Indonesian Journal of Electrical Engineering and Computer Science
Vol. 18, No. 3, June 2020, pp. 1559~1567
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v18.i3.pp1559-1567

A simulation study of first-order autoregressive to evaluate the performance of measurement error based symmetry triangular fuzzy number

Muhammad Shukri Che Lah, Nureize Arbaiy
Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn, Malaysia

Article Info

Article history:

Received Sep 14, 2019
Revised Dec 15, 2019
Accepted Dec 30, 2019

Keywords:

Error percentage
First-order autoregressive
Measurement error
Standard deviation
Triangular fuzzy number

ABSTRACT

Data collected by various data collection methods are often exposed to uncertainties that may affect the information presented by quantitative results. This also causes the forecasted model developed to be less precise 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 forecasting model to be less accurate. Traditional autoregressive (AR) model uses precise values and deals with the uncertainty normally in forecasting model. Fewer researches are focused on data preparation in time-series autoregressive for handling the uncertainties in data. Hence, this paper proposes a procedure to perform data preparation to handle uncertainty. The fuzzy data preparation involves the construction of fuzzy symmetric triangle numbers using percentage error and standard deviation method. The proposed approach is evaluated by using the simulation method for first-order autoregressive, AR (1) model in terms of forecasting accuracy performance. Simulation result demonstrates that the proposed approach obtains smaller error in forecasting and hence achieving better forecasting accuracy and dealing with uncertainty in the analysis.

Corresponding Author:

Nureize Arbaiy,
Faculty of Computer Science and Information Technology,
Universiti Tun Hussein Onn, Malaysia.
Email: nureize@uthm.edu.my

1. INTRODUCTION

In data science and analytic process, assembling an appropriate time series data input is a crucial task and challenging. In time series forecasting, past observations are collected and analysed to develop a suitable mathematical model which captures the underlying data generating process for the series [1, 2]. Time series observations are frequently encountered in many domains such as business, economics, industry, engineering and science, etc [1, 2]. There are various types of time series depending on the type of analysis and practical needs.

In time series observation, data inputs should be processed first for many reasons such as missing data, imbalance data, and incomplete data called data preparation. This is to avoid misleading results during analysis. Data which are collected from various measurements carries some degree of uncertainty [3] which giving rise to data uncertainty. It is important to describe the uncertainty in data to obtain realistic result from data analysis. One of the most challenging issues is interpreting evidence of observations that contain inherent measurement errors. Observation errors may derive from the measurement process [4]. These errors are the combined measure of the variations that exist in observed phenomenon and various factors that interfere with measurements [5, 6]. The error resulting from measured data can potentially lead to a bias in estimates [7, 8]. Since eliminating all measurement error is impossible to obtain totally precise value, estimation or approximation is used with certain limits.

The uncertainties contained in the historical data which is used to build a forecasting model can affect the model's performance. The existence of inherent uncertainties makes the conventional analysis incapable to deal with such data [9, 10]. Time-series records single-point data values which only best suit to the conventional time-series analysis such as autoregressive. However, most studies focus on model uncertainty regardless of data uncertainty. The data-processing carried out may not always take care of uncertainty. While uncertainty in the input data is not sufficiently handled, this creates more errors to be included in the predicted model. Standard procedures are also very limited to be followed to address the uncertainty in data. Since these uncertainties may have substantial implications for interpreting time-series forecasting in AR, a systematic technique for dealing with uncertainties during data preparation is necessary [11-14].

Conventional autoregressive (AR) method has limited capabilities to deal with uncertainty in data while building the time series model. It makes the conventional AR disregard inherent uncertainty in data preprocessing [10]. It is common in practice where the data are changing rapidly and has uncertainty. To address the limitation, the fuzzy theory is introduced in the AR model building which solves the uncertainty issues. In previous literatures, the issue of uncertainty has been studied by several authors by introducing fuzzy theory [14-18]. For example, fuzzy inference system was used to create a new air quality index while autoregressive model used to predict future air quality condition. The concept of fuzzy regression model was combined with the AR model to formulate the fuzzy AR model and applied in the forecasting [19-21]. Most of the researches deal with uncertainties in the AR model. However, only few discuss treatment to the fuzzy data during data preparation phase though it is important to treat the data before building the forecasting model. To overcome these issues, a new approach in data preparation is required to improve the accuracy in time-series forecasting with uncertainties.

