Outage Cost Modeling of High Energy Consumption Industries

Chaliew Ketkaew, Onurai Noohawm, and Dulpichet Rerkpreedapong

Abstract— *This paper presents modeling of customer outage cost of large industries in Thailand considered from electrical energy consumptions. The selected industries are classified into six Thai Standard Industrial Classification (TSIC) customer groups consisting of foods, textiles, chemical and rubber, iron and steel, cement and fabricated metals industries. The 279 factories in the central region of Thailand were surveyed to collect information for outage cost evaluation. Next, customer outage costs of those factories are modeled by using the statistic regression and fuzzy regression methods. Then, both modelings are analyzed and compared using statistical techniques. Finally, the method of selection between both modelings is presented in order to obtain a better model for industrial customers of each TSIC. Consequently, the customer outage cost assessment can be achieved from the proposed methodology, which is simple and offers a suitable model for individual industries.*

Keywords— **Customer outage cost, fuzzy regression, industrial customer groups.**

1. INTRODUCTION

All electric power utilities in Thailand aim to provide services to their customers with high efficiency, reliability and safety. A number of megaprojects have been implemented on power distribution infrastructures such as smart grids, risk and asset management, overhead to underground system conversion, micrograds, etc.

Customer outage cost is widely used by most utilities to make a decision on system investment. The outage cost is described as the damage costs due to power interruptions related to customers and utilities. The utilities can use it to prioritize locations or service areas for reliability improvement [4-6]. Also, it can be used by industries to evaluate the cost-benefit of their future projects [3,10].

The methods of outage cost modeling used in this paper are statistic regression and fuzzy regression. Generally, statistic regression is an effective tool to develop a mathematical model, in terms of regression parameters or regression coefficients, from the relationship between input and output data [8]. Meanwhile, the modeling resulted from statistic regression may result in a serious inaccuracy when the number of data is insufficient, or there is vagueness of the relationship between input and output variables [20]. On the other hand, many researches have shown that fuzzy regression can help fix these problems. Some applications of fuzzy regression are found in business pricing, health assessment [1], industrial process control [2], oil consumption forecasting and peak load estimation [12-16]. In this paper, the outage costs of customers by TSIC are modeled using both techniques. Then, the better suitable model is to be selected by statistical techniques for outage cost assessment of each TSIC.

2. METHOD

Customer Damage Function

Customer damage function is used to assess customer outage cost. Generally, sector customer damage functions (SCDF) are used to develop composite customer damage function (CCDF) [4]. SCDF is developed from data collected from customers via questionnaires. The damage costs are surveyed and evaluated for several outage durations, i.e., 1, 30, 60 minutes, etc. In this paper, the individual customer damage functions (CDF) as given in equation (1) are derived from the collected data of surveyed factories, and then the SCDF could be produced from those functions.

$$CDF(t) = DC.t + IC \tag{1}$$

where

CDF(t) is Customer damage function (Baht/interruption)
IC is Initial outage cost (Baht)
DC is Duration outage cost (Baht/hour)
t is Average outage duration time per interruption (minute)

Initial outage cost (IC) represents sudden losses and product impairment of industrial customers when the production lines are interrupted due to both momentary and sustained interruptions. In this paper, the initial outage cost is expressed in Baht per interruption, associated with the loss of production opportunity during the restart process. Duration outage cost (DC) represents duration losses caused by only sustained power

C. Ketkaew (corresponding author) is with Ragamangala University of Technology Suvarnabhumi, Suphanburi campus, 450 M.6, Suphanburi-Chainat Road, Yan-Yao, Samchuk, Suphanburi, Thailand 72130. E-mail: kchaliew@yahoo.com.

O. Noohawm and D. Rerkpreedapong are with the Department of Electrical Engineering, Kasetsart University, 50 Ngam Wong Wan Road, Lat-Yao, Chatuchak, Bangkok, Thailand 10900. E-mail: nonurai@hotmail.com and dulpichet.r@ku.ac.th.

interruptions including loss of production opportunity during a period of power outages.

