

Autoregressive Modeling with Error Percentage Spread based Triangular Fuzzy Number

Hamijah Mohd Rahman, Nureize Arbaiy, Chuah Chai Wen, Riswan Efendi

Abstract: Data collected by various methods are often prone to uncertainty of measurement which may affect the information conveyed by the quantitative result. This causes the developed predicted model to be less accurate because of the uncertainty contained in the input data used. Hence, preparing the data by means of handling inherent uncertainties is necessary to avoid the developed prediction model to be less accurate. In this paper, the standard autoregressive model is extended to the case where inherent uncertainty exist in the time series data input is handled by triangular fuzzy number. A systematic strategy to construct a symmetry triangular fuzzy number based on percentage error method to build the autoregressive model is presented. Three different spreads of 1%, 3% and 5% are evaluated under percentage error method. This method is applied to forecast the exchange rate of Association of South East Asian Nation (ASEAN) based on time series data. The enhancement made in data preparation of building fuzzy triangles in this study affirms that the proposed method can produce a better accuracy in predicting as compared to the standard auto regressive model. Importantly, the difficulties to build a triangular fuzzy number to treat the fuzziness which is contained in data is addressed. From the result, we could rank the best percentage error spread which gives higher accuracy among 1%, 3% and 5% model.

Index Terms: L Autoregressive, Error Percentage, Triangular Fuzzy Number, Uncertainty.

I. INTRODUCTION

Assembling an appropriate time series data input is a crucial task and challenging. In time series observation, data input should be accomplished first in data preparation to avoid misleading result during prediction analysis. Additionally, the historical data used for prediction are exposed to uncertainty which can affect the performance of prediction. Data which are collected from various measurements carries some degree of uncertainty [1].

To date, in the context of economic system such as exchange rate, the historical data has variation such as high low value instead of precise series [2].

Revised Manuscript Received on June 22, 2019 .

HamijahMohd Rahman, Faculty of Computer Science and Information Technology, UniversitiTun Hussein Onn, Parit Raja 86400 BatuPahat, Johor

Nureize Arbaiy, Faculty of Computer Science and Information Technology, UniversitiTun Hussein Onn, Parit Raja 86400 BatuPahat, Johor

Chuah Chai Wen, Faculty of Computer Science and Information Technology, UniversitiTun Hussein Onn, Parit Raja 86400 BatuPahat, Johor

Riswan Efendi, Faculty of Science and Technology, State Islamic University of Sultan SyarifKasim Riau, Panam, 28293, Indonesia.

In previous literatures, the issue of uncertainty has been studied by several authors by introducing fuzzy theory [3-6]. Most of the researches deal with uncertainties in the auto regression model. Few discuss the fuzzy data treatment during data preparation though it is important to treat the data before prediction can be modeled. Since the existence of uncertainty in data have substantial implications for interpreting the prediction model [7], a systematic data preparation is presented in this paper for dealing the uncertainty contained in the data. Thus, this paper extends the existing auto regression model to use treated data known as triangular fuzzy number which treats the uncertainty during the data preparation process.

In this method, the percentage error is utilized to determine the spread of a triangular fuzzy number. Commonly triangular fuzzy number is developed based on expert's definition [8]. It makes the description of triangular fuzzy number troublesome when the expert's definition is unavailable or hard to be described. Hence, a percentage error method is introduced in this study to ease the development of triangular fuzzy number to treat the fuzziness in data input. The symmetry fuzzy number is then used into auto regression model to build a prediction method. The emphasis made in the provision of fuzzy data and model prediction is very important to improve the predictor value and to achieve prediction accuracy. Time series exchange rate dataset of Association of South East Asian Nation (ASEAN) countries is used for experimental. A comparison between Triangular Fuzzy Number Auto regression model and conventional model are given to illustrate the performance of the proposed model in predicting the exchange rate.

The structure of this paper is divided into five sections. Section II explains concept of time series auto regression model. Section III is spent to describe the proposed method in solving the uncertainty problem. Section IV illustrate the proposed model through application in exchange rate and analyses the experimental result. Finally, conclusion is given in Section V.

II. TIME SERIES AUTO REGRESSION MODEL

Auto regression or AR (p) is used in predicting the next period which substitute the previous data to the model [9].

