

JIRSS (2020)

Vol. 19, No. 01, pp 163-183

DOI:10.29252/jirss.19.1.163

Parameter Estimation of Some Archimedean Copulas Based on Minimum Cramér-von-Mises Distance

Selim Orhun Susam ¹

¹ Department of Econometrics, Munzur University, Turkey.

Received: 31/01/2019, Revision received: 02/08/2019, Published online: 26/08/2019

Abstract. The purpose of this paper is to introduce a new estimation method for estimating the Archimedean copula dependence parameter in the non-parametric setting. The estimation of the dependence parameter has been selected as the value that minimizes the Cramér-von-Mises distance which measures the distance between Empirical Bernstein Kendall distribution function and true Kendall distribution function. A Monte Carlo study is performed to measure the performance of the new estimator and compared to conventional estimation methods. In terms of estimation performance, simulation results show that the proposed Minimum Cramér-von-Mises estimation method has a good performance for low dependence and a small sample size when compared with the other estimation methods. The new minimum distance estimation of the dependence parameter is applied to model the dependence of two real data sets.

Keywords. Cramér-von-Mises, Archimedean Copula, Parameter Estimation, Bernstein Polynomials

MSC: 62G05, 62G32

Corresponding Author: Selim Orhun Susam (orhunsusam@munzur.edu.tr)

1 Introduction

Copula models are popular tools for describing the dependency of multivariate data where the univariate distribution functions are combined with joint distribution function by Sklar's theorem (Sklar (1959)). Consider random variables X and Y with joint cumulative distribution function H and marginals F and G , respectively. Then there exist a copula C such that $H(x, y) = C(F(x), G(y))$, for all x, y in \mathfrak{X} .

An important family of the copula is known as Archimedean copula. This copula class is characterized by generator function φ . In application, φ belongs to the parametric family φ_θ for which the parameter θ needs to be estimated. This paper aims to investigate the estimation of parameter θ and to present a Monte Carlo simulation study on the performance of proposed estimators for some Archimedean copulas. For three popular parametric Archimedean copulas, classical maximum likelihood and inversion of Kendall's tau estimators are compared with the minimum Cramér-von-Mises estimator.

In the literature, there are many studies for estimating the dependency parameter. Deheuvels (1978), Genest et al. (1995), Joe (1996), Joe (1997), Joe (2005), Fermanian et al. (2004), Kim et al. (2007), Najafabadi et al. (2013), Biau et al. (1985), Tsukahara (2005), Mendes et al. (2007) and Weib (2011) studied the estimation methods and gave some comparisons among them. This paper is closely related to the works of Tsukahara (2005), Mendes et al. (2007), Biau et al. (1985) and Weib (2011). The similarity with these articles is the usage of the minimum distance method in parameter estimation. Tsukahara (2005) investigated the limiting behavior of the empirical copula process and applied it to prove some asymptotic properties of a minimum distance estimator for an Euclidean parameter in a copula model. Mendes et al. (2007) obtained robust estimators for copula parameters through the minimization of weighted goodness-of-fit statistics. Biau et al. (1985) introduced a minimum L_1 distance estimate for parametric copula densities. Weib (2011) presented a comprehensive Monte Carlo simulation study on the performance of minimum-distance and maximum-likelihood estimators for bivariate parametric copulas. In particular, they considered Cramér-von-Mises, Kolmogorov-Smirnov and L_1 -variants of the CvM-statistic based on the empirical copula process, Kendall's dependence function and Rosenblatt's probability integral transform. The dissimilarity of this study from other minimum distance estimation methods is the usage of Bernstein polynomials for estimation of Kendall distribution function.

The remainder of the study is organized as follows. In Section 2, we give the general concept of Archimedean copulas and their properties. Also, we review the estimation

methods of the dependence parameter, and in addition to the new flexible estimation method, the dependence parameter of Archimedean copula is introduced. In section 3, the estimation results of Monte Carlo simulation are given. Section 4 is devoted to real data applications. Lastly, the final section is devoted to the conclusion.

2 Archimedean Copulas

Archimedean copula models form a useful family of copula models in which the dependence structure can be characterized by a univariate distribution function (Nelsen (2006), Section 4). One of their important features that separates this class from the others is that it has a generator function φ which is used to construct an Archimedean copula. Archimedean copula with generator function φ , $C : [0, 1]^2 \rightarrow [0, 1]$ is defined by

$$C(u, v) = \varphi^{[-1]}{\varphi(u) + \varphi(v)}; \quad u, v \in [0, 1].$$

This generator function and its pseudo-inverse function are defined in Nelsen (2006) by the following definition,

Definition 2.1. A generator function φ is a continuous, strictly decreasing convex function defined from $[0, 1]$ to $[0, \infty)$ such that $\varphi(1) = 0$. If $\varphi(0) = \infty$, then the generator is called as a strict generator. The pseudo inverse of φ is the function $\varphi^{[-1]}$, defined on $[0, \infty)$ to $[0, 1]$ is given by

$$\varphi^{[-1]} = \begin{cases} \varphi^{-1}(t) & 0 \leq t \leq \varphi(0) \\ 0 & \varphi(0) \leq t < \infty. \end{cases}$$

An Archimedean copula can be reduced to a univariate distribution function through generator function. Genest et al. (1993) showed that the function φ can be obtained by the univariate distribution function $K(t) = P(C(u, v) \leq t)$ Remarkably, there is a link between the function φ and $K(t)$ such as

$$K(t) = t - \frac{\varphi(t)}{\varphi'(t)} \quad t \in (0, 1).$$

$K(t)$, called as Kendall distribution function, identifies the generator function $\varphi(t)$ and so the dependence structure of the Archimedean copula family. In Table 1, some Archimedean copulas with a generator function and Kendall distribution function are summarized.

Table 1: Archimedean copulas with generator function.