Fuzzy Regression

Fuzzy linear regression is first developed by Tanaka et al [12]. This method is based on the principle of fuzzy sets (L.A. Zadeh) [11]. Statistic regression is based on probability theory, while fuzzy regression is based on possibility theory and fuzzy set theory [6]. Equation (2) shows the fuzzy regression model.

$$Y = A_0 + A_1x_1 + A_2x_2 + ... + A_nx_n \qquad (2)$$

where Y are fuzzy outputs or dependent variables
A are fuzzy parameters
x are independent variables

Then, linear programming is used to optimize the performance index (J) as shown in equation (3).

$$\text{Minimize } J = \sum_{j=0}^{N}\left(c_j \sum_{i=1}^{M} \left| x_{ij} \right| \right)$$

$$\text{Subject to} \quad \sum_{j=0}^{N} \alpha_j x_{ij} + (1-h)\sum_{j=0}^{N} c_j \left| x_{ij} \right| \geq y_i \qquad (3)$$

$$\sum_{j=0}^{N} \alpha_j x_{ij} - (1-h)\sum_{j=0}^{N} c_j \left| x_{ij} \right| \leq y_i$$

$$c_j \geq 0, \ \alpha_j \in \Re, \ j = 0, 1, 2, ..., N \quad x_{i0} = 1, \ i = 0, 1, 2, ..., M \quad 0 \leq h \leq 1$$

The performance index J is minimized subject to constraints as given in (3) where h as the membership value between 0 and 1 is the threshold level, and determines the degree of fitting with human decision making. C_j is the width or spread, and α_j is the center of triangle base. For fuzzy parameters, x is independent variable, and y is dependent variable where M is the number of data and N is the number of x (independent variable).

Conventional Regression

Conventional regression or statistic regression is regression analysis. This method is based on the analysis of relationship between the variables [7-9]. Simple linear regression could be modeled from dependent variables and independent variables. If there are many independent variables it would be modeled by multiple linear regression as presented by equation (4).

$$Y = \beta_0 + \beta_1X_{1i} + \beta_2X_{2i} + ... + \beta_pX_{pi} + u_i \quad (4)$$

where Y is dependent or response variable
$X_1, X_2, ..., X_p$ are independent variables
$\beta_0, \beta_1, ..., \beta_n$ are constants or the model partial regression coeffients
u_i is random disturbance term
p is number of independent variables
i is number of observations

Analysis of Variance

Single factor analysis of variance (ANOVA) or one way ANOVA is the analysis of differential level of data by considering only one factor. The objective of variance analysis is the test of the several means of populations to compare between the groups of populations. The number of population groups have to be equal or greater than three groups of populations or treatments [9,14,19].

First, determine the null hypothesis H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_n$. Next, write the alternative hypothesis H_1: $\mu_i \neq \mu_j$ at least one pair of means, $i \neq j$, i, j = 1, 2, ..., n. Then, specify the α level : $\alpha = 0.05$. Lastly, calculate the appropriate statistic

where μ_1, μ_2, μ_3 and μ_n are the means of populations
n is number of group of populations
α is the significance level.

Test of Hypothesis

If the significance level in ANOVA results less than 0.05, the null hypothesis will be rejected. Tukey test is chosen for hypothesis test because it is the pairwise comparison method to test significant differences between groups [14,19]. For Tukey test, John W. Tukey [21] proposed the method for multiple comparison procedure in 1953 [22]. The test depends on studentized range distribution to show significant levels of group comparisons [23].

3. CALCULATION

Data Collection

The data of six industries shown in Table 1 were collected in 2008. The 279 surveyed factories can be classified into 6 TSIC groups including 10 subgroups as shown in Table 1. The key factors to be considered in regression models are customer outage cost per interruption, electricity peak demand, distribution transformer kVA, backup generator kVA, number of staff and restarting time of production process.

