27_Conf_2014_Premium analysis for copula model: A case study for Malaysian motor insurance claims

By Yulia Resti

AIP Conference Proceedings

Premium analysis for copula model: A case study for Malaysian motor insurance claims

Yulia Resti, Noriszura Ismail, and Saiful Hafizah Jaaman

Citation: AIP Conference Proceedings 1602, 1072 (2014); doi: 10.1063/1.4882617
View online: http://dx.doi.org/10.1063/1.4882617
View Table of Contents: http://scitation.aip.org/content/aip/proceeding/aipcp/1602?ver=pdfcov
Published by the AIP Publishing

Premium Analysis for Copula Model: A Case Study for Malaysian Motor Insurance Claims

Yulia Resti[a], Noriszura Ismail[b] and Saiful Hafizah Jaaman[b]

[a]*Jurusan Matematika, Fakultas Matematika & Ilmu Pengetahuan Alam, Universitas Sriwijaya, Indonesia*
[b]*School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia*

Abstract. This study performs premium analysis for copula models with regression marginals. For illustration purpose, the copula models are fitted to the Malaysian motor insurance claims data. In this study, we consider copula models from Archimedean and Elliptical families, and marginal distributions of Gamma and Inverse Gaussian regression models. The simulated results from independent model, which is obtained from fitting regression models separately to each claim category, and dependent model, which is obtained from fitting copula models to all claim categories, are compared. The results show that the dependent model using Frank copula is the best model since the risk premiums estimated under this model are closely approximate to the actual claims experience relative to the other copula models.

Keywords: premium analysis, copula, regression, insurance claims, Elliptical, Archimedean.
PACS: 02.50.-r, 02.50.Tt, 02.70.Rr

INTRODUCTION

The global financial crisis in 2008 resulted in a negative impact on the majority of sectors of financial industry in Malaysia. This impact can be seen in the decreasing stock price index of capital market and the increasing interest rate of financial industry, resulting in decreasing loan demands and increasing non-performing loans. In insurance sector, life insurers faced difficulties in producing adequate investment earning which is required for fulfilling policyholders' obligations, while non-life insurers encountered problems in the increased number of cases of fraudulent claims.

Insurance premiums for motor vehicle in Malaysia are determined by tariff structure set by General Insurance Association of Malaysia (PIAM). Premium rates determined from tariff structure may cause insurers to be less competitive and may not protect public interests in terms of providing 'fair' premium rates. Bank Negara Malaysia has also reported that motor insurance tariff in Malaysia has not been revised for more than 30 years. Therefore, an adjustment of motor premium rates is required, and the adjusted rates should be reviewed periodically to ensure that the rates continue to reflect actual claims experience, where vehicle owners with good claims experience (low risks) enjoy better premium rates, and vice versa. In other words, premium should be paid at the rate closest to the actual claims experience.

A road vehicle accident may produce three dependent types of claims namely third party bodily injury (TPBI), own damage (OD), and third party property damage (TPPD). If there is more than one claim type, the independent assumption for claim types can lead to over- or underestimated premium.

METHODOLOGY

Risk Premium Calculation

Risk premium for the i-th risk class, $i=1,2,\ldots,n$, can be equated as the product of estimated claim frequency and estimated average claim cost (severity) for all claim categories [1-5]. As such, if we have three claim categories, the risk premium is

$$\tilde{r}_i = \tilde{f}_{i1}\tilde{c}_{i1} + \tilde{f}_{i2}\tilde{c}_{i2} + \tilde{f}_{i3}\tilde{c}_{i3} \tag{1}$$

Proceedings of the 3rd International Conference on Mathematical Sciences
AIP Conf. Proc. 1602, 1072-1077 (2014); doi: 10.1063/1.4882617

where $\tilde{f}_{ij}$ is the estimated claim frequency and $\tilde{c}_{ij}$ is the estimated claim severity for class i and category j, j=1,2,3. The calculation of premium for both independent and dependent models are performed using (1), but the claim costs for dependent model are assumed dependent between categories, and the claims are modeled using Elliptical and Archimedean copula families. Further studies on Elliptical and Archimedean copula families can be found in [6-10].

