Nonparametric Distance Test and Kolmogorov-Smirnov Test to Detect Positive Quadrant Dependence in Gaussian Copula Distribution

by
Lie Leslie Liu

Submitted in partial fulfilment
of the requirement for the degree of
Master of Fine Arts in Statistics, Hunter College
The City University of New York
2017

Contents

Chapter 1

Introduction

There has been growing interest in concepts and testing methods of positive or negative quadrant dependence (PQD or NQD) in finance, insurance and actuarial science. The common testing tools for PQD available in the literature include Pearson's product-moment correlation, Kendall's tau, distance correlation, likelihood ratio dependence, Kolmogorov-Smirnov (KS) test, Cramer-von Mises test, Anderson-Darling test coupling with copula-based data. The primary aim of this paper is to use real data (closing prices of two stocks) to test PQD against independence by using a nonparametric distance test then a KS-type test. Both are based on the past literature. In the introductory section, one will review certain commonly used concepts, such as PQD, copula, PUOD and PLOD, then will review past literature on measuring dependence results on PQD.

Throughout the paper, one should use the following notations: H will be the joint distribution while A and B are the unknown marginal distributions for random variables (x,y). Let C be any copulas and I is denoted as independent where $all\ u, v \in [0, 1]\ and\ I(u, v) = uv$. F and G will serve as cumulative distribution functions.

Lehmann (1966) formally defined the condition of PQD as: when X, Y are the random variables, H(x, y) ≥ A(x)B(y) for all (x, y) ∈ R. For NQD, the sign "≥" will be reversed. Additionally, Cov[A(x), B(y)] ≥ 0 in the condition of PQD while equality only if X, Y are independent when F and G are nondecreasing real functions (see Shetty, 2003 and Pandit and Kumari, 2013).

PQD exists not only in a Euclidean space but also in a n-dimension space of random variables. First of all, PQD is a property of the underlying copulas, which are in the forms of dynamic correlations (including nonlinear). Recall that a bivariate copula is a function C in the unit square [0,1]x[0,1] with the following properties:

1. For $all\ u, v \in [0, 1]$, C(u,0)=0 and C(0,v)=0 .
2. When C is increasing, $u_1, u_2,\ v_1, v_2 \in [0, 1]$ such that $u_1 \le u_2$ and $v_1 \le v_2$ then $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \ge 0$
3. For $all\ u, v \in [0, 1]$, C(u,1)=u and C(1,v)=v. If either one of the marginal is 1, then the joint distribution behaves as a univariate distribution of opposite variable.

Sklar's Theorem states that every bivariate joint distribution can be composed of some copula C and marginal distribution functions. When

$$C(u,v)=H(A(u),\ B(v)),$$

then PQD can be expressed as follows:

$$C(u, v) \ge I(u, v) \text{ for all } u, v \in [0, 1].$$

Recall that given H, the copula is unique on the cartesian product of the ranges of the marginal CDF's, which are continuous. The uniqueness of copulas, in other words, is that any random pair (x,y) will be invariant in the monotonically increasing transformation of the marginals (Genest and Favre, 2007).

Per Joe (1997) and Nelsen (1999), a direct extension of PQD is positive orthant dependence (POD), including positively lower orthant dependent (PLOD) and positively upper orthant dependent (PUOD). Their definitions are as follows (see Denuit, 2004): the PLOD is

$$C(u) \ge \prod_{j=1}^{n} u_{j,}$$

where the PUOD is given by

$$\overline{C}(u) \geq \prod_{j=1}^{n}(1-u_j), \text{ for all } u \in [0,1]^n .$$

Additionally, the equations are no longer equivalent when $n \geq 3$ (Denuit, 2004). Fernández and González-Barrios(2004) incorporated the Glivenko Cantelli theorem to test multidimensional dependency. The PUOD, PLOD and their corresponding tests are beyond the scope of this paper.

Regarding measuring tests for two random variables, Lehmann (1966) stated that Kendall's τ , Spearman's ρ , normal score and Blomqvist's q are non-negative if the joint distribution is positively quadrant dependent. The mentioned tests detect the condition where large (small respectively) values of Y tend to associate with large (small respectively) values of X. For details regarding the proofs, lemmas and corollaries see Lehmann (1966).

In light of Lemann's statements on PQD testing, many researchers have investigated statistics of independence against PQD. Both Kendall's τ and Spearman's ρ are common nonparametric measurements (or rank tests) of dependence. Ledwina (1986) evaluated the Bahadur efficiency of the two mentioned rank tests, to detect PQD in the Morgenstern, Woodworth, Farlie and Marshall-Olkin distributions. By calculating the Bahadur efficiency, the p-value reveals the convergence rates to reject the null hypothesis (Ledwina, 1986) . Nowadays researchers prefer Monte Carlo methods to Bahadur efficiency computations when testing asymptotic properties due to the calculation complexity. Additionally, the study has introduced the statistics L^+ related to the monotonic dependence function and found that both Kendall's τ and Spearman's ρ are "asymptotically locally powerful."

Schriever (1987) showed that rank statistics (including Kendall's τ, Spearman's ρ, Fisher-Yates' normal score and many more) are unbiased as measures of dependence preserving the ordering stochastically (the $\geq_a$; or "more associated than PQD). The positive dependency with ordering-preserving, in this research, is more difficult to assess than PQD is.