Variables

The key factors mentioned above are used as the input and output variables of the regression models as follows. One output variable is:

- $y_{TSIC\,i}$: Industrial customer outage cost (Baht/interruption) where i is TSIC Code

and five input variables are:

- x_1 : Distribution transformer rating (kVA)
- x_2 : Number of staffs (person)
- x_3 : Generator rating (kVA)
- x_4 : Peak demand (kW)
- x_5 : Restarting time (minute).

Table 1. TSIC Code and descriptions of surveyed factories

TSIC	Group	Example of sub-group	Freq.	Percent
31	Food, beverage and tobacco	TSIC 31149 : Preserved and canned seafood	55	19.71
32	Textiles and leather products	TSIC 32113 : Spinning cotton and synthetic fibers TSIC 32115 : Weaving and textiles	24	8.60
35	Chemical and rubber	TSIC 35130: Synthetic resin, plastic materials and synthetic fibers, except glass TSIC 35510: Tire rubber and inner tube TSIC 35609: Other plastic products	85	30.47
36	Non metallic mineral	TSIC 36921: Cement	17	6.09
37	Basic material	TSIC 37110: Basic iron and stell TSIC 37120: Casting iron and steel TSIC 37200: Casting of non-ferrous metals	22	7.89
38	Fabricated metals	TSIC 38320: radio, television and communication product TSIC 38439: Automotive and machinery part	76	27.24
Total			279	100

* Reference as TSIC manual guide book on B.E.2549 (2006).

Outage Cost Estimation

The high energy consumption industrial customers of Provincial Electricity Authority (PEA) were surveyed in 2009. The data were collected by questionnaires including some from postal mails and onsite interviews. Then, these data as classified by TSIC groups are performed to develop CDF of each factory. The resulting CDF is the time dependent function where the time duration of interruptions is used in this case. To get the numerical results of CDF, the time duration is used at 36 minutes which comes from the average interruption duration of power distribution systems in the central area of Thailand [17]. Then, those CDF are used to evaluate the outage cost per interruption assigned as the dependent output variable ($y_{TSIC\ i}$) of the regression models against five independent input variables, which are distribution transformer kVA (x_1), number of staffs (x_2), generator kVA (x_3), peak demand (x_4) and restarting time (x_5).

First, modeling the TSIC outage cost by multiple linear regression, a version of statistic regression, is performed. Next, modeling based on fuzzy regression is achieved for a comparison. The latter technique will get an advantage when inadequacy of data, vagueness of data, or distortion of data are presented. Therefore, an algorithm to select the better suitable modeling for each group of industrial customers is required. In this paper, the analysis of variance (ANOVA) and hypothesis testing are employed to analyze the collected data. There are three populations or treatments for ANOVA. The first one is the outage cost estimated by fuzzy regression (FR), and the second is those from statistic regression (SR). The last one is the actual outage cost (AC) of each factory. The conditions of hypothesis testing are given as follows.

The null hypothesis H_0: $\mu_{AC} = \mu_{FR} = \mu_{SR}$

The alternative hypothesis H_1: $\mu_i \neq \mu_j$

for at least one pair of outage cost values where $i \neq j$, and i, j = AC, FR, SR.

To be able to conclude whether the null hypothesis is accepted or rejected, the ANOVA results are considered as follows. If the P-value is higher than α level: $\alpha = 0.05$, the hypothesis is not rejected. Then, the values of Mean Absolute Percentage Error (MAPE) of both FR and SR models are determined. Which model has a lower value of MAPE will be selected. On the contrary, if the hypothesis is rejected. All pairs of outage cost values among FR, SR and AC will be compared by using hypothesis testing with Tukey method to select the better model from the statistic or fuzzy regression method according to the P-value of each pair.

4. RESULTS AND DISCUSSIONS

The histograms and probabilities plots of dependent variable or $y_{TSIC\ i}$ are shown in Fig. 1 to display the line of normal curve of dependent variable for each TSIC group from multiple linear regression.

