

The Interval Autoregressive Time Series Model

Xun Wang

Department of Applied Mathematics
Beijing University of Technology
Beijing, China 100124
Email: wangxun@emails.bjut.edu.cn

Shoumei Li

Department of Applied Mathematics
Beijing University of Technology
Beijing, China 100124
Email: lisma@bjut.edu.cn

Abstract—This paper mainly suggests a new type of interval time series: interval autoregressive (IAR) model. Firstly we state why we should introduce the interval time series models. Then we give necessary definitions about random intervals and interval time series. Thirdly, we introduce some methods of efficiency evaluation for forecasting of interval time series. And then we discuss parameter estimation and forecasting in IAR model, in which the methods of parameter estimation are based on the evaluation forecasting for interval data. Furthermore, we give the simulation results and apply it to real data from Shanghai Stock Index, which is to illustrate our modeling methodology. This model makes it possible for decision makers to forecast the best and worst possible situations based on interval-valued observations.

Keywords: Interval time series; IAR model; parameter estimation; forecasting of stock price

I. INTRODUCTION

It is well-known that traditional time series models play an important role in many application areas. For instance, Box-Jenkins provided autoregressive integrated moving average (ARIMA) time series model in [4], which has been applied in engineering, economics, foreign current exchange, data analysis of stock markets, etc. In those traditional times series models, authors assumed that that the future values of a time series have a determined function relationship with current, past values and white noises. However, there are a great deal of economic problems that can not be explained and handled by these models since the actions of individuals in investment are affected not only by the random factors in financial market but also by the mental factors of investors. Therefore, interval-valued time series emerges quite naturally in some situations, which indicates uncertainty and variability.

Song and Chissom [11]-[13] believed in that fuzzy numbers would be better to describ the mental factors. So they introduced the concepts of fuzzy time series, and they assumed that the model parameters were fuzzy numbers. Tseng *et. al.* [15] built a fuzzy ARIMA model and gave the methods of parameter estimation and forecasting, and applied it to the foreign current exchange market.

The sample data usually are one-dimensional or multidimensional numbers in traditional time series analysis. However, in practice, the sample data sometimes are intervals in economics. For example, when the economists are surveyed to predict the economic increasing rate, they tend to give the answers like “5% to 7%”, “5.2% to 6.7%” and so on . So we get the interval-valued sample data. Therefore, it is necessary

to build the models of interval time series. The interval time series models show up the information in the form of a series of interval-valued data and make predictions based on the interval-valued data. Furthermore, such models also provide the upper boundary and lower boundary of possible selections for decision-makers or investors.

Hsu and Wu did a preliminary study on interval time series, and gave three evaluation criteria of forecasting efficiency for interval time series in [7], [8], that is, mean square error of interval (MSEI), mean relative interval error (MRIE), mean ratio of symmetric difference. Based on their works, we develop a new interval time series model in this paper ,which is called interval autoregressive (IAR) model .

This paper is organized as follows. Some definitions of random interval and interval time series are given in Section 2, especially interval autoregressive model is formulated and proposed. In Section 3, the methods of efficiency evaluation of forecasting for interval time series are presented. Methods of parameter estimation in IAR model are given in Section 4. In Section 5 and Section 6, simulation result and real data analysis are given respectively, which illustrate our modeling methodology.

II. RANDOM INTERVALS AND INTERVAL TIME SERIES

Random interval, as well as its operations, is the foundation in terms of analyzing the interval time series. Therefore, before the discussions of interval time series, we give some definitions of random interval and its operations.

Assume that $(\Omega, \mathcal{A}, P)$ is a probability space, $\mathbb{R}$ is the set of all real numbers.

Definition 2.1 The interval $X = [a, b]$ is called a random interval, if $a, b: \Omega \rightarrow \mathbb{R}$ are real-valued random variables with $a(\omega) \leq b(\omega)$ for all $\omega \in \Omega$. Then $c = (a + b)/2$ is called the center of the random interval X , $r = (b - a)/2$ is called the radius of the random interval X , and X is also denoted as $X = (c; r)$.

Let $X = [a, b] = (c; r)$ be a random interval, the interval length of X is $2r = b - a$, and denoted as $\|X\| = 2r$.

Remark 2.2 Actually the above concept of a random interval is the special case of that of a set-valued random variable (or a random set, or a multifunction in literature). Readers may refer to [2], [3], [6], [10], [16]. It was defined in a general way as follows.

