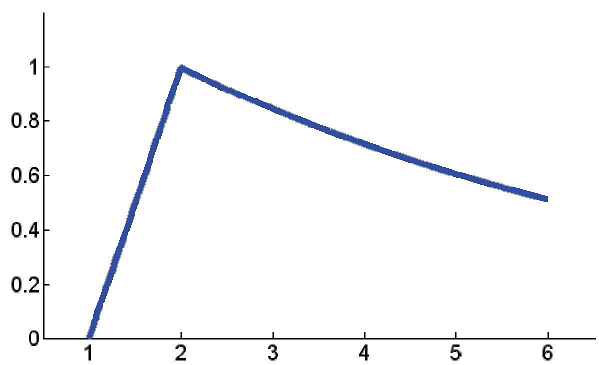

Figure 2.1 LR-type interval $\mathbf{X}$.

Definition 2.3 *The length of interval,* $\|\boldsymbol{X}\|$

Let $\boldsymbol{X} = (m, l, u)_{LR}$ *be an interval data, the length of interval* $\boldsymbol{X}$ *is* $\|X\|_{Fuzzy} = l + u$*, the right length of interval* $\boldsymbol{X}$ *is* $\|X\|_R = u$*, and the left length of interval* $\boldsymbol{X}$ *is* $\|X\|_L = l$*. For simplification, the length of interval* $\boldsymbol{X}$ *is denoted as* $\|X\|$ *instead of* $\|X\|_{Fuzzy}$*.*

Definition 2.4 *The operation of interval data*

Let $X_1 = (m_1, l_1, u_1)_{LR}$ *and* $X_2 = (m_2, l_2, u_2)_{LR}$ *be interval data. The interval addition, scalar multiplication and interval subtraction are defined as follows:*

Interval addition:

$$X_1 \oplus X_2 = (m_1, l_1, u_1)_{LR} \oplus (m_2, l_2, u_2)_{LR} = (m_1 + m_2, l_1 + l_2, u_1 + u_2)_{LR}.$$

Scalar multiplication:

$$k\boldsymbol{X} = k(m, l, u)_{LR} = \begin{cases} (km, kl, ku)_{LR} & \text{for } k > 0 \\ (km, |k|u, |k|l)_{LR} & \text{for } k < 0 \end{cases}, \text{ where } k \text{ is a scalar.}$$

Interval subtraction:

$$\begin{aligned} X_1 \ominus X_2 = X_1 \oplus (-X_2) &= (m_1, l_1, u_1)_{LR} \oplus (-m_2, u_2, l_2)_{LR} \\ &= (m_1 - m_2, l_1 + u_2, u_1 + l_2)_{LR} \end{aligned}$$

Because interval data have the behavior of fuzzy numbers, the operations of interval data are identical with the fuzzy procedures. Here it must be clarified that the subtraction of two interval data, $X_1 \ominus X_2$, is totally different from the subtraction of two real data, $x_1 - x_2$. The subtraction of interval data is an interval which is merely the difference of two intervals, but the subtraction of real data can be viewed as the distance of two points. The following define the distance between two interval data.

Definition 2.5 *The distance between two interval,* $D(\boldsymbol{X}_1, \boldsymbol{X}_2)$

Let $\boldsymbol{X}_1 = (m_1, l_1, u_1)_{LR}$ *and* $\boldsymbol{X}_2 = (m_2, l_2, u_2)_{LR}$ *be two interval data, then the distance between* $\boldsymbol{X}_1$ *and* $\boldsymbol{X}_2$ *is defined as*

$$D(\boldsymbol{X}_1, \boldsymbol{X}_2) = \left((m_1 - m_2)^2 + ((m_1 - l_1) - (m_2 - l_2))^2 + ((m_1 + u_1) - (m_2 + u_2))^2\right)^{1/2}.$$

Example 2.2 Let $\boldsymbol{X}_1 = (2, 1, 1)_{LR}$, $\boldsymbol{X}_1 = (3, 1, 3)_{LR}$, then

$$\boldsymbol{X}_1 \oplus \boldsymbol{X}_2 = (2,1,1)_{LR} \oplus (3,1,3)_{LR} = (2+3, 1+1, 1+3)_{LR} = (5, 2, 4)_{LR},$$

$$3\boldsymbol{X}_1 = 3(2,1,1)_{LR} = (3\cdot 2, 3\cdot 1, 3\cdot 1)_{LR} = (6,3,3)_{LR},$$

$$\boldsymbol{X}_1 \Theta \boldsymbol{X}_2 = (2, 1, 1)_{LR} \Theta (3, 1, 3)_{LR} = (2-3, 1+3, 1+1)_{LR} = (-1, 4, 2)_{LR},$$

$$D(\boldsymbol{X}_1, \boldsymbol{X}_2) = \sqrt{(2-3)^2 + (1-2)^2 + (3-6)^2} = \sqrt{11}. \qquad \blacklozenge$$

Since the types of membership functions are various, appropriate membership functions are chosen to advance the accuracy of the distance between intervals. In the dissertation, we assume that interval data have the characteristics of fuzzy number and their membership functions perform almost flat. For the reason that the interval data adopt the uniform membership function. Then interval data $\boldsymbol{X}$ with lower boundary a and upper boundary b is a dynamic range which is denoted as $\boldsymbol{X} = [\, a, b \,]$, we may collect data in real cases to get a and b. After that, l and u are obtained respectively.

Definition 2.6 *Interval time series*

An interval time series is a sequence of interval data, $\mathbf{X}_t = [a_t, b_t] = (m_t, l_t, u_t)_{LR}$, $t = 1,2,3,\ldots$, *denoted as* $\{\mathbf{X}_t\} = \{\mathbf{X}_t = [a_t, b_t] = (m_t, l_t, u_t)_{LR} \mid t = 1,2,3,\ldots\}$.

2.2.3 The Forecasting Models of Interval Time Series

The interval time series is an analytical method which applies intervals to the analysis of time series incorporating with the interval operations so as to solve the uncertainty of the data. A traditional time series is defined as $\{X_t = x_t, t = 1,2,3,\ldots\}$, and the corresponding prediction $\hat{X}_t = E[X_t | X_{t-1}, X_{t-2}, \ldots, X_1]$ is a point prediction. The prediction model for a traditional time series is not capable of being applied directly to the forecasting of interval time series $\{\mathbf{X}_t\} = \{\mathbf{X}_t = [a_t, b_t] = (m_t, l_t, u_t)_{LR} \mid t = 1,2,3,\ldots\}$. The following introduces some forecasting models of interval time series.

(1) *Interval moving average of order k (IMA)*

Let $\hat{m}_t = \dfrac{m_{t-1} + \ldots + m_{t-k}}{k}$, $\hat{l}_t = \dfrac{l_{t-1} + \cdots + l_{t-k}}{k}$, $\hat{u}_t = \dfrac{u_{t-1} + \cdots + u_{t-k}}{k}$, $t = k+1, k+2, k+3, \ldots$, then the forecasting of interval time series is $\hat{\mathbf{X}}_t = E[\mathbf{X}_t \mid \mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \ldots, \mathbf{X}_{t-k}] = (\hat{m}_t, \hat{l}_t, \hat{u}_t)_{LR}$.

