Recognizing and visualizing copulas: an approach using local Gaussian approximation

Geir Drage Berentsen, Bård Støve*, Dag Tjøstheim and Tommy Nordbø
University of Bergen, Department of Mathematics,
PO Box 7803, 5020 Bergen, Norway

Abstract

In this paper we examine the relationship between a newly developed local dependence measure, the local Gaussian correlation, and standard copula theory. We are able to describe characteristics of the dependence structure in different copula models in terms of the local Gaussian correlation. Further, we construct a goodness-of-fit test for bivariate copula models. An essential ingredient of this test is the use of a canonical local Gaussian correlation and Gaussian pseudo-observations which make the test independent of the margins, so that it is a genuine test of the copula structure. A Monte Carlo study reveals that the test performs very well compared to a commonly used alternative test. We also propose two types of diagnostic plots which can be used to investigate the cause of a rejected null. Finally, our methods are applied to a "classical" insurance data set.

Keywords: Copulas, goodness-of-fit, local Gaussian correlation

*Corresponding author. Tel. +47 55 58 28 86. E-mail: Bard.Stove@math.uib.no.

1 Introduction

There are two interrelated issues of copula theory that we will look at in this paper: i) visualizing and quantifying the non-linear dependence structure of a copula and ii) employing this to recognize and specify a copula model from given data via a goodness-of-fit test. Both of these issues will be explored using the new tool of local Gaussian correlation in Tjøstheim and Hufthammer (2013). The local Gaussian correlation is a non-linear dependence measure, but it retains the standard correlation interpretation based on a family of local Gaussian approximations.

Typically a copula model contains a few (often only one) parameters that describe the dependence structure. A problem is that the parameters are difficult to interpret. In what way do they measure dependence? A very crude characterization of a copula model is obtained by simulating observations from it and subsequently looking at the resulting scatter diagram, but a scatter diagram is not a very precise quantification of dependence.

Tjøstheim and Hufthammer (2013) examine a local correlation measure that is meant to give a precise mathematical description of non-linear dependence. A brief survey of this measure is given in Section 2. In Section 3 it will be shown how it can be used to precisely characterize and visualize the dependence structure for a number of standard copulas. This measure is defined for bivariate variables, and consequently, with the exception of the brief discussion in Section 7, we consider bivariate copula models in this paper. With the growing interest in pair-copulae constructions (see e.g. Aas et al. (2009)) it is imperative to understand the dependence structure of the bivariate copulas that forms the building blocks of the multivariate model.

Many proposals have been made for goodness-of-fit-testing of copula models, which dates back to Deheuvels (1979). Genest et al. (2009) review and perform a power study of available goodness-of-fit tests for copulas, and a similar study is undertaken by Berg (2009). Further, in goodness-of-fit testing, when a model is rejected, a problem is to identify the cause of the rejection. This problem has been recognized by Berg (2009): "When doing model evaluation... there is still an unsatisfied need for intuitive and informative diagnostic plots."

In this paper we introduce a new goodness-of-fit test for bivariate copula models based on the local Gaussian correlation. The test is carried out using observations that are obtained

via a rank-based transformation of the original observations. The test is based on calculating the difference between the local Gaussian correlation estimated non-parametrically for the transformed observations and estimated by using an analytical expression valid under the null hypothesis of a specific copula. An important asset of the test is that diagnostic plots that were asked for above can be obtained by plotting these estimates together. We also propose a second type of diagnostic plot which displays the results of a "local goodness-of-fit" test. Implementation issues of the goodness-of-fit test are discussed in Section 4 where a simulation study is conducted to assess the power and level of the proposed test, and the diagnostic plots are discussed in Section 5. A practical data example is given in Section 6, and finally, we conclude with possible extensions to the multivariate case in Section 7.

2 Local Gaussian approximation

Let $X = (X_1, X_2)$ be a two-dimensional random variable with density $f(x) = f(x_1, x_2)$. In this section we describe how f can be approximated locally in a neighbourhood of each point $x = (x_1, x_2)$ by a Gaussian bivariate density

$$\psi(v, \mu(x), \Sigma(x)) = \frac{1}{2\pi|\Sigma(x)|^{1/2}} \exp\left[-\frac{1}{2}(v - \mu(x))^T \Sigma^{-1}(x)(v - \mu(x))\right], \tag{2.1}$$

where $v = (v_1, v_2)^T$ is the running variable, $\mu(x) = (\mu_1(x), \mu_2(x))^T$ is the local mean vector and $\Sigma(x) = (\sigma_{ij}(x))$ is the local covariance matrix. With $\sigma_i^2(x) = \sigma_{ii}(x)$, we define the local correlation at the point x by $\rho(x) = \frac{\sigma_{12}(x)}{\sigma_1(x)\sigma_2(x)}$, and in terms of the local correlation, ψ may be written as

$$\begin{aligned}\psi(v, \mu_1(x), \mu_2(x), \sigma_1^2(x), \sigma_2^2(x), \rho(x)) = \\ \frac{1}{2\pi\sigma_1(x)\sigma_2(x)\sqrt{1-\rho^2(x)}} \exp\Bigg\{ -\frac{1}{2(1-\rho^2(x))} \times \Bigg[\left(\frac{v_1 - \mu_1(x)}{\sigma_1(x)}\right)^2 \\ - 2\rho(x)\left(\frac{v_1 - \mu_1(x)}{\sigma_1(x)}\right)\left(\frac{v_2 - \mu_2(x)}{\sigma_2(x)}\right) + \left(\frac{v_2 - \mu_2(x)}{\sigma_2(x)}\right)^2 \Bigg] \Bigg\}. \end{aligned} \tag{2.2}$$

First note that the representation in (2.2) is not well-defined unless extra conditions are imposed. We need to construct a Gaussian approximation that approximates $f(x)$ in a

neighborhood of x and such that (2.2) holds at x. In the case of $X \sim \mathcal{N}(\mu, \Sigma)$ this is trivially obtained by taking one Gaussian; i.e., $\mu(x) = \mu$ and $\Sigma(x) = \Sigma$ for all x. In fact, these relationships may be taken as definitions of the local parameters for a Gaussian distribution.

In Tjøstheim and Hufthammer (2013) it was demonstrated that for a given neighbourhood characterized by a bandwidth parameter b the local population parameters $\gamma(x) = (\mu(x), \Sigma(x))$ or $\gamma(x) = (\mu_1(x), \mu_2(x), \sigma_1^2(x), \sigma_2^2(x), \rho(x))$ can be defined by minimizing a likelihood related penalty function q given by

$$q = \int K_b(v - x) \left[\psi(v, \gamma(x)) - \log \psi(v, \gamma(x)) f(v)\right] dv, \tag{2.3}$$

where again b is the bandwidth parameter, and $K_b(v - x) = b^{-1}K(b^{-1}(v - x))$ with K being a kernel function. Such a penalty function was used in Hjort and Jones (1996) for density estimation purposes. They argue that it can be interpreted as a locally weighted Kullback-Leibler criterion for measuring the distance between f and $\psi(\cdot, \gamma(x))$. We define the population value $\gamma(x) = \gamma_b(x)$ as the the minimizers of this penalty function. It then satisfies the set of equations

$$\int K_b(v - x) \frac{\partial}{\partial \gamma_j} \log \psi(v, \gamma(x)) \left[f(v) - \psi(v, \gamma(x))\right] dv = 0, \quad j = 1, \ldots, 5. \tag{2.4}$$

It is assumed that there is a bandwidth b_0 such that there exists a unique solution of the minimization problem for any b with $0 < b < b_0$.

It is easy to find examples where (2.4) is satisfied with a unique $\gamma_b(x)$. A trivial example is when $X \sim \mathcal{N}(\mu, \Sigma)$ is Gaussian, where $\gamma_b(x) = \gamma(x) = (\mu, \Sigma)$. The next step is introducing a piecewise linear function of a Gaussian variable $Z \sim \mathcal{N}(\mu, \Sigma)$, as defined in Tjøstheim and Hufthammer (2013). Since it will be used later, we restate it here. We take $\mu = 0$ and $\Sigma = I_2$, the identity matrix of dimension 2. Let $R_i, i = 1, \ldots, k$ be a set of non-overlapping regions of $\mathbb{R}^2$ such that $\mathbb{R}^2 = \cup_{i=1}^k R_i$. Further, let a_i and A_i be a corresponding set of vectors and matrices in $\mathbb{R}^2$ such that A_i is non-singular and define the piecewise linear function

$$X = g(Z) = \sum_{i=1}^{k} (a_i + A_i Z) 1(Z \in R_i), \tag{2.5}$$

where $1(\cdot)$ is the indicator function. Let S_i be the region defined by $S_i = \{x : x = a_i +$

$A_i z, z \in R_i\}$. It is assumed that (2.5) is one-to-one in the sense that $S_i \cap S_j = \emptyset$ for $i \neq j$ and $\cup_{i=1}^k S_i = \mathbb{R}^2$. To see that the linear step function (2.5) can be used to obtain a solution of (2.4) let x be a point in the interior of S_i and let the kernel function K have a compact support. If $v - x$ is in the support of K_b, then b can be made small enough so that $v \in S_i$. Under this restriction on b, $\gamma_b(x) = \gamma(x) \equiv \gamma_i = (\mu_i, \Sigma_i)$ where $\mu_i = a_i$ and $\Sigma_i = A_i A_i^T$ as defined in (2.5). Thus, in this sense, for a fixed but small b, there exists a local Gaussian approximation $\psi(x, \gamma_b)$ of f, with corresponding local means $\mu_{i,b}(x)$, variances $\sigma_{i,b}^2(x)$, $i = 1, 2$, and correlation $\rho_b(x)$.

