Chapter 16
Building Fuzzy Autocorrelation Model and Its Application to the Analysis of Stock Price Time-Series Data

Yoshiyuki Yabuuchi and Junzo Watada

Abstract. The objective of economic analysis is to interpret the past, present or future economic state by analyzing economic data. Economic analyses are typically based on the time-series data or the cross-section data. Time-series analysis plays a pivotal role in analyzing time-series data. Nevertheless, economic systems are complex ones because they involve human behaviors and are affected by many factors. When a system includes substantial uncertainty, such as those concerning human behaviors, it is advantageous to employ a fuzzy system approach to such analysis. In this paper, we compare two fuzzy time-series models, namely a fuzzy autoregressive model proposed by Ozawa *et al.* and a fuzzy autocorrelation model proposed by Yabuuchi and Watada. Both models are built based on the concepts of fuzzy systems. In an analysis of the Nikkei Stock Average, we compare the effectiveness of the two models. Finally, we analyze tick-by-tick data of stock dealing by applying fuzzy autocorrelation model.

Keywords: fuzzy AR model, fuzzy autocorrelation, possibility, economic analysis.

1 Introduction

Many econometric models including the time-series model have been proposed for evaluating economic systems. In this paper, we propose the fuzzy time-series model, which is employed in the analysis of an economic system.

Yoshiyuki Yabuuchi
Shimonoseki City University, Faculty of Economics; 2-1-1 Daigaku-cho, Shimonoseki, Yamaguchi 751-8510 Japan
e-mail: yabuuchi@shimonoseki-cu.ac.jp

Junzo Watada
Waseda University, Graduate School of Information, Production and Systems; 2-4 Hibikino, Wakamatsu, Kitakyushu, Fukuoka 808-0196 Japan
e-mail: junzow@osb.att.ne.jp

W. Pedrycz & S.-M. Chen (Eds.): Time Series Analysis, Model. & Applications, ISRL 47, pp. 347–367.
DOI: 10.1007/978-3-642-33439-9_16

The objective of economic analysis is to precisely understand the past and present states of an economic system based on the statistical data [1, 21]. However, in an economic system, the state is closely related to many factors that are triggered by the aggregation of numerous human behaviors [24].

Therefore, it becomes insufficient to interpret such an economic system with the use of conventional statistical methods. Instead, it is desirable to apply the concept of the fuzzy system theory which can handle the ambiguity of a structure when we analyze an economic system with many vague factors.

In addition, illustrating a time-series system and making a prediction by a time-series model, estimates are far from observed values. It is a natural interpretation of one reason that a possibility of a time-series system makes data fluctuating. Under the consideration that a Box-Jenkins model can describe a time-series with high accuracy, we propose a fuzzy autocorrelation model based on a Box-Jenkins model to describe a time-series system.

Ozawa *et al.* have proposed a fuzzy autoregressive model [11]. Their model describe the system in terms of error term instead of the coefficient mentioned above. The Ozawa's model, the autoregressive parameter is constructed by including time-series data, minimizing the vagueness of the model, and has real values. In addition, their model express the vagueness of the time-series system by the error term written by fuzzy numbers. The characteristics of these two models are compared by using numerical examples.

We analyze the Nikkei stock average by employing both the fuzzy autoregressive model which was first proposed by Ozawa *et al.* [11], and the fuzzy autocorrelation model, which is being proposed in this paper. Furthermore, the fuzzy autocorrelation model is applied to an economic analysis by analyzing the tick-by-tick data of stock prices. This enables us to forecast a future trend by a sequential prediction that fits the present condition.

The structure of this chapter is organized as follows: In Section 2, a fuzzy time-series analysis based on the fuzzy auto-correlation model will be built by expanding the Box-Jenkins model. In Section 3, we compare the characteristics of the fuzzy autoregressive model with those of the fuzzy autocorrelation model by analyzing the Nikkei stock average. In Section 4, we analyze the tick-by-tick data of stocks by applying the fuzzy autocorrelation model.

2 Fuzzy Time-Series Model

This section elaborates on various fuzzy time-series analysis models. Then, the fuzzy autoregressive model proposed by Ozawa *et al.* and our fuzzy autocorrelation model are described.

2.1 Various Fuzzy Time-Series Analysis Models

Let us review several fuzzy time-series models before discussing our fuzzy autocorrelation model and the conventional fuzzy autoregressive model.

L.A. Zadeh proposed fuzzy set theory [29] in 1965. It helps deal with qualitative data such as linguistic ones by using membership functions.

Based on these concepts, a fuzzy time-series analysis can be built. The models are classified into three groups: (1) a regression model-based analysis, (2) a Box-Jenkins model-based analysis and (3) a fuzzy reasoning(If-Then rule)-based analysis.

These three groups and others are reviewed below.

2.1.1 Regressive Model-Based Analysis of Fuzzy Time-Series Data

In general, a fuzzy regression-based time-series analysis employs the fuzzy regression model proposed by Tanaka *et al.* [16], in which the vagueness included in the target system is addressed with the use of the fuzzy regression coefficients

$$Y_i = (a_1, c_1)x_{i1} + (a_2, c_2)x_{i2} + \cdots + (a_p, c_p)x_{ip} = (\mathbf{a}, \mathbf{c})\mathbf{x}_i \tag{1}$$

where, $\mathbf{x}_i$ are explanatory variables, $\mathbf{a} = [a_1, a_2, \cdots, a_p]$ denotes the center position of the fuzzy coefficients in the vector and $\mathbf{c} = [c_1, c_2, \cdots, c_p]$ are the widths of the fuzzy coefficients. The fuzzy regression model basically deals with the vagueness included in the system and treated with intervals that can express all of the possibilities by including all of the samples. Therefore, the estimated values, Y_i, of the observed values y_i are expressed as fuzzy numbers, and the coefficients are viewed as fuzzy coefficients. In this formulation, the fuzzy coefficients $(\mathbf{a}, \mathbf{c})$ can be obtained by solving a certain Linear Programming (LP) problem.

Watada *et al.* employ the fuzzy regression model to time-series data, and using fuzzy coefficients, they build a fuzzy regression model that includes all the vagueness of time-series system [22, 23]. However, they change the formulation of the fuzzy regression model to account the smoothing of time-series data.

2.1.2 Box-Jenkins Model-Based Analysis of Fuzzy Time-Series Data

Similarly to Box-Jenkins model, fuzzy time-series analysis includes two models based on fuzzy numbers and one model dealing with numeric data. Models dealing with fuzzy numbers are proposed by Yabuuchi *et al.* and Ozawa *et al.*. Yabuuchi *et al.* interpreted time-series data from a possibilistic point of view, define fuzzy autocorrelation coefficients based on fuzzified data, and build a fuzzy autocorrelation model using these values [24].

The fuzzy autocorrelation model aims to capture the present state or the future state of the time-series process by using past fuzzy time-series data similar to the Box-Jenkins model. Therefore, the fuzzy autocorrelation model is obtained from a fuzzy autocorrelation coefficient that expresses the relationship between fuzzy time-series data by a fuzzy operation. The proposed fuzzy autocorrelation model is explained below.