(2) *Weighted interval moving average of order k (WIMA)*

Let $\hat{m}_t = \sum_{i=t-k}^{t-1} p_i m_i$, $\hat{l}_t = \sum_{i=t-k}^{t-1} p_{Li} l_i$, and $\hat{u}_t = \sum_{i=t-k}^{t-1} p_{Ri} u_i$, for $t = k+1, k+2, k+3, \ldots$, where $p_i = \dfrac{f_i}{\sum_{j=t-k}^{t-1} f_j}$, $f_i = \dfrac{\|\boldsymbol{\Omega}\|}{\|\boldsymbol{X}_i\|}$, $p_{Li} = \dfrac{\|\mathbf{X}_i\|_L}{\sum_{j=t-k}^{t-1} \|\mathbf{X}_j\|_L}$, and $p_{Ri} = \dfrac{\|\mathbf{X}_i\|_R}{\sum_{j=t-k}^{t-1} \|\mathbf{X}_j\|_R}$, for $i = t-k, t-k+1, \ldots, t-1$, and $\boldsymbol{\Omega} = \left[\min_{t-k \le j \le t-1} \{a_j\}, \max_{t-k \le j \le t-1} \{b_j\} \right]$. Then the forecasting of interval time series is $\hat{\mathbf{X}}_t = E[\mathbf{X}_t \mid \mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \ldots, \mathbf{X}_{t-k}] = (\hat{m}_t, \hat{l}_t, \hat{u}_t)_{LR}$.

In fact, this method improves the *IMA*. Since the larger the interval is, the more information it will contain. So we will give l_i and u_i in the estimators of $\hat{l}_t$ and $\hat{u}_t$ larger weights which are $p_{Li} = \|\mathbf{X}_i\|_L \Big/ \sum_{j=t-k}^{t-1} \|\mathbf{X}_j\|_L$ and $p_{Ri} = \|\mathbf{X}_i\|_R \Big/ \sum_{j=t-k}^{t-1} \|\mathbf{X}_j\|_R$ respectively. While from the scale point of view, too large length of an interval will reduce the importance of location of mean value. That is why we use $f_i = \|\boldsymbol{\Omega}\| / \|\mathbf{X}_i\|$ to decrease its weight.

(3) *ARIMA interval forecasting (ARIMA)*

$\{m_t\}$, $\{l_t\}$, and $\{u_t\}$ are time series resulted from $ARIMA(p_m, d_m, q_m)$, $ARIMA(p_l, d_l, q_l)$, and $ARIMA(p_u, d_u, q_u)$ models respectively by

$$\left(1 - \sum\nolimits_{i=1}^{p} \phi_i L^i\right)(1-L)^d m_t = \left(1 + \sum\nolimits_{i=1}^{q} \theta_i L^i\right) \varepsilon_{m,t}$$

$$\left(1 - \sum\nolimits_{i=1}^{p} \beta_i L^i\right)(1-L)^d l_t = \left(1 + \sum\nolimits_{i=1}^{q} \alpha_i L^i\right) \varepsilon_{l,t}$$

$$\left(1 - \sum\nolimits_{i=1}^{p} \varphi_i L^i\right)(1-L)^d u_t = \left(1 + \sum\nolimits_{i=1}^{q} \psi_i L^i\right) \varepsilon_{u,t}$$

where $\varepsilon_{m,t} \sim WN(0, \sigma_m^2)$, $\varepsilon_{l,t} \sim WN(0, \sigma_l^2)$, and $\varepsilon_{u,t} \sim WN(0, \sigma_u^2)$. The t–step advanced forecasts can be obtained by

$$\hat{m}_t = E\left[m_t \,|\, m_{t-1}, m_{t-2}, \ldots, m_{t-k}\right]$$

$$\hat{l}_t = E\left[l_t \,|\, l_{t-1}, l_{t-2}, \ldots, l_{t-k}\right]$$

$$\hat{u}_t = E\left[u_t \,|\, u_{t-1}, u_{t-2}, \ldots, u_{t-k}\right]$$

Then the l–step advanced forecasting of interval time series is $\hat{\mathbf{X}}_t = \left(\hat{m}_t, \hat{l}_t, \hat{u}_t\right)_{LR}$.

Example 2.3 Let $\{X_t\} = \{[1, 2], [2, 4], [3, 4], [4, 6], [3, 7]\}$ and $l = u$. To be precise, $\boldsymbol{X}_1 = (1.5, 0.5, 0.5)_{LR}$, $\boldsymbol{X}_2 = (3, 1, 1)_{LR}$, $\boldsymbol{X}_3 = (3.5, 0.5, 0.5)_{LR}$, $\boldsymbol{X}_4 = (5, 1, 1)_{LR}$, and $\boldsymbol{X}_5 = (5, 2, 2)_{LR}$. By the interval moving average of order 5, it can be obtained by

$$\begin{aligned}\hat{\mathbf{X}}_6 &= \left(\frac{1.5+3+3.5+5+5}{5}, \frac{0.5+1+0.5+1+2}{5}, \frac{0.5+1+0.5+1+2}{5}\right)_{LR} \\ &= (3.6, 1, 1)_{LR} = [3.6-1, 3.6+1] = [2.6, 4.6].\end{aligned}$$

Since $\boldsymbol{\Omega} = [1, 7]$, the full length is $\|\boldsymbol{\Omega}\| = 6$, and $f_1 = \frac{6}{1}$, $f_2 = \frac{6}{2}$, $f_3 = \frac{6}{1}$, $f_4 = \frac{6}{2}$, $f_5 = \frac{6}{4}$, we have $p_1 = \dfrac{\frac{6}{1}}{\frac{6}{1}+\frac{6}{2}+\frac{6}{1}+\frac{6}{2}+\frac{6}{4}} = 0.31$, $p_2 = 0.15$, $p_3 = 0.31$, $p_4 = 0.15$, $p_5 = 0.08$. Then $\hat{m}_6 = 0.31 \times 1.5 + 0.15 \times 3 + 0.31 \times 3.5 + 0.15 \times 5 + 0.08 \times 0.5 = 2.79$. Because of $l = u$, $p_{L1} = p_{R1} = 0.5/(0.5 + 1 + 0.5 + 1 + 2) = 0.1$, $p_{L2} = p_{R2} = 0.2$, $p_{L3} = p_{R3} = 0.1$, $p_{L4} = p_{R4} = 0.2$, and $p_{L5} = p_{R5} = 0.4$. Then $\hat{l}_6 = \hat{u}_6 = 0.1 \times 0.5 + 0.2 \times 1 + 0.1 \times 0.5 + 0.2 \times 1 + 0.4 \times 2 = 1.3$. By the means of the weighted interval moving average of order 5, we have

$$\hat{\boldsymbol{X}}_6 = (2.79, 1.3, 1.3)_{LR} = [1.69, 4.09].$$ ♦

2.3 The Efficiency Analysis of Interval Time Series Forecasting

The reliability of the forecast interval is the most concern of the analysts. In a traditional forecasting of time series, it is to compare the distance between the actual value and the predicted value to assess the reliability of forecasting. With regard to the interval forecasting, not only the forecasting of interval lengths, l and u, we are also concerned with the position disparity, m, between the forecast interval and the actual interval. Therefore, the traditional methods for evaluating the forecasting efficiency of time series are insufficient to analyze the forecasting efficiency of interval time series. Therefore, the following defines the criteria for analyzing the efficiency of interval forecasting.

2.3.1 **The Mean Squared Error of Interval**

The conventional mean squared error is a statistic often used to measure the difference between an estimator and the true value of the quantity to be estimated. The following extends this traditional method to carry out the efficiency analysis of interval time series forecasting.

