

Fitting Compound Archimedean Copulas to Data for Modeling Electricity Demand

Moshe Kelner¹, Zinoviy Landsman² & Udi E. Makov²

¹ Department of Statistics, Haifa University, Israel

² Actuarial Research Center, Haifa University, Israel

Correspondence: Moshe Kelner, Department of Statistics, Haifa University, Israel. E-mail: msh.kelner@gmail.com

Received: June 6, 2021 Accepted: August 7, 2021 Online Published: August 15, 2021

doi:10.5539/ijsp.v10n5p20 URL: <https://doi.org/10.5539/ijsp.v10n5p20>

Abstract

Modeling dependence between random variables is accomplished effectively by using copula functions. Practitioners often rely on the single parameter Archimedean family which contains a large number of functions, exhibiting a variety of dependence structures. In this work we propose the use of the multiple-parameter compound Archimedean family, which extends the original family and allows more elaborate dependence structures. In particular, we use a copula of this type to model the dependence structure between the minimum daily electricity demand and the maximum daily temperature. It is shown that the compound Archimedean copula enhances the flexibility of the dependence structure and provides a better fit to the data.

Keywords: compound archimedean copula, dependence structure, electricity demand, Kendall's tau, Gini's gamma

1. Introduction

Modeling the dependence structure between random variables is a long-standing research interest. Commonly, two random variables can be described by a bivariate distribution, such as normal, log normal or gamma. The key characteristic of this concept is that both of the random variables have the same parametric univariate distribution. This restriction does not exist when using copula functions (Genest and Favre, (2007)). These functions provide a formidable route to create multidimensional distributions with any given marginals and enable an auspicious way for describing dependence structures between them. Sklar's theorem (1959) established the foundation for the concept. By Sklar, every two-dimensional joint distribution function $H(x, y) = P(X \leq x, Y \leq y)$, where X and Y are continuous random variables distributed $F(x) = P(X \leq x)$ and $G(y) = P(Y \leq y)$, can be expressed in terms of the separate parts of its marginal univariate distributions and copula C , which captures the interdependencies between them, such as

$$H(x, y) = C(F(x), G(y)), \quad x, y \in (-\infty, \infty). \quad (1)$$

An extensive study of the topic was conducted by Nelsen (2006), who established that the copula functions must meet three properties, for $u = F(x)$ and $v = G(y)$,

1. When one of the marginals has zero value then $C(u, 0) = C(0, v) = 0$.
2. In cases where one of the marginals is equal to one, then $C(u, 1) = u$ and $C(1, v) = v$.
3. C is 2-increasing function, i.e. for $u_1, u_2, v_1, v_2 \in I$, with $u_1 \leq u_2$, and $v_1 \leq v_2$, $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$.

Numerous types of copula functions have been described in the literature, with one of the most prominent groups of copulas being the Archimedean family, which is characterized by its generator function. A two-dimensional Archimedean copula is denoted by:

$$C_\theta(u, v) = \varphi_\theta^{-1}(\varphi_\theta(u) + \varphi_\theta(v)), \quad (2)$$

where $\varphi_\theta(t)$ is the Archimedean generator and θ is the dependence parameter. The generator φ is a decreasing and convex function with domain $(0, 1]$ and range $(0, \infty)$ such that $\varphi_\theta(1) = 0$. Numerous attributes of the Archimedean copulas and their generators were established by Nelsen (2006), in particular we note:

- For every copula C and $u, v \in I^2$, $W(u, v) \leq C(u, v) \leq M(u, v)$, for $W(u, v) = \max(u + v - 1, 0)$ and $M(u, v) = \min(u, v)$ which is the copula version of Fréchet-Hoeffding bounds inequality.

- The joint distribution of U and the Archimedean copula $C(U, V)$ for $\{(u, v) \in I^2 | u \leq s, C(u, v) \leq t\}$ is denoted by

$$K'C(s, t) = \begin{cases} s, & s \leq t \\ t - \frac{\varphi_\theta(t) - \varphi_\theta(s)}{\varphi'_\theta(t)}, & s > t \end{cases}$$

- For an Archimedean copula C and u, v in I^2 , $\varphi'_\theta(u) \frac{dC(u,v)}{dv} = \varphi'_\theta(v) \frac{dC(u,v)}{du}$.

