A comparison of two graphical methods for detecting dependence

Una comparación de dos métodos gráficos para detectar dependencia

Julieth V. Guarín-Escudero[a]*

Mario C. Jaramillo-Elorza[b]

Carlos M. Lopera-Gómez[c]

Abstract

Copulas have become a useful tool for modeling data when the dependence among random variables exists and the multivariate normality assumption is not fulfilled. The copulas have been applied in several fields. In finance, copulas are used in asset modeling and risk management. In biomedical studies, copulas are used to model correlated lifetimes and competitive risks [1]. In engineering, copulas are used in multivariate process control and hydrological modeling [2]. The interest in modeling multivariate problems involving dependent variables is generalized in several areas, making this methodology in a convenient way to model the dependence structure of random variables. However, in practice a first step before modeling phenomena through copulas is to assess whether there is dependence among the variables involved. In this paper some graphical methods to detect dependence are discussed and their performance will be evaluated through a simulation study. An application of graphical methods presented to insurance data is illustrated.

Keywords: Copula, graphics, dependence.

Resumen

Las cópulas se han convertido en una herramienta útil para modelar datos cuando existe una dependencia entre las variables aleatorias y el supuesto de normalidad no se cumple. Las cópulas se han aplicado en diversos campos, tales como finanzas, estudios biomédicos y en ingeniería. El interés en modelar problemas multivariados que involucran variables dependientes se generaliza en diversas áreas, haciendo de esta metodología una forma conveniente para modelar la estructura de dependencia entre las variables aleatorias. Sin embargo, en la práctica un primer paso antes de empezar a modelar fenómenos mediante cópulas es evaluar si existe dependencia entre las variables involucradas y en qué grado. En este artículo algunos métodos gráficos para detectar dependencia son discutidos y el desempeño de los mismos se evaluará a través de un estudio de simulación. Se ilustran los métodos gráficos presentados mediante una aplicación a datos de seguros.

Palabras clave: Cópula, gráficos, dependencia.

a Estudiante de Maestría, Escuela de Estadística, Universidad Nacional de Colombia, Sede Medellín.
*Correo electrónico: jvguarine@unal.edu.co
b Profesor asociado, Escuela de Estadística, Universidad Nacional de Colombia, Sede Medellín.
c Profesor asociado, Escuela de Estadística, Universidad Nacional de Colombia, Sede Medellín.

1. INTRODUCTION

In probability theory the functions called copulas can to represent distribution functions and they are a convenient way to model the dependence structure of random variables [1]. This concept allows building models beyond the standards in the analysis of dependence among variables, further, allows to capture non-linear dependence relationships and only need to specify the copula and marginal function associated with the random variables involved [3].

Before starting to fit models to a set of random variables, an analysis of the type and degree of dependence among them should be realized. In statistics, descriptive and graphical analysis plays an important role because it is the basis for realize and propose more complex models.

To study the dependence among variables some graphical methods like the $\mathcal{X}$-plot and the K-plot (Kendall plot) have been developed. The former was initially proposed by [4] and the latter was proposed by [5]. Some applications of these methods can be found in [6] where the relationship between oil price variation and stock indices is measured, in [7] where the relationship between storm characteristics are analyzed, and in [8] where the dependence between the infiltration index and the maximum rainfall intensity in an hydrological application.

In this paper both graphical methods are analyzed and compared through a simulation study with the traditional scatter plot. In particular we study the effect of some factors that can affect the performance of the dependence graphs.

This paper is organized as follows. Section 2 in- troduces the concept of copulas and the parameter of dependence used in this work, the Kendall's t, its most relevant properties and its form according to the copula used. Section 3 presents the definitions of the graphical methods to detect dependence explored in this paper, $\mathcal{X}$-plot and K-plot. In section 4, a simulation studio to assess the behavior of methods to detect dependence in comparison with the scatter plot is performed. Section 5 presents an application of both methods to real data. Finally, section 6 concludes this paper.

2. COPULAS

Suppose that C_α is a distribution function with density c_α over $[0, 1]^2$ for $\alpha \in \mathbb{R}$.. Denote (T_1, T_2) the failure times, and denote (S_1, S_2), (f_1, f_2) the corresponding marginal survival and density functions. If (T_1, T_2) comes from a copula C_α, for any $\alpha \in \mathbb{R}$., the joint survival and density functions of (T_1, T_2) are given by

$$S(t_1, t_2) = C_\alpha (S_1(t_1), S_2(t_2)) \quad t_1, t_2 \geq 0, \qquad (1)$$

$$f(t_1, t_2) = c_\alpha (S_1(t_1), S_2(t_2)) f_1(t_1) f_2(t_2) \ t_1, t_2 \geq 0, \qquad (2)$$

where α represents the dependency parameter bet- ween T_1 and T_2.

We introduce the Archimedean family of copulas, because is the most used copula family. A bivariate distribution belonging to the family of Archimedean copula models has the representation(2)

$$C_\alpha (u, v) = \phi_\alpha^{-1} [\phi_\alpha (u) + \phi_\alpha (v)], \quad 0 \leq u, v \leq 1, \qquad (3)$$

where ϕ is a convex and decreasing function such that $\phi \geq 0$, $\phi(1) = 0.$, $f(1) = 0$. The ϕ function is named *generator* of the C_α copula and the inverse of the generator, ϕ^{-1}, is the *Laplace transform* of a latent variable denoted as γ , which induces the dependency α. Thus, the selection of a generator results in several families of copulas. In table **1**, we show the forms for bivariate survival functions in three Archimedean copula families. Additionally, in table **2**, we show the generators and the Laplace transform for the considered families.

Family Copula	Parameter Space	Bivariate Copula $C_\alpha(u,v)$
Clayton	$\alpha > 1$	$\{u^{1-\alpha} + v^{1-\alpha} - 1\}^{1/(1-\alpha)}$
Gumbel	$0 < \alpha < 1$	$\exp\left\{-\left[(-\ln u)^{1/\alpha} + (-\ln v)^{1/\alpha}\right]^\alpha\right\}$
Frank	$\alpha > 0$	$\log_\alpha \{1 + (\alpha^u - 1)(\alpha^v - 1)/(\alpha - 1)\}$
Joe	$\alpha > 1$	$1 - [(1-u)^\alpha + (1-v)^\alpha - (1-u)^\alpha (1-v)^\alpha]^{1/\alpha}$

Table 1. Common Archimedean copulas.

Family Copula	Generator $\phi(t)$	Laplace Transform $(\tau(s)=\phi^{-1}(s))$
Clayton	$t^{1-\alpha}-1$	$(1+s)^{1/(1-\alpha)}$
Gumbel	$(-\ln t)^{1/\alpha}$	$\exp(-s^\alpha)$
Frank	$\ln\frac{\alpha^t-1}{\alpha-1}$	$\log_\alpha\{1-(1-\alpha)e^s\}$
Joe	$-\ln(1-(1-t)^\alpha)$	No closed form

Table 2. Generators and their Laplace Transforms.

In this work several copulas of the archimedian class are used. This class groups a large number of copula families with simple analytical properties [9]. Archimedian copulas also can describe a great diversity of dependency structures [10]. In addition, Gaussian copula is included as an alternative frequently used in literature. The Gaussian copula is a one-parameter family for pairs of random variables $(u; v)$. It takes the form [11]:

$$C_\rho(u,v) = \Phi_2\left(\Phi^{-1}(u),\Phi^{-1}(v)\right), \qquad (4)$$

where ρ is the correlation coefficient , $\rho = corr(u,v)$, Φ_2 is the bivariate normal distribution function and Φ is the univariate normal distribution function.

2.1. Kendall's τ

The Kendall's τ is perhaps the best alternative to use instead of linear correlation coefficient as a measure of dependence for variables that do not belong to the elliptical family [12].

Let (X_1,Y_1) and (X_2,Y_2) be a bivariate random sample of a joint and continuous distribution function $H(X,Y)$. Then Kendall's τ is defined as the probability of concordance less the probability of discordance [3]:

$$\tau = P[(X_1-X_2)(Y_1-Y_2)>0] - P[(X_1-X_2)(Y_1-Y_2)<0] \qquad (5)$$

Theorem 2.1. *[13] Property of invariance of Kendall's τ. Let (X_1,Y_1) y (X_2,Y_2) be a bivariate random sample of a joint and continuous distribution function, H(X,Y), let g and h two increasing functions, then $\tau[g(X),h(Y)] = \tau(X,Y)$. In [13] can be seen the proof of this theorem.*

As Kendall's τ is invariant to strictly increasing transformations, the following theorem provides an expression of this parameter in terms of copulas.

Theorem 2.2. *[14] Let X, Y continuous random variables whose copula is C. Then Kendall's τ for X and Y , τ (X,Y) is given by:*

$$\tau = 4\int_0^1\int_0^1 C(u,v)dC(u,v) - 1. \qquad (6)$$

3. GRAPHICAL METHODS FOR DETECTING DEPENDENCE

In this section both graphical methods that will be seen throughout this work are defined.

3.1. $\mathcal{X}$-plot

The $\mathcal{X}$-plot was originally proposed by [4]. Its construction is based on the $\mathcal{X}$-square statistic for independence.

Let $(X_1,Y_1),...,(X_n,Y_n)$ be a bivariate random sample of a joint and continuous distribution function, $H(X,Y)$, and let $I(A)$ be the indicator function of the event A. For each observation (x_i, y_i) the following procedure is performed: [15]

$$H_i = \frac{1}{n-1}\sum_{j\neq i}I(X_j\leq X_i, Y_j\leq Y_i) \qquad (7)$$

$$F_i = \frac{1}{n-1}\sum_{j\neq i}I(X_j\leq X_i)$$

$$G_i = \frac{1}{n-1}\sum_{j\neq i}I(Y_j\leq Y_i)$$

None of these quantities exclusively depend of the observations ranks. [4] proposed to plot the pairs (λ_i,χ_i), where:

$$\chi_i = \frac{H_i-F_iG_i}{\sqrt{F_i(1-F_i)G_i(1-G_i)}}, \qquad (8)$$

$$\lambda_i = 4\,\mathrm{sign}\left(\tilde{F}_i\tilde{G}_i\right)\mathrm{máx}\left(\tilde{F}_i^2,\tilde{G}_i^2\right),$$

and $\tilde{F}_i = F_i - 1/2$, $\tilde{G}_i = G_i - 1/2$ for $i\in\{1,\dots,n\}$.

