

COMPARISON OF NET PREMIUM RESERVES FOR MULTIPLE-LIFE PRODUCTS USING COPULAS WITH AGE DIFFERENCE

DION KRISNADI^{1,*}, GRADY MATTHIAS OKTAVIAN², DINA STEFANI³
AND HELENA MARGARETHA²

¹School of Information System and Technology

²Faculty of Science and Technology

Universitas Pelita Harapan

Jl. M. H. Thamrin Boulevard 1100 Lippo Village, Tangerang 15811, Indonesia

go9605@student.uph.edu; helena.margaretha@uph.edu

*Corresponding author: Dion.krisnadi@uph.edu

³Mathematics Department

SPK SMAK 8 Penabur

Jl. Tanjung Duren Raya No. 4, Jakarta Barat, DKI Jakarta 11470, Indonesia

dina.stefani@bpkpenaburjakarta.or.id

Received February 2023; accepted April 2023

ABSTRACT. *Multiple-life insurance products provide cover for more than one individual, typically a married couple. A common practice to assume independence between both future lifetimes is not realistic. Copula is a statistical tool which has been widely used to model dependency between bivariate random variables. Four Archimedean copulas were used, namely Gumbel, Frank, Clayton, and Joe. Furthermore, dependency factors in the form of age difference and the gender of the older individual were incorporated to obtain four more copulas. By comparing log-likelihood, AIC, MAPE, and the validity of probability value, the best copulas obtained were Frank, both with and without the dependency factors. NPRs were calculated with the two copulas and the independent model for a hypothetical multiple-life insurance. Friedman test was used to show that the NPRs from the three models for both joint-life and last-survivor products were significantly different for a couple with age difference of -10 to 10 .*

Keywords: Archimedean copula, Copula modeling, Age difference modeling, Multiple-life dependency modeling, Net premium reserve, Multiple-life product

1. Introduction. In life insurance companies that issue multiple-life insurance products, actuaries must model mortality of married couples to estimate the required premiums and reserves. A common practice in modeling the survival function of two (or more) individuals is to assume independence between individuals [1,2]. However, the independence assumption may not be able to capture the dependency between married couples, resulting in less realistic model and estimation [3-5]. Therefore, a more realistic dependency model is needed with dependency factors that can somehow quantify lifestyle habits of a married couple. Youn and Shemyakin proposed age difference as a dependency factor [4], which was further extended by Dufresne et al. to quantify the gender of the older individual [5].

Copula is a statistical tool that has been commonly used to model the dependency between random variables [5], especially when the marginal distribution of each random variable is known [2]. There are a lot of copula models, each with their own dependency structure that are classified into families. Among those, a family called Archimedean copula is widely used to model bivariate lifetimes [5]. In this paper, the four Archimedean

copulas in Dufresne (Gumbel, Frank, Clayton, and Joe) [5] are used to model the dependency between the lifetimes of married couples. In addition to vanilla copulas, copulas which incorporated age difference and the gender of the older individual as dependency factors [5] are also built. To choose the best copula model log-likelihood, AIC, and mean absolute percentage error (MAPE) are used. Empirical copula is considered as the true value in MAPE. The best model will be used to calculate the net premium reserves (NPR) of a hypothetical term multiple-life policy. Friedman test is conducted to conclude whether the NPRs from the best copula are significantly different from the simple independent model.

This paper is structured in four chapters. Chapter 2 discusses the problem of multiple-life insurance products, and the methodology and data used in this paper. The main results and analysis are in Chapter 3. Finally, Chapter 4 states the concluding remarks as well as recommendations for further research in this topic.

2. Problem Statement and Preliminaries. Multiple-life insurance products provide cover for more than one individual, typically a married couple, while treating them as one party. This type of products has been of interest, either offered as it is or as a part of another plan [6]. Generally, there are two types of multiple-life insurance depending on when the benefits will be paid to the insured. A joint-life insurance pays the benefit upon the first death, while a last-survivor insurance pays the benefit upon the last death.

One essential step for an insurer is to estimate reserve, which is the amount of money needed to meet its liabilities. This estimate must be done effectively, that is, it must be sufficient to cover all liabilities and ensure solvency, but also realistic to enable investments. One key risk affecting the reserve of a life insurer is the mortality risk [7]. Therefore, in multiple-life insurance, it is necessary to model and assess joint mortality risk, in the form of joint future lifetime, of married couples properly so that reserves can be estimated effectively.