Copula	$\varphi(t)$	$K(t)$	$\tau(\theta)$	Range of θ
Clayton	$(t^{-\theta} - 1)/\theta$	$t + \frac{t(1-t^\theta)}{\theta}$	$\frac{\theta}{\theta+2}$	$(-\infty, \infty) - \{0\}$
Frank	$-\log\left(\frac{\exp(t\theta)-1}{\exp(\theta)-1}\right)$	$t - \frac{(\exp(t\theta)-1)\log\left(\frac{\exp(-t\theta)-1}{\exp(-\theta)-1}\right)}{\theta}$	$1 + 4\theta^{-1}(D_1^*(\theta) - 1)$	$(-\infty, \infty) - \{0\}$
Gumbel	$(-\log(t))^\theta$	$t - \frac{t\log(t)}{\theta}$	$\frac{\theta-1}{\theta}$	$[1, \infty)$
Independence	$-\log(t)$	$t - t\log(t)$	-	-

* D_1 is the first Debye function of order 1, $D_1(x) = x^{-1} \int_0^x t(\exp(t) - 1)^{-1} dt$

Genest et al. (1993) proposed a nonparametric procedure using an empirical estimate \hat{K}_n of K . Let $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ be a random sample of size n from a pair (X, Y) with the copula C . For $i = 1, \dots, n$, consider the pseudo observations $T_i = \sum_{j=1}^n I(X_i < X_j, Y_i < Y_j)/(n-1)$. The Kendall distribution function, $K(t)$ is then estimated by the empirical distribution function

$$K_n(t) = \sum_{i=1}^n I(T_i \leq t)/n.$$

Susam et al. (2018) defined the Bernstein estimator of order ($m > 0$) of the Kendall distribution function K as,

$$K_{m,n}(t) = \sum_{k=0}^m K_n(k/m) P_{k,m}(t),$$

where $P_{k,m}(t) = \binom{m}{k} t^k (1-t)^{m-k}$ is the Binomial probability and K_n is the empirical distribution function of $K(t)$. They conclude that Bernstein Kendall estimator outperms the classical Kendall estimator. Because, Bernstein Kendall estimator is flexible with regard to its polynomial degree.

2.1 Parameter Estimation via Inversion of Kendall's Tau

If C is a bivariate Archimedean copula with generator function φ , Genest et al. (1986) defined Kendall's tau as

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

which can be computed for some Archimedean copulas, see Table 1. Genest et al. (1993) introduce a method-of-moments estimator for bivariate one-parameter Archimedean copulas based on Kendall's tau. Let $\tau = h(\theta)$ be the Kendall's tau and let τ_n be the empirical Kendall's tau based on a random sample of size n . Then the moment estimator of θ denoted by $\hat{\theta}$ is given by $\hat{\theta} = h^{-1}(\tau_n)$, $\tau_n = 4\bar{T} - 1$, where, $\bar{T} = \frac{1}{n} \sum_{i=1}^n T_i$. $h(\theta)$ function is also listed in Table 1. $\sqrt{n}(\hat{\theta} - \theta)$ is asymptotically normal with zero mean. (See, Kojadinovic et al. (2010)).

2.2 Maximum-Likelihood Estimation of Parameter

The most popular method for estimating the dependence parameter θ is based on Maximum Likelihood principle. This method is applicable for all absolute continuous copula classes. Suppose that copula $C(u, v)$ has a density $c(u, v)$ defined by

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v} \quad u, v \in (0, 1).$$

When the margins F and G are known, the log-likelihood function for θ is defined as

$$l(\theta) = \sum_{i=1}^n \log(c(F(x_i), G(y_i))).$$

Oakes (1994) proposed to replace $F(x)$, $G(y)$ with their empirical counterparts $F_n(x)$ and $G_n(y)$, respectively. Genest et al. (1995) defined empirical estimates of marginals as

$$F_n(x) = \frac{1}{n+1} \sum_{i=1}^n I(X_i \leq x), \quad G_n(y) = \frac{1}{n+1} \sum_{i=1}^n I(Y_i \leq y),$$

where I denotes indicator function. Then, Log-maximum pseudo likelihood function $l^*(\theta)$ is given by

$$l^*(\theta) = \sum_{i=1}^n \log(c(u_i, v_i)),$$

where $u_i = \frac{\text{rank}(X_i)}{n+1}$ and $v_i = \frac{\text{rank}(Y_i)}{n+1}$. The maximum likelihood estimator is the value $\hat{\theta}_n^{MLE}$ that maximizes $l^*(\theta)$. Genest et al. (1995) showed that $\hat{\theta}_n^{MLE}$ is a consistent estimator.

2.3 Minimum Cramér-von-Mises Estimator Based on Kendall Distribution Function

Let $K(t)$ be Kendall distribution function associated with copula C and let $K_n(t)$ be the empirical Kendall distribution function. The Cramér-von-Mises (CvM) distance between $K(t)$ and $K_n(t)$ is defined by

$$CvM = \int_0^1 (K_n(t) - K(t))^2 dt.$$

The estimation of the dependence parameter θ can be selected as the value that minimizes the CvM function. Weib (2011) proposed estimation methods of Archimedean copula dependence parameter using Cramér-von-Mises distance. They used classical Empirical estimation methods of Kendall distribution function in the estimation of the dependence parameter. In contrast to Weib (2011), we use the Bernstein Empirical estimation of Kendall distribution function instead of the classical empirical estimation. CvM distances for Bernstein estimation of Kendall distribution function for Gumbel, Clayton and Frank copulas are defined in next Lemma.

Lemma 2.1. *Let $K_{Gu}(t)$, $K_{Cl}(t)$ and $K_{Fr}(t)$ be Kendall distributions of the Gumbel, Clayton and Frank copulas and let $K_{m,n}(t)$ be the Bernstein estimator of Kendall's distribution function. The Cramer-von-Mises (CvM) distances between $K_{m,n}(t)$ and $K_{Gu}(t)$, $K_{Cl}(t)$, $K_{Fr}(t)$, denoted by CvM_{Gu} , CvM_{Cl} , CvM_{Fr} are given by*

$$\begin{aligned} CvM_{Gu} &= \sum_{k=0}^m \binom{m}{k}^2 K_n^2\left(\frac{k}{m}\right) Beta(2k+1, 2m-2k+1) \\ &+ 2 \sum_{k=0}^{m-1} \sum_{s=k+1}^m \binom{m}{k} \binom{m}{s} K_n\left(\frac{k}{m}\right) K_n\left(\frac{s}{m}\right) Beta(k+s+1, 2m-k-s+1) \\ &- 2 \sum_{k=0}^m \left(\binom{m}{k} K_n\left(\frac{k}{m}\right) \left(Beta(k+2, m-k+1) - \sum_{i=0}^{m-k} \frac{(-1)^i}{\theta} \binom{m-k}{i} \left(-\frac{1}{(k+i+2)^2} \right) \right) \right) \\ &+ \frac{2+6\theta+9\theta^2}{27\theta^2}. \end{aligned}$$