λ_i is a measure of distance from the observation (X_i,Y_i) to data center [15].

All values of λ_i must be in the interval $[-1;1]$ [14]. The χ-plot is a scatter-plot of the pairs (λ_i, χ_i), $i = 1,...,n$. If the data constitute a bivariate sample with independent continuous marginals, the values of λ_i will be evenly distributed. However, if X and Y are associated, the values of λ_i will tend to form groups, in particular, positive values of λ_i indicate that X_i and Y_i are relatively larger or smaller (at the same time) than the median, while negative values of λ_i correspond to X_i and Y_i located on opposite sides with respect to their median [15].

The horizontal lines on the graph are given by $\chi = -c_p/n^{1/2}$ and $\chi = c_p/n^{1/2}$, where c_p is selected so that approximately $100p$ % of the pairs (λ_i, χ_i) are between the two horizontal lines. For $p = 0.90, 0.95, 0.99$ the values of c_p are 1.54, 1.78 y 2.18, respectively [4]. Using the Monte Carlo method you can calculate other c_p values. It is recommended to draw only those pairs such that $|\lambda_i| < 4\left(\frac{1}{n-1} - \frac{1}{2}\right)^2$, in order to avoid misleading observations [14].

3.2 *K*-plot

The *K*-plot (Kendall-plot) was created by [5]. This tool is based on the ranks of observations using the integral transformation of multivariate probabilities, producing a similar graph to conventional Q-Q plot [15].

Let $(X_1,Y_1),..., (X_n,Y_n)$ be a random sample of a joint and continuous distribution function, $H(X,Y)$. To build the *K*-plot we proceed as follows:

1. For each $1 \leq i \leq n$ compute H_i (as in the χ - plot).

2. Sort the H_i values such that $H_{(1)} \leq \cdots \leq H_{(n)}$.

3. Plot the pairs $(W_{i:n}, H_{(i)})$, where $W_{i:n}$ is the expectation of the ith order statistic in a sample of size n, which is calculated as follows:

$$W_{i:n} = n\binom{n-1}{i-1} \int_0^1 w\left[K_0(w)\right]^{i-1} \left[1 - K_0(w)\right]^{n-i} dK_0(w)$$
(9)

with

$$K_0(w) = w - w\log(w) \quad 0 \leq w \leq 1. \quad (10)$$

When the scatter plot of $H_{(i)}$ against $W_{i:n}$ moves away from the diagonal, then there is an indication of a functional dependence between the two variables involved.

4. SIMULATION STUDY

In this section we present a simulation study to evaluate the development of the proposed graphical methods. In particular, we want to study the effect of some factors that can affect the performance of the dependence graphs such as: dependence level, sample size and the chosen copula to construct the bivariate function. In addition, we show the implementation of χ-plot and *K*-plot through the package CDVine of R [16].

The scope of the study is intended to cover several scenarios, where the scatter plot is compared with the χ-plot and *K*-plot, for which the sample size in 20, 50, 100 and 200 is varied, and the dependence parameter values (τ Kendall) of 0.3, 0.5 and 0.8 were considered. In addition, for the data generation Clayton, Frank, Gaussian, Gumbel and Joe copulas were used.

In total 60 simulation scenarios were obtained, which are summarized in the following table:

Table 3. Simulation Scenarios

Copula	τ	n	Copula	τ	n
Clayton	0.3	20	Clayton	0.3	50
Clayton	0.5	20	Clayton	0.5	50
Clayton	0.8	20	Clayton	0.8	50
Frank	0.3	20	Frank	0.3	50
Frank	0.5	20	Frank	0.5	50
Frank	0.8	20	Frank	0.8	50
Gaussian	0.3	20	Gaussian	0.3	50
Gaussian	0.5	20	Gaussian	0.5	50
Gaussian	0.8	20	Gaussian	0.8	50
Gumbel	0.3	20	Gumbel	0.3	50
Gumbel	0.5	20	Gumbel	0.5	50
Gumbel	0.8	20	Gumbel	0.8	50
Joe	0.3	20	Joe	0.3	50
Joe	0.5	20	Joe	0.5	50
Joe	0.8	20	Joe	0.8	50
Clayton	0.3	100	Clayton	0.3	200

Copula	τ	n	Copula	τ	n
Clayton	0.5	100	Clayton	0.5	200
Clayton	0.8	100	Clayton	0.8	200
Frank	0.3	100	Frank	0.3	200
Frank	0.5	100	Frank	0.5	200
Frank	0.8	100	Frank	0.8	200
Gaussian	0.3	100	Gaussian	0.3	200
Gaussian	0.5	100	Gaussian	0.5	200
Gaussian	0.8	100	Gaussian	0.8	200
Gumbel	0.3	100	Gumbel	0.3	200
Gumbel	0.5	100	Gumbel	0.5	200
Gumbel	0.8	100	Gumbel	0.8	200
Joe	0.3	100	Joe	0.3	200
Joe	0.5	100	Joe	0.5	200
Joe	0.8	100	Joe	0.8	200

4.1 Analysis of Results

In the following section the results of the simulation study are presented. The main objective is to evaluate the performance of graphics to detect dependence under the scenarios described in the previous section.

4.1.1 Sample size n = 20

Figures **1** to **5** show the graphics performance when the sample size is $n = 20$ and varying the parameter dependence τ, under the considered copula families.