Two main approaches have been extensively researched to model joint future lifetimes, the semi-Markov and copula approach [8]. Between the two, much research can be found in the past few years that use the copula approach [2]. [9] gave an extensive review on the use of copula, especially Archimedean copula, in both life and non-life insurance. Furthermore, since the major task in applying copula is how to identify the best copula [10], an extensive study on estimating and evaluating the best copula is also presented in [9]. The results show that MLE outperforms the other estimation methods, whereas squared difference between fitted and empirical copula can measure the overall fit of a copula.

Copula has been used to calculate reserve in life insurance. Gaussian, Archimedean, and t -copulas were used to model dependency among various assets and liabilities in life insurance companies [7]. Based on the AIC value, t -copula was then used to analyze the required capital for a life insurer in various scenarios. In multiple-life insurance, the use of Gaussian and Archimedean copula to analyze reserve was investigated by [11] and [2], respectively. The reserve estimated using copula to model joint future lifetimes was compared with reserve estimated using independent assumption. Both concluded that the copula approach that should be used as the resulting estimates was more effective. The use of Archimedean copula to model joint future lifetime was further investigated in [5]. Dependency factors were added to improve the dependency structure of the copula. Goodness-of-fit test showed that the proposed model outperformed the regular copula.

This paper will focus on modelling bivariate future lifetimes of husband and wife using copulas and analyzing the reserves estimated from the best copula. Four copulas from the Archimedean family are used, namely Gumbel, Frank, Clayton, and Joe. The copula distribution function for the four copulas and their parameter α can be seen in Table 1,

TABLE 1. Four copula functions from the Archimedean family

Copula	Copula distribution function	
Gumbel	$C_\alpha(u, v) = \exp\left(-\left[(-\ln(u))^\alpha + (-\ln(v))^\alpha\right]^{\frac{1}{\alpha}}\right)$	$\alpha \geq 1$
Frank	$C_\alpha(u, v) = -\frac{1}{\alpha} \ln\left(1 + \frac{(e^{-\alpha u}-1)(e^{-\alpha v}-1)}{(e^{-\alpha}-1)}\right)$	$\alpha \neq 0$
Clayton	$C_\alpha(u, v) = (u^{-\alpha} + v^{-\alpha} - 1)^{-\frac{1}{\alpha}}$	$\alpha > 0$
Joe	$C_\alpha(u, v) = 1 - \left((1-u)^\alpha + (1-v)^\alpha - (1-u)^\alpha(1-v)^\alpha\right)^{\frac{1}{\alpha}}$	$\alpha > 1$

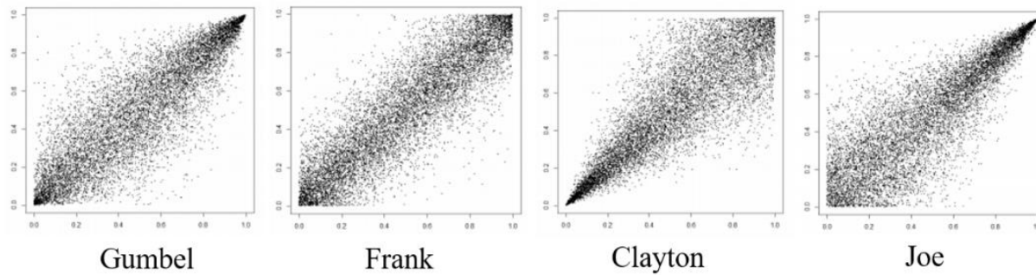


FIGURE 1. Scatter plot for the four Archimedean copulas

where u and v are the marginal distribution functions for each individual lifetime. The dependency nature of each copula can be seen from the scatter plot in Figure 1.

In addition to the four copulas, another four copulas are constructed by modifying the parameter α . Dependency factors in the form of age difference (d) and the gender of the older individual ($|d|$) are included. The α parameter can then be rewritten as Equation (1) for Frank and Clayton, and as Equation (2) for Gumbel and Joe copulas [5].

$$\alpha(d) = \frac{\beta_0}{1 + \beta_1 d + \beta_2 |d|}, \tag{1}$$

$$\alpha(d) = 1 + \frac{\beta_0}{1 + \beta_1 d + \beta_2 |d|}. \tag{2}$$

The approach used to estimate the parameters of the copula is the inference functions for margins (IFM) approach, which is computationally more convenient [5,12]. It consists of two steps. First, maximum likelihood estimation (MLE) is used to estimate the marginal survival function for each individual that is assumed to follow Gompertz survival model.