It was shown in Tjøstheim and Hufthammer (2013) that once a unique population vector $\gamma_b(x)$ exists, under weak regularity conditions one can let $b \to 0$ to obtain a local population vector $\gamma(x)$ defined at a point x. The population vectors $\gamma_b(x)$ and $\gamma(x)$ can both be consistently estimated by using a local log-likelihood function defined by

$$L\Big(X_1, \ldots, X_n, \gamma_b(x)\Big) = n^{-1} \sum_i K_b(X_i - x) \log \psi(X_i, \gamma_b(x)) - \int K_b(v - x) \psi(v, \gamma_b(x))\, \mathrm{d}v, \quad (2.6)$$

for given observations $X_1, \ldots, X_n$. This likelihood is taken from Hjort and Jones (1996) where it was used for density estimation. Here, the X_i's are iid observations or more generally from an ergodic time series $\{X_t\}$. In the latter case (2.6) could be thought of as a marginal local likelihood function.

The numerical maximization of the local likelihood (2.6) leads to local likelihood estimates $\gamma_{n,b}(x)$, including estimates $\rho_{n,b}(x)$ of the local correlation. It is shown in Tjøstheim and Hufthammer (2013) that under relatively weak regularity conditions $\gamma_{n,b}(x) \to \gamma_b(x)$ for b fixed, and $\gamma_{n,b}(x) \to \gamma(x)$ almost surely for $b = b_n$ tending to zero. In addition asymptotic normality is demonstrated in that paper. Further, equation (2.6) is consistent with (2.3). An R-package "localgauss" has been developed for finding the local likelihood estimates $\gamma_{n,b}(x)$ including $\rho_{n,b}(x)$ and is publicly available for Linux and Windows at the Comprehensive R Archive Network (CRAN, `http://CRAN.R-project.org/`). For a detailed description, see Berentsen et al. (2014). Many examples of real and simulated data are given in that paper, and in Tjøstheim and Hufthammer (2013), Støve and Tjøstheim (2014), Støve et al. (2014) and Berentsen and Tjøstheim (2014). The quantity $\rho(x)$ has a precise interpretation as a local Gaussian correlation. A number of properties has been given in Tjøstheim and

Hufthammer(2013, see especially Section 6).

3 Local Gaussian correlation for copula models

We start by a generalization of the so-called Rosenblatt (1952) transformation:

Lemma 3.1. *Let X have a density $f_X(x)$ on $\mathbb{R}^2$ with cumulative distribution function $F_X(x) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f_X(w_1, w_2)\,\mathrm{d}w_1\,\mathrm{d}w_2$. Then there exists a one-to-one function g such that $X = g(Z)$, where $Z \sim \mathcal{N}(0, I_2)$.*

Proof. We have $f_X(x) = f_{X_1}(x_1)f_{X_2|X_1}(x_2|x_1)$. Then $U_1 = F_{X_1}(X_1)$ is uniform. There also exists a standard normal variable Z_1 such that $U_1 = \Phi(Z_1)$, where Φ is the cumulative distribution function of the standard normal density. Hence, $X_1 = F_{X_1}^{-1}(\Phi(Z_1))$. In the same manner, there exists a uniform variable U_2 independent of U_1 (see Rosenblatt (1952)) such that $U_2 = F_{X_2|X_1}(X_2|X_1)$, and there exists a $Z_2 \sim \mathcal{N}(0,1)$ independent of Z_1 such that $U_2 = \Phi(Z_2)$, and hence

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} F_{X_1}^{-1}(\Phi(Z_1)) \\ F_{X_2|X_1}^{-1}(\Phi(Z_2)|F_{X_1}^{-1}(\Phi(Z_1)) \end{bmatrix} \doteq g(Z), \tag{3.1}$$

where $F_{X_2|X_1}^{-1}$ is interpreted as the inverse of $F_{X_2|X_1}$ with X_1 fixed (i.e., with U_1, Z_1 fixed). Here g is one-to-one due to the strict monotonicity of F_X. □

As pointed out in Rosenblatt (1952) this representation is non-unique, since we also have

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} F_{X_1|X_2}^{-1}(\Phi(Z_1)|F_{X_2}^{-1}(\Phi(Z_2)) \\ F_{X_2}^{-1}(\Phi(Z_2)) \end{bmatrix} \doteq g'(Z'), \tag{3.2}$$

where in general $g \neq g'$ and $Z \neq Z'$.

Assuming g to be continously differentiable at z and Taylor expanding, $X = g(Z) = g(z) + \frac{\partial g}{\partial z}(z)(Z - z) + o_p(|Z - z|)$, where $\frac{\partial g}{\partial z}$ is the Jacobi matrix. For this representation it is tempting to define the local mean $\mu(x)$ and $\Sigma(x)$ of the density of X at the point x as the mean and covariance of the Gaussian variable $U_z(Z) = g(z) + \frac{\partial g}{\partial z}(z)(Z - z)$. These are expressed as functions of x using $z = h(x) = g^{-1}(x)$. Since $\mathbb{E}(Z) = 0$ and $\Sigma(Z) = I_2$ this

results in

$$\mu(x) = g(z) - \frac{\partial g}{\partial z}(z)z = x - \left(\frac{\partial h}{\partial x}(x)\right)^{-1} h(x) \tag{3.3}$$

and

$$\Sigma(x) = \frac{\partial g}{\partial z}(z)\left(\frac{\partial g}{\partial z}(z)\right)^T = \left(\frac{\partial h}{\partial x}(x)\right)^{-1}\left(\left(\frac{\partial h}{\partial x}(x)\right)^{-1}\right)^T. \tag{3.4}$$

It is an easy matter to verify that $f_{U_z(Z)} = \psi(v, \mu(x), \Sigma(x))$ yields a representation of type (2.1). The problem is that the non-uniqueness of the two Rosenblatt representations means that in general $\mu'(x) \neq \mu(x)$, $\Sigma'(x) \neq \Sigma(x)$ and $\rho'(x) \neq \rho(x)$.

We now look at (3.1) and (3.2) and examine under what conditions these two transformations will give rise to a unique local Gaussian correlation that can be expressed analytically and can be used to recognize copulas. By (3.1) and (3.4) the matrix

$$\frac{\partial h}{\partial x}(x) = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} \\ \frac{\partial h_2}{\partial x_1} & \frac{\partial h_2}{\partial x_2} \end{bmatrix} \tag{3.5}$$

is lower triangular and

$$\left(\frac{\partial h}{\partial x}(x)\right)^{-1} = \left(\frac{\partial h_1}{\partial x_1}\frac{\partial h_2}{\partial x_2}\right)^{-1} \begin{bmatrix} \frac{\partial h_2}{\partial x_2} & 0 \\ \frac{-\partial h_2}{\partial x_1} & \frac{\partial h_1}{\partial x_1} \end{bmatrix},$$

which, using (3.4), results in the following local Gaussian correlation

$$\rho(x) = \rho(x_1, x_2) = \frac{\Sigma_{12}(x)}{\sqrt{\Sigma_{11}(x)\Sigma_{22}(x)}} = \frac{-\frac{\partial h_2}{\partial x_1}}{\sqrt{\left(\frac{\partial h_1}{\partial x_1}\right)^2 + \left(\frac{\partial h_2}{\partial x_1}\right)^2}}, \tag{3.6}$$

where we return to its validity and uniqueness below. Next, consider a continuous random variable $X = (X_1, X_2)$ with joint cumulative distribution function F and margins $F_{X_1}(x_1) = F_1(x_1)$ and $F_{X_2}(x_1) = F_2(x_2)$. Due to the representation theorem of Sklar (1959), F can be written as

$$F(x_1, x_2) = C((F_1(x_1), F_2(x_2)), \tag{3.7}$$

where the copula $C : [0,1]^2 \to [0,1]$ is a unique bivariate distribution function with uniform margins. Since any continuous distribution function F has the representation (3.7) we may re-express (3.1) and thus $\rho(x_1, x_2)$ in terms of the copula C and the margins F_1 and F_2. By

standard theory, $F_{2|1}^{-1}(x_2|x_1)$ may be written as

$$F_{2|1}^{-1}(x_2|x_1) = F_2^{-1}\left(C_1^{-1}(F_1(x_1), x_2)\right),$$

where $C_1^{-1}(u, v)$ is interpreted as the inverse of $C_1(u, v) = \frac{\partial}{\partial u}C(u, v)$ with u fixed. It follows that (3.1) may be written as

$$g(Z) = \begin{bmatrix} F_1^{-1}(\Phi(Z_1)) \\ F_2^{-1}\left(C_1^{-1}\left(\Phi(Z_1), \Phi(Z_2)\right)\right) \end{bmatrix}. \tag{3.8}$$

Note that this transformation (only with $\Phi(Z_1)$ and $\Phi(Z_2)$ replaced by two independent uniform $[0, 1]$ variables) is a standard way of sampling from the distribution $C(F_1(x_1), F_2(x_2))$ (see e.g. Nelsen, 2006, page 35-37). In the continuous case, g is one-to-one if the copula density $c(u_1, u_2)$ satisfies $c(u_1, u_2) > 0$ for all points $(u_1, u_2) \in [0, 1]^2$ (This guarantees the invertibility of $C_1(u_1, u_2)$ with respect to u_2). The inverse $h = g^{-1}$ is then given by

$$h(X) = \begin{bmatrix} h_1(X_1, X_2) \\ h_2(X_1, X_2) \end{bmatrix} = \begin{bmatrix} \Phi^{-1}\left(F_1(X_1)\right) \\ \Phi^{-1}\left(C_1\left(F_1(X_1), F_2(X_2)\right)\right) \end{bmatrix}. \tag{3.9}$$