$$\begin{aligned}
 CvM_{Cl} = & \sum_{k=0}^m \binom{m}{k}^2 K_n^2\left(\frac{k}{m}\right) Beta(2k + 1, 2m - 2k + 1) \\
 & + 2 \sum_{k=0}^{m-1} \sum_{s=k+1}^m \binom{m}{k} \binom{m}{s} K_n\left(\frac{k}{m}\right) K_n\left(\frac{s}{m}\right) Beta(k + s + 1, 2m - k - s + 1) \\
 & - 2 \sum_{k=0}^m \left(\binom{m}{k} K_n\left(\frac{k}{m}\right) \left(Beta(k + 2, m - k + 1) \left(\frac{\theta + 1}{\theta}\right) - \frac{\beta(k + \theta + 2, m - k + 1)}{\theta} \right) \right) \\
 & + \frac{17 + 13\theta + 2\theta^2}{27(1 + \theta) + 6\theta^2}.
 \end{aligned}$$

$$\begin{aligned}
 CvM_{Fr} = & \sum_{k=0}^m \binom{m}{k}^2 K_n^2\left(\frac{k}{m}\right) Beta(2k + 1, 2m - 2k + 1) \\
 & + 2 \sum_{k=0}^{m-1} \sum_{s=k+1}^m \binom{m}{k} \binom{m}{s} K_n\left(\frac{k}{m}\right) K_n\left(\frac{s}{m}\right) Beta(k + s + 1, 2m - k - s + 1) \\
 & - 2 \sum_{k=0}^m \left(\binom{m}{k} K_n\left(\frac{k}{m}\right) \left(Beta(k + 2, m - k + 1) - \sum_{i=0}^{m-k} \frac{(-1)^i}{\theta} \binom{m-k}{i} A(t, k, i) \right) \right) \\
 & + B(t, \theta),
 \end{aligned}$$

where $A(t, k, j)$ and $B(t, \theta)$ are numerical solutions of the $\int_0^1 t^{k+j} (\exp(t * \theta) - 1) \log\left(\frac{\exp(-t\theta)-1}{\exp(-\theta)-1}\right)$ and $\int_0^1 \left(t - \frac{\exp(t\theta)-1}{\theta} \log\left(\frac{\exp(-t\theta)-1}{\exp(-\theta)-1}\right)\right)^2 dt$ respectively. See the appendix for detailed calculations.

Then minimum Cramér-von-Mises estimation of parameter θ for Gumbel, Clayton and Frank copulas are defined as

$$\hat{\theta}_{n,m}^{CvM} = \operatorname{argmin}_{\theta \in \Theta} \{CvM_{(*)}\},$$

where m denotes the degree of the Bernstein polynomial and also this non-linear optimization problem can easily be solved by Statistical programming language R or Mathematica.

From Leblanc (2012) and Susam et al. (2018), we know that $\hat{K}_{m,n}(t)$ is a consistent estimator. Also, Duchesne et al. (1997) show that $\hat{\theta}_n^{CvM}$ is a consistent estimator.

3 Simulation Study

Monte Carlo simulation study was conducted to compare Inversion of Kendall's tau estimator, Maximum likelihood estimator and Minimum Cramér-von-Mises estimator of dependence parameter θ of Gumbel, Clayton, Frank copulas, and to assess estimated MSE's of estimations. The primary purpose is achieved by comparing the true parameter with the parameters estimated with the three estimation methods. The data are generated from three Archimedean copulas such as Clayton, Gumbel, and Frank with Kendall's Tau 0.1, 0.2, ..., 0.8. These distributions cover different shapes and characteristics. The Clayton copula exhibits strong left tail dependence. Gumbel copula cannot model negative dependence, but it contrasts to Clayton, Gumbel has strong right tail dependence. Moreover, Frank copula exhibits symmetric and weak tail dependence in both tails. 10,000 Monte Carlo samples of sizes $n = 75$ and 150 are generated from each type of Archimedean copulas. All computations were performed in R. We used the package "copula" for generating the data. Also, "nloptr" package was used for solving non-linear equations.

Table 2: Estimated MSE's scores of Gumbel copula parameter estimators for sample size $n=75$.

τ	$\hat{\theta}_n^\tau$	$\hat{\theta}_n^{MLE}$	$\hat{\theta}_{n,m=10}^{CvM}$	$\hat{\theta}_{n,m=20}^{CvM}$	$\hat{\theta}_{n,m=30}^{CvM}$
0.1	0.0089	0.0096	0.0239	0.0116	0.0081
0.2	0.0159	0.0171	0.0320	0.0191	0.0114
0.3	0.0255	0.0282	0.0452	0.0287	0.0245
0.4	0.0407	0.0426	0.0639	0.0429	0.0375
0.5	0.0696	0.0706	0.0949	0.0697	0.0622
0.6	0.1229	0.1129	0.1450	0.1149	0.1094
0.7	0.2544	0.2162	0.2411	0.2229	0.2145
0.8	0.5999	0.4525	0.5154	0.5096	0.5848

The results given in Tables 2-7 show that, as expected, estimated MSE's of parameter estimation of all Archimedean copulas decrease as sample size increases. This is required by the consistency of the Bernstein estimate of Kendall distribution function. On the other hand, estimated MSE's scores increase with increasing Kendall's tau for all Archimedean copulas and estimation methods. For sample size 75, Minimum Cramér-von-Mises estimator outperforms other estimation methods. Minimum Cramér-von-Mises performed better than others due to its flexibility since by increasing polynomial

degree estimated MSE's scores can be decreased for Minimum Cramér-von-Mises estimator. For sample size 150, the Maximum Likelihood estimator outperforms the other estimators.

Table 3: Estimated MSE's scores of Gumbel copula parameter estimators for sample size n=150.