Frequency and Severity Models

Based on past literatures, claim frequency and severity for each category can be estimated using negative binomial [11-14] and gamma [15-16] regression models respectively. If the random variable for claim count, Y_i, is distributed as negative binomial regression, the probability mass function (p.m.f.) is

$$\Pr(Y_i = y_i \mid \mu_i, a) = \frac{\Gamma(y_i + a^{-1})}{y_i!\Gamma(a^{-1})}\left(\frac{\mu_i}{\mu_i + a^{-1}}\right)^{y_i}\left(\frac{a^{-1}}{\mu_i + a^{-1}}\right)^{a^{-1}}, \quad y_i = 0,1,2,\ldots \tag{2}$$

with mean $E(Y_i) = \mu_i$ and variance $Var(Y_i) = (1 + a\mu_i)\mu_i$, where a denotes the dispersion parameter. Negative binomial regression reduces to Poisson regression in the limit as $a \to 0$. If $a > 0$, the variance exceeds the mean, and negative binomial regression allows overdispersion.

If C_i is the random variable for claim severity and follows gamma regression, the p.m.f. is

$$f(c_i \mid \mu_i, v) = \frac{1}{\Gamma(v)}\left(\frac{v}{\mu_i}\right)^{v} c_i^{v-1} \exp\left(-\frac{vc_i}{\mu_i}\right), \quad c_i > 0 \tag{3}$$

with mean $E(C_i) = \mu_i$ and variance $Var(C_i) = v^{-1}\mu_i^2$, where denotes the scale parameter.

The covariates for both negative binomial and gamma regression models can be included via a log link,

$$\mu_i = \exp\left(\sum_k \beta_k x_{ik}\right) = \exp(\mathbf{x}_i^T \boldsymbol{\beta}),$$

where $\boldsymbol{\beta}$ is the vector of regression parameters and $\mathbf{x_i}$ is the vector of explanatory variables.

Copula Model

Under dependent model, claim severities are assumed dependent between categories and the claims for all categories are modeled using copula [17-18]. Consider the distribution function (d.f.) of a trivariate distribution, F. The idea of Sklar's Theorem for a trivariate distribution is to represent F in two parts; marginal d.f., F_j, and copula d.f., H. Both F_j and H can be connected in a trivariate d.f.

$$F(c_1, c_2, c_3) = H[F_1(c_1), F_2(c_2), F_3(c_3)] = H(u_1, u_2, u_3) \tag{4}$$

where U_1, U_2 and U_3 are standard uniform random variables.

The trivariate d.f. of an Elliptical copula is

$$H(u_1, u_2, u_3; \mathbf{R}) = F[F_1^{-1}(u_1), F_2^{-1}(u_2), F_3^{-1}(u_3)] \tag{5}$$

where $F_j^{-1}(u_j)$ is the inverse of an Elliptical c.d.f. and $\mathbf{R}$ is the correlation matrix. The $\mathbf{R}$ can be obtained using

$$\tau(X, Y) = \frac{2}{\pi}\arcsin[\rho(X, Y)] \tag{6}$$

In this study, we use normal and t copula, which belong to Elliptical family, for fitting claim severities. The trivariate d.f. of an Archimedean copula can be constructed through generator, φ

$$H(u_1, u_2, u_3) = \phi^{-1}[\phi(u_1) + \phi(u_2) + \phi(u_3)] \tag{7}$$

where ϕ^{-1} is the inverse generator. In this study, we use Clayton, Frank and Gumbel copula, which belong to Archimedean family, for fitting claim severities. The generator and inverse generator for Clayton, Frank and Gumbel copula respectively are

$$\phi_C(u) = u^{-\alpha} - 1, \quad \phi_C^{-1}(u) = (u+1)^{-\frac{1}{\alpha}} \tag{8}$$

$$\phi_F(u) = \ln\left(\frac{e^{-\alpha u} - 1}{e^{-\alpha}}\right), \quad \phi_F^{-1}(u) = \frac{1}{\alpha}\ln[1 + e^{u}(e^{-\alpha} - 1)] \tag{9}$$

$$\phi_G(u) = (-\ln u)^{\alpha}, \quad \phi_G^{-1}(u) = e^{-u^{\frac{1}{\alpha}}} \tag{10}$$

Premium Analysis

The risk premium from both independent and dependent assumptions can be analyzed using (11), which is the quadratic distance (qd) between model (M) and actual claims experience (ACE). The best model is when M is the closest to ACE. The quadratic distance is defined as

$$qd = \sum_{i=1}^{n}[(M - ACE)^2]^{1/2} \tag{11}$$

RESULTS

Negative binomial and gamma regression models for each category of claims are shown in Table (1) and Table (2). The negative binomial regressions are fitted to claim frequencies whereas the gamma regressions are fitted to claim severities.