Table 2. Mean Absolute Percentage Error

TSIC Code	Mean Absolute Percentage Error (MAPE)	
	Fuzzy Regression	Statistic Regression
31	13.07	7.93
32	127.83	14.86
35	24.33	6.61
36	4.30	3.40
37	316.64	524.36
38	34.33	122.45

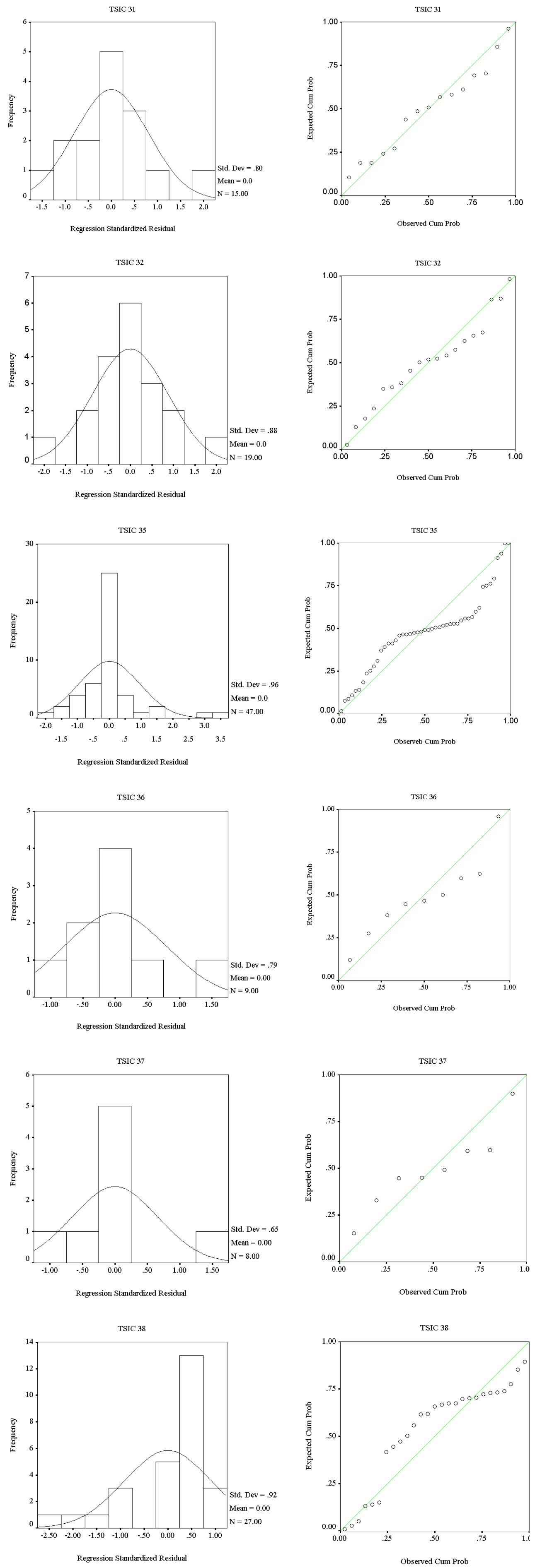

Fig.1. Distribution Histogram and Probability

After the customer outage costs of six TSIC groups are modeled using fuzzy regression and statistic regression, the MAPE resulted from both models are calculated and presented in Table 2. In case of MAPE more than one hundred percent, it can happen in the case that has a large difference between the estimated and observed data. For example, The MAPE in TSIC 37 models is more than one hundred percent. Because of the internal complicated data of industry group which are several sub-categories in this TSIC such as TSIC 37110 as manufacturer of basic iron and steel, TSIC 37120 as casting iron and steel and TSIC 37200 as casting of non-ferrous metals which they are much different in production process.

A summary of statistical values of industrial customer outage costs from AC, FR, and SR for each TSIC group are shown in Table 3, Table 5, Table 7, Table 9, Table 11 and Table 13.