This paper presents a fuzzy data preparation procedure which involve fuzzy data transformation from precise value to fuzzy number to obtain appropriate fuzzy values. The transformation involved building a triangular fuzzy number (TFN) by using percentage error and standard deviation methods. These two methods are associated with measurement error. Simulation study is performed to obtain empirical results. The emphasis made in this study on the preparation of fuzzy data is very important to increase predictive value and to achieve predictive accuracy. The result indicates the improvement in the accuracy achievement by employing the proposed approach. The rest of the paper is structured as follows. Section 2 provides a brief review of the related work. The proposed approach is explained in Section 3. Section 4 illustrates empirical studies with developed triangular fuzzy number using simulation method. Finally, concluding remarks are presented in Section 5.

2. RELATED WORK

In forecasting, data collection is a preliminary step before data pre-processing and preparation. Relevant data must be available and correct in order to have a reliable forecasting value. Observing, interviewing and questionnaire are several common techniques used in data collection. Certainly, assembling appropriate data from different sources are often challenging and time-consuming task [22]. Data preparation is the next step that changes the data according to the format that corresponds to the needs of the researcher. Data preparation is a crucial task and should be accomplished first to prevent mistakes. By organizing the data correctly also can save a lot of time during the step of analysis.

However, it is a common situation where the process in data collection and preparation is exposed to uncertainty and measurement error. Uncertainty may come from expert differences or due to systematic differences within the dataset [22, 23]. Random error and systematic error are included in the measurement error which can affect all measurement [23]. Thus, to overcome the problem in data preparation, TFN was introduced and applied in different applications [10, 24-27]. The implementations of triangular fuzzy number in these researches show that the fuzzy number is more realistic in describing the physical world than single-valued number.

Time series modelling has the fundamental importance of various domain domains and has contributed to many active researches works to date. Many important models have been proposed in the literature to improve the accuracy and efficiency of the time series and its forecasting model [1, 2, 28]. For example, the Autoregressive model [28] has been widely known to contribute to real-world problems [2], [13], although its prediction accuracy is still a critical issue. On the other hand, the ultimate aim of time series modelling is to carefully collect and analyze the past time series observations to develop a suitable model that describes the pattern that exists in the series [4]. A successful time series forecasting depends on the fitting of the appropriate model. Since the model also depends on the data, then data preparation should be carefully performed [3].

In addition, an efficient method is necessary to handle the complexity and adequacy of the uncertainty associated with real-world problems while forecasting. The traditional forecasting method is incapable to work with data which contain uncertainties [9, 10, 14]. However, fuzzy set theory has been identified to solve the problem of uncertainty, and deal with incomplete, inaccurate information of both qualitative and quantitative by nature. Previously fuzzy theory provides a remarkable solution to deal with fuzzy data. Fuzzy theories play a significant role to treat the uncertainties in real world. Although previous studies have used fuzzy data to deal with uncertainties, only few are intensely providing a systematic procedure which converts real data to fuzzy data. Hence, this study gives a new procedure of fuzzy data preparation for time series analysis.

3. PROPOSED FUZZY DATA PREPARATION PROCEDURE FOR FORECASTING AR(P) MODELS

This section discusses the procedure of fuzzy data preparation. The uncertainty in data is treated by the fuzzy approach by constructing a symmetric triangular fuzzy number (STFN) based on Percentage Error (PE) and Standard Deviation (SD) methods. These new methods of PE and SD are given subsequently. The simulation procedure of two proposed approach of building STFN is perfomed in first-order autoregressive, written as AR(1). The outcome variable in an AR(1) process at some point in time t is related only to time periods that are one period apart. Both methods are simulated and validated using generated datasets. The generated data is used to provide the flexibility to generate the number of samples and the AR(1) model itself.

In this study, a tool named as AR Generator (Argen) is developed to assist the simulation process. By using Argen, simulation datasets are generated by random function and build fuzzy triangles from single value data using PE and SD methods. Argen is a software tool that capable to accept the AR(1) model and the number of required samples as input. Using the given information, the tool generates datasets and builds fuzzy triangles. Argen allows importing the generated data into comma-separated values (CSV) for further analysis. Figure 1 shows the interface of Argen.