Definition 2.7 *Mean squared error of interval position (MSEP)*

Let $\{\boldsymbol{X}_t = (m_t, l_t, u_t)_{LR}\}$ *be an interval time series and* $\hat{\boldsymbol{X}}_t = (\hat{m}_t, \hat{l}_t, \hat{u}_t)_{LR}$ *be the forecast interval, then the mean squared error of interval position (MSEP) is defined by*

$$MSEP = \frac{1}{s}\sum_{t=1}^{s}(m_{n+t} - \hat{m}_{n+t})^2$$

where n denotes the current time, s is the number of the preceding intervals, and $\hat{m}_t$, $\hat{l}_t$, *and* $\hat{u}_t$ *are the estimations of* m_t, l_t, *and* u_t *respectively.*

Definition 2.8 *Mean squared error of interval (MSEI)*

Let $\{\boldsymbol{X}_t = (m_t, l_t, u_t)_{LR}\}$ *be an interval time series and* $\hat{\boldsymbol{X}}_t = (\hat{m}_t, \hat{l}_t, \hat{u}_t)_{LR}$ *be the forecast interval, the mean squared error of interval (MSEI) is defined by*

$$\begin{aligned} MSEI &= \frac{1}{s}\sum_{t=1}^{s} D^2\left(\mathbf{X}_{n+t}, \hat{\mathbf{X}}_{n+t}\right) \\ &= \frac{1}{s}\sum_{t=1}^{s}\left((m_{n+t} - \hat{m}_{n+t})^2 + ((m_{n+t} - l_{n+t}) - (\hat{m}_{n+t} - \hat{l}_{n+t}))^2 + ((m_{n+t} + u_{n+t}) - (\hat{m}_{n+t} + \hat{u}_{n+t}))^2\right) \end{aligned}$$

where n denotes the current time, s is the number of the preceding intervals, and $\hat{m}_t$, $\hat{l}_t$, *and* $\hat{u}_t$ *are the estimations of* m_t, l_t, *and* u_t *respectively.*

Remark 3.1

When $l = u$ in Definition 2.8, $\hat{l} = \hat{u}$. Then the mean squared error of interval can be simplified as follows:

$$MSEI = \frac{1}{s}\sum_{t=1}^{s}\left(3(m_{n+t} - \hat{m}_{n+t})^2 + 2(l_{n+t} - \hat{l}_{n+t})^2\right) . \qquad \blacklozenge$$

Example 2.4 Suppose the interval time series are $\boldsymbol{X}_1 = (\ 5,\ 1,\ 1\ )_{LR}$ and $\boldsymbol{X}_2 = (\ 6,\ 1,\ 2\ )_{LR}$, and the forecast intervals are $\hat{\boldsymbol{X}}_1 = (\ 3.5,\ 0.7,\ 1.9\ )_{LR}$ and $\hat{\boldsymbol{X}}_2 = (\ 6,\ 2,\ 1.6\ )_{LR}$. Then the mean squared error of interval position and the mean squared error of interval are given by

$$MSEP = \frac{1}{2}\left((5-3.5)^2 + (6-6)^2\right) = 1.125,$$

$$\begin{aligned} MSEI &= \frac{1}{2}(((5-3.5)^2 + (4-2.8)^2 + (6-5.4)^2) + ((6-6)^2 + (5-4)^2 + (8-7.6)^2)) \\ &= 2.605. \end{aligned} \qquad \blacklozenge$$

2.3.2 **The Mean Relative Interval Error**

While considering the efficiency of interval forecasting, the most important consideration is whether the forecast interval does cover the actual interval. Explicitly speaking, the forecast result is better if not only the mean value $\hat{m}$ of the forecast interval $\hat{\boldsymbol{X}}$ is closer to the mean value m of the actual interval $\boldsymbol{X}$ but also their overlap is larger. For this reason, the following proposes a technique to evaluate forecasting performance for interval data.

Suppose the actual interval data $\boldsymbol{X} = (m,\ l,\ u)_{LR}$ and the forecast interval $\hat{\boldsymbol{X}} = (\hat{m}, \hat{l}, \hat{u})_{LR}$ with $m \le \hat{m}$, then the minimum distance between $\boldsymbol{X}$ and $\hat{\boldsymbol{X}}$ can be defined by

$$d = \min_{x\in \boldsymbol{X},\, \hat{x}\in \hat{\boldsymbol{X}}} |x - \hat{x}| . \tag{1}$$

This definition was introduced by Denœux and Masson [11]. The Eq. (1) can be expanded for the application of interval to derive the following equation:

$$d = \max(0,\ (\hat{m} - m) - (u + \hat{l}\,)). \tag{2}$$

From Eq. (2), if $(\hat{m}-m)-(u+\hat{l})>0$, it means $\boldsymbol{X}$ and $\hat{\boldsymbol{X}}$ are disjoint. On the other hand, if $(\hat{m}-m)-(u+\hat{l})\le 0$, there is an overlap of $\boldsymbol{X}$ and $\hat{\boldsymbol{X}}$. By combining the mean value of the interval and the distances between the mean value and the boundaries of the interval, there are three decision conditions: (1) when $\frac{|m-\hat{m}|}{u+\hat{l}}\le 1$, there is an overlap of the forecast interval and the actual interval, which means that the interval forecasting is good; (2) when $\frac{|m-\hat{m}|}{u+\hat{l}}<<1$, it means that there is a larger overlap so that the interval forecasting is much better; (3) when $\frac{|m-\hat{m}|}{u+\hat{l}}>1$, the forecast interval and the actual interval are completely separated so that the interval forecasting is undesirable. Certainly, when $m\ge\hat{m}$, $\frac{|m-\hat{m}|}{\hat{u}+l}$ can be discussed with the same argument.

Because $\left\|\boldsymbol{X}\ominus\hat{\boldsymbol{X}}\right\|_L = l+\hat{u}$ and $\left\|\boldsymbol{X}\ominus\hat{\boldsymbol{X}}\right\|_R = \hat{l}+u$, $\frac{|m-\hat{m}|}{\left\|\boldsymbol{X}\ominus\hat{\boldsymbol{X}}\right\|_L}$ and $\frac{|m-\hat{m}|}{\left\|\boldsymbol{X}\ominus\hat{\boldsymbol{X}}\right\|_R}$ can be the criteria for evaluating the interval forecasting. Therefore, the following definition is proposed to be another criterion for analyzing the integrated efficiency of interval forecasting.

Definition 2.9 *Mean relative interval error (MRIE)*

Let $\{\boldsymbol{X}_t=(m_t,l_t,u_t)_{LR}\}$ *be an interval time series and* $\hat{\boldsymbol{X}}_t=(\hat{m}_t,\hat{l}_t,\hat{u}_t)_{LR}$ *be the forecast interval, the mean relative interval error (MRIE) is given by*

$$MRIE=\frac{1}{s}\sum_{t=l}^{s}\frac{|m_{n+t}-\hat{m}_{n+t}|}{\left\|\boldsymbol{X}_{n+t}\ominus\hat{\boldsymbol{X}}_{n+t}\right\|_{(*)}} ,$$

where n represents the current time, s is the number of the preceding intervals, and $\hat{m}_t$ *is the estimation of* m_t.