In contrast to the above model, Ozawa *et al.* proposed a fuzzy autoregressive model [11] that expresses the possibilities of fuzzified difference sequences. The fuzzy autoregressive model is written down as follows:

$$\left.\begin{aligned}\tilde{\mathbf{Z}}_t = \phi_1\mathbf{Z}_{t-1} + \cdots + \phi_p\mathbf{Z}_{t-p} + \mathbf{u}\\ \mathbf{Z}_t \subseteq \tilde{\mathbf{Z}}_t,\ \tilde{\mathbf{Z}}_t = (\tilde{\alpha}_t, \tilde{\beta}_t, \tilde{\delta}_t)\end{aligned}\right\} \tag{2}$$

A time-series data is fuzzified to be used by the fuzzy autoregressive model:

$$\mathbf{Y}_t = \left(z_t \min_{i=1}^{3} z_{t+2-i} / \max_{i=1}^{3} z_{t+2-i},\ z_t,\ z_t \max_{i=1}^{3} z_{t+2-i} / \min_{i=1}^{3} z_{t+2-i} \right) \tag{3}$$

Using the differenced series $\mathbf{Z}_t$, which transformed fuzzy time-series data $\mathbf{Y}_t$ to detrending series, a fuzzy autoregressive model is formed. The fuzzy autoregressive model illustrate the relationship between fuzzy time-series data by a real-valued autoregressive parameter ϕ, and the vagueness of the system by introducing a triangular fuzzy number $\mathbf{u} = (u_\alpha, u_\beta, u_\gamma)$.

F.M. Tseng *et al.* proposed a fuzzy ARIMA model [18]. The fuzzy ARIMA model expresses the possibilities of the time-series system with fuzzy coefficients of the model, which is similar to other fuzzy time-series models.

The fuzzy ARIMA model is expressed as follows:

$$\begin{aligned}\tilde{z}_t =& (a_1, c_1)z_{t-1} + \cdots + (a_p, c_p)z_{t-p}\\ &+ \varepsilon_t - (a_{p+1}, c_{p+1})\varepsilon_{t-1} - \cdots - (a_{p+q}, c_{p+q})\varepsilon_{t-q}\end{aligned} \tag{4}$$

where $z_t = \nabla_d(Z_t - \mu)$, α_i denotes the center position of the fuzzy number and c_i stands for the vagueness of the fuzzy number. In addition, ε_t are independent and identically distributed normal random variables with a zero mean and variance of σ^2.

Although the fuzzy ARIMA model does not deal with seasonal variations, the fuzzy ARIMA model is developed as a fuzzy regression model and has fuzzy parameters [19]. Therefore, F.M. Tseng *et al.* proposed the fuzzy seasonal ARIMA model to deal with the seasonal variations. The fuzzy ARIMA model combines the fuzzy regression model and the seasonal ARIMA model [19].

This Box-Jenkins model based fuzzy time-series analysis models have a fuzzy numbers output to illustrate the possibilities of the system. Additionally, because these models are formulated by the LP problem, the required number of time-series data is less than that required to construct a statistical model.

2.1.3 Fuzzy Reasoning(If-Then Rule)-Based Analysis

Song *et al.* proposed a fuzzy reasoning (IF-Then rule)-based model, which is based on a fuzzy time-series model [13, 14, 15].

Song and Chisson describe the relationships between fuzzy time-series data using IF-Then rules and build a fuzzy time-series model to express these rules [13, 14, 15]. Various models have been proposed based on Song *et al.*'s fuzzy time-series model. Yu proposed the method to employ Song *et al.*'s fuzzy time-series model after assigning weighs to emphasize data near to the time point in the fuzzy time-series data [28].

Teoh *et al.* improved a model to accept intuitive and subjective opinions using fuzzy logical rules proposed by Song *et al.*'s model. Cheng *et al.* indicated that Song *et al.* did not consider any logical relations among rules and proposed a novel forecasting model that considers the similarities among fuzzy logical relations [3].

Teoh *et al.* created fuzzy logical relations based on rough sets theory and built a model with the rules generated by Song *et al.*'s model [17].

Additional fuzzy time-series models that improve upon Song *et al.*'s model have been proposed as well [2, 6]

2.1.4 Some Other Fuzzy Time-Series Models

Many other useful fuzzy time-series models have been proposed. R. Dong *et al.* proposed a granular time series approach that employs a fuzzy clustering technique to construct an original series [4]. This model uses long-term forecasting and trend forecasting.

M. Khashei *et al.* proposed a hybrid model, in which the ARIMA models are integrated with artificial neural networks and fuzzy logic to move beyond the linearity of a model [8]. However, O. Valenzuela *et al.*'s model is an integrated ARMA model, using both neural networks and fuzzy logic [20]. K. Lukoseviciute *et al.*'s model is employed using an evolutionary algorithm [10].

T. Partal *et al.* proposed a method that combines a discrete wavelet transform and a neuro-fuzzy method [12].

C.H.L. Lee *et al.*'s approach employs the Japanese candlestick theory [9]. The theory assumes that the candlestick patterns reflect the psychology of the market, and the investors can make their investment decision based on the identified candlestick patterns. With this approach, a vague candlestick patterns is transformed by fuzzy linguistic variables, and the financial time series data is transformed by fuzzy candlestick patterns. The objective of this approach is to understand the vagueness of investors and markets, which is expressed with fuzzy linguistic variables.

J.T Yao *et al.* proposed the rough set model [27]. The rough set model captures the vagueness and uncertainty of time-series data.

2.2 Fuzzy Autoregressive Model

Let us assume that all the fuzzy time-series data $\mathbf{Z}_t$ be defined by triangular fuzzy numbers. Triangular fuzzy numbers are defined by three parameters, and are denoted as $\mathbf{Z}_t = (\alpha_t, \beta_t, \delta_t), (\alpha_t \leq \beta_t \leq \delta_t)$. The inclusion relation of triangular fuzzy numbers is defined through the following inequalities:

$$\mathbf{Z}_t \subseteq \mathbf{Z}_s \Leftrightarrow \{\alpha_t \geq \alpha_s,\ \delta_t \leq \delta_s\} \tag{5}$$

In the same way, the arithmetic operations are defined as follows:

$$\mathbf{Z}_t + \mathbf{Z}_s = (\alpha_t + \alpha_s, \beta_t + \beta_s, \delta_t + \delta_s)$$
$$\mathbf{Z}_t - \mathbf{Z}_s = (\alpha_t - \alpha_s, \beta_t - \beta_s, \delta_t - \delta_s)$$
$$p \cdot \mathbf{Z}_t = \begin{cases} (p \times \alpha_t, p \times \beta_t, p \times \delta_t), \ p \geq 0 \\ (p \times \delta_t, p \times \beta_t, p \times \alpha_t), \ p < 0, \end{cases}$$

where p is a real number.

