Fuzzy AR Model using Possibility Regression Analysis

Kazuhiro Ozawa[1] and Takahide Niimura[2]

1 Faculty of Economics, Hosei University, 4342 Aihara, Machida, Tokyo 194-0298, Japan
2 Dept. of Electrical and Computer Eng., University of British Columbia, Vancouver, B.C., Canada, V6T 1Z4

Abstract

The authors present time series data analysis and estimation procedure under uncertain conditions. A new approach is introduced based on possibility theory and fuzzy autoregression, and it is applied to the analysis of time-series data.

1. Introduction

The authors introduce a new approach based on possibility theory [1] and fuzzy autoregression, and apply it to the analysis and estimation of time-series data. Traditional methods are based on statistical and probabilistic approaches but it may not be quite suitable to apply purely mathematical/stochastic models to the data generated by human activities. The proposed fuzzy model assumes that the time-series data reflects human decision-making activities rather than the stochastic process assumed in conventional models. This approach can represent the rich information contained in the original data set, including a certain level of possibility, while conventional models are aimed at converging to a single point at a time representing the most probable point of occurrence.

2. Fuzzy AR Model

2.1 Simple Fuzzy AR Model (FAR-S)

A fuzzy autoregression (FAR) model represents a possibility of occurrence of a certain set of data in a future when the present data are dependent to some degree on the past data. Let x be a time-series datum observed, then the fuzzy AR model is represented by the following equation:

$$X(t) \equiv A_0 + A_1 \cdot x(t-1) + A_2 \cdot x(t-2) + \cdots + A_m \cdot x(t-m), \tag{1}$$

where t is discrete time of an equal interval, m is the order of model description ($m \leq t$), A_i is fuzzy coefficient, and x is observed variable. The estimated value $X(t)$ is a fuzzy number and is capitalized to distinguish from x. There can be various ways to define the fuzzy parameters [2]. One of the simplest ways is to use symmetric triangular fuzzy sets represented by two crisp parameters such as

$$A_i = < a_i, c_i >, \quad (c_i > 0) \tag{2}$$

For the matter of convenience we call this fuzzy autoregression model as FAR-S model.

The crisp parameters a_i and c_i will be determined so that the model will include all or most of the observed data such as,

$$X \supseteq x(t). \tag{3}$$

To minimize the vagueness of the model, linear programming is applied to determine the parameters such that

$$\text{Minimize: } \sum_{t=m+1}^{n} \{c_0 + c_1 \cdot |x(t-1)| + \cdot \cdot \cdot + c_m \cdot |x(t-m)|\} \tag{4}$$

Subject to:

$$a_0 + a_1 \cdot x(m) + a_2 \cdot x(m-1) + \cdots + a_m \cdot x(1)$$
$$+ c_0 + c_1 \cdot |x(m)| + c_2 \cdot |x(m-1)| + \cdots + c_m \cdot |x(1)| \geq x(m+1)$$
$$\vdots$$
$$a_0 + a_1 \cdot x(n-1) + a_2 \cdot x(n-2) + \cdots + a_m \cdot x(n-m)$$
$$+ c_0 + c_1 \cdot |x(n-1)| + c_2 \cdot |x(n-2)| + \cdots + c_m \cdot |x(n-m)| \geq x(n)$$
$$a_0 + a_1 \cdot x(m) + a_2 \cdot x(m-1) + \cdots + a_m \cdot x(1)$$
$$-\{c_0 + c_1 \cdot |x(m)| + c_2 \cdot |x(m-1)| + \cdots + c_m \cdot |x(1)|\} \leq x(m+1)$$
$$\vdots$$
$$a_0 + a_1 \cdot x(n-1) + a_2 \cdot x(n-2) + \cdots + a_m \cdot x(n-m)$$

$$-\{c_0+c_1\cdot|x(n-1)|+c_2\cdot|x(n-2)|+\cdots+c_m\cdot|x(n-m)|\}\le x(n)$$

and

$$c_0, c_1, c_2, \ldots, c_m \ge 0$$

Where n is the total number of observed data.

