Int. J. Advance Soft Compu. Appl, Vol. 8, No. 3, December 2016
ISSN 2074-8523

Predictive Modelling for Motor Insurance Claims Using Artificial Neural Networks

Zuriahati Mohd Yunos[1], Aida Ali[2], Siti Mariyam Shamsyuddin[2],
Noriszura Ismail[3], Roselina Salleh@Sallehuddin[1]

[1]Faculty of Computing
Universiti Teknologi Malaysia, Johor, Malaysia
[2]UTM Big Data Centre,
Universiti Teknologi Malaysia, Johor Malaysia
[3]Pusat Pengajian Sains Matematik,
Universiti Kebangsaan Malaysia, Selangor, Malaysia
e-mail: zuriahati@utm.my, mariyam@utm.my, aida@utm.my,
roselina@utm.my, ni@ukm.my

Abstract

The expected claim frequency and the expected claim severity are used in predictive modelling for motor insurance claims. There are two category of claims were considered, namely, third party property damage (TPPD) and own damage (OD). Data sets from the year 2001 to 2003 are used to develop the predictive model. The main issues in modelling the motor insurance claims are related to the nature of insurance data, such as huge information, uncertainty, imprecise and incomplete information; and classical statistical techniques which cannot handle the extreme value in the insurance data. This paper proposes the back propagation neural network (BPNN) model as a tool to model the problem. A detailed explanation of how the BPNN model solves the issues is provided.

Keywords: *predictive modelling, claim frequency, claim severity, back propagation neural network.*

1 Introduction

Predictive modelling is a process that involves problem identification, analysis of data, model development, prediction and validation [25]. Predictive modelling in the insurance industry helps actuaries and other insurance analysis employing predictive models to enhance business operations that were previously using human expertise. Historically, actuaries performed their duties using pencil and paper before the advent of computers. Today, more advanced computing tools are available [16]. Predictive modelling has provided a set of instruments to the insurance companies for a variety of intentions from pricing to underwriting and claim handling [5]. Moreover, the influences of predictive modelling are also dependent on the quality of the data used to generate the model. Insurance is a unique type of agreement between the insurer or insurance company and the insured or client in which the insurers permit that upon the occurrence of specific events, whether to make a payment to clients or cover the specific costs. There are two types of insurance that are life insurance and general insurance, and motor insurance is under general insurance. This research focuses to develop a predictive model for motor insurance claim by estimating the two important components, namely, claim frequency and claim severity [9], [30], [23], [21], [11], [14]. Claim frequency is defined as the number of claims per exposure unit, whereas claim severity is defined as the average claim cost per claim [14]. The modelling of claim frequency and claim severity needed an information of exposure, number of claims and the amount of the claim (cost). The expected of claim frequency and claim severity can be calculated through a process of identifying grouping risk, which having the same characteristics is also known as risk classification. Thus, this study investigates how capable artificial neural networks (ANN) is as a potential technique to be applied in modelling the motor insurance claims problem. The utility of ANNs has been demonstrated in the insurance industry such as [8], [16] and also was found in [4], [13], [19], [31], [32]. In this paper, we intend to highlight the importance of the ANN approach in modelling claim frequency and claim severity. The remainder of this paper is organized as follows: Section 2 presents related works. Section 3 discusses the BPNN model development is described. In Section 4 presents the experimental results and analysis. Finally, the conclusion is provided in Section 5.

2 Related Works

Two important issues that were highlighted in motor insurance claims are the data and techniques, and these issues are interrelated among others. The first issue is related to the characteristics of insurance data which contain huge information or large number of variables, uncertainty, information is very noisy and incomplete information and this was agreed by [11], [15], [12], [26], [10]. Hence, the development of the predictive model in motor insurance claims requires sufficient data to acquire an accurate model. Besides that, the predictive models are used for risk classification and to determine the premium [9], [30]. The development of predictive models in motor insurance claims is known as risk classification. Risk classification is defined as the formulation of different premiums for the same coverage based on group characteristics. The group characteristic refer to the data or variables and is called rating variables or rating factors. The problem is how to determine and choose the significant rating factors and rating classes in risk classification [22], [17] and this is related to the data issue and the techniques. It has been acknowledged by [22], [17]. Another study also found a premium can be determined by a number of factors such as the vehicle cubic capacity, the geographic zone where the insured lives, the aged of the insured [29], [23], [21], [11. Another problem is the existing of extreme values in the data which cannot be ignored or dealt as outliers [18], [1]. Extreme values or outliers can attract the insurers because it can help the insurers to determine the amount of the highest demands and protect themselves in the future. The second issue is related to the complexity of statistical analysis that has become more apparent. Due to this, actuaries had to solve the problem of finding a model that can explain realistically the event of risk [6], [9] and a model that able to handle complex problems in exploiting varying information [28].