Let $\mathfrak{X}$ be a separable Banach space with its dual space $\mathfrak{X}^*$. A set-valued mapping X from Ω to the family of all nonempty closed subsets of $\mathfrak{X}$ is called a set-valued random variable if, for each open subset O of $\mathfrak{X}$,

$$X^{-1}(O) = \{\omega \in \Omega : X(\omega) \cap O \neq \emptyset\} \in \mathcal{A}.$$

If X takes compact convex subsets of $\mathfrak{X}$, then that X is a set-valued random variable is equivalent to that, for any $x^* \in \mathfrak{X}^*$, its support function

$$s(x^*, X(\omega)) = \sup\{ \langle x^*, x \rangle : x \in X(\omega) \}$$

is a real-valued random variable.

So it is obvious that, if we limit $\mathfrak{X} = \mathbb{R}$, X takes compact convex set values in $\mathbb{R}$, then the concept of a random interval above is consistent with that of a set-valued random variable.

Definition 2.3 Let $X_1 = [a_1, b_1] = (c_1; r_1)$, $X_2 = [a_2, b_2] = (c_2; r_2)$ be two random intervals. The interval addition, scalar multiplication and interval subtraction are defined as

(i) interval addition:

$$X_1 \oplus X_2 = (c_1; r_1) \oplus (c_2; r_2) = (c_1 + c_2; r_1 + r_2);$$

(ii) scalar multiplication:

$$k \cdot X_1 = kX_1 = k(c_1; r_1) = (kc_1; |k|r_1), \forall k \in \mathbb{R};$$

(iii) interval subtraction:

$$X_1 \ominus X_2 = X_1 \oplus (-X_2) = (c_1 - c_2; r_1 + r_2).$$

Remark 2.4 (1) Note that above definitions of interval addition, scalar multiplication are the special case of the definitions of addition and scalar multiplication of set-valued random variables in a separable Banach space.

(2) Let $\mathbf{K}_{kc}(\mathbb{R})$ be the set of all finite intervals of $\mathbb{R}$. We define interval addition and scalar multiplication on $\mathbf{K}_{kc}(\mathbb{R})$ as above definition. But we have to notice that $(\mathbf{K}_{kc}(\mathbb{R}), \oplus, \cdot)$ is not a linear space! This is because for any $[a, b]$ with $a < b$, $[a, b] \oplus (-[a, b]) \neq \{0\}$. Thus, above interval subtraction operator is not inverse operator of interval addition.

In the set theory, set difference $A \setminus B$ is defined by $A \setminus B = \{x : x \in A \text{ and } x \notin B\}$. While we use this definition of the set difference on the family of closed intervals, the difference of two closed intervals is a half-closed interval. Since we want it to be a closed set, we have to revise the definition of interval difference as follows.

Definition 2.5 Let $X_1 = [a_1, b_1] = (c_1; r_1)$, $X_2 = [a_2, b_2] = (c_2; r_2)$ be two random intervals, then the interval difference $X_1 - X_2$ is defined by

$$X_1 - X_2 = \text{cl}(X_1 \cap X_2^c),$$

where $\text{cl}(A)$ means the closure of the set A .

If $a_1 \leq a_2 \leq b_1 \leq b_2$, then $X_1 - X_2 = [a_1, b_1] - [a_2, b_2] = [a_1, a_2]$ by the above definition.

Definition 2.6 Let $X_1 = [a_1; b_1] = (c_1; r_1)$, $X_2 = [a_2; b_2] = (c_2; r_2)$ be two random intervals, then the symmetric difference is defined by $X_1 \triangle X_2 = (X_1 - X_2) \cup (X_2 - X_1)$.

If $a_1 \leq a_2 \leq b_1 \leq b_2$, then by the above definition, $X_1 \triangle X_2 = (X_1 - X_2) \cup (X_2 - X_1) = [a_1, a_2] \cup [b_1, b_2]$.

Definition 2.7 Let $X = (c; r)$, c and r are two real-valued random variables, then $X = (c; r)$ is called a generalized random interval.

Obviously, if r is positive, $X = (c; r)$ is a random interval defined in Definition 2.1. Similarly, we may define the addition and scalar multiplication for generalized random interval as we do in definition 2.3.