In 2013, Pandit and Kumari proposed U statistics corresponding to the kernel Φ_k (inspired by the research from Kochar and Gupta, 1987 and Shetty Pandit, 2003) with respect to Kendall's τ in the Morgenstern and exponential distributions. The proposed u_{k+1} statistics, proposed by Kochar and Gupta (1987), in a continuous exponential distribution shows that Kendall type tests are distribution-free, unbiased and consistent. Considering both Morgenstern and exponential distributions along with Gumbel model, researchers, including Shetty and Pandit (2003), have proposed distribution-free rank tests to test on PQD based on ordering of observations in sub-samples (based on the r^{th} and s^{th} matched pairs). The above mentioned researchers have concluded that their proposed test (modified from U statistic) is unbiased, consistent for testing on PQD.

The nonparametric testing and theory of copulas, after the publishing of Joe (1997) and of Nelsen (1999), have significantly impacted the testing of PQD. The use of copulas can be considered a flexible modeling, but not assessing, tool to detect dependence. Detailed explanations of copulas, their properties and applications will be beyond the scope of this paper (see Joe,1997; Nelsen, 1999 and Genest and Favre, 2007).

A nonparametric testing proposal of Denuit and Scaillet (2004) included distance and intersection-union tests, based on the minimum of t-statistics, in the copula based data. In their testing procedure, Denuit and Scaillet (2004) used the weight $\omega\,(d,\ d-i, \widehat{V}_K)$, in the distribution and the p-value selection, proposed by Kodde and Palm(1986). The numerical determination of weight and the upper and lower bound critical values in aWald test should be used for jointly testing nonlinear equality or inequality in an asymptotic distribution (Kodde and Palm.1986). According to Scaillet (2005), the drawbacks of mentioned proposed tests include test inconsistency related to selection of the chosen points.

To achieve test consistency, Scaillet (2005) proposed a Kolmogorov-Smirnov (KS) type test by using a weighted supremum test statistic $\sqrt{n}\, sup(D_n w)$ to test PQD in two copula-based real datasets. The researcher used simulation-based multiplier and bootstrap methods to simulate P-values in Frank, Gaussian and Farlie-Gumbel-Morgenstern copulas for statistical inference. Gijbels et al. (2010) and Gijbels and Sznajder (2013) have pointed out that the multiplier and bootstrap resampling methods in Scaillet (2005)'s study are not conducted under the null hypothesis. Additionally, the criticism from Ledwina and Wyłupek (2014) is that the KS type test rejects the null hypothesis with large numbers. Regardless the disapprovals, the use of multipliers and bootstrap methods for p-values and goodness-of-fit on KS test have been adapted into the R packages, "Copula" and "VineCopula."

Gijbels et al. (2010) and Gijbels and Sznajder (2013) considered KS, Cramér–von Mises (CvM) and Anderson-Darling (AD) test statistics to detect PQD by conducting a cross-estimator analysis against the two proposed methods of Scaillet (2005). Based on the cross-estimator analysis, the researchers (Gijbels et al., 2010 and Gijbels and Sznajder, 2013) have recommended the use of mirror reflection shrunken or local linear shrunken estimators coupled with AD or CvM distance measures. Per Ledwina and Wyłupek (2014), the aforementioned tests have rejected the null hypothesis with large values of CvM and AD and that the p-values (Gijbels and Sznajder, 2013) of any swapped pair orders, such as (B, C) and (C, B), are only slightly different.

From the viewpoints of Ledwina and Wyłupek (2014), the previous methods of detecting PQD (from 2004-2010) are based solely on a measure of discrepancy between the copula estimator and the independence copula. Alternatively, they proposed two new tests M_d and AD* by elaborating the previous independence tests (proposed by Kallenberg and Ledwina (1999), Janic-Wróblewska et al (2004) and so on) to investigate the structure of C(u,v)and its relation to I. Their motivation was to study the "successive correlation coefficient of $f_r(U)$ and $f_s(V)$ for a rich class of function $f_j{}'$s in $L_2([0, 1],\ du)$" where their null hypothesis is Cov[$f_1(U)$, $f_2(V)$] ≥ 0. Kallenberg and Ledwina (1999) recommended the use of orthonormal Legendre polynomials in an exponential distribution to test independence against PQD. On the other hand, Janic-Wróblewska et al (2004) used rank tests, including their modified V^+ test, to show its improvement in standard tests for linear correlation in their generated data over other tests. Ledwina and Wyłupek (2014) found that the AD* model (their modified version) is more stable than both the AD and the M_d when the given fixed number of dyadic points is 31. Their additional finding is that AD* outperforms AD when C>I is large.

Chapter 2

Methods

For the two tests in this paper, one will use a chessboard-like grid (details located in Chapter 3 Properties and Simulations). Let's consider a fixed number of distinct points, d.

$$y_i = (y_{i1}, \cdots, y_{in}) \in R\ n,\ i = 1, \cdots, d.$$

The first method to detect PQD is based on the proposed distance test in Denuit and Scaillet (2004). The joint empirical distribution is

$$\widehat{F}(y_i) = \frac{1}{T} \sum_{t=1}^{T} \prod_{j=1}^{n} 1\ (Y_{jt} \le y_{ij}),\ i = 1, \cdots, d.$$

Additionally, the individual empirical distribution is given by

$$\widehat{F}_j(y_{ij}) = \frac{1}{T} \sum_{t=1}^{T} 1\ (Y_{jt} \le y_{ij}),\ i = 1, \cdots, d \text{ and } j = 1, \cdots, n.$$

Their null hypothesis of PQD is

$$H_0 : D_F \ \ge\ 0, \text{ where } D_F = (D_F^1, \cdots, D_F^d) \text{ at d fixed points.}$$

The alternative hypothesis is D_F unrestricted.

