Australian Mathematical Society
Australian and New Zealand Industrial and Applied Mathematics

Proceedings of the 49th ANZIAM Conference

Newcastle, New South Wales
3–7 February 2013

3 Conference Details and History

3.1 Organising Committee

- Jon Borwein (University of Newcastle) — Meeting director
- Juliane Turner (University of Newcastle) – General administration
- David Allingham (University of Newcastle) – Technical administration
- Roslyn Hickson (University of Newcastle) – Conference advisor
- Natashia Boland (University of Newcastle) – General assistance
- Mike Meylan (University of Newcastle) – Programme organiser
- Bishnu Lamichhane (University of Newcastle) – Programme organiser
- Masoud Talebian (University of Newcastle) – Treasurer
- Philip Broadbridge (La Trobe University) – SIAM representative

3.2 Invited Speakers Committee

- Larry Forbes (University of Tasmania) — Chairperson
- Jon Borwein (University of Newcastle)
- Carlo Laing (Massey University)
- Natashia Boland (University of Newcastle)
- Geoffry Mercer (Australian National University)
- Irene Hudson (University of Newcastle)
- Kerry Landman (University of Melbourne)
- Kerrie Mengersen (Queensland Institute of Technology)
- Mary Myerscough (University of Sydney)
- John Sader (University of Melbourne)

3.3 Invited Speakers

- Heinz Bauschke (The University of British Columbia, Okanagan) [Page 35]
- Keith Devlin (Stanford University) [Page 35]
- Matthias Ehrgott (University of Auckland) [Page 36]
  CARMA/AMSI Speaker for Mathematics of Planet Earth
- G. Bard Ermentrout (University of Pittsburgh) [Page 36]
- Chris Glasbey (BIOSS Edinburgh) [Page 37]
- Graeme Hocking (Murdoch University) [Page 37]
- Hanna Kokko (The Australian National University) [Page 38]
- Robert McKibbin (Massey University) [Page 38]
  ANZIAM Medalist
- Matthew Simpson (Queensland University of Technology) [Page 39]
  Michell Medalist

The Linear Regression and Fuzzy Linear Regression based Medical Service Value Models for Informal Workers in Thailand

Wiyada Kumam

Department of Mathematics, Faculty of Science,

King Mongkut's University of Technology Thonburi (KMUTT),

Bangkok, 10140, Thailand

Tel: +662-470-8822 E-mail address: wiyada.kum@mail.rmutt.ac.th

Adisak Pongpullponsak (Corresponding author)

Department of Mathematics, Faculty of Science,

King Mongkut's University of Technology Thonburi (KMUTT),

Bangkok, 10140, Thailand

Tel: +662-470-8822 E-mail address: iadinsak@kmutt.ac.th

Abstract

The purpose of this research is to develop an estimation model for non-surgical medical service value of informal workers for the social security system in Thailand. In the study, the data of workers in year 2010 provided by the Social Security Office was analyzed and used to create the medical service value model. Two methodologies, linear regression and fuzzy linear regression have been chosen to develop the model, and then the estimates obtained from each model are compared to the actual costs from hospitals. The results demonstrated that the medical service value model established from fuzzy linear regression method gave the closest estimates to the real expenses.

Keywords: informal workers, linear regression, fuzzy linear regression, medical service value

1. Introduction

The social security system is established to insure employers in case of illness, retiring and disability from work. The insurance is commonly involving with 3 sides: government, employer and employee. Fund management will be collected through subvention which is mandatory for persons who have income. The social security system in Thailand has initiated since 1952 administrating employees' compensation in case of illness or accident due to work. Until 1990, the House of Representatives has approved and confirmed the Social Security Act draft resulting in the Social Security Act A.D. 1992. It defines the mutual fund to carry employment insurance for illness, disability, death which does not be work-related, parturition, child allowance, and unemployment and pension benefits. Social security outset in the project usually provides protection limited only for some certain employments such as manual labors in establishments with more than 20 employees. In the present days, the system has been expanded into various public sectors in order to cover benefits of the social security to workers who receive regular incomes and members in family during working age as well as out of work, disability or elderly. The project is planned to primarily carry out as mandatory until it has been successfully settled down and the performance reaches acceptable level, the system will be extended to form a voluntary social security scheme, where the weaver freelance can be included in this program.