2.2 Fuzzy AR Model based on Crisp AR (FAR-A)

The fuzzy AR model described in 2.1 is obviously simple and powerful as it will be shown later, but one of the problem is that the center value a_i of the fuzzy parameters does not necessarily represent the most possible data in the time series. On the other hand, conventional autoregression methods are quite established and some popular programs are publicly accessible by World Wide Web (e.g., Web Decomp [3]). Therefore, it is reasonable to apply time-series data obtained by a conventional AR as part of the fuzzy time-series model. We refer such conventional AR-based fuzzy model as FAR-A in this paper. In the FAR-A model, we use the results of conventional AR for a median value a_i of triangular fuzzy number. By determining the center value, the triangular fuzzy set, representing the fuzzy parameters, need not to be symmetrical, and it has an advantage of reflecting the possibility distribution of the time-series data better than FAR-S. Fig. 1 shows the unsymmetrical fuzzy set used in the FAR-A model.

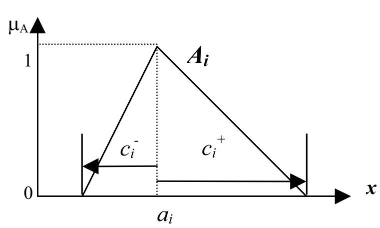


Fig. 1 Unsymmetrical fuzzy set

Having the two different parameters c_i^+, and c_i^- (c_i^+ for upper deviation and c_i^- for lower deviation), the FAR-A parameters are determined by solving the extended linear programming problem such as

Minimize:

$$\sum_{t=m+1}^{n}\{c_0^+ + c_1^+\cdot|x(t-1)|+\cdots+c_m^+\cdot|x(t-m)| + c_0^- + c_1^-\cdot|x(t-1)|+\cdots+c_m^-\cdot|x(t-m)|\} \tag{5}$$

Subject to:

$$a_0+a_1\cdot x(m)+a_2\cdot x(m-1)+\cdots+a_m\cdot x(1)$$
$$+c_0^+ + c_1^+\cdot|x(m)|+c_2^+\cdot|x(m-1)|+\cdots+c_m^+\cdot|x(1)|\ge x(m+1)$$

$$a_0+a_1\cdot x(m+1)+a_2\cdot x(m)+\cdots+a_m\cdot x(2)$$
$$+c_0^+ + c_1^+\cdot|x(m+1)|+c_2^+\cdot|x(m)|+\cdots+c_m^+\cdot|x(2)|\ge x(m+2)$$

$$\vdots$$

$$a_0+a_1\cdot x(n-1)+a_2\cdot x(n-2)+\cdots+a_m\cdot x(n-m)$$
$$+c_0^+ + c_1^+\cdot|x(n-1)|+c_2^+\cdot|x(n-2)|+\cdots+c_m^+\cdot|x(n-m)|\ge x(n)$$

$$a_0+a_1\cdot x(m)+a_2\cdot x(m-1)+\cdots+a_m\cdot x(1)$$
$$-\{c_0^- + c_1^-\cdot|x(m)|+c_2^-\cdot|x(m-1)|+\cdots+c_m^-\cdot|x(1)|\}\le x(m+1)$$

$$a_0+a_1\cdot x(m+1)+a_2\cdot x(m)+\cdots+a_m\cdot x(2)$$
$$-\{c_0^- + c_1^-\cdot|x(m+1)|+c_2^-\cdot|x(m)|+\cdots+c_m^-\cdot|x(2)|\}\le x(m+2)$$

$$\vdots$$

$$a_0+a_1\cdot x(n-1)+a_2\cdot x(n-2)+\cdots+a_m\cdot x(n-m)$$
$$-\{c_0^- + c_1^-\cdot|x(n-1)|+c_2^-\cdot|x(n-2)|+\cdots+c_m^-\cdot|x(n-m)|\}\le x(n)$$

and

$$c_0^+, c_0^-, c_1^+, c_1^-, \ldots, c_m^+, c_m^- \ge 0$$

where n is the total number of observed data.

