Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/120317

Dependence Between Report Lag and Claim Amounts in Property and Casualty Insurance

N. Ahmed* and L. Farid

Insurance and Actuarial Science Department, Faculty of Commerce, Cairo University, Giza, Egypt

Received: 25 Nov. 2022, Revised: 4 Jan. 2023, Accepted: 15 Jan. 2023.
Published online: 1 Sep. 2023.

Abstract: An essential tool for comprehending the variable of interest is possessing a variable that can interpret the behavior of another. The concept of dependency has been intensively researched throughout the years, with numerous models. Copulas are a comprehensive tool for simulating the interdependence of the variables. They provide alternative interpretations of the linear and non-linear relationship between associated random variables and their marginal. In Insurance sector, one of the most risks that the insurers face is holding inefficient provision amounts for claims. This paper explains the dependence structure between the claim amount and the report lag period for claims in the insurance companies.

Keywords: Report lag, claim amounts, copula model, loss provision, property and casualty insurance.

1 Introduction

In property and casualty insurance companies, the loss provision is a significant part of the financial statements because it represents the largest liability of the insurer's balance sheet. Insurance companies are required to maintain an adequate loss provision, which is a sufficient fund to settle all the outstanding claims. The following figure summarizes the claims procedure for the insurance companies. When an accident occurs, the policyholder reports the incident to the insurer who accordingly settles the payments required. Additionally, the claim file will be closed if the insurer determines that no additional payments are necessary for this claim. While a single claim may be subject to many payments [1, 2].

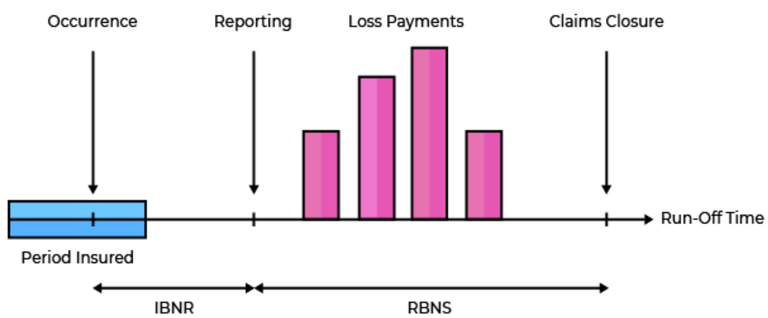

***Fig** 1: process of claim settlement*

Claim characteristics describe the speed of which claims occur, are reported, are settled and are occasionally revisited. If payment of claims is made by the insurer whenever a claim event takes place [3]. If the policyholder is involved in a car

*Corresponding author e-mail: Nourhan.hasan@foc.cu.edu.eg or Nourhan.hassan.90@hotmail.com

accident, for instance, payment is transferred directly to the accounts of the affected parties. In this case, the insurer has no potential liability for outstanding claims [4].

Insurers have outstanding claims liabilities due to the time-lag between the occurrence of the event and the reporting of the claim to the insurer. This lag is called the reporting lag or reporting delay [5]. While, the time between the reporting and the closure of the claim is called the settlement delay. During this period, the insurer determines which amounts should be paid to the insured. Loss provision or provision for outstanding claims includes the outstanding reported claims, which is also known as Reported but not Settled (RBNS), and the Incurred but not Reported (IBNR). The RBNS indicates the total amount of paid loss that will be needed to settle all reported claims, excluding any payments previously made on those claims [6]. On the other hand, The IBNR indicate the amount of paid loss that will be required when claim event has occurred, but the claim has not yet been reported to the insurer.

$$Total\ loss\ provision = IBNR + RBNS$$

1.1 Research problem

A provision in insurers is the sum of money kept aside by the company to cover unforeseen risks like claim occurrences or policy cancellations. The IBNR is one of the key provisions that an insurance company must maintain. These provisions are kept covering any claims that might have happened but haven't yet been reported to the insurer. Two aspects must be considered in order to decide how much money should be set aside for the IBNR: the length of time it takes to report a claim and the amount of the claims. Time is a crucial consideration when estimated IBNR claims since it is essential to figure out how much time will have passed before the claim is reported [7].