In Thailand, informal workers such as workers in fishing, forestry and agricultural services, part-time workers, sweatshop, hawkers etc. are as vital to the national economy as those labors whose are covered under the social security program. Nevertheless, the social security or any protection coverage has yet not been provided for these informal workers which in turn the burden of welfare of these workers falls to the government in different ways. For example, informal workers and members in families receive medical and health treatment under the universal coverage health insurance system, although they are able to pay social security contributions to get medical benefits from the social security system. For this reason, the government of Thailand has attempted to extend the social security program to informal sector workers covering 4 benefits, (1) costs of medical treatment, (2) compensation for unemployment, (3) funeral expenses and (4) financial aid. The study has started by a study of Pongpulponsak et al. (2010) who defined target groups that should be included in the program. The researchers have subsequently determined a model for estimating costs of medical treatment without surgery for informal workers.

Several studies have reported for an appropriate medical benefit for informal sector workers. For instance, Baker and Krueger (1995) established a model for estimating medical and health compensation for insured persons under the social security system. Ding and Zhu (2009) employed controlling of medical service value in revolution of health insurance system in China. In 2011, Galbraith and Stone proposed the abuse of regression in the National Health Service allocation formulae which is in response to the Department of Health's 2007 resource allocation research paper. Kazumitsu and Nawata (2008, 2009) analyzed hip fracture treatments in Japan by using discrete-type proportional hazard and probit models. One year late, they developed the discrete-type proportional hazard model for estimating duration of hospital stay for cataract patients. From their reports, it is found that duration of hospital stay should be taken into account for medical service value model.

We have previously developed a methodology, the linear regression based medical service value without surgery model, used for estimating medical costs of informal workers for the social security system in Thailand. Since the information used in the previous study are high variation and ambiguous, this study is aimed to analyze the data as in the previous study by fuzzy clustering method before using to establish a new method. The efficiency of a newly constructed model will then be compared with the previous model in order to select the most appropriate estimation method. Fuzzy clustering method is an effective methodology that has been popularly used to deal with fuzzy or ambiguous data. Chen et al. (2011) used fuzzy clustering method in clustering the data of flood damage into dependent variables and independent variables, where they were subsequently analyzed to construct a logistic regression based risk analysis model. Peduzzi et al. (1996) conducted a simulation study of the number of events per variable in logistic regression analysis using fuzzy clustering method for data allocation. McLay et al. (2012) used logistic regression method to analyze the volume and nature of emergency medical calls during severe weather events.

In estimation of medical service costs, it is found that if the ambiguous data has been used in the study, this might yield inaccurate results. To avoid such a problem, several researchers have adapted the principle of fuzzy for data analysis in their studies. Ho (2011) developed a method for optimal evaluation of infectious medical waste disposal companies using the fuzzy analytic hierarchy process. Bolotin (2005) studied fuzzification of linear regression models with indicator variables in medical decision making. Stefan (2010) defined tree types of fuzzy predictions of the observed variable in the classical regression model where unknown parameters and observations are crisp. Therefore, the aim of this study is to develop a non-surgical medical service value estimation model of informal workers for the social security, Thailand using the fuzzy logistic regression analysis method. Subsequently, the estimates obtained from using the newly constructed model will be compared with the results from our previous model.

2. Methods

2.1 Fuzzy set

From Lee (2009), an universal set X is defined in the universe of discourse and it includes all possible element relate with the given problem. If we define a set A in the universal set X, we see the following relationships $A \subseteq X$. In the case, we say a set A is included in the universal set X. If A is not included in X, this relation is represented as follows $A \not\subseteq X$. If an element x is included in the set A, this element is called as a member of the set A, this element is called as a member of the set and the following notation is used $x \in A$. If an element x is not included in the set A, we use the following notation $x \notin A$. If we use membership function (characteristic function or discrimination function), we can represent whether an element x is involved in a set A or not.

Definition 1 (membership function) For a set A, we define a membership function μ_A such as

$$\mu_A(x) = \begin{cases} 1, & if\ and\ only\ if\ \ x \in A \\ 0, & if\ and\ only\ if\ \ x \notin A \end{cases}$$

We can say that the function μA maps the element in the universal set X to the set {0, 1}.

$$\mu_A\colon \mathrm{X} \rightarrow \{0,\ 1\}.$$

membership function μA in crisp set maps whole members in universal set X to set {0, 1}

$$\mu_A\colon \mathrm{X} \rightarrow \{0,\ 1\}.$$

Definition 2 (membership function of fuzzy set) In fuzzy sets, each elements is mapped to [0, 1] by membership function.

$$\mu_A\colon \mathrm{X} \rightarrow \{0,\ 1\}.$$

where [0, 1] mean real numbers between 0 and 1 (including 0, 1).

