Sains Malaysiana 37(4)(2008): 412-419

Construction of an Insurance Scoring System Using Regression Models

(Pembinaan Sistem Skor Insurans Melalui Model Regresi)

NORISZURA ISMAIL & ABDUL AZIZ JEMAIN

ABSTRACT

This study suggests the regression models of Lognormal, Normal and Gamma for the construction of an insurance scoring system. Comparison between Lognormal, Normal and Gamma regression models were also carried out, and the comparison were centered upon three main elements; fitting procedures, parameter estimates and structure of scores. The main advantage of utilizing a scoring system is that the system may be used by insurers to differentiate between good and bad insureds and thus allowing the profitability of insureds to be predicted.

Keywords: Profitability; regression models; scoring system

ABSTRAK

Model regresi Lognormal, Normal dan Gamma dicadang untuk membina suatu sistem skor insurans. Perbandingan di antara model regresi Lognormal, Normal dan Gamma juga dilaksanakan, dan perbandingan ini tertumpu kepada tiga elemen utama; prosedur penyuaian, penganggar parameter dan struktur skor. Kelebihan utama sistem skor adalah ia boleh diterap oleh syarikat insurans untuk membezakan insud yang baik dan kurang baik dan membenarkan peramalan keberuntungan insud dilakukan.

Kata kunci: Keberuntungan; model regresi; sistem skor

INTRODUCTION

One of the most recent developments in the U.S. and European insurance industry today is the rapidly growing use of scoring system in pricing, underwriting and marketing of high volume and low premium insurance businesses. In the Asian markets however, the scoring system is still considered as relatively new, although several markets in the region have already started utilizing the system especially in its rating of motor insurance premium. In Singapore for instance, towards the end of 1992, the biggest private car insurer, NTUC Income, announced that it was changing from tariff system to scoring system as it was said that under the scoring system, owners of newer cars and more expensive models would probably pay lower premiums (Lawrence 1996).

Utilization of a scoring system provides several advantages in the pricing, underwriting and marketing of insurance businesses. One of it main advantages is that the scores may be used by insurers to differentiate between "good" and "bad" insureds and thus allowing the profitability of insureds to be predicted by using a specified list of rating factors such as driver's experience, vehicle characteristics and scope of coverage. In addition to distinguishing the risks of insureds, insurers may also employ the scores to determine the amount of premium to be charged on potential new clients.

Several studies have been carried out on the methodology and construction of scoring system. For examples, Coutts (1984) proposed Orthogonal Weighted Least Squares (OWLS) to convert premium amounts into scores and examined the impact of changing several input assumptions such as inflation rates, base periods of bodily injury claims, expenses and weights on the structure of scores. Brockman and Wright (1992) suggested Gamma regression model for converting premium amounts into scores, rationalizing that the variance of Gamma depends on weights or exposures and not on magnitude of premium amounts.

