ISSN 0973-1377 (Print)
ISSN 0973-7545 (Online)

International Journal of Applied Mathematics and Statistics

CESER Publications

योगः कर्मसु कौशलम्

Year 2015 Volume 53 Issue Number: 3

2015, Volume 53, Issue Number: 3

Table of Contents

Articles

 ISSN: 0973-7545

International Journal of
Applied Mathematics and Statistics™
ISSN: 0973-7545 (Online), ISSN 0973-1377 (Print)
www.ceser.in/ceserp/
www.ceserp.com/cp-jour/

International Journal of Applied Mathematics and Statistics™

The main aim of the **International Journal of Applied Mathematics and Statistics™** [ISSN 0973-1377 (Print); 0973-7545 (Online)] is to publish refereed, well-written original research articles, and studies that describe the latest research and developments in the area of applied mathematics and statistics. This is a broad-based journal covering all branches of mathematics, statistics and interdisciplinary research. ***International Journal of Applied Mathematics and Statistics (IJAMAS)*** is a peer-reviewed journal and published by **CESER Publications**.

Reviewed, Abstracted and **Indexed:** The **IJAMAS** is reviewed, abstracted and indexed by the **Mathematical Reviews; MathSciNet; Zentralblatt für Mathematik (Zentralblatt MATH); SCOPUS (Elsevier Bibliographic Databases); ERA: Excellence in Research for Australia (by Govt. of Australia); Statistical Theory and Method Abstracts (STMA-Z), Current Index to Statistics (CIS)** [The Current Index to Statistics (CIS) is a joint venture of the *American Statistical Association* & *Institute of Mathematical Statistics*, USA]**, International Abstracts in Operations Research (IAOR), Indian Science Abstracts; Academic keys; JournalSeek, Ulrich's Periodicals Directory, SCImago Journal & Country Rank, International Statistical Institute (ISI, Netherlands) Journal Index, IndexCopernicus, Getcited...**

 ISSN: 0973-7545

International Journal of Applied Mathematics and Statistics,
Int. J. Appl. Math. Stat.; Vol. 53; Issue No. 3; Year 2015,
ISSN 0973-1377 (Print), ISSN 0973-7545 (Online)

Trivariate Copula to Estimate Claim Costs: the Malaysian Motor Insurance Experience

Y.Resti[1] **, N.Ismail**[2] and **S S.H Jamaan**[2]

[1] Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sriwijaya
Jl. Raya Palembang-Prabumulih Km.32 Inderalaya 30662, Sumatera Selatan (Indonesia);
Email : farras_yani@yahoo.com

[2] Fakulti Sains dan Teknologi,Universiti Kebangsaan Malaysia,
43600 UKM Bangi, Selangor Darul Ehsan (Malaysia);
Email : ni@ukm.my

ABSTRACT

A road vehicle accident may produce three types of insurance claim 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 between insurance claim types can lead to over- or underestimated claims. This study proposes the application of trivariate copula from Archimedean family, namely Clayton, Frank and Gumbel, with regression marginals, namely Gamma and inverse Gaussian, to model dependencies among claim types in the Malaysian motor insurance claim costs data. The results of AIC and BIC from independent models indicate that Gamma regression is better than inverse Gaussian regression for all claim types. The results of AIC, BIC, plot of empirical versus fitted copula and plot of intermediate variable distribution versus intermediate variable from dependent models imply that Frank copula is the best model for accommodating dependences among insurance claim types. The simulation results imply that independent models for both TPBI and TPPD claims produce larger number of risk classes with overestimated claims.

Keywords: Dependence; trivariate copula; claim cost; claim type; SPLUS program.

Mathematics Subject Classification: 62J12, 62G99

Computing Classification System: I.4

1. INTRODUCTION

In actuarial practices, premium rating requires four basic elements agreed by most actuaries; to calculate fair premium rates so that high risk insureds pay higher premiums and vice versa, to provide sufficient funds for paying claims and expenses, to ensure safe contingency margins, and to produce reasonable returns. The second, third and fourth elements can be fulfilled by determining premium rates at overall levels, taking into account economic and other external factors such as inflations and government legislations, and involving only minimal statistical analysis. However, the first element requires the determination of premium rates at relative levels which involves risk classification procedures.

The purpose of risk classification is to divide insured risks into similar or homogeneous classes. The common methods studied by actuarial researchers can be written as regression models where the independent variables are risk or rating factors such as vehicle type, vehicle age, or insured's driving experience, and the dependent variable is either claim frequency (number of claims per exposure), or claim severity (average claim cost per claim), or loss ratio (claim cost per premium amount). This study focuses on risk classification of claim severity which is applied to motor insurance data.

In insurance practices, claim costs may give rise to multiple types. For the case of motor insurance experience, a vehicle crash may produces three types of dependent claims such as Third Party Bodily Injury (TPBI), Own Damage (OD) and Third Party Property Damage (TPPD). If there is more than one claim type, independent assumption among claim types may produce either over- or underestimated costs. Therefore, a copula model, which expresses the joint distribution of two or more random variables by separating the joint distribution into marginal distributions of individual variables and interdependent probability of individual variables, can be applied to model dependencies in claims data.