A fuzzy autoregressive model is specified as follows:

$$\tilde{\mathbf{Z}}_t = \phi_1 \mathbf{Z}_{t-1} + \cdots + \phi_p \mathbf{Z}_{t-p} + \mathbf{u}$$
$$\mathbf{Z}_t \subseteq \tilde{\mathbf{Z}}_t, \ \tilde{\mathbf{Z}}_t = (\tilde{\alpha}_t, \tilde{\beta}_t, \tilde{\delta}_t) \tag{6}$$

Based on (5) and (6), it is clear that the following relations hold.

$$\alpha_t \geq \tilde{\alpha}_t, \delta_t \leq \tilde{\delta}_t$$

Namely, the fuzzy time-series model includes all of the fuzzy data. The autoregressive parameters $\phi_1, \phi_2, \cdots, \phi_p$ have real values, and show the degree the fuzzy time-series data depend on the past. An error term is the constant specific to the model and refers to the part of the fuzzy data that do not depend on the past data. This term is defined as a triangular fuzzy number:

$$\mathbf{u} = (u_\alpha, u_\beta, u_\delta)$$

A fuzzy autoregressive model results through solving a problem of linear programming that minimizes that ambiguity of the model according to the inclusion condition (6) as follows:

$$\begin{aligned} &\text{minimize} \ \sum_{t=p+1}^{n} (\tilde{\delta}_t - \tilde{\alpha}_t) \\ &\text{subject to} \ \alpha_t \geq \tilde{\alpha}_t, \\ &\qquad \delta_t \leq \tilde{\delta}_t \\ &\qquad \quad (t = p+1, p+2, \cdots, n) \\ &\qquad u_\alpha \leq u_\delta \end{aligned} \tag{7}$$

2.3 Fuzzy Autocorrelation Model

Even if we had described the behavior of time-series system by using a time-series model, the estimated values should have a near value from the observed data. The understanding of the behavior is shaped so as that the possibility of the time-series system is making it natural. Our model describes the possibility of the time-series system by the coefficients. Let us employ a triangular fuzzy number here, since it is manageable.

When different sequences are employed, trend and noise can be easily removed. The autocorrelation makes the model describing the behaviors easily. Therefore, we fuzzify the Box-Jenkins model.

In the fuzzy autocorrelation model, the time-series data z_t are transformed into a fuzzy number to express the possibilities of the data. The following fuzzy equation shows the case in which only one time point before and after the time point t is taken into consideration in building a fuzzy number [23].

$$\mathbf{Y}_t = (Y_t^L, Y_t^C, Y_t^U) = (\min(z_{t-1}, z_t, z_{t+1}), z_t, \max(z_{t-1}, z_t, z_{t+1})) \tag{8}$$

Next, we employ a calculus of finite differences to filter out the time-series trend data, which enables us to use the first-order difference-equation to write the following:

$$\mathbf{Z}_t = (Z_t^L, Z_t^C, Z_t^U) = (\min(\mathbf{Y}_t - \mathbf{Y}_{t-1}), Y_t^C - Y_{t-1}^C, \max(\mathbf{Y}_t - \mathbf{Y}_{t-1})) \tag{9}$$

Generally, if we take the finite differences then we reduce the trend variation, and only an irregular pattern is included in the difference series. However, when we use the fuzzy operation, the ambiguity may increase, and the value of an autocorrelation coefficient may take values not lower than 1 or not greater than -1. To solve this problem in the case of the fuzzy operation, we adjust the width of a fuzzy number using α-cut when determing the difference series. An α-cut level h is determined from the value of the autocorrelation. When we calculate the fuzzy autocorrelation, we employ the usual fuzzy operation under the condition that the fuzzy autocorrelation of lag 0 is set to $\rho_0 = \lambda_0/\lambda_0 = (1,1,1)$, which results in the following linear programming to decide the value at the α-cut level. When we set the α-cut level to 1, the ambiguity of the fuzzy autocorrelation is the smallest, but we cannot obtain the fuzzy autocorrelation that reflects the possibility of the system. Therefore, we maximize the width of the autocorrelation. However, the size of the width is decided automatically as the value of autocorrelation should be included in [-1,1].

$$\begin{aligned} \underset{h}{\text{maximize}} \quad & \sum_{i=1}^{p} (\rho_i^U - \rho_i^L) \\ \text{subject to} \quad & \rho_i^U \le 1 \\ & \rho_i^L \ge -1 \\ & \rho_i^L \le \rho_i^C \le \rho_i^U \\ & \qquad (i = 1, 2, \cdots, p) \end{aligned} \tag{10}$$

We can define the fuzzy covariance and the fuzzy autocorrelation as follows:

$$\begin{aligned} \Lambda_k &\equiv Cov[\mathbf{Z}_t \mathbf{Z}_{t-k}] = E[\mathbf{Z}_t \mathbf{Z}_{t-k}] = [\lambda_k^L, \lambda_k^C, \lambda_k^U] \\ \mathbf{r}_k &= \Lambda_k / \Lambda_0 = [\rho_k^L, \rho_k^C, \rho_k^U] \end{aligned}$$

We adjust the ambiguity of the difference series by employing the α-cut level h, which is obtained by solving the above linear programming. Using the fuzzy autocorrelation coefficient which is calculated by employing α-cut level h, we redefine the Yule-Walker equations as in linear programming and calculate the partial autocorrelation.

We form the following autoregressive process.

$$\mathbf{Z}_t = \boldsymbol{\Phi}_1 \mathbf{Z}_{t-1} + \boldsymbol{\Phi}_2 \mathbf{Z}_{t-2} + \cdots + \boldsymbol{\Phi}_p \mathbf{Z}_{t-p}$$

where $\boldsymbol{\Phi} = [\phi^L, \phi^C, \phi^U]$ is a fuzzy partial autoregressive coefficient.

As mentioned above, the next observation value either exceeds the observed value at present by the size of the value of autocorrelation or it is less than the observed value. For this reason, autocorrelation is important to the time-series analysis. Therefore, we build a model that illustrates the ambiguity of the system captured by the fuzzy autocorrelation. The reason for the autocorrelation is also fuzzy autocorrelation, Yule-Walker equations can also be viewed as the fuzzy equation in the same way.

$$\mathbf{R}_t = \boldsymbol{\Phi}_1 \mathbf{r}_{t-1} + \boldsymbol{\Phi}_2 \mathbf{r}_{t-2} + \cdots + \boldsymbol{\Phi}_p \mathbf{r}_{t-p} \tag{11}$$

$\boldsymbol{\Phi}$ shown in (11) is an unknown coefficient. We are building the model in terms of fuzzy autocorrelation which can describe the ambiguity of the system. However, when the ambiguity of a model is large, the relationship between a model and a system becomes ambiguous. Therefore, the possibility of the system cannot be properly described. Therefore, to obtain the fuzzy partial autocorrelation coefficient, for which the ambiguity of a time-series model should be minimized, we have the following linear programming:

$$\begin{aligned}
\text{minimize } & \sum_{i=1}^{p} (\rho_t^U - \rho_t^L) \\
\text{subject to } & R_t^U \geq \rho_t^U \\
& R_t^C = \rho_t^C \\
& R_t^L \leq \rho_t^L \\
& \rho_t^L \leq \rho_t^C \leq \rho_t^U \\
& \quad (t = 1, 2, \cdots, p)
\end{aligned} \tag{12}$$

As mentioned above $\mathbf{R}$ is obtained by the fuzzy operation employing the fuzzy autocorrelation $\mathbf{r}$ and fuzzy partial autocorrelation $\boldsymbol{\Phi}$. R^L, R^C and R^U represent the lower limit, the center, and the upper limit of $\mathbf{R}$, respectively.

A fuzzy autocorrelation model expresses the possibility that the change of the system is realized in the data, which is different from the conventional statistical method. We are building a model that can show an ambiguous portion called a possibility that has not been clearly expressed through conventional statistics techniques.

The time-series data was fuzzified by using Equation (8), and a fuzzy operation was employed to calculate any coefficients. Then, the center of all coefficients is coincident to the non-fuzzy coefficient. In addition that, the center of our model is coincident to the non-fuzzy model, the autoregressive model by the constraint of LP problem (12).