This model is used in time series analysis which can investigate the data stationary based on the statistical data. The AR (p) can be written as following form:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon_t \quad (1)$$



where y_t denotes the observed value of time series data, β_0 is a constant, and β_1 is the coefficient respectively, and ϵ_t is error model at time- t , $t=1,2,\dots,k$, and observation is in time-series data.

Let us assume y_t depends on the previous data with time $t, y_{t-1}, y_{t-2}, \dots, y_{t-k}$, then the AR model used to forecast the next period with the substitution of previous data are as follow:

$$\tilde{Y}_{t+1} = \beta_0 + \beta_1 y_t + \epsilon_t \quad (2)$$

We can forecast a few period by repeating the same calculation

$$\tilde{Y}_{t+2} = \beta_0 + \beta_1 y_{t-1} + \epsilon_t$$

$$\tilde{Y}_{t+3} = \beta_0 + \beta_1 y_{t-2} + \epsilon_t$$

$$\tilde{Y}_{t+h} = \beta_0 + \beta_1 y_{t-i} + \epsilon_t \quad (3)$$

First order auto regressive model AR (1) is utilized in this study. The AR (1) model formulation includes three steps [10].

Step 1: AR (1) structure identification. Use Partial Autocorrelation (PACF) to develop the rough function. AR(1) model would likely be feasible for the dataset if there is significant spike at a lag of 1 and much lower spikes for the subsequent lags.

Step 2: Estimate constant, coefficient and P value. The P value for constant and coefficient should significantly $P < 0.05$.

Step 3: Prediction from the selection model.

Conventional AR method has limited capabilities to deal with uncertainty in data while modeling the time series data. It makes the conventional AR disregard inherent uncertainty in data [11]. It is common in practice where the data are changing rapidly and has uncertainty. Several researchers applied fuzzy theory in auto regression model to capture the uncertainty. For example, fuzzy auto regressive model was applied to identify the daily river flows by combining the auto regressive model with fuzzy inference system [4]. The fuzzy auto regressive model and conventional AR model was applied to asthmatic data by considering some factors of weather, temperature and atmospheric [5]. A combination of fuzzy logic and auto regressive model was proposed in air quality prediction based on the result of chemical toxicity test [6]. The fuzzy inference system was used to create a new air quality index while auto regressive model used to predict future air quality condition. The concept of fuzzy regression model was combined with the ARIMA model to formulate the fuzzy ARIMA model and applied in the prediction the foreign exchange market [10].

From these previous literatures, we observed that the fuzzy theory was applied in the model to reduce uncertainty and imprecision. However, the procedure on how to treat uncertainty during data preparation is not explained in detail. Hence, based on the formulation of AR (1) described in this section, we proposed the systematic strategy to build triangular fuzzy number to resolve the uncertainty exist in observation data.

III. AUTOREGRESSIVE MODELLING WITH ERROR PERCENTAGE

This section explains the systematic strategy to build autoregressive model which uses triangular fuzzy number to

treat the inherent uncertainties. Before proceed, we summarize the theoretical background of triangular fuzzy number.

According to [12], the theory of Triangular Fuzzy Number (TFN) can be defined as $\tilde{A} = (a, b, c)$ if its membership function is

$$\mu_A(x) = \left\{ \begin{array}{ll} 0 & x \leq a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ \frac{c-x}{c-b} & b \leq x \leq c \\ 0 & x > c \end{array} \right\} \quad (4)$$

While few researches explain in detail the procedure to construct the triangles, this paper proposes a strategy to construct the triangular fuzzy number (TFN) based on percentage error; that is 1%, 3% and 5% spreads. Having the procedure to construct the triangular is important to avoid misconstruction in fuzzy data treatment. The TFN is constructed based on percentage error of 1%, 3% and 5% spreads which is stimulated from confident interval concept. The 5% error and below is concentrated in this study [13, 14].

The strategy to treat the fuzzy data which build the TFN based on percentage error is presented in steps 1 - 7 as follows:

Step 1: Data collection

Collected data can be arranged in the format as in Table 1.

Table 1. Format for Input Data

Sample	Input
1	y_1
2	y_2
...	...
n	y_n

Step 2: Build Triangular Fuzzy Number

Determine the spread of TFN using 1%, 3% and 5% of measurement error. Assume p is spread of TFN. $p = 1\%, 3\%, 5\%$

The triangular fuzzy number is written as follows

$$\tilde{y}_t^p = [y_t - (y_t \cdot p), y_t, y_t + (y_t \cdot p)] \quad (5)$$

where p is the spread in the form of $y_t 0.01, y_t 0.03, y_t 0.05$.