τ	$\hat{\theta}_n^\tau$	$\hat{\theta}_n^{MLE}$	$\hat{\theta}_{n,m=10}^{CvM}$	$\hat{\theta}_{n,m=20}^{CvM}$	$\hat{\theta}_{n,m=30}^{CvM}$
0.1	0.0041	0.0044	0.0170	0.0068	0.0068
0.2	0.0071	0.0074	0.0234	0.0107	0.0079
0.3	0.0118	0.0117	0.0316	0.0159	0.0124
0.4	0.0188	0.0181	0.0450	0.0243	0.0193
0.5	0.0311	0.0281	0.0628	0.0375	0.0313
0.6	0.0577	0.0511	0.0945	0.0641	0.0575
0.7	0.1121	0.0945	0.1555	0.1158	0.1125
0.8	0.2706	0.2226	0.2859	0.2627	0.3144

Table 4: Estimated MSE's scores of Clayton copula parameter estimators for sample size n=75.

τ	$\hat{\theta}_n^\tau$	$\hat{\theta}_n^{MLE}$	$\hat{\theta}_{n,m=10}^{CvM}$	$\hat{\theta}_{n,m=20}^{CvM}$	$\hat{\theta}_{n,m=30}^{CvM}$
0.1	0.0366	0.0359	0.0756	0.0394	0.0303
0.2	0.0626	0.0606	0.0985	0.0631	0.0541
0.3	0.1089	0.0963	0.1507	0.1063	0.0953
0.4	0.1552	0.1386	0.1909	0.1479	0.1356
0.5	0.2930	0.2359	0.3347	0.2694	0.2210
0.6	0.4803	0.3770	0.5050	0.4346	0.4185
0.7	1.0071	0.7648	1.0038	0.8952	0.9030
0.8	2.6822	2.1405	2.5275	2.3589	2.6595

4 Real Data Examples

To demonstrate the new estimation methods presented in the previous sections, we fit the Gumbel, Clayton and Frank copula to the following two real data sets in this

section:

1. According to the manual of R's package "MASS", the US National Institute of Diabetes and Digestive and Kidney Diseases collected a data set from the population of women (at least 21 years old, of Pima Indian heritage and living near Phoenix, Arizona) who were tested for diabetes according to World Health Organization criteria. The data set "Pima.tr" contains a randomly selected set of 200 subjects.
2. According to the manual of R's package "boot", the excess return for the Acme Cleveland Corporation are recorded along with those for all stocks listed on the New York and American Stock Exchanges recorded over five years. These excess returns are relative to the return on a risk-less investment such a U.S. Treasury bills. The data set contains 60 samples of each variable.

Table 5: Estimated MSE's scores of Clayton copula parameter estimators for sample size $n=150$.

τ	$\hat{\theta}_n^\tau$	$\hat{\theta}_n^{MLE}$	$\hat{\theta}_{n,m=10}^{CoM}$	$\hat{\theta}_{n,m=20}^{CoM}$	$\hat{\theta}_{n,m=30}^{CoM}$
0.1	0.0167	0.0154	0.0555	0.0227	0.0138
0.2	0.0292	0.0261	0.0667	0.0337	0.0231
0.3	0.0454	0.0385	0.0914	0.0526	0.0427
0.4	0.0780	0.0602	0.1266	0.0842	0.0727
0.5	0.1263	0.1007	0.1937	0.1362	0.1193
0.6	0.2491	0.1946	0.3437	0.2592	0.2325
0.7	0.5183	0.3960	0.6444	0.5206	0.4881
0.8	1.3206	1.0284	1.4690	1.2783	1.3155

The scatterplots of the data sets are shown in Figure 1. In Figure 1, it is observed that there is a visible dependence structure of two data sets. Kendall's tau was measured as 0.1842 for variables between body mass index and dialectic blood pressure for Pima Indian women, and as 0.3649 for variables between the excess return of the market and excess return for the Acme Corporation.

We apply goodness of fit testing procedure based on parameter estimations methods introduced in section 2. Genest et al. (2008) proposed a test statistic for testing the null

Table 6: Estimated MSE's scores of Frank copula parameter estimators for sample size $n=75$.

τ	$\hat{\theta}_n^\tau$	$\hat{\theta}_n^{MLE}$	$\hat{\theta}_{n,m=10}^{CvM}$	$\hat{\theta}_{n,m=20}^{CvM}$	$\hat{\theta}_{n,m=30}^{CvM}$
0.1	0.5478	0.5592	0.8052	0.4417	0.4710
0.2	0.5675	0.5733	0.8467	0.5171	0.5673
0.3	0.6827	0.6820	0.9449	0.6246	0.6799
0.4	0.7618	0.7412	0.9538	0.6707	0.7217
0.5	1.0855	1.0342	1.2285	0.9217	1.0109
0.6	1.5555	1.4228	1.5745	1.3810	1.3867
0.7	2.6039	2.2988	2.4870	2.2182	2.8495
0.8	6.1027	5.1382	6.7891	11.0283	8.0771

Table 7: Estimated MSE's scores of Frank copula parameter estimators for sample size $n=150$.

τ	$\hat{\theta}_n^\tau$	$\hat{\theta}_n^{MLE}$	$\hat{\theta}_{n,m=10}^{CvM}$	$\hat{\theta}_{n,m=20}^{CvM}$	$\hat{\theta}_{n,m=30}^{CvM}$
0.1	0.2527	0.2572	0.6133	0.3513	0.2980
0.2	0.2859	0.2848	0.6319	0.3713	0.3298
0.3	0.3269	0.3225	0.6242	0.3887	0.3591
0.4	0.3931	0.3822	0.6660	0.4385	0.4156
0.5	0.5323	0.5214	0.8532	0.5749	0.5612
0.6	0.7310	0.6899	0.9752	0.7162	0.7217
0.7	1.2181	1.1520	1.5066	1.1826	1.1914
0.8	2.6545	2.3927	3.0553	3.2469	4.2071

hypothesis $H_0 : K(t) = K_{\hat{\theta}}(t)$ as

$$CvM = \int_0^1 n(K_n(t) - K_{\hat{\theta}}(t))^2 dK_{\hat{\theta}}(t),$$

P-values are obtained by parametric bootstrap method as described in Genest et al. (2008).