A copula model, which is a modeling tool in multivariate distributions, has become an increasingly popular approach for modeling dependences between multivariate random variables in many research areas. As examples, copula models were applied in financial studies for modeling asset allocations, credit scorings and default risks (Li 2000, Cherubini &Luciano 2001, Melchiori 2003, Micocci & Masala 2003), in biomedical researches for modeling correlated event times and competing risks (Meester & MacKay 1994, Andersen 2005). Further studies and applications of copula models can be found in (Genest & MacKay 1986, Frees & Valdez 1998).

Resti et al (2010) employed bivariate copula for estimating dependences in claim cost data, as well as Resti et al (2012). This article proposes trivariate copula from Archimedean family. In particular, we consider three types of Archimedean family copula, namely Frank, Clayton and Gumbel, for modeling claim cost dependencies and two types of regression models, namely Gamma and inverse Gaussian, for modeling claim cost marginals. The copula models are applied to the Malaysian motor insurance data.

2. MATERIALS AND METHODS

2.1. Fitting Marginal Regression Model

Most studies of claim cost modeling show that the claim cost distributions have positive ranges and are positively skewed. Based on these properties, Gamma and inverse Gaussian regression models have been fitted and several examples can be found in (McCullagh & Nelder 1989, Renshaw 1996, Ismail & Jemain 2006, Cheong et al. 2008).

Consider an insurance data with *n* risk classes, $i = 1, \cdots, n,$ and three claim types, $j = 1,2,3.$ Let C_{ij} be the random variable for claim severity (average claim cost per claim) in the i th risk class and j th claim type where the claim cost is represented by both paid and case estimates of outstanding, already adjusted and trended for inflation. For an easier illustration, we drop subscript j which indicates the j th claim type. If C_i is the random variable for claim severity and follows gamma regression model, the pdf is,

$$f(c_i \mid \mu_i, \omega) = \frac{1}{\Gamma(\omega)} \left(\frac{\omega}{\mu_i} \right)^{\omega} c_i^{\omega - 1} \exp\left(-\frac{\omega c_i}{\mu_i} \right), \qquad c_i > 0$$

with mean $E(C_i)=\mu_i$ and variance $Var(C_i)=\omega^{-1}\mu_i^2=\sigma^2\mu_i^2$, where $\omega=\sigma^{-2}$ denotes the scale parameter. If C_i is modeled as inverse Gaussian regression, the pdf is,

$$f(c_i|\mu_i,\sigma)=\frac{1}{\sigma\sqrt{2\pi c_i^3}}\exp\left(-\frac{1}{2c_i}\left[\frac{c_i-\mu_i}{\mu_i\sigma}\right]\right), \qquad c_i>0$$

with mean $E(C_i)=\mu_i$ and variance $Var(C_i)=\sigma^2\mu_i^3$, where σ denotes the scale parameter. The covariates for both regression models can be included via a log link $\mu_i=\exp\left(\sum_k \beta_k x_{ik}\right)=\exp(\mathbf{x_i^T\beta})$, where $\mathbf{\beta}$ is the vector of regression parameters and $\mathbf{x_i}$ the vector of explanatory variables. From the density function, log likelihoods for Gamma and inverse Gaussian regressions respectively are,

$$\log L(\mathbf{\beta},\omega)=\sum_{i=1}^{n}\left[(\omega-1)\log(c_i)-\log(\Gamma(\omega))+\omega\log(\omega)-\omega\log(\mu_i)-\frac{\omega c_i}{\mu_i}\right]$$

and

$$\log L(\mathbf{\beta},\sigma)=\sum_{i=1}^{n}\left[-\log(\sigma)-\frac{1}{2}\log(2\pi)-\frac{3}{2}\log(c_i)-\frac{1}{2c_i}\left(\frac{c_i-\mu_i}{\sigma\mu_i}\right)^2\right]$$

and the maximum likelihood estimates of $\mathbf{\beta}$ and ω (or σ) can be obtained by maximizing the log likelihoods.

2.3. Fitting Trivariate Copula

Consider the cdf of a trivarate distribution, F. The idea of Sklar's Theorem for a trivariate distribution is to represent F in two parts; marginal df, F_j, and copula df, H, which describes dependences between individual variables. Both F_j and H can be connected in a trivariate df,

$$F(c_1,c_2,c_3)=H(F_1(c_1),F_2(c_2),F_3(c_3))=H(u_1,u_2,u_3)$$

where U_1,U_2 and U_3 are standard uniform random variables. By differentiation, the density is,

$$f(c_1,c_2,c_3)=\frac{\partial^3 F(c_1,c_2,c_3)}{\partial c_1\partial c_2\partial c_3}=h(u_1,u_2,u_3)f_1(c_1)f_2(c_2)f_3(c_3)$$

where f_j is the marginal density and h the copula density.