1.2 Objective of the paper

The objective of this paper is to understand the reporting lag function of property and casualty insurance claims; in order to establish a relationship between delay it takes for an insured party to report a claim and the claim amount, and ascertain an effective copula that explains the dependence structure between them.

1.3 Research Methodology

This paper proposes a copula function. Copula is a Latin word which means a link, where it is a function that links a multidimensional distribution to its one-dimensional margins [8]. Copulas are a comprehensive tool to model the dependency between runoff triangles within the line of business and with different lines of business [9]. They provide alternative interpretations of the linear and non-linear relationship between associated random variables and their marginal.

It allows combining marginal distributions and decomposing joint distributions into marginal and dependence structure. It is a multivariate distribution function with uniform marginal densities $[0, 1]^n$ to $[0,1]$, that has the same properties as a joint cumulative distribution. It is a function that links a multidimensional distribution to its one-dimensional margins [10]. If F is a n-dimensional cumulative joint function with margins $F_{(1)}, F_{(2)},..., F_{(n)}$. And all margins are continuous, and then the joint distribution of n random variables $(y_{1,} y_{2,...,} y_n)$ can be represented by a unique copula function [11]:

$F(y_{1,} y_{2,...,} y_n) = C(F_{(1)}, F_{(2)},..., F_{(n)})$,

$u_i = F_i(y_i)$, $i = 1,\ldots,n$

Copula has three families; Elliptical (Normal Gaussian) Copula, t-student copula and Archimedean Copula. [12]. An Archimedean copula is constructed through a generator, a convex, and decreasing function ϕ with domain (0, 1] and range [0, ∞) such that [13]

$\phi(1) = 0$ and $\phi(0) = \infty$

$C(u_1, \ldots, u_n) = \phi^{-1}(\phi(u_1) + \cdots + \phi(u_n))$

Where;

$C(u)$ → the density function of copula

ϕ ⇒ the generator of Copula

There are three types of Archimedean copulas: the Clayton, Frank and Gumbel. [14].

Clayton copula

The Clayton copula is an asymmetric Archimedean copula, it demonstrates more dependency in the negative tail than in the positive tail. This copula is given by:

$$C(u_1, u_2; \theta) = \max\left((u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}}, 0\right)$$

θ ⇒ is the dependence parameter.

Gumbel Copula

The Gumbel copula is an asymmetric Archimedean copula that shows more dependence in the positive tail than in the negative. This copula is given by [5]:

$$C(u_1, u_2) = \exp\left(-\left[(-\ln u_1)^{\theta} + (-\ln u_2)^{\theta}\right]^{\frac{1}{\theta}}\right)$$

Frank Copula

The Frank copula is a symmetric Archimedean copula given by:

$$C(u_1, u_2) = \frac{-1}{\theta}\ln\left(1 + \frac{\left(e^{-\theta u_1} - 1\right)\left(e^{-\theta u_2} - 1\right)}{e^{-\theta} - 1}\right)$$

Normal Gaussian (Elliptical) Copula:

The Elliptical copula is a multivariate function of normal marginal distributions, assuming Y =($Y_{1,}\ Y_{2\,,\ldots,}\ Y_n$) is multivariate normal [15]

$$C(u_1, u_2; \rho) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} exp\left[\frac{r^2 - 2\rho rs + s^2}{2(1-\rho^2)}\right] drds$$

Where,

$\rho \Rightarrow$ Pearson correlation coefficient between marginals

$\Phi^{-1} \Rightarrow$The inverse of standard normal

T-student copula:

The t student copula is a function of t distributions, assuming Y =($Y_{1,}\ Y_{2\,,\ldots,}\ Y_n$) is multivariate t student marginals, where R is the correlation matrix between marginals, and y is a random variable with χ^2 distribution [15]:

$$C(u_1, u_2; \rho) = \int_{-\infty}^{t_v^{-1}(u_1)} \int_{-\infty}^{t_v^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} exp\left[1 + \frac{r^2 - 2\rho rs + s^2}{v(1-\rho^2)}\right]^{\frac{-(v+2)}{2}} drds$$

$t_v^{-1} \Rightarrow$The inverse of uni-variate student t-distribution function

$v \Rightarrow$ Degree of freedom $\rho \Rightarrow$ Pearson correlation coefficient between marginal

2. Materials and Methods

The copulas were fitted using the inference for margins technique. This technique requires fitting each marginal distribution individually using the maximum likelihood method and then utilizing the fitting results to fit the copula distributions. The Akaike Information criterion (AIC) was used to assess the effectiveness of the various copulas in interpreting the dependency between the variables, the copula with the lowest AIC value was considered to be the most effective.