In figures **1** to **5** with $n = 20$, the behavior of graphics to detect dependence is similar in all simulated copulas. When $\tau = 0.3$, the $\mathcal{X}$-plot and *K*-plot provide similar results to the traditionally used graph: the scatter plot. In this case the three graphics fail to detect dependencies between variables. When the dependence parameter τ increases to values of 0.5 and 0.8, again the three graphs behave similarly, all fail to detect such dependence between variables for all simulated copulas. In the case of the $\mathcal{X}$-plot with $\tau = 0.5$ and $\tau = 0.8$, most points fall outside the bands in all simulated copulas, indicating a clear dependence between variables. In the case of the graph *K*-plot, for $\tau = 0.5$ and $\tau = 0.8$ the points consistently fall away from the diagonal, which indicates dependence.
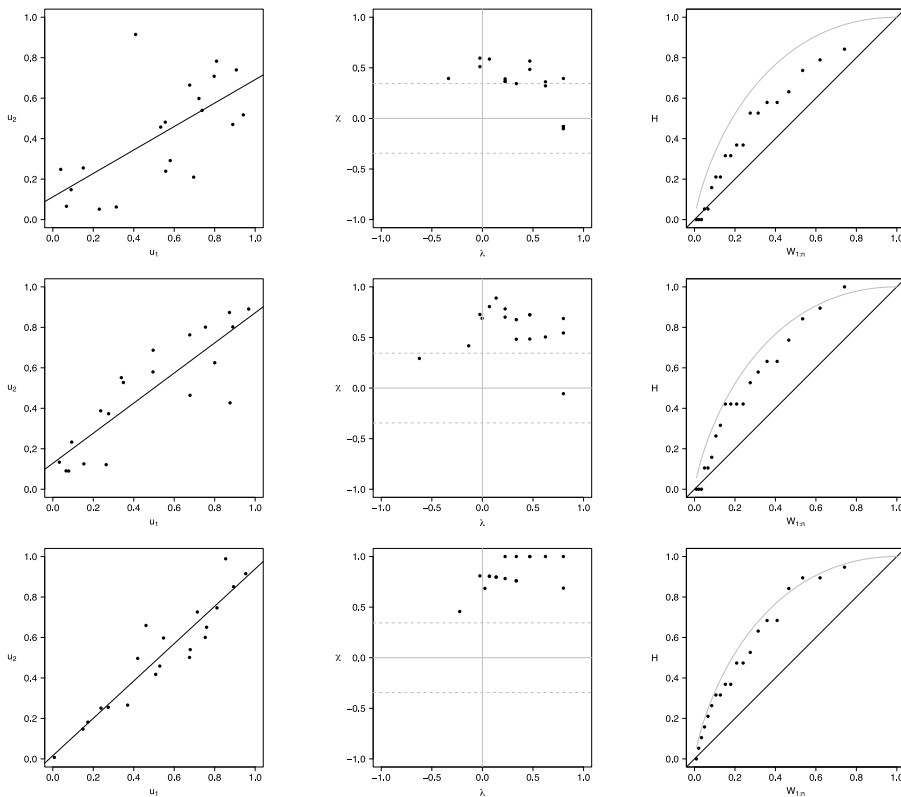
Figure 1. Scatter-plot (left), $\mathcal{X}$-plot (center) and K-plot (right) for n = 20 using the Clayton Copula with $\tau = 0.3$ (top), $\tau = 0.5$ (medium) and $\tau = 0.8$ (bottom)
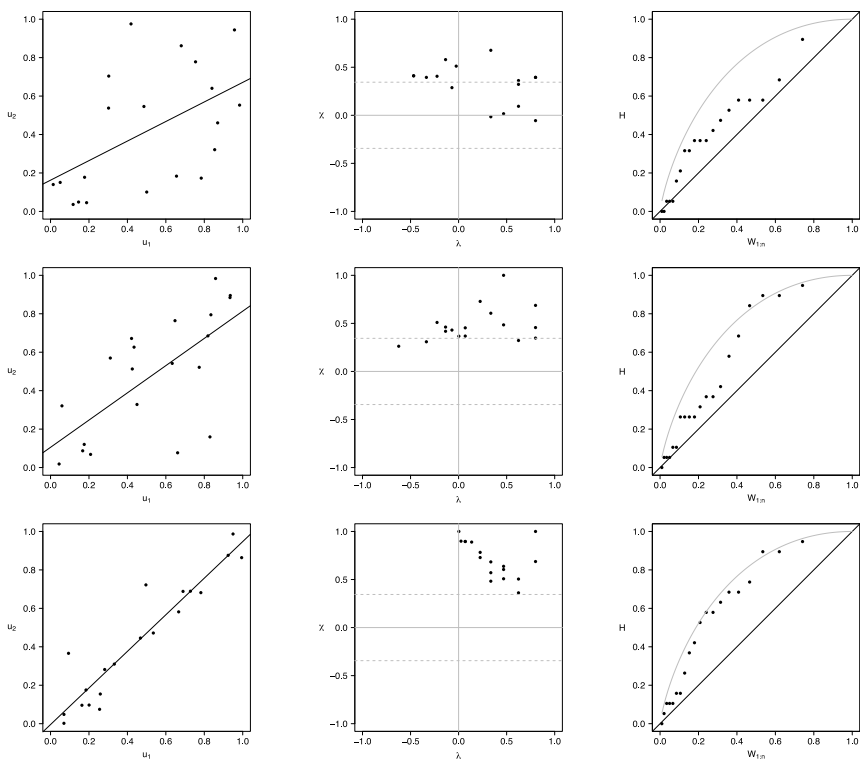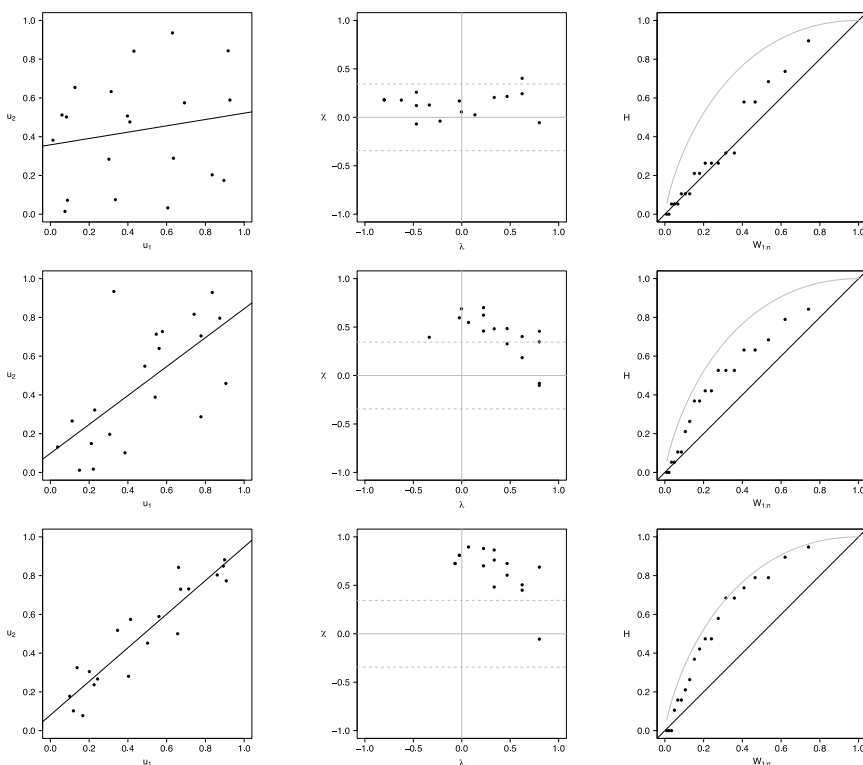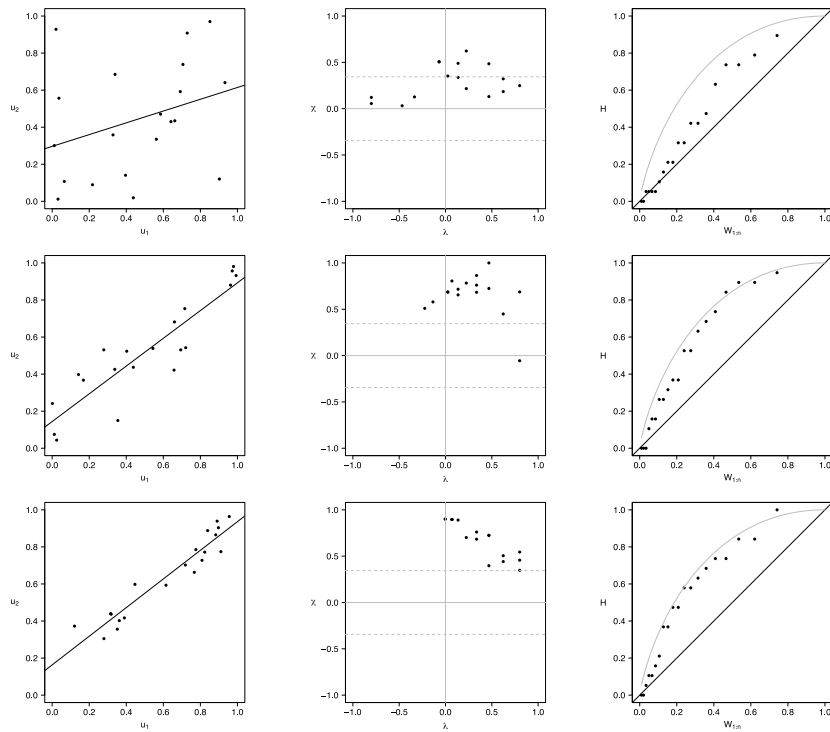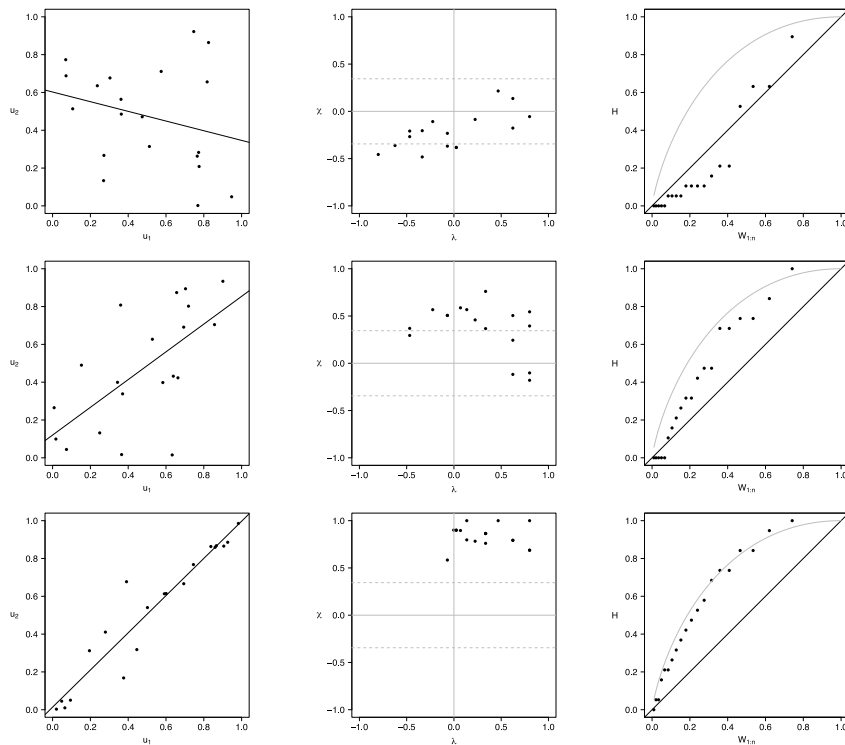
Figure 2. *S*catter-plot (left), χ-plot (center) and K-plot (right) for n = 20 using the Frank Copula
with τ = 0.3 (top), τ = 0.5 (medium) and τ = 0.8 (bottom)

Figure 3. Scatter-plot (left), χ-plot (center) and K-plot (right) for n = 20 using the Gaussian
Copula with τ =0.3 (top), τ = 0.5 (medium) and τ = 0.8 (bottom)

Figure 4. Scatter-plot (left), 𝒳-plot (center) and K-plot (right) for n = 20 using the Gumbel Copula with $\tau = 0.3$ (top), $\tau = 0.5$ (medium) and $\tau = 0.8$ (bottom)

Figure 5. Scatter-plot (left), 𝒳-plot (center) and K-plot (right) for n = 20 using the Joe Copula with $\tau = 0.3$ (top), $\tau = 0.5$ (medium) and $\tau = 0.8$ (bottom)

4.1.2 Sample size n = 50

Figures **6** to **10** show the graphics performance when the sample size is *n* = 50 and varying the parameter dependence τ, under the considered copula families.