Towards finding an expression for $\rho(x_1, x_2)$ using (3.6), let ϕ denote the standard normal density function and let $C_{11}(u_1, u_2) = \frac{\partial^2}{\partial u_1^2}C(u_1, u_2)$. Then by using the two partial derivatives of h, the local Gaussian correlation (3.6) for the model (3.7) may be written as

$$\rho(x_1, x_2) = \frac{-C_{11}(F_1(x_1), F_2(x_2))\phi\left(\Phi^{-1}\left(F_1(x_1)\right)\right)}{\sqrt{\phi^2\left(\Phi^{-1}\left(C_1(F_1(x_1), F_2(x_2))\right)\right) + C_{11}^2(F_1(x_1), F_2(x_2))\phi^2\left(\Phi^{-1}\left(F_1(x_1)\right)\right)}}. \tag{3.10}$$

However, repeating the above argument with the Rosenblatt representation (3.2) as a starting point instead of (3.1) leads to another local Gaussian correlation $\rho'(x_1, x_2)$ given by

$$\rho'(x_1, x_2) = \frac{-C_{22}(F_1(x_1), F_2(x_2))\phi\left(\Phi^{-1}\left(F_2(x_2)\right)\right)}{\sqrt{\phi^2\left(\Phi^{-1}\left(C_2(F_1(x_1), F_2(x_2))\right)\right) + C_{22}^2(F_1(x_1), F_2(x_2))\phi^2\left(\Phi^{-1}\left(F_2(x_2)\right)\right)}} \tag{3.11}$$

where $C_2(u_1, u_2) = \frac{\partial}{\partial u_2}C(u_1, u_2)$ and $C_{22}(u_1, u_2) = \frac{\partial^2}{\partial u_2^2}C(u_1, u_2)$.

Since the two Rosenblatt representations are bases for any representation of $f_X(x)$, (including a density generated by a general functional relationship $X = g(Z)$), we have unique-

ness at points where they coincide. The local parameters along such curves are consistent with the local parameters derived from the local penalty function (2.3). Indeed, for a point x where the Rosenblatt representations give a unique $\gamma(x) = (\mu(x), \Sigma(x))$ such that $f(x) = \psi(x, \gamma(x))$, a local Gaussian approximation with $\gamma_b(x)$ can be found that satisfies the equations (2.4) and that converges to $\gamma(x)$. Simply choose a linear stepwise representation (2.5), such that $x \in S_i$ for some i, and take $A_i = \Sigma^{1/2}(x)$ and $a_i = \mu(x)$. Then with a small enough bandwidth, $\gamma_b(x) = \gamma_i = (a_i, A_i A_i^T) = (\mu(x), \Sigma(x))$, and $\gamma_b(x) \to \gamma(x)$ trivially as $b \to 0$. If for a point x there is not a unique Rosenblatt representation, then such an approach is not possible since there is not a unique $\gamma(x)$ that could serve as a starting point for the construction. Nevertheless, for such points x, under the regularity conditions mentioned in Tjøstheim and Hufthammer (2013), there does exist a limiting unique minimizer $\gamma(x)$ and resulting $\rho(x)$ such that the local likelihood estimate $\gamma_{n,b}(x)$ (or $\rho_{n,b}(x)$) converges towards $\gamma(x)$ (or $\rho(x)$) (see Tjøstheim and Hufthammer (2013), Theorem 1-3).

Inspecting (3.10) and (3.11) it is seen that in the copula case, when the copula is exchangeable (i.e. $C(u_1, u_2) = C(u_2, u_1)$), the points where $\rho'(X_1, X_2) = \rho(X_1, X_2)$, are found along the curve defined by $F_1(x_1) = F_2(x_2)$. In the particular case of identical margins, which is true when X_1 and X_2 are exchangeable, we have equality along the diagonal $x_1 = x_2$.

Note that since $\phi\left(\Phi^{-1}\left(F_1(\cdot)\right)\right) > 0$, the sign of $\rho(x_1, x_2)$ is determined by the sign of $-C_{11}(F_1(x_1), F_2(x_2))$. To see that this is reasonable, consider a random variable X_1 positively related to the variable X_2 in the neighbourhood of (x_1, x_2), in the sense that $m(s) \doteq P(X_2 \leq x_2 | X_1 = s) = C_1(F_1(s), F_2(x_2))$ is decreasing as s increases (in a neighbourhood of x_1). Then since $m'(s) < 0$, we have that $-C_{11}(F_1(x_1), F_2(x_2)) > 0$ and thus $\rho(x_1, x_2) > 0$ in the neighbourhood of (x_1, x_2).

When X_1 and X_2 are independent their copula is the independence copula $C(u_1, u_2) = u_1 u_2$. Then $C_{11}(u_1, u_2) = 0$ which implies that $\rho(x_1, x_2) = 0$ along the curve $F_1(x_1) = F_2(x_2)$. In Tjøstheim and Hufthammer (2013) it is shown that independence implies $\rho(x) = 0$ everywhere and that a necessary and sufficient condition for independence is that $\rho(x) \equiv 0$, $\mu_i(x) \equiv \mu_i(x_i)$, $\sigma_i^2(x) \equiv \sigma_i^2(x_i)$, $i = 1, 2$.

Tjøstheim and Hufthammer (2013) consider the connection between $\rho(x_1, x_2)$ and the upper and lower tail dependence index given by

$$\lambda_u = \lim_{q \to 1^-} P(F_2(X_2) > q | F_1(X_1) > q) = \lim_{q \to 0^+} \frac{C(q, q)}{q} \doteq \lim_{q \to 0^+} \lambda_{u,q}, \tag{3.12}$$

$$\lambda_l = \lim_{q\to 0^+} P(F_2(X_2) \leq q | F_1(X_1) \leq q) = \lim_{q\to 0^+} \frac{C_S(q,q)}{q} \doteq \lim_{q\to 0^+} \lambda_{l,q}, \tag{3.13}$$

where C_S is the survival copula given by $C_S(u_1, u_2) = u_1 + u_2 - 1 + C(1-u_1, 1-u_2)$. Let $U(x) = (U_1(x), U_2(x))$ be the local Gaussian approximation of $f(x)$ at the point $x = (s, s)$, then the lower tail dependence index of $U(x)$ is given by

$$\lambda_l = \lambda_l(s,s) = 2 \lim_{u\to-\infty} \Phi\left(u\sqrt{\frac{1-\rho(s,s)}{1+\rho(s,s)}}\right),$$

and if $U(x)$ should approximate $f(x)$ in the tail as $s \to -\infty$, positive tail dependence would require $\rho(s,s) \to 1$. Thus by (3.10), $\rho(x_1, x_2)$ for copula models with lower tail dependence should satisfy

$$\lim_{s\to-\infty} \rho(s,s) = \lim_{q\to 0^+} \frac{-C_{11}(q,q)\phi\left(\Phi^{-1}(q)\right)}{\sqrt{\phi^2\left(\Phi^{-1}\left(C_1(q,q)\right)\right) + C_{11}^2(q,q)\phi^2\left(\Phi^{-1}(q)\right)}} = 1. \tag{3.14}$$

For exchangeable copulas with lower tail dependence λ_l, it can be shown that $\lim_{q\to 0^+} \phi^2\left(\Phi^{-1}\left(C_1(q,q)\right)\right) = \phi^2\left(\Phi^{-1}(\lambda_l/2)\right) \neq 0$. So for (3.14) to hold when $\lambda_l \neq 0$ we must have that $\lim_{q\to 0^+} -C_{11}(q,q)\phi\left(\Phi^{-1}(q)\right) = \infty$. This can for example be verified for the Clayton copula. For the speed at which $\rho(s,s) \to 1$ for the Clayton copula we refer to Figure 1 in the case of standard Gaussian margins.

3.1 Canonical local Gaussian correlation

Above we derived an analytical expression of the local correlation for a variable (X_1, X_2) with distribution function $C(F_1(X_1), F_2(X_2))$. In addition to the copula C, the resulting local Gaussian correlation $\rho(x)$ in (3.10) also depends on the margins F_1 and F_2. By Sklar's theorem (Sklar (1959)) the copula C contains all information on the dependence structure between X_1 and X_2 and it is therefore unfortunate that the local correlation depends on the margins when our goal is to characterize the copula and to use it in goodness-of-fit testing. However, it turns out that this problem can be circumvented, and that the dependence structure of the copula of X_1 and X_2 can be characterized by the local Gaussian correlation of the variable

$$Z = (Z_1, Z_2) = \left(\Phi^{-1}(F_1(X_1)), \Phi^{-1}(F_2(X_2))\right), \tag{3.15}$$

where, as before, Φ is the standard normal distribution function. Since the copula C of a continuous random variable $X = (X_1, X_2)$ is invariant to any continuous, strictly increasing transformations of X_1 and X_2, Z and X share the same copula. Moreover, since $F_1(X_1)$ and $F_2(X_2)$ are both marginally uniformly distributed, Z_1 and Z_2 are marginally standard normal distributed. It follows that the distribution function of Z is given by $F(z_1, z_2) = C(\Phi(z_1), \Phi(z_2))$. For a random variable $Z = (Z_1, Z_2)$ with distribution function $C(\Phi(z_1), \Phi(z_2)) = C_\theta(\Phi(z_1), \Phi(z_2))$, (3.10) simplifies to

$$\rho_\theta(z_1, z_2) \doteq \frac{-C_{11}(\Phi(z_1), \Phi(z_2))\phi(z_1)}{\sqrt{\phi^2\left(\Phi^{-1}\left(C_1(\Phi(z_1), \Phi(z_2))\right)\right) + C_{11}^2(\Phi(z_1), \Phi(z_2))\phi^2(z_1)}}. \tag{3.16}$$