Definition 2.8 The expectation of the random interval $X = [a, b] = (c; r)$ is defined by

$$E[X] = [E(a), E(b)],$$

where $E(a), E(b)$ are the expectations of real-valued random variables a and b respectively.

Obviously, we have $E[X] = (E(c); E(r))$. If $X_i = [a_i, b_i]$, $i = 1, 2$, then $E[X_1 \oplus X_2] = E[X_1] \oplus E[X_2]$, $E[kX_1] = kE[X_1]$, for any $k \in \mathbb{R}$.

Remark 2.9 (1) The expectation of a random interval above is consistent with Aumann integral of a compact convex set-valued random variable in $\mathbb{R}$ (cf. [3], [10]). Hence, we may obtain the same properties of the expectations of random intervals as that of Aumann integrals.

(2) In [2] and [10], authors defined the concepts of the distribution of a set-valued random variable, independence of a sequence of set-valued random variables. Here we used them to the random interval case and we omit them. By the Chapter 3 in [10], we know that, if two random intervals $X_1 = [a_1, b_1] = (c_1; r_1)$ and $X_2 = [a_2, b_2] = (c_2; r_2)$ are independent, so are a_1 and a_2 , b_1 and b_2 , c_1 and c_2 , r_1 and r_2 respectively.

Assume that a simple random sample $X_i = [a_i, b_i]$, $i = 1, 2, \dots, n$, is taken from a random interval $X = [a, b]$, i.e. $X_1, X_2, \dots, X_n$ are independent and identically distributed with X . Let

$$\bar{X} = \frac{1}{n} [X_1 \oplus X_2 \oplus \dots \oplus X_n]$$

denote the random sample average. By using the strong law of large numbers of set-valued random variables (cf. [2] and [10]), we may estimate the expectation $E[X]$ of random interval X by using the random sample average

$$\bar{X} = \left[\frac{1}{n} \sum_{i=1}^n a_i, \frac{1}{n} \sum_{i=1}^n b_i \right].$$

It is obvious that $\bar{X}$ is an unbiased estimator of $E[X]$.

Now we introduce the definition of interval autoregressive model (IAR).

Definition 2.10 An interval time series is a sequence of random intervals, $X_t = [a_t, b_t] = (c_t; r_t)$, $t = 1, 2, 3, \dots$, denoted by $\{X_t\} = \{X_t = [a_t, b_t] = (c_t; r_t) : t = 1, 2, 3, \dots\}$.

Assume that $\tilde{\varepsilon}_t = (c_{\varepsilon_t}; r_{\varepsilon_t})$, $t = 1, 2, \dots$ is a generalized random interval (r_{ε_t} may be negative), c_{ε_t} and r_{ε_t} are independent white noise processes $WN(0, \sigma_c^2)$ and $WN(0, \sigma_r^2)$ respectively, then the p -order equation

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p} + \tilde{\varepsilon}_t, t \in \mathbb{Z}$$

is called a p -order interval autoregressive model, denoted as IAR(p), where $a_1, a_2, \dots, a_p$ ($a_p \neq 0$) are called autoregressive coefficients of the IAR(p) model.

III. EFFICIENCY EVALUATION OF FORECASTING WITH INTERVAL TIME SERIES

A. The Mean Square Error of Interval Time Series

Let $X_t = [a_t, b_t] = (c_t; r_t)$, $t = 1, 2, \dots$, be an interval time series, and $\hat{X}_t = [\hat{a}_t, \hat{b}_t] = (\hat{c}_t; \hat{r}_t)$, $t = 1, 2, \dots$, be the forecast interval with respect to $X_t = [a_t, b_t] = (c_t; r_t)$. In the analysis and forecasting of interval time series, we may meet four forecasting situations:

- (1) If $\hat{a}_t \leq a_t \leq b_t \leq \hat{b}_t$, then the forecast interval is too wide, and denoted by FIW;
- (2) If $a_t \leq \hat{a}_t \leq \hat{b}_t \leq b_t$, then the forecast interval is too narrow, and denoted by FIN;
- (3) If $a_t \leq \hat{a}_t \leq b_t \leq \hat{b}_t$, then the forecast interval inclines to the right, and denoted by FIR;
- (4) If $\hat{a}_t \leq a_t \leq \hat{b}_t \leq b_t$, then the forecast interval inclines to the left, and denoted by FIL.