\tilde{y}_t^p is the estimated value of y_t in which the y_t is the observation value of a time series at period of $t, t = 1, 2, \dots, T$.

Step 3: Construct TFN based on Step 2 The TFN which is built from Step 2 can be arranged in the Table 2 data format.



Table 2. Format for Input Fuzzy Data

Sample	Input	TFN		
		1%	3%	5%
1	y_1	$[y_1^L 0.01, y_1^R 0.01]$	$[y_1^L 0.03, y_1^R 0.03]$	$[y_1^L 0.05, y_1^R 0.05]$
2	y_2	$[y_2^L 0.01, y_2^R 0.01]$	$[y_2^L 0.03, y_2^R 0.03]$	$[y_2^L 0.05, y_2^R 0.05]$
...
n	y_n	$[y_n^L 0.01, y_n^R 0.01]$	$[y_n^L 0.03, y_n^R 0.03]$	$[y_n^L 0.05, y_n^R 0.05]$

Step 4: Find average from predicted values, \bar{y}_t^s

Step 5: Forecast original data (single point) and find AR (1) model based on Eq. (6)

$$y_t = \beta_1 - \beta_2 y_{t-1} \quad (6)$$

Step 6: Forecast fuzzy data with spread, p and find AR (1) model based on Eq. (7)

$$y_t^s = (\beta_1^L, \beta_1^R) - (\beta_2^L, \beta_2^R) y_{t-1} \quad (7)$$

Step 7: Evaluate the performance between the AR (1) models based on single point (crisp) data and fuzzy data using Mean Square Error (MSE) as shown in equation (8):

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_1 - \hat{Y}_1)^2 \quad (8)$$

where y_1 is the actual observation value based on day one, $t = 1$ and \hat{Y}_1 is the predicted observation value based on day one. Smaller MSE indicates higher accuracy of the predicted model.

Step 7: Find best point from intervals.

Using triangular fuzzy number results in interval value which gives more flexibility in selecting the decision point. However, in this study, the best single point can also be identified as some decision maker may require single point decision rather than interval.

Table 3. AR (1) Identification

Exchange Rate	AR (1) Identification			
	Coefficient	P Value	Constant	P Value
BND	0.1895	0.000	1.10493	0.000
SGD	0.9912	0.000	0.0120908	0.000
THB	0.9985	0.000	0.051682	0.000
KHR	0.8491	0.000	611.929	0.000

Table 3 shows the dataset series which fulfills AR (1) requirement. As mentioned in AR (1) model formulation in Section II, the P value for constant and coefficient should significantly $P < 0.05$. From the result, it shows that the P value for both constant and coefficient of BND, SGD, THB and KHR are 0.000. Hence, we accept that the series of BND, SGD, THB and KHR have the AR (1) structure and could be used for next stage of building the TFN. Following the steps of building the TFN in Section III, AR (1) based on TFN can be written as follows:

$$\tilde{Y}_t = 1.3236 + 0.1895 Y_{t-1}$$

$$\tilde{Y}_t = [1.3103, 1.3368] + 0.1895 Y_{t-1} \quad (7)$$

Based on Equation (7), the proposed model has left (0.21132) and right (0.21571) spread of TFN values to

IV. NUMERICAL EXPERIMENT AND DISCUSSION

In this section, we present the implementation of the proposed model to predict ASEAN countries exchange rate datasets. Malaysian Ringgit (MYR), Brunei Dollar (BND), Indonesian Rupiah (IDR), Singapore Dollar (SGD), Thailand Baht (THB), Philippine Peso (PHP), Cambodian Riel (KHR), Myanmar Kyat (MMK), Laos Kip (LAK) and Vietnamese Dong (VND) are ten countries included in the ASEAN country. All are denominated in US Dollar (USD) as the USD is well known largest ASEAN trading partners.

The collected data are originally crisp daily data from 1st January 2015 until 30th September 2018 observation. Following the step of AR (1) identification, all the exchange rate data series were tested using PACF before estimate the coefficient and constant. The step of AR (1) identification is essential to identify the data stationarity[15]. From ten ASEAN countries, only four countries of BND, SGD, THB and KHR meet the requirement of AR (1) and are tabulated in Table 3.

contribute the \tilde{Y}_t of exchange rate which differ from actual AR (1) model that only has a value of constant. These intervals give a range of values instead of assigning the estimated rigidly crisp values[16].