$${}_t p_{x_j} = \mathbb{P}(T_{x_j} > t_j) = \exp\left(e^{-\frac{x_j - m_j}{s_j}} \left(1 - e^{-\frac{t_j}{s_j}}\right)\right), \tag{3}$$

where T_{x_j} is the lifetime random variable of individual with age x_j and $j = m, f$ denotes the gender, while m_j and s_j are the two parameters to be estimated. The estimated marginals are then used to conduct MLE for the copula parameter α or β_i , $i = 0, 1, 2$. Let $\bar{u} = {}_t p_{x_m}$ and $\bar{v} = {}_t p_{x_f}$ be the marginal survival, and the joint survival $\mathbb{P}(T_{x_m} > t_m, T_{x_f} > t_f) = \tilde{C}_\alpha(\bar{u}, \bar{v})$ can be defined as [5]

$$\tilde{C}_\alpha(\bar{u}, \bar{v}) = \bar{u} + \bar{v} - 1 + C_\alpha(1 - \bar{u}, 1 - \bar{v}). \tag{4}$$

Goodness-of-fit test was done by Dufresne et al. to compare the four copulas [5], and we refer to their conclusion to help determine the fit of the copula models. We choose the best copula by comparing the log-likelihood and AIC [10,13]. Furthermore, MAPE is also used.

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \tag{5}$$

where empirical copula is used as the actual value of the i th sample y_i , and the estimated value from the copula as \hat{y}_i . The empirical copula C_n is defined as [14]

$$C_n \left(\frac{i}{n}, \frac{j}{n} \right) = \frac{\text{number of pairs } (x, y) \text{ which qualifies } x \leq x_{(i)}, y \leq y_{(j)}}{n}, \tag{6}$$

with $(x_{(i)}, y_{(j)})$ as the ordered statistics of the data points. This non-parametric copula along with its variants can be deemed as “a consistent estimator of the true underlying copula” [12].

The best Archimedean copula along with its variant with (or without) the dependency factor d and the simple independent model are then used to calculate the NPRs of a 50-year term multiple-life policy with fixed premium under equivalence principle, payable at the beginning of each year. The insured will be paid a lump sum of 100 unit if their status failed within 50 years, payable at the end of year of death. The status of the insured party fails either at the first death for joint-life, or the second death for last-survival. The policy has a fixed annual interest rate of 0.001. The policy has an entry age of 40-50 for both genders with $d = x_m - x_f$ as the age difference between the husband and the wife. There are 21 values of d from -10 to 10 . The formula to calculate the NPR at time k (denoted as ${}_kV$) is

$${}_kV = (\text{APV of Benefit at } t = k) - P \times (\text{APV of Annuity at } t = k), \tag{7}$$

where APV is the actuarial present value and P is the premium. Friedman test is conducted for each d to see if there is a significant difference among the NPRs from the three models.

3. Main Results. The dataset is provided by an unnamed Canadian insurance company in an R package “CASdatasets”. This dataset has been used in past papers [3,5,8]. There are 14,889 couples that were observed for five years from 29 December 1988 to 31 December 1993. Each couple has the entry age, the time spent being observed, and the time of death (if applicable) for each individual. After removing duplicate rows and couples whose entry ages are outside of 40 to 110, 12,264 couples were obtained. The censoring status of each individual and the age difference between each couple were then determined. Individuals who died during the observation are given the censoring status of 0, while those who did not have the value of 1. There are 11,007 and 11,817 censored individuals for male and female, respectively, 10,748 couples with at least one individual are censored, and only 188 couples with both individuals died during observation. This shows the data is highly censored [5].

A sample of the data can be seen in Table 2, where the subscript m and f denote whether it is the husband or wife, respectively. Column x is the entry age, column t is the time to exit (either because of death or surviving the observation), column δ is the censoring status, and column d is the age difference $d = x_m - x_f$. A positive d indicates that the husband is older than the wife, and vice versa.