In figures **6** to **10** with *n* = 50, $\mathcal{X}$-plot and *K*-plot provide slightly different results when $\tau = 0.3$ compared to the previous case with *n* = 20. In this case, with the Clayton, Frank, Gumbel and Joe copulas in $\mathcal{X}$-plot about half of the points are outside the bands and around the half of the data is within the bands which indicate a low dependence between the random variables. The *K*-plot for Clayton and Gaussian copulas does not detect dependence between the variables because the points are very close to the diagonal. In the case

of the Frank, Gumbel and Joe copulas is observed that at the beginning, the points are near the diagonal but the rest of points are consistently going away which it would be a sign of low dependence between the variables. With $\tau = 0.3$ scatter plot does not detect dependence between the variables in any of the cases. When the parameter of dependence τ increases to values of 0.5 and 0.8 the three graphs behave similarly, all fail to detect such dependence between variables for all simulated copulas. Notice that in the $\mathcal{X}$-plot with $\tau = 0.5$ and $\tau = 0.8$ most points fall outside the bands in all simulated copulas, which indicates a clear dependence between the variables, while in the *K*-plot with $\tau = 0.5$ and $\tau = 0.8$ the points consistently fall away from the diagonal, which indicates dependence.
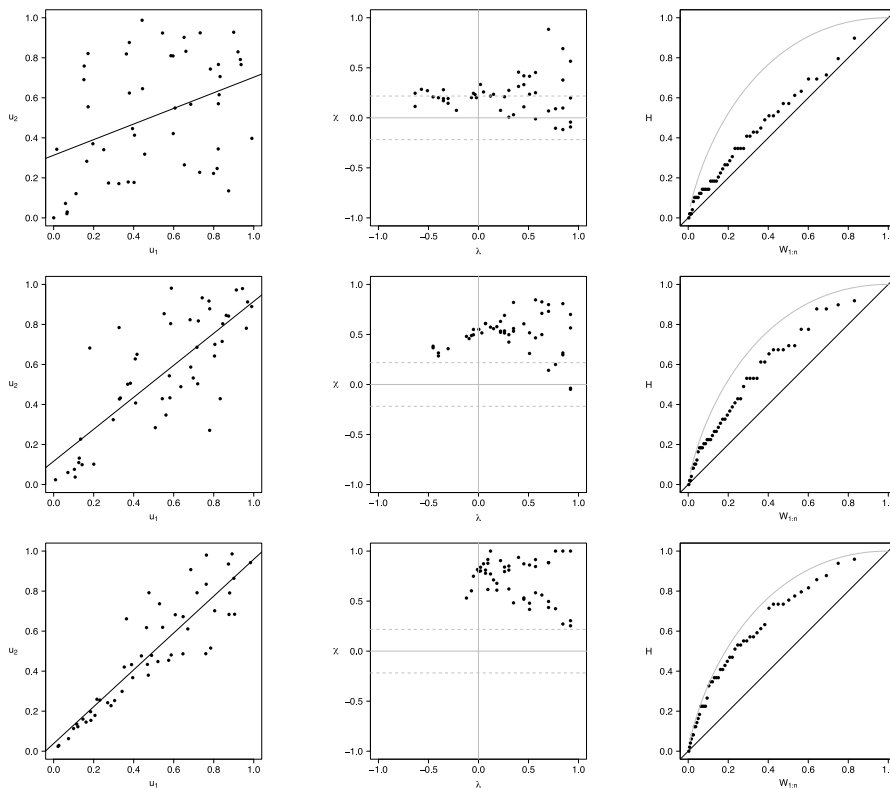

Figure 6. Scatter-plot (left), $\mathcal{X}$-plot (center) and K-plot (right) for n = 50 using the Clayton Copula with $\tau = 0.3$ (top), $\tau = 0.5$ (medium) and $\tau = 0.8$ (bottom)

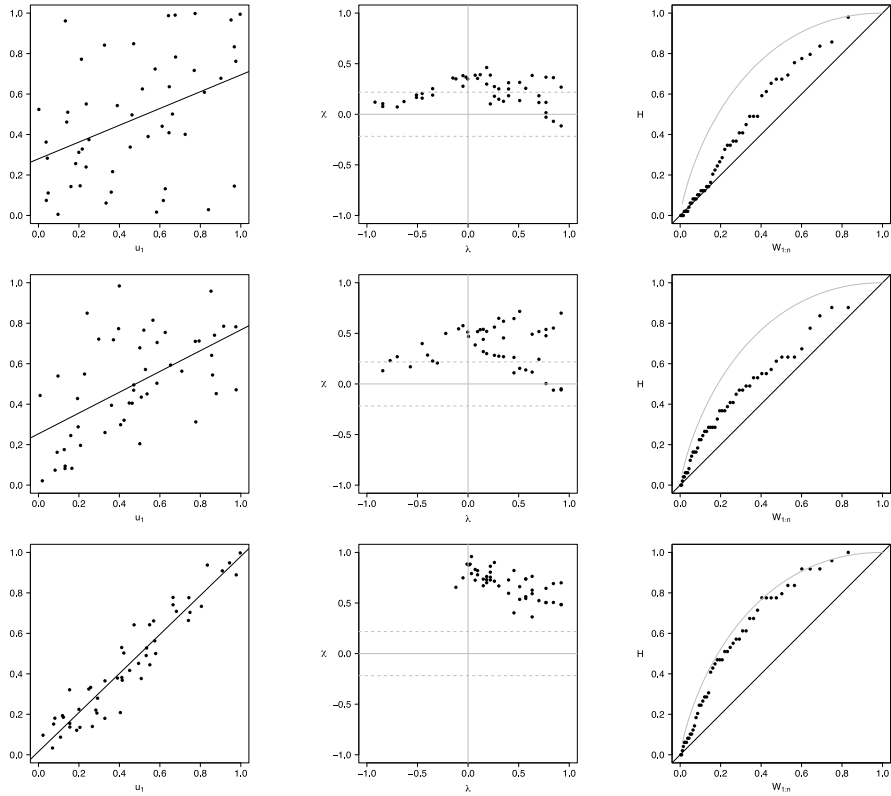**Figure 7.** Scatter-plot (left), χ-plot (center) and K-plot (right) for n = 50 using the Frank Copula with τ = 0.3 (top), τ= 0.5 (medium) and τ = 0.8 (bottom)
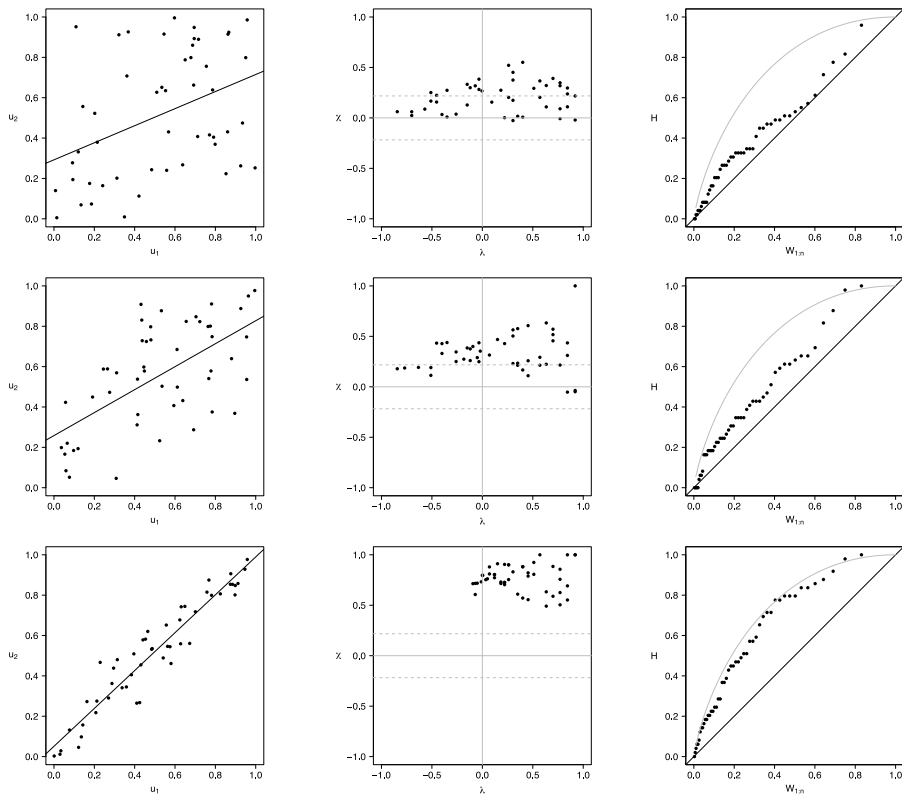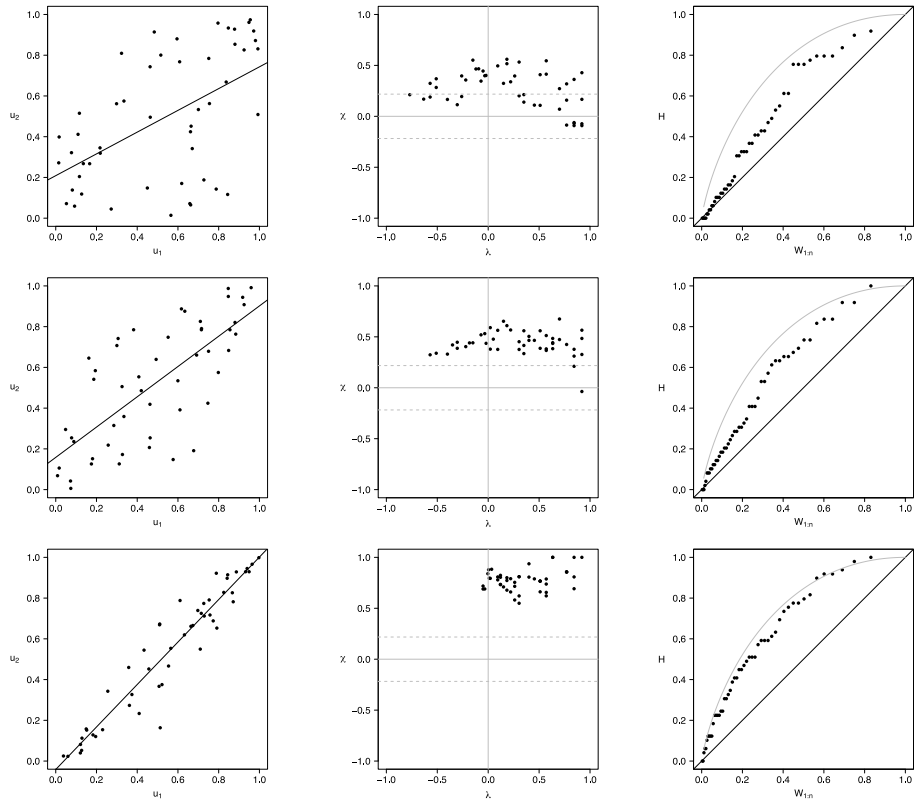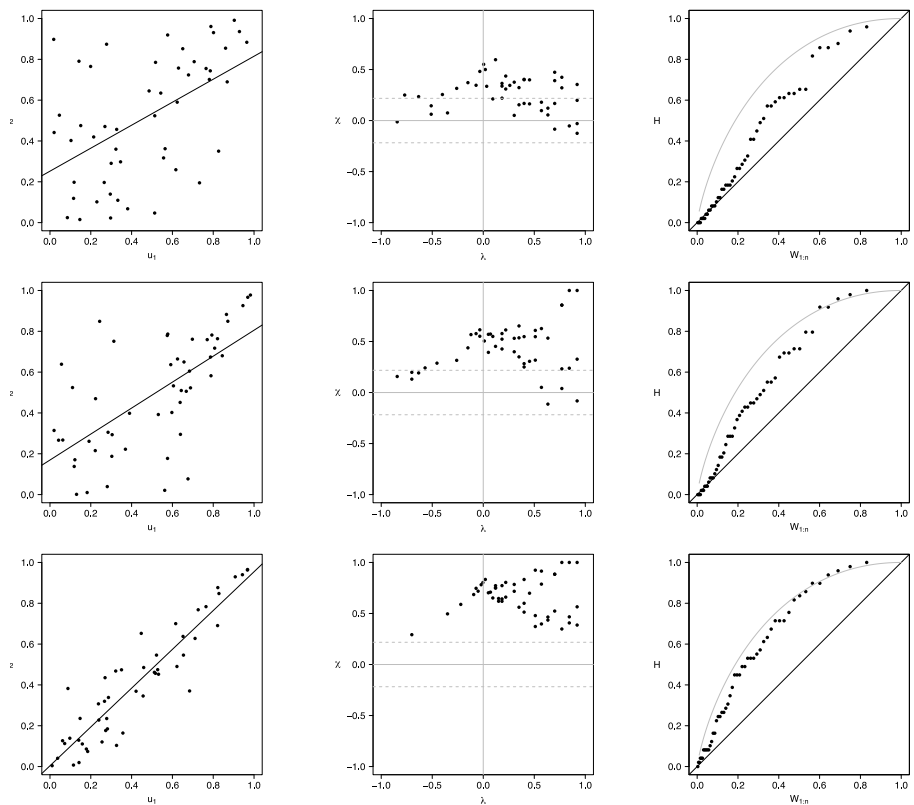