Both Denuit and Scaillet (2004) and Scaillet (2005) have applied the concept of $\sqrt{n}(D_n - D)$ converging weakly to a tight mean zero Gaussian process. Denuit and Scaillet (2004) provided the asymptotic distribution of $\sqrt{T}\,(\widehat{D}_F - D_F)$, which can be considered a d-dimensional normal distribution with mean zero and covariance matrix V_F. In the i.i.d. case, and n = 2 (bivariate case), the covariance matrix V_F has elements of

$$V_{F,kl} = b_k^T \ A_{kl}\ b_l, \text{ where k,l} = 1, \cdots, d.$$

$$b_i = (1 - F_2(y_{i2}) - F_1(y_{i1}), \text{ where i} = 1, \cdots, d.$$

Additionally, A_{kl} is equal to

$$\begin{bmatrix} F(y_{k1} \wedge y_{l1}, y_{k2} \wedge y_{l2}) - F(y_{k1}, y_{k2})F(Y_{l1}, y_{l2}) & \ldots & F(y_{k1}, y_{k2} \wedge y_{l2}) - F(y_{k1}, y_{k2})F_2(y_{l2}) \\ F(y_{l1} \wedge y_{k1}, y_{l2}) - F(y_{l1}, y_{l2})F(y_{k1}) & \ldots & F(y_{k1} \wedge y_{l1}, y_{k2} \wedge y_{l2}) - F_1(y_{k1})F_2(y_{l2}) \\ F(y_{l1}, y_{l2} \wedge y_{k2}) - F(y_{l1}, y_{l2})F_2(y_{k2}) & \ldots & F_2(y_{k2} \wedge y_{l2}) - F_2(y_{k2})F_2(y_{l2}) \end{bmatrix}$$

Let $\overline{D}_F$ be the solution to the constrained quadratic minimization problem

$$inf_D\, T\,(D - \widehat{D}_F)^T \widehat{V}_F^{-1} (D - \widehat{D}_F) \text{ s.t. D} \leq 0.$$

$$\widehat{\xi}_F = T\,(D - \widehat{D}_F)^T \widehat{V}_F^{-1} (D - \widehat{D}_F)$$

$$\Pr(\widehat{\xi}_F \geq x) \approx \sum_{i=1}^{d} \Pr(\chi_i^2 \geq x) * \ \omega\,(d,\ d-i, \widehat{V}_F),$$

where the weight $\omega\,(d,\ d-i, \widehat{V}_F)$ is the probability that $\overline{D}_F$ has exactly d − i negative elements. In Denuit and Scaillet (2004), the lower bound of p-values is given as (Kodde and Palm, 1986) :

$$\tfrac{1}{2}\Pr(\chi^2\ (\mathrm{d}-1) \geq \widehat{\xi}_F\ ) + \tfrac{1}{2}\Pr(\chi^2\ (\mathrm{d}) \geq \widehat{\xi}_F\ ).$$

The second method is based on the Kolmogorov-Smirnov type test from Scaillet (2005). Let

$$D_n(u,v) = uv - C_n(u,v),\ D\,(u,v) = uv - C\,(u,v)$$

and l^∞ ([0,1]x[0,1]) will be the set of all locally bounded real functions on the unit square. Then, $\sqrt{n}(D_n - D)$ "converges weakly to a tight mean zero gaussian process, Gc, in l^∞ ([0,1]x[0,1])."

The null hypothesis of dependence is defined as follows:

$$H_0:\ C(u,v) \geq uv \; for\ all\ (u,v) \in [0,1]^2,$$

$$H_1:\ C(u,v) < uv \; for\ all\ (u,v) \in [0,1]^2,$$

According to Scaillet (2005) , if H_0 is true then

$$\lim_{n\to\infty} P(reject\ H_0)\ \leq P(\overline{S} >\ c) = \alpha(c),\ where\ \ \overline{S} = sup\, \mathrm{Gc}(u, v)\ \text{for}\ for\ (u, v)\ \in\ [0, 1]$$

Otherwise,

$$\lim_{n\to\infty} P(reject\ H_0\ )\text{=}1.$$

Their test statistic is

$$S_n = \sqrt{n}\, sup\ \ D_n(u, v),\ for\ (u, v)\ \in\ [0, 1]\,.$$

When the critical value is c, then the test will reject H_0 if S_n > c. This is because there are different p-values that correspond to copula distributions. To find the appropriate critical value c, one relies on the standard bootstrap method. Since the testing distribution for this paper is Gaussian copula, the Chi Square test will be implemented identically to the one in Denuit and Scaillet (2004) by assuming that the critical value c is 0.05. In other words, one will derive KS test probabilities from the same dataset**.**

Chapter 3

Properties and Simulations

In this section, one will use a real life PQD dataset and the artificial pseudo-random numbers of NQD to verify the performance of both tests - distance test and Kolmogorov-Smirnov type test. The data is available at http://www.nasdaq.com/quotes/historical-quotes.aspx. The data was downloaded on May 15, 2017; thus, the extracted data is the closing stock prices, reflecting the period from November 15, 2015 to May 14, 2017. In other words, the sample size of paired data n=546. The two companies are Alibaba Group Holding Ltd (NYSE: BABA) and Charles Schwab Corp (NYSE:SCHW)

As indicated in chapter 2, the null hypothesis for both tests is non-independence. To evaluate PQD, it is defined as:

$$H_0 : \ C(u, v) \geq uv \ for \ all \ (u, v) \in [0, 1]^2,$$

For NQD, the sign will be changed to "$\leq$" as the null hypothesis. Otherwise, the alternative hypothesis will be unrestricted.