Estimations and Goodness of fit results for two data sets presented in Tables 8, 9. When Table 8 is examined, it is seen that CvM measure and associated P-values have approximately the same values in almost all cases. Despite this, CvM measure based

on minimum distance estimator is slightly lower than the other estimation methods. These results are similar to the results obtained by Monte Carlo simulations for a small sample size and $\tau = 0.4$ in section 3. In Table 9, small CvM measures are obtained for minimum distance estimators with polynomial degree $m = 30$. Unlike simulation results for big sample size and $\tau = 0.2$, minimum distance estimator has lower CvM measure than the others. Also, we note that CvM measure decreases with increasing polynomial degree for minimum distance estimator for two data sets. Also, Figures 2-3 visualizes the fit on the empirical λ function. In Figures 2-3, empirical λ functions are presented by solid lines and parametric estimation of λ functions with parameters estimated by minimum distance estimator are presented by dashed lines. Three parametric estimation of λ functions are presented in Figures 2-3. We estimated Clayton and Frank λ function's parameters using minimum distance estimator with Bernstein polynomial degree $m = 10, 20, 30$ for the two data sets. Estimated parameters are given in Tables 8-9. We preferred λ functions as a visual comparison on the goodness of fit performance for two real data sets; because $\lambda(t) = \frac{\varphi(t)}{\varphi'(t)}$ function is visually more informative than the Archimedean generator φ and Kendall distribution K . For more details about λ function, see Michiels et al. (2012). From Figures 2-3, when the polynomial degree m increases, parametric estimation of Clayton and Frank λ functions get closer to the empirical λ function of the two data sets.

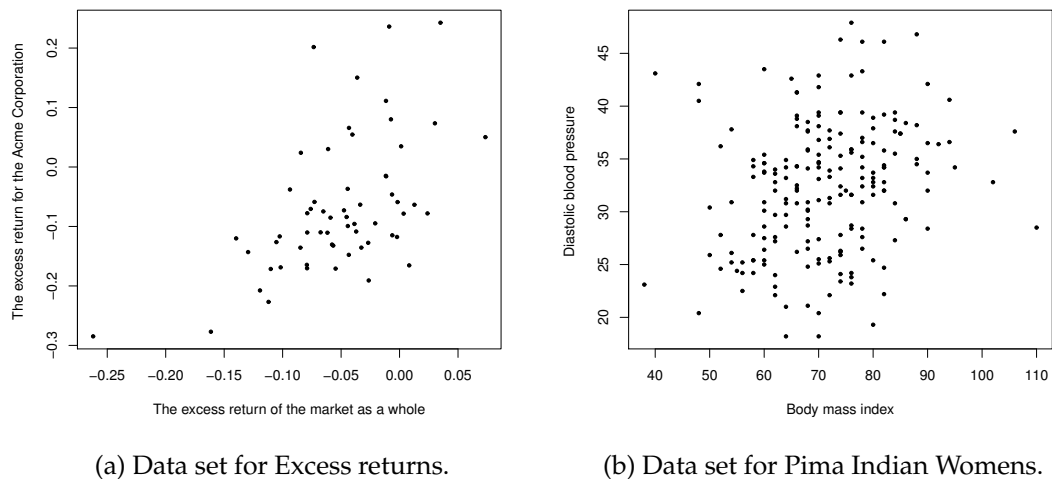


Figure 1: Scatter plots of the data sets.

Table 8: Estimation and goodness of fit results for excess return for the Acme Cleveland.

Copula		$\hat{\theta}_n^\tau$	$\hat{\theta}_n^{MLE}$	$\hat{\theta}_{n,m=10}^{CvM}$	$\hat{\theta}_{n,m=20}^{CvM}$	$\hat{\theta}_{n,m=30}^{CvM}$
Gumbel	Estimation	1.5747	1.5515	1.8118	1.6860	1.6731
	CvM	0.1714	0.1799	0.1365	0.1456	0.1470
	P-Value	0.0253	0.0172	0.0282	0.0328	0.0331
Clayton	Estimation	1.1494	1.3519	1.4678	1.2492	1.2303
	CvM	0.0463	0.0529	0.0642	0.0472	0.0417
	P-Value	0.7401	0.6320	0.4171	0.6286	0.7682
Frank	Estimation	3.6969	3.7978	4.4861	3.9375	3.9091
	CvM	0.1391	0.1361	0.1333	0.1332	0.1331
	P-Value	0.0415	0.0525	0.0373	0.0315	0.0561

Table 9: Estimation and goodness of fit results for Pima Indian Womens.

Copula		$\hat{\theta}_n^\tau$	$\hat{\theta}_n^{MLE}$	$\hat{\theta}_{n,m=10}^{CvM}$	$\hat{\theta}_{n,m=20}^{CvM}$	$\hat{\theta}_{n,m=30}^{CvM}$
Gumbel	Estimation	1.2258	1.1607	1.3064	1.2231	1.1953
	CvM	0.0602	0.0685	0.1261	0.0593	0.0559
	P-Value	0.4714	0.4971	0.4981	0.6535	0.6286
Clayton	Estimation	0.4517	0.3400	0.5576	0.4055	0.3489
	CvM	0.1517	0.0768	0.2776	0.1113	0.0730
	P-Value	0.0754	0.3145	0.0371	0.2031	0.3525
Frank	Estimation	1.7055	1.6517	2.0886	1.6083	1.5139
	CvM	0.0618	0.0548	0.1415	0.0501	0.0422
	P-Value	0.4334	0.5322	0.3651	0.8121	0.8020

5 Concluding Comments

In this study, we propose a method of estimating the parameter for some bivariate Archimedean copula. We estimate the dependence parameter using Cramér-von-Mises distance and Bernstein polynomial approximation. We also use the MSE values to measure the performance of the estimators. Our simulation suggests that the Minimum Cramér-von-Mises estimation method has a good performance for low dependence and small sample size when compared with the other estimation methods.

