ARIMA based Interval Type-2 Fuzzy Model
for Forecasting

*Saima H., J. Jaafar, S. Belhaouari
Department of Computer and Information Sciences
Universiti Teknologi PETRONAS,
Tronoh Perak, Malaysia

T.A. Jillani
Department of Computer Science
University of Karachi,
Karachi, Pakistan

ABSTRACT

To solve the chaotic and uncertain problems, researchers are focusing on the extensions of classical fuzzy model. At present Interval Type-2 Fuzzy logic Systems (IT2-FLS) are extensively used after the thriving exploitation of Type-2 FLS. Fuzzy time series models have been used for forecasting stock and FOREX indexes, enrollments, temperature, disease diagnosing and weather. In this paper a hybrid fuzzy time series model is proposed that will develop an Interval type 2 fuzzy model based on ARIMA. The proposed model will use ARIMA to select appropriate coefficients from the observed dataset. IT2-FLS is utilized here for handling the uncertainty in the time series data so that it may yield a more accurate forecasting result.

General Terms

Autoregressive Integrated Moving Average (ARIMA) models, Time Series Forecasting**.**

Keywords

Type-1 Fuzzy sets (T1-FS), Type-2 Fuzzy sets (T2-FS), Fuzzy Logic System (FLS), Interval Type-II Fuzzy Logic Systems (IT2-FLS).

1. INTRODUCTION

Dealing with only numerical data is one of the limitations of the linear models. A linear model may not be effectively constructed if the relationship between the variable is not linear. Also if multivariables are used in such a model for forecasting then calculating the parameters of the model is a difficult issue. Furthermore, the strong relationship among these variables may result in large errors. Being nonlinear, the fuzzy logic systems (FLS) have been combined with time series models to get rid of the problems occur with linear data and can also handle linguistic chaos and ambiguity. Autoregressive (AR), Multiple Regressions, Exponential Smoothing and Autoregressive Integrated Moving Average (ARIMA) are a few examples of time series models that have been combined with FLS. These Fuzzy time series models are used by many researchers to solve linear and nonlinear problems. Since the introduction of FLS, it has successfully been applied to a lot of time series problems including stock and FOREX indexes [3] [6] [7] [11], enrollments [2] [4] [5], temperature [8], weather forecasting [12] [10], and disease diagnosing [9].

Researchers are finding their way to solve the chaotic and uncertain problems using the extensions of classical fuzzy model. At present Interval Type-2 Fuzzy logic Systems (IT2-

FLS) are extensively used after the thriving exploitation of Type-2 FLS. The non-interval secondary membership function of a general Type 2 Fuzzy sets (T2-FS) made it computationally more complex. The reason that IT2 fuzzy sets are enjoying the center stage these days; is the non fuzziness of its secondary function [1] [13] [14].

With the intention to achieve better results of forecasting, numerous hybrid models have been developed [2] [3] [5] [6] [10] [12]. It has been deduced from their result that the hybrid models outperform the individual models. Keeping this concept, in this paper, we have proposed an efficient fuzzy time series model based on ARIMA and IT2-FLS to improve forecasting accuracy and handle forecasting problems.

The rest of the paper is organized as follow. In section 2, a brief literature review on time series analysis techniques and forecasting is presented. In section 3 and 4, short overview of ARIMA models and IT2-FS and IT2-FLS are presented respectively. A hybrid model is proposed in section 5 and future work is described in section 6.

2. LITERATURE REVIEW

Forecasting is a systematic way of anticipating future events and situations by estimating the past value of a variable(s). The management and scheduling done by forecasting helps to cope with the uncertainty of future. Fuzzy time series is used to handle forecasting problems where numerical as well as linguistic information pertaining to that problem is available. Thus, fuzzy time series forecasting not only covers the statistical flavor but also the linguistic chaos analysis as well.