In the recent years, Miller and Smith (2003) conducted an actuarial analysis of the relationship between credit-based insurance score and propensity of loss for private passenger automobile insurance, utilizing Poisson distribution for claim frequency analysis and Gamma distribution for average claim costs analysis. In their study, insurance scores were found to be correlated with propensity of loss and this correlation is primarily due to the correlation between insurance scores and claim frequency rather than average claim severities. Anderson et al. (2004) suggested Generalized Linear Modeling (GLM) for deriving scores, suggesting the fitting of frequency and severity separately for each claim type as the starting point. The expected claims costs resulted from frequency and severity fitting were then divided by the premiums to yield the expected loss ratios, and finally, the profitability scores were produced by rescaling the loss ratios. Wu and Lucker (2004) reviewed the basic structure of several insurance credit scoring models in the U.S. by dividing scoring algorithms into two main categories; rule-based approach which assigns scores directly to each rating factor, and formula approach which determines scores using mathematical formulae. In their study, the methods of minimum bias and GLM were suggested for rule-based approach whereas the methods of Neural Networks (NN) and Multivariate Adaptive Regression Splines (MARS) were suggested for formula approach. Wu and Guszcza (2004) studied the relationship between credit scores and insurance losses by fitting data and producing scores using data mining methodology and several predictive modeling techniques such as NN, GLM, Classification and Regression Trees (CART) and MARS. The results of their multivariate predictive modeling indicated that credit scores showed significant relationships with loss ratio, frequency and severity of an insurance losses. Vojtek and Kocenda (2006) reviewed several methods of credit scoring employed by banks such as linear discriminant analysis (LDA), logit analysis, *k*-nearest neighbor classifier (*k*-NN) and NN to evaluate the applications of loans in Czech and Slovak Republics. Based on their study, logit analysis and LDA methods were mostly used, CART and NN methods were used only as supporting tools, and *k*-NN method was rarely used in the process of selecting variables and evaluating the quality of credit scoring models. Recently, Karlis and Rahmouni (2007) predicted the number of defaults in loan applications by developing finite mixture of Poisson regression model to allow for over-dispersion and to present better interpretability of the results. Their study indicates that the finite mixture of Poisson regression model is more flexible than the Negative Binomial regression model especially if the data have a long right tail.

The objective of this study is to suggest the regression models of Lognormal, Normal and Gamma for construction of an insurance scoring system. Even though several actuarial studies have been carried out on the methodology of scoring system, the detailed procedure of these methods were not provided, except for Coutts (1984) who proposed the use of Orthogonal Weighted Least Squares (OWLS) to convert premium amounts into scores. Furthermore, the Lognormal model proposed in our study differs from the OWLS method suggested by Coutts (1984) in terms of fitting procedure. The OWLS method assumed that the weights were possible to be factorized and the fitted value were calculated by using estimated weights whereas in this study, the weights were not required to be factorized and were not replaced by the estimated weights. In addition to suggesting Lognormal, Normal and Gamma regression models for constructing the scoring system, comparison between Lognormal, Normal and Gamma regression models will also be carried out in this study, and the comparison will be centered upon three main elements; fitting procedures, parameter estimates and structure of scores. The main advantage of having a scoring comparison between Lognormal, Normal and Gamma regression models is that the comparison allows an insurer to choose the best regression model that fulfills the company's objectives and requirements.

METHODOLOGY

This section provides the methodology of constructing a scoring system based on three types of regression models; Lognormal, Normal and Gamma. Response variable, independent variables and weight for the regression models are premium amounts, rating factors and exposures and the datasets required are $(g_i, e_i)$where g_i and e_i respectively denote the premium amounts and the exposure ith observation or rating class, $i = 1,2,..., n$.

Table 1 shows the related rating factors, premium amounts and exposures for several rating classes which were used to construct the scoring system in this study. Premium amounts were written in Ringgit Malaysia (RM) currency and they were based on a motor insurance claims experience provided by an insurance company in Malaysia. Exposures were written in terms of number of vehicle years and the rating factors considered, which were further divided into several rating classes, consist of scope of coverage (comprehensive and non-comprehensive), vehicle make (local and foreign), use-gender (private-male, private-female and business), vehicle year (0-1, 2-3, 4-5 and 6+) and location (Central, North, East, South and East Malaysia). It should be noted that preliminary analysis such as one-way and two-way distributions across classes of each rating factors should be implemented prior to the construction of a scoring model to assure that the predictive power of the scoring model stays within a reasonable range of time.

LOGNORMAL MODEL

Let the relationship between premium amounts, g_i and scores, s_i, be written as,

$$g_i = b^{s_i}, \tag{1}$$

or,

$$\log_b g_i = s_i. \tag{2}$$

In this study, $b = 1.1$ is chosen for Equation (1) to accommodate the conversion of premiums which range from RM30 to RM3,000 into scores which range from 0 to 100. For example, the score that corresponds to the premium amount of RM3,000 is equal to 84.