TABLE (1). Negative binomial regression model

Parameter	TPBI claim			OD claim			TPPD claim		
	est.	***t*-ratio**	***p*-value**	**est.**	***t*-ratio**	***p*-value**	**est.**	***t*-ratio**	***p*-value**
Intercept	-4.33	-93.62	0.00	-2.50	-30.30	0.00	-3.38	-89.47	0.00
0-1 yr	-1.76	-24.20	0.00	-0.42	-4.96	0.00	-1.06	-17.31	0.00
2-3 yrs	-0.43	-6.85	0.00	0.20	2.44	0.01	-	-	-
4-5 yrs	-0.35	-5.55	0.00	0.15	1.89	0.06	-	-	-
6-7 yrs	-0.18	-2.98	0.00	-	-	-	-	-	-
0-1000 cc	0.10	1.74	0.08	-0.48	-5.00	0.00	-0.30	-4.73	0.00
1001-1300 cc	0.12	2.11	0.03	-0.22	-2.21	0.03	-	-	-
1501-1800 cc	-	-	-	0.19	2.06	0.04	-	-	-
1801+ cc	0.18	2.99	0.00	0.35	3.78	0.00	0.10	1.66	0.10
foreign	-	-	-	-0.16	-2.61	0.01	-0.19	-4.04	0.00
dispersion, a	0.01	2.32	0.02	0.04	4.29	0.00	0.02	3.53	0.00
Log L		-212.34			-328.14			-264.76	
AIC		442.67			676.27			541.53	
BIC		459.88			695.39			553.00	

TABLE(2). Gamma regression model

Parameter	TPBI claim			OD claim			TPPD claim		
	est.	*t*-ratio	*p*-value	est.	*t*-ratio	*p*-value	est.	*t*-ratio	*p*-value
Intercept	9.40	126.85	0.00	7.96	76.99	0.00	7.53	77.72	0.00
0-1 yr	-0.64	-3.20	0.00	0.28	2.58	0.01	-	-	-
2-3 yrs	-	-	-	-	-	-	-	-	-
4-5 yrs	-	-	-	-	-	-	0.26	1.87	0.06
6-7 yrs	-	-	-	-	-	-	-	-	-
0-1000 cc	-	-	-	-	-	-	-	-	-
1001-1300	-	-	-	-	-	-	-	-	-
1501-1800	-	-	-	0.51	4.43	0.00	-	-	-
1801+ cc	-	-	-	0.72	6.52	0.00	-	-	-
foreign	-	-	-	0.54	5.59	0.00	0.24	2.08	0.04
scale, σ	2728.99	4.87	0.00	616.36	4.93	0.00	418.33	4.91	0.00
Log L		-495.93			-443.45			-411.05	
AIC		997.87			898.90			830.10	
BIC		1003.60			910.37			837.75	

Table (3) shows the risk premium for independent and dependent models, whereas Table (4) provides the risk premium based on actual claims experience. Risk premiums for both independent and dependent models are calculated using (1). Under independent model, claim severities for each category are fitted separately to gamma regression models, whereas under dependent model, claim severities for all categories are fitted together using copula (normal and *t* copula from Elliptical family, and Frank, Clayton and Gumbel copula from Archimedean family) with gamma regression marginals. The best models for Elliptical and Archimedean families are Normal and Frank copula respectively. From Table (3), we can observe that the first rating class is for 0-1 year, 0-1000 cc and local vehicles, whereas the second rating class is for 0-1 year, 0-1000 cc and foreign vehicles. The estimated risk premiums under independent model for the first rating class is RM159, dependent model using normal copula is RM183, and dependent model using Frank copula is RM181. As for the second rating class, the estimated risk premiums under independent model is RM218, dependent-Normal copula model is RM228, and dependent-Frank copula model is RM207. The risk premiums for other risk classes are interpreted similarly.