Thus, for the variable Z defined by (3.15) the local Gaussian correlation is independent of the margins F_1 and F_2, and we therefore define (3.16) as the canonical local Gaussian correlation for the copula C. The subscript θ in $\rho_\theta(z_1, z_2)$ emphasizes that $\rho_\theta(z_1, z_2)$ only depends on the parameter(s) θ of the copula C. Of course, the variable Z can not be observed via the transformation (3.15) without knowledge of the margins F_1 and F_2. Nevertheless, as will be seen in Section 4, given observations $X_1, \ldots, X_n$ one can obtain an approximate sample from $C(\Phi(z_1), \Phi(z_2))$ via an empirical version of the transformation (3.15). Using this sample of Gaussian pseudo-observations, one can estimate $\rho_\theta(\cdot)$ by the local likelihood method described in Section 2 and subsequently compare this estimate with the analytic expression for $\rho_\theta(\cdot)$ for different copulas. The choice of Gaussian margins in the transformation (3.15) is not arbitrary. Since the copula of (X_1, X_2) is defined as the distribution function of $(U_1, U_2) = (F_1(X_1), F_2(X_2))$ one could, in principle, consider the local Gaussian correlation of the variable (U_1, U_2) directly. However, fitting a Gaussian density to finite support variables requires special considerations of boundary effects, which makes this approach unpractical and illogical. The choice of Gaussian margins is natural since we are dealing with local Gaussian approximations, and also because this choice simplifies (3.10) into (3.16). The point is that we choose to describe the dependence properties of different copulas C via the dependence properties of the corresponding distribution function $C(\Phi(z_1), \Phi(z_2))$ (or via an approximate sample from the same distribution). For this reason we will first derive the canonical local Gaussian correlation $\rho_\theta(z_1, z_2)$ for variables (Z_1, Z_2) with distribution function $C(\Phi(z_1), \Phi(z_2))$ for several well known copulas C. We return to the topic of recognizing copulas via a goodness-of-fit test in Section 4. In that section we will see that considering C

on a scale $C(\Phi(\cdot), \Phi(\cdot))$ may have distinct advantages compared to the scale $C(U(\cdot), U(\cdot))$ with $U(x) = x$ on $[0, 1]$.

3.2 Examples

All copulas considered in the following examples are exchangeable and since Z_1 and Z_2 are standard Gaussian (i.e. $F_1 = F_2 = \Phi$) $\rho_\theta(z_1, z_2)$ is well defined along the diagonal $z_1 = z_2$. In practice, given a copula, the formula (3.16) often becomes quite complicated. As a consequence, for the examples in Section 3.2.1 and Section 3.2.2, we only formulate the functions C_1 and C_{11} and refer to Figure 1 - 4 for the characteristics of $\rho_\theta(z_1, z_2)$ for each copula. In Figures a) 1 - 4 we have plotted $\rho_\theta(s, s)$ against s. The copula parameters in these plots are chosen so that they correspond to a specific value of Kendall's tau ($\tau = 0.2, 0.4, 0.6, 0.8$), which in general is uniquely related to the (one-parameter) copula C by the formula

$$\tau = m(\theta) = 4 \int \int_{[0,1]^2} C(u, v) \, \mathrm{d}C(u, v) - 1. \tag{3.17}$$

We have not been able to find an analytic expression of the local Gaussian correlation valid for general (z_1, z_2), but using the local likelihood algorithm the local correlation defined by penalty function (2.3)) can be estimated for all (z_1, z_2) for which there are data. Figures b) 1 - 4 display this estimate based on one realization of $n = 1000$ samples from each of the copula models $C(\Phi(z_1), \Phi(z_2))$ considered, with copula parameter corresponding to $\tau = 0.4$ and with bandwidth b chosen according to the procedure outlined in Section 4. There is some boundary bias in the estimation, but the estimated dependence patterns revealed in Figures b) 1 - 4 are consistent with the theoretical ones along the diagonal in Figures a) 1 - 4.

3.2.1 Archimedean copulas

An important class of copulas is the class of Archimedean copulas, which have been extensively studied. These copulas are completely defined by their so-called generator function φ, see its properties in e.g. Nelsen (2006). In the following three examples we consider the commonly used Archimedean copulas Clayton, Gumbel and Frank.

Example 3.1 (Clayton copula)**.** The Clayton copula is an asymmetric copula, exhibiting

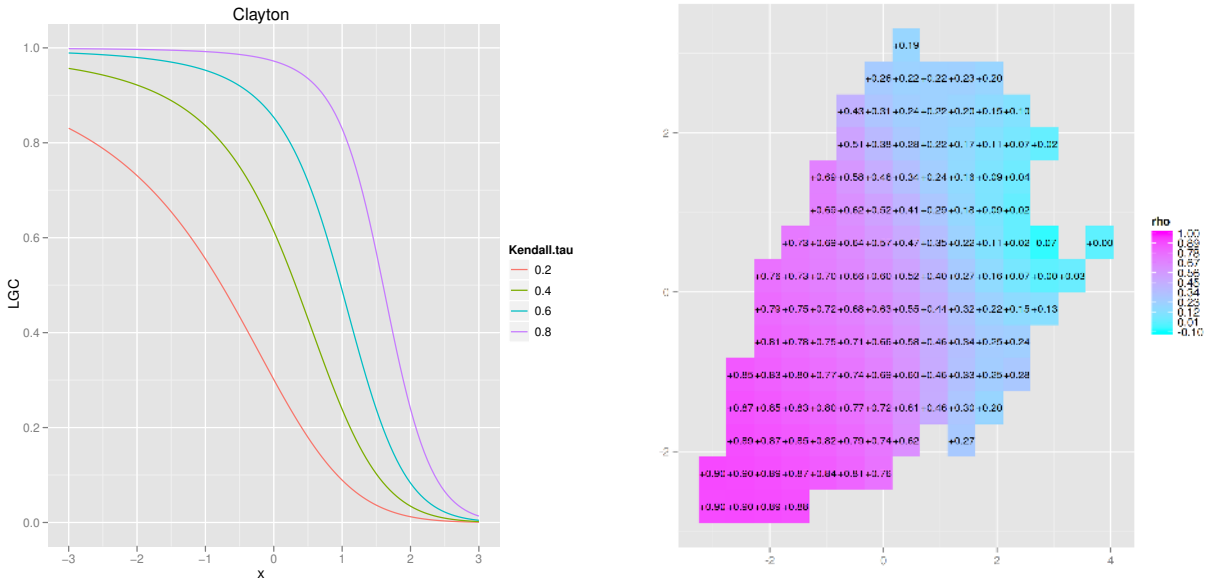
greater dependence in the negative tail than in the positive (i.e. lower tail dependence and no upper tail dependence).

$$C_\theta^{Cl}(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta},$$

with derivatives

$$C_1(u_1, u_2) = \left(1 + u_1^\theta(u_2^{-\theta} - 1)\right)^{-\frac{\theta+1}{\theta}}$$
$$C_{11}(u_1, u_2) = (\theta + 1)u_2^{\theta-1}(1 - u_2^{-\theta})(1 + u_1^\theta(u_2^{-\theta} - 1))^{-1/\theta-2}$$

This implies that $\rho_\theta(s, s) \to 0$ as $s \to \infty$ and $\rho_\theta(s, s) \to 1$ as $s \to -\infty$. These features can be seen in Figure 1. This plot (and the subsequent plots for the other copulas) gives a



(a) (b)

Figure 1: Local Gaussian correlation for $C(\Phi(z_1), \Phi(z_2))$ where C is the Clayton copula: (a) along the diagonal $z_1 = z_2$; (b) estimated based on $n = 1000$ observations with $\tau = 0.4$.

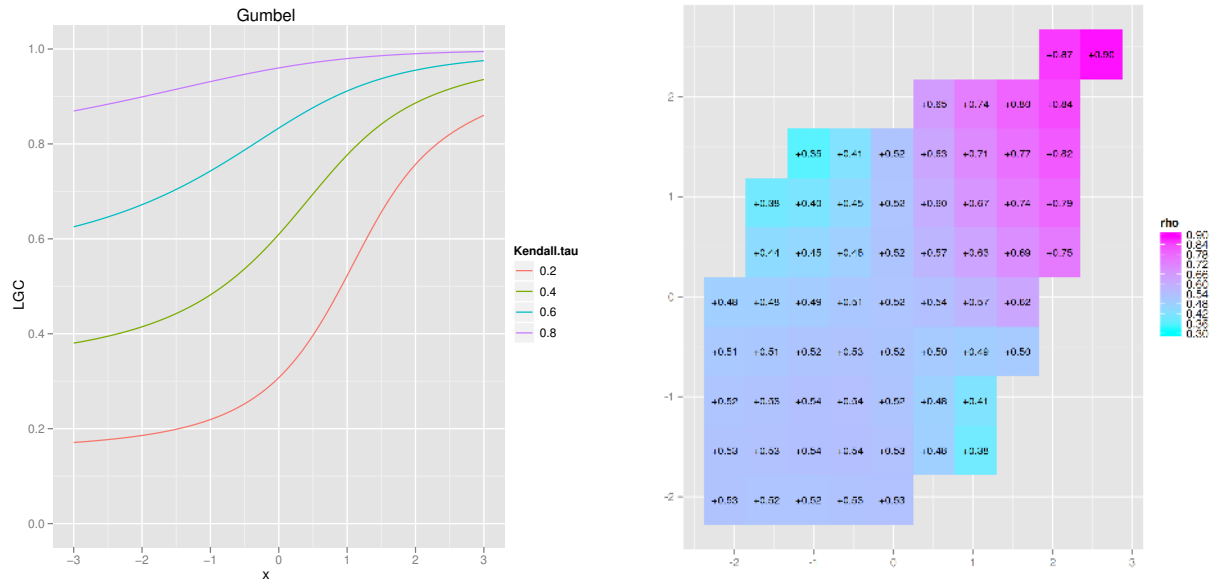
precise and interpretable characterization of the local dependence properties of the Clayton copula. It replaces an informal scatter plot. It also gives a vastly more detailed picture of the (asymmetric) dependence properties than the one-number characterization of the Kendall's

tau. The same values of Kendall's tau have been used in Figures 2-4, but the other copulas have very different local correlation curves associated with them, this difference forming the basis for the formal goodness-of-fit test in Section 4.