Assume that the distribution of premium, G_i, is Lognormal with parameters s_i and $e_i^{-1}\sigma^2$. Therefore, the distribution of $\log_{1.1} G_i$ is Normal with mean s_i and variance $e_i^{-1}\sigma^2$, where the density function is,

$$f(\log g_i; S_i) = \frac{1}{g_i\sqrt{2\pi\sigma^2 e_i^{-1}}}\exp\left(-\frac{e_i(\log g_i - s_i)^2}{2\sigma^2}\right).$$

The relationship between scores, s_i, and rating factors, x_{ij}, may be written in a linear function,

$$s_i = \mathbf{X_i}^{\mathrm{T}}\boldsymbol{\beta} = \sum_{j=1}^{p}\beta_j x_{ij}, \tag{3}$$

TABLE 1. Rating factors, exposures and premium amounts for Malaysian data

Rating factors					Exposure (vehicle-year)	Premium amount (RM)
Coverage	Vehicle make	Use-gender	Vehicle year	Location		
Comprehensive	Local	Private-male	0-1 year	Central	4243	1811
				North	2567	2012
				East	598	1927
				South	1281	1869
				East M'sia	219	983
			2-3 years	Central	6926	1704
				North	4896	1919
				East	1123	1854
				South	2865	1794
				East M'sia	679	1301
			4-5 years	Central	6286	1613
				North	4125	1840
				East	1152	1770
				South	2675	1687
				East M'sia	700	1162
			6+ years	Central	6905	1524
				North	5784	1790
				East	2156	1734
				South	3310	1633
				East M'sia	1406	1144
		Private-female	0-1 year	Central	2025	1256
				North	1635	1343
				East	301	1396
				South	608	1289
				East M'sia	126	787
			2-3 years	Central	3661	1210
				North	2619	1298
				East	527	1255
				South	1192	1212
				East M'sia	359	942
			4-5 years	Central	2939	1139
				North	1927	1243
				East	439	1125
				South	959	1176
				East M'sia	376	652
			6+ years	Central	2215	1072
				North	1989	1215
				East	581	1219
				South	937	1112
				East M'sia	589	623
			0-1 year	Central	290	722
		Business		North	66	547
				East	24	107
				South	52	685
				East M'sia	6	107
			2-3 years	Central	572	731
				North	148	630
				East	40	107
				South	91	657
				East M'sia	17	107
			4-5 years	Central	487	654
				North	100	549
				East	40	540
				South	59	571
				East M'sia	22	493
			6+ years	Central	468	567
				North	93	518
				East	33	562
				South	77	515
				East M'sia	25	402

where $\mathbf{x}_i$ denotes the vector of explanatory variables or rating factors that take the values of either one or zero, and $\boldsymbol{\beta}$ the vector of regression parameters. In other words, $\beta_j = 1,2,...,p$, represents the individual score of each rating factor and s_i represents the total scores of all rating factors.

The first derivatives of Equation (3) may be simplified into,

$$\frac{\partial s_i}{\partial \beta_j} = x_{ij} . \quad (4)$$

The solution for $\boldsymbol{\beta}$ may be obtained from the maximum likelihood equation,

$$\frac{\partial \ell}{\partial \beta_j} = \sum_i e_i (\log g_i - S_i) x_{ij} = 0 , \quad j = 1,2,...,p . \quad (5)$$

Since the maximum likelihood equation shown by Equation (5) is also equivalent to the Normal equation in standard weighted linear regression, $\boldsymbol{\beta}$ may be solved by using Normal equation.