3 A Numerical Example

In this section, we employ the Nikkei stock average which indicates the trend of the whole stock market as an index of the Japanese stock market. We use the monthly data from 1970 to 1998.

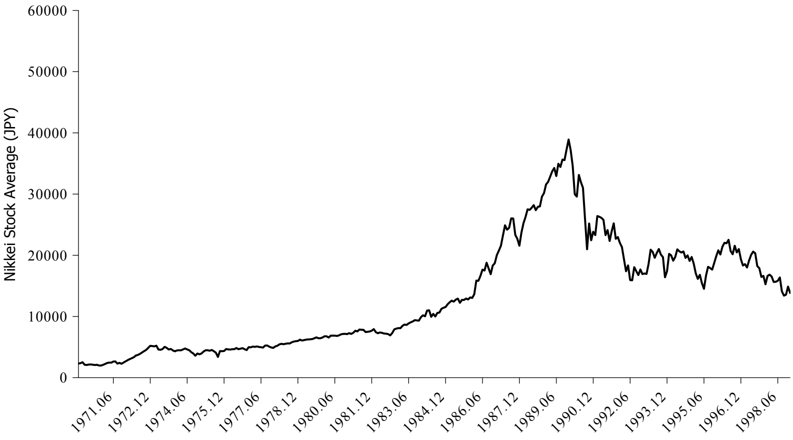

Fig. 1 Nikkei Stock Average in 1970 to 1998

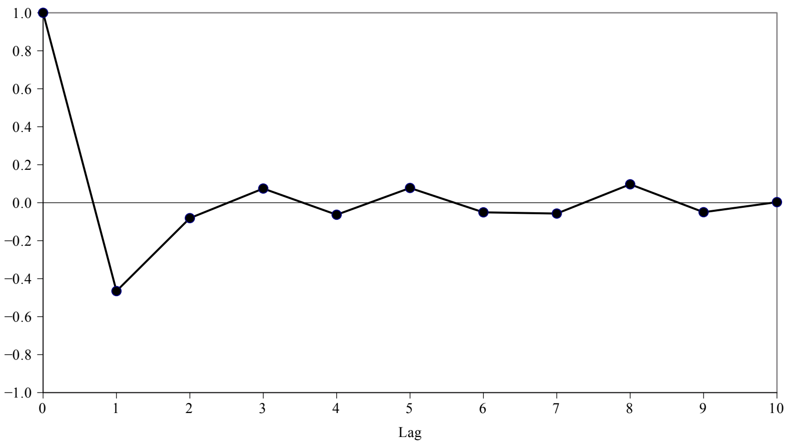

Fig. 2 Autocorrelation of Nikkei Stock Average

We show the sample autocorrelation coefficient at each time lag (Figure 2) to determine the order. Figure 2 shows the negative correlation in lags 1 and 2 where the sign of the autocorrelation coefficient changes from minus to plus. Because of this result, we analyze the Nikkei stock average by employing the AR(2) model of the second-order.

Furthermore, because of the existing seasonal variation in this data, we employ the calculation $\triangledown^2 \triangledown_{12} z_t$ which take the first-order seasonal difference of every 12-month period after taking the second-order difference.

$$\triangledown^2 \triangledown_{12} \tilde{\mathbf{Z}}_t = \phi_1 \triangledown^2 \triangledown_{12} \mathbf{Z}_{t-1} + \phi_2 \triangledown^2 \triangledown_{12} \mathbf{Z}_{t-2} + \mathbf{u}$$

where the data z_t in analysis are statistical data and the actual measurement. $\mathbf{u}$ is an error term of the model. For the ambiguity of the time-series system to reflect these data, we employ fuzzy numbers to deal with these data.

3.1 Fuzzy Autoregressive Model

We analyze the data by employing the fuzzy autoregressive model with the triangular fuzzy number. The following is the procedure to obtain the fuzzy autoregressive model.

Step 1. The original series is fuzzified with the use of Equation (3) and transformed difference without trends. We employ the difference $\nabla^2 \nabla_{12} \mathbf{Z}_t$ since the differebce series is detrending.

Step 2. In order to minimize the vagueness of the model, the autoregressive parameters ϕ_1, ϕ_2 and an error term $(u_\alpha, u_\beta, u_\delta)$ was determined by LP problem (7). Here, the constraint of LP problem (7) is expressed that the time-series model can include difference series.

Ozawa's model is obtained by using the above procedures. The coefficients of this model are determined as follows:

$$\triangledown^2 \triangledown_{12} \tilde{\mathbf{Z}}_t = -0.749 \triangledown^2 \triangledown_{12} \mathbf{Z}_{t-1} - 0.348 \triangledown^2 \triangledown_{12} \mathbf{Z}_{t-2} + (-0.143, -0.013, 0.117)$$

The model that is obtained by the fuzzy autoregressive model has a negative coefficient that is the same as the result that is obtained by the autocorrelation. An original series and its estimate are shown in Figure 3.

Figure 3 shows that the estimated model has a large width of possible values. Numerically, the width of the possibility of the model is 12,500JPY (on average), 41,340JPY at the maximum and 1,770JPY at the minimum. The ambiguity of this model is extremely large.

However, the central value of the estimated value shows a value that is almost the same as the original series. Because it showed a strongly oscillating tendency in the past, these results are with the large width of the model estimation. This can also be understood from the section of the error term of the model.

A result of the sequential prediction of this model is shown in Figure 4.

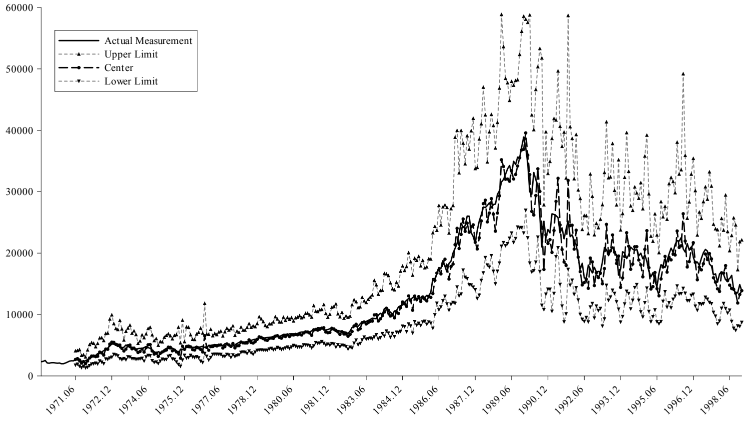

Fig. 3 The result of the Fuzzy Autoregressive Model

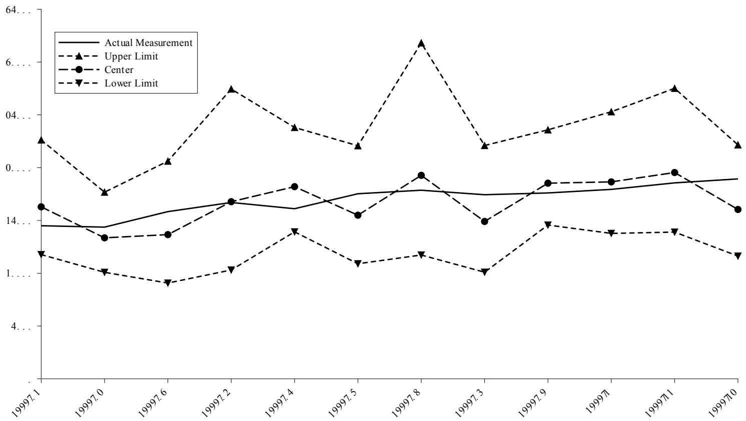

Fig. 4 The prediction result obtained by the Fuzzy Autoregressive Model (1999)

3.2 Fuzzy Autocorrelation Model

Next, we analyze the Nikkei stock average by employing the fuzzy autocorrelation model which is proposed in this paper. The following is the procedure to obtain our model.

