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

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

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:

\[
C_\theta(u, v) = \varphi_\theta^{-1}(\varphi_\theta(u) + \varphi_\theta(v)),
\]  

where \( \varphi_\theta(t) \) is the Archimedean generator and \( \theta \) is the dependence parameter. The generator \( \varphi \) 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 Fre’chet-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

$$K'(s, t) = \begin{cases} s, & s \leq t \\ t - \frac{\psi(t) - \psi(s)}{\psi'(s)}, & s > t \end{cases}$$

• For an Archimedean copula $C$ and $u, v$ in $I^2$, $\frac{dC(u,v)}{du} = \frac{\varphi_\theta(u)}{\varphi_\theta(t)} \frac{dC(u,v)}{dv}$.

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 $F(X_1, ..., X_n) = P(X_1 \geq x, ..., X_n \geq x) = h(\sum_{i=1}^n \lambda_i x_i), x_i \geq 0$, 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.

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 $\theta$, defined by

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

where $C$ denotes compound. They gave an example of a copula introduced by Nelsen (2006, 4.2.8), denoted by C8,

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

where they compounded its generator function

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

with a distribution of the dependence parameter $\theta$ given by,
We get an expression for Kendall’s tau for the original copula estimating \( \hat{\theta} \) and estimated the dependence parameter \( \hat{\theta} \) using the observed variables. Genest et. al. (2010) provided a detailed description of the applications of this empirical value of Kendall’s tau. They estimated Kendall’s tau using the observed variables \( \theta \) on the marginals distribution. The first step in their approach was to estimate the dependence parameter \( \theta \) on the marginals distribution. 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 \( L(\theta; U, V) = \prod_{i=1}^{n} \frac{1}{2} C(U_i, V_i) \), 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 maximim 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 \( \theta \) using the empirical value of Kendall’s tau. They estimated Kendall’s tau using the observed variables.

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 semi-parametric approach, which is the maximim 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 \( \theta \) using the empirical value of Kendall’s tau. They estimated Kendall’s tau using the observed variables.

Using Eq. (7) and Eq. (8) the Kendall’s tau of the compound copula is given by

\[
\tau = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt,
\]

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

\[
\tau = 1 - \frac{2(\theta + 2)}{3\theta}.
\]

Using Eq. (7) and Eq. (8) the Kendall’s tau of the compound copula is given by

\[
\tau_C = 1 + \frac{4}{6(a-b)} \left[ \frac{2(a-1)(b-1)\left(\ln\left(\frac{b}{a}\right)\right)}{+3(a+b-2)\ln\left(\frac{b}{a}\right)+b-a} + \frac{2(a-1)(b-1)+3(a+b)(\ln\left(\frac{b}{a}\right))}{(a-1)} \right]
\]

\[
= \left( \frac{n}{2} \right)^{-1} \sum_{i<j} \text{sign}\left( (X_i - X_j)(Y_i - Y_j) \right)
\]

where

\[
\text{sign}(X) = \begin{cases} 1, & X > 0 \\ 0, & X = 0 \\ -1, & X < 0 \end{cases}
\]

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 \((X_1, Y_1), \ldots, (X_n, Y_n)\) with associated ranks \((R_1, S_1), \ldots, (R_n, S_n)\), the empirical Gini’s gamma is

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

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

\[
\gamma = 4 \left[ \int_{t=0}^{1} C(t, 1-t) \, dt - \int_{t=0}^{1} [t - C(t, t)] \, dt \right],
\] (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 \(\theta\) either by equating the theoretical value of Kendall’s tau in Eq.(9) to its empirical value in Eq.(11) and obtain

\[
\hat{\theta}_\tau = \frac{4}{1 - 3\hat{\gamma}},
\] (14)

or repeat this procedure and estimate \(\theta\) with respect to the Ginis’s gamma by equating Eq.(12) to Eq.(13). Clearly these two approaches would typically yield different estimates of \(\theta\). 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.

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.
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

<table>
<thead>
<tr>
<th>Season</th>
<th>( n )</th>
<th>( \hat{\tau} )</th>
<th>( \hat{\theta}_\tau )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winter</td>
<td>1,073</td>
<td>−0.326</td>
<td>2.022</td>
</tr>
<tr>
<td>Spring</td>
<td>1,074</td>
<td>0.023</td>
<td>4.926</td>
</tr>
<tr>
<td>Summer</td>
<td>1,388</td>
<td>0.256</td>
<td>17.241</td>
</tr>
<tr>
<td>Autumn</td>
<td>698</td>
<td>0.125</td>
<td>6.4</td>
</tr>
</tbody>
</table>

Once \( \theta \) is estimated by \( \hat{\theta}_\tau \), the Gini’s gamma can be derived from the original copula using Eq.(13) and is denoted by \( \gamma_\tau \). 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 \( \gamma \) are identical since such a copula allows matching the empirical and theoretical values for both \( \tau \) and \( \gamma \). The results are introduced graphically in Figure 3.

Figure 3. Gini’s gamma dependence measure for empirical, original \( C^8 \) copula and compound copula values using \( \hat{\theta}_\tau \).
of Gini’s gamma, \( \hat{\gamma} \) for each season. The black hollow dots and the red pluses depict the values of \( \gamma \) 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 \( \gamma \). 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

\[
MAPE = \frac{100}{4} \sum_{i=1}^{4} \left| \frac{\hat{\gamma} - \gamma_i}{\hat{\gamma}} \right|, \tag{15}
\]

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

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

<table>
<thead>
<tr>
<th>Dependence measures</th>
<th>Season</th>
<th>( \hat{\gamma} )</th>
<th>( \gamma )</th>
<th>( \frac{\hat{\gamma} - \gamma_i}{\hat{\gamma}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma )</td>
<td>Winter</td>
<td>-0.34</td>
<td>-0.487</td>
<td>0.432</td>
</tr>
<tr>
<td></td>
<td>Summer</td>
<td>0.285</td>
<td>0.301</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>Spring</td>
<td>0.035</td>
<td>0.077</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>Autumn</td>
<td>0.129</td>
<td>0.153</td>
<td>0.186</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MAPE 46.9</td>
</tr>
</tbody>
</table>

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 \( \gamma_i \), 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, \( \tau \) and the Gini’s gamma, \( \gamma \). 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


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/).