Archimedean copulas are widely useful in diverse areas, including hydrology, medical science, accident analysis, insurance, banking, weather, climate change, and more (Bhatti and Do, (2019)). In actuarial science, Hendriks and Landsman (2017) investigated the dependence structure of multivariate random variable with the probability measure $\bar{F}(X_1, \dots, X_n) = P(X_1 \geq x, \dots, X_n \geq x_n) = h\left(\sum_{i=1}^n \lambda_i x_i\right)$, $x_i \geq 0$, and showed the intrinsic relation of this structure with survival Archimedean copula with a generator $\varphi = h^{-1}$. Alai and Landsman (2017) used this measure for modeling life dependence structure. Applications for portfolio allocations were introduced by Byun and Song (2021) who searched for appropriate copula functions in various dependence structures among asset return distributions. Chen and Guo (2019) exhibited various advantages of using copula in hydrology. They employed the Archimedean copula in order to describe the dependence structure between depth, volume, and duration of flows. Regarding energy research, applications of Archimedean copulas have been extensively studied in recent years. Schindler and Jung (2018) used copulas to fit bivariate distributions of wind vector components to estimate directional wind energy yields in Germany. A Wind-Wave energy model proposed by Wu et. al. (2019) used Archimedean copulas to analyse correlations between wind speed and wave height. Implementation of dependence structures of electricity demand, economic and sectoral indicators using members of the family were investigated by Evkaya et.al. (2018). Due to the increasing number of electric vehicles and smart cities, the estimation of the electricity demand has become progressively more important. For instance, an underestimation of peak demand might result in the shortage of electricity in different locations. Numerous other studies have investigated electricity peak demand, and used copula functions to research the dependence structures between components which affect it. Cecati et.al. (2015) and Wang et. al. (2016) suggested models for estimating demand behaviour and peak load. Sun et. al. (2019) gave a peak load estimate approach using copulas to model complex non-linear dependence structures among different consumers. Research on minimum electricity demand has received less attention, and even less so when it comes to dependence structures using copula functions. Minimum electricity demand is the lowest level of demand for any given day and is very important for a system design and day-to-day operation, since low demand results in the shut-down of power generating units which can have negative consequences when a sudden need for energy arises. In our work we take steps to fill this gap. We investigate the application of a compound Archimedean copula in electricity demand. The dependence structure between the minimum daily demand and the maximum daily temperature was examined. It is shown that the compound copula enriches the flexibility of the dependence structure and provides a better fit to the data than the original compound copula. In section 2, we provide an example of a compound Archimedean copula. Methods for fitting copula to data and in particular a nonparametric approach for fitting two dependence measures are outlined in section 3. The modeling of minimum electricity demand versus maximum temperature using the original and the compound copula is discussed in section 4. Conclusions are given in section 5.

2. An Example of a Compound Archimedean Copula

Kelner et. al. (2021) introduced a method for generating new Archimedean copulas having a more flexible dependence structure. They created new generators by compounding an existing generator with respect to $g_\eta(\theta)$, a probability density of the dependence parameter θ , defined by

$$\varphi_C(t) = \int_{\Theta} \varphi_\theta(t) g_\eta(\theta) d\theta, \tag{3}$$

where C denotes compound. They gave an example of a copula introduced by Nelsen (2006, 4.2.8), denoted by C_8 ,

$$C_\theta(u, v) = \max\left[\frac{\theta^2 uv - (1-u)(1-v)}{\theta^2 - (\theta-1)^2(1-u)(1-v)}, 0\right], \tag{4}$$

where they compounded its generator function

$$\varphi_\theta(t) = \frac{1-t}{1+(\theta-1)t}, \theta > 1, \tag{5}$$

with a distribution of the dependence parameter θ given by,

$$g_{\eta}(\theta) = \frac{1}{\ln(b) - \ln(a)} \frac{1}{\theta}, \quad a < \theta < b. \tag{6}$$