(*) Note: When $m_t\ge\hat{m}_t$, $\left\|\boldsymbol{X}_t\ominus\hat{\boldsymbol{X}}_t\right\|_{(*)}=\left\|\boldsymbol{X}_t\ominus\hat{\boldsymbol{X}}_t\right\|_{(L)}=l_t+\hat{u}_t$; on the other hand, when $m_t<\hat{m}_t$,

$\left\|\boldsymbol{X}_t\ominus\hat{\boldsymbol{X}}_t\right\|_{(*)}=\left\|\boldsymbol{X}_t\ominus\hat{\boldsymbol{X}}_t\right\|_R=u_t+\hat{l}_t$.

Example 2.5 Suppose the interval time series are $\boldsymbol{X}_1 = (4, 1, 1.5)_{LR}$ and $\boldsymbol{X}_2 = (3, 2, 1)_{LR}$, and their forecast intervals are $\hat{\boldsymbol{X}}_1 = (3, 0.7, 1)_{LR}$ and $\hat{\boldsymbol{X}}_2 = (4.2, 1.2, 0.8)_{LR}$, then the mean relative interval error is given by $MRIE = \frac{1}{2}\left(\frac{|4-3|}{1+1} + \frac{|3-4.2|}{1+1.2}\right) = 0.52$. ♦

2.4 SIMULATION ANALYSIS AND DISCUSSIONS

2.4.1 Simulations of Interval Time Series

In the analysis of traditional time series, the data of time series is sampled from the values present at discrete time points. Since time is a continuous variable, the data variation is not known during the time interval between two consecutive samples so that the forecasting by a set of discrete data may be too subjective and biased. Hence, the concept of interval time series is proposed to represent the data collected at any time point in the form of interval with the purpose of forecasting analysis. Taking stock market as an example, if it is desired to make a long-term prediction analysis for a certain stock, the traditional analysis is to take merely the daily closing prices into account for the analysis so that the predicted values will also be single numerical values. It seems to be formally accurate but lack of flexibility. If the daily highest and lowest prices of the stock are considered as the boundary values of an interval for analyzing the trend of the stock prices, the predicted values will be represented in the form of interval too. Therefore, the stock analysts can make comparatively objective decision according to the center position and length of the predicted value.

In the past, there was no forecasting analysis for interval time series as the same for the general time series. In order to analyze the forecasting efficiency of interval time series, it must first generate some stationary interval time series and non-stationary interval time series by simulation. Therefore, four interval time series consisting of 450 samples are generated by the following equations respectively,

$$Y_t = 0.8Y_{t-1} + \varepsilon_t, \quad 1 \le t \le 150 \tag{2.1}$$

$$Y_t = \begin{cases} 0.2Y_{t-1} + \varepsilon_t\,, & if \quad 1 \le t \le 150 \\ 0.5Y_{t-1} + \varepsilon_t\,, & if \quad 151 \le t \le 300 \\ 0.8Y_{t-1} + \varepsilon_t, & if \quad 301 \le t \le 450 \end{cases} \tag{2.2}$$

$$Y_t = \begin{cases} 1 + 0.8Y_{t-1} + \varepsilon_t\,, & if \quad 1 \le t \le 150 \\ 0 + 0.8Y_{t-1} + \varepsilon_t\,, & if \quad 151 \le t \le 300 \\ -1 + 0.8Y_{t-1} + \varepsilon_t, & if \quad 301 \le t \le 450 \end{cases} \tag{2.3}$$

$$Y_t = \begin{cases} \sigma_t\varepsilon_t\,, \sigma_t^2 = 5 + 0.5Y_{t-1}^2\,, & if \quad 1 \le t \le 150 \\ \sigma_t\varepsilon_t\,, \sigma_t^2 = 5 + 0.2Y_{t-1}^2\,, & if \quad 151 \le t \le 300 \\ \sigma_t\varepsilon_t\,, \sigma_t^2 = 10 + 0.6Y_{t-1}^2\,, & if \quad 301 \le t \le 450 \end{cases} \tag{2.4}$$

where $\varepsilon_t \sim WN(0,1)$. The model (2.1) is a model of $AR(1)$; the model (2.2) and the model (2.3) are the threshold models which combine three models of $AR(1)$ with different sets of coefficients; the model (2.4) is a model of $ARCH(1)$ designed by three different variations.

If the simulated data of each model are collected at 450 time points, the data are also merely 450 single values. Thus, the simulation is carried out by repeating 30 times at each time point. Assume $\{Y_{ij}\}_{30\times 450}$ be the point time series. Since the collected data of 30 samples at each of 450 time points are variable, the minimum and maximum at each time point can be considered as the boundary of the interval data at each time point. Let $a_t = \min_{1\le i\le 30} Y_{ij}$, $b_t = \max_{1\le i\le 30} Y_{ij}$, and $m_t = \dfrac{\sum_{i=1}^{30} Y_{ij}}{30}$ for $t = 1, 2, \ldots, 450$, then $l_t = m_t - a_t$, $u_t = b_t - m_t$ for $t = 1, 2, \ldots, 450$, and $\{\boldsymbol{X}_t = (m_t, l_t, u_t)_{LR} \mid t = 1, 2, 3, \ldots, 450\}$ is an interval time series. Figure 2.2 (a)–(d) show the upper and lower bounds of the interval time series generated by the models (2.1), (2.2), (2.3) and (2.4) respectively. Obviously, the interval time series of the model (2.1) is stationary, while the interval time series of the models (2.2), (2.3) and (2.4) are not stationary.