Table 3. Statistical Values of TSIC 31

Groups	*Count*	*Sum*	*Average*	*Variance*
AC	55	12,259,514	222,900	379,370,284,797
FR	55	16,402,109	298,220	221,789,199,603
SR	55	5,303,731	96,431	210,807,647,982

Table 4. ANOVA results of TSIC 31

Source of Variation	*Sum square*	*Degree of freedom*	*Mean square*	*F*	*P-value*
Between Groups	1,143,745,430,683	2	571,872,715,342	2.1129	0.1242
Within Groups	43,846,225,148,604	162	270,655,710,794		
Total	44,989,970,579,287	164			

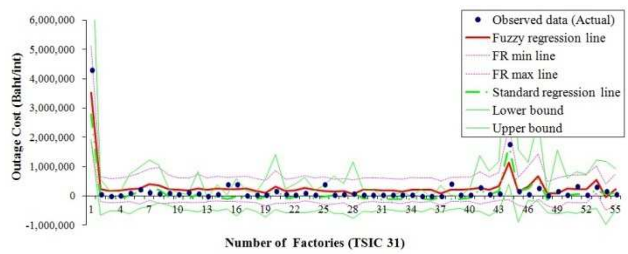

Fig.2. FR and SR Estimation for TSIC 31

According to ANOVA results in Table 4, the P-value is greater than 0.05 so the null hypothesis is accepted for TSIC 31. Then, MAPE of statistic regression in Table 2 is found lower than that of fuzzy regression, so the SR model is selected for TSIC 31 customer group.

Table 5. Statistical Values of TSIC 32

Groups	*Count*	*Sum*	*Average*	*Variance*
AC	24	1,597,643	66,568	14,450,594,062
FR	24	3,761,034	156,710	37,013,285,472
SR	24	1,083,947	45,164	3,426,610,590

Table 6. ANOVA results of TSIC 32

Source of Variation	*Sum square*	*Degree of freedom*	*Mean square*	*F*	*P-value*
Between Groups	168,207 ,583,377	2	84,103 ,791,688	4.5966	0.0134
Within Groups	1,262,481 ,272,873	69	18,296 ,830,042		
Total	1,430,688 ,856,250	71			

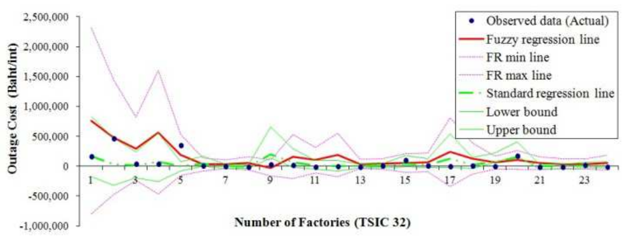

Fig.3. FR and SR Estimation for TSIC 32

According to ANOVA results in Table 6, the P-value is less than 0.05 so the null hypothesis is rejected for TSIC 32. Thus, the hypothesis testing by Tukey method is used to compare the pairs of treatments. The results are shown as follows:

	P-value	Result
AC and FR	0.0613	$\mu_{AC} = \mu_{FR}$
AC and SR	0.8478	$\mu_{AC} = \mu_{SR}$
FR and SR	0.0154	$\mu_{FR} \neq \mu_{SR}$

The significant results of pairs AC and FR, AC and SR have the P-value greater than 0.05 so both statistic and fuzzy regressions are applicable. However, the model resulted from statistic regression is better recommended for TSIC 32 customer group because of a higher of P-value.