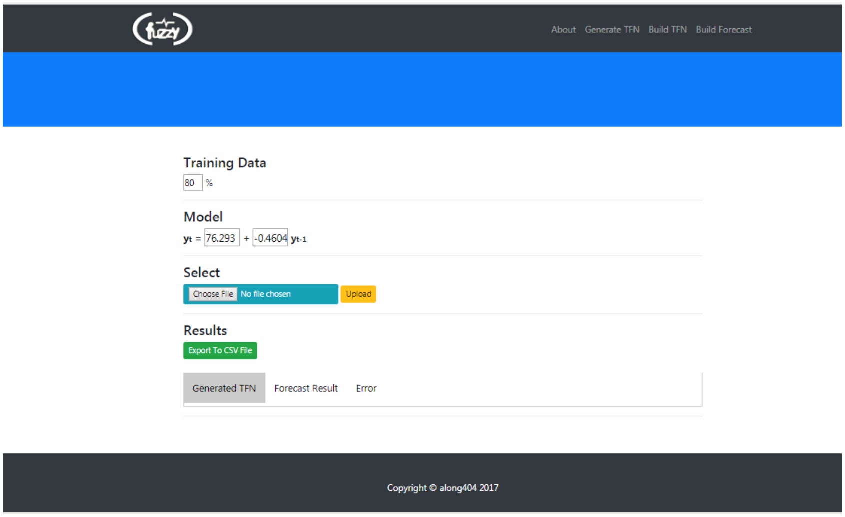

Figure 1. Interface of Argen

3.1. Construction of Symmetry Triangular Fuzzy Number Based on Percentage Error and Standard Deviation

Data preparation is very important phase to be considered before going to another analytical processes [29]. Data preparation transforms the data sets so that their information content is best exposed to the next process. Many studies [14-18] in AR frequently uses single point values as data input in building forecasting model. However, most single point values use the average data where this data will affect the standard deviation. A small standard deviation can be a goal in certain situations where the results are restricted [30], thus the data can be considered as good. Unfortunately, if the value of standard deviation is large, the probability of error is high. In addition, some data is collected several times over a period of time. Therefore, it makes the data can not represent the nature of the data and is less suitable for forecasting.

Figure 2 shows the transformation of the single point data to the symmetry triangular fuzzy number. Theoretically within the confidence interval and the standard deviation, the smaller the spread, the better the result is [31, 32]. However, in order to build the spread, there is a very limited standard procedure to be used. For example, [10] explained the triangular fuzzy number can be formed using low-high data. Unfortunely, it is inappropriate to construct TFN based on low-high data because that might be leading to bias data used. In addition, not all data has a low and high value. It makes difficulty when the original data contains only one value. Motivated by this situation, the procedure to transform crisp data to fuzzy data is presented. Fuzzy data transformation is performed by building the symmetry triangular fuzzy number using two methods of percentage error and standard deviation. In (1) shows the general equation for fuzzy triangle.

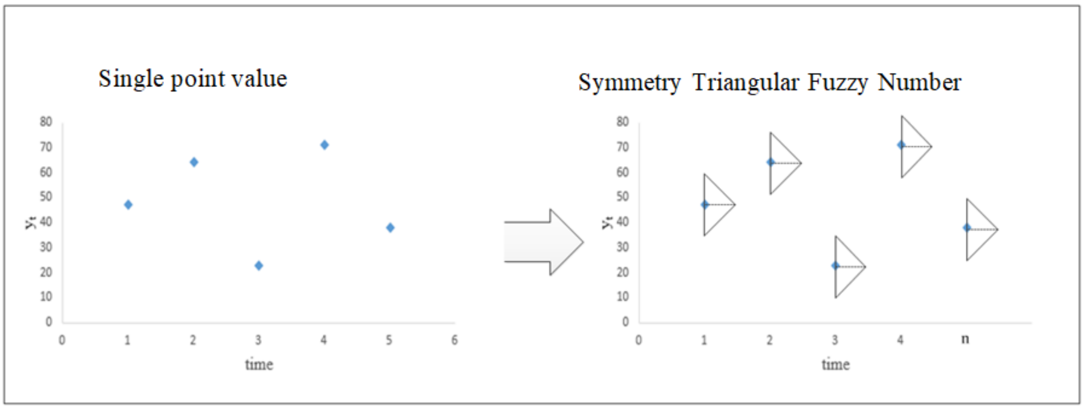

Figure 2. Transformation of the single point data into STFN

$$\tilde{y}_t^{\Delta} = [(y_t - \Delta), y_t, (y_t + \Delta)] \quad (1)$$

where $\tilde{y}_t^{\Delta}$ is a fuzzy time series data and y_t is a time series data at time, $t\ (t = 1,2, \dots, n)$. The spread of the TFN, Δ is characterized by $(y_t . p)$ for PE or s for SD. The following subsection describes the two proposed approach to build triangular fuzzy number.