This paper includes the distance test and Kolmogorov-Smirnow test, which are defined as:

$$inf_D \ T \ (D - \widehat{D}_F)^T \widehat{V}_F^{-1} (D - \widehat{D}_F) \text{ s.t. D } \leq 0$$

and

$$\text{KS} = \sqrt{n} \ sup \ (uv - \widehat{C}_n(u, v)) = \sqrt{n} \ max \left| uv - \widehat{C}_n(u, v) \right|, \text{ for } u, v \in [0, 1].$$

Chapter 4

Application

The approximation of Spearman's ρ between SCHW and BABA is 0.7153. The recommended copula selection in R for the data set is Gaussian copula with $\rho = 0.7014$. Thus, one can construct the Gaussian copula for the observed closing stock prices. The 2D and 3D images of copula-based data with a PQD property are as follows (see Figure 1).

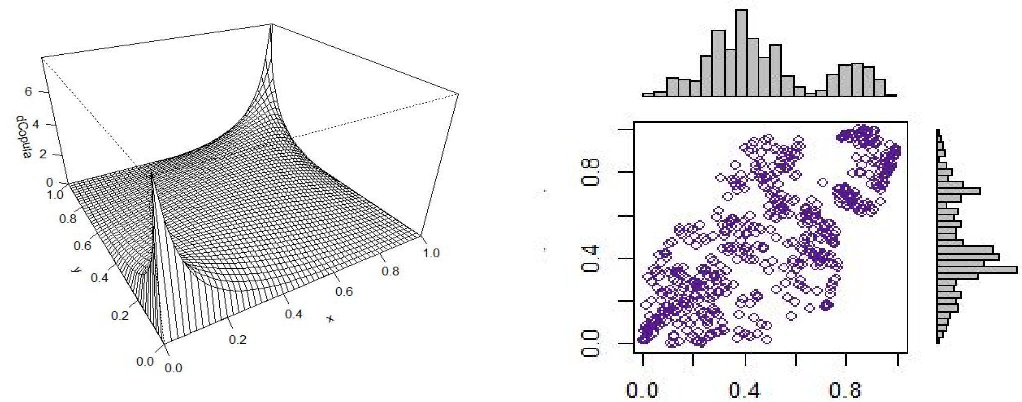


Figure 1: 3-D copula of stock closing prices (left) and 2D plot with histograms (right).

Regarding the Gaussian copula simulation of NQD, the ρ is equal to -0.4 (see Figure 2).

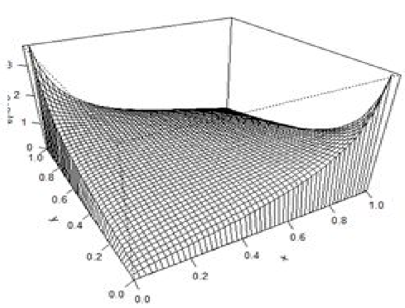


Figure 2: 3D-copula of NQD simulation.

The chessboard-like grid from Scaillet (2005) will be applied in both methods. It is made of the values (u,v) evenly placed inside {0.05, 0.010, . . ., 0.95}x{0.05, 0.010, . . ., 0.95}.

From Denuit and Scaillet's (2004) test statistic formulas, one obtains the results for the p-values of both PQD and NQD approaching 1. On the other hand, one obtains 0.2131 and 0.3676 as the p-values from KS type test for PQD and NQD, respectively. Corresponding to the specified significance level (0.05), the p-values computed from both methods indicate an acceptance of the null hypothesis of dependence.

Note that the theory underlying the bootstrap guarantee that the resulting significance levels are unbiased for a wide range of situations. When testing on the goodness-of-fit with bootstrap sampling (B = 50), the p-values on the KS test for PQD is zero while for NQD is 0.92. Per the computed outputs from R, the KS statistic for PQD is 1.4548 and NQD 0.4978. The null hypothesis of goodness-of-fit is independence. Then, the p-value results reject PQD as independent but accept NQD. Since bootstrap sampling is unbiased, there might be some technical issues in the computation.

Chapter 5

Conclusions

The main goal of this paper is to replicate and to assess a distance test and a Kolmogorov-Smirnov type test in detecting PQD and NQD against independence. One relies on the recently developed R packages ("Copula" and "VineCopula") and on past literature of PQD testing methods. Although the concepts of PQD and NQD were introduced back in 1966, the copula concepts in 1999 significantly impact the testing methods from earlier. The mentioned two tests were formulated in 2004-2005. Similar assessments of proposed testing methods, in the statistical literature, have been addressed by Gijbels et al. (2010), Gijbels and Sznajder (2013) and Ledwina and Wyłupek (2014).