Figure 8. Scatter-plot (left), χ-plot (center) and K-plot (right) for n = 50 using the Gaussian Copula with τ= 0.3 (top), τ= 0.5 (medium) and τ = 0.8 (bottom)

Figure 9. Scatter-plot (left), 𝒳-plot (center) and *K*-plot (right) for *n* = 50 using the Gumbel Copula with 𝜏= 0.3 (top), 𝜏 = 0.5 (medium) and 𝜏= 0.8 (bottom)

Figure 10. Scatter-plot (left), 𝒳-plot (center) and *K*-plot (right) for *n* = 50 using the Joe Copula with 𝜏 = 0.3 (top), 𝜏 = 0.5 (medium) and 𝜏 = 0.8 (bottom)

4.1.3 Sample size n = 100

Figures **11** to **15** show the graphics performance when the sample size is $n = 100$ and varying the parameter dependence τ, under the considered copula families.

The case with $n = 100$ is presented in figures **11** to **15**. In the χ-plot with $\tau = 0.3$, for all simulated copulas, about half of the points remain outside the bands and about half of the data is within the bands, which indicates a low dependence between the random variables. The *K*-plot for Clayton copula does not detect dependence between the variables because the points are very close to the diagonal. In the Frank, Gumbel, Gaussian and

Joe copulas is observed that at the beginning, the points are near the diagonal but the rest of points are consistently going away which it would be a sign of a low dependence between the variables. With $\tau = 0.3$ the scatter plot does not detect dependence between the variables in any of the cases. When the parameter of dependence τ increases to values of 0.5 and 0.8 the three graphs behave similarly and all of them fail to detect such dependence between the variables for all simulated copulas. In the case of the χ-plot with $\tau = 0.5$ and $\tau = 0.8$ most points fall outside the bands in all simulated copulas, which indicates a clear dependence between the variables. In the case of *K*-plot with $\tau = 0.5$ and $\tau = 0.8$ the points consistently fall away from the diagonal, which indicates dependence.
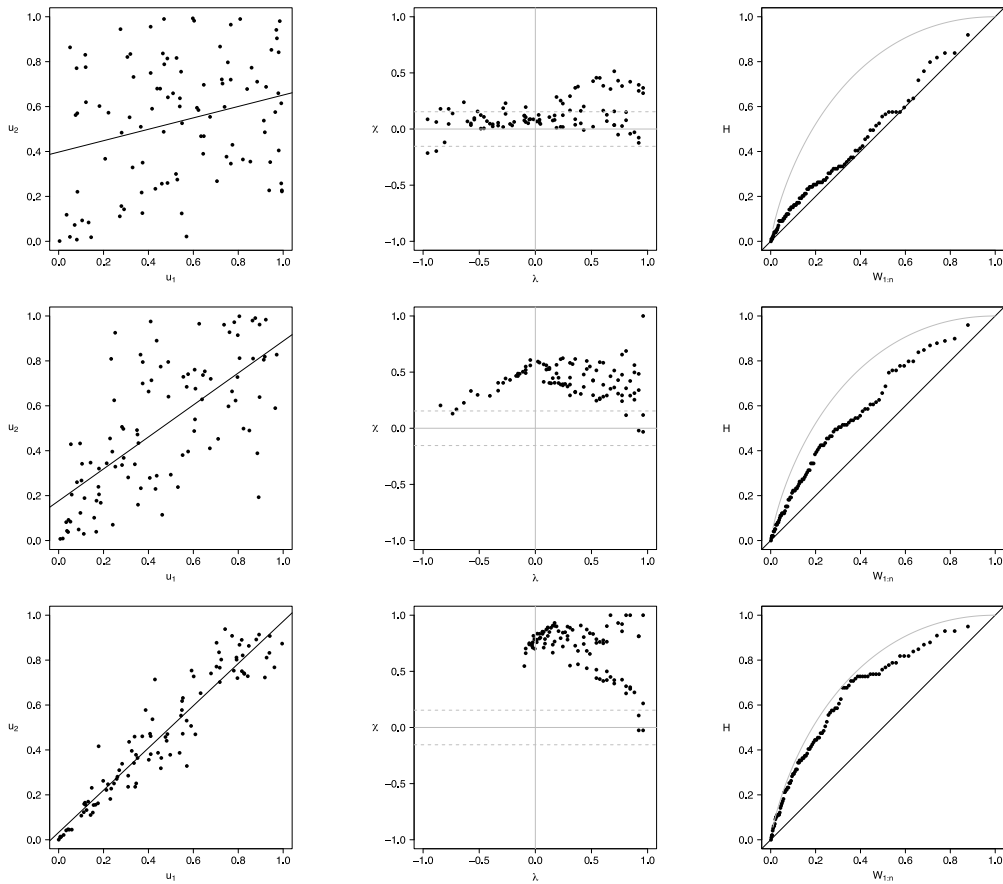