In the classical time series models, we use the distance between actual value and forecast value to describe the efficiency of prediction. In the interval time series models, what we are concerned about are not only the forecasting of interval length, but also the location disparity between the predicted interval and the actual interval. Therefore, it is necessary to define new criteria for analyzing of interval time series forecasting.

Assume that $X_t = (c_t; r_t)$, $t = 1, 2, \dots$, is an interval time series, $\hat{X}_t = (\hat{c}_t; \hat{r}_t)$, $t = 1, 2, \dots$, is the forecasting interval series of X_t , mean square error of interval is given in the following.

Definition 3.1 The mean squared error of interval position (MSEP) is given by

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

and the mean squared error of interval length (MSEL) is defined as

$$MSEL = \frac{1}{s} \sum_{t=1}^s (r_{n+t} - \hat{r}_{n+t})^2,$$

where n denotes the current time, and s is the number of the preceding intervals.

Definition 3.2 The error between the forecast interval $\{\hat{X}_t = (\hat{c}_t; \hat{r}_t)\}$ and the actual interval $\{X_t = (c_t; r_t)\}$ consists

of two parts: the position error and the length error. The mean square error of interval time series (MESI) is given by

$$\begin{aligned} MSEI &= MSEP + MSEL \\ &= \frac{1}{s} \sum_{t=1}^s (c_{n+t} - \hat{c}_{n+t})^2 + \frac{1}{s} \sum_{t=1}^s (r_{n+t} - \hat{r}_{n+t})^2, \end{aligned}$$

where n denotes the current time, and s is the number of the preceding intervals.

In practice, different people may have different standards in terms of the prediction of interval time series due to different situation of problems. One important case, we take into account, is that mean squared error of interval with respect to position and mean squared error of interval with respect to length should pose different influence to the total error. Thus we define a generalized mean squared error of interval time series as follows.

Definition 3.3 The generalized mean squared error of interval (GMSEI) between the forecast interval $\{\hat{X}_t = (\hat{c}_t; \hat{r}_t)\}$ and the actual interval $X_t = (c_t; r_t)$ of a time series is

$$\begin{aligned} GMSEI &= f(MSEP) + g(MSEL) \\ &= f\left(\frac{1}{s} \sum_{t=1}^s (c_{n+t} - \hat{c}_{n+t})^2\right) \\ &\quad + g\left(\frac{1}{s} \sum_{t=1}^s (r_{n+t} - \hat{r}_{n+t})^2\right), \end{aligned}$$

where f and g are strictly increasing functions with $f(0) = 0$, $g(0) = 0$, n is the current time and s is the number of the preceding intervals.

For instance, if we let $f(x) = e^x - 1$, $g(x) = x$, $\forall x \geq 0$, that is

$$GMSEI = \exp(MSEP) - 1 + MSEL.$$

Under this evaluation criteria of forecasting error, we believe that the $MSEP$ is more significant than $MSEI$ in terms of contribution to the total error.

B. Other Evaluation Criteria

Symmetric difference between the actual interval and the forecast interval shows their non-intersecting parts. The smaller the length of their symmetric difference is, the more the forecast interval covers the actual interval. By virtue of this character of symmetric difference, we offer another method of the efficiency evaluation for the interval time series forecasting, i.e. mean ratio of symmetric difference.

Definition 3.4 Let $\{X_t = (c_t; r_t), t = 1, 2, \dots, s\}$, be an interval time series, and $\{\hat{X}_t = (\hat{c}_t; \hat{r}_t), t = 1, 2, \dots, s\}$, be the forecast interval time series. Then the mean ratio of symmetric difference, denoted by MRSD, is defined as

$$MRSD = \frac{1}{s} \sum_{t=1}^s \frac{\|X_{n+t} \Delta \hat{X}_{n+t}\|}{\|X_{n+t}\|}.$$

Definition 3.5 [The efficiency of MRSD] Let $\{X_t = (c_t; r_t)\}$ be an interval time series, and the forecast interval

time series $\{\hat{X}_{1t} = (\hat{c}_{1t}; \hat{r}_{1t})\}$ and $\{\hat{X}_{2t} = (\hat{c}_{2t}; \hat{r}_{2t})\}$ are obtained by two different forecast methods. If the MRSD of $\{\hat{X}_{1t}\}$ (denoted as $MRSD_1$) is smaller than that of $\{\hat{X}_{2t}\}$ (denoted as $MRSD_2$), i.e. $MRSD_1 < MRSD_2$, we say the forecast interval $\{\hat{X}_{1t}\}$ is more efficient than the forecast interval $\{\hat{X}_{2t}\}$.