This technique was applied to data from a property and casualty insurance company operating in Egypt, which included date of loss, reported date and amount per claim information. From this data set we were able to calculate report lag by determining number of days between when claim was incurred and when it was reported. The amount considered claim amount paid. The data statistical results for the two variables under consideration; report delay and claim amount were as follows;

Table 1: *descriptive statistics of the data*

	Report Lag	Claims Amount
Mean	2.654	6504
Standard Deviation	2.31343	5838.319
Skewness	1.28	1.33
Kurtosis	1.11	1.24
Sample Size	8441	8441
Median	2	4775

Minimum	0	27
Maximum	10	26464
Range	10	26437
Standard Error	0.03	63.55

Pearson's Correlation Coefficient = 0.04106869
Kendall's Tau = 0.03495474
Spearman's Rho = 0.04848236

2.1 Fitting Marginal Distributions

This step is to fit the marginal distributions for two variables: report lag and claim amounts

Report Lag Distribution

The report delay was interpreted as a discrete distribution in terms of days. Regarding this, the negative binomial distribution and the geometric distribution are the best fit discrete distributions; based on easy-fit software.

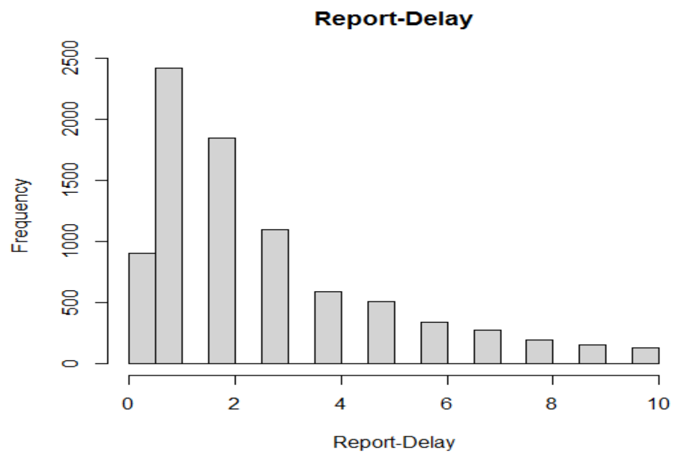

Fig.2: *illustration for report delay data*

The negative binomial distribution

Using the maximum likelihood method to fit the negative binomial distribution to the report delay data, the following results were obtained;

Parameters:

Table 2: *Results for the negative binomial fit to the report delay variable*

	Estimate	**Std. Error**
Size	2.892992	0.09522284
Mu	2.653856	0.02455322

The negative binomial density curve superimposed over the histogram of the observed data to highlight how the negative binomial distribution fitted on the data as shown in the following graph:

The Geometric distribution

Fitting the geometric distribution through the maximum likelihood method, the following results were obtained.

Parameters:

Table 3: Results for the geometric distribution fit to the report delay variable

	Parameter estimate	Std. Error
Probability	0.2736941	0.002538773

The following figure represent graphical illustration for both distributions fitted on report delay data where the negative binomial is black, and the geometric distribution is red