Table 7. Statistical Values of TSIC 35

Groups	*Count*	*Sum*	*Average*	*Variance*
AC	85	5,447,058	64,083	19,666,687,385
FR	85	18,682,428	219,793	228,580,296,986
SR	85	7,545,328	88,769	134,504,363,257

Table 8. ANOVA results of TSIC 35

Source of Variation	*Sum square*	*Degree of freedom*	*Mean square*	*F*	*P-value*
Between Groups	1,190,638 ,241,045	2	595,319 ,120,522	4.6661	0.0102
Within Groups	32,151,113 ,200,718	252	127,583 ,782,543		
Total	33,341,751 ,441,762	254			

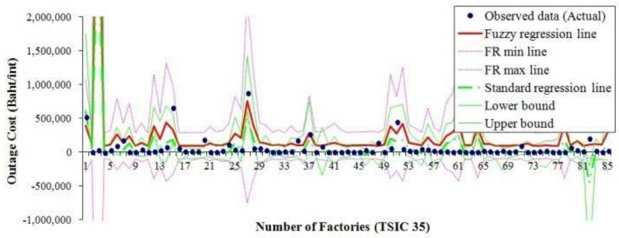

Fig.4. FR and SR Estimation for TSIC 35

According to ANOVA results in Table 8, the P-value is less than 0.05 so the null hypothesis is rejected for TSIC 35. Thus, the hypothesis testing by Tukey method is used to compare the pairs of treatments. The results are shown as follows:

	P-value	Result
AC and FR	0.0134	$\mu_{AC} \neq \mu_{FR}$
AC and SR	0.8942	$\mu_{AC} = \mu_{SR}$
FR and SR	0.0460	$\mu_{FR} \neq \mu_{SR}$

The significant result of pair AC and SR has the P-value greater than 0.05 so the model resulted from statistic regression is selected for TSIC 35 customer group.

Table 9. Statistical Values of TSIC 36

Groups	*Count*	*Sum*	*Average*	*Variance*
AC	17	9,379,412	551,730	1,126,985,471,107
FR	17	9,024,046	530,826	991,094,786,265
SR	17	9,116,261	536,251	1,133,541,051,991

Table 10. ANOVA results of TSIC 36

Source of Variation	*Sum square*	*Degree of freedom*	*Mean square*	*F*	*P-value*
Between Groups	4,000 ,718,421	2	2,000 ,359,210	0.0018	0.9982
Within Groups	52,025,940 ,949,810	48	1,083,873 ,769,788		
Total	52,029,941 ,668,231	50			

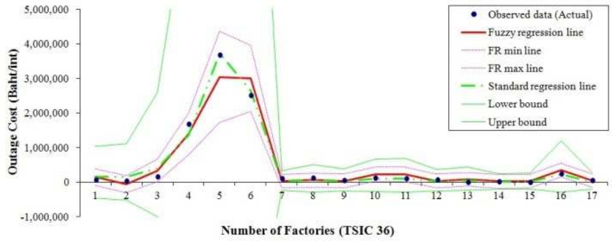

Fig.5. FR and SR Estimation for TSIC 36.

According to ANOVA results in Table 10, the P-value is greater than 0.05 so the null hypothesis is accepted for TSIC 36. Thus, MAPE in Table 2 is considered. It is found that the MAPE of statistic regression is lower than that of fuzzy regression, so the SR model is selected for TSIC 36 customer group.

Table 11. Statistical Values of TSIC 37

Groups	*Count*	*Sum*	*Average*	*Variance*
AC	22	11,727,237	533,056	2,445,715,349,382
FR	22	14,957,018	679,864	1,652,806,416,678
SR	22	-9,587,286	-435,786	959,107,950,584

Table 12. ANOVA results of TSIC 37

Source of Variation	*Sum square*	*Degree of freedom*	*Mean square*	*F*	*P-value*
Between Groups	16,169,139 ,959,289	2	8,084,569 ,979,645	4.7955	0.0115
Within Groups	106,210,224 ,049,522	63	1,685,876 ,572,215		
Total	122,379,364 ,008,811	65			

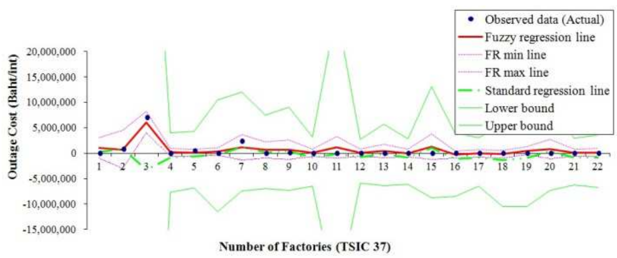

Fig.6. FR and SR Estimation for TSIC 37.