Hsu and Wu [7] also referred to another evaluation criterion, called mean relative interval error (MRIE). Roughly speaking, we believe that a forecast result is better if the center of this forecast interval is closer to the center of the actual interval, and the intersection of their intervals is larger. The definition of mean relative interval error is as follows.

Definition 3.6 Let $\varepsilon_t = \frac{|c_t - \hat{c}_t|}{r_t + \hat{r}_t}$ be the relative error between the forecast interval $\hat{X}_t$ and the actual interval X_t , the mean relative interval error (MRIE) is given by

$$MRIE = \frac{1}{s} \sum_{t=1}^s \frac{|c_t - \hat{c}_t|}{r_t + \hat{r}_t}.$$

In practical use, we may be confronted with the case that the importance of FIR errors and FIL errors are different. With respect to how to treat this case, we shall discuss it in future works since the paper limitation of this paper.

IV. PARAMETER ESTIMATION IN IAR MODEL

In this section, we will estimate the order p , autoregressive coefficients $a_1, a_2, \dots, a_p$ and the variances of white noises σ_c^2, σ_r^2 in IAR(p) model

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p} + \tilde{\varepsilon}_t, \quad t = 1, 2, \dots$$

through the observation data $x_1, x_2, \dots, x_N$.

A. Methods of coefficient estimation

Firstly, we assume that the order p is known. Now we will choose an evaluation criterion of prediction efficiency. Without loss of generality, we take mean square error of interval (MSEI) as an example here, if other evaluation criteria are taken, a similar conclusion can be derived.

The basic idea of parameter estimation here is similar to the least squares estimation in the classic statistics.

Now we use the linear combination of $x_{j-1}, x_{j-2}, \dots, x_{j-p}$ with coefficients $a_1, a_2, \dots, a_p$, i.e. $a_1 x_{j-1} + a_2 x_{j-2} + \dots + a_p x_{j-p}$, to predict x_j , the reasonable estimators of $a_1, a_2, \dots, a_p$ should minimize the square error

$$\begin{aligned} & S(a_1, a_2, \dots, a_p) \\ &= \sum_{j=p+1}^N [c_j - (a_1 c_{j-1} + a_2 c_{j-2} + \dots + a_p c_{j-p})]^2 \\ &+ \sum_{j=p+1}^N [r_j - (|a_1| r_{j-1} + |a_2| r_{j-2} + \dots + |a_p| r_{j-p})]^2. \end{aligned}$$

Thus, we take the values $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_p$ which minimize the $S(a_1, a_2, \dots, a_p)$ as the estimators of autoregressive coefficients. And the estimators of the variances of the white noises,

$c_{\varepsilon_t}, c_{\varepsilon_t}$ are

$$\hat{\sigma}_c^2 = \frac{1}{N-p} \sum_{j=p+1}^N [c_j - (\hat{a}_1 c_{j-1} + \hat{a}_2 c_{j-2} + \dots + \hat{a}_p c_{j-p})]^2,$$

$$\hat{\sigma}_r^2 = \frac{1}{N-p} \sum_{j=p+1}^N [r_j - (|\hat{a}_1| r_{j-1} + |\hat{a}_2| r_{j-2} + \dots + |\hat{a}_p| r_{j-p})]^2.$$

B. Order Selection

Parameter estimation in IAR model described above is based on that the order p is known. But order p is usually unknown when we cope with some practical problems. Therefore, order p should be estimated.

In terms of order selection problem in traditional single-valued time series models, one of the most applicable criteria is Akaike information criterion (AIC), which was given by Akaike in 1973 [1]. Here we use an interval version of AIC to select an order p . We assume order $p \leq P_0$, where $P_0 \in \mathbb{N}$. The AIC statistics, defined as

$$\begin{aligned} AIC(k) &= \ln \frac{S(\hat{a}_1, \hat{a}_2, \dots, \hat{a}_p)}{N-k} + \frac{4k}{N} \\ &= \ln (\hat{\sigma}_c^2 + \hat{\sigma}_r^2) + \frac{4k}{N}, \end{aligned}$$

where $k = 0, 1, \dots, P_0$. We select the value of p as the order in our model which minimizes $AIC(k)$.