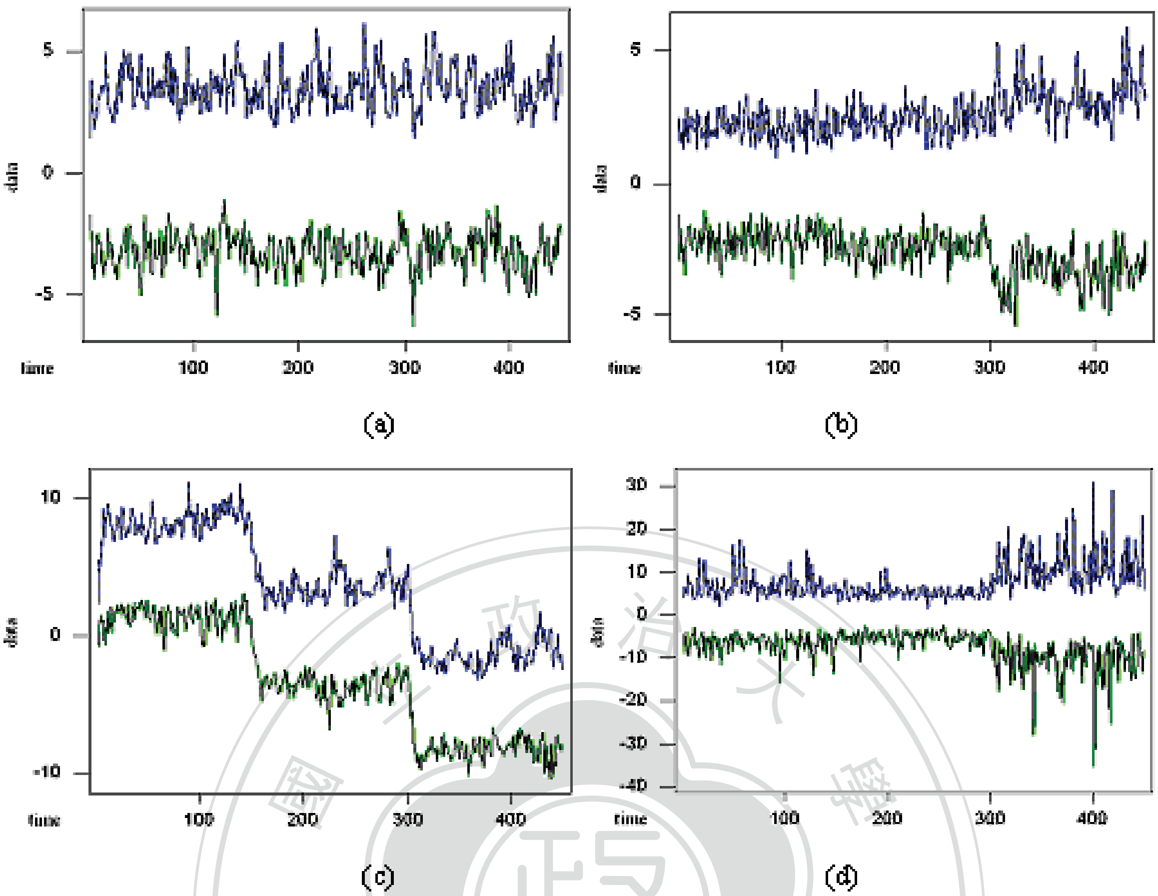

Figure 2.2 The interval time series generated by the models.

2.4.2 **Model Construction of Interval Time Series**

In the model construction of interval time series, it must be cautious of two situations which will be described as follows separately. Firstly, since the collected data are in the form of interval, if the mean values m_t of the intervals $\boldsymbol{X}_t$ for all t are unknown and the tendency of the interval time series is unable to be known effectively, it may take the middle value $\frac{a_t + b_t}{2}$ to be the interval center c_t for each $\boldsymbol{X}_t$ and thus get the radius of interval which is $r_t = c_t - a_t = b_t - c_t$. Then the interval moving average of order k is employed for finding the averages of the centers and radii of the interval time series respectively to perform the forecasting. If it concerns the range of the interval time series, it can apply the weighted average to the center and radius of every interval, that is, it can utilize the weighted interval moving average of order k to make the forecasting.

Secondly, if the patterns of the interval time series { m_t }, { l_t }, { u_t }, and the tendencies of *ACF* and *PACF* are observed and compared with the theories of *ACF* and *PACF*, it can determine a few tentative models. According to the *AIC* criterion by Akaike [2], it can define a criterion for the decision of the best model. The criterion is given by

$$AIC = n \ln \hat{\sigma}^2 + 2(p + q),$$

where n denotes the size of samples or the degree of freedom, $\hat{\sigma}^2$ denotes the residual variance, and $p + q$ is the number of the parameters of the model. Therefore, the criterion for the best model is decided by the smallest value of *AIC* and the *ARIMA* process is applied to forecast.

2.4.3 **The Comparison and Analysis of the Forecast Results**

As soon as the most suitable model of interval time series is built, it can proceed to forecast based on this model. If $\boldsymbol{X}_n$ is the last observed interval in the interval time series $\{\boldsymbol{X}_t\}$, $\hat{\boldsymbol{X}}_{n+s} = E[\boldsymbol{X}_{n+s} \mid \boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n] = (\hat{m}_{n+s}, \hat{l}_{n+s}, \hat{u}_{n+s})_{LR}$ for $s \geq 1$ are the forecast intervals of the $(n+s)\,th$ actual intervals X_{n+s} respectively. According to the interval time series of 450 samples generated in Section 2.4.1, the first 444 samples are assumed to be the observed intervals which are used to predict the last 6 data, *i.e.*, $n = 444$ and s = 1, 2, 3, 4, 5, 6. And MINITAB 14 is utilized to obtain the last 6 forecasting values.

Table 2.1–2.4 list the six forecast intervals $\hat{X}_{n+s}$, s = 1, 2, 3, 4, 5, 6, respectively for the four simulation data of interval time series obtained by the proposed forecasting methods. Table 2.5–2.8 also list the efficiency analyses of interval time series forecasting for the four models respectively.

Table 2.1 The comparison of the forecast and simulated values for the last 6 intervals in Model (2.1).

Model (2.1) (m_t, l_t, u_t)	***IMA*** $(\hat{m}_t, \hat{l}_t, \hat{u}_t)$	***WIMA*** $(\hat{m}_t, \hat{l}_t, \hat{u}_t)$	***ARIMA*** of (m, l, u) ((1,0,0),(1,0,0),(1,0,0)) $(\hat{m}_t, \hat{l}_t, \hat{u}_t)$
(-0.09,2.72,3.09)	(0.03,3.22,3.45)	(0.03,3.34,3.62)	(-0.17,3.23,2.91)
(-0.03,2.24,2.79)	(0.03, 3.22,3.45)	(0.03,3.35, 3.63)	(-0.13,3.22,3.18)
(0.31,3.68,2.94)	(0.03, 3.22,3.45)	(0.03, 3.35, 3.63)	(-0.10,3.22,3.32)
(0.41,2.53,4.53)	(0.03, 3.22,3.45)	(0.03, 3.35, 3.63)	(-0.07,3.22,3.38)
(0.31,2.54,4.59)	(0.03, 3.22,3.46)	(0.03, 3.35, 3.63)	(-0.05,3.22,3.41)
(-0.29,2.21,3.48)	(0.03, 3.22,3.46)	(0.04, 3.35, 3.63)	(-0.04,3.22,3.43)

Table 2.2 The comparison of the forecast and simulated values for the last 6 intervals in Model (2.2).

Model (2.2) (m_t, l_t, u_t)	***IMA*** $(\hat{m}_t, \hat{l}_t, \hat{u}_t)$	***WIMA*** $(\hat{m}_t, \hat{l}_t, \hat{u}_t)$	***ARIMA*** of (m, l, u) ((1,0,0),(2,0,0),(1,1,0)) $(\hat{m}_t, \hat{l}_t, \hat{u}_t)$
(0.15,3.12,4.60)	(-0.01,2.59,2.56)	(0.00,2.81,2.71)	(-0.06,3.11,3.60)
(0.34,3.81,4.80)	(-0.01,2.60,2.56)	(0.00, 2.81,2.79)	(-0.04,3.04, 3.60)
(0.49,2.27,3.96)	(-0.01,2.59,2.56)	(0.00, 2.81,2.79)	(-0.03,2.90, 3.60)
(0.38,3.78,3.14)	(-0.00,2.60,2.56)	(-0.01, 2.81,2.79)	(-0.02,2.82, 3.60)
(0.26,2.45,2.90)	(-0.00,2.60,2.56)	(0.00, 2.81,2.80)	(-0.01,2.76, 3.60)
(0.23,3.42,3.11)	(-0.00,2.60,2.57)	(0.00, 2.81,2.80)	(-0.01,2.71, 3.60)

Table 2.3 The comparison of the forecast and simulated values for the last 6 intervals in Model (2.3).