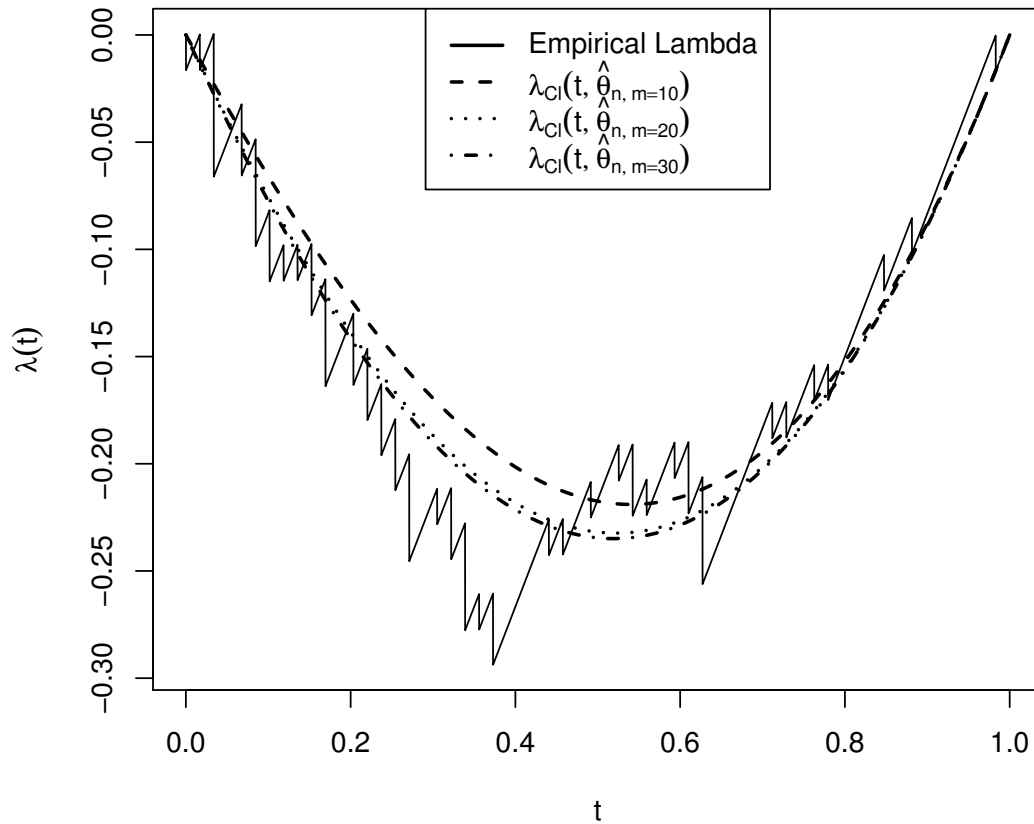


Figure 2: Empirical λ and parametric estimation of Clayton λ functions with parameter obtained by minimum distance estimator for the Acme Cleveland Corporation data set.

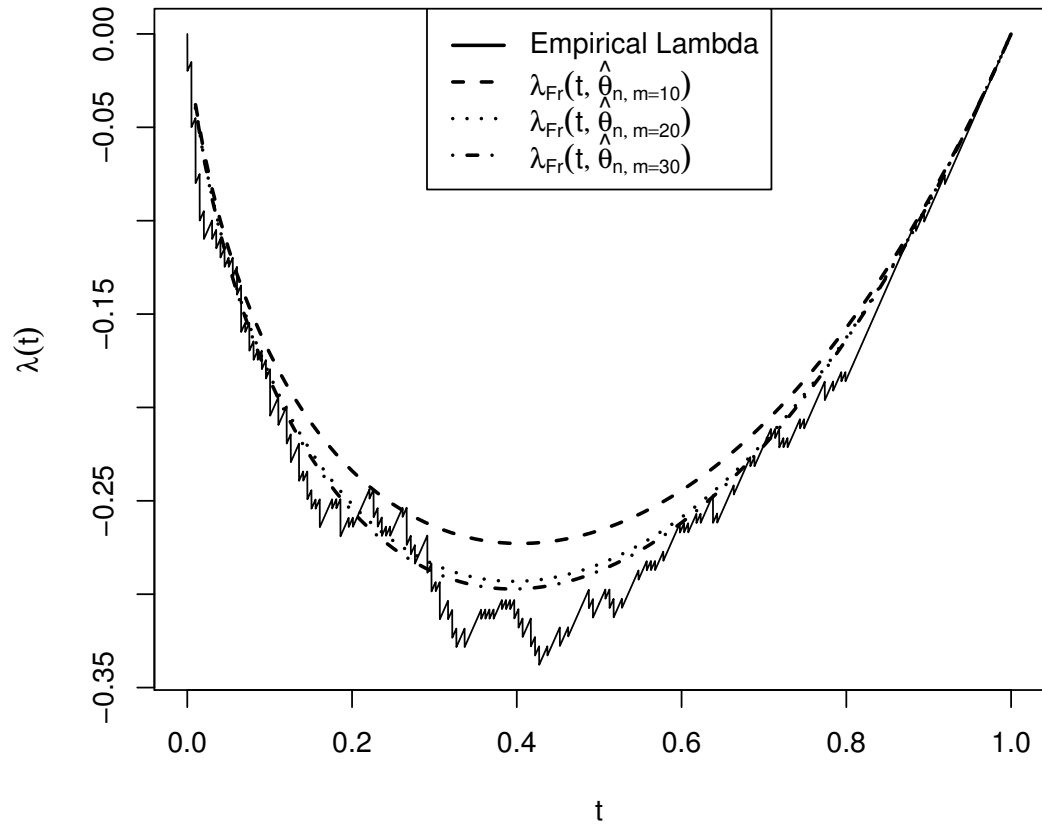


Figure 3: Empirical λ and parametric estimation of Frank λ functions with parameter obtained by minimum distance estimator for the Pima Indian woman data set.

However, this result is no longer valid at higher Kendall's tau values. We established estimations and Goodness of fit results for two data sets. CvM measure based on the minimum distance estimator is slightly lower than the other estimation methods for a small sample size. Unlike simulation results for big sample size, minimum distance estimator has lower CvM measure than the others. In practice, weight selection is important for specific types of dependence. For example, Clayton copula has a heavy concentration of probability near $(0, 0)$. On the contrary, Gumbel copula has a heavy concentration of probability near $(1, 1)$. Also, Frank copula exhibits symmetric dependence. If one can propose a method, which minimizes "weighted" Cramér-von-Mises distance according to the dependence structure of data, it should most probably decrease estimated MSE's scores.