The idea of fuzzy time series based on the historical enrollments of the University of Alabama is practiced in [20]. The time-invariant fuzzy time series model and the time-variant fuzzy time series model based on the fuzzy set theory for forecasting the enrollments of the University of Alabama is presented in [4], [20]. A heuristic model of fuzzy time series model has been developed to improve forecasting [21] where the improvement in forecasting result is achieved by introducing heuristic rules. Weighted models are proposed in [6] to resolve the recurrence and weighting issues in fuzzy time series forecasting. These models demonstrate similarity to the weight functions in local regression models; though, both are dissimilar.

The local regression models concentrated on fitting using a small portion of the data, whilst the weighted fuzzy time series models established fuzzy relationships using the promising data from the entire database. ARIMA is integrated with fuzzy regression model in [18] for forecasting the foreign exchange

market. It is deduced from the results that it takes less observation to estimate a model than ARIMA. The forecasted results were made good by narrowing the fuzzy interval (Upper and lower bound). When fuzzy intervals were wide deleting its upper or lower bound the interval were made narrow. This gave better performance. However deleting the bounds may cause incomplete or missing interval/data. Two new multivariate fuzzy time series forecasting methods are presented in [22]. These methods assume m-factors with one significant main factor. Stochastic fuzzy dependence of order k is presumed to define general methods of multivariate fuzzy time series forecasting and control. A refined fuzzy time series model with improved defuzzification model is presented in [23] for stock exchange forecasting. Results with better accuracy were obtained by proposing a heuristic approach of fuzzy metric. T2-FLS is applied to forecast Mackay-Glass time series [19]. Using T1-FLS, a T2-FLS is formed by incorporating noise information. The FLSs were designed on a single realization for practice. However, for different realizations of same data set, different FLSs parameters must be choose to obtain improved forecasting [19].

3. A BRIEF OVERVIEW OF AUTO REGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA)

Time series is a unique sequence of data taken at a specific interval of time. This data may be stationary or non-stationary. The data must first be made stationary by numerous methods, as forecasting cannot be done with data comprising non-stationary behavior. ARIMA is a non-stationary model that can be reduced to a stationary time series by differencing. The model has three components i.e. Autoregressive, Integrated and Moving Average.
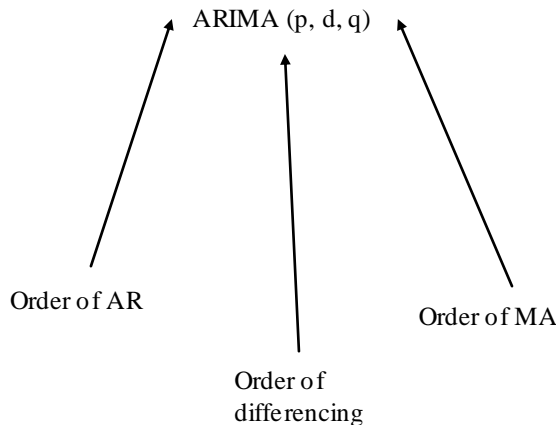
ARIMA (p, d, q)

Order of AR

Order of differencing

Order of MA

Fig 1: A General View of ARIMA models.

A time-series Y_t is generated by an ARIMA (p, d, q) if

$$\varphi(L)(1 - L)^d Y_t = \theta_q(L)\varepsilon_t \qquad (1)$$

where $\varphi(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p$ and $\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_q L^q$

are polynomials in L of degree p and q, L is the backward shift operator, p, d, q are integers, Y_t denotes the observed value of time-series data, $t = 1, 2, \ldots, k$, and time-series data are the observations.

Being one of the best forecasting models, it can be best illustrate by Box-Jenkins [17] four steps.

1. Identification finds the appropriate value of differencing d. Practically, one or two levels of differencing are sufficient to make the time series stationary.
2. Estimation of the unknown model parameters.
3. Diagnostic statistics help to judge the adequacy of the model.
4. Forecasting from the selection mode.