TABLE 2. A snapshot of the processed data

x_m	x_f	t_m	t_f	δ_m	δ_f	d
40.5767	45.2911	5.0055	5.0055	1	1	-4.7144
⋮						
104.8826	88.9249	5.0055	1.045	1	0	15.9577

The data was first used to estimate the parameters of the marginal survival functions in (3) using MLE. The contribution of the i th individual towards the likelihood function is

$$L_j^i(m, s) = \left[t_j^i p_{x_j^i}(m, s) \right]^{\delta_j^i} \left[f_{x_j^i}(t_j^i, m, s) \right]^{1-\delta_j^i}, \tag{8}$$

where $j = m, f$ denotes the gender of the individual. The first term in (8) is the contribution of the individual who survived the observation, while the second term is the contribution for those who died. The four estimated parameters, $\hat{m}_m, \hat{m}_f, \hat{s}_m,$ and $\hat{s}_f,$ can be seen in Table 3. The estimation was done using “optim” function in R.

TABLE 3. The estimated parameters for gompertz marginal survival functions

Parameters	m_m	s_m	m_f	s_f
Estimates	86.29132	10.17565	92.017339	7.962881

Using the estimated parameters, marginal survival functions were calculated and used to estimate the parameters of each copula. The contribution of the i th couple towards the likelihood function is

$$L_j^i(\alpha) = \left[\frac{\partial^2 \tilde{C}_\alpha(u_i, v_i)}{\partial u \partial v} \right]^{(1-\delta_m^i)(1-\delta_f^i)} \times \left[\frac{\partial \tilde{C}_\alpha(u_i, v_i)}{\partial u} \right]^{(1-\delta_m^i)\delta_f^i} \times \left[\frac{\partial \tilde{C}_\alpha(u_i, v_i)}{\partial v} \right]^{\delta_m^i(1-\delta_f^i)} \times \left[\tilde{C}_\alpha(u_i, v_i) \right]^{\delta_m^i \delta_f^i}. \tag{9}$$

As in the case of marginal distribution, the four terms in (9) are associated with the censoring status of each individual. The estimated parameters for each copula, both with and without the dependence factor d , can be seen in Table 4. Estimation was again done using “optim” function in R. It should be noted that there are notable differences with the results from past research [5]. This may be caused by the differences in the data processing steps or in the use of R “optim” function. The Nelder-Mead and BFGS methods are used to obtain $\hat{\alpha}$ and $\hat{\beta}_i, i = 0, 1, 2,$ respectively. After a few experiments, it was found that the estimation for $\hat{\beta}_i, i = 0, 1, 2$ was influenced greatly by the choice of the starting value for each parameter.

TABLE 4. The estimated parameters for each copula

Estimated copula parameter	$\hat{\alpha}(d)$			$\hat{\alpha}$
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	
Gumbel	0.5463732	0.08966002	0.041001593	1.4662745
Frank	3.4772704	0.03028555	-0.003317607	3.3055115
Clayton	0.2369763	0.02043875	-0.010975339	0.2132795
Joe	2.2888205	0.07162282	0.040160301	2.7724875

The results in Table 4 were used to calculate the log-likelihood and AIC. In addition, MAPE was also calculated by computing the distance between each copula model with the empirical copula. In this process, observations from each individual were converted into pseudo-observations $x_j^i = \frac{n\hat{F}_j(x_j^i)}{n+1}$ where j denotes the gender and $\hat{F}_j(x_j^i)$ denotes the empirical marginal distribution. The pseudo-observations were then used to calculate the empirical copula using the *copula* package in R [15-18] and to re-estimate the parameters of each copula. The two copulas were then used to calculate MAPE. The three metrics

TABLE 5. Comparisons of the eight copulas

	Log-likelihood	AIC	MAPE
Gumbel	-1492.283	2986.566	9.996420e-02
Frank	-1491.294	2984.588	6.864099e-02
Clayton	-1505.053	3012.105	1.372324e-01
Joe	-1499.938	3001.876	1.810615e-01
Gumbel(d)	-1489.297	2984.595	9.750004e-02
Frank(d)	-1490.154	2986.308	6.820497e-02
Clayton(d)	-1555.192	3116.384	1.463071e-01
Joe(d)	-1496.169	2998.339	1.786735e-01

can be seen in Table 5, where “(d)” denotes the copula variant with the dependency factor d .

Table 5 shows that Gumbel and Frank have the best performance. They have the highest log-likelihood, and the smallest AIC and MAPE. This observation is in line with the results of Dufresne et al. [5]. However, Gumbel(d) could produce a joint survival function larger than 1, which was not found in the other copulas. This may be caused by the optimization method. Therefore, we used Frank and Frank(d) as the best copulas to calculate the NPRs.