Model (2.3) (m_t, l_t, u_t)	***IMA*** $(\hat{m}_t, \hat{l}_t, \hat{u}_t)$	***WIMA*** $(\hat{m}_t, \hat{l}_t, \hat{u}_t)$	***ARIMA*** of (m, l, u) ((1,1,0),(1,0,0),(1,0,0)) $(\hat{m}_t, \hat{l}_t, \hat{u}_t)$
(-4.99,2.24,5.07)	(0.05,3.47,3.50)	(0.05,3.66,3.72)	(-4.96,3.31,3.85)
(-4.76,3.18,3.06)	(0.05,3.47,3.50)	(0.04,3.66,3.72)	(-4.96,3.39,3.71)
(-4.84,3.57,2.90)	(0.05,3.48,3.50)	(0.04,3.66,3.72)	(-4.96,3.43,3.63)
(-4.87,3.11,3.48)	(0.04,3.48,3.50)	(0.03,3.67,3.72)	(-4.96,3.45,3.56)
(-5.00,2.90,3.07)	(0.04,3.48,3.51)	(0.02,3.67,3.72)	(-4.96,3.46,3.54)
(-4.94,3.47,2.47)	(0.03,3.48,3.51)	(0.01,3.67,3.73)	(-4.96,3.46,3.52)

Table 2.4 The comparison of the forecast and simulated values for the last 6 intervals in Model (2.4).

Model (2.4) (m_t, l_t, u_t)	***IMA*** $(\hat{m}_t, \hat{l}_t, \hat{u}_t)$	***WIMA*** $(\hat{m}_t, \hat{l}_t, \hat{u}_t)$	***ARIMA*** of (m, l, u)* ((0,0,2),(2,1,0),(1,0,1)) $(\hat{m}_t, \hat{l}_t, \hat{u}_t)$
(0.50,4.56,6.20)	(0.04,7.45,7.59)	(0.04,9.38,9.58)	(0.12,9.90,11.78)
(-0.01,7.74,10.96)	(0.04,7.46,7.60)	(0.04,9.39,9.58)	(-0.03,9.87,11.11)
(0.82,10.62,10.13)	(0.04,7.46,7.60)	(0.04,9.39,9.59)	(0.06,8.84,11.07)
(1.30,14.59,21.91)	(0.04,7.46,7.61)	(0.04,9.39,9.59)	(0.06,9.32,11.07)
(-0.62,9.43,11.31)	(0.04,7.46,7.61)	(0.04,9.40,9.60)	(0.06,9.42,11.07)
(0.10,8.28,5.75)	(0.04,7.46,7.61)	(0.04,9.40,9.60)	(0.06,923,11.07)

**Note: The pattern of the interval time series of Model (2.4) shows the interval data presenting a significant change after the 300th sample. Because the trend of the time series { m_t }, { l_t }, and { u_t } has the same tendency, the data before and after the 300th sample belong to two models of different types. Therefore, it only uses the last 144 samples to predict the last 6 forecasting.*

Table 2.5 The comparison of the efficiency of the forecasting methods for Model (2.1).

Model (2.1)	***IMA***	***WIMA***	***ARIMA*** of (m, l, u) ((1,0,0),(1,0,0),(1,0,0))
MSEP	0.07	0.07	0.10
MSEI	1.50	1.65	1.73
MRIE	0.04	0.04	0.05

Table 2.6 The comparison of the efficiency of the forecasting methods for Model (2.2).

Model (2.2)	***IMA***	***WIMA***	***ARIMA*** of (m, l, u) ((1,0,0),(2,0,0),(0,1,1))
MSEP	0.11	0.11	0.12
MSEI	3.40	2.74	1.35
MRIE	0.06	0.05	0.05

Table 2.7 The comparison of the efficiency of the forecasting methods for Model (2.3).

Model (2.3)	***IMA***	***WIMA***	***ARIMA*** of (m, l, u) ((1,1,0),(1,0,0),(1,0,0))
MSEP	24.44	24.33	0.01
MSEI	71.95	72.20	0.85
MRIE	0.73	0.71	0.01

Table 2.8 The comparison of the efficiency of the forecasting methods for Model (2.4).

Model (2.4)	***IMA***	***WIMA***	***ARIMA*** of (m, l, u) ((0,0,2),(2,1,0),(1,0,1))
MSEP	0.47	0.47	0.46
MSEI	56.53	43.80	41.73
MRIE	0.03	0.03	0.02

With regard to the stationary interval time series, such as the model (2.1), all of the proposed methods achieve excellent forecast results, as Table 2.5 demonstrates. But for the non-stationary interval time series whose the expectation at each point is zero, *i.e.* $E(Y_t) = 0$, for $t = 1, 2, \ldots, 450$, such as the model (2.2), the three proposed methods of interval forecasting attain an excellent position forecasting, *i.e.* with small *MSEP*; but for *MSEI*, only *ARIMA* interval forecasting can make a better forecasting than the others, as Table 2.6 shows.

For the model (2.3) that the interval time series is non-stationary and monotonically decreasing, *ARIMA* interval forecasting attains better *MSEP* and superior *MSEI* than the interval moving average and the weighted interval moving average can do, as indicated in Table 2.7. It means that *ARIMA* interval forecasting achieves a supreme interval forecasting than the other forecasting methods do. While considering the interval time series of the non-stationary model which was simulated by *ARCH*(*1*), such as the model (2.4), the interval moving average, the weighted interval moving average, and *ARIMA* interval forecasting all attains poor *MSEI*. Although these forecasting methods cannot present satisfying forecast results, *ARIMA* interval forecasting is still better than the interval moving average and the weighted interval moving average. Furthermore, all of the forecasting methods achieve very similar position forecasting $\{\hat{m}_t\}$ because they have very close *MSEP* values, as illustrated in Table 2.8.

Calculating the mean relative interval error for every model, it can be found that the mean relative interval error is smaller as long as there is more overlap range of the forecast interval and the actual interval. Among the four models, only for the case of the model (2.3), the interval moving average and the weighted interval moving average performs a worse *MRIE*. That is because the overlap range of the forecast interval and the actual interval is little so that the mean relative interval error is larger.

When we forecast the mean value of interval based on the four simulated interval time series, the *MSEP* values of the models (2.1), (2.2), and (2.3) exhibit that they all generate a very good forecast result. Since the process is same as the traditional point prediction did, we can say that this three proposed forecasting methods have the same forecasting efficiency as the traditional prediction methods. Furthermore, the interval forecasting also provides the length forecasting of interval, that is, we can tell how much the future variation is.

In the model (2.1) which is a the stationary time series, the *MSEP*, *MSEI*, and *MRIE* values show that all the three forecasting methods result in an excellent forecast result. But in the threshold model (2.2), the *MSEI* value of *IMA* is larger than those of the other two methods. This means that *IMA* results in a worse forecasting of interval length than the other two methods. On the other hand, *WIMA* which is *IMA* with fuzzy weights could produce a better forecast result. In the threshold model (2.3), *IMA* and *WIMA* cannot provide a good position forecasting ($\hat{m}$). Since their *MSEP* values are much larger than *ARIMA* interval forecasting's, their *MSEI* values are relatively larger. It explains that it is not a good interval forecasting even though it produces a better length forecast because the overlap part of the forecast interval and the actual interval is small when $\hat{m}$ is far deviated from m.

In the forecasting of model (2.4), the *MSEP* values of the three forecasting methods are close, that is, they have almost the same efficiency of position forecasting. Nevertheless, the *MSEI* values exhibit that *WIMA* and *ARIMA* interval forecasting both have a better length forecasting. Therefore, in these simulated models, *ARIMA* interval forecasting can provide a better interval forecasting than the other two forecasting methods.

2.5 TWO CASE ANALYSES

In this section we use two practical cases to demonstrate the forecasting methods. One is the monthly trading values of China Steel stock. The other is the daily temperatures in Taipei.