NORMAL MODEL

Assume that the distribution of premium, G_i, is Normal with mean δ_i and variance $e_i^{-1}\sigma^2$, where the density function is,

$$f(g_i; \delta) = \frac{1}{\sqrt{2\pi\sigma^2 e_i^{-1}}} \exp\left(-\frac{e_i(g_i - \delta_i)^2}{2\sigma^2}\right).$$

The conversion of premium amounts into scores may be implemented by letting the relationship between scores, S_i, and fitted premium, δ_i, to be written in a log-linear function or multiplicative form. If the base value is equal to 1.1, the fitted premium is,

$$\delta^i = (1.1)^{S_i} \quad (6)$$

where $s_i = \mathbf{X}_i^T \boldsymbol{\beta} = \sum_{j=1}^{p} \beta_j x_{ij}$.

The first derivatives of Equation (6) is equal to,

$$\frac{\partial \delta_i}{\partial \beta_j} = \log(1.1)\delta_i x_{ij} \quad (7)$$

and the solution for $\boldsymbol{\beta}$ may be obtained from the maximum likelihood equation,

$$\frac{\partial \ell}{\partial \beta_j} = \sum_i e_i (g_i - \delta_i)\delta_i x_{ij} = 0 , \quad j = 1,2,...,p. \quad (8)$$

The maximum likelihood equation shown by Equation (8) is not quite straightforward to be solved compared to the Normal equation shown by Equation (5). However, since Equation (8) is equivalent to the weighted least squares, the fitting procedure may be carried out by using an iterative method of weighted least squares (McCullagh & Nelder 1989; Mildenhall 1999; Dobson 2002; Ismail & Jemain 2005; Ismail & Jemain 2007). In this study, the iterative weighted least squares procedure was performed by using SPLUS programming.

GAMMA MODEL

The construction of scoring system based on Gamma model is also similar to Normal model. Assume that the distribution of premium, G_i, is Gamma with mean δ_i and variance $v^{-1}\delta_i^2$, where the density function is,

$$f(g_i; \delta_i) = \frac{1}{g_i \Gamma(v)} \left(\frac{v g_i}{\delta_i}\right)^v \exp\left(-\frac{v g_i}{\delta_i}\right)$$

and v denotes the index parameter.

The conversion of premium amounts into scores may also be implemented by letting the relationship between scores, S_i, and fitted premium, δ_i, to be written in a log-linear function or multiplicative form which is equal to Equation (6). Therefore, the first derivative of Equation (6) is also the same as Equation (7).

Assume that the index parameter, v, varies within classes, so that the index parameter can be written as $v_i = e_i\sigma^2$ and the equation for variance of response variable is equal to $\sigma^2\delta_i^2 e_i^{-1}$. By using maximum likelihood method, the solution for $\boldsymbol{\beta}$ may be obtained through the maximum likelihood equation,

$$\frac{\partial \ell}{\partial \beta_j} = \sum_i \frac{e_i (g_i - \delta_i) x_{ij}}{\delta_i} , \quad j = 1,2,...,p. \quad (9)$$

Again, the maximum likelihood equation shown by Equation (9) is not quite straightforward to be solved compared to the Normal equation shown by Equation (5). However, since Equation (9) is also equivalent to the weighted least squares, the fitting procedure for Gamma model may be carried out by using an iterative method of weighted least squares. In this study, the iterative least squares procedure was employed by using SPLUS programming which is similar to the Normal model.

RESULTS

SCORING SYSTEM BASED ON LOGNORMAL MODEL

The best model for Lognormal regression may be determined by using standard analysis of variance. Based on the results of variance analysis, all of the rating factors were significant and 89.3% of the model's variations ($R^2 =$ 0.893) can be explained by using the same rating factors.

Parameter estimates for the best regression model are shown in Table 2. In order to provide significant effects for all individual regression parameters, the class for 2-3 year old vehicle was combined with 0-1 year old vehicle (intercept), and the classes for East location and South location were combined with Central location (intercept).

Construction of scoring system requires the negative estimates to be converted into positive values and the conversion process can be performed by using the following procedures. First, the smallest negative estimate of each rating factor was transformed into zero by adding an appropriate positive value. Next, the same positive value was added to the rest of the estimates categorized under the same rating factor. Finally, the intercept was deducted by the total positive values which were added to all estimates. The final scores were then rounded into whole numbers in order to provide easier calculation for premium amount and nicer interpretation for degree of risks relativities. Original estimates, modified estimates and final scores are shown in Table 3.