The suggestion of ANN approach to motor insurance claim is the use of past experience to train the network to provide more consistent and reliable evaluations on claim frequency and claim severity. ANN, have been massive and it can process a large volume of uncertainty, inaccurate and lacking data, and also to look for estimate [26], [5], [4], [13], [16]. Furthermore, ANN has been applied in the fields of finance and economics, and many businesses use the ANN in their decision support systems. The input of training data and output for claim frequency and claim severity is shown in Table 1.1 and 1.2. For claim frequency, each data is used to determine the number of claims made by the insured (clients) and as claim severity the data are used to compute the amount claimed by the insured or amount paid by the insurer to insured. Following are the abilities of using ANN which can improve the prediction with promising results.

i. ANN is capable to handle a nonlinear and easily learn rich representations through the mapping of inputs to outputs.

ii. ANN are flexible in how they can be used such as pattern recognition, time series, and image processing and so on.

iii. ANN is more successful in terms of speed, simplicity and capacity compared to traditional statistical models.

iv. The performance of the ANN can be improved by tuning the parameters through trial-and-error.

v. MATLAB provides ANN toolbox which can easily be used for training and testing.

The implementation will be discussed in the next section.

Table 1.1**:** Input and output variables for TPPD, TPBI and Theft claim

Notation	**Claim Frequency**		**Claim Severity**	
	Input nodes	**Output nodes**	**Input nodes**	**Output nodes**
F1	Coverage	Claim frequency	Coverage	Claim severity
F2	Vehicle made		Vehicle made	
F3	Vehicle cc		Vehicle cc	
F4	Vehicle year		Vehicle year	
F5	Location		Location	
F6	Exposure		-	
F7	-		Number of claim	

Table 1.2: Input and output variables for OD claim

Notation	**Claim Frequency**		**Claim Severity**	
	Input nodes	**Output nodes**	**Input nodes**	**Output nodes**
F1	Coverage	Claim frequency	Coverage	Claim severity
F2	Vehicle made		Vehicle made	
F3	Vehicle cc		Vehicle cc	
F4	Vehicle year		Vehicle year	
F5	Location		Location	
F6	Exposure		-	
F7	-		Number of claim	

3 Model Development Using ANN

BPNN is one of the algorithm in ANN with a three-layer network structure of a back propagation (BP) learning algorithm is choose to model the motor insurance claims. There are two main steps involved; the ANN architecture and the BP algorithm. The essential steps for designing BPNN model are summarized in Table 1.3. In particular, step 1 to step 4 are carried out on data pre-processing, where the raw data is scaled and normalized to an appropriate format to facilitate the predicting process. Step 5, which is the step that designs the ANN model, involves the determination of the following variables:

i. number of input nodes

ii. number of hidden layers and hidden nodes

iii. number of output nodes

iv. activation function

Table 1.3: Steps in designing a BPNN model

Step 1	Variable selection
Step 2	Data collection
Step 3	Data normalization
Step 4	Data division: training and testing
Step 5	Determine the : - number of input nodes - number of hidden layers - number of hidden nodes - number of output nodes - activation function
Step 6	ANN training by applying BP algorithm : - set the learning rate and momentum - set the number of training iterations
Step 7	Model evaluation

In Step 2, data normalization is implemented to smooth out the data, resulting in better data generalization and improved performance. The normalization function is based on the maximum and minimum values that could be set and is suggested by [20]. The normalization formula used is given in Equation (1) :

$$X_{new} = \frac{X_t - X_{\min}}{X_{\max} - X_{\min}}(D_{\max} - D_{\min}) + D_{\min} \qquad (1)$$

where X_t is the value will be normalized, $X_{\min}$ is the minimum value of the statistic variable, and $X_{\max}$ is the maximum value of the statistic variable. $D_{\max}$ and $D_{\min}$ are the maximum and the minimum values needed for normalization. The values of $D_{\max}$ and are set to 0.95 and 0.0, respectively and these values are used in this study. The normalization equation is selected due to the range of the activation function are between 0 to 1, and utilized in the BPNN. Hence, the outputs of a set date should be scaled to within this range.