According to ANOVA results in Table 12, the P-value is less than 0.05 so the null hypothesis is rejected for TSIC 37. Thus, the hypothesis testing by Tukey method is used to compare the pairs of treatments. The results are shown as follows:

	P-value	Result
AC and FR	0.9255	$\mu_{AC} = \mu_{FR}$
AC and SR	0.0419	$\mu_{AC} \neq \mu_{SR}$
FR and SR	0.0161	$\mu_{FR} \neq \mu_{SR}$

The significant result of pair AC and FR has the P-value greater than 0.05 so the model resulted from fuzzy regression is selected for TSIC 37 customer group.

Table 13. Statistical Values of TSIC 38

Groups	*Count*	*Sum*	*Average*	*Variance*
AC	76	73,361,764	965,286	16,286,343 ,096,486
FR	76	96,958,818	1,275,774	5,898,747 ,682,194
SR	76	30,497,494	401,283	11,862,743 ,385,749

Table 14. ANOVA results of TSIC 38

Source of Variation	*Sum square*	*Degree of freedom*	*Mean square*	*F*	*P-value*
Between Groups	29,874,008 ,986,360	2	14,937,004 ,493,180	1.3161	0.2702
Within Groups	2,553,587,562 ,332,150	225	11,349,278 ,054,810		
Total	2,583,461,571 ,318,510	227			

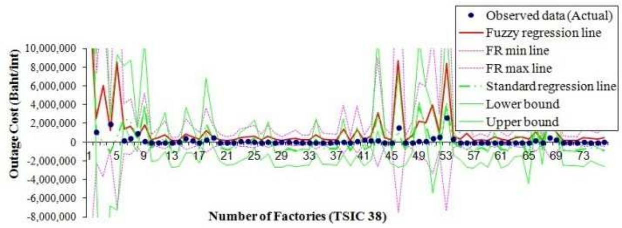

Fig.7. FR and SR Estimation for TSIC 38.

According to ANOVA results in Table 14, the P-value is greater than 0.05 so the null hypothesis is accepted for TSIC 38. Thus, MAPE in Table 2 is considered. It is found that the MAPE of fuzzy regression is lower than that of statistic regression, so the FR model is selected for TSIC 38 customer group.

5. CONCLUSION

This paper presents the modeling of customer outage costs as classified into six TSIC groups. From the above results, the statistic regression method is appropriate for modeling the customer outage costs of TSIC 31, TSIC 35 and TSIC 36. On the other hand, the fuzzy regression method is suitable for modeling the customer outage cost of TSIC 37 and TSIC 38, while that of TSIC 32 may be modeled by fuzzy or statistic regression.

The customer outage cost models of the six TSIC groups result in the outage cost in Baht per interruption. Those models are simple to apply, although some of input variables are not available. Especially for fuzzy regression models, the constant parameter (A_0) alone can be used to assess the outage costs with an acceptable error. Also, outage cost models classified by TSIC will offer a higher accuracy that is very helpful to electric utilities for making a decision on maintenance planning and reliability improvement projects.

ACKNOWLEDGMENT

The authors would like to thank all industrial customers in the central region of Thailand, who responded to postal questionnaires and onsite interviews. We also would like to acknowledge the financial support of Rajamangala University of Technology Suvarnabhumi (RMUTSB), Kasetsart University (KU), and Provincial Electricity Authority of Thailand (PEA).

REFERENCES

[1] Dom, R. M., Kareem, S. A., Rasak, I. A. and Abidin, B. 2008, A learning system prediction method using fuzzy regression, *In Proceedings of the International MultiConference of Engineers and Computer Scientists*. Hongkong, PR China, 19-21 March.