3.1.1. Percentage Error (PE)

Based on 95% confidence interval, only 5% spread and below is considered the best distribution to use. Thus, the spread of STFN is adjusted to 5%, 3% and 1% according to percentage error. From (1), the STFN for percentage error method is rewritten as follows:

$$\tilde{y}_t^{p} = [y_t - y_t . p, y_t, y_t + y_t . p] \quad (2)$$

where $\tilde{y}_t^{p}$ is a fuzzy time series data at time, t with STFN form with spread, p $(p = 0.01, 0.03, 0.05)$. y_t is a time series data at time, $t\ (t = 1,2, \dots, n)$. Figure 3 shows the STFN with a spread that is built using PE method.

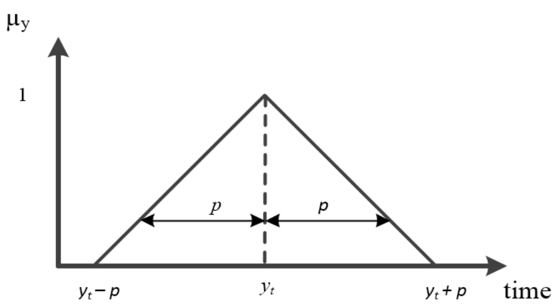

Figure 3. STFN with spread, p of PE method

3.1.2. Standard Deviation

Standard deviation data is considered as a spread because the nature of standard deviation which measures the spread of a data distribution. The deviation data can be left or right from the center. Thus, based on (1) the STFN form for standard deviation is rewritten as follows:

$$\tilde{y}_t^s = [y_t - s, y_t, y_t + s] \tag{3}$$

where $\tilde{y}_t^s$ is a fuzzy time series data at time, t with spread, s of y_t. y_t is a time series data at time, t ($t = 1,2, \dots , n$). Figure 4 shows STFN whereby the spread is determined by using a standard deviation method.

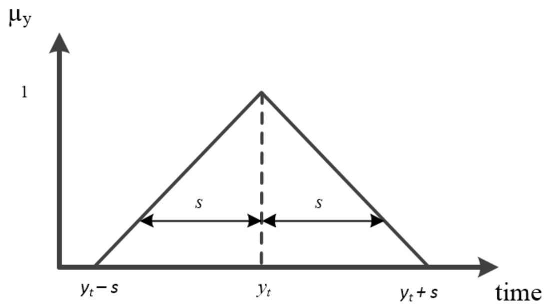

Figure 4. STFN with spread, s of standard deviation method

Percentage error method in this study uses three values of 1%, 3%, and 5% to determine the fuzzy triangle's spread, p. Then, three variants of a symmetry fuzzy triangle with the spread, p is produced. Meanwhile, the standard deviation method produces one value spread, s. All four variants of STFN is shown in Table 1.

Since the two methods of PE and SD produces the STFN with different spread value, there is a need to evaluate the best STFN among them. Section 3.2 give details to evaluate the best STFN through simulation technique.

Table 1. Symmetry Triangular Fuzzy Number for y_t.

Method	Percentage error						Standard deviation	
y_t	$y_t^{p_{0.05}}$		$y_t^{p_{0.03}}$		$y_t^{p_{0.01}}$		y_t^{S}	
	Left	Right	Left	Right	Left	Right	Left	Right
y_1	$y_1 + p_{0.05}$	$y_1 - p_{0.05}$	$y_1 + p_{0.03}$	$y_1 - p_{0.03}$	$y_1 + p_{0.01}$	$y_1 - p_{0.01}$	$y_1 + s$	$y_1 - s$
y_2	$y_2 + p_{0.05}$	$y_2 - p_{0.05}$	$y_2 + p_{0.03}$	$y_2 - p_{0.03}$	$y_2 + p_{0.01}$	$y_2 - p_{0.01}$	$y_2 + s$	$y_2 - s$
y_3	$y_3 + p_{0.05}$	$y_3 - p_{0.05}$	$y_3 + p_{0.03}$	$y_3 - p_{0.03}$	$y_3 + p_{0.01}$	$y_3 - p_{0.01}$	$y_3 + s$	$y_3 - s$
…	…	…	…	…	…	…	…	…
y_{99}	$y_{99} + p_{0.05}$	$y_{99} - p_{0.05}$	$y_{99} + p_{0.03}$	$y_{99} - p_{0.03}$	$y_{99} + p_{0.01}$	$y_{99} - p_{0.01}$	$y_{99} + s$	$y_{99} - s$
y_{100}	$y_{100} + p_{0.05}$	$y_{100} - p_{0.05}$	$y_{100} + p_{0.03}$	$y_{100} - p_{0.03}$	$y_{100} + p_{0.01}$	$y_{100} - p_{0.01}$	$y_{100} + s$	$y_{100} - s$