Example 3.2 (Gumbel copula)**.** The Gumbel copula is also an asymmetric copula, exhibiting greater dependence in the positive tail than in the negative (i.e. upper tail dependence and no lower tail dependence). The Gumbel copula can be written as

$$C_\theta^{Gu}(u_1, u_2) = \exp\Big[-((-\ln u_1)^\theta + (-\ln u_2)^\theta)^{1/\theta}\Big].$$

The functions C_1 and C_{11} are quite complicated and are therefore not given here. The characteristics of $\rho_\theta(z_1, z_2)$ for the Gumbel copula can be seen in Figure 2 where we clearly see the upper tail dependence numerically quantified in terms of the local correlation.



(a) (b)

Figure 2: Local Gaussian correlation for $C(\Phi(z_1), \Phi(z_2))$ where C is the Gumbel copula (a) along the diagonal $z_1 = z_2$; (b) estimated based on $n = 1000$ observations with $\tau = 0.4$.

An alternative to modeling variables (X_1, X_2) with upper tail dependence by the Gumbel copula is to model $(-X_1, -X_2)$ by the Clayton copula. What separates the resulting dependence structure in these two approaches? From Figure 1 and 2 we see that the dependence

structure in the "non-dependent" tail of these two copulas is quite different. For the Clayton copula $\rho_\theta(z_1, z_2)$ in the upper tail approaches zero faster than in the lower tail of for the Gumbel copula with corresponding values of τ. This is consistent with a plot (not shown) of the upper tail dependence coefficient $\lambda_{u,q}$ given by (3.12) for the Clayton copula together with the lower tail coefficient $\lambda_{l,q}$ given by (3.13) for the Gumbel copula, where both copula parameters are chosen to correspond to $\tau = 0.4$. Then $\lambda_{u,q} < \lambda_{l,q}$ as $q \to 0$ indicating that the upper tail dependence in the Clayton copula vanishes faster than the lower tail dependence in the Gumbel copula. This distinction plays an important role in the empirical example of Section 6.

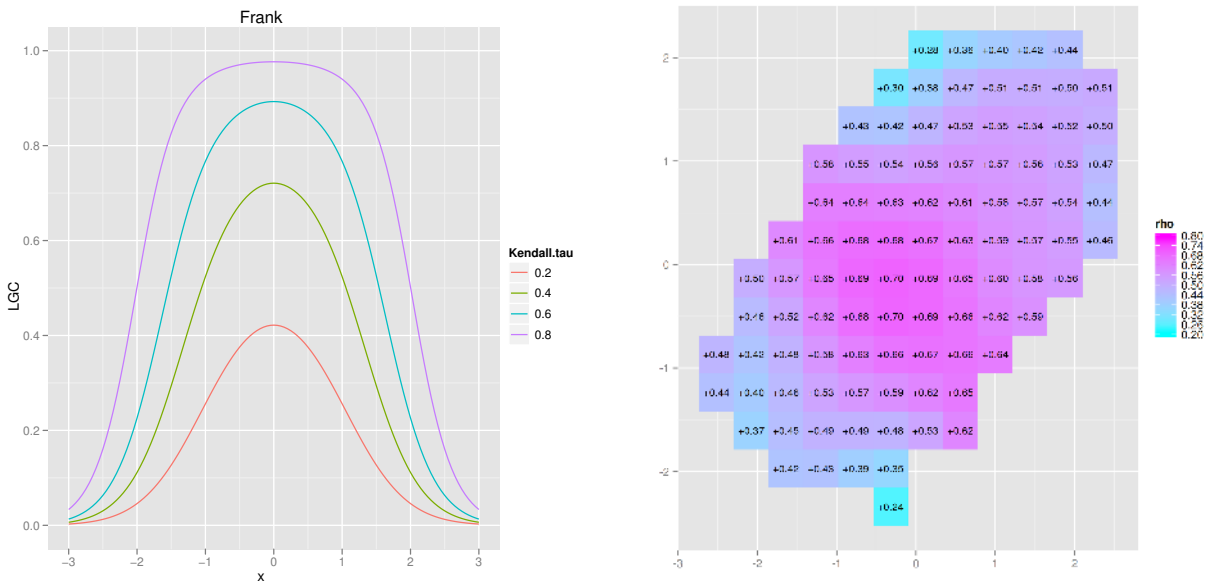
Example 3.3 (Frank copula)**.** Define $q_t = e^{-\theta t} - 1$. The Frank copula may be written as

$$C_\theta^{Fr}(u_1, u_2) = -\theta^{-1} \ln \{1 + q_{u_1} q_{u_2} / q_1\}$$

The derivatives C_1 and C_1 are

$$C_1(u_1, u_2) = \frac{q_{u_1} q_{u_2} + q_{u_2}}{q_{u_1} q_{u_2} + q_1}$$
$$C_{11}(u_1, u_2) = \frac{q'_{u_1} q_{u_2} (q_1 - q_{u_2})}{(q_{u_1} q_{u_2} + q_1)^2}$$

Thus $\rho_\theta(z_1, z_2)$ goes to zero in both the upper and lower tail. This feature is reflected in Figure 3; close to constant dependence in the center which vanish in the tails.



(a) (b)

Figure 3: Local Gaussian correlation for $C(\Phi(z_1), \Phi(z_2))$ where C is the Frank copula: (a) along the diagonal $z_1 = z_2$; (b) estimated based on $n = 1000$ observations with $\tau = 0.4$.

3.2.2 Elliptical copulas

Elliptical copulas are the copulas of elliptical distributions. The key advantage of elliptical copulas is that one can specify different levels of global correlation between the marginals, but a disadvantage is that elliptical copulas typically do not have simple closed form expressions. The most commonly used elliptical distributions are the Gaussian and Student-t distributions.

Example 3.4 (Gaussian copula)**.** For a given correlation matrix $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ the Gaussian copula with correlation matrix Σ can be written as

$$C_{\Sigma}^{Gauss}(u_1, u_2) = \Phi_{\Sigma}(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) \tag{3.18}$$

where Φ_{Σ} is the joint bivariate distribution function of a Gaussian variable with mean vector zero and correlation matrix Σ. In general, when (Z_1, Z_2) is Gaussian with mean vector zero and correlation matrix Σ, then $Z_1 | Z_2 = z_2 \sim N(\rho z_2, 1 - \rho^2)$. It follows that for the Gaussian

copula

$$C_1(u_1,u_2) = P(U_2 \le u_2 | U_1 = u_1) = P(\Phi^{-1}(U_2) \le \Phi^{-1}(u_2) | \Phi^{-1}(U_1) = \Phi^{-1}(u_1))$$
$$= \Phi\left(\frac{\Phi^{-1}(u_2) - \rho\Phi^{-1}(u_1)}{\sqrt{1-\rho^2}}\right).$$

Letting $R = \frac{\Phi^{-1}(u_2)-\rho\Phi^{-1}(u_1)}{\sqrt{1-\rho^2}}$ and differentiating this expression once more with respect to u_1 we get

$$C_{11}(u_1,u_2) = \frac{-\rho}{\sqrt{1-\rho^2}\phi(\Phi^{-1}(u_1))}\phi(R).$$

Thus for a Gaussian copula model $C^{Gauss}(F_1(x_1), F_2(x_2))$ with arbitrary margins(!) F_1 and F_2, the local Gaussian correlation is given by (3.10) and reduces to

$$\rho(x_1,x_2) = \frac{\rho}{\sqrt{1-\rho^2+\rho^2}} = \rho. \tag{3.19}$$

This is of course valid for all (x_1, x_2), not only on a curve $F_1(x_1) = F_2(x_2)$, and it shows that a constant local Gaussian correlation is a feature of the Gaussian copula, i.e. more general than for the bivariate Gaussian distribution. (It is in fact a consequence of the invariance noted for (3.16)). Note that the local mean and local variance are not in general constant for non-Gaussian marginals. It remains to prove the converse statement that $\rho(x) = c$, $(-1 < c < 1,\ c \neq 0)$ implies the Gaussian copula.