[2] Nasiri, M., Shanbeh, M. and Tavanai, H. 2005, Comparison of statistic regression, fuzzy regression and artificial Neural Network Modeling Methodologies in Polyester Dyeing, *In Proceedings of the International Conference on Computational Intelligence for Modelling,Control and Automation, and International Conference.* Vienna, Austria, 28-30 November.

[3] Ketkaew, C., Noohawm, O. and Rerkpreedapong, D. 2013, Rice mill industry maintenance planning using Reliability Centered Maintenance (RCM), *In Proceedings of the IASTED International Conference Power and Energy Systems.* Phuket, Thailand, 10-12 April.

[4] Billinton, R., 1996. Reliability Evaluation of Power Systems. New York and London: Plenum.

[5] Hokierti, J., 2006. Power Distribution System Reliability and Maintenance. Thailand: Kasetsart University.

[6] Brown, R. E., 2009. Electric Power Distribution reliability. Second Edition, USA:CRC Press.

[7] Neter, J., Wasserman, W. and Kutner, M. H. 1989. Applied Linear Regression Models. USA: IRWIN.

[8] Chatterjee, S. and Price, B. 1991. Regression Analysis by Example. USA: Wiley-Intersceince.

[9] Wanichbancha, K., 2001. Principle of Statistic, Thailand: Chulalongkorn University.

[10] Smith, A. M. and Hinchcliffe, G. R. 2003. RCM-Gateway to World Class Maintenance. USA: Elsevier Butterworth-Heinemann.

[11] Zadeh, L. A. (1978), Fuzzy sets as a basic for a theory of possibility, Fuzzy Sets and Systems, vol. 1, pp. 3 - 28.

[12] Tanaka, H., Uejima, S. and Asai, K. (1982), Linear regression analysis with fuzzy model, IEEE Transactions on System, Man and Cybernetics, vol. 12 n. 6, November/December, pp. 903 – 907.

[13] Ip, K.W., Kwong, C.K. and Wong, Y.W. (2003), Fuzzy regression approach to modeling transfer moulding for microchip encapsulation, Journal of Materials Processing Technology, vol. 140, pp. 147 – 151.

[14] Azadeh, A., Khakestani, M. and Saberi, M. (2009), A flexible fuzzy regression algorithm for forecasting oil consumption. Energy Policy, vol. 37, pp. 5567-5579.

[15] Kim, K. J., Moskowitz, H. and Koksalan, M. (1996), Fuzzy versus statistical linear regression, European Journal of Operational Research, vol.92 , pp. 417 – 434.

[16] Nazarko, J. and Zalewski, W. (1996), The fuzzy regression approach to peak load estimatation in power distribution systems, IEEE Transaction on Power System, vol. 14 n. 3, August 1999, pp. 809-814.

[17] IEEE Committee Report (1998), Report on reliability survey of industrial plants, Part II: Cost of power outages, plant restart time, critical service loss duration time, and type of loads lost versus time of power outages (IEEE Std 493-1997). Appendix A, pp. 234-241.

[18] Sullivan, M. and Keane, D. (1995). Outage cost estimation guidebook. Electric Power Research Institute, TR-106082, December, from the World Wide Web: http://www.epriweb.com/public/TR-106082.pdf.

[19] Elvers, G. C. (2013). Using SPSS for one way analysis of variance. University of dayton, from the World Wide Web: http://academic.udayton.edu/ gregelvers/psy216/spss/1wayanova.htm

[20] Shapiro, A. F. (2005). Fuzzy regression models. Pennstate University, from the World Wide Web: http://www.soa.org/library/research/actuarial-research.../arch06v40n1-ii.pdf

[21] Brillinger, D. (2002), John W.Tukey: His Life and Professional Contributions, The Annals of Statistics, Vol.30, pp.1535-1575.

[22] Montgomery, D., 2001. Design and Analysis of Experiment. fifth edition, John Wiley & Sons.

[23] George, D. and Mallery, P., 1999. SPSS for Windows Step by Step. Allyn and Bacon.