In the early stage of data preparation, ten ASEAN countries exchange rate series are examined. However, after the stage of AR(1) identification, only BND, SGD, THB and KHR exchange rates series meet the requirement of AR (1) model. Then, TFN based on 1%, 3% and 5% spread are constructed to these four exchange rate 996 dataset series. The sample of TFN construction for exchange rate series is shown in Table4.

Table 4. Percentage Error based TFN

Sample	Actual	TFN – 1% y_t 0.01	TFN – 3% y_t 0.03	TFN – 5% y_t 0.05
1	1.3236	[1.310364, 1.336836]	[1.283892, 1.363308]	[1.25742, 1.38978]
2	1.3312	[1.31788, 1.344512]	[1.291264, 1.371136]	[1.26464, 1.39776]
...
995	1.3665	[1.352835, 1.380165]	[1.325505, 1.407495]	[1.298175, 1.434825]
996	1.1549	[1.143351, 1.166449]	[1.120253, 1.189547]	[1.097155, 1.212645]

Then, the proposed models based on 1%, 3% and 5% spread are compared with the standard AR (1) using mean squared error (MSE). Table5 shows the MSE comparison between the actual model and proposed model based on TFN spread of 1%, 3% and 5%.

Table 5. MSE Comparison

TFN	BND		SGD		THB		KHR	
	Trainin ng	Testin g	Trainin g	Testing	Trainin g	Testing	Trainin ng	Testin g
y_t	0.0008	0.0247	0.00002	0.00001	0.0084	0.01	306.47	178.92
5%	0.0009	0.0244	0.00003	0.00002	0.0156	0.0162	412.80	264.29
3%	0.0008	0.0246	0.00002	0.00001	0.0093	0.0108	321.06	186.05
1%	0.00078	0.0247	0.00002	0.00001	0.0084	0.01000	306.78	178.40

Table5 shows the series of dataset which meet the AR (1) requirement divide into training and testing. The training and testing data are tested by the MSE to analyze the accuracy. From the result, it shows that the smallest MSE (0.0244) for BND dataset is belong to 5% TFN when compared with actual MSE which is 0.0247. For SGD, TFN with 1% spread has similar MSE with actual MSE (0.000013). But, when compare with TFN with 3% spread and 5% spread, TFN 1% spread has smallest MSE rather than the other spreads. For THB and KHR, the TFN 1% spread also gives smaller MSE rather than TFN 3% and 5%. From the result, we may conclude that TFN with 1% spread could give higher accuracy in prediction model as it gives smallest value of MSE for testing. The result of TFN 1% spread is in interval form which has a range of values. Based on the range of values, we want to find which point that gives the best MSE value. The objective of point estimation is to obtain a single number from the interval which will represent the best point among the range of number that has smallest MSE [12]. Table 6 shows the result of best point for each series of exchange rate respectively.

Table 6. Best Point from Intervals

Exchange rate series	Actual Value	Interval	Best point	MSE
BND	1.1	[1.05,1.16]	1.12	0.000631
SGD	0.012	[0.0119, 0.0122]	0.0121	0.000026
THB	0.0517	[0.0512, 0.0522]	0.0512	0.008391
KHR	611	[605, 618]	612	306.4548

Table6 shows the best point for each datasets based on the interval result in which has the smallest MSE value. For BND, the smallest MSE is TFN 5% spread where the result is in interval values. Based on this TFN 5% spread, we find the best point from interval [1.05, 1.16] is 1.12. For SGD, THB and KHR, the smallest MSE is belonging to TFN 1% spread. Hence, the best point finding is based on the interval of 1%. From the result, it shows that the best point for SGD is 0.0121 from the interval of [0.0119, 0.0122]. While, 0.0512 is the best point for THB based on the interval of [0.0512, 0.0522]. As for KHR, the best point is 612 based on the interval of [605, 618].