Note that the aforementioned tests in this paper are applied on the Gaussian copula-based real life data and simulation with a bootstrap sample of 50. The p-values are based on the Chi Square test of independence. According to the p-value result, distance test and KS type test have worked well in detecting the PQD and NQD. When applied with the built-in goodness-of-fit test function in R, the p-value results for copula-based data with the PQD or NQD properties indicate that the KS type test method warrants further investigation. The goodness-of-fit test function was recently added to R according to the writings of Berg (2009) and Genest and Favre (2007).

Based on past literature, the apparent simplicity of the KS test can be misleading (Gijbels et al., 2010; Gijbels and Sznajder, 2013; and Ledwina and Wyłupek, 2014). Based on criticisms of independence not deriving from the null hypothesis, the author of this paper has attempted to derive KS test probabilities from the copula of original dataset. Per past literature, the theory underlying the KS test model must be derived from another dataset for the standard KS test probabilities to be applied.

Appendix

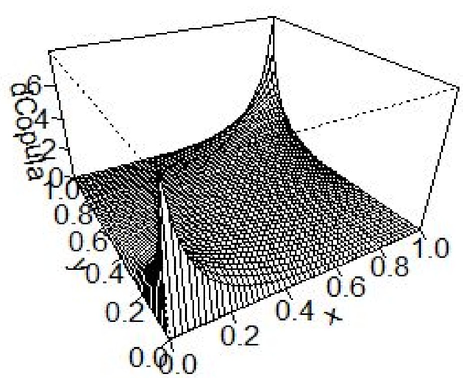
```
##################2 real stock datasets for 1.5-year daily prices##########

#since the last closing prices of May 15, 2017 were not finalized, we will
start with row from 2 to 547;
#Column 2 reflects the closing prices
baba.price<- read.csv("C:/Users/Leslie/Downloads/HistoricalQuotes
(1).csv",header = T,sep = ",")[2:547,2]
schw<- read.csv("C:/Users/Leslie/Downloads/HistoricalQuotes (2).csv",
header=T,sep="," )[2:547,2]
#datawhich is downloaded from
http://www.nasdaq.com/quotes/historical-quotes.aspx showed the daily closing
prices of
#Alibaba Group Holding Ltd(NYSE: baba) and Charles Schwab Corp (NYSE:SCHW)
reflecting the period from November 15, 2015 to May 14,2017.


Rho.s.b.<-cor(schw,baba.price,method = "spearman")
Rho.s.b.

## [1] 0.7153408

intactM<-as.matrix(cbind(schw,baba.price))
library(VineCopula)

## Warning: package 'VineCopula' was built under R version 3.1.3

library(copula)

## Warning: package 'copula' was built under R version 3.1.3

##
## Attaching package: 'copula'

## The following object is masked from 'package:VineCopula':
##
##     fitCopula

set.seed(1001)
mm<-pobs(intactM)#create pseudo observations.
u1<-mm[,1]
u2<-mm[,2]
select.cop<-BiCopSelect(u1, u2,familyset = 1:10)#familyset can be NA but it
takes much longer.
select.cop

## $p.value.indeptest
## [1] NA
##
```

```
## $family
## [1] 1
##
## $par
## [1] 0.7014716
##
## $par2
## [1] 0
##
## attr(,"class")
## [1] "BiCop"
```

```
#based on the copula outputs, family=1 Gaussian copula; rho=0.7014716; Df=0
#construct Gaussian copula but let the fit.cop to set rho based on mm.
g.cop<-normalCopula(dim = 2)
fit.cop<-fitCopula(g.cop,mm,method='ml')
rho.g.cop<-coef(fit.cop)
rho.g.cop
```

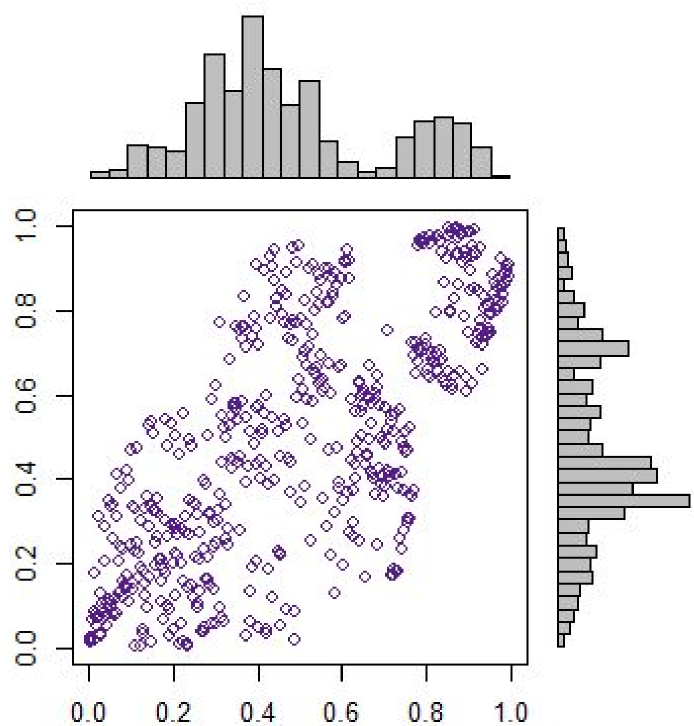
```
##     rho.1
## 0.7014668
```

```
#plot 3 D copula.

persp(normalCopula(dim=2, rho.g.cop),dCopula)
```




```
#slightly off when using rho from original data VS the one from select.cop()

PQD.matrix<-round(mm, digits=4)
```

```
PQD.cop.df<-data.frame(PQD.matrix)
colnames(PQD.cop.df)<-c("x","y")
#print(PQD.cop.df)
#str(PQD.cop.df)

#####To plot copula with histograms

xhist <- hist(intactM[,1], breaks=30, plot=FALSE)
yhist <- hist(intactM[,2], breaks=30, plot=FALSE)
top <- max(c(xhist$counts, yhist$counts))
xrange <- c(0,1)
yrange <- c(0,1)
nf <- layout(matrix(c(2,0,1,3),2,2,byrow=TRUE), c(3,1), c(1,3), TRUE)
par(mar=c(3,3,1,1))
plot(u1, u2, xlab="u1 (SCHW)", ylab="U2 (BABA)", col="purple4")
par(mar=c(0,3,1,1))
barplot(xhist$counts, axes=FALSE, ylim=c(0, top), space=0)
par(mar=c(3,0,1,1))
barplot(yhist$counts, axes=FALSE, xlim=c(0, top), space=0, horiz=TRUE)
```