This article proposes the application of Frank, Clayton and Gumbel copula, which belong to the Archimedean family, for modeling dependences among several claim types. For a trivariate case, an Archimedean copula can be constructed through a generator, φ,

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

where φ^{-1} is the inverse generator. A generator uniquely determines an Archimedean copula. The generator and inverse generator of Clayton, Frank and Gumbel copula with space parameter, α, respectively are $\phi_{Cl}(u)=u^{-\alpha}-1$, $\phi_{Cl}^{-1}(u)=(u+1)^{-\frac{1}{\alpha}}$, $\phi_{Fr}(u)=\ln\left(\frac{e^{-\alpha u}-1}{e^{-\alpha}}\right)$, $\phi_{Fr}^{-1}(u)=-\frac{1}{\alpha}\ln(1+e^{u}[e^{\alpha}-1])$, $\phi_{Gu}(u)=[-\ln(u)]^{\alpha}$ and $\phi_{Gu}^{-1}(u)=\exp\left[(-u)^{-\frac{1}{\alpha}}\right]$.

From equation above, the density of a trivariate Clayton copula is,

$$h(u_1,u_2,u_3)=\frac{\partial^3 H(u_1,u_2,u_3)}{\partial u_1\partial u_2\partial u_3}=(1+3\alpha+2\alpha^2)(u_{i1}u_{i2}u_{i3})^{-(\alpha+1)}(u_{i1}{}^{-\alpha}+u_{i2}{}^{-\alpha}+u_{i3}{}^{-\alpha}-2)^{-\left(\frac{1}{\alpha}+3\right)}$$

The trivariate Frank and Gumbel copula can also be derived in a similar manner.

2.4. Maximum Likelihood Estimation

Consider a dataset of insurance claims where the costs are divided into three types. Assume that the three types of claim costs are C_{i1}, C_{i2} and C_{i3}, and each claim type is modeled by Gamma regression with pdf. Assume that the dependencies among C_{i1}, C_{i2} and C_{i3} are fitted through a trivariate Clayton copula with density (12). In addition, let $\boldsymbol{\beta}=(\boldsymbol{\beta}_1^{\mathbf{T}},\boldsymbol{\beta}_2^{\mathbf{T}},\boldsymbol{\beta}_3^{\mathbf{T}})^{\mathbf{T}}$ be the vector of regression parameters where $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_3$ are regression vectors from C_{i1}, C_{i2} and C_{i3}, $\boldsymbol{\omega}=(\boldsymbol{\omega}_1,\boldsymbol{\omega}_2,\boldsymbol{\omega}_3)^{\mathbf{T}}$ the vector of scale parameters where $\boldsymbol{\omega}_1,\boldsymbol{\omega}_2$ and $\boldsymbol{\omega}_3$ are scale parameters from claim costs C_{i1}, C_{i2} and C_{i3}, and α the Clayton copula parameter. The maximum likelihood estimates can be obtained by maximizing the likelihood function of the density,

$$L(\boldsymbol{\beta},\boldsymbol{\omega},\alpha)=\prod_i f(c_{i1},c_{i2},c_{i3};\boldsymbol{\beta},\boldsymbol{\omega},\alpha)$$

$$=\prod_i [f(c_{i1};\boldsymbol{\beta}_1,\omega_1)f(c_{i2};\boldsymbol{\beta}_2,\omega_2)f(c_{i3};\boldsymbol{\beta}_3,\omega_3)h_{Cl}(u_{i1},u_{i2},u_{i3};\alpha)]$$

or the log likelihood,

$$\begin{aligned}\ell(\boldsymbol{\beta},\boldsymbol{\omega},\alpha)&=\sum_{i=1}^n\log(f(c_{i1};\beta_1,\omega_1))+\sum_{i=1}^n\log(f(c_{i2};\beta_2,\omega_2))+\sum_{i=1}^n\log(f(c_{i3};\beta_3,\omega_3))+\sum_{i=1}^n\log(h_{Cl}(u_{i1},u_{i2},u_{i3};\alpha))\\
&=\sum_{i=1}^n\left[-\log(\Gamma(\omega_1)+(\omega_1-1)\log(c_{i1}))+\omega_1\log(\omega_1)-\omega_1\log(\mu_{i1})-\frac{\omega_1 c_{i1}}{\mu_{i1}}\right]\\
&+\sum_{i=1}^n\left[-\log(\Gamma(\omega_2)+(\omega_2-1)\log(c_{i2}))+\omega_2\log(\omega_2)-\omega_2\log(\mu_{i2})-\frac{\omega_2 c_{i2}}{\mu_{i2}}\right]\\
&+\sum_{i=1}^n\left[-\log(\Gamma(\omega_3)+(\omega_3-1)\log(c_{i3}))+\omega_3\log(\omega_3)-\omega_3\log(\mu_{i3})-\frac{\omega_3 c_{i3}}{\mu_{i3}}\right]\\
&+\sum_{i=1}^n\left[\log(1+3\alpha+2\alpha^2)-(\alpha+1)(\log(u_{i1})+\log(u_{i2})+\log(u_{i3}))-\left(\frac{1}{\alpha}+3\right)\log(u_{i1}^{-\alpha}+u_{i2}^{-\alpha}+u_{i3}^{-\alpha}-2)\right]\end{aligned}$$

The maximum likelihood estimates for trivariate Frank and Gumbel copula can also be obtained in a similar manner.