Then using Eq.(3) the compound generator takes the form

$$\varphi_C(t) = 1 + \frac{\ln\left(\frac{((a-1)t+1)}{((b-1)t+1)}\right)}{\ln\left(\frac{b}{a}\right)}, \quad b > a. \tag{7}$$

Applying the equation introduced by Genest and Mackay (1986b) for establishing Kendall’s tau

$$\tau = 1 + 4 \int_0^1 \frac{\varphi_{\theta}(t)}{\varphi'_{\theta}(t)} dt, \tag{8}$$

we get an expression for Kendall’s tau for the original copula

$$\tau = 1 - \frac{2(\theta + 2)}{3\theta}. \tag{9}$$

Using Eq.(7) and Eq.(8) the Kendalls tau of the compound copula is given by,

$$\tau_C = 1 + \frac{4}{6(a-b)} \left[\begin{array}{l} 2(a-1)(b-1)\left(\ln\left(\frac{b}{a}\right)\right) \\ + 3(a+b-2)\ln\left(\frac{b}{a}\right) + b-a \\ + (2(a-1)(b-1) + 3(a+b))\ln\left(\frac{a}{b}\right) \\ - \frac{2(b-1)}{(a-1)} + \frac{3(a+b)}{(a-1)} - \frac{3(a+b)}{(b-1)} + \frac{2(a-1)}{(b-1)} \\ + 6\ln\left(\frac{b}{a}\right) + \frac{(3a-b-2)\ln(a)}{(a-1)^2} + \frac{(a+2-3b)\ln(b)}{(b-1)^2} \\ - \frac{6a}{(a-1)} + \frac{6b}{(b-1)} \end{array} \right]. \tag{10}$$

3. Fitting Copula to Empirical Data Methods

Using copulas for modeling dependence and identifying the empirical associations between random variables has become a useful tool in many applied fields. The remaining challenge is how to fit a copula function to given data sets. The answers to this have attracted widespread attention over the years and various methods are available. Joe (1996) and Moore et. al. (2017) gave an in-depth introduction to three estimate approaches. These include the parametric, semi-parametric and non-parametric. The most effective parametric approaches are the maximum likelihood estimation (MLE), where $\hat{\theta} = \arg \max_{\theta} \sum_{i=1}^n \log L(U_i, V_i; \theta)$, and $L(\theta; U, V) = \prod_{i=1}^n \frac{d}{du} \frac{d}{dv} C(U_i, V_i)$, or the two stage inference for margins (IFM) method, in which the parameters of the marginals are estimated and used for estimating the copula dependence parameter. The semi-parametric approach, which is the maximum pseudo-likelihood (MPL), introduced by Genest et. al. (1995) is comparable to the MLE with a converted pseudo data. As for the non-parametric approach, Genest and Rivers (1993) developed a method for determining a suitable Archimedean copula for particular data. They provided an approach for estimating the dependence structure of Archimedean copulas, which is robust to outliers and has no need for information on the marginals distribution. The first step in their approach was to estimate the dependence parameter θ using the empirical value of kendall’s tau. They estimated Kendall’s tau using the observed variables

$$\begin{aligned} \hat{\tau} &= \binom{n}{2}^{-1} \sum_{i < j} \text{sign} [(X_i - X_j)(Y_i - Y_j)] \\ &= \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \left(\sum_{j=i+1}^n \text{sign} [(X_i - X_j)(Y_i - Y_j)] \right), \end{aligned} \tag{11}$$

where

$$\text{sign}(X) = \left\{ \begin{array}{l} 1, X > 0 \\ 0, X = 0 \\ -1, X < 0 \end{array} \right\},$$

and estimated the dependence parameter $\hat{\theta}$ using $\hat{\tau}$ by equating Eq.(11) to Eq.(8). In our work we use the same method for estimating $\hat{\theta}$. We present another dependence measure, the Gini’s gamma, which is a nonparametric measure which represents rank association of random variables. Genest et. al. (2010) provided a detailed description of the applications of this

measure and explored its finite-sample moments and asymptotic distributions under independence. Nelsen (1998) showed the empirical form of the measure: For a random sample $(X_1, Y_1), \dots, (X_n, Y_n)$ with associated ranks $(R_1, S_1), \dots, (R_n, S_n)$, the empirical Gini's gamma is

$$\hat{\gamma} = \frac{2}{n^2} \sum_{i=1}^n \{(n+1 - R_i) - S_i\} - |R_i - S_i| \tag{12}$$