3. Method of Constructing a Fuzzy AR Model and Its Parameter Identification

3.1 Electricity Consumption Model

The most direct method of constructing a fuzzy AR model would be to apply intuition and experience of the observer on the given set of data. However, to obtain an effective model, the modeler needs to have an experience and data-specific knowledge, and the model structure could not be implied quantitatively. Therefore, in this paper we partly follow the conventional approach to determine the model structure of the fuzzy AR model. In a typical conventional approach, the structure of autoregression model for time-series data can be determined by removing trends and other seasonal fluctuations from the original data to extract steady-state data, and by observing the autocorrelation function values. Parameters are optimized subsequently and the model is evaluated in the final stage. The model is successively modified, for example, by changing the orders of differences [4] to improve accuracy.

Fig. 2 shows an example of raw data of electricity consumption. In this paper, we have used the data recorded at California Power Exchange for the performance evaluation of model and forecast. This particular set of data is the maximum consumption of electrical energy (MWh) recorded everyday for a three month period (January - March 1999) and available from the California Power Exchange web site [5].

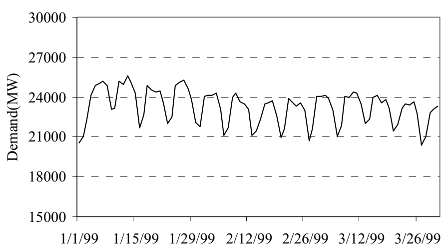


Fig. 2 Recorded electricity consumption in California

Fig. 3 shows the autocorrelation function values of the same data. The data in the first two months are analyzed. It is obvious that the weekly cyclic pattern should be reflected in the model structure. Most of the industrial plants, volume consumers of electricity, shutdown on weekends, start-up on Monday, and their activity peaks in mid-week. The autocorrelation function values justify such heuristic knowledge of electricity consumption. Thus, we make use of the series of t-1 and t-7 time lags such as

$$X(t) \equiv A_0 + A_1 \cdot x(t-1) + A_2 \cdot x(t-7) \tag{6}$$

where A_0, A_1, and A_2 are fuzzy numbers and time series $x(t)$ are real numbers. Therefore, the resultant estimated value $X(t)$ will also be a fuzzy number.

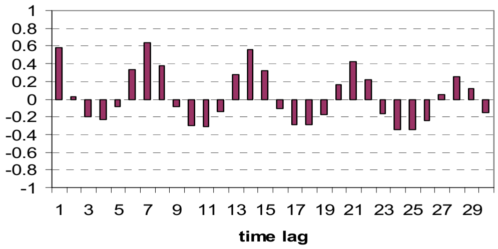


Fig. 3 Autocorrelation function values of the data

3.2 Parameter Identification

Parameters of fuzzy sets that construct the target fuzzy AR model can be determined by linear programming. A similar method has been proposed as possibility regression [6] and many applications have been developed. In an analogous fashion, in this paper, we apply LP to determine the parameters of the fuzzy autoregression model. The target of the linear programming problem is to minimize the region of the fuzzy time-series while covering the given time-series as much as possible. The variables to be determined are the crisp values that define the fuzzy numbers. For the California PX case, the LP problem using the data for the first two months (52 points) of those shown in Fig. 2.

The results of the optimization are shown in Table 1. Underlined parameters in FAR-A column have been obtained by conventional AR.