4. AN OVERVIEW OF INTERVAL TYPE 2 FUZZY LOGIC SETS AND SYSTEMS (IT2-FS AND IT2-FLS)

Type-1 fuzzy sets cannot handle the uncertainty exist in data since its membership grades are crisp value [13] [14]. T2-FS sets are introduced as an extension of T1-FS with a third dimension. The additional dimension in T2-FS helps in handling more uncertainties than T1-FS [14]. Instead of using T2-FS with complex computation, IT2-FS can be preferred with much simplified computation [13]. This paper is not supposed to give a background of T2 FS, that can be found in [1],[7],[15] and [16]. An IT2 FS can be defined as

Definition: A T2-FS denoted by $\tilde{A}$ will be an IT2-FS, when all membership functions $\mu_{\tilde{A}}(x, u) = 1$, where $x \in X$ and $u \in J\_x \subseteq [0,1]$, i.e.

$$\tilde{A} = \{((x, u), 1) \mid x \in X, u \in Jx, Jx \subseteq [0,1]\},$$

$\check{A}$ can also be expressed as

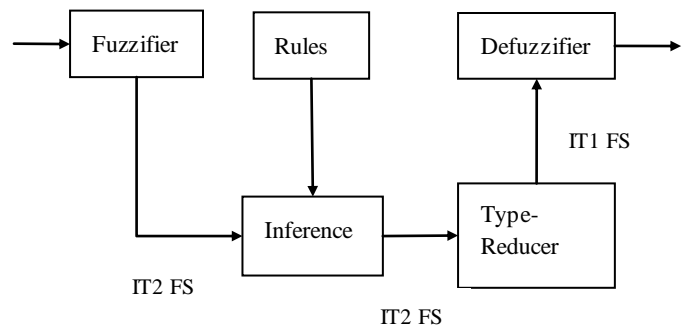$$\check{A} = \int_{x \in X} \int_{u \in J_x} 1/(x, u), J_x \subseteq [0,1],$$

Fig 2: A General view of Interval Type 2 Fuzzy Logic models.

The rule base and inference engine are activated by these sets to produce IT2 FS. A type-reducer in the IT2-FLS reduces these fuzzy sets to IT1 FS, which are further defuzzified to get a crisp value. The defuzzified value can be obtained by taking the average of the two end-points of the interval [13].
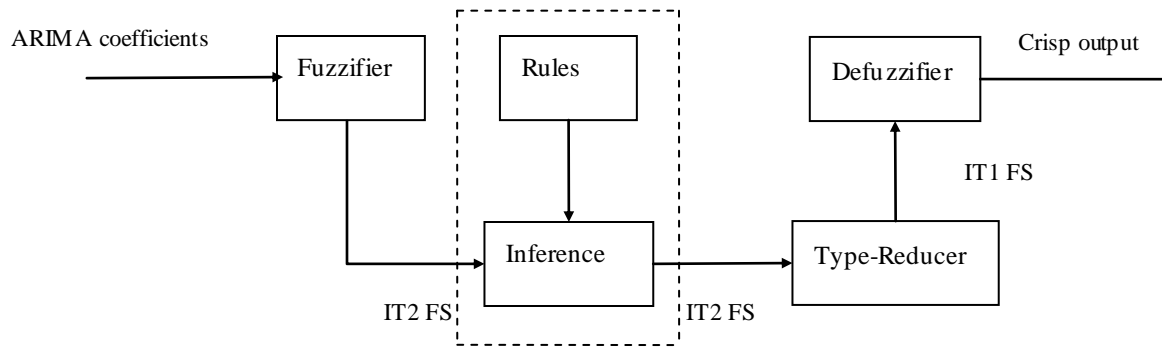
Fig 3: Proposed Interval Type 2 Fuzzy Model based on ARIMA

5. PROPOSED MODEL BASED ON IT2-FLS AND ARIMA MODELS

To tackle various forecasting problems and improve forecasting accuracy an efficient fuzzy time series model based on IT2-FLS and ARIMA is proposed in this section. The novel IT2-FARIMA model will estimate the parameters using ARIMA (see Fig 3). These estimated parameters will be fuzzified by IT2-FLS and will be then used for forecasting.

5.1 The proposed model consists of following steps

5.1.1 Fit variable/coefficient through ARIMA model

The observed time series data must first be transformed into a reduced data set since dealing with large data requires high processing power and large amount of memory. To select an optimum set of parameters a predictive model is used. ARIMA in the proposed model will ensure the appropriate parameters for the available dataset. With the help of various statistical measurements an appropriate model and its coefficient will be estimated.