TABLE 2. Parameter estimates for Lognormal model

Parameters		Estimates	Std.dev.	*p*-values
β_1	Intercept	78.81	0.26	0.00
β_2	Non-comprehensive	-14.52	0.43	0.00
β_3	Foreign	4.23	0.26	0.00
β_4	Female	-4.30	0.28	0.00
β_5	Business	-9.25	0.53	0.00
β_6	4-5 years	-1.17	0.33	0.02
β_7	6+ years	-1.56	0.30	0.01
β_8	North	0.84	0.29	0.04
β_9	East Malaysia	-4.18	0.45	0.00

TABLE 3. Original estimates, modified estimates and final scores

Parameters	Original estimates	Modified estimates	Final scores
Intercept (Minimum score)	78.81	49.30	49
Coverage:			
Comprehensive	0.00		
Non-comprehensive	-14.52	14.52	15 0
Vehicle make:			
Local	0.00	0.00	0
Foreign	4.23	4.23	4
Use-gender:			
Private-male	0.00	9.25	9
Private-female	-4.30	4.95	5
Business	- 9.25	0.00	0
Vehicle year:			
0-1 year & 2-3 years	0.00	1.56	2
4-5 years	-1.17	0.39	0
6+ years	-1.56	0.00	0
Vehicle location:			
Central, East & South	0.00	4.18	4
North	0.84	5.02	5
East Malaysia	-4.18	0.00	0

The final scores shown in Table 3 clearly specify and summarize the degree of relative risks associated to each rating factor. For instance, the risks for foreign vehicles are relatively higher by four points compared to local vehicles, and the risks for male and female drivers who used their cars for private purposes are relatively higher by nine and five points compared to drivers who used their cars for business purposes.

Goodness-of-fit of the scores in Table 3 may be tested by using two methods; comparing the ratio of fitted over actual premium income, and comparing the difference between fitted and actual premium income. Table 4 shows the total difference of premium income and the overall ratio of premium income for the scores.

Based on Table 4, the total income of fitted premiums was understated by RM560,380 or 0.2% of the total income of actual premiums. Therefore, the fitted premiums for all classes were suggested to be multiplied by a correction factor of 1.002 to match their values with the actual premiums.

Besides differentiating between good and bad insureds, scoring system may also be used by insurers to calculate the amount of premium to be charged on each potential client. The procedure for converting scores into premium amounts involves two basic steps. First, the scores for each rating factor were recorded and aggregated. Then, the aggregate scores were converted into premium amount by using a scoring conversion table, a table listing the aggregate scores with associated monetary values. Table 5 shows a scoring conversion table which was constructed by using Equation (1).

COMPARISON OF SCORING SYSTEM BASED ON LOGNORMAL, NORMAL AND GAMMA MODELS

Comparison of parameter estimates resulted from Lognormal, Normal and Gamma regression models are shown in Table 5.

Based on Table 6, parameter estimates for Lognormal, Normal dan Gamma models provide similar values, except for β_2 and β_5 which produced larger values in Normal and Gamma models compared to Lognormal model.

Comparison of scoring system resulted from Lognormal, Normal and Gamma regression models are shown in Table 7. Scores for Lognormal model range from 49 to 84, scores for Normal model range from 53 to 84 and scores for Gamma model range from 51 to 85. In addition, the lowest minimum score is produced by Lognormal model. Based on minimum score and range of score, if an insurer is planning to lower its premium rates for low risks classes, Lognormal model may be an appropriate model for this purpose.