C. Diagnostic Checking

Typically, the goodness of fit of a statistical model to a set of data is judged by comparing the observed values with the corresponding predicted values obtained from the fitted model. If the fitted model is appropriate, the residuals should behave in a manner that is consistent with the model.

By the interval observation data $x_1, x_2, \dots, x_N$, if now we have the estimation $\hat{p}$ and $(\hat{a}_1, \hat{a}_2, \dots, \hat{a}_p)$ of the order p and the autoregressive coefficient $a_1, a_2, \dots, a_p$ of IAR model

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p} + \tilde{\varepsilon}_t, \quad t = 1, 2, \dots,$$

the residuals are defined as follows

$$\hat{\varepsilon}_{ct} = c_t - \sum_{j=1}^p \hat{a}_j c_{t-j}, \quad t = \hat{p} + 1, \hat{p} + 2, \dots, N,$$

$$\hat{\varepsilon}_{rt} = r_t - \sum_{j=1}^p \hat{a}_j r_{t-j}, \quad t = \hat{p} + 1, \hat{p} + 2, \dots, N.$$

If we can check that $\{\hat{\varepsilon}_{ct}\}$ and $\{\hat{\varepsilon}_{rt}\}$ are white noise processes, i.e. $\{\hat{\varepsilon}_{ct}\} \sim WN(0, \hat{\sigma}_c^2)$, $\{\hat{\varepsilon}_{rt}\} \sim WN(0, \hat{\sigma}_r^2)$, then the assertion that the above model is the true process in terms of generating $x_1, x_2, \dots, x_N$ is acceptable.

V. SIMULATIONS

In this section, we conduct simulation experiments to illustrate the methodology of parameter estimation in *IAR* model. We indicate the mean square error of interval (*MSEI*) as the evaluation criterion in this section.

Our simulated data is generated from the *IAR* model

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \tilde{\varepsilon}_t, \quad t = 3, 4, \dots, n$$

(the order is $p=2$), where $\tilde{\varepsilon}_t = (c_{\varepsilon_t}; r_{\varepsilon_t})$, $t = 3, 4, \dots, n$ is a generalized random interval, $c_{\varepsilon_t}, r_{\varepsilon_t}$ are independent normal white noise processes $WN(0, \sigma_c^2)$, $WN(0, \sigma_r^2)$, and let $a_1 = 0.85, a_2 = 0.152, \sigma_c = \sigma_r = 0.1$. Fix $X_1 = (10; 1), X_2 = (10.1; 0.9)$, then we can compute

$$\begin{aligned} X_t &= a_1 X_{t-1} + a_2 X_{t-2} + \tilde{\varepsilon}_t \\ &= 0.85 X_{t-1} + 0.152 X_{t-2} + \tilde{\varepsilon}_t, \quad t = 3, 4, \dots, n, \end{aligned}$$

where $\tilde{\varepsilon}_t = (c_{\varepsilon_t}; r_{\varepsilon_t})$ is the generalized random intervals and $c_{\varepsilon_t}, r_{\varepsilon_t}$ are generated by normal random variables $N(0, \sigma_c^2) = N(0, 0.1^2)$, $N(0, \sigma_r^2) = N(0, 0.1^2)$ respectively.

Take the sample size n to be 50, Fig.1. shows a simulated orbit generated by the above *IAR* model.

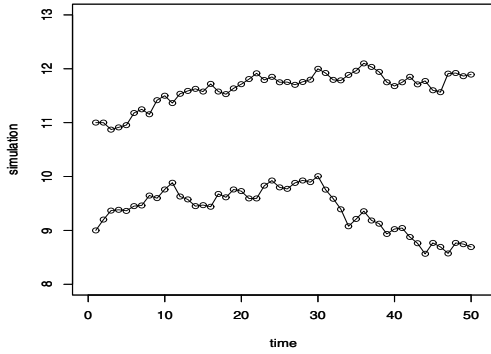


Fig.1.

We use the method introduced in Section 4.1 to estimate the autoregressive coefficient.