In our study, the trivariate copula models are fitted using *SPLUS program* with optimization procedure (*nlm* or *optim* functions). To provide better convergence, the estimates of regression models from the independent assumption are used as initial values. The variance of parameters can be estimated using diagonal elements of the inverse of negative Hessian matrix. For the case of a Clayton copula with gamma regression marginals, the elements of Hessian matrix contain the second derivatives of log likelihood in (13) with their corresponding parameters.

2.5. Goodness of Fit Measures

Several goodness of fit measures can be used for the Archimedean family copula such as log likelihood, Akaike Information Criteria (AIC), Schwartz Bayesian Information Criterion (BIC), plot of empirical copula, H_e, versus fitted copula, H_s, and plot of intermediate variable distribution, $K(z)$, versus intermediate variable, z. Let n be the number of observations, m the number of estimated parameters and ℓ the log likelihood. The AIC and BIC can be calculated respectively as $AIC = -2\ell + 2m$ and $BIC = -2\ell + m\ln(n)$. For the plot of H_e versus H_s, the empirical copula can be defined as,

$$H_e(u_{i1},\cdots,u_{i3}) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{I}\left(\frac{R_{i1}}{n+1} \le u_{i1},\cdots,\frac{R_{i3}}{n+1} \le u_{i3}\right)$$

where $u_{i1},\cdots,u_{i3} \in (0,1)$ and $\mathbf{I}(.)$ denotes the indicator function, while $R_{i1},\cdots,R_{i3}$ respectively denote the ranks of $c_{i1},\cdots,c_{i3}$. For the plot of $K(z)$ versus z, let $Z_i = F(C_{i1},C_{i2},C_{i3})$ be the intermediate variable of C_{i1},C_{i2} and C_{i3} which is defined as,

$$Z_i = \frac{\text{number}\{(C_{1i},C_{2i},C_{3i})\ \text{until}\ (C_{1j} < C_{1i}, C_{2j} < C_{2i}, C_{3j} < C_{3i})\}}{(n-1)}$$

where $i \neq j$. The df, $K(z) = P(Z \le z)$, for Clayton, Frank and Gumbel copula respectively are [22],

$$K_{Cl}(z) = \frac{z(1+\alpha - z^{\alpha})}{\alpha}\ ,\ K_{Fr}(z) = \left(z + \frac{e^{-\alpha z}-1}{\alpha e^{-\alpha z}}\right)\log\left(\frac{e^{-\alpha z}-1}{e^{-\alpha}-1}\right) \text{ and } K_{Gu}(z) = \frac{z(\alpha - \log(z))}{\alpha}\ .$$

A better copula model is indicated by a larger log likelihood, and a smaller AIC and BIC. For the plots of H_e versus H_s and $K(z)$ versus z, a better model is indicated by a closer distance between the points and the diagonal line. The closeness of the points $(b_1,b_2,\cdots,b_n)$ with the diagonal line for H_e versus H_s plot can be measured using a quadratic distance,

$$q_d(H_e,H_s) = \left[\sum_{i=1}^{n}(H_e(b_i) - H_s(b_i))^2\right]^{1/2}$$

A similar way can also be used to calculate q_d for $K(z)$ versus z plot.

2.6. Simulation Procedure

Assuming the claim types are independent, the predicted claim costs can be simulated separately for each claim type using Gamma or inverse Gaussian regression models. Let c_{lij}, $l = 1,2,\cdots,s,$ be the simulated claim cost for the *i*th risk class and *j*th claim type, where s is the number of iterations. The claim cost can be predicted as $\tilde{c}_{ij} = \frac{1}{s}\sum_{l} c_{lij}$ .If the claim types are assumed dependent, the predicted claim costs can be simulated using the Archimedean family copula with regression marginals. Let $(c_{li1},c_{li2},c_{li3})$, $l = 1,2,\cdots,s,$ be the simulated claim costs. The claim costs for all claim types are predicted as $(\tilde{c}_{li1},\tilde{c}_{li2},\tilde{c}_{li3}) = \frac{1}{s}\sum_{l}(c_{li1},c_{li2},c_{li3})$.

3. RESULTS AND DISCUSSIONS

The data for private car insurance portfolios compiled from ten general insurance companies in Malaysia for 2000-2003 is considered. The sample data, which is supplied by Insurance Services Malaysia (ISM) Berhad, contains 1,009,175 policies with 117,586 (or 9.7%) claims. The claims can be categorized into three types namely Own Damage (OD), Third Party Property Damage (TPPD) and Third Party Bodily Injury (TPBI), while the rating factors are vehicle make, vehicle cubic capacity (cc) and vehicle age. Further information on the rating factors and classes are shown in Table 1.