2.5.1 The Monthly Trading Value of the Stock

The practical data comes from the report of monthly trading values of individual stock provided by Taiwan Stock Exchange Corporation [38]. The highest prices (b_t), the lowest prices (a_t) and the weighted average prices (m_t) of monthly trading values of China Steel stock from April 1999 to September 2006 are collected to form the interval time series. The chart of the interval time series is shown in Figure 2.3. The monthly prices from April 1999 to March 2006 are used as the observed interval time series $\{\boldsymbol{X}_t, t = 1, 2, \ldots, 84\}$. Then the proposed methods are applied to perform the interval forecasting for the last 6 intervals (from April 2006 to September 2006). After performing the forecasting on the interval time series $\{\boldsymbol{X}_t\}$, Table 2.9 lists the actual monthly trading values of China Steel stock from April 2006 to September 2006 and the forecast intervals by the proposed interval forecasting methods. Furthermore, Table 2.10 illustrates the comparison of the efficiency of the interval forecasting methods.

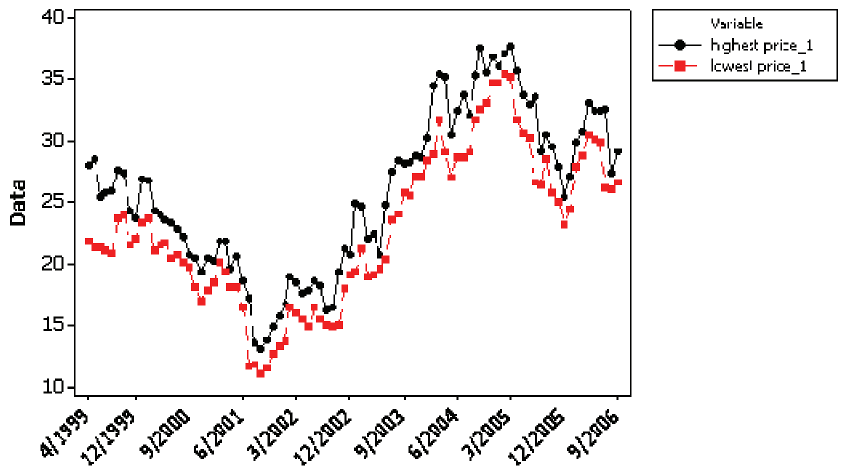

Figure 2.3 The chart of the monthly trading values of China Steel stock from 4/1999 to 9/2006.

Table 2.9 The monthly trading values of China Steel stock and their forecasts by the proposed interval forecasting methods for the latter 6 periods from 4/2006 to 9/2006.

Actual Trading Values (m_t, l_t, u_t)	***IMA*** $(\hat{m}_t, \hat{l}_t, \hat{u}_t)$	***WIMA*** $(\hat{m}_t, \hat{l}_t, \hat{u}_t)$	***ARIMA*** of (m, l, u) ((1,0,1),(1,0,0), (0,1,1)) $(\hat{m}_t, \hat{l}_t, \hat{u}_t)$
(31.67,1.22,1.33)	(23.99,1.60,1.39)	(23.38,2.09,1.62)	(29.41,1.29,1.39)
(31.37,1.37,1.03)	(23.97,1.58,1.38)	(23.37,2.05,1.60)	(29.24,1.49,1.39)
(30.77,0.97,1.63)	(23.95,1.54,1.37)	(23.36,1.95,1.58)	(29.08,1.56,1.39)
(29.39,3.29,3.16)	(23.95,1.52,1.37)	(23.35,1.93,1.58)	(28.93,1.59,1.39)
(26.39,0.39,0.96)	(23.95,1.51,1.36)	(23.35,1.90,1.58)	(28.79,1.60,1.39)
(27.63,1.08,1.37)	(23.95,1.48,1.36)	(23.34,1.87,1.87)	(28.64,1.60,1.39)

Table 2.10 The comparison of efficiency of the proposed interval forecasting methods for the interval time series of monthly trading values of China Steel stock.

Criteria	***IMA***	***WIMA***	***ARIMA*** of (m, l, u) ((1,0,1),(1,0,0), (0,1,1))
MSEP	34.89	42.00	3.25
MSEI	109.59	133.93	10.67
MRIE	2.10	2.13	0.62

From the above analyses, the forecast results of this interval time series are extremely similar with that of the model (2.3). It explains that the interval time series is non-stationary but steadily increasing or decreasing. As shown in Table 2.9, the interval moving average and the weighted interval moving average underestimate the forecast interval. Thus, they cannot provide a good forecast result. In contrast, the *ARIMA* interval forecasting can attain a better forecast interval since the *MSEP* and *MSEI* values by *ARIMA* are smaller than the *MSEP* and *MSEI* values obtained by the interval moving average and the weighted interval moving average, as shown in Table 2.10. In addition, the mean relative interval error by *ARIMA* is fairly small, which means the overlap ranges of the forecast intervals and the actual intervals are larger. Therefore, it can conclude that *ARIMA* interval forecasting is a superior forecasting method. On the other hand, the mean relative interval errors of the interval moving average and the weighted interval moving average are relatively large, which demonstrates few overlap part of the forecast intervals and the actual intervals. Hence, these two forecasting methods are not suitable for this case.

2.5.2 **The Daily Temperatures in Taipei**

The interval data used in this section are the daily temperatures provided by the Central Weather Bureau in Taiwan. The interval time series are composed of the highest temperature (b_t), the lowest temperature (a_t) and the daily average temperature (m_t) from April 1, 2008 to April 30, 2008 in Taipei. The boundaries of the interval time series is shown in Figure 2.4. Especially, we use the interval center $c_t = \frac{a_t + b_t}{2}$ and the radius of interval $r_t = l_t = u_t$ which is mentioned in Section 2.4.2 to carry on forecasting. The proposed methods are applied to perform the interval forecasting for the last 6 intervals (from April 25, 2008 to April 30, 2008) and the corresponding forecast results with the actual data are listed in Table 2.11 and Table 2.12. The comparisons of the efficiency of the interval forecasting methods are shown in Table 2.13 and Table 2.14.

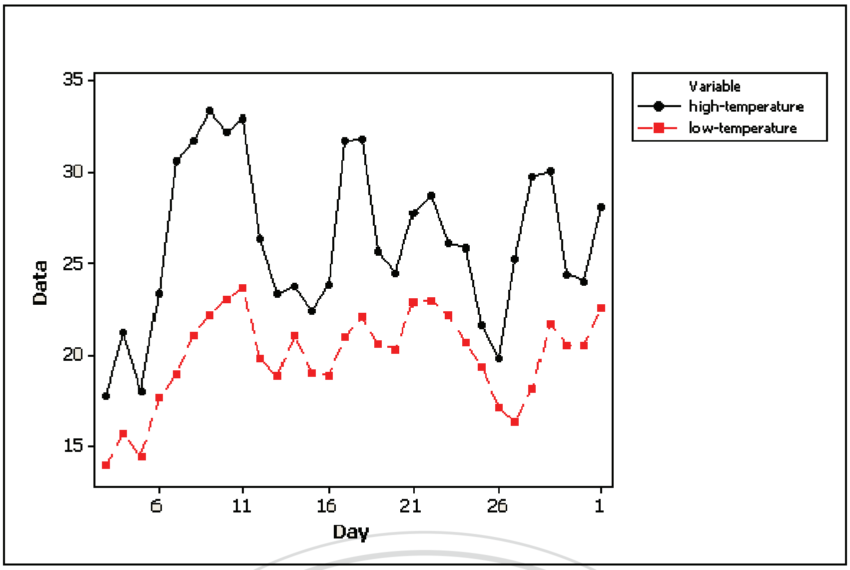

Figure 2.4 The chart of the daily temperatures in Taipei from 4/1/2008 to 4/30/2008.