Figure 11. Scatter-plot (left), χ-plot (center) and *K*- plot (right) for $n = 100$ using the Clayton Copula with $\tau = 0.3$ (top), $\tau = 0.5$ (medium) and $\tau = 0.8$ (bottom)
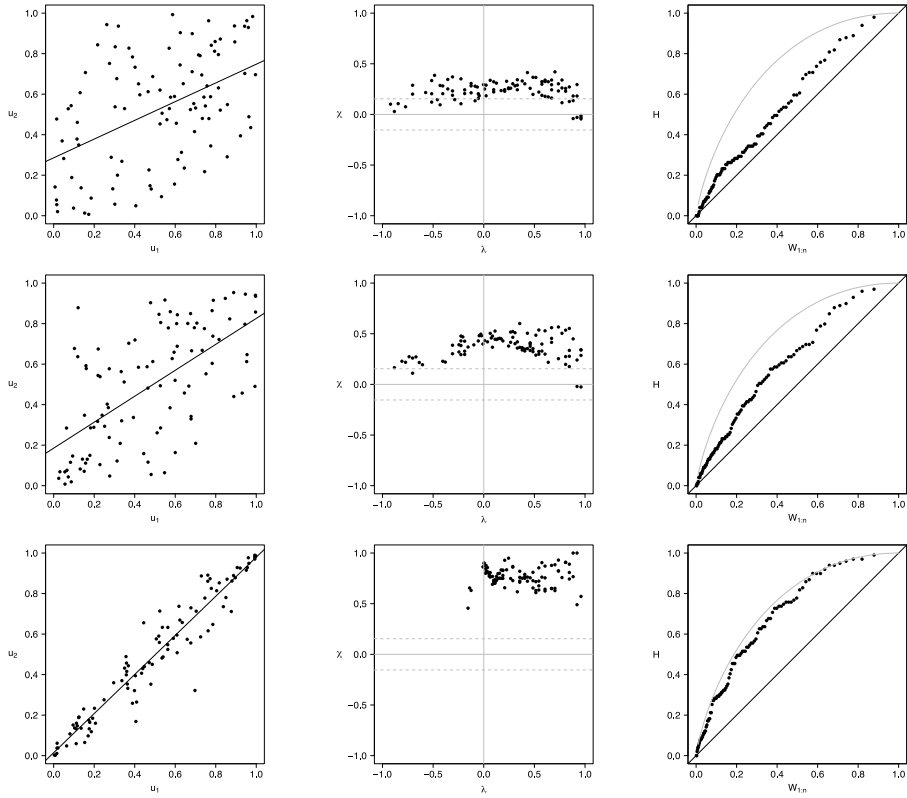
Figure 12. Scatter-plot (left), χ-plot (center) and *K*-plot (right) for $n = 100$ using the Frank Copula with $\tau = 0.3$ (top), $\tau = 0.5$ (medium) and $\tau = 0.8$ (bottom)
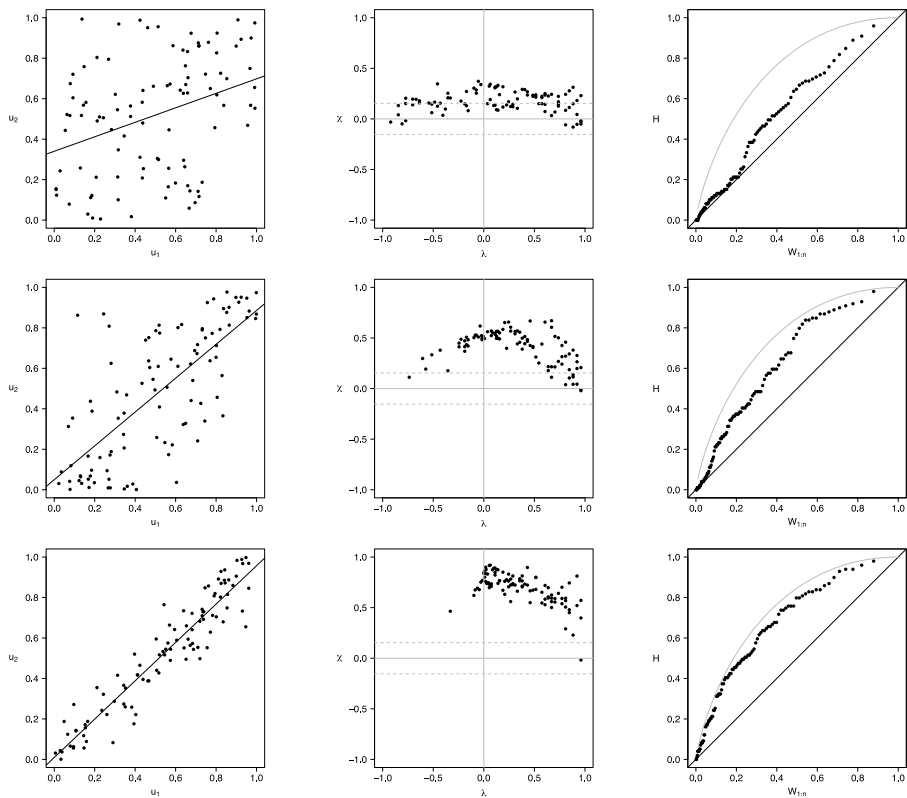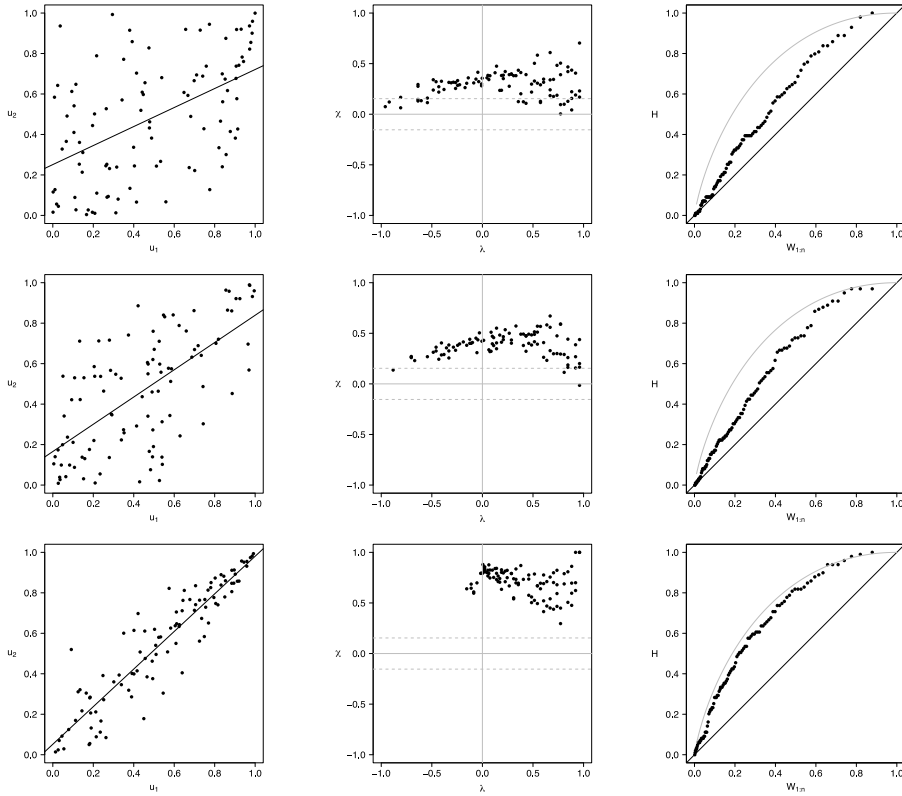
Figure 13. Scatter-plot (left), χ-plot (center) and *K*- plot (right) for $n = 100$ using the Gaussian Copula with $\tau = 0.3$ (top), $\tau = 0.5$ (medium) and $\tau = 0.8$ (bottom)
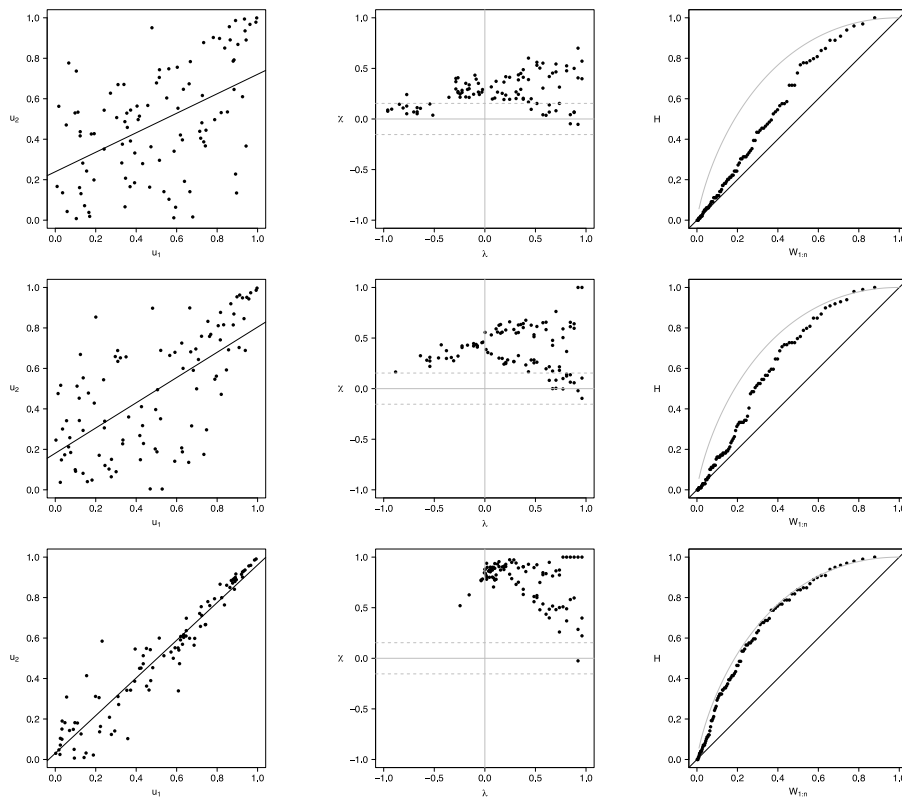
Figure 14. Scatter-plot (left), χ-plot (center) and K- plot (right) for $n = 100$ using the Gumbel Copula with $\tau = 0.3$ (top), $\tau = 0.5$ (medium) and $\tau = 0.8$ (bottom)

Figure 15. Scatter-plot (left), χ-plot (center) and K-plot (right) for $n = 100$ using the Joe Copula with $\tau = 0.3$ (top), $\tau = 0.5$ (medium) and $\tau = 0.8$ (bottom)

4.1.4. Sample size n = 200

Figures **16** to **20** show the graphics performance when the sample size is $n = 100$ and varying the parameter dependence τ, under the considered copula families.

Figures **16** to **20** show the case of $n = 200$, in Joe copula the $\mathcal{X}$-plot with $\tau = 0.3$ about half of the points are outside bands and about half of the data is within the bands, which indicates a low dependence between the random variables. With the Clayton, Frank, Gumbel and Gaussian copulas $\mathcal{X}$-plot detects dependence between the variables since most points fall outside the bands. The K-plot with all simulated copulas is observed that at the beginning, the points are near the diagonal