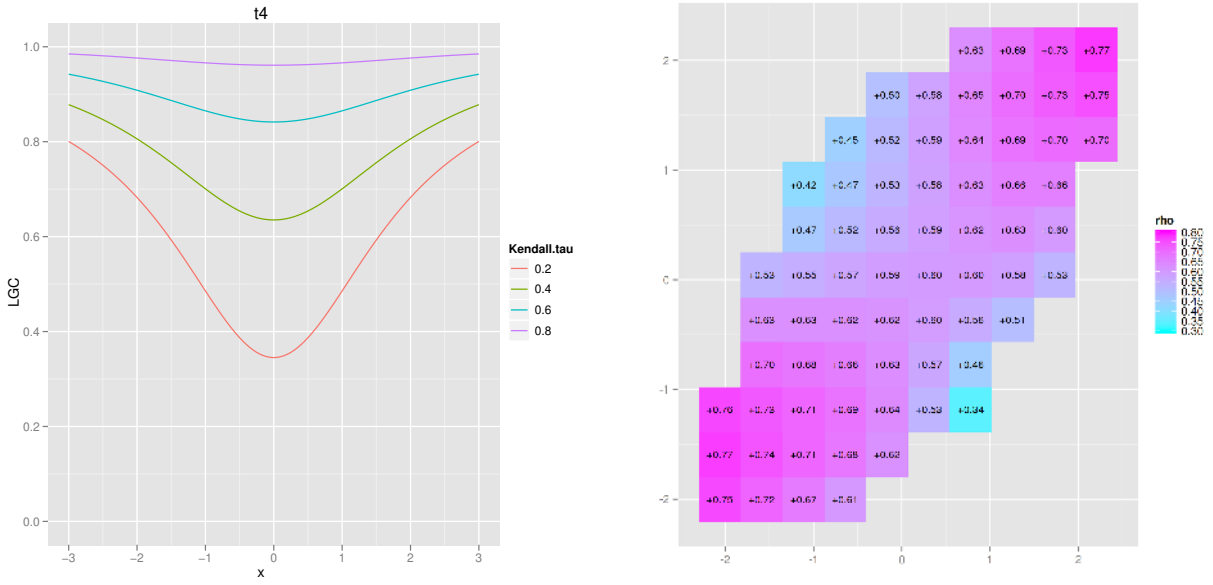
Example 3.5 (T-copula)**.** In the case that (X_1, X_2) is t-distributed with ν degrees of freedom and correlation coefficient ρ, we have that $X_1|X_2 = x_2$ is t-distributed with $\nu + 1$ degrees of freedom, expected value ρx_2 and variance $\left(\frac{\nu+x_2^2}{\nu+1}\right)(1-\rho^2)$. With t_ν as the standard t-distribution function, a similar argument as for the Gaussian copula leads to

$$C_1(u_1,u_2) = t_{\nu+1}\left(\frac{t_\nu^{-1}(u_2) - \rho t_\nu^{-1}(u_1)}{\sqrt{\frac{(\nu+t_\nu^{-1}(u_1)^2)(1-\rho^2)}{\nu+1}}}\right) \doteq t_{\nu+1}(R),$$

and with f_{t_v} as the standard t-density function and $d = \sqrt{\frac{(\nu+t_\nu^{-1}(u_1)^2)(1-\rho^2)}{\nu+1}}$

$$C_{11}(u_1,u_2) = \frac{\partial R}{\partial u_1} f_{t_{\nu+1}}(R) = \frac{-f_{t_{\nu+1}}(R)}{f_{t_\nu}(t_\nu^{-1}(u_1))d^2}\left(\rho d + \frac{1-\rho^2}{\nu+1}t_\nu^{-1}(u_1)R\right),$$

No simple formula for $\rho_\theta(z_1, z_2)$ comes as a result of this. In Figure 4 a) we see that $\rho_\theta(s, s)$ increases towards each tail which is consistent with the t-copula having both upper and lower tail-dependence.



(a) (b)

Figure 4: Local Gaussian correlation for $C(\Phi(z_1), \Phi(z_2))$ where C is the Student t-copula: (a) along the diagonal $z_1 = z_2$; (b) Estimated based on $n = 1000$ observations with $\tau = 0.4$.

4 Recognizing copulas

Given iid observations $X_1, \ldots, X_n$ from $F(x_1, x_2) = C(F_1(x_1), F_2(x_1))$ consider the issue of using local Gaussian correlation to test the null hypothesis

$$H_0 : C \in \mathcal{C}, \quad \mathcal{C} = \{C_\theta : \theta \in \Theta\}, \tag{4.1}$$

where Θ is the parameter space. This null hypothesis should not be confused with the compound null hypothesis $H_0 \cap H_0^{'}$, where $H_0^{'}$ is the additional parametric assumption

$$H_0^{'} : F_1 \in \mathcal{F}_1, \ F_2 \in \mathcal{F}_2.$$

that is, the margins also belong to some parametric class. In classical goodness-of-fit testing for copulas, $H_0^{'}$ is avoided by considering (uniform) pseudo-observations. In our approach we avoid the additional assumption $H_0^{'}$ by considering a rank-based transformation of the observations to Gaussian pseudo-observations.

4.1 Uniform pseudo-observations

Since we only consider one-parameter copulas in this paper we emphasize that the copula parameter θ is a scalar. When estimating θ under H_0, a full maximum likelihood approach or the "Inference Functions for Margins" (IFM) approach (see Joe, 1997) requires the additional assumption $H_0^{'}$. This assumption can be avoided by replacing F_j in the likelihood by the empirical distribution function (X_{ij} being the ith observation of X_j, $j = 1, 2$.)

$$F_{n,j}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(X_{ij} \leq x), \quad j = 1, 2. \tag{4.2}$$

This method is denoted the pseudo-likelihood (Demarta and McNeil, 2005) or the canonical maximum likelihood (Romano, 2002). To avoid that the copula density blows up at the boundary of $[0, 1]^2$ one typically base the pseudo-likelihood estimation on the scaled ranks $\hat{U}_1 = (\hat{U}_{11}, \hat{U}_{12}) \ldots, \hat{U}_n = (\hat{U}_{n1}, \hat{U}_{n2})$ where $\hat{U}_{ij} = nF_{n,j}(X_{ij})/(n+1)$. This transformation can be seen as an empirical version of the marginal probability integral transformation given by $U_{ij} = F_j(X_{ij})$. Since the distribution function of $U = (F_1(X_1), F_2(X_2))$ is the very definition

of the copula C, one can interpret $\hat{U}_1, \ldots, \hat{U}_n$ approximately as a sample from the underlying copula C. The observations $\hat{U}_1, \ldots, \hat{U}_n$ are therefore often referred to as (uniform) pseudo-observations. By using the pseudo-observations one could also estimate the copula parameter using the relation to Kendall's tau given by equation (3.17). For other rank-based estimators see Tsukahara (2005) and Chen et al. (2006).

4.2 Gaussian pseudo-observations

Most copula goodness-of-fit tests are carried out using the pseudo-observations described above to avoid the additional assumption $H_0^{'}$ (See e.g. Genest et al. (2009) for an overview). In our approach we first obtain "Gaussian pseudo-observations" $\hat{Z}_1 = (\hat{Z}_{11}, \hat{Z}_{12}) \ldots, \hat{Z}_n = (\hat{Z}_{n1}, \hat{Z}_{n2})$ by applying Φ^{-1} to the pseudo-observations $\hat{U}_1, \ldots, \hat{U}_n$, that is

$$\hat{Z}_{ij} = \Phi^{-1}\left(nF_{n,j}(X_{ij})/(n+1)\right) = \Phi^{-1}(\hat{U}_{ij}). \tag{4.3}$$

In the same sense that the pseudo-observations $\hat{U}_1, \ldots, \hat{U}_n$ can be interpreted as a sample from the underlying copula C, the Gaussian pseudo-observations $\hat{Z}_1, \ldots, \hat{Z}_n$ can be interpreted as a sample from $C(\Phi(z_1), \Phi(z_2))$ for which we have defined the canonical local Gaussian correlation $\rho_\theta(z_1, z_2)$ by (3.16). A non-parametric estimate of $\rho_\theta(z_1, z_2)$ can then be obtained by replacing $X_1, \ldots, X_n$ with $\hat{Z}_1, \ldots, \hat{Z}_n$ in the local likelihood described in Section 2. As before we denote this estimate $\rho_{n,b}(z_1, z_2)$ for a point (z_1, z_2). From the results of Section 3 and the fact that $\hat{Z}_1, \ldots, \hat{Z}_n$ approaches a real sample from $C(\Phi(z_1), \Phi(z_2))$ as $n \to \infty$ it follows that $\rho_{n,b}(z_1, z_2)$ is consistent with estimating $\rho_\theta(z_1, z_2)$ along the line $z_1 = z_2$ for an exchangeable copula C. A parametric estimate $\rho_{\theta_n}(z_1, z_2)$ of $\rho_\theta(z_1, z_2)$ can be obtained by replacing θ in (3.16) with θ_n, where $\theta_n = \theta_n(\hat{U}_1, \ldots, \hat{U}_n)$ is an estimate of the copula parameter under H_0 based on the pseudo-observations $\hat{U}_1, \ldots, \hat{U}_n$. Note that both $\rho_{n,b}(\cdot)$ and $\rho_{\theta_n}(\cdot)$ are based on the pseudo-observations, and thus (for any fixed bandwidth b) only depends on the ranks of the original observations. This means that a copula goodness-of-fit test based on $\rho_{n,b}(\cdot)$ and $\rho_{\theta_n}(\cdot)$ does not require any additional assumption about the marginal distributions F_1 and F_2 other than continuity.