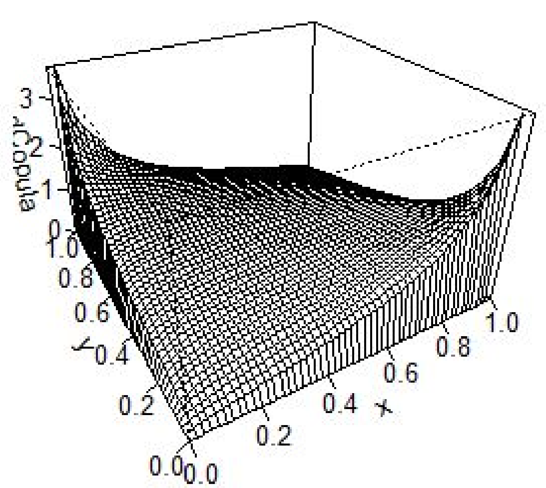



```
##build the chessboard-like grid####
a=seq(0.05, 0.95, by = 0.05)
b=a
sizeT=nrow(intactM)
grid<-data.matrix(expand.grid(a,b))# to matrix of the chessboard-like grid
```

```
#dim(grid)#[1] 361   2
#dim(PQD.matrix)#[1] 546   2

#####nonparametric distance test######
jntllF <- rep(0,4);
for (i in 1:361) {
  jntllF[i] <- sum((PQD.matrix[,1] <= grid[i,1]) * (PQD.matrix[,2] <=
grid[i,2]))/sizeT;
}
#print(jntllF)

indllF <- matrix(rep(0,722),nrow=361,ncol=2);
for (i in 1:361) {
  for (j in 1:2) {
    indllF[i,j] <- sum((PQD.matrix[,j] <= grid[i,j]))/sizeT;
  }
}
#print(indllF)

Vmatrix <- matrix(rep(0,361*361),nrow=361,ncol=361);
for (k in 1:361) {
  bveck = matrix(c(1, - indllF[k,2], -indllF[k,1]),nrow=3);
  for (l in 1:361) {
    bvecl = matrix(c(1, - indllF[l,2], -indllF[l,1]),nrow=3);
    Amat <- matrix(rep(0,3*3),nrow=3,ncol=3);
    Amat[1,1] = sum((PQD.matrix[,1] <= min(grid[k,1],grid[l,1])) *
(PQD.matrix[,2] <= min(grid[k,2],grid[l,2])))/sizeT - jntllF[k]*jntllF[l];
    Amat[1,2] = sum((PQD.matrix[,1] <= min(grid[k,1],grid[l,1])) *
(PQD.matrix[,2] <= grid[k,2]))/sizeT - jntllF[k]*indllF[l,1];
    Amat[1,3] = sum((PQD.matrix[,1] <= grid[k,1]) * (PQD.matrix[,2] <=
min(grid[k,2],grid[l,2])))/sizeT - jntllF[k]*indllF[l,2];
    Amat[2,1] = sum((PQD.matrix[,1] <= min(grid[k,1],grid[l,1])) *
(PQD.matrix[,2] <= grid[l,2]))/sizeT - jntllF[l]*indllF[k,1];
    Amat[2,2] = sum((PQD.matrix[,1] <= min(grid[k,1],grid[l,1])))/sizeT -
indllF[k,1]*indllF[l,1];
    Amat[2,3] = sum((PQD.matrix[,1] <= min(grid[k,1],grid[l,1])) *
(PQD.matrix[,2] <= min(grid[k,2],grid[l,2])))/sizeT -
indllF[k,1]*indllF[l,2];
    Amat[3,1] = sum((PQD.matrix[,1] <= grid[l,1]) * (PQD.matrix[,2] <=
min(grid[k,2],grid[l,2])))/sizeT - jntllF[l]*indllF[k,2];
    Amat[3,2] = sum((PQD.matrix[,1] <= min(grid[k,1],grid[l,1])) *
(PQD.matrix[,2] <= min(grid[k,2],grid[l,2])))/sizeT -
indllF[l,1]*indllF[k,2];
    Amat[3,3] = sum((PQD.matrix[,2] <= min(grid[k,2],grid[l,2])))/sizeT -
indllF[k,2]*indllF[l,2];
    Vmatrix[k,l] = t(bveck) %*% Amat %*% bvecl;
  }
```

```
}
V.tbl <- data.frame(Vmatrix); #the matrix computation will take a while here.
V.tbl <- round(V.tbl, digits = 4);
#print(xtable(V.tbl,digits = 4)) #Too big to print

Dvec = rep(0,4);
for (i in 1:361) {
  Dvec[i] = jntllF[i]- indllF [i,1]*indllF[i,2];
}
#print(Dvec)
library(OpenMx)

## Warning: package 'OpenMx' was built under R version 3.2.3

## OpenMx is not compiled to take advantage of computers with multiple cores.

intUniG <- sqrt(sizeT)*Dvec/sqrt(diag2vec(Vmatrix));
1 - pnorm(min(intUniG))

## [1] 0.9999457

#0.9999547(~1)=p-value. We cannot reject the null hypothesis of dependence.


##############KS test#######################
# our formula : Ks=sqrt(n) *sup(uv-c-hat(uv)) or Ks=max[sqrt(n)
*(uv-c-hat(uv))]

Dvec.Ks.test<-Dvec

intUniG.KS<- Dvec.Ks.test /sqrt(diag2vec(Vmatrix));
1 - pnorm(max(abs(intUniG.KS)))

## [1] 0.2130865

#[1] 0.2130865 as p-vaule.  cannot reject H0 of dependence.

##########simulation of NQD#####

rho.nqd<--0.4
normal.cop <- normalCopula(param=rho.nqd, dim=2)#contruct normal copula with
rho=-0.4
set.seed(987)
NQD.cop.mat<-rCopula(sizeT, normal.cop)
colnames(NQD.cop.mat)<-c("a","b")
NQD.cop.df<-data.frame(NQD.cop.mat)
NQD.cop.mat<-round(NQD.cop.mat, digits = 4)
```

```
persp(normalCopula(dim=2, rho.nqd),dCopula)