TABLE (3). Independent and dependent premium

Exposure	Fitted count			Fitted cost									Fitted premium		
				Independent			Normal			Frank			Ind	Normal	Frank
	TPBI	OD	TPPD	TPBI	OD	TPPD	TPBI	OD	TPPD	TPBI	OD	TPPD			
65923	165	2200	576	6374	3790	1863	6634	4484	1897	6276	4441	1946	159	183	181
990	2	28	7	6374	6503	2368	6701	6984	2373	6261	6299	2292	218	228	207
47611	122	2061	562	6374	3790	1863	6606	4466	1909	6344	4445	1948	202	232	231
435	1	16	4	6374	6503	2368	6679	6988	2367	6234	6318	2302	279	294	268
32659	74	1761	385	6374	3790	1863	6610	4507	1905	6313	4447	1935	241	281	277
389															
3	9	179	38	6374	6503	2368	6578	7002	2360	6279	6292	2314	336	360	326
29670	67	1935	350	6374	6311	1863	6609	6254	1886	6308	5752	1951	448	445	412
20298	46	1128	198	6374	10829	2368	6657	9723	2361	6242	8193	2305	639	577	491
5360	15	410	70	6374	7785	1863	6590	7157	1891	6286	6617	1941	637	589	548
25914	70	1690	279	6374	13360	2368	6647	11152	2371	6371	9425	2290	914	769	655
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
26166	345	2213	737	12088	8185	2368	12177	7284	2363	12177	7284	2363	918	844	844
1744	27	203	66	12088	5884	1863	12130	5388	1891	12130	5388	1891	946	893	893
55830	880	5542	1737	12088	10097	2368	12140	8317	2376	12140	8317	2376	1267	1090	1090

TABLE (4). Premium from actual claims experience

Exposure	Actual count			Actual cost			Actual premium
	TPBI	OD	TPPD	TPBI	OD	TPPD	
65923	191	3181	579	6870	4514	2730	262
990	3	20	8	925	3799	399	83
47611	117	2190	690	8998	6915	3351	389
435	4	33	4	1748	4342	1196	356
32659	82	1730	496	7341	4304	1968	276
3893	6	166	38	7862	5548	2191	270
29670	65	1362	301	8960	6075	1948	318
20298	33	894	155	5915	8686	2374	410
5360	7	392	68	3344	5250	1491	407
25914	50	1476	218	12220	20772	4686	1246
:	:	:	:	:	:	:	::
26166	345	2091	869	31256	7584	3402	1131
1744	28	331	104	7116	3009	1642	783
55830	667	4794	1749	22024	8437	3844	1108

The risk premiums provided in Table (3) and Table (4) can be analyzed using (11). Table (5) provides the quadratic distance (qd) of risk premium between model and actual claims experience. It can be observed that the risk premiums from dependent-Frank copula model are closest to the risk premiums from actual claims experience.

TABLE (5). Premium quadratic distance

Independent Model	Dependent Model	
	Normal Copula	Frank Copula
1332	1001	982

Figure 1. presents the risk premiums from actual claims experience (PTS), independent model (MTB), dependent model using Normal copula (MCN) and dependent model using Frank copula (MCF) for all rating classes. The graphs show that the risk premiums from dependent model using Frank copula (MCF) has the closest distance to the actual claims experience (PTS).

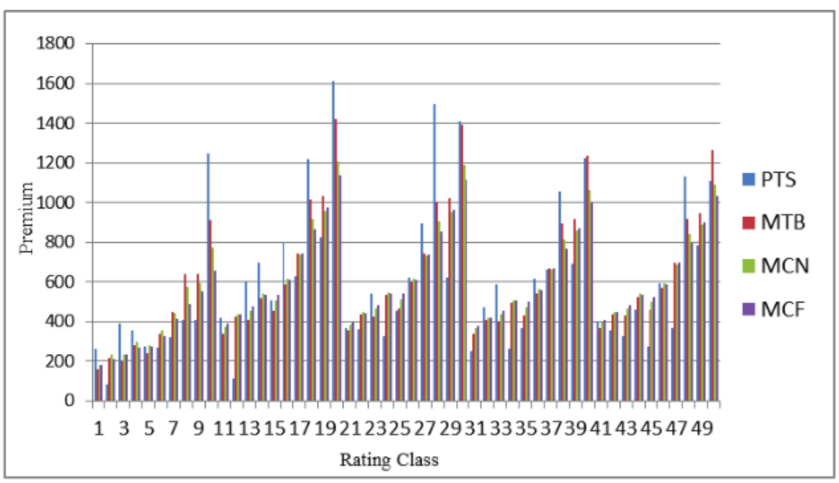

FIGURE 1. Risk premium from actual claims experience, independent and dependent models

CONCLUSIONS

This study has performed premium analysis for copula models with regression marginals. The analysis has been performed using Malaysian motor insurance claims data. Several copula models from Archimedean and Elliptical families have been considered, namely normal, *t*, Clayton, Frank and Gumbel. The best model from Elliptical family is normal copula, while the best model from Archimedean family is Frank copula. The results showed that the premium estimated from Frank copula (dependent model) has the closest distance to the actual claims experience, and therefore, can be considered as the best model for modeling dependent claims data.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support received in the form of research grants (FRGS/1/2013/SG04/UKM/02/5 and LRGS/TD/2011/UKM/ICT/03/02) from Ministry of Higher Education (MOHE), Malaysia. The authors also gratefully acknowledge Insurance Services Malaysia Berhad (ISM) for supplying the data.