NPR is the amount of money a life insurer must provide to ensure that on average it can pay its liabilities. It needs to be estimated for each year the insurance is in force. To calculate NPR, we need to calculate the net premium for the insurance, the APV of benefit payable to the insured, and the APV of annuity paid by the insured. Furthermore, since two types of multiple-life insurance were considered, the premium and NPR calculations were done twice for joint-life and last-survivor. We rewrite Equation (7) using the actuarial notation to calculate ${}_kV$ for joint-life and last-survivor insurance as (10) and (11).

$$\begin{aligned}
{}_kV &= 100A_{x+k,y+k:\overline{50-k}|} - P\ddot{a}_{x+k,y+k:\overline{50-k}|} \\
&= 100 \sum_{i=0}^{50-k} v^{k+1} ({}_i p_{x_m x_f} - {}_{i+1} p_{x_m x_f}) - P \sum_{i=0}^{50-k} v^k {}_i p_{x_m x_f}, \quad (10)
\end{aligned}$$

$$\begin{aligned}
{}_kV &= 100A_{x+k,y+k:\overline{50-k}|} - P\ddot{a}_{x+k,y+k:\overline{50-k}|} \\
&= 100 \sum_{i=0}^{50-k} v^{k+1} ({}_{i+1} q_{x_m x_f} - {}_i q_{x_m x_f}) - P \sum_{i=0}^{50-k} v^k {}_i p_{x_m x_f}. \quad (11)
\end{aligned}$$

As an example, we present the calculation for male age 40 and female age 40. The premiums for joint-life product were 2.32, 2.08, and 2.23 for independent, Frank, and Frank(d), respectively. The premiums for last-survivor product were 0.84, 1.01, and 0.90. Using these premiums and Equations (10) and (11), NPRs for $k = 0, 1, \dots, 50$ are shown in Table 6. These reserves were then used in Friedman test. The test was done 21 times for each value of d from -10 to 10 . The results can be seen in Table 7.

From Table 7, H_0 are rejected for all d . The null hypothesis in Friedman states that the treatments have equal effect, while the alternative states at least one treatment has different effect. Therefore, we can conclude that the NPR calculated using copula is significantly different from the ones using independent model for both joint-life and last-survivor products.

4. Conclusions. Dependency between future lifetimes of married couples has been modeled using copula. From the four Archimedean copulas, the best one according to log-likelihood, AIC, MAPE, and the validity of probability value was Frank copula. We calculated NPR of a hypothetical multiple-life product using independent model and Frank

TABLE 6. NPRs calculations for male and female age 40

t	Joint-life			Last survivor		
	Independent	Frank(a)	Frank(b)	Independent	Frank(a)	Frank(b)
0	0	0	0	0	0	0
1	2.19692	1.97051	2.11335	0.81950	0.98461	0.87881
			⋮			
49	17.91664	16.44181	17.47679	0.26010	1.80832	0.72881
50	0	0	0	0	0	0

TABLE 7. Friedman test results

x_m	x_f	d	p-value		x_m	x_f	d	p-value	
			Joint-life	Last survivor				Joint-life	Last survivor
40	40	0	5.24E-22	4.74E-21	40	40	0	5.24E-22	4.74E-21
40	41	-1	1.08E-21	5.24E-22	41	40	1	4.74E-21	4.74E-21
40	42	-2	5.24E-22	2.47E-22	42	40	2	5.24E-22	5.24E-22
40	43	-3	5.24E-22	3.65E-21	43	40	3	4.74E-21	5.24E-22
40	44	-4	1.08E-21	5.24E-22	44	40	4	5.24E-22	2.47E-22
40	45	-5	2.47E-22	1.08E-21	45	40	5	5.24E-22	2.47E-22
40	46	-6	5.24E-22	1.93E-22	46	40	6	5.24E-22	1.08E-21
40	47	-7	2.47E-22	5.24E-22	47	40	7	5.24E-22	4.74E-21
40	48	-8	1.08E-21	5.24E-22	48	40	8	2.47E-22	4.74E-21
40	49	-9	5.24E-22	5.24E-22	49	40	9	5.24E-22	5.24E-22
40	50	-10	1.08E-21	5.24E-22	50	40	10	5.24E-22	5.24E-22

copula, both with and without the additional dependency factors in the form of age difference and the gender of the older individual. Friedman test was done to conclude that the NPR from the copula model is significantly different from the independent model for all age differences between male and female from -10 to 10 , both for joint-life and last-survivor products. From the Friedman test, we can conclude that copula modelling should be used to obtain more realistic estimates of premium and reserves. More research can be done to consider other dependency factors between a married couple into a copula model, such as the length of time they have been married, or the time the remaining individual has been left behind. In addition, analysis of tail dependence may also be done to further evaluate the fit of the copula.