3.2. Simulation Procedure for $AR(1)_p$

Simulation is the process of designing a real-world system and perform experiments to understand the behaviour of the system or to evaluate various strategies [32]. In the other words, simulation is the imitation of the operation of a real-world process time [33]. Simulation is used to analyse the response of a system over its input variation. Therefore, simulation needs to be done before the actual situation occurs to reduce the effect of the event. The simulation is necessary to obtain the empirical result. Figure 5 shows the flowchart of simulation.

Based on Table 1, there are four TFNs to be produced using two proposed methods. The following procedure is then used to determine the best TFN among the four.

Step 1. Provide AR(1) model as shown in (1)
Step 2. Generate error, a_t of AR(1) model for t, ($t = 1, 2, \dots, n$), with k experiment ($k = 100, 200, \dots$) and determine values of y_t.
Step 3. Build triangular fuzzy number, $\tilde{y}_t^A$ using proposed approaches as in Section 3.1.
Step 4. Estimate parameters model AR(1) for each $\tilde{y}_t^p$.

Step 5. Find parameters average (coefficient and constant) of each $\tilde{y}_t^p$.
Step 6. Validate parameters average (coefficient and constant) of each $\tilde{y}_t^p$.

Based on the procedures given, it is expected to provide guidance to provide time series data that contain uncertainty or fuzzy information for forecasting.

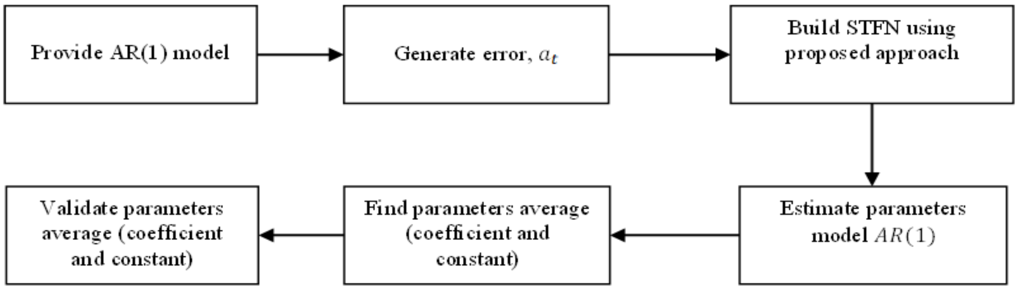

Figure 5. Simulation's flowchart

4. NUMERICAL EXAMPLE

This section gives explanation with numerical example. The simulation is performed based on the following steps with experiment $k = 1$ and $p = 0.05$.

Step 1. Provide AR(1) model. In this example, the AR(1) model used is as follows:

$$y_t = 0.8y_{t-1} + a_t \tag{4}$$

Step 2. Based on model selected in Step 1, error a_t with sample size, $t = 100$ and number of experiments, $k = 100$ is generated.

Step 3. Build triangular fuzzy number which uses 5% percentage error $y_t^{p_{0.05}}$. Table 2 shows the produced spread for $\tilde{y}_t^{p_{0.05}}$.

Table 1. The Spread for $y_t^{p_{0.05}}$

	a_t	y_t	$y_t^{p_{0.05}}$	
			Left	Right
y_1	0.2601	0.2601	0.2471	0.2731
y_2	0.9684	1.1765	1.1177	1.2353
y_3	0.1531	1.0943	1.0396	1.149
…	…	…	…	…
y_{99}	0.3595	2.3325	2.2159	2.4491
y_{100}	0.8148	2.6808	2.5468	2.8149

Step 4. Parameters model AR(1) is estimated for each $y_t^{p_{0.05}}$. Table 3 shows the coefficient's and constant's value for $y_t^{p_{0.05}}$.