Definition 3 (α-cut set) The α-cut set A_α is made up of members whose membership is not less than α.

$$A_\alpha = \{x \in X | \mu_A(x) \geq \alpha\}$$

note that α is arbitrary. This α − cut set is a crisp set.

Definition 4 (Fuzzy number) If a fuzzy set is convex and normalized, and its membership function is defined in $\Re$ and piecewise continuous, it is called as "fuzzy number". So fuzzy number (fuzzy set) represents a real number interval whose boundary is fuzzy (figure 1)

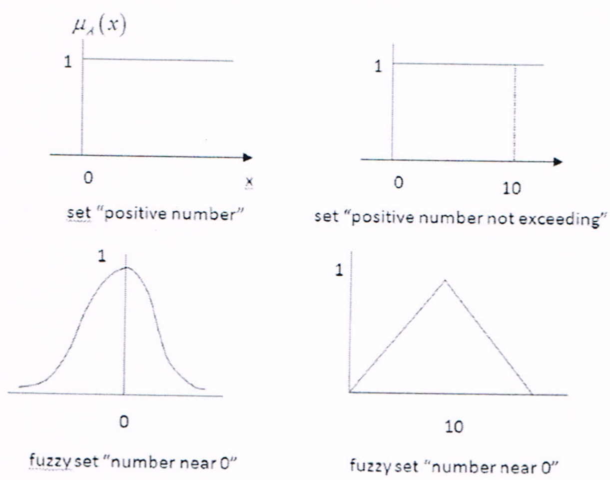

Figure 1 Sets denoting intervals and fuzzy numbers.

Definition 5 (Triangular Fuzzy number) It is fuzzy number represented with three points as follows:

$A = (a_1, a_2, a_3)$, this representation is interpreted as membership functions (figure 2).

$$\mu_A(x) = \begin{cases} 0 & , \quad x < a_1 \\ \frac{x-a_1}{a_2-a_1} & , \quad a_1 \le x \le a_2 \\ \frac{a_3-x}{a_3-a_2} & , \quad a_1 \le x \le a_2 \\ 0 & , \quad x > a_3 \end{cases}$$

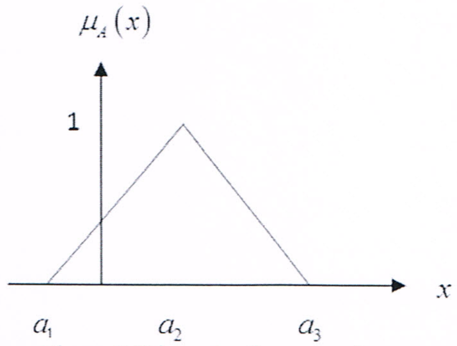

Figure 2 Triangular fuzzy number

Definition 2.1.6. (Trapezoidal Fuzzy number) We can define trapezoidal fuzzy number A as

$A = (a_1, a_2, a_3, a_4)$, the membership functions of this fuzzy number will be interpreted as follows(figure 3).

$$\mu_A(x) = \begin{cases} 0 & , \quad x < a_1 \\ \frac{x-a_1}{a_2-a_1} & , \quad a_1 \le x \le a_2 \\ 1 & , \quad a_2 \le x \le a_3 \\ \frac{a_4-x}{a_4-a_3} & , \quad a_1 \le x \le a_2 \\ 0 & , \quad x > a_4 \end{cases}$$

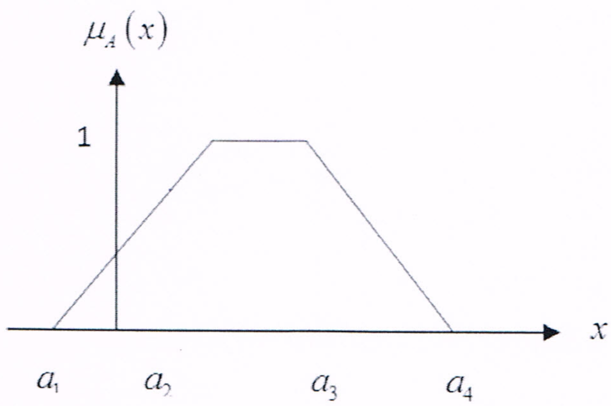

Figure 3 Trapezoidal fuzzy number $A = (a_1, a_2, a_3, a_4)$

2.2 The medical service value model

The data used in the study, is obtained from the surveys of informal workers of the social security, Thailand in 2010, included sex, age, weight, height, education, occupation, number of family members, income, number of medical visiting, length of hospital stay, costs of medical care.