```




```
#####nonparametric distance test for NQD######
jntllF.NQD <- rep(0,4);
for (i in 1:361) {
  jntllF.NQD[i] <- sum((NQD.cop.mat[,1] <= grid[i,1]) * (NQD.cop.mat[,2] <=
grid[i,2]))/sizeT;
}
#print(jntllF.NQD)

indllF.NQD <- matrix(rep(0,722),nrow=361,ncol=2);
for (i in 1:361) {
  for (j in 1:2) {
    indllF.NQD[i,j] <- sum((NQD.cop.mat[,j] <= grid[i,j]))/sizeT;
  }
}
#print(indllF.NQD)

Vmat.NQD <- matrix(rep(0,361*361),nrow=361,ncol=361);
for (k in 1:361) {
  NQD.bveck= matrix(c(1, - indllF.NQD[k,2], -indllF.NQD[k,1]),nrow=3);
  for (l in 1:361) {
    NQD.bvecl = matrix(c(1, - indllF.NQD[l,2], -indllF.NQD[l,1]),nrow=3);
   NQD.amat <- matrix(rep(0,3*3),nrow=3,ncol=3);
   NQD.amat[1,1] = sum((NQD.cop.mat[,1] <= min(grid[k,1],grid[l,1])) *
(NQD.cop.mat[,2] <= min(grid[k,2],grid[l,2])))/sizeT -
```

```
jntllF.NQD[k]*jntllF.NQD[l];
   NQD.amat[1,2] = sum((NQD.cop.mat[,1] <= min(grid[k,1],grid[l,1])) *
(NQD.cop.mat[,2] <= grid[k,2]))/sizeT - jntllF.NQD[k]*indllF.NQD[l,1];
   NQD.amat[1,3] = sum((NQD.cop.mat[,1] <= grid[k,1]) * (NQD.cop.mat[,2] <=
min(grid[k,2],grid[l,2])))/sizeT - jntllF.NQD[k]*indllF.NQD[l,2];
   NQD.amat[2,1] = sum((NQD.cop.mat[,1] <= min(grid[k,1],grid[l,1])) *
(NQD.cop.mat[,2] <= grid[l,2]))/sizeT - jntllF.NQD[l]*indllF.NQD[k,1];
   NQD.amat[2,2] = sum((NQD.cop.mat[,1] <= min(grid[k,1],grid[l,1])))/sizeT -
indllF.NQD[k,1]*indllF.NQD[l,1];
   NQD.amat[2,3] = sum((NQD.cop.mat[,1] <= min(grid[k,1],grid[l,1])) *
(NQD.cop.mat[,2] <= min(grid[k,2],grid[l,2])))/sizeT -
indllF.NQD[k,1]*indllF.NQD[l,2];
   NQD.amat[3,1] = sum((NQD.cop.mat[,1] <= grid[l,1]) * (NQD.cop.mat[,2] <=
min(grid[k,2],grid[l,2])))/sizeT - jntllF.NQD[l]*indllF.NQD[k,2];
   NQD.amat[3,2] = sum((NQD.cop.mat[,1] <= min(grid[k,1],grid[l,1])) *
(NQD.cop.mat[,2] <= min(grid[k,2],grid[l,2])))/sizeT -
indllF.NQD[l,1]*indllF.NQD[k,2];
   NQD.amat[3,3] = sum((NQD.cop.mat[,2] <= min(grid[k,2],grid[l,2])))/sizeT -
indllF.NQD[k,2]*indllF.NQD[l,2];
    Vmat.NQD[k,l] = t(NQD.bveck) %*%NQD.amat %*% NQD.bvecl;
  }
}
V.tblNQD<- data.frame(Vmat.NQD);
V.tblNQD <- round(V.tblNQD, digits = 4);
#print(xtable(V.tblNQD,digits = 4)) #Too big to print

Dvec.NQD = rep(0,4);
for (i in 1:361) {
  Dvec.NQD[i] = jntllF.NQD[i]- indllF.NQD [i,1]*indllF.NQD[i,2];
}
#print(Dvec.NQD)

intUniGNQD<- sqrt(sizeT)*Dvec.NQD/sqrt(diag2vec(Vmat.NQD));
1 - pnorm(min(intUniGNQD))

## [1] 1

#[1] 1=p-value. We cannot reject the null hypothesis of dependence.

#############KS test on NQD#################

intUniG.KS.nqd<- Dvec.NQD /sqrt(diag2vec(Vmat.NQD));
1 - pnorm(max(abs(intUniG.KS.nqd)))

## [1] 0.3676387

##[1] 0.363537. We cannot reject the null hypothesis of dependence.
###########Bootstrap for P-value on goodness-of-fit; build in function in
```

```r
R#####
set.seed(20175)
#Setting B (bootstrap) to be lower than 100=saving a lot of time.
GOF.realdata<-BiCopGofTest(u1, u2, family = 1, method = "kendall",B=50)
GOF.realdata
```

```
## $p.value.CvM
## [1] 0
##
## $p.value.KS
## [1] 0
##
## $statistic.CvM
## [1] 0.2693744
##
## $statistic.KS
## [1] 1.454775
```

```r
set.seed(0307)
GOF.sim<-BiCopGofTest(NQD.cop.mat[,1], NQD.cop.mat[,2], family = 1,
                         method = "kendall",B=50)
GOF.sim
```

```
## $p.value.CvM
## [1] 0.92
##
## $p.value.KS
## [1] 0.94
##
## $statistic.CvM
## [1] 0.03435594
##
## $statistic.KS
## [1] 0.4978293
```