Another important step is data division. The data is broken down into two parts, training set and testing set. The training set is used for model development, and testing set is used to prediction. Basically, there is no guideline to divide the data. However, [33] suggested that the data need to be divided by a ratio written in percentages, such as 90%:10%, 80%:20% and 70%:30%, with a total of 100% for the combined ratio. The percentage ratio of 70%: 30% is used and utilized in a cross validation technique to build an appropriate model and to avoid over-fitting [24], [33]. Table 1.4 shows the data division with a ratio of 70%:30% for each category of claims.

A single hidden layer is used and the determination of the number of hidden nodes is done via a trial-and-error method [33]. The related illustration is given in Fig.

1.1. Basically, the purpose of a hidden layer is to detect the features, to capture the data pattern, and to perform the complicated non-linear mapping between the input and the output variables. Hence, we applied [33] suggestion to determine the number of hidden nodes, whether they are "*n*", or "*2n*", or "*2n+1*", where "*n*" is the number of input nodes. The number of input nodes is predetermined by trial and error in the proposal stage based on the data given by Insurance Services Malaysia Berhad (ISM).

Another factor to be considered is the learning rate (η) and the momentum (α) where the value is in the range of 0 to 1. By using different values of learning rate and momentum, the learning process can be speeded up, causing the network convergence to be either slower or faster. However, the choice of learning rate and momentum can be very sensitive [33]. Hence, a simple way to choose the learning rate and the momentum is through a trial and error-method.

The BP algorithm involves two phases, the forward phase and the backward phase. In the forward phase, the activations are propagated from the input to the output layer, while in the backward phase, if the output pattern is different from the desired output, the error between actual and predicted values in the output layer is calculated and propagated backwards to modify the weights and bias values. The most popular error function used for the output layer is the mean sum squared error. The activation function ensures the relationship between the input and the output of a node.

The network is trained with a pre-defined stopping criterion; either the number of iterations has been reached or when the total sum of square errors is lowers than a pre-determined value. This is the core part of ANN. The summary of the BPNN architecture and parameters used are shown in Table 1.5 and Table 1.6 describes the tested network structures. By applying the parameters show in Table 1.5 and depends on the process of trial and error with some considerations to give the best predictive result. Fig. 1.2 illustrates the network structure model with different number of hidden layer based on Table 1.6.

Table 1.4: Data division with 70%:30%

Category of claims		**Actual**
TPPD	Training	741
	Testing	318
	Total data	1059
OD	Training	386
	Testing	166
	Total data	552