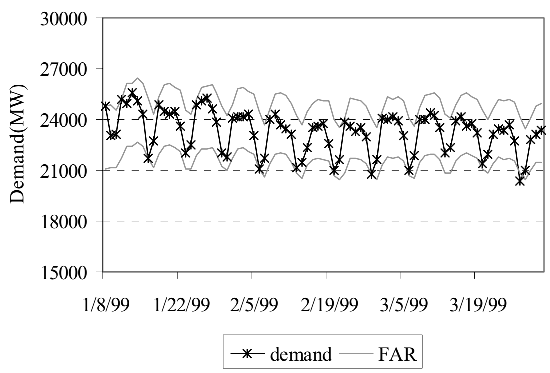
Table 1. Parameters of fuzzy numbers

	FAR-S	FAR-A
a_0	9182	<u>2078</u>
a_1	0.295	<u>0.312</u>
a_2	0.309	<u>0.597</u>
c_0^+	0	0
c_0^-		0
c_1^+	0.074	0.104
c_1^-		0.071
c_2^+	0	0
c_2^-		0

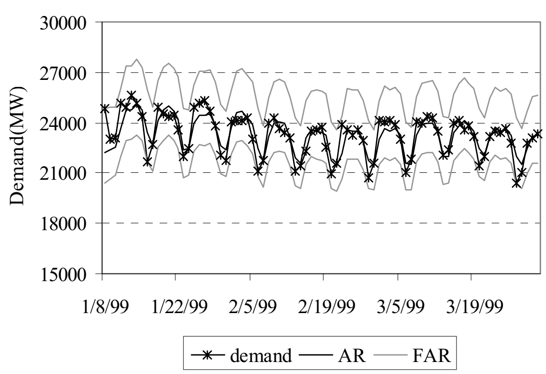
Note that c_0 and c_2 parameters for A_0 and A_2 are both zero. This fact indicates that the fuzziness is only represented by A_1 for this particular set of data.

3.3 Forecasting by Fuzzy AR Model

In this section, we apply the fuzzy AR models obtained above to a forecast of electricity consumption data in a near future. Fig. 4 shows the results of forecast. Model parameters are determined as above using the first 52 data (i.e., from $t = 8$ to $t = 59$; Note that the data for $t = 1, \ldots, 7$ period are not represented by this model). The estimated data are in the period from $t = 60$ to $t = 90$. Points plotted on thick real lines are the recorded original data, including the forecasted period, and faint lines show the encompassed possibility region of data. As evident in the figures, most of the data in the forecasted period are covered by possible regions in both FAR-S and FAR-A models.



(a) Time series data of FAR-S



(b) Time series data of FAR-A

Fig. 4 Time series data of electricity consumption represented by fuzzy autoregression models

The time series by the conventional (crisp) AR model is plotted as the center value of the FAR-A model. In the conventional approach, the difference between the estimation and the real data are interpreted as errors. There are various ways to make the crisp AR model more accurate but we are not concerned as far as the proposed FAR models (mostly) cover the possible regions of the target data. The criteria that decide whether or not the FAR models are acceptable or useful depend on what the decision-maker expects on the information represented by the time-series data. For example, in FAR-A model, the possibility distribution that are wider on the high demand side ($c_i^+ > c_i^-$) may indicate a larger volatility to push up the demand than to pull down the demand.

4. Conclusions

In this paper, the authors have introduced an alternative approach for time-series data analysis based on possibility theory and fuzzy autoregression, and applied it to the analysis of time-series data of electric power consumption. Two models have been developed, and both models have been applied to the forecasting of real-life data. The proposed fuzzy autoregression models preserve the rich information of the original data including uncertainty. Such information is useful when a degree of uncertainty is important for a decision making. Modeling and forecasting the highly volatile electricity prices under the deregulated market will be an interesting and challenging issue for the fuzzy autoregression.

5. References

[1] L. A. Zadeh, "Fuzzy Sets as a Basis for a Theory of Possibility," *Fuzzy Sets and Systems*, pp.3-28, Vol.1, No.1, 1978.

[2] D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York 1980.

[3] http://ssnt.ism.ac.jp/inets/inets2.html

[4] G.E.P. Box and G.M. Jenkins, *Time Series Analysis, Forecasting and Control*, Holden-Day, Revised Edition, 1982.

[5] http://www.calpx.com (closed)

[6] H. Tanaka, S. Uejima, and K. Asai, "Linear Regression Analysis with Fuzzy Model," *IEEE Trans. SMC*, Vol.12, No.6, 1982, pp. 903-907.