but the rest of points are consistently going away, which it would be a sign of a low dependence between the variables. With $\tau = 0.3$ scatter plot does not detect dependence between the variables in any of the cases, which makes the $\mathcal{X}$-plot and K-plot good alternatives for detecting dependence when n is large even when the dependence is low. When the parameter of dependence τ increases to values of 0.5 and 0.8 the three graphs behave similarly and all fail to detect such dependence between the variables for all simulated copulas. In the case of $\mathcal{X}$-plot with $\tau = 0.5$ and $\tau = 0.8$ most points fall outside the bands in all simulated copulas, which in- dicates a clear dependence between the variables. In the case of K-plot for $\tau = 0.5$ and $\tau = 0.8$ points are consistently away from the diagonal, which indicates dependence.
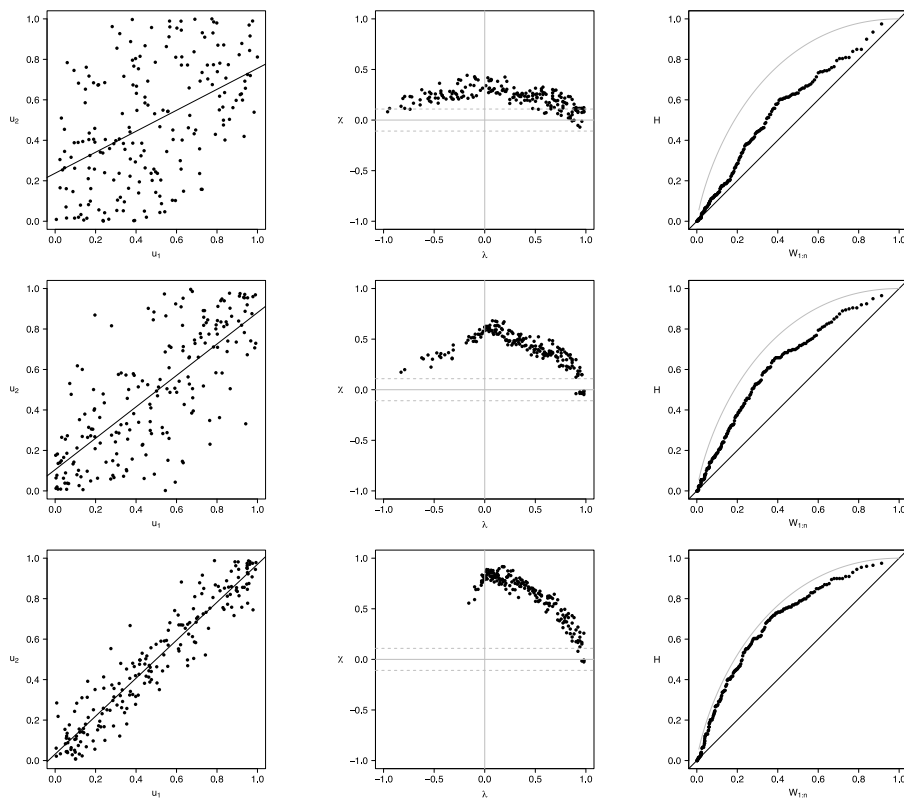

Figure 16. Scatter-plot (left), $\mathcal{X}$-plot (center) and K- plot (right) for $n = 200$ using the Clayton Copula with $\tau = 0.3$ (top), $\tau = 0.5$ (medium) and $\tau = 0.8$ (bottom)
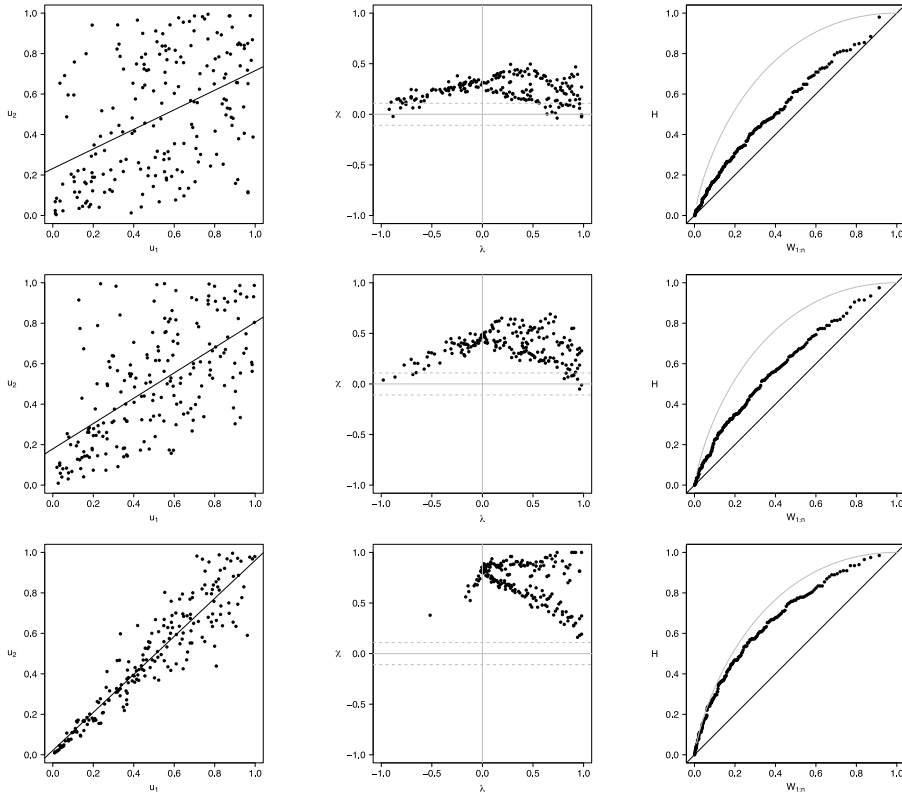
Figure 17. Scatter-plot (left), χ-plot (center) and K-plot (right) for $n = 200$ using the Frank Copula with $\tau = 0.3$ (top), $\tau = 0.5$ (medium) and $\tau = 0.8$ (bottom)
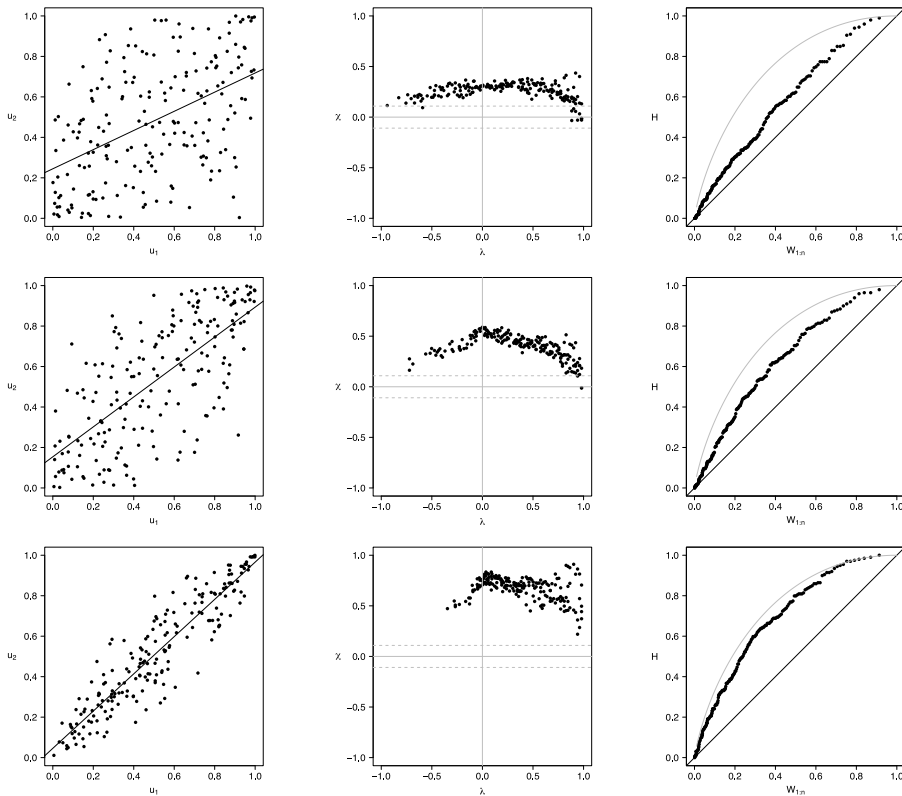

Figure 18. Scatter-plot (left), χ-plot (center) and K- plot (right) for $n = 200$ using the Gaussian Copula with $\tau = 0.3$ (top), $\tau = 0.5$ (medium) and $\tau = 0.8$ (bottom)
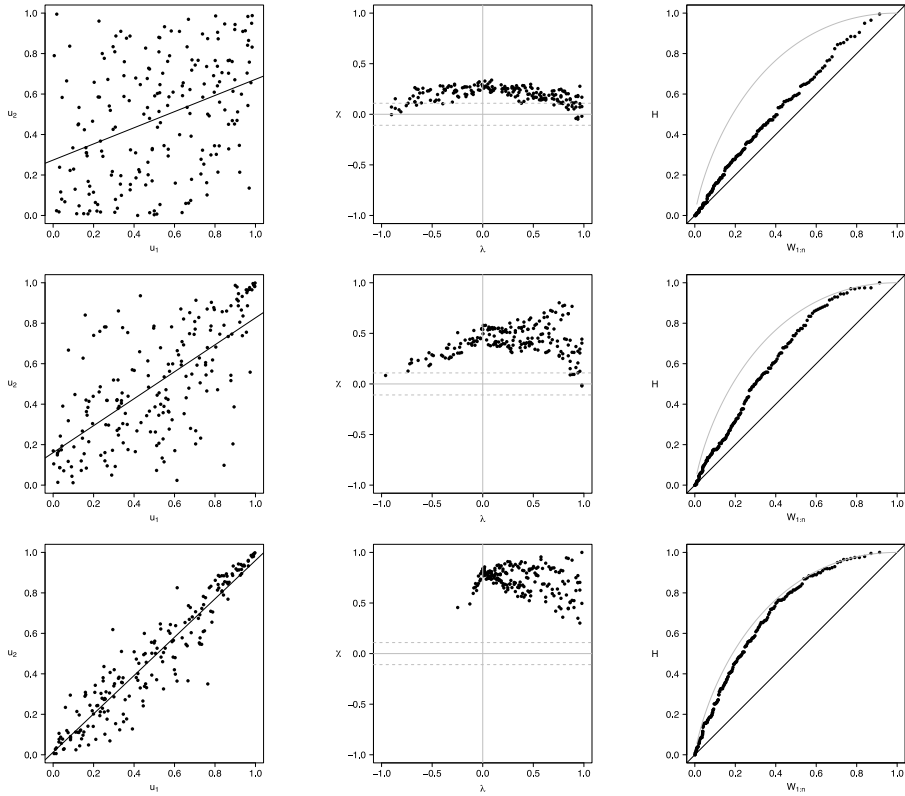
Figure 19. Scatter-plot (left), χ-plot (center) and *K*- plot (right) for *n* = 200 using the Gumbel Copula with τ = 0.3 (top), τ = 0.5 (medium) and τ = 0.8 (bottom)