Using copula functions, Nelsen (1998) showed that this measure can be represented as

$$\gamma = 4 \left[\int_{t=0}^1 C(t, 1-t) dt - \int_{t=0}^1 [t - C(t, t)] dt \right], \tag{13}$$

for all $t \in (0, 1)$. Notice that both of the dependence measures, Kendall's tau and Gini's gamma, depend only on the copula function and not on the marginals.

If we were to use a single parameter C8 copula we could estimate θ either by equating the theoretical value of Kendall's tau in Eq.(9) to its empirical value in Eq.(11) and obtain

$$\hat{\theta}_\tau = \frac{4}{1 - 3\hat{\gamma}}, \tag{14}$$

or repeat this procedure and estimate θ with respect to the Gini's gamma by equating Eq.(12) to Eq.(13). Clearly these two approaches would typically yield different estimates of θ . In other words, there is no $\hat{\theta}$ which could guarantee that the values of these two dependence measures are equal to their empirical values. This, of-course, is achievable when a two-parameter compound copula is considered. Equating Eq.(10) to Eq.(11) allows multiple parameter pairs (a, b) . The chosen pair, say (\hat{a}, \hat{b}) , is the one which would assure that Eq.(12) is equal to Eq.(13). Clearly the compound copula with parameters (\hat{a}, \hat{b}) exhibiting the dependence expressed by the data. This will be demonstrated in section four.

4. Application of Daily Electrical Demand Data

This study examines daily electricity demand data in Israel, which involves the minimum demand and maximum temperature in the working hours (08:00 to 20:00), from the year 2005 to 2020, excluding weekends and holidays, where weekdays companies are not operating and electricity demand is lower compared to the weekdays (a total of 4,233 observations). The data is divided into four seasons, Winter (December-February), Spring (March-May), Summer (June-September) and Autumn (October-November). Figure 1 introduces the distribution of the minimum demand in the four seasons.

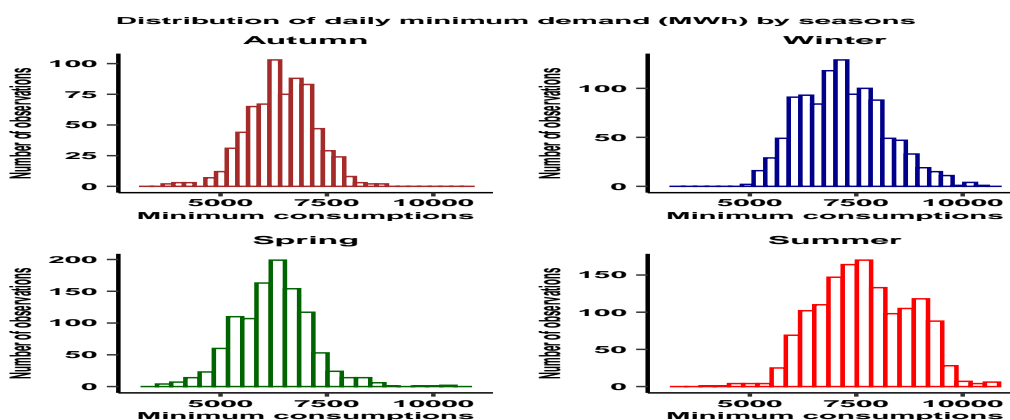


Figure 1. Distributions of minimum demand by season

It shows that the minimal electricity demand during winter and summer is higher than the minimum demand during the transitional seasons, with the summer being the highest. This is due to the use of household appliances such as heaters in winter and cooling air conditioners in summer, which affect the total and minimum electricity demand. Figure 2 introduces seasonal scatter plots for daily minimum demand vs. daily maximum temperature.