Step 1. The original series is fuzzified by Equation (8) and transformed difference without trends. We employ the difference $\nabla^2\nabla_{12}\mathbf{Z}_t$ since the differebce series is detrending.

Step 2. The fuzzy autocorrelation coefficients was obtained by fuzzy operation. The fuzzy autocorrelation coefficient has a large vagueness because fuzzy operation. Therefore, the vagueness is managed using an $\alpha-$cut method without removing characteristics of a fuzzy autocorrelation coefficients. The $\alpha-$cut value h is determined by LP problem (10). In this model, we set the $\alpha-$cut level 0.978. The fuzzy autocorrelation showed minus in the correlation of lag 2 is similar to the case of the autocorrelation.

Step 3. The fuzzy autocorrelation coefficients is determined by LP problem (12).

In the estimated fuzzy autocorrelation model, the coefficient was determined as follows:

$$\begin{aligned}\bigtriangledown^2 \bigtriangledown_{12} \tilde{\mathbf{Z}}_t &= (-1,-0.642,-0.642) \bigtriangledown^2 \bigtriangledown_{12} \mathbf{Z}_{t-1} \\ &\quad +(-0.607,-0.380,-0.321) \bigtriangledown^2 \bigtriangledown_{12} \mathbf{Z}_{t-2}\end{aligned}$$

The model that is obtained by the fuzzy autocorrelation model has a negative coefficient that is the same as the result that is obtained by the fuzzy autoregressive model.

The original series and estimated series are shown in Figure 6.

As shown in Figure 6, the estimated model has a small width and results in the low level of fuzziness. Numerically, the width of the possibility of the model is 2,500JPY on average, 18,000JPY at the maximum and 100JPY at the minimum. In Figure 6, the width of the results produced by the model is smaller than that shown in Figure 3.

A result obtained by the sequential prediction of this model is shown in Figure 7.

There is a point that is the predicted value which differs from the original series in Figure 7. Because the fuzzy autocorrelation model places stress on the fluctuation of a system unlike the fuzzy autoregressive model which includes all of the possibilities of the system, this model produces a large error.

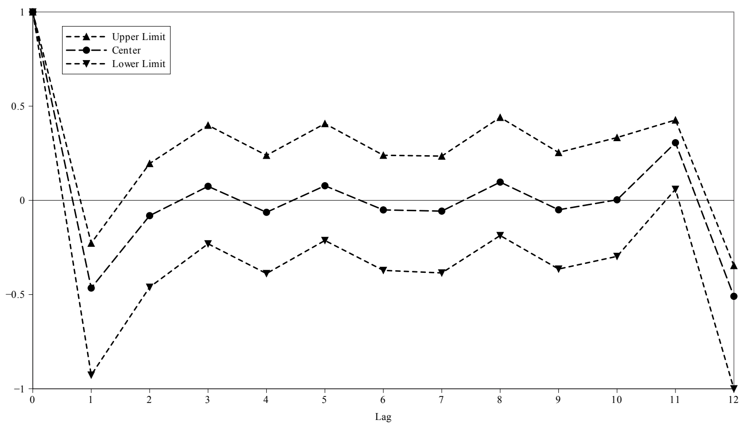

Fig. 5 Fuzzy Autocorrelation

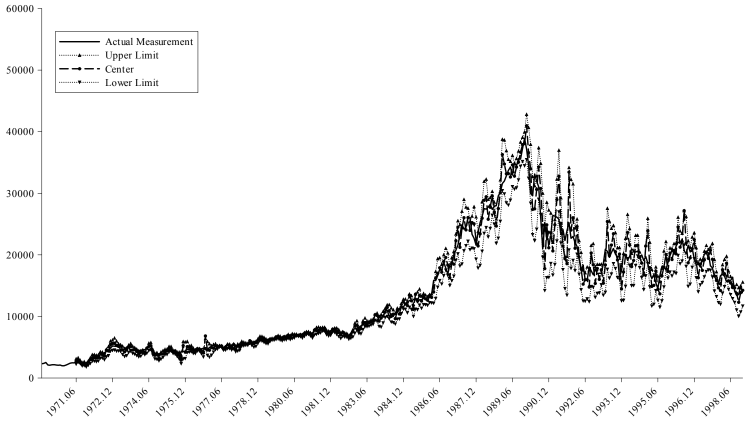

Fig. 6 The result produced by the Fuzzy Autocorrelation Model

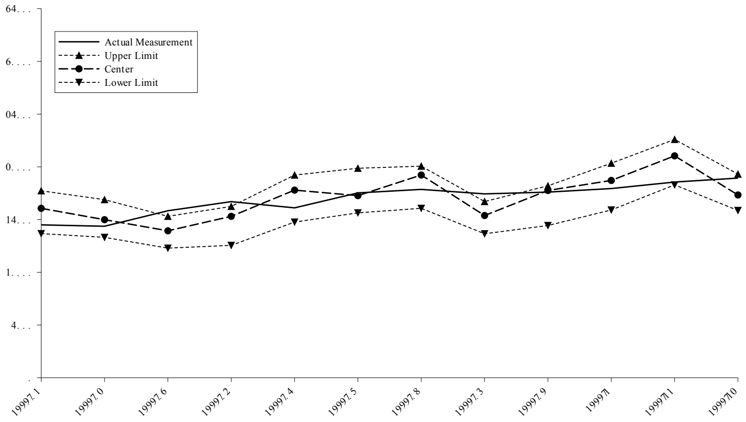

Fig. 7 The predicted result - Fuzzy Autocorrelation Model (1999)

Here, the difference between the fuzzy autoregressive model and our model are summarized as followings:

The fuzzy autoregressive model The fuzzy autoregressive model is determined by including relation between this and difference series. At that time, non-fuzzy coefficients ϕ and fuzzy error term $\mathbf{u}$ are determined by LP in order to minimize the vagueness of model.

Our fuzzy autocorrelation model First, the fuzzy autocorrelation coefficient is obtained from fuzzy difference series by LP. Next, the fuzzy autoregressive coefficient is obtained from a fuzzy autocorrelation coefficient in order to describe the behavior of a fuzzy difference series with high dimensional accuracy by LP.

The difference of two models is illustrated in Figures 3 and 6.

4 Economic Analysis Based on Tick-by-Tick Data

Tick-by-tick data are the record of stock dealing transactions. Let us analyze tick-by-tick data of stocks by applying the fuzzy autocorrelation model.

Tick-by-tick data record five items: the trading date, stock brand, traded time, dealing price and dealing amount.