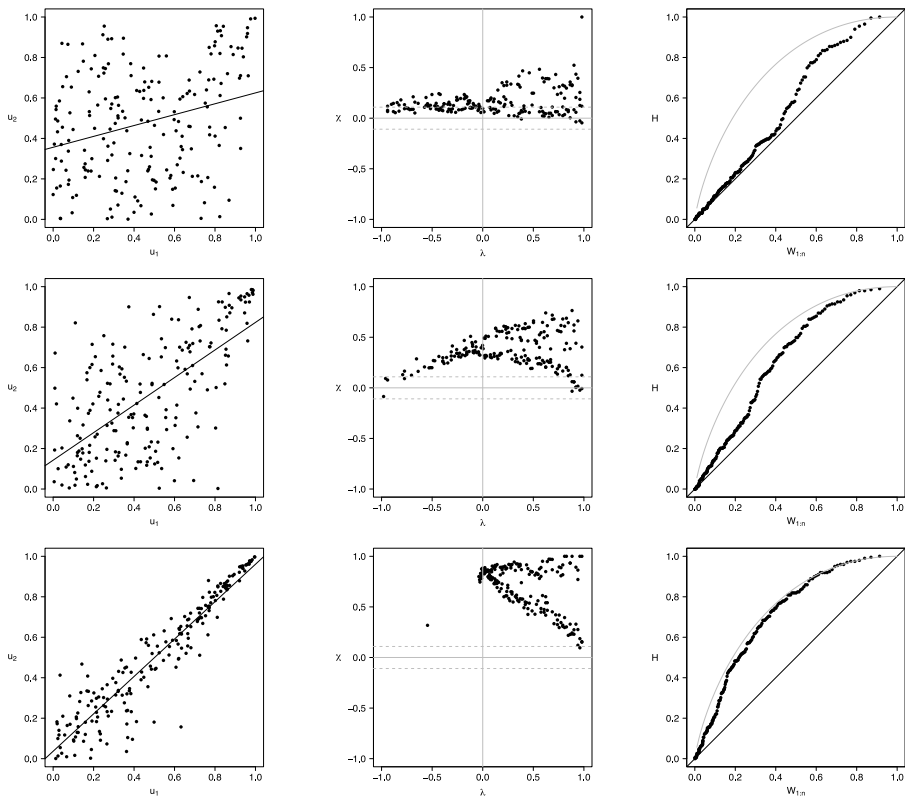**Figure 20.** Scatter-plot (left), χ-plot (center) and *K*-plot (right) for *n* = 200 using the Joe Copula with τ = 0.3 (top), τ = 0.5 (medium) and τ = 0.8 (bottom)

5. APPLICATION TO REAL DATA

In this section, we present an application of graphical methods to detect dependence previously shown to insurance data [17], comparing the results obtained with the traditional scatter plot. A random sam Application to real data
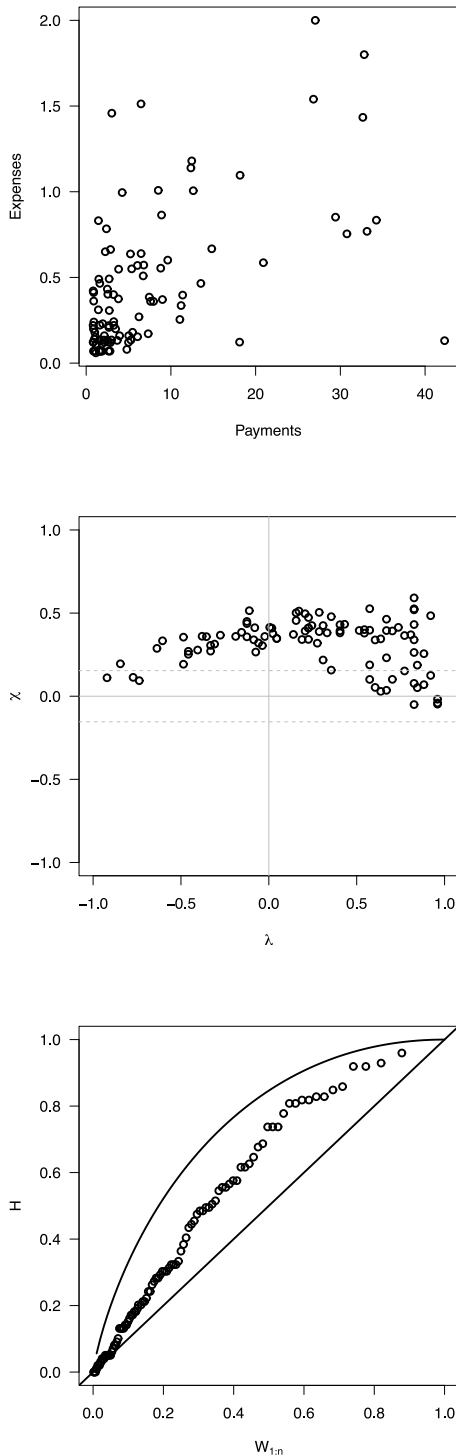
In this section, we present an application of graphical methods to detect dependence previously shown to insurance data [17], comparing the results obtained with the traditional scatter plot. A random sam ple of size 100 of the data was used, consisting of payments and expenses of claims (in millions of pesos) in property insurance policies [17]. The results are shown below:

In figure 5 it can be observed that the Chiplot and the K-plot are able to detect dependence between the two variables used (payments and expenses of claims), in particular, the Chi-plot shows a clear dependence, since most of the observations are outside of the bands. In addition it can be affirmed that the parameter of dependence is high, due to the form of the graphs obtained. In this case the scatter plot is not as clear and precise as the proposed methods.

6. CONCLUSIONS

Graphical methods for detecting dependency studied in this work provide a useful alternative tool to scatter plot traditionally used, since they are simple to interpretate and clearly show if there is dependence between the variables studied.

In simulated scenarios with a small sample size ($n = 20$) the χ-plot and the K-plot achieve the same results as the scatter plot, that is, when the parameter of dependence is low the three methods fail to detect dependence, while under moderate or high dependence the three methods can detect such dependences.

In the simulated scenarios with sample sizes moderate to large ($n \geq 50$) and under low dependence, the χ-plot and the K-plot detect such dependence in at least some of the studied copulas families while the scatter plot does not in any of the cases. On the other hand when the parameter of dependence is moderate to high the three methods can detect such dependences.

In general, the Chi-Plot and K-Plot graphs have the advantage that by increasing the sample size, their performance improves and they manage to detect dependence even when the dependency parameter is $\tau = 0.3$, a result that is not achieved

Figure 21. Scatter-plot (top), χ-plot (medium) and K- plot (bottom) using financial data.

with the scatter plot, since it can not detect dependence when the dependency parameter is low even if the sample size is large. Additionally, the archimedian copulas have a better behavior than the Gaussian copula to detect dependence when the sample sizes are small.

In the application to real data presented in section 5, it can be observed that the $\mathcal{X}$-Plot and the K-plot have a better performance than the scatter plot, since they could detect the dependence between the variables, which was not clear in the scatter plot analysis.

REFERENCES

[1] Escarela, G. and Hernández, A. "Modelado de parejas aleatorias usando cópulas", *Revista Colombiana de Estadística* **32**(1), 33–58, 2009.

[2] Genest, C. and Favre, A. "Everything you always wanted to know about copula modeling but were afraid to ask", *Journal of Hydrologic Engineering* **12**(4), 347–368, 2007.

[3] Nelsen, R. *An Introduction to Copulas*, Sprin- ger Science & Business Media, 2007.

[4] Fisher, N. and Switzer, P. "Chi-plots for asses- sing dependence", *Biometrika* **72**(2), 253–265, 1985.

[5] Genest, C. and Boies, J. "Detecting dependence with Kendall plots", *The American Statistician* **57**(4), 275–284, 2003.

[6] Nguyen, C. C. and Bhatti, M. I. "Copula model dependency between oil prices and stock markets: Evidence from China and Vietnam". *Jour- nal of International Financial Markets, Institu- tions and Money*, **22**(4), 758–773, 2012.

[7] Vandenberghe, S., Verhoest, N. E. C., and De Baets, B. "Fitting bivariate copulas to the dependence structure between storm characteristics: A detailed analysis based on 105 year 10 min rainfall". *Water resources research*, **46**(1), 2010.

[8] Gargouri-Ellouze, E., and Bargaoui, Z. "Investigation with Kendall plots of infiltration index?maximum rainfall intensity relationship for regionalization". *Physics and Chemistry of the Earth, Parts A/B/C*, **34**(10), 642-653, 2009.

[9] Genest, C. and Mackay, R. J. "Copules archimédiennes et familles de lois bidimensionne- lles dont les marges sont données", *Canadian Journal of Statistics* **14**(2), 145–159, 1986.

[10] Evin, G., Favre A.C. and Genest, C. "Comparison of goodness-of-fit tests adapted to copulas", *Geophysical Research Abstracts*, 2005.

[11] De Matteis, R. "Fitting copulas to data". *Insti- tute of Mathematics of the University of Zürich*, 2001.

[12] Embrechts, P., Lindskog, F. and McNeil, A. "Modelling dependence with copulas and applications to risk management", *Technical Re- port, Department of Mathematics, ETH Zurich*, 2001.

[13] Joe, H. *Multivariate models and dependence concepts*, Chapman and Hall/CRC, 1997.

[14] Cintas del Río, R. "Teoría de cópulas y control de riesgo financiero", PhD thesis, Universidad Complutense de Madrid, 2007.

[15] Moreno, D. C. "Método para elegir una cópula Arquimediana óptima", Master's thesis, Universidad Nacional de Colombia, 2012.

[16] R Core Team *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2015.

[17] Lopera, C.M., Jaramillo, M.C. and Arcila, L.D. "Selección de un Modelo Cópula para el Ajuste de Datos Bivariados Dependientes", *Dyna* **76**(158), 253–263, 2009.