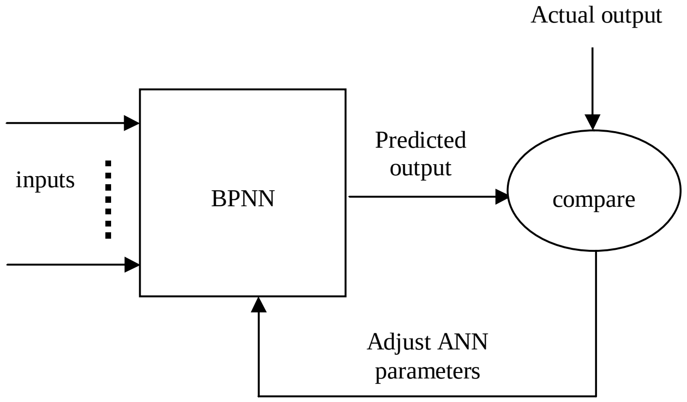


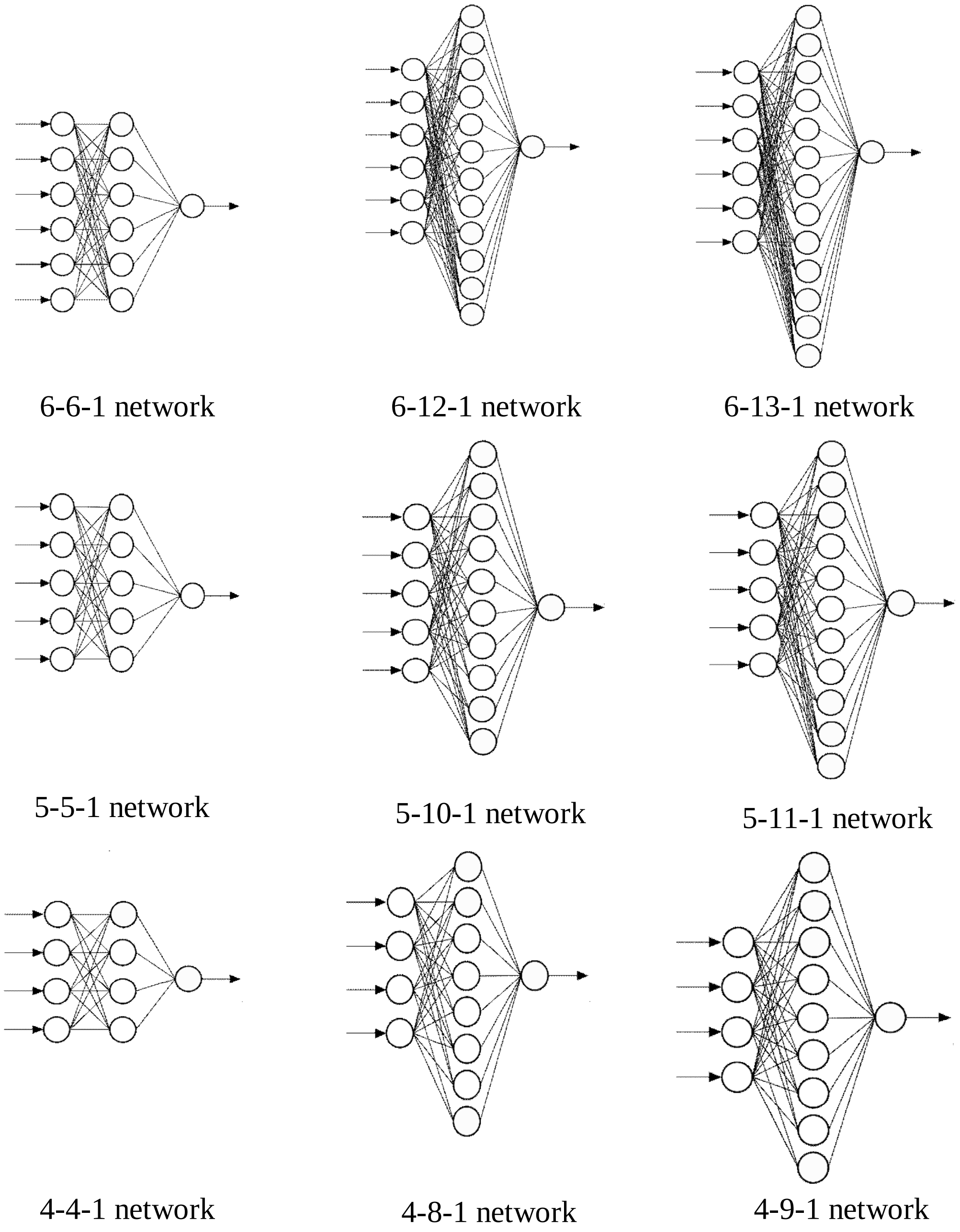
Fig. 1.1: BPNN proposed model for motor insurance claim

Table 1.5: Summary of standard BPNN architecture and parameters

Number of input nodes	4,5 and 6	
Number of hidden layer	1	
Number of hidden nodes	See Table 1.4 (tested network structures)	
Number of output nodes	1	
Learning rate	0.3	
Momentum	0.9	
Activation function	**Input to hidden layer**	sigmoid
	Hidden layer to output	sigmoid
Error performance	Mean square of error (MSE)	

Table 1.6: Tested network structures

4-4-1	4-8-1	4-9-1
5-5-1	5-10-1	5-11-1
6-6-1	6-10-1	6-12-1



6-6-1 network

6-12-1 network

6-13-1 network

5-5-1 network

5-10-1 network

5-11-1 network

4-4-1 network

4-8-1 network

4-9-1 network

Fig. 1.2**:** Network structure model with different number of hidden layer

The final step is model evaluation and validation. Four statistical methods were used to measure the constructed models such as mean squared of error (MSE), root mean square of error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). The MAPE measures the absolute error as a percentage. A lower percent or closer to zero deviation implies a more accurate prediction. The best model is chosen based on four error measurements, RMSE, MSE, MAPE, and MAE. If the results given by the four error measurements are inconsistent, then MAPE is chosen as a benchmark. These statistical methods were chosen because of the wide applicability [27], [2], [7], [3], [33].