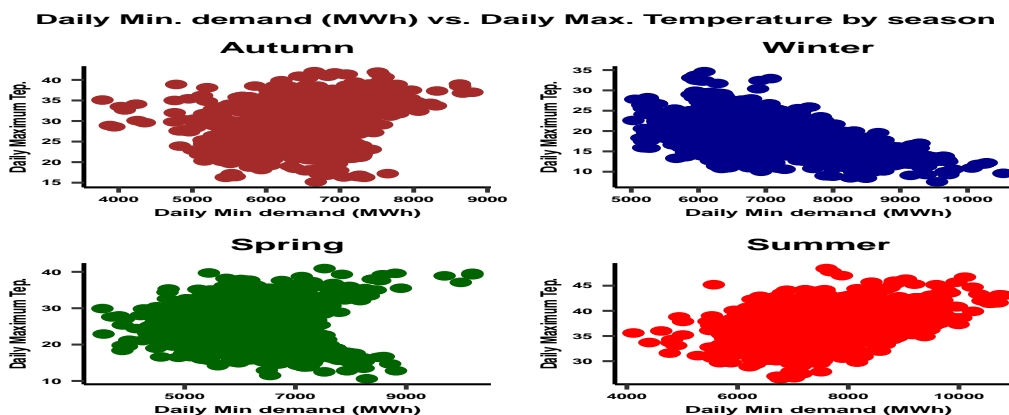


Figure 2. Daily minimum demand vs. Daily maximum temperature by season

It is clear from figure 2 that during the winter, there is a negative dependence between the minimum daily electricity demand and the maximum daily temperature, and a positive dependence during the summer. In the other season, a correlation is not noticeable.

For each separate season we calculated the empirical values of Kendall’s tau and the estimated dependence parameter using Eq.(11) and Eq.(14), respectively. The results are shown in Table 1.

Table 1. Empirical values of Kendall’s tau $\hat{\tau}$ and the estimated dependence parameters $\hat{\theta}_{\tau}$ by season

Season	n	$\hat{\tau}$	$\hat{\theta}_{\tau}$
Winter	1,073	-0.326	2.022
Spring	1,074	0.023	4.926
Summer	1,388	0.256	17.241
Autumn	698	0.125	6.4

Once θ is estimated by $\hat{\theta}_{\tau}$, the Gini’s gamma can be derived from the original copula using Eq.(13) and is denoted by γ_{τ} . Clearly, this value is bound to be different then the empirical Gini estimate $\hat{\gamma}$. We note that for the compound copula the empirical and the theoretical values of γ are identical since such a copula allows matching the empirical and theoretical values for both τ and γ . The results are introduced graphically in Figure 3.

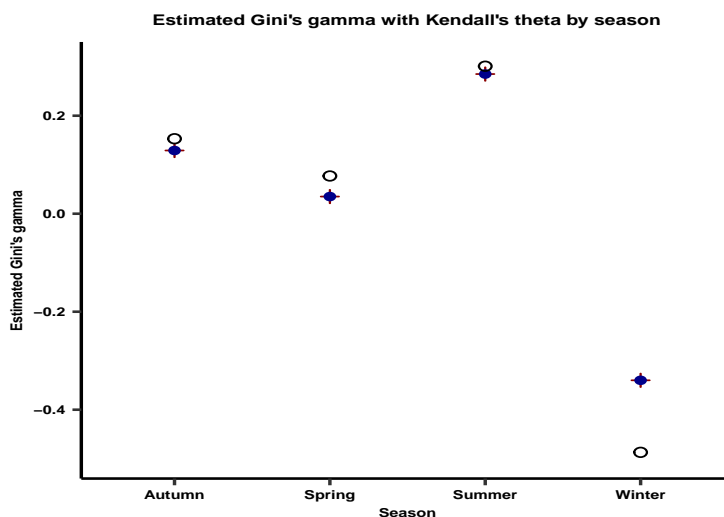


Figure 3. Gini’s gamma dependence measure for empirical, original C8 copula and compound copula values using $\hat{\theta}_{\tau}$