4.3 A Goodness-of-fit test based on local Gaussian correlation

Having established a non-parametric and parametric estimate of the canonical local Gaussian correlation, we propose to base a goodness-of-fit test on the process

$$P_n(\cdot) = \rho_{n,b}(\cdot) - \rho_{\theta_n}(\cdot). \tag{4.4}$$

(An alternative would have been to smooth ρ_{θ_n} with a kernel operator as in Härdle and Mammen (1993)). Recall that $\rho_{n,b}(\cdot)$ is only consistent with $\rho_\theta(\cdot)$ along the curve defined by $z_1 = z_2$. We therefore aggregate P_n^2 along the diagonal by

$$T_n = \int_{z_{\alpha/2}}^{z_{1-\alpha/2}} P_n(t,t)^2 \, \mathrm{d}t, \tag{4.5}$$

where $z_{\alpha/2}$ and $z_{1-\alpha/2}$ are the $\alpha/2$ and $1-\alpha/2$ quantiles of the standard normal distribution (we typically use $\alpha = 0.05$). Here large values of T_n lead to the rejection of H_0. There has been much work on the asymptotic theory of test functionals such as T_n in a goodness-of-fit context, see e.g. Härdle and Mammen (1993) and Gao et al. (2009), but we do not pursue this theory here. There are several reasons for this. By the construction of T_n it is clear that its asymptotic distribution (when scaled properly by some function $\delta(n,b)$) in general depends on the underlying copula and the parameter θ, which in turns means that critical values are difficult to tabulate by means of the asymptotic properties. Moreover, it is known (see e.g. Härdle and Mammen (1993), Gao et al. (2009), Terasvirta et al. (2010) chapter 7.7) that in general the asymptotics of functional tests like T_n are not very accurate and much better results are obtained by bootstrapping. In this paper we therefore directly adopt the parametric bootstrap proposed by Genest et al. (2009) (see also Stute et al. (1993)) to obtain approximate P-values. The bootstrap procedure is as follows:

Parametric bootstrap

1. Convert the observations $X_1, \ldots, X_n$ into pseudo-observations $\hat{U}_1, \ldots, \hat{U}_n$.
2. Estimate θ under H_0 by a pseudo-observation-based estimator $\theta_n = \theta_n(\hat{U}_1, \ldots, \hat{U}_n)$. Obtain $\rho_{\theta_n}(\cdot)$ by replacing θ in (3.16) with θ_n.
3. Convert the pseudo-observations into Gaussian pseudo observations $\hat{Z}_1 =$

$\Phi^{-1}(\hat{U}_1), \ldots, \hat{Z}_n = \Phi^{-1}(\hat{U}_n)$. Obtain $\rho_{n,b}(\cdot)$ by local likelihood using the Gaussian pseudo-observations $\hat{Z}_1, \ldots, \hat{Z}_n$ as observations.

4. Compute the value of T_n.

5. For some large integer R, repeat the following steps for every $k \in \{1, \ldots, R\}$:

   (a) Generate a random sample $U_{1k}^*, \ldots, U_{nk}^*$ from the copula C_{θ_n}, and compute the associated pseudo-observations $\hat{U}_{1k}^*, \ldots, \hat{U}_{nk}^*$.

   (b) Compute $T_{n,k}^*$ by repeating step 2-4 for the new pseudo-observations $\hat{U}_{1k}^*, \ldots, \hat{U}_{nk}^*$.

The P-value for this test can then be approximated by $R^{-1} \sum_{k=1}^{R} \mathbf{1}(T_{n,k}^* > T_n)$.

In the Monte Carlo study in Section 4.5 we only consider one-parameter copulas so we have chosen to estimate θ by $\theta_n = m^{-1}(\hat{\tau})$ where $\hat{\tau}$ is the sample Kendall's tau and m is defined by (3.17). A general framework for the validity of the parametric bootstrap can be found in Genest et al. (2009). It is clear, however, that the process $P_n(\cdot)$ given by (4.4) does not fall into the category of processes considered there due to the bandwidth parameter b involved in the estimation of $\rho_{n,b}(\cdot)$. To establish a theoretical framework for the validity in our case would require a separate investigation, possibly based on a decomposition of $P_n(\cdot)$ into $\rho_{n,b}(\cdot) - \rho_\theta(\cdot) - (\rho_{\theta_n}(\cdot) - \rho_\theta(\cdot))$. Still, the results of the Monte Carlo study in Section 4.5 for the level of the test indicate validity for our approach. See also Berg (2009), where the parametric bootstrap has been used successfully in a number of different approaches.

The method described above can only be used when the analytic expression (3.16) is available. It is in the examples considered in this paper, but since this is not always the case, we provide a second bootstrap procedure which does not rely on the analytical expression (3.16) but where we instead estimate $\rho_\theta(\cdot)$ by Monte Carlo approximation:

Double parametric bootstrap

1. Convert the observations $X_1, \ldots, X_n$ into pseudo-observations $\hat{U}_1, \ldots, \hat{U}_n$.

2. Estimate θ under H_0 by a pseudo-observation-based estimator $\theta_n = \theta_n(\hat{U}_1, \ldots, \hat{U}_n)$.

3. Obtain $\rho_{n,b}(\cdot)$ by local likelihood using the Gaussian pseudo-observations $\hat{Z}_1 = \Phi^{-1}(\hat{U}_1), \ldots, \hat{Z}_n = \Phi^{-1}(\hat{U}_n)$ as observations.

4. For some (preferable large) integer $m \geq n$:

   (a) Generate a random sample $V_1^*, \ldots, V_m^*$ from C_{θ_n}.

   (b) Approximate $\rho_{\theta_n}(\cdot)$ by $\rho_{m,b(m)}(\cdot)$ where $\rho_{m,b(m)}(\cdot)$ is obtained by local likelihood using the observations $\Phi^{-1}(V_1^*), \ldots, \Phi^{-1}(V_m^*)$.

   (c) Compute the corresponding value of T_n.

5. For some large integer R, repeat the following steps for every $k \in \{1, \ldots, R\}$:

   (a) Generate a random sample $U_{1k}^*, \ldots, U_{nk}^*$ from the copula C_{θ_n}, and compute the associated pseudo-observations $\hat{U}_{1k}^*, \ldots, \hat{U}_{nk}^*$.

   (b) Compute $T_{n,k}^*$ by repeating step 2-4 for the new pseudo-observations $\hat{U}_{1k}^*, \ldots, \hat{U}_{nk}^*$.

The P-value for this test can then be approximated by $R^{-1} \sum_{k=1}^{R} \mathbf{1}(T_{n,k}^* > T_n)$. Note that this procedure can be extended to functionals obtained by integrating not only on the diagonal (t, t) but to points (t_1, t_2) with $t_1 \neq t_2$.

In Genest et al. (2009) a similar bootstrap procedure is used for a number of test statistics in the context of copula goodness-of-fit testing. There it is concluded that for the double bootstrap to be efficient the number m of repetitions must be substantially larger than the sample size n (minimum $m = 2500$ when $n = 150$). In our case we can expect that even larger values of m is required since a larger m is balanced out by a smaller bandwidth $b = b(m)$. This makes the double-bootstrap computational demanding and, consequently, we only considered the one-level parametric bootstrap in the simulation study in Section 4.5.

4.4 Choice of bandwidth

When testing $H_0 : C \in \mathcal{C}$ we may choose the bandwidth such that it is optimal if H_0 is true. In general, for a variable Z with distribution function $C(\Phi(z_1), \Phi(z_2))$ the mean integrated squared error of the local likelihood estimate $\rho_{n,b}(\cdot)$ along $z_1 = z_2$ is given by

$$\mathrm{MISE}(\rho_{n,b}(\cdot)) = E\left(\int \left(\rho_{n,b}(t,t) - \rho_\theta(t,t)\right)^2 \mathrm{d}t\right) \tag{4.6}$$

where the expectation is with respect to the distribution function $F(z) = C_\theta(\Phi(z_1), \Phi(z_2))$. Since $\rho_{\theta_n}(\cdot)$ has the ordinary parametric convergence rate it converges faster than $\rho_{n,b}(\cdot)$

(under H_0). It is therefore reasonable to choose b as the minimizer of

$$\widehat{\mathrm{MISE}}(\rho_{n,b}(\cdot)) = E^* \left(\int \left(\rho_{n,b}(t,t) - \rho_{\theta_n}(t,t)\right)^2 \mathrm{d}t \right) \tag{4.7}$$

where the expectation E^* is with respect to the distribution function $F^*(z) = C_{\theta_n}(\Phi(z_1), \Phi(z_2))$ estimated under H_0. The bandwidth b is chosen as the minimizer of (4.7) which can be approximated by Monte Carlo integration.

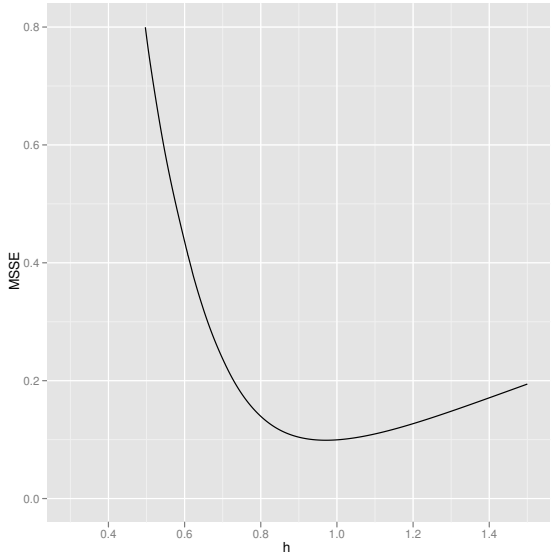


Figure 5: $\widehat{\mathrm{MISE}}(\rho_{n,b}(\cdot))$ versus b for the Clayton copula.

Table 1 reports bandwidth estimates based on minimizing (4.7). For simplicity we take $b_1 = b_2 = b$. The copula parameter and the re-sampling distribution $F^*(z) = C_{\theta_n}(\Phi(z_1), \Phi(z_2))$ is estimated from the associated pseudo-observations of a single sample from five different copulas. We consider two different sample sizes ($n = 250, 500$) and two degrees of global dependence (Kendall's tau $\tau = 0.2, 0.4$). For comparison, the minimizer of (4.6) (which can be computed knowing the real value of θ) is given in parentheses. To avoid unreliable estimates of $\rho_{n,b}(\cdot)$ in the tails we used finite integration limits in (4.6) and (4.7). We used the limit $(-1.8, 1.8)$ for $n = 250$ and $(-2, 2)$ for $n = 500$. Figure 5 displays $\widehat{\mathrm{MISE}}(\rho_{n,b}(\cdot))$ as a function of b when C is the Clayton copula.