5.1.2 Fuzzify the fitted coefficients

The estimated parameters will be fuzzified in the fuzzifier block of IT2 FLSs.

5.1.3 Establish the Fuzzy logic relationship

The model will be evaluated using Takagi-Sugeno [25] fuzzy rule based model, which is better than Mamdani's model [24] in the sense that it reduces the number of required rules.

5.1.4 Reduce the fuzzified set

Type-reduction technique extracts an interval for the output value from the uncertain region. Among various type-reduction methods the output will be reduced by a centroid type-reduction technique [27]. The centroid of IT2 FS was developed in [27] so as to measure and compute the uncertainty of that FS [26].

5.1.5 Defuzzify to get the forecasted result

At the end the type-reduced output will be defuzzify to get a crisp output in the defuzzifier block.

5.2 Problems that can be handled with the proposed model

5.2.1 Reduce dataset

One of the limitations of ARIMA model is that it cannot create accurate result if the observed data is too small. But sometimes in short term forecasting, the historical data may have less values [18]. In Fig.4, the graph is built with 22 observations and it can be seen that even after first differencing, the result is not satisfactory, (see Fig.5). At first differe the graph become stationary at mean but still without constant variability (Non-homoscedastic). In such a situation the model cannot estimate the accurate coefficients, resulted in parameter uncertainty.
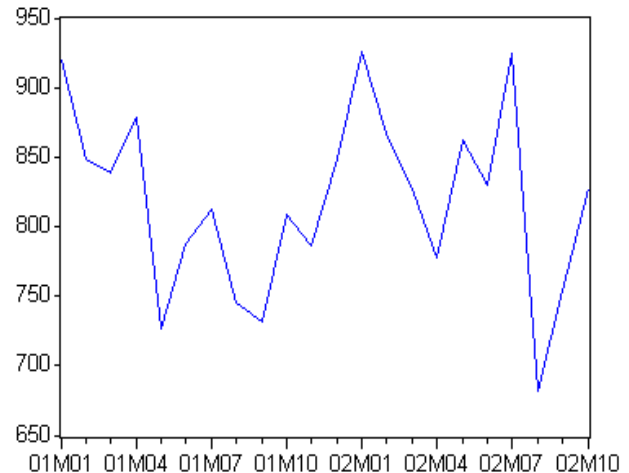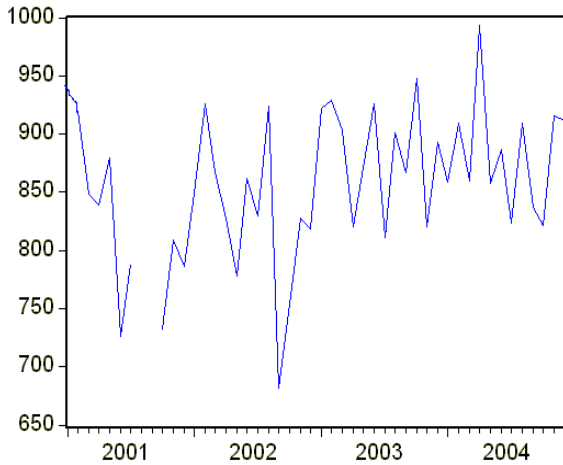

Fig 4: Shows limitation od ARIMA with less data.

5.2.2 Missing data

The second issue is of handling the missing data (see Fig. 6). In practical, missing data occurs frequently. The major reason of the missing data is the non-functionality of the acquisition.
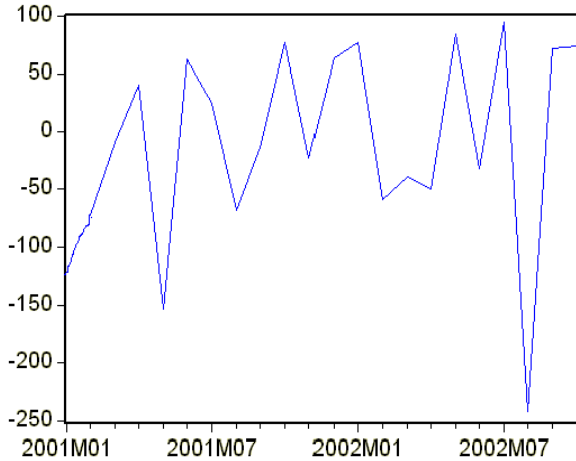
Fig 5: Shows limitation of ARIMA with less data.