Estimation of non-surgical medical expenses of patients can be done from analysis of length of illness and hospital stay, which is discrete random variable, where it equates to 1, 2, 3 etc. Nawata et al. (2008, 2009) analyzed the length of stay in hospital, which is depending on the severity of the case, using the discrete-type proportional hazard model. Thus, let the leaving rate, designated as $h_i(t)$, be a conditional probability that the i^{th} patient staying in a hospital on the t^{th} day will leave the hospital on that day. Therefore, the probability of the i patient to leave hospital on the t day is a function of $h_i(t)$ and give by

$$p_i(t) = \begin{cases} h_i(t), & t = 1 \\ [\prod_{s=1}^{t-1}\{1 - h_i(s)\}]h_i(t), & t \geq 2, i = 1, 2, \ldots, n \end{cases} \quad (1)$$

where n is number of patients, s is number of days staying in the hospital an $s = 1, 2, \ldots, t-1$. According to health benefits of the social security system, Thailand, the patient can claim for reimbursement at actual payment but not exceed 12,000 baht per days and not more than twice a year. Given that there is no limitation of the length of hospital staying for insured person. Let T is the maximum number of days that patient could stay in a hospital, and let $p_i(T+1)$ is the probability that the patient i^{th} will stay in hospital more than T days. Then,

$$p_i(T+1) = \prod_{s=1}^{T}\{1 - h_i(s)\}, \qquad t \geq 2, i = 1, 2, \ldots, n \ . \quad (2)$$

Let v_i is random variable of medical expenses of patients i^{th}. From the continuous proportional hazard models by Nawata et al.(2008, 2009), we obtain an equation of risk incidence for various characteristics of patients as below;

$$h_i(t) = d_t \exp(v_i'\beta) \qquad t = 1, 2, 3, \ldots, T, \quad (3)$$

when d_t is the rate of patient staying in a hospital on day t^{th}, and β is regression coefficients of patient condition.

2.3 Analysis of Regression

The aim of regression analysis is to estimate the parameters on the basis of empirical data. The linear form of regression analysis can be written as

$$y = \beta_0 + \beta_i x_i \qquad for\ i=1, 2, \ldots, n, \quad (4)$$

where y is an output variable, x_i is input variable, and β_i is parameter of the most frequent mathematical form use in regression analysis. Then, the equation of linear regression will be

$$y_j = \beta_0 + \beta_{ij}x_{ij} + \varepsilon_{ij}; \qquad i = 1, 2, \ldots, n \quad and\ j = 1, 2, \ldots, m. \quad (5)$$

If m is sample size, n is number of variables and ε_{ij} is error of the equation.

In 2004, Dennis and Wage (2000) studied health insurance and pension plans to investigate the relationship between employee compensation and small business owner income by using regression analysis. From the equation 3 of Pongpulponsak's report (2010), estimation of $v_i'\beta$ using regression method and expected value of non-surgical medical service of patients when they are not admitted to a hospital for treatment can be expressed by

$$E(C_{IN}(t)) = \int_0^t v_i h_i(t) dv_i = d_i \int_0^t v_i \exp(v_i'\beta) dv_i \, . \qquad (6)$$

2.4 Analysis of fuzzy linear regression

The fuzzy regression model is presented by

$$\tilde{y}_j = \beta_0 + \beta_{ij}\tilde{x}_{ij} + \varepsilon_{ij}; \qquad i = 1, 2, \dots, n \quad and \ j = 1, 2, \dots, m. \qquad (7)$$

where β_{ij} for all i = 1,2,3,...,n and j=1, 2, ..., m are the regression coefficients, $\tilde{x}_{ij}$ for all i = 1,2,3,...,n and j=1, 2, ..., m are fuzzy input independent variables and $\tilde{y}$ is a fuzzy output dependent variable. From the study of Stefan in year 2010, the fuzzy regression model was developed from observations and unknown regression parameters were fuzzy numbers. Hence, the predictions of the observed variables would be in fuzzy numbers too.

3. Results

3.1 Model 1 by regression method

From Pongpullponsak et al. (2010), $v_i'\beta$ can be estimated by regression method.

$$v_i'\beta = 3.812 - 0.209x_1 - 0.374x_2 - 0.0000004605x_3 - 0.007x_4 \ , (8)$$

where x_1 is number of family members, x_2 is sex, x_3 is income, x_4 is weight.