Step 5. Find parameters average (coefficient and constant) of each $y_t^{p_{0.05}}$. Table 4 shows the average coefficient and average constant value for $y_t^{p_{0.05}}$.

Table 2. Coefficient, $\emptyset_1$ and Constant, $\emptyset_2$ for $y_t^{p_{0.05}}$

	y_t		$y_t^{p_{0.05}}$			
			Left		Right	
	$\emptyset_1$	$\emptyset_2$	$\emptyset_1$	$\emptyset_2$	$\emptyset_1$	$\emptyset_2$
k_1	0.8556	0.3176	0.8556	0.3017	0.8556	0.3335
k_2	0.9241	0.1648	0.9241	0.1565	0.9241	0.1730
k_3	0.8740	0.2744	0.8740	0.2607	0.8740	0.2881
…	...	...	...	...	...	…
k_{99}	0.7555	0.5361	0.7555	0.5093	0.7555	0.5629
k_{100}	0.9199	0.1745	0.9199	0.1658	0.9199	0.1832

Table 4. Average Coefficient, $\emptyset_1$ and Average Constant, $\emptyset_2$ for $y_t^{p_{0.05}}$

	y_t		$y_t^{0.05}$	
	$\emptyset_1$	$\emptyset_2$	$\emptyset_1$	$\emptyset_2$
k_1	0.8556	0.3176	0.8556	0.4684
k_2	0.9241	0.1648	0.9241	0.1647
k_3	0.8740	0.2744	0.8740	0.2744
…	...	...	...	...
k_{99}	0.7555	0.5361	0.7555	0.5361
k_{100}	0.9199	0.1745	0.9199	0.1742

Step 6. Validate parameters average (coefficient and constant) of each $\tilde{y}_t^p$.

The experiment continues until $k = 100$. The result for one hundred experiment is presented and discussed in Section 5.

5. RESULTS AND ANALYSIS

In this section, parameters validation for the AR(1) model and the results obtained will be addressed and discussed to choose the best STFN. With reference to Table 1, four STFNs have been obtained and tested to obtain parameter values for coefficients and constants. All four values of STFN coefficients will be compared with the coefficient value using conventional methods. Table 5 shows the minimum, maximum and mean values for one hundred experiments.

Table 5. Minimum, Maximum and Average of Coefficient, $\emptyset_1$

	y_t	$y_t^{p_{0.05}}$	$y_t^{p_{0.03}}$	$y_t^{p_{0.01}}$	y_t^S
Minimum	0.7172	0.7172	0.7172	0.7172	0.7172
Maximum	0.9602	0.9602	0.9602	0.9602	0.9601
Average	0.8829	0.8829	0.8829	0.8829	0.8829

From Table 5, the coefficient values produced by the four STFNs are the same and cause difficulty in choosing the best method. Therefore, a constant value has been used to select the best STFN. Table 6 shows the minimum, maximum and average values for one hundred experiments.

Table 6. Minimum, Maximum and Average of Constant, $\emptyset_2$

	y_t	$y_t^{p_{0.05}}$	$y_t^{p_{0.03}}$	$y_t^{p_{0.01}}$	y_t^S
Minimum	0.08443	0.08444	0.08444	0.08445	0.08446
Maximum	0.71358	0.71357	0.71354	0.71355	0.71357
Average	0.26568	0.26568	0.26577	0.26567	0.26568

Table 6 shows that there is slightly difference in all constant values. From observation on Table 5 and Table 6, $y_t^{0.01}$ indicate the best STFN because the coefficient criterion is closer to 0.8 as shown in (4), and the constant value is close to zero. However, other STFNs are also acceptable due to the very small differences in value.

6. CONCLUSION

Autoregressive forecasting is an interesting topic for researchers to find. There are many models produced to predict autoregressive but only certain researchers who focus on data preparation. This study presents a fuzzy data preparation by using symmetry triangular fuzzy number which are induced from percentage errors and standard deviation methods. In addition, this study also proposes a procedure to simulate the generated dataset to obtain the best STFN between the proposed STFN. Four STFN variants were generated and simulation results showed STFN with 1% spread achieved the best results. However, based on Table 5 and Table 6 also shows that other STFNs are acceptable because they are not much different from conventional models. The efficiency of percentage error and standard deviation based triangular fuzzy number for AR(1) model was shown based on the accuracy analysis. The results reveal that the proposed approach can achieve a similar and better result when compared with conventional AR(1). The results also obtain empirical evidence to support the case for the proposed methods.