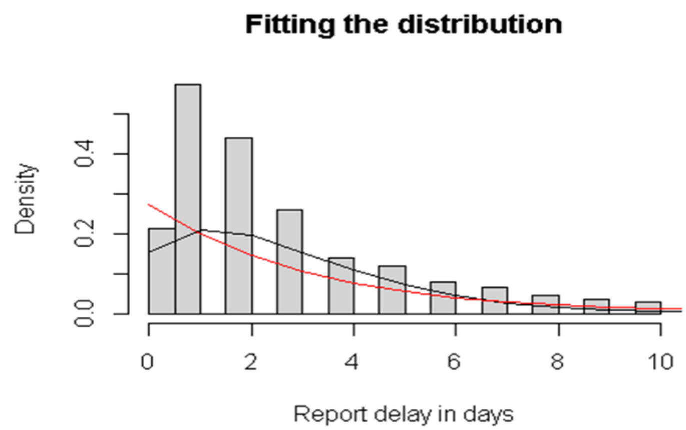

Fig.3: *Graphical illustration for the negative binomial and geometric distributions fit to the report lag variable.*

Report delay distribution marginal fit summary
The following table compares the log likelihoods, AIC, and BIC for the three distributions:

Table 4: Results for the AIC and BIC statistics for the fitted marginal distributions

Distribution	Log likelihood	AIC	BIC
Negative binomial	-17386.61	34777.22	34791.3
Geometric	-18100.54	36203.08	36210.12

It was noted that both the AIC and the BIC values for the negative binomial were less than that of the geometric distribution implying that the negative binomial distribution was the better fit for the data.

Claim amount
The claim amount was interpreted as a continuous distribution. Regard this, the most two fit of continuous distributions based on easy fit software are the log normal distribution and the Weibull distribution.

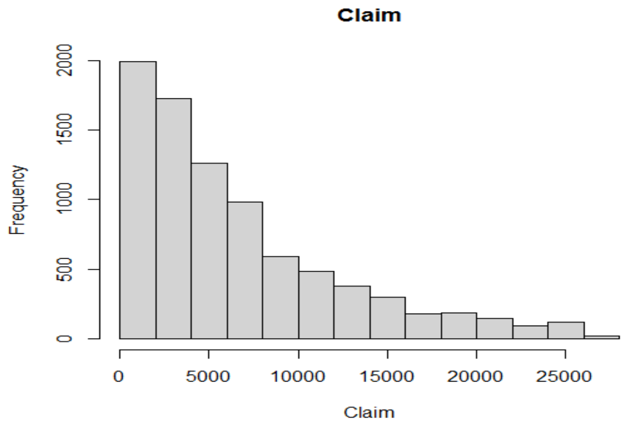

Fig.4: *illustration for Claim amounts data*

The lognormal distribution

Using the maximum likelihood method to fit the lognormal distribution, the following results were obtained.

Parameters:

Table 5: *Results for the lognormal distribution fit to the claim amount variable*

	Estimate	Standard Error
Mean-log	8.317397	0.011648240
Sd-log	1.070181	0.008236517

The Weibull Distribution

Using the maximum likelihood method to fit the Weibull distribution, the following results were obtained;

Parameters:

Table 6: *Results for the Weibull distribution fit to the claim amount variable*

	Estimate	Standard Error
Shape	1.119761	0.009493078
Scale	6786.39089	69.73720582

The following figure represent graphical illustration for both distributions fitted on claim amounts data where the lognormal is black, and the Weibull distribution is blue.

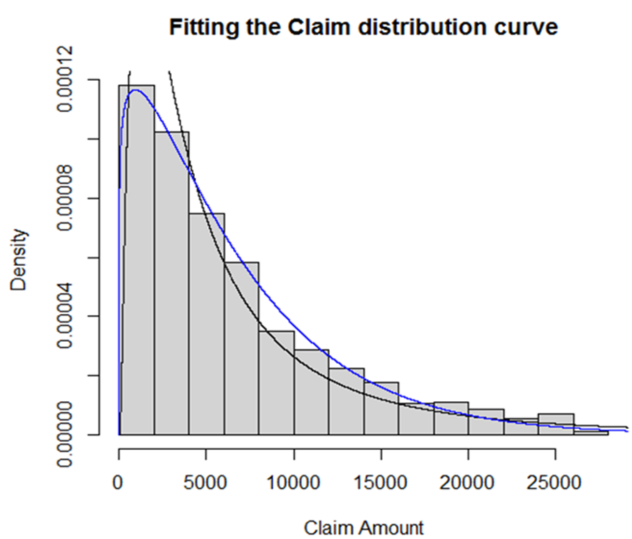

Fig.5: *Graphical illustration for the lognormal and Weibull distributions fit to the claim amount variable*