Substituting (8) into (6), where d1, d2, d3 are equal to 0.11, 0.08 and 0.19 day/person/year, so if a patient stays in a hospital for 1, 2 and 3 days, the medical costs will be 1405, 2763 and 3410 baht, respectively.

3.2 Model 2 by fuzzy linear regression method

The next step is to establish the medical service value model by using the data from a questionnaire surveying the medical care of informal workers. We hypothesize that there are several factors related to medical care of informal workers including age, weight, height, education, occupation, number of family members, income of the family, number of receiving medical examination and number of days staying in a hospital. All these factors have been used in finding of $v_i'\beta$ by fuzzy linear regression. So we obtain

$$\tilde{y} = (v'\beta) = 1.223 - 0.003\tilde{x}_1 - 0.00000046699\tilde{x}_2 \qquad (9)$$

where x_1 is age, x_2 is income. Using Matlab, the medical services value model can be created based on fuzzy rule, "if age and income to cost", as shown in Figure 1.

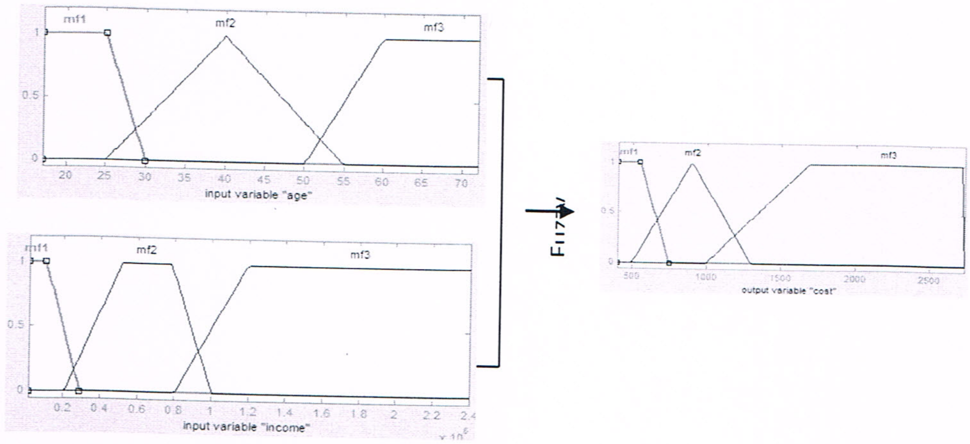

Figure 1 The Fuzzy rule of this study

Substituting the values from result of fuzzy rule into equation (6), yields the expected value of medical expenses for patient who receives treatment without surgery at registered hospital per day, where d_1, d_2, d_3 are equals to 0.11, 0.08 and 0.19 day/person/year, respectively. Hence, at significant level ($\alpha = 0.01$) when a patient stays in a hospital for 1 day the estimate of medical costs will be 1021 to 1023 baht. If a patient stays in a hospital for 2 and 3 days, the estimate of medical cost will be 3342 to 3345 baht and 7232 to 7259 baht, respectively.

4. Conclusion and discussion

In this study, we develop the medical service value model for estimating non-surgical medical expenses, including admission to a hospital and medical treatment, of informal workers in Thailand. Using the regression method (model 1) where d_1, d_2, d_3 are equal to 0.11, 0.08, 0.19 day/person/year, respectively, the expected medical costs for the length of hospital stay at 1, 2 and 3 days are 1405, 2763 and 3410 baht, respectively. In case of the fuzzy linear regression method (model 2), it started from

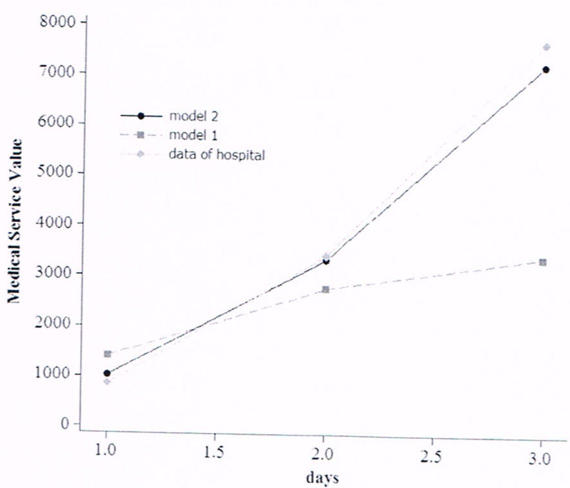

Figure 2 Comparisons of the non-surgical medical costs of informal workers between the actual expenses from data hospitals in Thailand, the estimates from the regression model (model 1) and from the fuzzy linear regression model (model 2)