ACKNOWLEDGEMENTS

The author would like to extend its appreciation to the Ministry of Higher Education (MOHE) and Universiti Tun Hussein Onn Malaysia (UTHM). This research is supported by Tier 1 Research grant (Vot U895) and Geran Penyelidikan Pascasiswazah (GPPS) grant (Vot U975). The author thanks to the anonymous viewers for the feedback.

REFERENCES

[1] G. P. Zhang, “A neural network ensemble method with jittered training data for time series forecasting,” *Inf. Sci. (Ny).*, vol. 177, no. 23, pp. 5329-5346, 2007.
[2] G. P. Zhang, “Time series forecasting using a hybrid ARIMA and neural network model,” *Neurocomputing*, vol. 50, pp. 159-175, 2003.
[3] W. Eugster, "Eddy covariance: A practical guide to measurement and data analysis", no. May 2014. *Springer Science & Business Media*, 2012.
[4] J. S. Clark and O. N. Bjornstad, “*Population time series: process variability, observation errors, missing values, lags, and hidden states*,” vol. 85, no. 11, pp. 3140-3150, 2013.
[5] H. W. Coleman and W. G. Steele, “Experimentation, Validation and Uncertainty Analysis for Engineers,” *J. Chem. Inf. Model*., 2018.
[6] S. S. Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, "Measurement Errors in Surveys". *John Wiley & Sons*, 2011, 2004.
[7] A. Ferrero and S. Salicone, “Modeling and processing measurement uncertainty within the theory of evidence: Mathematics of random-fuzzy variables,” *IEEE Trans. Instrum*. Meas., vol. 56, no. 3, pp. 704-716, 2007.
[8] A. Ferrero and S. Salicone, “An innovative approach to the determination of uncertainty in measurements based on fuzzy variables,” *IEEE Trans. Instrum. Meas*., vol. 52, no. 4, pp. 1174-1181, 2003.
[9] N. Arbaiy, N. A. Samsudin, J. Watada, and P. C. Lin, “An Enhanced Possibilistic Programming Model with Fuzzy Random Confidence-Interval for Multi-Objective Problem,” *Innov. Comput. Optim. Its Appl*., no. January, pp. 217-235, 2018.
[10] R. Efendi, N. Arbaiy, and M. M. Deris, “A new procedure in stock market forecasting based on fuzzy random auto-regression time series model,” *Inf. Sci. (Ny).*, vol. 441, pp. 113-132, 2018.
[11] P.-C. Chang, J.-L. Wu, and J.-J. Lin, “A Takagi-Sugeno fuzzy model combined with a support vector regression for stock trading forecasting,” *Appl. Soft Comput*., vol. 38, pp. 831-842, 2016.
[12] L. Maciel, F. Gomide, and R. Ballini, “Evolving Fuzzy-GARCH Approach for Financial Volatility Modeling and Forecasting,” *Comput. Econ*., vol. 48, no. 3, pp. 379-398, 2016.
[13] P. Singh, “An Efficient Method for Forecasting Using Fuzzy Time Series,” in *Emerging Research on Applied Fuzzy Sets and Intuitionistic Fuzzy Matrices, IGI Global*, 2017, p. 18.
[14] Z. Ismail, R. Efendi, and M. M. Deris, “Application of Fuzzy Time Series Approach in Electric Load Forecasting,” *New Math. Nat. Comput*., vol. 11, no. 03, pp. 229-248, 2015.
[15] Z. ISMAIL, R. EFENDI, and M. M. DERIS, “Inter-Quartile Range Approach To Length–Interval Adjustment of Enrollment Data in Fuzzy Time Series Forecasting,” *Int. J. Comput. Intell. Appl*., vol. 12, no. 03, p. 1350016, 2013.
[16] M. Y. Chen and B. T. Chen, “A hybrid fuzzy time series model based on granular computing for stock price forecasting,” *Inf. Sci. (Ny).*, vol. 294, pp. 227-241, 2015.
[17] R. EFENDI, Z. ISMAIL, and M. M. DERIS, “Improved Weight Fuzzy Time Series As Used in the Exchange Rates Forecasting of Us Dollar To Ringgit Malaysia,” *Int. J. Comput. Intell. Appl*., vol. 12, no. 01, p. 1350005, 2013.
[18] R. Efendi, Z. Ismail, and M. M. Deris, “A new linguistic out-sample approach of fuzzy time series for daily forecasting of Malaysian electricity load demand,” *Appl. Soft Comput. J*., vol. 28, pp. 422-430, 2015.