4 Experimental Results and Analysis

This study discusses the two models namely, claim frequency and claim severity and were tested on two category of claims which are TPPD and OD. The development of BPNN model is done through trial-and-error with the aim to obtain the best predictive result for claim frequency and claim severity. Thus, nine networks have been developed (see Table 1.4) and tested on the two models.

Thus, from the nine network structures, the best model is chosen based on the smallest error value of the MSE, RMSE, MAE and MAPE. The results of the best BPNN model for claim frequency and claim severity for each category of claims are outlined in Table 1.7 and Table 1.8, respectively. The experimental result reveals that the number of input nodes and hidden nodes, as well as the parameters chosen influence the predictive accuracy. As a result, the best networks for each category of claims are shown in Fig. 1.3 and 1.4. The criterion used to determine the best BPNN model by looking at the lowest error value given by MSE, RMSE, MAE and MAPE. However, if the result produced by the four error measurement is inconsistent, then MAPE is chosen. As a conclusion, the BPNN is capable of producing a reliable prediction for motor insurance claims.

Table 1.7: Best BPNN model for claim frequency

Category of claims	Network	Claim frequency			
		MSE	RMSE	MAE	MAPE
TPPD	6-13-1	369.50	19.22	10.73	0.2191
OD	4-9-1	3383.56	58.17	32.49	0.2169

Note: Best parameters chosen through trial-and error with learning rate = 0.3, momentum = 0.7

Table 1.8: Best BPNN model for claim severity

Category of claims	Network	Claim severity			
		MSE	RMSE	MAE	MAPE
TPPD	6-12-1	2870793.89	1694.34	1105.20	0.6515
OD	4-8-1	8441272.93	2905.39	2097.01	0.3261

Note: *Best parameters chosen through trial-and error with learning rate = 0.3, momentum = 0.7*