Table 2.11 The daily temperature in Taipei and their forecasts by the proposed interval forecasting methods for the latter 6 periods from 4/25/2008 to 4/30/2008. (X = (m, l, u)$_{LR}$)

Actual Temperatures (m_t, l_t, u_t)	***IMA*** $(\hat{m}_t, \hat{l}_t, \hat{u}_t)$	***WIMA*** $(\hat{m}_t, \hat{l}_t, \hat{u}_t)$	***ARIMA* of** (m, l, u) $(\hat{m}_t, \hat{l}_t, \hat{u}_t)$
(20.73,4.43,4.47)	(22.37,2.48,3.63)	(21.44,3.03,4.54)	(20.20,1.84,3.40)
(23.47,5.27,6.23)	(22.64,2.50,3.71)	(21.74,3.07,4.59)	(22.09,2.09,3.59)
(24.88,3.18,5.12)	(22.83,2.50,3.73)	(21.88,3.10,4.64)	(22.09,2.24,3.59)
(22.35,1.85,2.05)	(23.12,2.54,3.80)	(22.24,3.13,4.70)	(22.09,2.31,3.59)
(22.31,1.81,1.69)	(23.25,2.56,3.82)	(22.33,3.16,4.74)	(22.09,2.36,3.59)
(24.64,2.04,3.36)	(23.24,2.47,3.68)	(22.31,3.05,4.57)	(22.09,2.38,3.59)

Table 2.12 The daily temperature in Taipei and their forecasts by the proposed interval forecasting methods for the latter 6 periods from 4/25/2008 to 4/30/2008. (X = (c, r, r)$_{LR}$)

Actual Temperatures (c_t, r_t, r_t)	***IMA*** $(\hat{c}_t, \hat{r}_t, \hat{r}_t)$	***WIMA*** $(\hat{c}_t, \hat{r}_t, \hat{r}_t)$	***ARIMA* of** (c, r, r) $(\hat{c}_t, \hat{r}_t, \hat{r}_t)$
(20.75,4.45,4.45)	(22.95,3.06,3.06)	(21.88,3.75,3.75)	(20.52,1.84,1.84)
(23.95,5.75,5.75)	(23.24,3.10,3.10)	(22.20,3.79,3.79)	(22.60,2.17,2.17)
(25.85,4.15,4.15)	(23.44,3.12,3.12)	(22.35,3.83,3.83)	(22.60,2.40,2.40)
(22.45,1.95,1.95)	(23.75,3.17,3.17)	(22.72,3.88,3.88)	(22.60,2.55,2.55)
(22.25,1.75,1.75)	(23.88,3.19,3.19)	(22.81,3.91,3.91)	(22.60,2.65,2.65)
(25.30,2.70,2.70)	(23.84,3.07,3.07)	(22.77,3.77,3.77)	(22.60,2.71,2.71)

Table 2.13 The comparison of efficiency of the proposed interval forecasting methods for the interval time series of daily temperature in Taipei. ($X = (m, l, u)_{LR}$)

Criteria (*m*, *l*, *u*)	***IMA***	***WIMA***	***ARIMA***
MSEP	1.84	2.99	2.76
MSEI	12.26	14.56	14.03
MRIE	0.21	0.17	0.20

Table 2.14 The comparison of efficiency of the proposed interval forecasting methods for the interval time series of daily temperature in Taipei. ($X = (c, r, r)_{LR}$)

Criteria (*c*, *r*, *r*)	***IMA***	***WIMA***	***ARIMA***
MSEP	2.94	3.90	3.13
MSEI	13.39	16.34	17.89
MRIE	0.26	0.22	0.22

Since the mean of the daily average temperatures (m_t) is about 23°C and the fluctuation of temperatures is around 23°C in Figure 2.4, the interval moving average and the *ARIMA* of (*m*, *l*, *u*) have better forecast results. Then the *MSEP* values of the interval moving average and the *ARIMA* of (*m*, *l*, *u*) are lesser than the others. Besides, the *MSEI* value of the interval moving average is the smallest value. So by analyzing their *MSEIs*, the interval moving average is the better forecast method than the others. Additionally, we found that these four forecasting methods differ little in the mean relative interval error (*MRIE*). From Table 2.13 and Table 2.14, the *ARIMA* of (*m*, *l, u*) offers a better forecast result than the *ARIMA* of (*c*, *r*, *r*). By mean of evaluating *MSEI* and *MRIE*, the *ARIMA* of (*m*, *l, u*) and the interval moving average can present a good forecast result.

2.6 CONCLUSIONS

This chapter discusses the establishment of forecast model and the forecasting of interval time series by means of the interval operations. From the research results, it is found that the forecasting of *ARIMA* interval time series achieves more accurate forecast, no matter in the comparisons of the mean squared error of interval or the mean relative interval error, than the traditional forecasting methods such as the moving average and the weighted moving average, *etc*. Especially for the threshold time series, the forecast results achieve much better forecast.

In fact, using interval data to establish model and forecast interval, we can find that the forecasting in each step is carried out by means of intervals so as to increase the objectiveness of the forecasting results. In the general aspect, the "*intervalization*" seems to be a very reasonable phenomenon too. But on the contrary, if the concept of dealing with numerical data does not change and the forecasting method does not make a breakthrough, it often frustrates the objectivity of measurements and the possibility of long-term forecasting. If we measure interval time series by means of the centers and the lengths of intervals, it demonstrates clearly that interval time series has better forecasting ability than the traditional *ARIMA* method does. However, according to the interval operations and the *ARIMA* method, it is noteworthy that if we can establish a good model construction, we can make a superior interval forecasting for the interval time series of stock trading values. For investors, it not only provides a new forecasting method but also offers a more flexible forecast result. Therefore, investors can make more objective judgments under correct information.

Although the approaches proposed in this chapter effectively perform interval forecasting, some problems are remained to be solved and some improvement could be done for further research, described respectively as follows.

(1) There are so many unpredictable factors on the monthly trading values of stocks, such as

trading volumes, exchange rates, interest rates and even the influence of the government policy etc. Consequently, in respect to interval time series proposed in this chapter, it only considers the monthly highest prices and the monthly lowest prices as the range of the monthly trading prices caused by all factors. If it needs to make the result more accurate, it must find out the key factors of influencing the interval range.

(2) In order to achieve a more accurate result of interval time series forecasting, it needs to make the collected data stationary for further analysis. But, how to judge if an interval time series is stationary? It could find another approach to judge the stationariness of interval time series by other interval operations.

(3) Because the research of interval data forecasting is rare in the past, the practical interval data are few too. Interval time series data should be generated by simulation. But the simulation method will influence the objectiveness of the forecasting methods. Therefore, we can consider other simulation approaches, such as Bootstrap, Bayesian, etc, to generate simulated interval time series for further analysis. The variety of the simulation methods should contribute to the improvement of the forecasting methods.

(4) In the analysis and forecasting of interval time series, how to estimate the forecasting accuracy of interval data is an important issue. There are found four forecasting situations, which are the forecast interval is too wide, the forecast interval is too narrow, the forecast interval inclines to the right, and the forecast interval inclines to the left respectively. From the overlap parts and the non-overlap parts of the actual intervals and the forecast intervals, it should be defined a criterion which is more sufficient to show the efficiency of interval forecasting.