REFERENCES

- [1] J. Li and A. Ng, *SOA Exam MLC Study Manual, Fall 2017*, ACTEX Learning, 2017.
- [2] H. Safari-Katesari and S. Zaroudi, Analyzing the impact of dependency on conditional survival functions using copulas, *Statistics in Transition New Series*, vol.22, no.1, pp.217-226, DOI: 10.21307/STATTRANS-2021-013, 2021.
- [3] E. W. Frees, J. Carriere and E. Valdez, Annuity valuation with dependent mortality, *J. Risk. Insur.*, vol.63, no.2, pp.229-261, 1996.
- [4] H. Youn and A. Shemyakin, Statistical aspects of joint life insurance pricing, *Proceedings of the Business and Statistics Section of the American Statistical Association*, pp.34-38, 1999.
- [5] F. Dufresne, E. Hashorva, G. Ratovomirija and Y. Toukourou, On age difference in joint lifetime modelling with life insurance annuity applications, *Annals of Actuarial Science*, vol.12, no.2, pp.350-371, DOI: 10.1017/S1748499518000076, 2018.
- [6] D. C. M. Dickson, M. R. Hardy and H. R. Waters, *Actuarial Mathematics for Life Contingent Risks*, Cambridge University Press, 2019.

- [7] S. Benson et al., Copula models of economic capital for life insurance companies, *Applied Econometrics*, vol.58, pp.32-54, <https://ideas.repec.org/a/ris/apltrx/0393.html>, 2020.
- [8] F. Gobbi, N. Kolev and S. Mulinacci, Joint life insurance pricing using extended Marshall-Olkin models, *ASTIN Bulletin: The Journal of the IAA*, vol.49, no.2, pp.409-432, DOI: 10.1017/ASB.2019.3, 2019.
- [9] T. D. Kularatne, J. Li and D. Pitt, On the use of Archimedean copulas for insurance modelling, *Annals of Actuarial Science*, vol.15, no.1, pp.57-81, DOI: 10.1017/S1748499520000147, 2021.
- [10] I. Ghosh, D. Watts and S. Chakraborty, Modeling bivariate dependency in insurance data via copula: A brief study, *Journal of Risk and Financial Management*, vol.15, no.8, 329, DOI: 10.3390/JRFM15080329, 2022.
- [11] I. Lee, H. Lee and H. T. Kim, Analysis of reserves in multiple life insurance using copula, *Commun. Stat. Appl. Methods*, vol.21, no.1, pp.23-43, DOI: 10.5351/CSAM.2014.21.1.023, 2014.
- [12] C. Genest, B. Rémillard and D. Beaudoin, Goodness-of-fit tests for copulas: A review and a power study, *Insur. Math. Econ.*, vol.44, no.2, pp.199-213, 2009.
- [13] Y. Zou, X. Ye, K. Henrickson, J. Tang and Y. Wang, Jointly analyzing freeway traffic incident clearance and response time using a copula-based approach, *Transp. Res. Part C Emerg. Technol.*, vol.86, pp.171-182, DOI: 10.1016/j.trc.2017.11.004, 2018.
- [14] R. B. Nelsen, *An Introduction to Copulas*, Springer, New York, 2007.
- [15] M. Hofert, I. Kojadinovic, M. Maechler and J. Yan, *copula: Multivariate Dependence with Copulas*, R Package Version 1.1-0, 2022.
- [16] M. Hofert and M. Mächler, Nested Archimedean copulas meet {R}: The {nacopula} package, *J. Stat. Softw.*, vol.39, no.9, pp.1-20, 2011.
- [17] J. Yan, Enjoy the joy of copulas: With a package {copula}, *J. Stat. Softw.*, vol.21, no.4, pp.1-21, 2007.
- [18] I. Kojadinovic and J. Yan, Modeling multivariate distributions with continuous margins using the {copula} {R} package, *J. Stat. Softw.*, vol.34, no.9, pp.1-20, 2010.