Summary of claims distribution marginal fit

From graphical illustrations it clearly shows that both distributions could be considered a reasonable fit for the claim amount data. Comparison between the log likelihoods, AIC and BIC for the two distributions was as follows;

Table 7: *Summary of claims distribution marginal fit*

Distribution	Loglikelihood	AIC	BIC
Lognormal	-82470.29	165517.9	165532
Weibull	-82756.94	164944.6	164958.7

It was noted that both the AIC and the BIC values for the Weibull were less than that of the Lognormal distribution implying that the log normal distribution was the better fit for the data.

Fitting the Copula Distributions

The maximum likelihood method was employed to the copulas given the results obtained for the marginal distributions. The copula distributions first had to be converted to their respective probability density functions in order to obtain the likelihood function. The log likelihood function was then obtained for easier optimization. It is important to note that the likelihood and the log likelihood functions were not easily maximized and therefore for some of the copulas; a numeric iterative method – The Newton Raphson method was applied to obtain the parameter estimates. The variables used in the copula functions were the distribution values of the observed instances using the distribution functions that were maximized. The results for the copulas were as follows;

Gumbel Copula

The main results for the Gumbel Barnett Copula were as follows:

Table 8: *Results for the Gumbel Barnett Copula*

Method of likelihood maximization	Newton Raphson
Log-Likelihood	6003.885
Estimate	1.33464
Std. error	0.01834
AIC	-12005.77

the parameter value obtained is outside the bounds required for the Gumbel Barnet copula; (0, 1). In this case, we find the maximum likelihood estimator within these bounds by plotting the graph of parameter values against the log likelihood to obtain the parameter value that maximizes the copula. The graph obtained is as below;

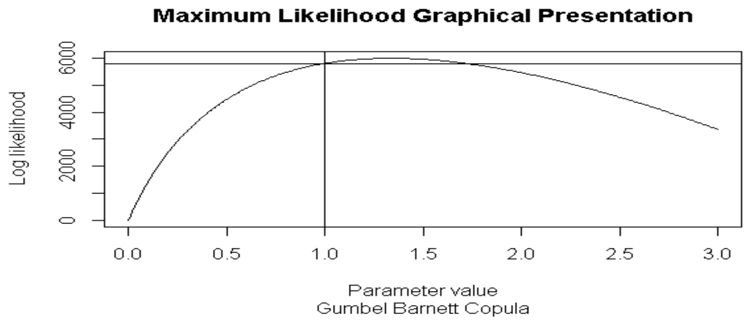

Fig.6: *The graphical illustration of the maximum likelihood estimates for the Gumbel Barnett Copula*

This shows that the parameter value that maximizes the log likelihood within the bounds is 1. Hence rather than 1.24 as obtained by the numerical analysis we settle for 1.

Frank Copula

The results for Franks Copula were as follows:

Table 9: *Results for the Frank Copula*

Method of likelihood maximization	**Newton Raphson**
Log-Likelihood	15.20435
Estimate	0.39269
Std. error	0.07123
AIC	-28.40869

The result for the parameter estimates from the numerical approach was further complemented using the graphical approach as shown below;