Fig. 1.   Shows limitation of ARIMA when missing values are present in the dataset.

5.2.3 Model's Integration

Models are integrated for the purpose of obtaining improved results than the individual. Since the real time series data contain linear and non linear patterns, the best approach would be the combination of linear and non linear model. The resulted single model would be able to deal with both types of relationships.

In [10] the Artificial Neural network is combined with Fuzzy Inference System to produce a hybrid ANFIS model for forecasting. Both the models are non-linear. The same data were used by $ARIMA\ (2.1.1)$ model to evaluate performance criteria. It is evidence from the result that ARIMA performed well at testing data where the performance of the ANFIS model was good in training data.

At feature extraction linear models perform well [12], where as in ANFIS model it is done by non linear model. That is the reason that testing data done by ARIMA gave good results as compared to ANFIS. Another weakness of ANFIS is its complex computations.

The performance comparisons of ANFIS and ARIMA models is done on Mean Absolute Error (MAE), Root Mean Squared Error

(RMSE), and Coefficient of Determination (R^2).are shown below

Table 1. Comparison of ANFIS and ARIMA

	MAE		RMSE	
	ANFIS	ARIMA	ANFIS	ARIMA
Training data	1.2552	1.523	1.6312	2.075
Test data	1.3212	1.178	1.7176	1.0709

5.3 A Research Approach for Handling These Problems

The above all three issues can be handled by the proposed hybrid model of Interval Type 2 Fuzzy model based on ARIMA (IT2FARIMA). Issue 1 and 2 are the limitations of ARIMA which would be effectively handled by IT2-FS. The last issue of improving the accuracy would be solved by combining the two dissimilar models. The propose model will combine the ARIMA model with IT2-FLS, where the former is a linear model and later is a nonlinear one. The sequence for hybridizing the model will follow the conventional hybrid model, which starts with a linear model, followed by a nonlinear model to model the residual. The model will be evaluated by Mean Squared Error (MSE), RMSE and MAE.

Measurements formulas discuss in this paper are,

$$MSE = \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}$$

$$MAE = \frac{\sum_{i=1}^{N} |y_i - \hat{y}_i|}{N}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}}$$

$$R^2 = 1 - \frac{\sum_{t=1}^{N}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{N}(y_t - \overline{y}_t)^2} = 1 - \frac{E}{\sum_{t=1}^{N}(y_t - \overline{y}_t)^2}$$

where y_i, $\hat{y}_i$, $\overline{y}_i$ are actual, forecasted and mean values respectively, and N is the forecast number.

6. FUTURE WORK

Among different optimization techniques, evolutionary and particle swarm intelligence techniques, can be used to select best model parameters. Furthermore the proposed model is considering the observations of a single time series, the model can be extended to accept more than one factors affecting forecasting.

7. REFERENCES

[1] R. John, S. Coupland. "Type-2 fuzzy logic : A historical view," *IEEE Computational Intelligence Magazine* , February 2007, vol. 2, pp. 57-62.

[2] S.M. Chen, "Forecasting enrollments based on fuzzy time series," *Fuzzy Sets and Systems* 1996, vol. 81, pp 311-319.

[3] F.M. Tseng, G.-H. Tzeng, H.-C. Yu, BJ.C. Yuan," Fuzzy ARIMA model for forecasting the foreign exchange market," Fuzzy *Sets and System*s, 2001, vol. 118, pp 9-19.

[4] Q. Song, B.S. Chissom, "Forecasting enrollments with fuzzy time series-part 1," *Fuzzy Sets and Systems*, 1993, vol. 54, pp. l-9.

[5] Q. Song, B.S. Chissom, "Forecasting enrollments with fuzzy time series-part 2," Fuzzy *Sets and Systems* 1994, vol. 62, pp. 1-8.