The graph above represents the values of $\hat{\gamma}$ vs. γ_{τ} separated into four seasons. The blue dots depict the empirical values

of Gini’s gamma, $\hat{\gamma}$ for each season. The black hollow dots and the red pluses depict the values of γ_τ with respect to the original copula and the compound copula, respectively. In this graph, the accuracy is denoted by the closeness of $\hat{\gamma}$ to γ_τ . It is easy to see that the compound copula symbolized by the red pluses offers an accurate match. As for the original copula, which is symbolized by the black hollow dot, absolute accuracy of $\hat{\gamma}$ is unattainable.

For measuring the deviance between the empirical and the estimated Gini’s gamma, we used the Mean Average Percentage Error (MAPE) given by

$$MAPE = \frac{100}{4} \sum_{i=1}^4 \left| \frac{\hat{\gamma} - \gamma_\tau}{\hat{\gamma}} \right|, \tag{15}$$

where i symbolizes the season. The values of $\hat{\gamma}, \gamma_\tau$ and the MAPE for each season are presented in Table 3.

Table 3. Empirical and estimated values for Gini’s gamma calculated by $\hat{\theta}_\tau$, by season

Dependence measures	Season	$\hat{\gamma}$	γ_τ	$\left \frac{\hat{\gamma} - \gamma_\tau}{\hat{\gamma}} \right $	
γ	Winter	-0.34	-0.487	0.432	
	Summer	0.285	0.301	0.056	
	Spring	0.035	0.077	1.2	
	Autumn	0.129	0.153	0.186	<i>MAPE</i>
					46.9

The results of the MAPE show that using the original single parameter copula, the average deviation between the empirical values of Gini’s gamma $\hat{\gamma}$ and the estimate values of Gini’s gamma γ_τ , is 46.9%, which means high inaccuracy. High deviations occur in the winter and the spring season, i.e. in December-May, which are the coldest months of the year.

5. Conclusions

In this paper, we introduce an application of an extension of Archimedean copulas, namely the compound copula, which is governed by more than a single parameter, thus offering a higher degree of flexibility in term of fitting a copula which accurately reflects dependence measures. We used daily data of electricity demand and applied a univariate copula function and a compound copula for fitting the dependence structure between the daily minimum demand and daily maximum temperature. The data was separated into four seasons and the study was conducted for two different dependence measures: the Kendall’s tau, τ and the Gini’s gamma, γ . We have shown that for a univariate copula it is impossible to find a value of the dependence parameter which allows accurate fitting with respect to these dependence measures, unlike the compound Archimedean copula which offers a perfect fit.

References

Alai, D. H., & Landsman, Z. (2017). Lifetime dependence models generated by monotone functions. *Scandinavian Actuarial Journal*, 7, 576-604. <https://doi.org/10.1080/03461238.2017.140584>

Bhatti, M. I., & Do, H. Q. (2019). Recent development in copula and its applications to the energy, forestry and environmental sciences. *International Journal of Hydrogen Energy*, 44, 19453-19473. <https://doi.org/10.1016/j.ijhydene.2019.06.015>

Byun, K., & Song, S. (2021). Value at Risk of portfolios using copulas. *Communications for Statistical Applications and Methods*, 28, 59-79. <https://doi.org/10.29220/CSAM.2021.28.1.059>

Cecati, C., Kolbusz, J., Rozycki, P., Siano, P., & Wilamowski, B. (2015). A Novel RBF Training Algorithm for Short-Term Electric Load Forecasting and Comparative Studies. (2015). *IEEE Transactions on Industrial Electronics*, 62, 6519 - 6529. <https://doi.org/10.1109/TIE.2015.2424399>

Chen, L., & Guo, S. (2019). *Copulas and Its Application in Hydrology and Water Resources*. Springer, New York.

Cossette, H., Marceau, E., & Mtalai, I. (2019). *Collective risk models with dependence*, 87, 153-168. <https://doi.org/10.1016/j.insmathco.2019.04.008>

Durante, F., & Sempì, C. (2015). *Principles of copula theory*. Chapman and Hall, New York.