Now we simulate $\{X_t\}$ 500 times. Then we obtain 500 different estimators $\hat{a}_1, \hat{a}_2$ of autoregressive coefficients a_1, a_2 . The mean values of this 500 different estimators are 0.847818 and 0.1541758 (the actual values are 0.85 and 0.152). Also, we get 500 estimators $\hat{\sigma}_c, \hat{\sigma}_r$ of standard deviation of white noise σ_c, σ_r , their mean value are 0.0989, 0.0989 (the actual values are 0.1, 0.1).

Again, we simulate $\{X_t\}$ 1000 times. Then we get 1000 different estimators $\hat{a}_1, \hat{a}_2$ of autoregressive coefficients a_1, a_2 . The mean values of this 1000 different estimators are 0.8494906 and 0.1524187 (the actual values are 0.85 and 0.152). Also, we get 1000 estimators $\hat{\sigma}_c, \hat{\sigma}_r$ of standard deviation of white noise σ_c, σ_r , their mean value is 0.0991782, 0.0991782 (the actual values are 0.1, 0.1).

Such results show that the method of parameter estimation in Section 4.1 is fairly good.

VI. EMPIRICAL STUDY

In this section, we use an example to illustrate the parameter estimation and forecasting technique in *IAR* model.

While using the traditional time series models to forecast the stock price, we usually take the closing prices as the analyzing data. Then the information of fluctuations of stock price during each day is not able to be utilized. And the forecast value is also single-valued, so the information provided to the decision makers is also lack of flexibility. In the following, we will take the highest and the lowest daily price as the upper and lower boundary of the interval respectively, and the forecast will also be the interval-valued.

Here we collect data of the highest and lowest daily price of the Shanghai Stock Index from April 1 to June 17 in the year of 2009 (53 trading days in all). See Fig.2.

To select the order, $p = 2$, using the method described in Section 4.1, we obtain the estimated values of autoregressive coefficients, which are $\hat{a}_1 = 1.033, \hat{a}_2 = -0.03$.

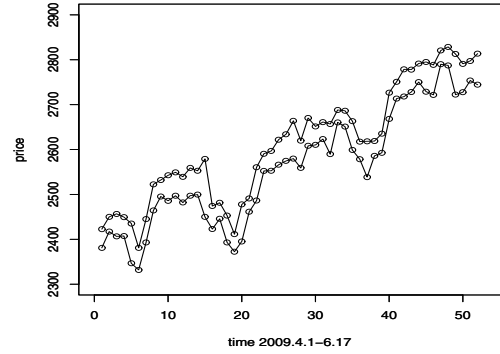


Fig.2.

Then check the residuals

$$\hat{\varepsilon}_{ct} = c_t - \sum_{j=1}^2 \hat{a}_j c_{t-j}, \quad t = 3, 4, \dots, 52,$$

$$\hat{\varepsilon}_{rt} = r_t - \sum_{j=1}^2 \hat{a}_j r_{t-j}, \quad t = 3, 4, \dots, 52,$$

by the method introduced in Section 4.3, $\{\hat{\varepsilon}_{ct}\}$ and $\{\hat{\varepsilon}_{rt}\}$ are white noise processes.

Now we can forecast the stock prices. We use the linear combination of the stock prices of the preceding two trade days X_{t-1}, X_{t-2} with coefficients which are the autoregressive coefficients $\hat{a}_1, \hat{a}_2$:

$$\begin{aligned} \widetilde{X}_t &= \hat{a}_1 X_{t-1} + \hat{a}_2 X_{t-2} \\ &= 1.033 X_{t-1} - 0.03 X_{t-2}, \quad t = 53, 54, \dots \end{aligned}$$

as the forecast interval series.

We compare the forecast interval with the actual interval of the latter ten trading days.

The forecast intervals of latter ten trading days are
[2752.169, 2822.535], [2821.281, 2865.241],
[2856.997, 2896.092], [2893.885, 2933.112],

[2849.159,2950.418], [2890.083,2931.713],
 [2918.681,2956.444], [2919.119,2946.144],
 [2927.386,2987.030], [2963.269,3006.872].

The actual intervals of the latter ten trading days are

[2810.85,2855.42], [2847.36,2886.50],
 [2884.13,2923.24], [2841.90,2941.06],
 [2880.29,2923.47], [2909.09,2946.90],
 [2910.35,2937.61], [2918.39,2976.92],
 [2953.36,2997.27], [2947.69,3009.71].