Table 1 Example of Tick-by-Tick Data

Traded date	Stock name record	Traded time	Traded price	Traded amount
20080602	13010	0900	200.000	6000
20080602	13010	0900	200.000	4000
20080602	13010	0900	200.000	10000
20080602	13010	0900	200.000	1000
20080602	13010	0900	200.000	4000
⋮	⋮	⋮	⋮	⋮
20080602	13010	0900	201.000	1000
20080602	13010	0900	201.000	1000
20080602	13010	0901	202.000	1000
⋮	⋮	⋮	⋮	⋮

We will use the stock trading record. Table 1 shows some of the tick-by-tick data, which is the trading data from 2nd June 2008 recorded from 9 am. From the stating time at 9 am, 6,000, 4,000 and 10,000 shares of brand code 13010 is traded at 200JPY. We have all of the trading records. At a later time, the records have different values. The trading variables change in real time and the stock is traded by the real time price and amount. These real time trading situations are shown in the tick-by-tick data.

However, it is not possible to use the tick-by-tick data directly in a time-series analysis, and the upper, lower and average prices of the same stock at the same time are employed to create fuzzy numbers of time-series data.

In this chapter, we employ one week tick-by-tick data from 7th July 2008 (Monday) to 11th July 2008 (Friday). When nothing was traded for a certain time period, we used the previous price.

First, we apply AR model to analyze the tick-by-tick data. When we took a 5 minutes difference of the central value, we could remove its trend. Figure 8 shows the correlogram time series figure of the tick-by-tick data. As it shows one-time previous value ($\Delta_5 Z_{t-1}$), two-time previous value ($\Delta_5 Z_{t-2}$) and three-time previous value ($\Delta_5 Z_{t-3}$) have a large correlation, the following AR(3) model was employed.

$$\Delta_5 \tilde{Z}_t = \phi_1 \Delta_5 Z_{t-1} + \phi_2 \Delta_5 Z_{t-2} + \phi_3 \Delta_5 Z_{t-3}$$

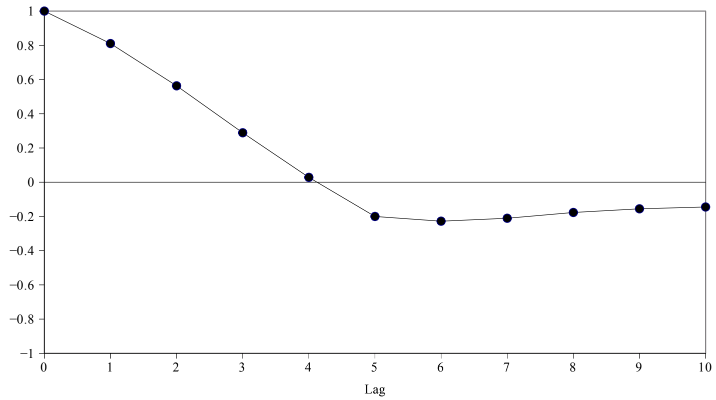

Fig. 8 Correlogram of AR model

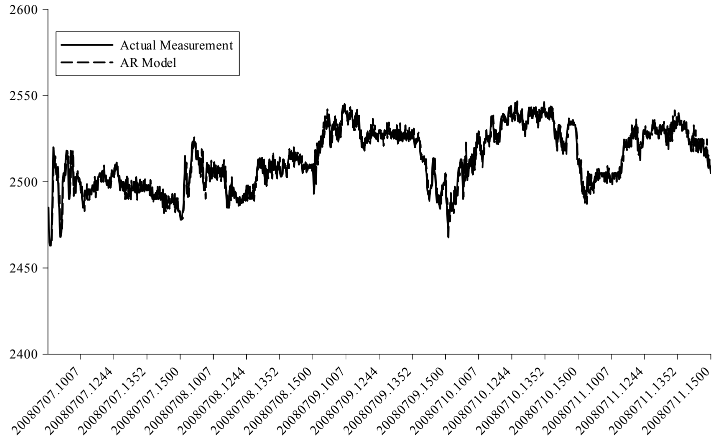

Fig. 9 Tick-by-tick data and the AR(3) model

Solving the Yule-Walker equation we form the following AR(3) model.

$$\Delta_5\tilde{Z}_t = 0.970\Delta_5 Z_{t-1} - 0.041\Delta_5 Z_{t-2} - 0.224\Delta_5 Z_{t-3}$$

$$\phi_1 = 0.970,\ \phi_2 = -0.041,\ \phi_3 = -0.224$$

Figure 9 shows the forecasted result based on the AR(3) model. Tick-by-tick data are given as fuzzy numbers but when using interval values, it is not possible to distinguish between the observed and forecasted values. Therefore, Figure 9 shows only the center value of the traded values and the forecasted values of the AR(3) model.

Figure 9 highlights the success of the forecast but the forecast delay and some error are recognized in the figure. Therefore, in the figure the movement of lines of real prices and forecasted values appear to have thick lines. Additionally, in some places the predicted stock prices are too high.

Figure 10 shows the forecast result by fuzzy autocorrelation model in which the autocorrelation coefficients are shown as fuzzy numbers. A black up-pointing triangle and a black down-pointing triangle in the figure denote the upper and lower values of the autocorrelation coefficients, respectively. The circle in the boxes are the center values of the fuzzy autocorrelation coefficients. Similar to the AR model, the one-time previous value ($\Delta_5 \mathbf{Z}_{t-1}$), two-time previous value ($\Delta_5 \mathbf{Z}_{t-2}$) and three-time previous value ($\Delta_5 \mathbf{Z}_{t-3}$) exhibit a high correlation. Therefore, the following fuzzy autocorrelation model, FAR(3) was employed.

$$\tilde{\mathbf{Z}}_t = \Phi_1 \Delta_5 \mathbf{Z}_{t-1} + \Phi_2 \Delta_5 \mathbf{Z}_{t-2} + \Phi_3 \Delta_5 \mathbf{Z}_{t-3}$$

The fuzzy autocorrelation model was obtained by solving the fuzzy Yule-Walker equation.

$$\begin{aligned} \tilde{\mathbf{Z}}_t = & [0.892, 0.970, 1] \Delta_5 \mathbf{Z}_{t-1} \\ & + [0.041, 0.041, 0.103] \Delta_5 \mathbf{Z}_{t-2} \\ & + [-0.224, -0.224, -0.224] \Delta_5 \mathbf{Z}_{t-3} \end{aligned}$$

Figure 11 shows the forecasting result obtained by the fuzzy autocorrelation model. The figure shows the results are acceptable. The result shows that the high accuracy of the predicted values in comparison with the results produced by the probabilistic AR(3) model.

$$\begin{aligned} \Phi_1 &= [\ 0.892,\ 0.970,\ 1\ ] \\ \Phi_2 &= [-0.041, -0.041,\ 0.103] \\ \Phi_3 &= [-0.224, -0.224, -0.224] \end{aligned}$$

To validate the model, we compared between the forecasted values and the real traded result for the last 30 minutes from 14:30 to 15:00 on 11th July 2008 and the next 30 minutes from 9:00 to 9:30 on 14th July 2008. Figures 12 and 13 show the results of the AR(3) model and the fuzzy autocorrelation model. In these figures, the values forecasted by the model based on the tick-by-tick data from 7th July to 11th July are applied to 14th July. Shown are the latter 30 minutes. In each figure, the observed values are shown by a solid line and the forecasted values by a dotted line.

First, the result using AR(3) looks one timing proceeding in Figure 12. At some points, when the real stock price came down, the forecasted value increased. At another points when the real stock price does not increase, the model forecasted the increasing of the price. Additionally when the stock price fell, the forecasted stock price decreased too rapidly.