V. CONCLUSION

The data preparation steps to treat data which contain uncertainty (known as Fuzzy data) is proposed and defined to be incorporated in AR(1) modeling. Data uncertainty were handled by implementing the Triangular Fuzzy Number which is characterized by different spread of 1%, 3% and 5% based on percentage error method. The performance of the proposed model was discussed using an application of prediction ASEAN countries exchange rate. The efficiency of percentage error based triangular fuzzy number for auto regressive model was shown based on the accuracy analysis. The results reveal that the proposed percentage error based triangular fuzzy number can achieve similar and better result when compare with conventional Auto Regression model, AR (1). Besides, the proposed model could be ranked to see which spread gives higher accuracy regarding on the smallest MSE. Further, the method also determines the best point from the interval in which has smallest MSE.

VI. ACKNOWLEDGMENT

The authors would like to express an appreciation to the University Tun Hussein Onn Malaysia (UTHM) and for funding project through the grant of TIER 1 Vot U895.

REFERENCES

1. Aubinet, M., Vesala, T., & Papale, D. (Eds.). Eddy covariance: a practical guide to measurement and data analysis. Springer Science & Business Media, (2012).
2. Jung, H. Y., Lee, W. J., & Yoon, J. H. A unified approach to asymptotic behaviors for the autoregressive model with fuzzy data. *Information Sciences*, 257, 127-137, (2014).
3. Ozawa, K., Watanabe, T., & Kanke, M. Fuzzy auto-regressive model and its applications. In *Knowledge-Based Intelligent Electronic Systems, 1997. KES'97. Proceedings., 1997 First International Conference on (Vol. 1, pp. 112-117)*. IEEE. (1997).
4. Greco, R. A fuzzy-autoregressive model of daily river flows. *Computers & geosciences*, 43, 17-23. (2012).
5. Kaku, Y., Kuramoto, K., Kobashi, S., & Hata, Y. Asthmatic attacks prediction considering weather factors based on Fuzzy-AR model. In *Fuzzy Systems (FUZZ-IEEE), 2012 IEEE International Conference on (pp. 1-4)*. IEEE. (2012).
6. Carbajal-Hernández, J. J., Sánchez-Fernández, L. P., Carrasco-Ochoa, J. A., & Martínez-Trinidad, J. F. Assessment and prediction of air quality using fuzzy logic and autoregressive models. *Atmospheric Environment*, 60, 37-50. (2012).
7. Jurado, K., Ludvigson, S. C., & Ng, S. Measuring uncertainty. *American Economic Review*, 105(3), 1177-1216. (2015).
8. Rahmani, A., HosseinzadehLotfi, F., Rostamy-Malkhalifeh, M., & Allahviranloo, T. (2016). A new method for defuzzification and ranking of fuzzy numbers based on the statistical beta distribution. *Advances in Fuzzy Systems*, (2016).
9. Dai, X., Liu, J., & Zhang, H. Application of AR model in the analysis of preearthquake ionospheric anomalies. *Mathematical Problems in Engineering*, (2015).
10. Tseng, F. M., Tzeng, G. H., Yu, H. C., & Yuan, B. J. Fuzzy ARIMA model for forecasting the foreign exchange market. *Fuzzy sets and systems*, 118(1), 9-19. (2001).
11. Pouzols, F. M., & Barros, A. B. Automatic clustering-based identification of autoregressive fuzzy inference models for time series. *Neurocomputing*, 73(10-12), 1937-1949. (2010).
12. Anand, M.C.J., & Bharatraj, J. Theory of Triangular Fuzzy Number, proceeding NCATM, (2017).
13. Bezruck, V. M., Belov, Y. N., Voitovich, O. A., Netrobenko, K. A., Tikhonov, V. A., Rudnev, G. A., & Khomenko, S. I. Application of autoregressive model for recognition of meteorological objects. In *Radar Symposium (IRS), 2010 11th International. (pp. 1-3)*. IEEE. (2010)
14. Ganapathy, S., Thomas, S., Motlicek, P., & Hermansky, H. Applications of signal analysis using autoregressive models for amplitude modulation. In *Applications of Signal Processing to Audio and Acoustics, WASPAA'09. IEEE Workshop on (pp. 341-344)*. IEEE. (2009)
15. Efendi, R., Arbaiy, N., & Deris, M. M. A new procedure in stock market forecasting based on fuzzy random auto-regression time series model. *Information Sciences*; 441, 113-132. (2018)
16. Nureize, A., & Watada, J. Building fuzzy random objective function for interval fuzzy goal programming. In *Industrial Engineering and Engineering Management (IEEM), 2010 IEEE International Conference on (pp. 980-984)*. IEEE. (2010)