Acknowledgments

We thank the anonymous referees and the editor for their helpful suggestions which improved the presentation of the paper.

References

- Biau, G., and Wegkamp, M. (2005), Minimum distance estimation of copula densities. *Statistics & Probability letters*, **73**, 105–114.
- Deheuvels, P. (1978), Caractérisation complète des lois extrêmes multivariées et de la convergence des types extrêmes. *Publications de l'Institut de Statistique de l'Université de Paris*, **3**, 1–36.
- Duchesne, T., Rioux, J., and Luong, A. (1997), Minimum Cramér-von Mises distance methods for complete and grouped data. *Communications in statistics-theory and methods*, **26**, 401–420.
- Fermanian, J. D., Radulovic, D., and Wegkamp, M. (2004), Weak convergence of empirical copulaprocesses. *Bernoulli*, **10**, 847–860.
- Genest, C., and Mackay, R. J. (1986), Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *Canadian journal of statistics*, **14**, 145–159.
- Genest, C., and Rivest, L. (1993), Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association*, **88**, 1034–1043.

- Genest, C., Molina, J., and Lallena, J. (1995), De l'impossibilite de construire des lois a marges multidimensionnelles donnees a partir de copules. *Comptes rendus de l'Academie des Sciences*, **320**, 723–726.
- Genest, C., Rémillard, B., and Beaudoin, D. (2008), Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, **44**, 199–213.
- Joe, H. (1978), Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, **94**, 401–419.
- Joe, H., and Xu, J. (1996), The estimation method of inference functions for margins for multivariate models. *Technical Report*, **166**, The University of British Columbia.
- Joe, H. (1997), *Multivariate Models and Dependence Concepts*. London, England: Chapman and Hall.
- Joe, H. (2005), Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, **94**, 401–419.
- Kim, G., Silvapulle, M., and Silvapul, P. (2007), Comparison of semiparametric and parametric methods for estimating copulas. *Communications in Statistics: Simulation and Computation*, **51**, 2836–2850.
- Kojadinovic, I., and Yan, J. (2010), Comparison of three semiparametric methods for estimating dependence parameters in copula models. *Insurance: Mathematics and Economics*, **47**, 52–63.
- Leblanc, A. (2012), On estimating distribution functions using bernstein polynomials. *Annals of the Institute of Statistical Mathematics*, **64**, 919–943.
- Mendes, B., De Melo B., and Nelsen, R. (2007), Robust fits for copula models. *Communications in statistics: Simulation and computation*, **36**(5), 997–1017.
- Michiels, F., Koch, I., and De Schepper, A. (2012), How to improve the fit of Archimedean copulas by means of transforms. *Statistical Papers*, **53**, 345–355.
- Najafabadi, A., Farid-Rohani, M., and Qazvini, M. (2013), A GLM-Based Method to Estimate a Copula's Parameter(s). *Journal of the Iranian statistical society*, **12**(2), 321–334.
- Nelsen, R. B. (2006), *An introduction to copulas*. New York, USA: Springer.
- Oakes, D. (1994), Multivariate survival distributions. *Journal of Nonparametric Statistics*, **3**, 343–354.

Sklar, A. (1959), Fonctions de répartition n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, **8**, 229–231.

Susam, S. O., and Ucer, B.U. (2018), Testing independence for Archimedean copula based on Bernstein estimate of Kendall distribution function. *Journal of Statistical Computation and Simulation*, **88**, 2589–2599.

Tsukahara, H. (2005), Semiparametric estimation in copula models. *Canadian Journal of Statistics*, **33**(3), 357–375.

Weib, G. (2011), Copula parameter estimation by maximum-likelihood and minimum-distance estimators: a simulation study. *Computational Statistics*, **26**, 31–54.

Appendix

Proof. 1 (Proof of Lemma 2.1). Cramer-Von-Mises distance CvM_G can be written as

$$CvM_{Gu} = \int_0^1 (K_{n,m}(t) - K_{Gu}(t))^2 dt.$$

Kendall distribution function of Gumbel copula is give as

$$K_{Gu}(t) = t(1 - \frac{t \log(t)}{\theta}).$$

Then Cramér-von-Mises distance is given as

$$\begin{aligned} CvM_{Gu} &= \int_0^1 (K_{n,m}(t) - t(1 - \frac{t \log(t)}{\theta}))^2 dt \\ &= \int_0^1 K_{n,m}^2(t) dt + n \int_0^1 (t(1 - \frac{t \log(t)}{\hat{\theta}}))^2 dt - 2n \int_0^1 K_{n,m}(t)(t(1 - \frac{t \log(t)}{\theta})) dt \\ &= \int_0^1 (\sum_{k=0}^m \binom{m}{k} t^k (1-t)^{m-k} K_n(\frac{k}{m}))^2 dt + n \int_0^1 (t(1 - \frac{t \log(t)}{\theta}))^2 dt \\ &\quad - 2 \sum_{k=0}^m K_n(\frac{k}{m}) \binom{m}{k} \int_0^1 t^k (1-t)^{m-k} (t(1 - \frac{t \log(t)}{\theta})) dt \\ &= I_1 + I_2 - I_3. \end{aligned}$$

Now, we calculate part of I_1 . We know that, $(a_1+a_2+\dots+a_n)^2 = \sum_{i=1}^n a_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n a_i a_j$, then we can write

$$\begin{aligned} I_1 &= \sum_{k=0}^m \binom{m}{k}^2 K_n^2\left(\frac{k}{m}\right) \int_0^1 t^{2k}(1-t)^{2m-2k} dt \\ &+ 2 \sum_{k=0}^{m-1} \sum_{s=k+1}^m \binom{m}{k} K_n\left(\frac{k}{m}\right) \binom{m}{s} K_n\left(\frac{s}{m}\right) \int_0^1 t^{k+s}(1-t)^{2m-k-s} dt \\ &= \sum_{k=0}^m \binom{m}{k}^2 K_n^2\left(\frac{k}{m}\right) \beta(2k+1, 2m-2k+1) \\ &+ 2 \sum_{k=0}^{m-1} \sum_{s=k+1}^m \binom{m}{k} K_n\left(\frac{k}{m}\right) \binom{m}{s} K_n\left(\frac{s}{m}\right) \beta(k+s+1, 2m-k-s+1). \end{aligned}$$