However, Figure 13 shows the interval values because the fuzzy autocorrelation model is an interval model. Figure 13 illustrates looks one timing proceeding similar

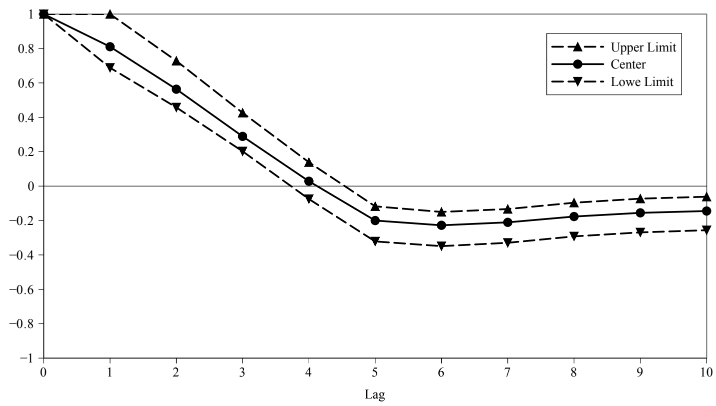

Fig. 10 Correlogram of Fuzzy Autocorrelation Model

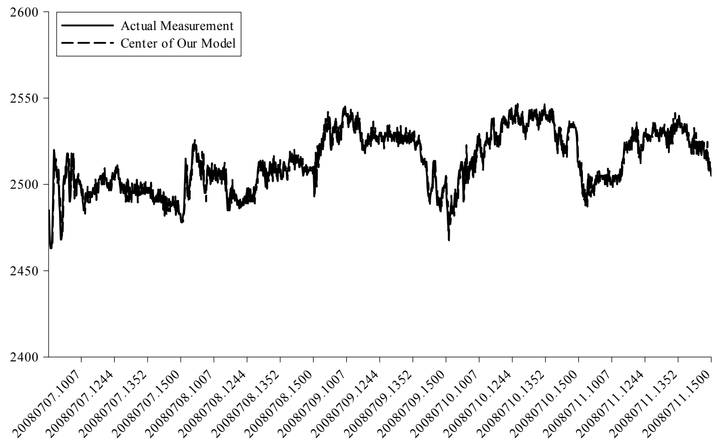

Fig. 11 Center Values of Fuzzy Autocorrelation Model of Tick-by-Tick Data

to the AR(3). Furthermore, for some points, the change of stock prices is forecasted to be too large.

For example, at the closing time on 11th July, the forecasting error looks small at the center of the model but the vagueness of the forecasted value is large. We interpret this as the last minute push of dealings that disturbed the forecasting of the time series system. That is, even though the stock price stopped moving, the forecasted value increased or decreased with the real value. This tendency of the forecasted result explicitly shows the forecasted value of the stocks starting on 14th July.

Overall it was found that the fuzzy autocorrelation model could describe the movement of the time series system well.

We can summarize the result as follows: It is difficult to employ this method to stock trading even though the forecasted values are shown in the interval and even though almost all the values are included in the intervals. If the fuzzy autocorrelation model has the constraint $R_t^C = \rho_t^C (t = 1,2,\cdots,p)$, then the center value of the fuzzy autocorrelation model is coincident to the autoregressive model. Therefore, our model illustrate the possibilities of the the time-series data.

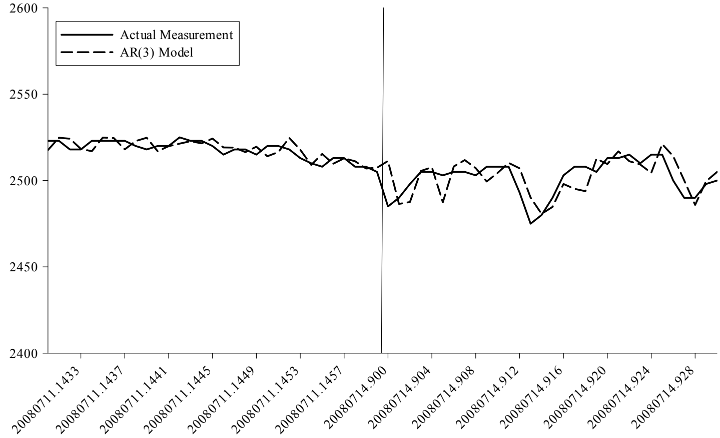

Fig. 12 Validation of AR(3) Model

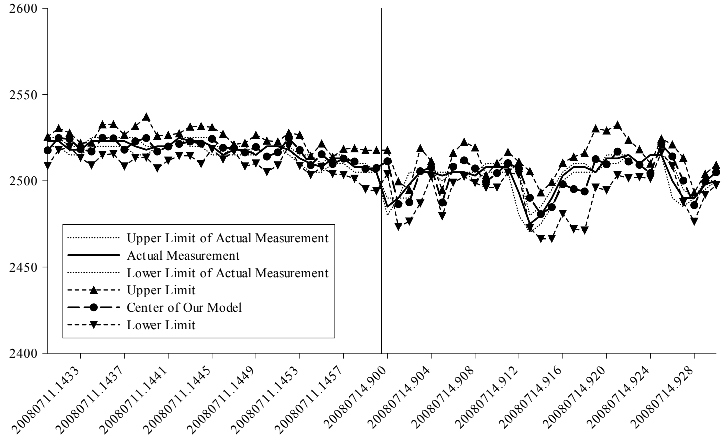

Fig. 13 Validation of the Fuzzy Autocorrelation Model

5 Model Interpretation

In this section, we interpret two fuzzy time-series models, the fuzzy autoregressive model proposed by Ozawa *et al.* and the fuzzy autocorrelation model proposed by Yabuuchi and Watada.

Let us compare the results between a fuzzy autoregressive model and a fuzzy autocorrelation model employing the triangular fuzzy number, which are illustrated in Figures 3 and 6. These figures show that the fuzzy autoregressive model is estimated as the model with a large width of possibility and the fuzzy autocorrelation is estimated as the model with a small width. The ambiguity of the fuzzy autocorrelation model is little, and this model estimated the original series more correctly. However, the fuzzy autoregressive model includes the entire fluctuation of an original series by estimating its fluctuation with a large value.

Let us compare the results of sequential prediction which are illustrated in Figures 4 and 7. Though the width of the possibility of the fuzzy autoregressive model was larger, it should be predicted that the estimated central value corresponds to the actual measurement. In the case of the fuzzy autocorrelation model, the estimated value is different from the original series at three points. Because the fuzzy autocorrelation describes the fluctuation of a system, this is considered to be the cause that an error produces.

The return and the amount of funds for an investor, a company, etc. are greatly influenced by the Nikkei stock average. Therefore, if a prediction is greatly different from a reality, it can cause great damage; in particular, in the case of a company, it may cause a business downsizing or the laying off of company employees. Sometimes, the Japanese economy may be inactive because of the influence of a wrong prediction. Therefore, the lower such a risk is, the more favorable it becomes for the company and the investor when they consider the future return, future prospects, and so on. However, the monthly fluctuation of the real Nikkei stock average is almost less than 10,000JPY. Therefore, for these reasons, we could regard a fuzzy autocorrelation model as being suitable for estimation and prediction of the Nikkei stock average.