Chapter 6

Bibliography

1. Berg, D. (2009). Copula goodness-of-fit testing: an overview and power comparison," European Journal of Finance 15 (7-8): 675-701.
2. Denuit, M., Scaillet, O. (2004). "Nonparametric tests for positive quadrant dependence," Journal of Financial Econometrics 2(3): 422–450.
3. Fernández, B.F, González-Barrios, J.M. (2004). "Multidimensional dependency measures," Journal of Multivariate Analysis 89 (2): 351-370
4. Genest, C., Favre, A.C. (2007). "Everything you always wanted to know about copula modeling but were afraid to ask," Journal of Hydrologic Engineering 12(4): 347-368.
5. Gijbels, I.,Omelka, M, Sznajder, D (2010). "Positive quadrant dependence tests for copulas," Canadian Journal of Statistics 38 (4): 555-581.
6. Gijbels, I., Sznajder, D. (2013). "Positive quadrant dependence testing and constrained copula estimation," Canadian Journal of Statistics 41(1): 36-64.
7. Janic-Wróblewska, A., Kallenberg, W.C.M., Ledwina, T. (2004). "Detecting positive quadrant dependence and positive function dependence," Insurance Mathematics and Economics 34(3): 467-487.
8. Joe, H. (1997). Multivariate models and dependence concepts, Chapman and Hall, London.
9. Kallenberg, W.C.M, Ledwina, T. (1999). "Data-Driven Rank Tests for Independence," Journal of the American Statistical Association 94(445): 285-301.
10. Kochar, S. C., Gupta, R. P. (1987). "Competitors of the Kendall-Tau Test for Testing Independence Against Positive Quadrant Dependence," Biometrika 74(3): 664-666.
11. Kodde, D.A., Palm, F,C. (1986). "Wald criteria for jointly testing equality and inequality restrictions," Econometrica, 54(5): 1243-1248.
12. Ledwina, T. (1986). "Large deviations and Bahadur Slope of some rank tests of independence," Indian Journal of Statistics (Sankhhya), 48(A,2): 188-207.
13. Ledwina, T., Wyłupek, G. (2014). "Validation of positive quadrant dependence," Insurance Mathematics and Economics 56: 38-47.
14. Lehmann, E.L. (1966). "Some concepts of dependence," Annals of Mathematical Statistics. 37: 1137-1153.
15. Nelson, R. B. (1999). An introduction to copulas, Springer-Verlag New York, Inc.
16. Pandit,V.P., Kumari, S. (2013). "A class of distribution-free tests for independence against positive quadrant dependence," Pakistan Journal of Statistics and Operation Research 9(4): 373-379.
17. Scaillet, O. (2005). "A Kolmogorov-Smirnov type test for positive quadrant dependence," Canadian Journal of Statistics 33(3): 415-427.

18. Schriever, B. F. (1987). “An Ordering for Positive Dependence,” The Annals of Statistics 15(3): 1208-1214.
19. Shetty, I., Pandit, P.(2003). “ Distribution-free tests for independence against positive quadrant dependence: A generalization,” Statistical Methods and Applications 12(1): 5-17
20. Yan, J. (2007). “Enjoy the joy of copulas: with a package copula,” Journal of Statistical Software 21(4): 1-21.