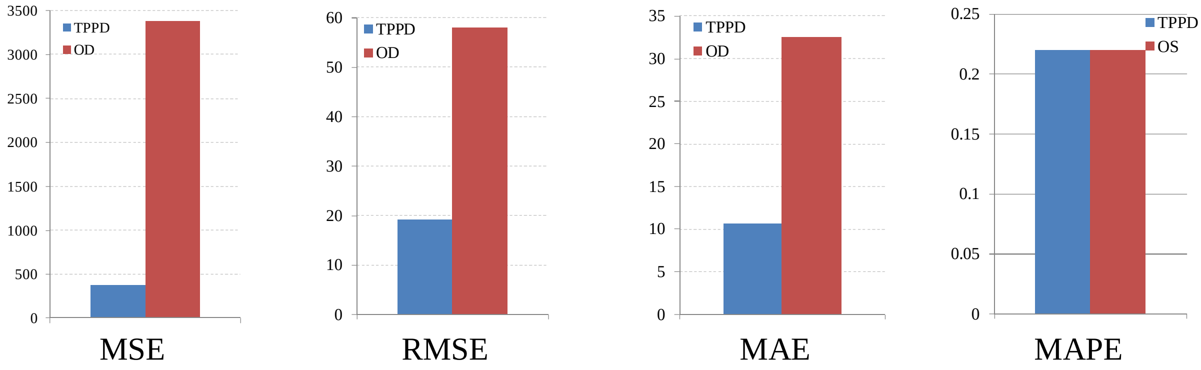


Fig. 1.3: Comparison based on error measurements for claim frequency

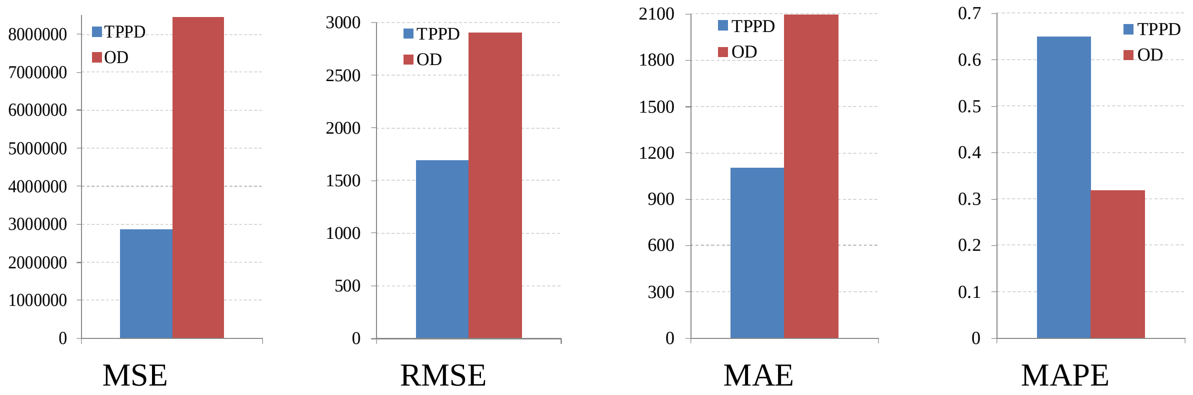


Fig. 1.4: Comparison based on error measurements for claim severity

Fig. 1.3 and 1.4 show that among the four error measurements, MAPE error is the easier to understand and represent a smallest value of error for claim frequency and claim severity.

5 Conclusion

In this paper, we use BPNN as a learning tool for motor insurance claim in predictive modelling. The model development using this has been discussed in details. It is concluded that the BPNN model is successful in predictive modeling the Malaysian motor insurance claims by using several of network structures. Several factors that significantly influenced the performance of the BPNN have also been discussed, namely the network structure (number of input nodes and number of hidden nodes), data preprocessing, the parameters and the error measurement. The main advantage of using BPNN is that the model is capable of dealing with non-linear data.

ACKNOWLEDGEMENTS

This research is supported by the Ministry of Higher Education of Malaysia under a grant number (R.J130000.7828.4F606). The authors would like to thank the Ministry of Higher Education of Malaysia (MOHE), UTM Big Data Centre

(UTM-BDC) of Ibnu Sina Institute for Scientific and Industrial Research (ISI-SIR), Universiti Teknologi Malaysia for their support in this study.

References

[1] Aftab, S., Abbas, W., Bilal, M.M., Hussain, T., Shoaib, M. and Mehmood, S.H. 2013. Data mining in insurance claims (DMICS) two-way mining for extreme values, International *Conference on Digital Information Management (ICDIM),* pp. 1-6, 10-12.

[2] Azlan, M. Z., Habibollah, H. and Safian, S. 2010. Prediction of surface roughness in the end milling machining using Artificial Neural Network, *Expert Systems with Applications*, 37(2):1755-1768.

[3] Bahia, H. S. I. 2013. Using Artificial Neural Network Modelling in Forecasting Revenue: Case Study in National Insurance Company/Iraq, *International Journal of Intelligence Science*, 3(3):136-143.

[4] Baser, F. and Apaydin, A. 2010. Calculating Insurance Claim Reserves with Hybrid Fuzzy Least Squares Regression Analysis. *Gazi University Journal of Science*, 23(2):163-170.

[5] Batty, M., Tripathi, A., Kroll, A., Peter Wu, C-S., Moore, D., Stehno, C., Lau, L., Guszcza, J. and Katcher, M. 2010. Predictive Modelling for Life Insurance, *Deloitte Consulting LLP* (April).

[6] Christmann, A. (2004). An approach to model complex high-dimensional insurance data**,** *Allgemeines Statistisches Archiv*, 88: 375-397.

[7] Chu, F. L. 2009. Forecasting tourism demand with ARMA-based methods, *Tourism Management*, 30:740-751.

[8] Dalkilic, T. E., Tank, F. and Kula, K. S. 2009. Neural networks approach for determining total claim amounts in insurance, *Insurance: Mathematics and Economics*, 45(2):236-241.

[9] David, M. 2015. Auto insurance premium calculation using generalized linear models, *Procedia Economics and Finance,* 20:147-156.

[10]Dugas, C., Chapados, N., Ducharme, R., Saint-Mleux, X. and Vincent, P. 2011. A high-order feature synthesis and selection algorithm applied to insurance risk modelling, *International Journal of Business Intelligence and Data Mining*, 6(3): 237-258.

[11]Frees, E. W., Peng Shi, P. and Valdez, E. A. 2008. Actuarial applications of a hierarchical insurance claims model, *ASTIN Bulletin*, 39 (1):165–197.

[12]Frees, E. 2014. Frequency and severity models. In E. Edward, G. Meyers, and R. A. Derrig (Eds.), Predictive Modelling Applications in Actuarial Science, Cambridge. Cambridge University Press.