In the case of the fuzzy autocorrelation model, we should not include all of the economic data within the model, but we should construct the model by employing the fuzzy autocorrelation to include the possibility of the fluctuation of data. We could show the effectiveness of the economic analysis using the Nikkei stock average by the fuzzy autocorrelation model, because the proposed model could illustrate the fluctuation in the system.

Next, the forecast of stock prices based on the tick-by-tick data of stocks are provided using both the autoregressive model and probabilistic autocorrelation model.

In the analysis of the tick-by-tick data of stocks, past traded data from three times, $\Delta_5 Z_{t-1}$, $\Delta_5 Z_{t-2}$, $\Delta_5 Z_{t-3}$ were used to describe successfully the movement of stock prices.

The purpose of employing tick-by-tick data for stock price forecasting is to validate the forecasting precision of the fuzzy autocorrelation model in stock trading. Figure 13 showed the stock prices of both the model building data and the last 30 minutes for validation. In this validation the figures showed that the fuzzy autocorrelation model could describe successfully the price movement of the time series system.

6 Conclusions

In this paper, we proposed a fuzzy autocorrelation model, which is based on a Box-Jenkins model and describes a time-series system by using fuzzy autocorrelation coefficients. The proposed model has compared with a fuzzy autoregressive model proposed by Ozawa *et al.* through numerical examples, the Nikkei stock average and the tick-by-tick data of stocks.

We showed the width of a fuzzy autocorrelation model is smaller than the one of a fuzzy autoregressive model in the analysis of the Nikkei stock average. But, the center of a fuzzy autoregressive model could estimate the original series.

The fuzzy autocorrelation model had the higher accuracy of the predicted values in the analysis of the tick-by-tick data of stock.

Finally, we can conclude that the fuzzy autocorrelation model can describe the movement of the time-series system well.

References

1. Box, G.P., Jenkins, G.M.: Time Series Analysis: Forcasting and Control. Holden-day Inc., San Francisco (1976)
2. Chen, S.M.: Forecasting enrollments based on fuzzy time series. Fuzzy Sets and Systems 81(3), 311–319 (1996)
3. Cheng, Y.C., Li, S.T., Li, H.Y., Chen, P.C.: A Novel Forecasting Model Based on Similarity Measure for Fuzzy Time Series. Proceedings of Business and Information 8 (2011)
4. Dong, R., Pedrycz, W.: A granular time series approach to long-term forecasting and trend forecasting. Physica A: Statistical Mechanics and its Applications 387(13), 3253–3270 (2008)
5. Dubois, D., Prade, H.: Fuzzy Sets and Systems, Theory and Application. Academic Press, New York (1980)
6. Ismail, Z., Efendi, R.: Enrollment Forecasting based on Modified Weight Fuzzy Time Series. Journal of Artificial Intelligence 4(1), 110–118 (2011)
7. Kariya, T., Liu, R.Y.: Asset Pricing -Discrete Time Approach. Kluwer Academic Publishers (2002)
8. Khashei, M., Bijari, M., Ardali, G.A.R.: Improvement of Auto-Regressive Integrated Moving Average models using Fuzzy logic and Artificial Neural Networks (ANNs). Neurocomputing 72(4-6), 956–967 (2009)
9. Lee, C.H.L., Liu, A., Chen, W.S.: Pattern discovery of fuzzy time series for financial prediction. IEEE Transactions on Knowledge and Data Engineering 18(5), 613–625 (2006)
10. Lukoseviciute, K., Ragulskis, M.: Evolutionary algorithms for the selection of time lags for time series forecasting by fuzzy inference systems. Neurocomputing 12(10-12), 2077–2088 (2010)
11. Ozawa, K., Watanabe, T., Kanke, M.: Forecasting Fuzzy Time Series with Fuzzy AR Model. In: Proceedings, the 20th International Conference on Computers & Industrial Engineering, pp. 105–108 (1996)

12. Partal, T., Kişi, Ö.: Wavelet and neuro-fuzzy conjunction model for precipitation forecasting. Journal of Hydrology 342(1-2), 199–212 (2007)
13. Song, Q., Chissom, B.S.: Forecasting enrollments with fuzzy time series – Part I. Fuzzy Sets and Systems 54(1), 1–9 (1993)
14. Song, Q., Chissom, B.S.: Fuzzy times series and its models. Fuzzy Sets and Systems 54(3), 269–277 (1993)
15. Song, Q., Chissom, B.S.: Forecasting enrollments with fuzzy time series – Part II. Fuzzy Sets and Systems 62(1), 1–8 (1994)
16. Tanaka, H., Watada, J.: Possibilistic Linear Systems and Their Application to The Linear Regression Model. Fuzzy Sets and Systems 27, 275–289 (1988)
17. Teoh, H.J., Cheng, C.H., Chu, H.H., Chen, J.S.: Fuzzy time series model based on probabilistic approach and rough set rule induction for empirical research in stock markets. Data & Knowledge Engineering 67, 103–117 (2008)
18. Tseng, F.M., Tzeng, G.H., Yu, H.C., Yuan, B.J.C.: Fuzzy ARIMA Model for forecasting the foreign exchange market. Fuzzy Sets and Systems 118(1), 9–19 (2001)
19. Tseng, F.M., Tzeng, G.H.: A fuzzy seasonal ARIMA model for forecasting. Fuzzy Sets and Systems 126(3), 367–376 (2002)
20. Valenzuela, O., Rojas, I., Rojas, F., Pomares, H., Herrera, L.J., Guillen, A., Marquez, L., Pasadas, M.: Hybridization of intelligent techniques and ARIMA models for time series prediction. Fuzzy Sets and Systems 159(7), 821–845 (2008)
21. Vandaele, W.: Applied Time Series and Box-Jenkins Model. Academic Press (1983)
22. Watada, J.: Fuzzy time series analysis and forecasting of sales volume. In: Kacprzyk, J., Fedrizzi, M. (eds.) Fuzzy Regression Analysis, pp. 211–227. Omnitec Press, Physica-Verlag, Warsaw, Heidelberg (1992)
23. Watada, J.: Possibilistic Time-series Analysis and Its Analysis of Consumption. In: Dubois, D., Prade, H., Yager, R.R. (eds.), pp. 187–217. John Wiley & Sons, INC. (1996)
24. Yabuuchi, Y., Watada, J., Toyoura, Y.: Fuzzy AR model of stock price. Scientiae Mathematicae Japonicae 60(2), 303–310 (2004)
25. Yabuuchi, Y., Watada, J.: Possibilistic Forecasting Model and its Application to Analyze the Economy in Japan. In: Proceedings of Eighth International Conference on Knowledge-Based Intelligent Information & Engineering Systems, Part.III, pp. 151–158 (2004)
26. Yabuuchi, Y., Watada, J.: Fuzzy Robust Regression Model by Possibility Maximization. Journal of Advanced Computational Intelligence and Intelligent Informatics 15(4), 479–484 (2011)
27. Yao, J.T., Herbert, J.P.: Financial time-series analysis with rough sets. Applied Soft Computing 9(3), 1000–1007 (2009)
28. Yu, H.K.: Weighted fuzzy time series models for TAIEX forecasting. Phisica A: Statistical Mechanics and its Applications 349(3-4), 609–624 (2005)
29. Zadeh, L.A.: Fuzzy sets. Information and Control 8(3), 338–353 (1965)