REFERENCES

1. M.H. Brockman and T.S. Wright, *Journal of the Institute of Actuaries* **119**(3), 457-543 (1992).
2. S. Haberman and A.E. Renshaw, *The Statistician* **45**(4), 407-436 (1996).
3. A.E. Renshaw, *ASTIN Bulletin* **24**(2), 265-285 (1994).
4. S.M. Coutts, *Journal of the Institute of Actuaries* **111**, 87-148 (1984).
5. P.W. Cheong, A.A. Jemain and N. Ismail, *Journal of Quality Measurement and Analysis* **4**(1), 11-24 (2008).
6. U. Cherubini, E. Luciano and W. Vecchiato, *Copula Methods in Finance*, John Wiley and Son Ltd, 2004.
7. P. Embrechts, F. Lindskog and A. McNeil, "Modeling dependence with copulas and applications to risk management" in Rachev, S. (Ed.), *Handbook of Heavy Tailed Distribution in Finance*, Elsevier, 2003, pp. 329-384.
8. G. Wong, M.F. Lambert and A.V. Metcalfe, *Australian and New Zealand Industrial and Applied Mathematics Journal* **49**, 306-323 (2008).
9. J. Yan, "Modeling with copulas and engineering applications" in Pham, H. (Ed.), *Handbook in Engineering Statistics*, Springer, 2006, pp. 973-990.
10. J. Yan, *Journal of Statistical Software* **21**(4), 1-21 (2007).
11. M. Aitkin, D. Anderson, B. Francis and J. Hinde, *Statistical modelling in GLIM*, New York: Oxford University Press, 1990.
12. A.C. Cameron and P.K. Trivedi, *Journal of Applied Econometrics* **1**, 29-53 (1986).
13. N. Ismail and A.A. Jemain, *Casualty Actuarial Society Forum Winter*, 103-158 (2007).
14. P. McCullagh and J.A. Nelder, *Generalized Linear Models*, 2nd Edition. London: Chapman and Hall, 1989.
15. N. Ismail and A.A. Jemain, *Journal of Modern Applied Statistical Methods* **5**(2), 513-528 (2006).
16. Y. Resti, N. Ismail and S.H. Jaaman, *Journal of Mathematics and Statistics* **6**(2), 136-142 (2010).
17. E.W. Frees and E.A. Valdez, *North American Actuarial Journal* **2**(1), 1-25 (1998).
18. E.W. Frees and E.A. Valdez, *Journal of the American Statistical Association* **103**(484), 1457–1469 (2008).

27_Conf_2014_Premium analysis for copula model: A case study for Malaysian motor insurance claims

ORIGINALITY REPORT

16%
SIMILARITY INDEX

PRIMARY SOURCES

1	www.thescipub.com Internet	73 words — 3%
2	thescipub.com Internet	72 words — 3%
3	aip.scitation.org Internet	54 words — 2%
4	www.ceser.in Internet	43 words — 2%
5	www.tandfonline.com Internet	41 words — 1%
6	www.naturalspublishing.com Internet	37 words — 1%
7	research-api.cbs.dk Internet	24 words — 1%
8	Dawi Karomati Baroroh, Alexander Hantya Mahardhika. "SEM PLS Models for Performance Analysis of Manufacturing Companies", 2018 4th International Conference on Science and Technology (ICST), 2018 Crossref	20 words — 1%

9	Resti. "Handling the Dependence of Claim Severities with Copula Models", Journal of Mathematics and Statistics, 2010 Crossref	20 words — 1%
10	www.wseas.us Internet	16 words — 1%
11	doczz.net Internet	12 words — < 1%
12	Emmanuel Nii Martey, Nii Attoh-Okine. "Analysis of train derailment severity using vine copula quantile regression modeling", Transportation Research Part C: Emerging Technologies, 2019 Crossref	10 words — < 1%
13	ejurnal.mipa.unsri.ac.id Internet	9 words — < 1%
14	pdffox.com Internet	9 words — < 1%
15	González-López, V. A.. "Preface of the "Symposium on highlights in copula modeling"", AIP Conference Proceedings, 2013. Crossref	4 words — < 1%