Now, we calculate the part of I_3

$$\begin{aligned} I_3 &= 2 \sum_{k=0}^m \binom{m}{k} K_n\left(\frac{k}{m}\right) \int_0^1 t^k(1-t)^{m-k} \left(t(1 - \frac{t \log(t)}{\theta})\right) dt \\ &= 2 \sum_{k=0}^m \binom{m}{k} K_n\left(\frac{k}{m}\right) \left(\int_0^1 t^{k+1}(1-t)^{m-k} dt - \int_0^1 t^{k+1}(1-t)^{m-k} \frac{\log(t)}{\theta} dt \right) \\ &= 2 \sum_{k=0}^m \binom{m}{k} K_n\left(\frac{k}{m}\right) \left(\beta(k+2, m-k+1) - \int_0^1 t^{k+1}(1-t)^{m-k} \frac{\log(t)}{\theta} dt \right). \end{aligned}$$

If we get binomial expansion term $(1-t)^{m-k}$ in the right side of equation

$$I_3 = 2 \sum_{k=0}^m \left(\binom{m}{k} K_n\left(\frac{k}{m}\right) \left(\beta(k+2, m-k+1) - \sum_{i=0}^{m-k} \frac{(-1)^i}{\theta} \binom{m-k}{i} \int_0^1 t^{k+1+i} \log(t) dt \right) \right).$$

Because $\log(0)$ does not exist, 0 is not the domain of the integrand. This is an improper integral, so we use

$$I_3 = 2 \sum_{k=0}^m \left(\binom{m}{k} K_n\left(\frac{k}{m}\right) \left(\beta(k+2, m-k+1) - \sum_{i=0}^{m-k} \frac{(-1)^i}{\theta} \binom{m-k}{i} \lim_{a \rightarrow 0} \int_a^1 t^{k+1+i} \log(t) dt \right) \right).$$

Recalling that $\log(1) = 0$, we get partial integral of integrand ($u = \log(t)$, $dv = t^{k+1+i} dt$)

$$I_3 = 2 \sum_{k=0}^m \left(\binom{m}{k} K_n \left(\frac{k}{m} \right) \left(\beta(k+2, m-k+1) - \sum_{i=0}^{m-k} \frac{(-1)^i}{\theta} \binom{m-k}{i} \left(-\frac{1}{(k+i+2)^2} \right. \right. \right. \\ \left. \left. \left. - \lim_{a \rightarrow 0} \left(\frac{\log(a)}{k+i+2} \right) \right) \right) \right),$$

$\lim_{a \rightarrow 0} \left(\frac{\log(a)}{k+i+2} \right)$ is indeterminate because it equals to $\frac{-\infty}{+\infty}$. Applying L'Hopital's rule gives us

$$I_3 = 2 \sum_{k=0}^m \left(\binom{m}{k} K_n \left(\frac{k}{m} \right) \left(\beta(k+2, m-k+1) - \sum_{i=0}^{m-k} \frac{(-1)^i}{\theta} \binom{m-k}{i} \left(-\frac{1}{(k+i+2)^2} \right) \right) \right).$$

Finally, after some algebra I_2 is calculated by

$$\frac{2 + 6\theta + 9\theta^2}{27\theta^2}.$$

And also, Cramér-Von-Mises distance CvM_C can be written as

$$CvM_{Cl} = \int_0^1 (K_{n,m}(t) - K_{Cl}(t))^2 dt.$$

$K_{Cl}(t)$ can be rewritten as

$$K_{Cl}(t) = t + \frac{t}{\theta} - \frac{t^{\theta+1}}{\theta}.$$

Then Cramér-von-Mises distance is given as

$$\begin{aligned} CvM_{Cl} &= \int_0^1 (K_{n,m}(t) - (t + \frac{t}{\theta} - \frac{t^{\theta+1}}{\theta}))^2 dt \\ &= \int_0^1 K_{n,m}^2(t) dt + n \int_0^1 (t + \frac{t}{\theta} - \frac{t^{\theta+1}}{\theta})^2 dt - 2n \int_0^1 K_{n,m}(t) (t + \frac{t}{\theta} - \frac{t^{\theta+1}}{\theta}) dt \\ &= \int_0^1 \left(\sum_{k=0}^m \binom{m}{k} t_k (1-t)^{m-k} K_n \left(\frac{k}{m} \right) \right)^2 dt + n \int_0^1 (t + \frac{t}{\theta} - \frac{t^{\theta+1}}{\theta})^2 dt \\ &\quad - 2 \sum_{k=0}^m \binom{m}{k} K_n \left(\frac{k}{m} \right) \int_0^1 t^k (1-t)^{m-k} (t + \frac{t}{\theta} - \frac{t^{\theta+1}}{\theta}) dt \\ &= I_1 + I_2 - I_3, \end{aligned}$$

part of the I_1 is the same as in the proof of CvM_{Gu} . Now we calculate the part of I_3

$$\begin{aligned}
 I_3 &= 2 \sum_{k=0}^m \binom{m}{k} K_n\left(\frac{k}{m}\right) \int_0^1 t^k (1-t)^{m-k} \left(t + \frac{t}{\theta} - \frac{t^{\theta+1}}{\theta}\right) dt \\
 &= 2 \sum_{k=0}^m \binom{m}{k} K_n\left(\frac{k}{m}\right) \left(\left(1 + \frac{1}{\theta}\right) \int_0^1 t^{k+1} (1-t)^{m-k} dt - \frac{1}{\theta} \int_0^1 t^{k+\theta+1} (1-t)^{m-k} dt \right) \\
 &= 2 \sum_{k=0}^m \binom{m}{k} K_n\left(\frac{k}{m}\right) \left(\frac{\theta+1}{\theta} \beta(k+2, m-k+1) - \frac{1}{\theta} \beta(k+\theta+2, m-k+1) \right).
 \end{aligned}$$

Finally, after some algebra I_2 is calculated as

$$\frac{17 + 13\theta + 2\theta^2}{27(1 + \theta) + 6\theta^2}.$$