In terms of risks relativities, both Lognormal and Gamma models resulted in a relatively higher score for male driver, female driver and comprehensive coverage. Therefore, if an insurer is interested to charge higher premium for male driver, female driver and comprehensive coverage, both Lognormal and Gamma models may be

TABLE 4. Total premium income difference and overall premium income ratio

		Value
Total number of business/policy/exposure	$\sum_{i=1}^{240} e_i$	170,749
Total income from fitted premiums	$\sum_{i=1}^{240} e_i \hat{g}_i$	RM 275,269,816
Total income from actual premiums	$\sum_{i=1}^{240} e_i g_i$	RM 275,830,196
Total premium income difference	$\sum_{i=1}^{240} e_i (\hat{g}_i - g_i)$	RM 560,380
Overall premium income ratio	$\dfrac{\sum_{i=1}^{240} e_i \hat{g}_i}{\sum_{i=1}^{240} e_i g_i}$	0.998

TABLE 5. Scoring conversion table

Aggregate scores	Premium amounts (RM)	Aggregate scores	Premium amounts (RM)
49	107	67	595
50	118	68	654
51	129	69	719
52	142	70	791
53	157	71	870
54	172	72	958
55	189	73	1053
56	208	74	1159
57	229	75	1274
58	252	76	1402
59	277	77	1542
60	305	78	1696
61	336	79	1866
62	369	80	2052
63	406	81	2258
64	447	82	2484
65	491	83	2732
66	540	84	3005

TABLE 6. Estimates for Lognormal, Normal and Gamma regression models

Parameters		Lognormal			Normal			Gamma		
		Est.	std. error	*p*-value	Est.	std. error	*p*-value	Est.	std. error	*p*-value
β_1	Intercept	78.81	0.26	0.00	79.02	0.01	0.00	78.89	0.02	0.00
β_2	Non-comp	-14.52	0.43	0.00	-12.79	0.05	0.00	-13.71	0.03	0.00
β_3	Foreign	4.23	0.26	0.00	4.02	0.01	0.00	4.19	0.02	0.00
β_4	Female	-4.30	0.28	0.00	-4.03	0.01	0.00	-4.25	0.02	0.00
β_5	Business	-9.25	0.53	0.00	-7.40	0.03	0.00	-8.55	0.04	0.00
β_6	4-5 years	-1.17	0.33	0.02	-1.17	0.01	0.00	-1.17	0.02	0.00
β_7	6+ years	-1.56	0.30	0.01	-2.10	0.01	0.00	-1.73	0.02	0.00
β_8	North	0.84	0.29	0.04	0.49	0.01	0.00	0.81	0.02	0.00
β_9	East M'sia	-4.18	0.45	0.00	-4.01	0.03	0.00	-4.21	0.03	0.00

TABLE 7. Scoring system for Lognormal, Normal and Gamma regression models

Rating factors	Scores		
	Lognormal	Normal	Gamma
Minimum scores	49	53	51
Coverage:			
Comprehensive	15	13	14
Non-comprehensive	0	0	0
Vehicle make:			
Local	0	0	0
Foreign	4	4	4
Use-gender:			
Private-male	9	7	9
Private-female	5	3	4
Business	0	0	0
Vehicle year:			
0-1 year	2	2	2
2-3 years	2	2	2
4-5 years	0	1	1
6+ years	0	0	0
Location:			
Central	4	4	4
North	5	5	5
East	4	4	4
South	4	4	4
East Malaysia	0	0	0

suitable for fulfilling this strategy. However, the difference between Lognormal and Gamma model is that the scores for low risks classes provided by Gamma is slightly higher compared to Lognormal.

CONCLUSION

This paper discusses the methodology of constructing insurance scoring system using regression models of Lognormal, Normal and Gamma. The main advantage of utilizing scoring system is that the system may be used by insurers to differentiate between good and bad insureds and thus allowing the profitability of insureds to be predicted. In addition, scoring system has an operational advantage of reducing premium calculations and can be treated as a more sophisticated device for customers to assess their individual risks.