Table 1 : Rating factors and classes

Rating factors	Rating classes
Vehicle age	0-1 year 2-3 years 4-5 years 6-7 years 8+ years
Vehicle cubic capacity	0-1000 cc 1001-1300 cc 1301-1500 cc 1501-1800 cc 1801+ cc
Vehicle make	Local Foreign

The 3D scatterplots for TPBI, OD and TPPD claims (in log scale) are shown in Figure 1. It can be observed that the TPBI, OD and TPPD claims have a moderately strong trivariate relationship.

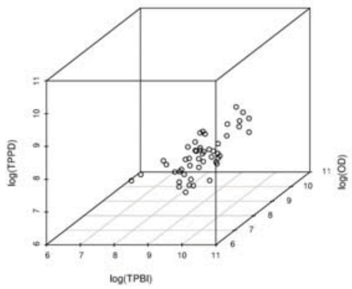

Figure.1 Scatterplots (in log scale) for TPBI, OD, TPPD

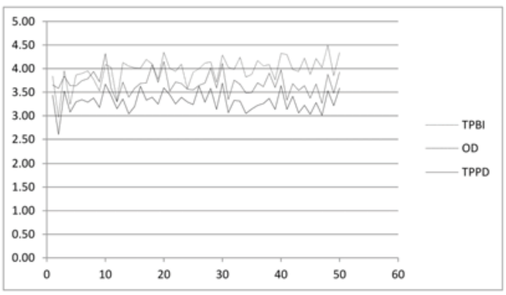

Figure 2. Claim severities (in log scale) for all claim types

Figure 2 provides the line plots for TPBI, OD and TPPD claims (in log scale) which are plotted according to their respective risk classes. It can be seen that the bivariate dependences between TPBI-OD, TPBI-TPPD and OD-TPPD claims are similar in both low and high severities, indicating possible applications of Frank, Clayton or Gumbel copula from the Archimedean family that has a single space parameter, α, for explaining the dependences among the three claim types. In particular, Clayton copula assumes stronger dependences among low severities and almost no dependence among high severities, whereas Gumbel copula assumes stronger dependences among high severities and almost no dependence among low severities. The Frank copula assumes symmetric dependence structures where both high and low severities exhibit approximately the same dependences.

Table 2 : Independent models for claim severity

Parameter	TPBI		OD		TPPD	
	est.	*p*-value	est.	*p*-value	est.	*p*-value
	Gamma regression model					
Intercept	9.40	0.00	7.96	0.00	7.53	0.00
0-1 yr	-0.64	0.00	0.28	0.01	-	-
4-5 yr	-	-	-	-	0.26	0.06
1501-1800 cc	-	-	0.51	0.00	-	-
1800+ cc	-	-	0.72	0.00	-	-
ω foreign	-	-	0.54	0.00	0.24	0.04
scale	2728.99	0.00	616.36	0.00	418.33	0.00
Log Likelihood	-495.93		-443.45		-411.05	
AIC	997.87		898.90		830.10	
BIC	1003.60		910.37		837.75	
	Inverse Gaussian regression model					
Intercept	9.40	0.00	8.17	0.00	7.52	0.00
0-1 yr	-0.64	0.00	-	-	-	-
1501-1800 cc	-	-	0.41	0.00	-	-
1800+ cc	-	-	0.59	0.01	-	-
foreign	-	-	0.34	0.00	0.35	0.01
ω, scale	0.46	0.00	0.03	0.00	0.00	0.00
Log Likelihood	-506.51		-445.43		-412.49	
AIC	1019.03		900.86		830.59	
BIC	1024.76		910.42		838.32	

Table 2 shows the parameters and *t*-ratios for Gamma and inverse Gaussian regression models which are fitted under the assumption of independent claim types, i.e. TPBI, OD and TPPD claims are fitted separately to the regression models. The results indicate that different regression models have either the same or different significant factors. As an example, vehicle age 0-1 year is the only significant factor for both Gamma and inverse regression models in TPBI claims, whereas in TPPD claims, vehicle age 4-5 years is a significant factor for Gamma regression but not for inverse Gaussian regression. For each claim type, the fitted claim cost is calculated as $\hat{c}_i = \exp\left(\sum_k \beta_k x_{ik}\right)$, where β_k is the regression parameters and x_{ik} the explanatory variables with values of zero or one. As an example, the fitted TPBI claims for vehicles age 0-1 year from Gamma regression model is $\hat{c}_i = \exp(9.40 - 0.64(1)) = \text{RM } 6,374.$ The results in Table 2 can also be used to compare the relative risk of each rating factor, and therefore, identifying low or high risk vehicles. As an example, the fitted TPBI claim for vehicles age 2+ years from Gamma regression model is $\hat{c}_i = \exp(9.40 - 0.64(0)) = \text{RM } 12,088,$ indicating that new vehicles (age 0-1 year) have lower risks than older vehicles (age 2 years and above). The AIC and BIC for all claim types imply that Gamma regression is a better model than inverse Gaussian regression. Assuming that the claim types are dependent, the Archimedean family copula (Clayton, Frank and Gumbel) are then fitted to TPBI, OD and TPPD claims using Gamma regression as marginals.