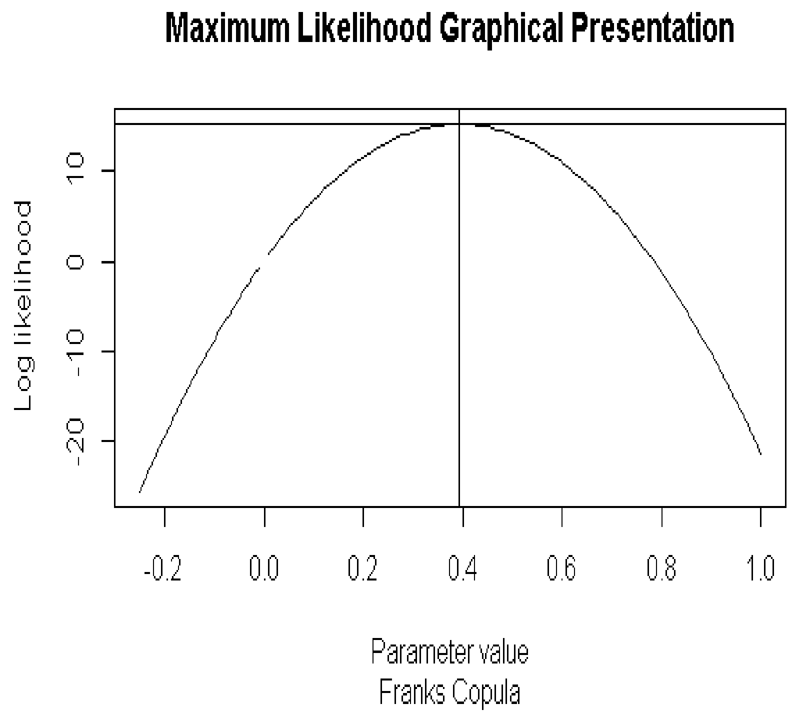

Fig .7: *The graphical illustration of the maximum likelihood estimates for Frank Copula*

Nelson Number 10

As a result of the Nelson number 10 copula, the following results were obtained:

Table 10: *Results for the Nelson 10 Copula*

Method of likelihood maximization	**Newton Raphson**
Log-Likelihood	5917.646
Estimate	-1.078120
Std. error	0.007324
AIC	-11833.29

the parameter value obtained is outside the bounds required for the nelson 10 copula: (0, 1). In this case, we find the maximum likelihood estimator within these bounds by plotting the graph of parameter values against the log likelihood to obtain the parameter value that maximizes the copula. The graph obtained is as below;

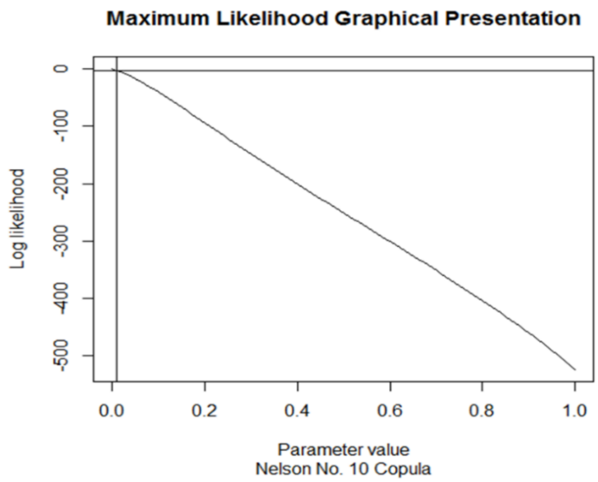

Fig.8: *The graphical illustration of the maximum likelihood estimates for the Nelson No. 10 Copula*

This shows that the parameter value that maximizes the log likelihood within the bounds is 1We choose max value between 0 and 1, as figure indicates max on 0.01

Clayton Copula

The results for Clayton Copula were as follows:

Table 11: *Results for the Clayton Copula*

Method of likelihood maximization	**Newton Raphson**
Log-Likelihood	4.011901
Estimate	0.03938
Std. error	0.01431
AIC	-6.023802

When the maximum likelihood graphical approach was considered, the results were as below;

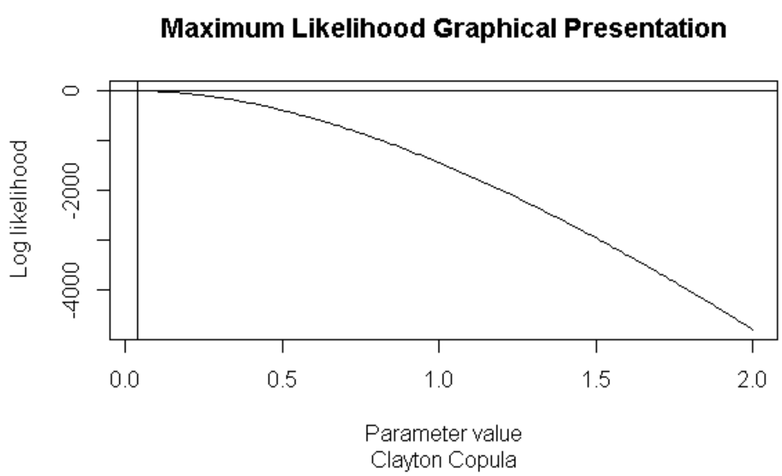

Fig.9: *The graphical representation of the maximum likelihood estimates for Clayton Copula*