[13]Ibiwoye, A., Ajibola, O. O. E. and Sogunro, A. B. 2012. Artificial Neural Network Model for Predicting Insurance Insolvency, *International Journal of Management and Business Research*, 2(1):59- 68.

[14]Ismail, N. and Jemain, A. A. 2008. Construction of an insurance scoring system using regression models, *Sains Malaysia*, 37(4):412-41.

[15]Kim, J. H. T. and Kim, J. 2014. Fuzzy regression towards a general insurance application. *Journal of Application Mathematics & Informatics*, 32(3-4):343 - 357.

[16]Kitchen, F. L. 2009. Financial implications of artificial neural networks in automobile insurance underwriting, *International Journal of Electronic Finance*, 3(3):311-319.

[17]Laas, D., Schmeiser, H. and Wagner, J. 2014. Empirical Findings on Motor Insurance Pricing in Germany, Austria, and Switzerland. *Geneva Papers on Risk and Insurance - Issues and Practice*, In Press.

[18]Lennon, H. 2011. *Generalized Linear Models and their Extensions for Insurance Data*. Master of Degree, University of Manchester.

[19]Lin, Y-J., Huang, C-S. and Lin, C-C. 2008. Determination of Insurance Policy Using Neural Networks and Simplified Models with Factor Analysis Technique", *WSEAS Transaction on Information science and Applications*, 5(10):1405-1415.

[20]Liew, T. and Chen, W. Y. 1998. Intelligence detection of drill wear. *Journal of Mechanical Systems and Signal Processing*, 12:863–873.

[21]Mohamed, M. A., Ismail, H., Razali, A. M., Ismail, N. and Ganiyat, A. U. 2011. Own damage, third party property damage claims and Malaysian motor insurance: An empirical examination, *Australian Journal of Basic and Applied Sciences*, 5(7):1190-1198.

[22]Pelessoni, R. and Picech, L. (1998). Some applications of unsupervised neural networks in rate making procedure, *The General Insurance Convention and Astin Colloquium*, 2: 550-567.

[23]Pinquet, J. 2012. Experience rating in non-life insurance. *Working Papers hal-00677100*, HAL.

[24]Rodriguez, J. D., Perez, A. and Lozano, J. A. 2010. Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3):569-575.

[25]Rosenberg, M. A., Frees, E. W., Sun, J., Johnson Jr., P. H. and Robinson, J. (2007). Predictive Modelling with Longitudinal Data, *North American Actuarial Journal*, 11(3):54-69.

[26] Salcedo-Sanz, S., Cuadra, L., Portilla-Figueras, J. A., Jiménez-Fernández, S. and Alexandre-Cortizo, E. 2012. A review of computational intelligence

algorithms in insurance applications, *Statistical and Soft Computing Approaches in Insurance Problems* (pp. 1 – 50), Nova Science Publishers.

[27]Sallehuddin, R., Shamsuddin, S.M. and Mohd Hashim, S.Z. 2010. Forecasting Small Data Set Using Hybrid Cooperative Feature Selection, in *Computer Modelling and Simulation (UKSim), 2010 12th International Conference on*, pp. 80-85.

[28]Shapiro, A. F. 2009. Fuzzy random variables, *Insurance: Mathematics and Economics*, 44:307-314.

[29]Shi, P. and Valdez, E. A. 2014. Multivariate negative binomial models for insurance claim counts, *Insurance: Mathematics and Economics*, 55:18-29.

[30]Shi, P., Feng, X. and Ivantsova, A. 2015. Dependent frequency–severity modelling of insurance claims, *Insurance: Mathematics and Economics*, 64:417-428.

[31]Viaene, S., Dedene, G. and Derrig, R. A. 2005. Auto claim fraud detection using Bayesian learning neural networks, *Expert Systems with Applications*, 29:653-666.

[32]Yeo, A. C. and Smith, K. A. 2003. An integrated data mining approach to premium pricing for automobile insurance industry. *Intelligent techniques in the insurance industry: theory and applications*, World Scientific Press.

[33]Zhang, G., Patuwo, B. E. and Hu, M. Y. 1998. Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14(1): 35–62.

[34]Zhang G. P. 2003. Time series forecasting using a hybrid ARIMA and neural network model, *Neurocomputing*, 50:159-175.