Table 3 : Dependent models for claim severity

Claim types	Parameters	Clayton Copula		Frank Copula		Gumbel Copula	
		est.	*p*-value	est.	*p*-value	est.	*p*-value
TPBI	Intercept	9.37	0.00	9.38	0.00	9.43	0.00
	0-1 yr	-0.34	0.00	-0.63	0.00	-0.73	0.00
	v, scale	3005.36	0.00	3047.32	0.00	2924.24	0.00
OD	Intercept	8.09	8.09	8.21	8.21	8.18	0.00
	0-1 yr	0.38	0.38	0.19	0.19	0.24	0.01
	1501-1800 cc	0.24	0.24	0.26	0.26	0.26	0.00
	1800+ cc	0.50	0.50	0.40	0.40	0.40	0.00
	foreign	0.50	0.50	0.35	0.35	0.35	0.00
	v, scale	675.59	0.00	773.11	0.00	773.11	0.00
TPPD	Intercept	7.59	0.00	7.57	0.00	7.57	0.00
	4-5 yr	0.20	0.04	0.19	0.19	0.23	0.01
	foreign	0.19	0.03	0.17	0.02	0.24	0.00
	v, scale	482.44	0.00	470.14	0.00	484.16	0.00
α , copula		1.33	0.00	6.86	0.00	1.93	0.00
Log likelihood		-1322.14		-1314.99		-1322.57	
AIC		2672.29		2657.97		2673.15	
BIC		2699.06		2684.74		2699.91	

Table 3 provides the parameters and *t*-ratios for Clayton, Frank and Gumbel copula models. Based on the results, the following observations are obtained:

i. Each copula model provides different parameter estimates.
ii. All copula models produce significant copula parameters, i.e. the *p*-values for α are 0.00.
iii. Frank copula has the largest log likelihood compared to Clayton and Gumbel copula.
iv. Frank copula has the smallest AIC and BIC compared to Clayton and Gumbel copula.

The goodness-of-fit of the fitted copula can also be illustrated using the q-q plots of H_e versus H_s and $K(z)$ versus z. The plots and quadratic distance (qd) measures are shown in Table 4 and Figures 3-4 respectively.

Table 4 : quadratic distance measures

Copula	H_e versus H_s	*K(z)* vs. z
Clayton	4.20	1.13
Frank	1.31	0.68
Gumbel	3.83	1.18

The results indicate that Frank copula has the smallest q_d in both plots. Therefore, based on log likelihood, AIC, BIC, and q_d from both H_e versus H_s plot and $K(z)$ versus z plot, the best model for handling dependences among the three claim types is Frank copula. This result also indicates that there are symmetric dependences among TPBI, OD and TPPD claims at both low and high costs. The results in Table 3 can also be used to compare the relative risks of each rating factor, and therefore, identifying low or high risk vehicles. As an example, the fitted TPBI claim from Frank copula for vehicles age 0-1 year and 2+ years respectively are $\tilde{c}_i = \exp(9.38 - 0.63(1)) = \text{RM } 6{,}311$ and $\tilde{c}_i = \exp(9.38 - 0.63(0)) = \text{RM } 11{,}849$, also indicating that new vehicles (age 0-1 year) have lower risks than older vehicles (age 2+ years). However, the difference between independent and dependent (Frank copula) models is that the low risk vehicles in dependent model have lower claims $[\exp(8.75)]$ than independent model $[\exp(8.76)]$. The high risk vehicles in dependent model also have lower claims $[\exp(9.38)]$ than independent model $\exp[(9.40)]$.