[6] K. Yu, "Weighted fuzzy time series models for TAlEX forecasting," *Physical A,* 2005, vol. 349, pp. 609-624.

[7] K. Huamg, Hui-Kuang Yu, "A Type 2 fuzzy time series model for stock index forecasting," *Physical A* 2005, vol. 353, pp. 445-462.

[8] S.-M. Chen, J.R. Hwang, "Temperature prediction using fuzzy time series," *IEEE Trans. Syst. Man, Cybem.* 2000, Part B vol. 30 (2), pp. 263-275.

[9] D. Delen, G. Waller, A. Kadam, "Predicting Breast Cancer Survivability: A Comparison of Three Data Mining Methods," Artificial *Intelligence in Medicine*, 2005, vol. 34, pp. 113-127.

[10] M. Tektaş, "Weather forecasting using ANFIS and ARIMA models. A case study for Istanbul," *Environmental Research, Engineering and Management*, 2010, vol. 51, pp. 5-10.

[11] N. Merh, V.P. Saxena, K. R. Pardasani, "A comparison between Hybrid Approaches of ANN and ARIMA for Indian stock trend forecasting," *Business Intelligence Journal* 2010, vol. 3, No 2, pp. 23-43.

[12] E. Cadenas, W. Rivera, "Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA_ANN model," *Renewable Energy* 2010, vol. 35, pp 2732-2738.

[13] D. Wu, "A brief Tutorial on Interval type-2 fuzzy sets and systems," in *Fuzzy Sets and Systems*.

[14] J.M. Mendal, R.I. John and F. Liu, "Interval type 2 fuzzy logic systems made simple," *IEEE Trans. On Fuzzy Systems*, Dec. 2006, vol. 14, pp. 808-821.

[15] N.N. Karnik and J.M. Mendel, "An Introduction to Type-2 Fuzzy Logic Systems*," Tech. Rep.*, University of Southern California, 1998.

[16] R.I. John, "Type 2 Fuzzy Sets: An Appraisal of Theory and Applications," International Journal of Uncertainty, Fuzziness and Knowledge Based Systems 1998, vol. 6, no. 6, pp. 563–576.

[17] G.P. Box, G.M. Jenkins, Time Series Analysis: Forecasting and Control, *Holden-day Inc.*, San Francisco, CA, 1976.

[18] F.M. Tseng, G.H. Tzeng, H.C. Yu, B.J.C. Yuan, "Fuzzy ARIMA model for forecasting the foreign exchange market," *Fuzzy Sets and Systems* 2000, vol. 118, pp. 9-19.

[19] N.N. Karnik and J.M Mendel, "Applications of type-2 fuzzy logic systems to forecasting of time-series," *Information Sciences*, 1999, vol. 120, pp. 89-111.

[20] Q. Song and B.S. Chissom, "New models for forecasting enrollments: fuzzy time series and neural network approaches, " *ERIC*, 1993 p. 27, http://www.eric.ed.gov

[21] K. Huarng (2002), "Heuristic models of fuzzy time series for forecasting," *Fuzzy Sets and Systems*, vol. 123, no. 3, pp.369-386.

[22] T.A. Jilani and S. Burney, "Multivariate stochastic fuzzy forecasting models," *Expert Systems with Applications*, vol.35, 2008, pp. 691–700.

[23] T.A. Jilani and S. Burney, "A refined fuzzy time series model for stock market forecasting"*, Physica-A-Statistical mechanics and its applications*, 387, 2008, pp. 2857-2862.

[24] Mamdani, E. H. and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *Int. J. Man-machine Studies*, Vol. 7, 1–13, 1975.

[25] Takagi, T. and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Trans. Systems, Man, and Cybernetics*, Vol. 15, 116–132, 1985.

[26] H. Wu, J.M. Mendel, "Uncertainty bounds and their use in the design of interval type-2 fuzzy logic systems," *IEEE Transactions on Fuzzy Systems* October 2002, pp. 622-639.

[27] N.N. Karnik, J.M. Mendel, "Centroid of a type-2 fuzzy set," *Information Sciences* 2001, vol.132, pp. 195-220.