The mean square error of interval (MSEI) of this ten-day forecasting is 729.2548. As is shown in Fig.3., the red lines depict the forecast interval series while the black lines describe the actual interval series.

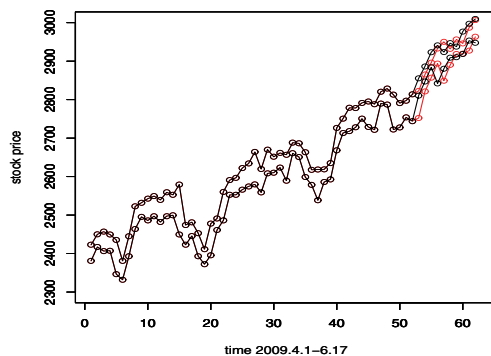


Fig.3.

VII. CONCLUSIONS

In this paper, based on the concepts of interval time series and efficiency evaluation of forecasting for interval time series, we suggest a new type of interval time series (*IAR* model), and present the methods of parameter estimation and forecasting in this model. Furthermore, we apply it to forecast the Shanghai Stock Index.

Both the examples regarding simulated data and real data show that the methods of parameter estimation are pretty good, which is based on the efficiency evaluation of forecasting for interval-valued data. The evaluation criterion we chosen is *MSEI*, and the parameter estimation based on other evaluation criteria can also be derived.

One advantage of the *IAR* model is that it is possible for decision makers to know the best and worst possible situations, and it requires fewer amount of observations than traditional time series models.

ACKNOWLEDGMENT

This research is partially supported by PHR(IHLB), NSFC(No.10771010) and NSFB(No.1112001).

REFERENCES

- [1] H. Akaike, "Information theory and an extension of the maximum likelihood principle", in The Proceeding of the 2nd International Symposium on Information Theory, B.N. Petrov and F. Csaki(eds), Akademia Kiado, Budapest, 1973.
- [2] Z. Artstein and R.A. Vitale, "A strong law of large numbers for random compact sets", Ann. Probab., vol:3, pp. 879-882, 1975.

- [3] R. Aumann, "Integrals of set valued functions", J. Math. Anal. Appl., vol: 12, pp. 1-12, 1965.
- [4] G.P. Box and G.M. Jenkins, Time Series Analysis: Forecasting and Control, Holden-day Inc., San Francisco, CA, 1976.
- [5] P.J. Brockwell and R.A. Davis, Introduction to time series and forecasting, Springer, 2002.
- [6] F. Hiai and H. Umegaki, "Integrals, conditional expectations and martingales of multivalued functions", J. Multivar. Anal., vol:7, pp. 149-182, 1977.
- [7] H.L. Hsu and B. Wu, "Evaluating forecasting performance for interval data", Computers and Mathematics with Applications, vol: 56, pp. 2155-2163, 2008.
- [8] H.L. Hsu and B. Wu, "On the forecasting accuracy of fuzzy time series", to appear, 2010.
- [9] T.L. Lai and H. Xing, Statistical Model and Methods for Financial Markets, Springer, 2007.
- [10] S. Li, Y. Ogura and V. Kreinovich, Limit Theorems and Applications of Set-Valued and Fuzzy Set-Valued Random Variables, Kluwer Academic Publishers, Dordrecht, 2002.
- [11] Q. Song and B.S. Chissom, "Forecasting enrollments with fuzzy time series - part 1", Fuzzy Set and Systems, vol:54, pp. 1-9, 1993.
- [12] Q. Song and B.S. Chissom, "Fuzzy time series and its models", Fuzzy Set and Systems, vol:54, pp. 269-277, 1993.
- [13] Q. Song and B.S. Chissom, "Forecasting enrollments with fuzzy time series - part 2", Fuzzy Set and Systems, vol:62, pp. 1-8, 1994.
- [14] R.S. Tsay, Analysis of Financial Time Series, Wiley, New York, 2005.
- [15] F. Tseng, G. Tzeng, H. Wu and B. Yuan, "Fuzzy ARIMA model for forecasting the foreign exchange market", Fuzzy Sets and Systems, vol:118, pp. 9-19, 2001.
- [16] W. Zhang, S. Li, Z. Wang and Y. Gao, Set-Valued Stochastic Processes, Science Publisher (in Chinese), 2007.