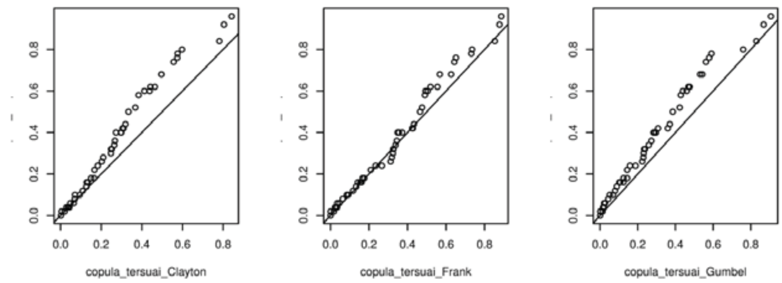

Figure 3. Empirical vs. fitted Clayton, Frank and Gumbel copula

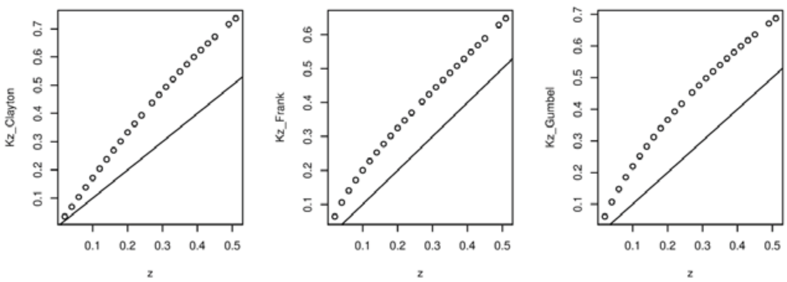

Figure 4. *K(z)* vs. *z* of Clayton, Frank and Gumbel copula

The simulation results for independent and dependent (Frank copula) models are provided in Table 5. For each claim type, the independent model produces several classes of both over- and underestimated claims. As an example, consider the first rating class which consists of 0-1 year, 0-1000 cc and local vehicles. Compared to the dependent model, the independent model produces overestimated TPBI claims with difference and ratio of RM139 and 1.02 respectively, but underestimated OD claims with difference and ratio of -RM611 and 0.86 respectively. As for TPPD claims, the independent model also produces underestimated claim with –RM82 difference and 0.96 ratio. In general, both TPBI and TPPD claims produce more classes of overestimated claims (44 TPBI and 37 TPPD rating classes) than underestimated claims (6 TPBI and 13 TPPD rating classes). On the contrary, OD claims produce almost equal number of classes for under- and overestimated claims (23 underestimated and 22 overestimated rating classes).

4. CONCLUSION

In real practices, independent assumptions between motor insurance claim types may not be true for all cases as an incidence of vehicle crash may resulted in more than one claim types and an incurred claim from one type may has an impact on the incurrence of a claim from another type. This article proposes the trivariate copula from Archimedean family, namely Clayton, Frank and Gumbel, for estimating dependences among claim types. The independent models are investigated by fitting Gamma and inverse Gaussian regression models separately to all claim types. The resulting parameters are then used as initial values for estimating the dependent copula models. Based on the results of AIC and BIC of independent model, Gamma regression is better than inverse Gaussian regression for all claim types. Therefore, Gamma regressions are used as marginals for fitting Clayton, Frank and Gumbel copula in the dependent models. Based on the results of AIC, BIC, plot of empirical versus fitted copula, and plot of *K(z)* versus *z*, Frank copula is the best model for accommodating dependences among claim types. This result also indicates the existence of symmetric dependences in both low and high claims among claim types for the Malaysian motor insurance data. The simulation results imply that both TPBI and TPPD claims produce more classes of overestimated claims, whereas OD claims produce almost equal number of classes for under- and overestimated claims. For future study, a similar approach can be applied to other lines of insurance or any other costs data (besides insurance) with covariates, as long as information on marginal data and their related covariates are available. Further applications can also be performed to other copula functions for modeling the dependences, or other regression models for modeling the marginals.

5. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support received in the form of research grants (GUP-2012-024 and LRGS/TD/2011/UKM/ICT/03/02) from the Ministry of Higher Education (MOHE), Malaysia. The authors are also pleasured to thank Insurance Services Malaysia Berhad (ISM) for supplying the data.

6. REFERENCES

Andersen, A.W., 2005. Two-stage estimation in copula models used in family studies. *Lifetime Data Analysis* 11: 333–350.

Cheong, P.W., Jemain, A.A. & Ismail, N., 2008. Practice and pricing in non-life insurance: the Malaysian experience. *Journal of Quality Measurement and Analysis* 4(1): 11-24.

Cherubini, U., Luciano, E., 2001. Value at risk trade off and capital allocation with copulas. *Economic Notes, Review of Banking, Finance and Monetary Economics* 30(2): 235-256. DOI: 10.1111/j.0391-5026.2001.00055.x

Frees, E.W., Valdez, E.A., 1998. Understanding relationships using copulas, North *American Actuarial Journal* 2(1): 1-25.

Genest, C., MacKay, J., 1986. The joy of copulas: Bivariate distributions with uniform marginals, *American Statistics* 40: 280-285.

Ismail, N., Jemain, A.A., 2006. A comparison of risk classification methods for claim severity data. *Journal of Modern Applied Statistical Methods* 5(2): 513-528.

Li, D.X., 2000. On default correlation: a copula function approach. *The Risk Metrics Group, Working Paper* 2000; 99-07.

McCullagh, P., Nelder, J.A., 1989. *Generalized Linear Models*. Ed.2nd. Chapman and Hall: London.

Meester, S.G., MacKay, J., 1994. A parametric model for cluster correlated categorical data. *Biometric* 1994; 50: 954-963.

Melchiori, R.M., 2003. *Which Archimedean copula is the right one?* 2003. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1123135 [22 January 2010]

Micocci M., Masal, G., 2003. Pricing pension funds guarantees using a copula approach. *Proceedings of the AFIR International Congress*. Maastricht, Netherlands, 17-19 September.

Renshaw, A.E., 1994. Modeling the claims process in the presence of covariates. *ASTIN Bulletin* 24(2): 265-285.

Resti, Y., Ismail, I. & Jaaman, S.H., 2010. Handling the dependence of claim severities with copula models. *Journal of Mathematics and Statistics* 6(2): 136-142.

Resti, Y., Ismail, I. & Jaaman, S.H., 2012. Mathematical modeling for claim severities using normal and t copulas. *International Journal of Applied Mathematics and Statistics* 27(3): 8-19.