[19] A. Marszałek and T. Burczyński, “Modeling and forecasting financial time series with ordered fuzzy candlesticks,” *Inf. Sci. (Ny).*, vol. 273, pp. 144-155, 2014.
[20] Bor-Sen Chen, Sen-Chueh Peng, and Ku-Chen Wang, “Traffic modeling, prediction, and congestion control for high-speed networks: a fuzzy AR approach,” *IEEE Trans. Fuzzy Syst*., vol. 8, no. 5, pp. 491-508, 2002.
[21] K. Ozawa, ’Takahide Niimura, and T. Nakashima, “Fuzzy Time-Series Model of Electric Power Consumption,” *J. Adv. Comput. Intell. Intell. Informatics*, vol. 4, no. 3, pp. 188-194, 2016.
[22] J. E. Hanke and D. W. Wichern, "UPPER SADDLE RIVER, NEW JERSEY 07458 Ninth Edition BUSINESS FORECASTING". 2009.
[23] J. Buonaccorsi, “Measurement Error : Models, "Methods and Applications,” *Chapman and Hall/CRC*.
[24] J. S. Yao, J. S. Su, and T. S. Shih, “Fuzzy system reliability analysis using triangular fuzzy numbers based on statistical data,” *J. Inf. Sci. Eng*., vol. 24, no. 5, pp. 1521-1535, 2008.
[25] A. Gani, Nagoor and S. . Assarudeen, Mohamed, “A new operation on triangular fuzzy number for solving fuzzy linear programming problem A New Operation on Triangular Fuzzy Number for Solving Fuzzy Linear Programming Problem,” *Appl. Math. Sci.* , vol. 6 no. 11, no. September 2014, pp. 525-532 A, 2012.
[26] J. Wang, D. Ding, O. Liu, and M. Li, “A synthetic method for knowledge management performance evaluation based on triangular fuzzy number and group support systems,” *Appl. Soft Comput. J*., vol. 39, pp. 11-20, 2016.
[27] E. ATIFY, C. DAOUI, and A. BOUMEZZOUGH, “Using A Fuzzy Number Error Correction Approach to Improve Algorithms in Blind Identification,” *Indones. J. Electr. Eng. Comput. Sci*., vol. 3, no. 2, p. 410, 2016.
[28] M. Geurts, G. E. P. Box, and G. M. Jenkins, “Time Series Analysis: Forecasting and Control,” *J. Mark. Res.*, vol. 14, no. 2, p. 269, 2006.
[29] F. Tseng, G. Tzeng, H. Yu, and B. J. C. Yuan, “Fuzzy ARIMA model for forecasting the foreign exchange market,” vol. 118, pp. 9-19, 2001.
[30] D. Pyle, "Data Preparation for Data Mining". *Morgan Kaufman Publisher*, 1999.
[31] D. J. Rumsey, "Statistics For Dummies", *2-eBook Bundle Statistics For Dummies*. 2011.
[32] J.-B. du Prel, G. Hommel, B. Röhrig, and M. Blettner, “Confidence Interval or P-Value? Part 4 of a Series on Evaluation of Scientific Publications,” *Dtsch. Aerzteblatt Online*, vol. 106, no. 19, pp. 335-339, 2009.

BIOGRAPHIES OF AUTHORS

Muhammad Shukri Che Lah is currently aa Master degree student and Graduate Research Assistant in the Faculty of Computer Science and Information Technology Universiti Tun Hussein Onn, Malaysia (UTHM). He received his Bachelors degrees of Software Engineering in Faculty of Computer Science and Information Technology Universiti Tun Hussein Onn. His research focused on fuzzy data preparation and his current research work is on the data preparation on time series.

Nureize Arbaiy is currently with Software Engineering Department at the Faculty of Computer Science and Information Technology, University Tun Hussein Onn Malaysia (UTHM). She received her Dr. of Engineering from the Graduate School of Information, Production, and Systems at Waseda University, Japan. She received her B.S degree from University of Technology Malaysia in 2001 and M.S degree from University Utara of Malaysia in 2004. Her research interests are on multi-criteria decision making, fuzzy logic, expert system, fuzzy regression analysis, and possibilistic theory.