Evkaya, O., Yozgatligil, C., & Kestel, S. (2018). Measuring Dependence between Electricity Consumption and Economic Indicators via Copulas: Turkish Case. *Gazi University Journal of Science*, 31, 1284-1296.

- Genest, C., & Favre, A. (2007). Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. *Journal of Hydrologic Engineering*, 12, 347-368. [https://doi.org/10.1061/\(asce\)1084-0699\(2007\)12:4\(347\)](https://doi.org/10.1061/(asce)1084-0699(2007)12:4(347))
- Genest, C., & MacKay, R. J. (1986b). The joy of copulas: Bivariate distributions with uniform marginal. *Amer. Statist.*, 40, 280-283. <https://doi.org/10.2307/2684602>
- Genest, C., & Neslehova, J., & Ghorbal, B. N. (2010). Spearman's footrule and Gini's gamma: a review with complements. *Journal of Nonparametric Statistics*, 22, 937-954. <https://doi.org/10.1080/10485250903499667>
- Genest, C., & Rivest, L. P. (1993). Statistical Inference Procedures for Bivariate Archimedean Copulas. *Journal of the American Statistical Association*, 88, 1034-1043. <https://doi.org/10.2307/2290796>
- Genest, C., Ghoudi, K., & Rivest P. L. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82, 543-552. <https://doi.org/10.1093/biomet/82.3.543>
- Hendriks, H., & Landsman, Z. (2017). A generalizations of multivariate Pareto distributions: tail risk measures, divided differences, asymptotic. *Scandinavian Actuarial Journal*, 9, 785-803. <https://doi.org/10.1080/03461238.2016.1255249>
- Hennessy, D., & Harvey, L. (2002). The use of Archimedean copula to model portfolio allocations. *Mathematical Finance*, 12, 143-154. <https://doi.org/10.1111/1467-9965.00136>
- Joe, H. (1997). *Multivariate models and dependence concepts*. Chapman and Hall, New York.
- Joe, H., & Xu, J. J. (1996). The Estimation Method of Inference Functions for Margins for Multivariate Models. *Technical Report No.*, 166. <https://doi.org/10.14288/1.0225985>
- Kelner, M., & Landsman, Z., & Makov, E. U. (2021). Compound Archimedean Copulas. *International Journal of Statistics and Probability*, 10, 126-134. <https://doi.org/10.5539/ijsp.v10n3p126>
- Moore, D., McCabe P. G., & Craig, B. (2017). *Introduction to the Practice of Statistics*. Ninth edition, Springer, New York.
- Nelsen, R. B. (1998). Concordance and Gini's measure of association. *Journal of Nonparametric Statistics*, 9, 227-238. <https://doi.org/10.1080/10485259808832744>
- Nelsen, R. B. (2006). *An Introduction to Copulas*. Second ed. Springer, New York.
- Schindler, D., & Joung, C. (2018). Copula-based estimation of directional wind energy yield: A case study from Germany. *Ann. Inst. Statist. Math.*, 59, 487-498. <https://doi.org/10.1016/j.enconman.2018.05.071>
- Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229-231.
- Sun, M., Wang, Y., Strbac, G., & Kang, C. (2019). Probabilistic Peak Load Estimation in Smart Cities Using Smart Meter Data. *IEEE Transaction on Industrial Electronics*, 66, 1608-1618.
- Wang, Y., Zhou, N., Ju, P., & Zhang, X. (2016). Clustering of Electricity Consumption Behavior Dynamics Toward Big Data Applications. *IEEE TRANSACTIONS ON SMART GRID*, 7, 2437-2447. <https://doi.org/10.1109/TSG.2016.2548565>
- Wu, F., Chen, Q., Kang, C., & Xia, Q. (2019). Wind-Wave Coupling Model for Wave Energy Forecast. *IEEE Transactions on Sustainable Energy*, 10, 586-595. <https://doi.org/10.1109/TSTE.2018.2839358>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).