Table 5. Simulated Claim Costs

Independent Model (IM)			Dependent Model (DM)			Difference (IM-DM)			Ratio (IM/DM)		
TPBI	OD	TPPD	TPBI	OD	TPPD	TPBI	OD	TPPD	TPBI	OD	TPPD
6415	3830	1864	6276	4441	1946	139	-611	-82	1.02	0.86	0.96
6299	6517	2357	6261	6299	2292	38	218	65	1.01	1.03	1.03
6318	3820	1845	6344	4445	1948	-26	-625	-103	1.00	0.86	0.95
6401	6488	2356	6234	6318	2302	167	170	54	1.03	1.03	1.02
6405	3784	1868	6313	4447	1935	92	-663	-67	1.01	0.85	0.97
6347	6473	2361	6279	6292	2314	68	181	47	1.01	1.03	1.02
6328	6313	1861	6308	5752	1951	20	561	-90	1.00	1.10	0.95
6388	10844	2377	6242	8193	2305	146	2651	72	1.02	1.32	1.03
6345	7777	1858	6286	6617	1941	59	1160	-83	1.01	1.18	0.96
6425	13317	2390	6371	9425	2290	54	3892	100	1.01	1.41	1.04
12152	2853	1869	11824	3664	1923	328	-811	-54	1.03	0.78	0.97
12045	4959	2370	11827	5237	2302	218	-278	68	1.02	0.95	1.03
12068	2863	1852	11840	3683	1944	228	-820	-92	1.02	0.78	0.95
12170	4902	2363	11689	5233	2269	481	-331	94	1.04	0.94	1.04
12070	2893	1862	11779	3669	1951	291	-776	-89	1.02	0.79	0.95
12090	4898	2386	11876	5212	2303	214	-314	83	1.02	0.94	1.04
12097	4754	1855	11828	4802	1947	269	-48	-92	1.02	0.99	0.95
12008	8174	2372	11945	6747	2296	63	1427	76	1.01	1.21	1.03
12109	5921	1859	11797	5486	1952	312	435	-93	1.03	1.08	0.95
12198	10093	2361	11712	7765	2307	486	2328	54	1.04	1.30	1.02
12087	2858	2406	11786	3678	2352	301	-820	54	1.03	0.78	1.02
12121	4917	3057	11926	5237	2771	195	-320	286	1.02	0.94	1.10
12042	2837	2411	11929	3689	2338	113	-852	73	1.01	0.77	1.03
12095	4932	3090	11898	5230	2793	197	-298	297	1.02	0.94	1.11
11997	2854	2409	11814	3697	2367	183	-843	42	1.02	0.77	1.02
11996	4922	3075	11885	5216	2782	111	-294	293	1.01	0.94	1.11
12199	4776	2414	11760	4783	2346	439	-7	68	1.04	1.00	1.03
12140	8216	3068	11781	6777	2778	359	1439	290	1.03	1.21	1.10
12156	5879	2411	11789	5486	2340	367	393	71	1.03	1.07	1.03
12133	10079	3084	11855	7770	2787	278	2309	297	1.02	1.30	1.11
12185	2873	1867	11899	3669	1941	286	-796	-74	1.02	0.78	0.96
12014	4934	2369	11733	5205	2301	281	-271	68	1.02	0.95	1.03
12146	2849	1864	11941	3676	1944	205	-827	-80	1.02	0.78	0.96
12112	4932	2382	11809	5211	2315	303	-279	67	1.03	0.95	1.03
12013	2855	1860	11831	3660	1938	182	-805	-78	1.02	0.78	0.96
12019	4914	2373	11948	5196	2321	71	-282	52	1.01	0.95	1.02
12080	4757	1854	12149	4641	146	-69	116	1708	0.99	1.02	12.70
12132	8200	2375	11948	7252	165	184	948	2210	1.02	1.13	14.39
12087	5864	1850	12162	5379	83	-75	485	1767	0.99	1.09	22.29
12136	10094	2378	12087	8343	394	49	1751	1984	1.00	1.21	6.04
12018	2856	1874	12108	3343	82	-90	-487	1792	0.99	0.85	22.85
12215	4884	2365	12015	5235	97	200	-351	2268	1.02	0.93	24.38
12029	2881	1872	12020	3375	572	9	-494	1300	1.00	0.85	3.27
12196	4918	2378	12055	5203	230	141	-285	2148	1.01	0.95	10.34
12061	2871	1857	12075	3329	1081	-14	-458	776	1.00	0.86	1.72
12105	4934	2367	12159	5235	323	-54	-301	2044	1.00	0.94	7.33
12077	4777	1854	12180	4670	155	-103	107	1699	0.99	1.02	11.96
11988	8174	2371	12177	7284	344	-189	890	2027	0.98	1.12	6.89
12088	5850	1868	12130	5388	28	-42	462	1840	1.00	1.09	66.71
11972	10046	2383	12140	8317	880	-168	1729	1503	0.99	1.21	2.71