data clustering using the fuzzy clustering method on the basis of fuzzy equivalent relation, which results in 3 data groups. Each data group is then used in establishing a linear regression model. The results obtained from the fuzzy linear regression model or model 2 is then used to estimate the medical value of informal workers in case of treatment without surgery. At $\alpha = 0.01$ where d_1, d_2, d_3 are equal to 0.11, 0.08 and 0.19 day/person/year, respectively, the medical costs of the length of hospital stay at 1, 2 and 3 days are estimated at 1021 to 1023, 3342 to 3345 and 7232 to 7259 baht, respectively. From figure 2, comparisons of the estimated medical values from each method with the actual expenses from data hospitals in Thailand revealed that using the model 2, the fuzzy linear regression method, gives the medical values closer to the actual costs than those of the model 1. This is because the model 2 is established from the data that have been dealt with fuzzy clustering method to solve the problem of data ambiguity. This leads into no outliers in the data, there by the model yields more accurate estimated values. For the future work, the overview of the medical service value will be considered for the most suitable medical service value model including expenses of medical treatment with surgery. This information will be contributed in setting up the social insurance of informal workers by the Social Security Office in Thailand.

5. Acknowledgements

The first author would like to thank the Rajamangala University of Technology Thanyaburi (RMUTT) for financial support for the Ph.D. program at King Mongkut's University of Technology Thonburi (KMUTT).

6. References

Baker, L. C., & Krueger A. B. (1995), Medical costs in workers' compensation insurance, *Journal of Health Economics*. 14, 531-*549.*

Bolotin, A. (2005). Proceedings of the 2005 International Conference on Computational Intelligence for Modeling, Control and Automation, and International Conference Intelligent Agents. *Web Technologies and Internet Commerce (CIMCA-IAWTIC'05)*

Chen, J., Zhao, S., & Wang, H. (2011), Risk Analysis of Flood Disaster Based on Fuzzy Clustering. *Energy Procedia.* 5, 1915-1919

Dennis, W. J. & Wages, Jr. (2000), Health Insurance and Pension Plans: The Relationship Between Employee

Ho, C. C. (2011). Optimal evaluation of infectious medical waste disposal companies using the fuzzy analytic hierarchy process. Waste Management. 1553-1559

Galbraith, J., & Stone, M. (2011). The abuse of regression in the National Health Service allocation formulae: response to the Department of Health's 2007 'resource allocation research paper. *Journal of the Royal Statistical Society*. 517-528

Jihong, D., & Minglai, Z. (2009). A theoretical investigation of the reformed public health insurance in urban China. *Higher Education Press and Springer-Verlag.* 4, 1-29

Laura, A. M., Edward, L. B., & Brooks, J. P. (2012), Analyzing the volume and nature of emergency medical calls during severe weather events using regression methodologies. Socio-Economic Planning Sciences, 46, 55-66.

Lee, K. H., First Course on Fuzzy Theory and Applications}, Springer Berlin Heidelberg NewYork., (2009).

MINITAB 16 Order *Number 100004968850, Single License, MINITAB Thailand. , February 02, 2010.*

Nawata, K. et al. (2008). An analysis of hip fracture treatments in Japan by the discrete-type proportional hazard and ordered probit models. *Mathematics and Computers in Simulation*, 78, 303-312

Nawata, K. et al. (2009). An analysis of the length of hospital stay for cataract patients in Japan using the discrete-type proportional hazard model. *Mathematics and Computers in Simulation*, 79, 2889-2896.

Peduzzi, P., et al. (1996), A Simulation Study of the Number of Events per Variable in Logistic Regression Analysis. *J Clin Epidemiol,* 49,12, 1373-1379.

Pongpullponsak, A. et al. (2010). Research to define the target group of the Social Security Office, *Full report* Compensation and Small Business Owner Income. *Small Business Economics,* 15, *247-263.*

Stefan, V. (2010), Fuzzy Predictions in Regression Models. *Journal of Applied Mathematics*, 3, 245-252.

The Math WorksTM, MATLAB, 7.6.0(R2009a), *License Number 350306, February 12, 2009.*

Yang, M.S. *(*1993), A Survey of Fuzzy Clustering. *Math. Comput. Modelling, 18, 11, 1–16*