Relationship between aggregate scores and rating factors in Lognormal model was suggested to be written in a linear function or additive form, whereas relationship between aggregate scores and rating factors in Normal and Gamma models were proposed to be written in a log-linear function or multiplicative form. Regression parameters for Lognormal model were calculated by using standard Normal equation, whereas regression parameters for Normal and Gamma models were estimated by using the iterative weighted least squares procedure.

The best regression model for Lognormal model was selected by implementing standard analysis of variance. Goodness-of-fit of the scoring estimates were then tested by comparing the ratio of fitted over actual premium income and by comparing the difference between fitted and actual premium income.

Besides distinguishing the risks of insureds, another advantage of using scoring system is that the system enables the premium amount to be calculated easily. Hence, scoring system can also be used by insurers to examine the effect of various input assumptions, such as assumptions for risk and gross premium estimation, and assumptions for scoring system construction. A good example on the use of scores for examining various input assumptions was provided by Coutts (1984), who investigated changes of assumptions in the elements of inflation rates, base periods of bodily injury claims, expenses and weights.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support received in the form of a research grant (IRPA RMK8: 09-02-02-0112-EA274) from the Ministry of Science, Technology and Innovation, Malaysia. The authors also thank Insurance Services Malaysia for the data.

REFERENCES

Anderson, D., Feldblum, S., Modlin, C., Schirmacher, D., Schirmacher, E. & Thandi, N. 2004. A practitioner's guide to generalized linear models. *Casualty Actuarial Society Discussion Paper Program* 1-115.

Brockman, M.H. & Wright, T.S. 1992. Statistical motor rating: Making effective use of your data. *Journal of the Institute of Actuaries* 119(3): 457-543.

Coutts, S.M. 1984. Motor insurance rating, an actuarial approach. *Journal of the Institute of Actuaries* 111: 87-148.

Dobson, A.J. 2002. *An introduction to generalized linear models* (2nd edition). New York: Chapman & Hall.

Ismail, N. & Jemain, A.A. 2005. Bridging minimum bias and maximum likelihood methods through weighted equation. *Casualty Actuarial Society Forum* Spring: 367-394.

Ismail, N. & Jemain, A.A. 2007. Handling overdispersion with Negative Binomial and Generalized Poisson regression models. *Casualty Actuarial Society Forum Winter*: 103-158.

Karlis, D. & Rahmouni, M. 2007. Analysis of defaulters' behavior using the Poisson-mixture approach. *IMA Journal of Management Mathematics* 18: 297-311.

Lawrence, B. 1996. Motor insurance in Singapore. In Low Chan Kee (ed.). *Actuarial and insurance practices in Singapore*. Singapore: Addison-Wesley. pp. 191-216.

McCullagh, P. & Nelder, J.A. 1989. *Generalized linear model* (2nd ed.). London: Chapman & Hall.

Mildenhall, S.J. 1999. A systematic relationship between minimum bias and generalized linear models. *Proceedings of the Casualty Actuarial Society* 86(164): 93-487.

Miller, M.J. & Smith, R.A. 2003. The relationship of credit-based insurance scores to private passenger automobile insurance loss propensity. *Presentation to NAIC*. July, 2003.

Vojtek, M. & Kocenda, E. 2006. Credit scoring models. *Czech Journal of Economics and Finance* 56(3-4): 152-167.

Wu, C.P. & Lucker, J.R. 2004. A view inside the Black Box: A review and analysis of personal lines insurance credit scoring models filed in the state of Virginia. *Casualty Actuarial Society Forum* Winter: 251-290.

Wu, C.P. & Guszcza, J.C. 2004. Does credit score really explain in losses? Multivariate analysis from a data mining point of view. *Casualty Actuarial Society Forum* Winter: 113-138.

Pusat Pengajian Sains Matematik
Fakulti Sains dan Teknologi
Universiti Kebangsaan Malaysia
43600 Bangi, Selangor D. E.
Malaysia

Received: 26 September 2007
Accepted: 4 January 2008