Summary of the copula fits to the data

The comparison of the copulas' AIC test statistics yielded the following results:

Table 12: *Results for the AIC statistics for the four copulas*

Copula	AIC
Gumbel Barnett Copula	-12005.77
Nelson Number 10	-11833.29
Franks	-28.40869
Clayton Copula	-6.023802

3. Results and Conclusion

It is observed that the two variables claim amount and report lag exists a positive dependence. This was illustrated by the correlation measures; Pearson's correlation coefficient, Kendall's tau and spearman's rho. In addition to the correlation measures being positive; it was also noted that the magnitude of the measures was significantly small. This implied existence of weak dependence.
And a comparison of the AIC values of each of the copulas was conducted. The Gumbel Barnett copula had the smallest AIC value while the clayton copula had the largest AIC value. It is an asymmetric Archimedean copula that exhibits greater dependence in the positive tail than the negative. This result implied that according to the comparisons of the AIC values for the different copulas, the Gumbel Barnett Copula is the best option in modeling dependence between the variables report delay and claim amount. Finally, it is concluded that there is weak dependence between report lag and claim amounts in property and casualty insurance company.

References

[1] Institute and Faculty of Actuaries. ST7 General Insurance: Reserving and capital modelling. The Actuarial Education Company (ActEd). United Kingdom. (2013).

[2] Gao, G., Gao, & Zhang. Bayesian claims reserving methods in non-life insurance with STAN. Springer Singapore. (2018).

[3] Institute and Faculty of Actuaries. CT6 Statistical Methods. The Actuarial Education Company (ActEd). United Kingdom. (2013).

[4] Zaroudi, S., Faridrohani, M. R., Behzadi, M. H., & Safari-Katesari, H. Copula-based Modeling for IBNR Claim Loss Reserving. 2203.12750. (2022).

[5] Ahn, S., Badescu, A. L., Cheung, E. C., & Kim, J. R. An IBNR-RBNS insurance risk model with marked Poisson arrivals. Insurance: Mathematics and Economics. (2017).

[6] Amayi, S. O. Modeling Dependence Between Report Lag and Claim Amounts Using Copula Models. M.A. thesis, University of Nairobi, Kenya, (2016).

[7] Jaworski, P., Durante, F., Hardle, W. K., & Rychlik, T. Copula theory and its applications. Berlin: Springer. (Vol. 198). (2010).

[8] Haugh, M. An Introduction to Copulas. Quantitative Risk Management. Spring. (2016).

[9] Abdallah, A., Boucher, J. P., Cossette, H., & Trufin, J. Sarmanov family of bivariate distributions for multivariate loss reserving analysis. North American Actuarial Journal, 20(2), 184-200. (2016).

[10] Abdallah, A., Boucher, J. P., & Cossette, H. Modeling dependence between loss triangles with hierarchical Archimedean copulas. ASTIN Bulletin: The Journal of the IAA, 45(3), 577-599. (2015).

[11] Kes, K. Claims reserving using copulas. M.A. thesis, University of Amsterdam, Netherland, (2017).

[12] Shi, P., & Frees, E. W. Dependent loss reserving using copulas. ASTIN Bulletin: The Journal of the IAA, 41(2), 449-486. (2011).

[13] Zhang, Y., & Dukic, V. Predicting multivariate insurance loss payments under the bayesian copula framework. Journal of Risk and Insurance, 80(4), 891-919. (2013).

[14] Soliman, A. Applying the Integrated Risk Management (IRM) Framework for Evaluating & Controlling Financial Risks. PhD thesis, Cairo University, Egypt, (2017).

[15] Merz, M., Wüthrich, M. V., & Hashorva, E. Dependence modelling in multivariate claims run-off triangles. Annals of Actuarial Science, 7(1), 3-25. (2013).