Not surprisingly, neither (4.6) nor (4.7) has a minimum for the Gaussian copula (both

decrease as b increases). This is a result of the local Gaussian likelihood being equivalent to the global Gaussian likelihood when $b \to \infty$. However, it is not recommended to use a very large bandwidth when testing for the Gaussian copula, since too much smoothing results in poor power when the null hypothesis is false. Another alternative bandwidth algorithm is the general likelihood cross-validation method proposed in Berentsen and Tjøstheim (2014).

Copula	$\tau = 0.2$		$\tau = 0.4$	
	$n = 250$	$n = 500$	$n = 250$	$n = 500$
Clayton	0.9731(1.0536)	0.9710(0.9515)	0.8465(0.8525)	0.8116(0.7955)
Gumbel	1.1030(1.0674)	0.9198(0.9350)	1.0084(0.9602)	0.9116(0.8896)
Frank	1.4065(1.7846)	1.0166(1.0423)	0.8545(0.7824)	0.7165(0.7122)
Gaussian	∞	∞	∞	∞
t4	0.9176(0.8728)	0.8290(0.7778)	0.8786(0.7977)	0.7675(0.7055)

Table 1: Estimated bandwidth based on minimizing $\widehat{\text{MISE}}(\rho_{n,b}(\cdot))$ for a single sample from each copula model for $n = 250, 500$ and $\tau = 0.2, 0.4$. The minimizer of $\text{MISE}(\rho_{n,b}(\cdot))$ is given in the parentheses.

4.5 Simulation study

A Monte Carlo study is performed to assess the finite-sample properties of the proposed goodness-of-fit test (4.5) (based on the one-level parametric bootstrap). In order to examine its performance, we compare it with a much used test proposed by Genest and Rémillard (2008). This test is chosen because of its very good overall performance in the simulation studies of Genest et al. (2009) and Berg (2009). It stands out as one of the best. Their test is based on the empirical copula process

$$C_n(u) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(\hat{U}_{i1} \leq u_1, \hat{U}_{i2} \leq u_2), \tag{4.8}$$

where $\hat{U}_i = (\hat{U}_{i1}, \hat{U}_{i2})$ $i = 1, \ldots, n$ are the pseudo-observations and $u = (u_1, u_2) \in [0, 1]^2$. A natural test consists in comparing a distance between C_n and an estimate C_{θ_n} of C obtained

under H_0. Then a goodness-of-fit test may be based on the Cramer-von-Mises type statistic

$$A_n = n \int_{[0,1]^2} \{C_n(u) - C_{\theta_n}(u)\}^2 \mathrm{d}C_n(u). \tag{4.9}$$

Further one proceed by parametric bootstrap analogue to the first procedure described in Section 4.3 to find an approximate P-value for the test. If an analytical expression for C_θ is not available one may resort to a double parametric bootstrap analogue to the second bootstrap procedure described in Section 4.3. For a more detailed description of these test procedures we refer to Genest et al. (2009) or Berg (2009). The test based on (4.9) and similar tests are good in discriminating between copulas with different asymmetries. They are not so good in discriminating between copulas whose main difference is expressed in the tail structure, such as a Gaussian copula versus a Student t_4-copula. We believe that the reason for this is that the tail behavior in (4.9) is measured on a $[0, 1]$ scale where tail differences are compressed. This is very different if Gaussian pseudo-observations and the local correlation are used. Tail differences are expressed much more clearly and in fact lead to a very dramatic increase of discriminatory power as will be seen in the simulation experiment reported in Table 2. (Possibly the use of an analogue version of (4.9) with Gaussian pseudo-observations instead of traditional pseudo-observations may also lead to an increase of power, but we have not examined this).

In particular, we are interested in the test's ability to maintain its prescribed level (arbitrarily fixed at 5 % throughout the study) and the power against a variety of fixed alternatives. The simulation design is as follows:

- Five H_0 copulas: Clayton, Gumbel, Frank, Gaussian and Student t with 4 degrees of freedom
- Five H_1 copulas: Clayton, Gumbel, Frank, Gaussian and Student t with 4 degrees of freedom
- Two degrees of global dependence: Kendall's tau $\tau = \{0.2, 0.4\}$
- Two sample sizes: $n = \{250, 500\}$

For every combination of the above set-up 1000 samples of size n is drawn from the copula C under H_1 with dependence parameter corresponding to τ. The test statistic (4.5) and

the alternative test statistic (4.9) are then computed under H_0, and P-values are estimated using the parametric bootstrap procedure described in Section 4.3. The number of bootstrap samples was fixed at $R = 1000$. In the estimation of (4.5) the bandwidth $b = b_1 = b_2$ is taken from Table 1, except when H_0 is Gaussian in which case we put $b_1 = b_2 = 1$.

Table 2 reports the level and power of the test (4.5) and the A_n-based test (4.9) (in parenthesis). Each line of the table shows the percentage of rejections of H_0 associated with the two tests for the different combinations described above. The nominal levels match relatively well the prescribed size of 5%, but seems to be a little more volatile compared to the nominal level of the A_n-based test. We did a rerun of the level for the Student test for $n = 500$, $\tau = 0.4$ (the case where the nominal level deviated the most from the prescribed level) but based on 5000 test decisions and with $R = 2000$ bootstrap replicas for each test. For prescribed levels 1%, 5% and 10% the corresponding nominal levels where 1.36%, 6.12% and 10.84%. This indicates that 1000 bootstrap replicas may not quite be sufficient for constructing the null distribution of $P_n(\cdot)$.

The power of our proposed test is very good compared to the A_n-based test, and the power of the A_n-based test in our simulation study corresponds very well with the power found in similar studies by Genest et al. (2009) and Berg (2009). Note that for testing the Gaussian and Student hypothesis, powers are in general lower than for testing Clayton, Gumbel and Frank hypothesis. This is also in line with the previously mentioned studies.

There were only two cases where the power did not increase with the level of dependence. This happened when H_0 was the Gaussian copula and H_1 the t-copula with 4 degrees of freedom, and when H_0 was the t-copula with 4 degrees of freedom and H_1 the Gaussian copula. This can be explained by Figure 4 where we see that the local Gaussian correlation for the Student t-copula becomes more constant as the level of dependence increase and thus resembling more the Gaussian structure.

Copula under H_0	True copula	$\tau = 0.2$		$\tau = 0.4$	
		$n = 250$	$n = 500$	$n = 250$	$n = 500$
Clayton	Clayton	**5.9(4.6)**	**5.4(5.0)**	**5.3(5.2)**	**4.1(5.0)**
	Gumbel	97.3(90.0)	100.0(99.8)	100.0(100.0)	100.0(100.0)
	Frank	77.4(60.3)	95.9(88.8)	96.7(97.6)	99.6(100.0)
	Gaussian	74.8(58.2)	93.4(77.8)	97.4(95.4)	100.0(100.0)
	Student 4 df	83.3(69.2)	98.2(84.6)	99.5(97.8)	100.0(100.0)
Gumbel	Clayton	98.3(81.2)	100.0(99.0)	100.0(99.8)	100.0(100.0)
	Gumbel	**6.1(5.8)**	**6.1(5.2)**	**6.1(5.2)**	**5.2(4.8)**
	Frank	65.7(20.2)	90.7(49.0)	90.9(50.2)	99.9(92.2)
	Gaussian	55.6(11.4)	81.6(36.0)	76.2(27.8)	96.4(69.0)
	Student 4 df	43.5(20.2)	74.0(54.6)	75.1(36.6)	97.8(80.4)
Frank	Clayton	74.4(50.6)	94.0(85.6)	98.8(95.8)	100.0(100.0)
	Gumbel	56.5(40.4)	90.2(62.8)	92.3(76.0)	99.8(97.4)
	Frank	**3.1(4.8)**	**3.8(4.6)**	**4.7(4.6)**	**4.6(5.0)**
	Gaussian	11.4(8.0)	25.1(15.6)	49.7(19.2)	76.3(49.6)
	Student 4 df	84.4(27.8)	99.7(52.0)	97.2(46.0)	100.0(87.0)
Gaussian	Clayton	66.0(44.2)	91.7(73.6)	98.4(93.4)	100.0(100.0)
	Gumbel	31.1(33.2)	67.4(42.8)	56.0(58.2)	92.8(82.4)
	Frank	7.7(7.6)	14.6(7.0)	29.5(21.4)	65.6(35.4)
	Gaussian	**6.3(5.2)**	**6.2(4.8)**	**5.8(5.0)**	**6.3(5.4)**
	Student 4 df	35.0(20.6)	82.5(26.6)	14.5(21.2)	66.2(23.8)
Student 4 df	Clayton	81.8(35.8)	97.8(69.8)	98.9(88.4)	100.0(99.6)
	Gumbel	55.7(23.0)	81.9(34.0)	67.5(45.8)	95.2(63.4)
	Frank	81.6(9.2)	98.3(16.8)	94.1(26.8)	100.0(48.4)
	Gaussian	72.4(5.2)	93.8(7.8)	66.3(4.0)	94.6(2.8)
	Student 4 df	**6.5(5.4)**	**5.7(5.0)**	**6.2(4.8)**	**7.6(4.8)**

Table 2: Percentage of rejection of H_0 by the T_n-based test and the A_n-based test (in parenthesis) for data sets of different sizes arising from different copula models with dependence $\tau = 0.2$ or $\tau = 0.4$.

5 Visualizing departures from H_0

An advantage of using the local correlation is that if the null is rejected, the cause of the rejection can be investigated by visually comparing the non-parametric estimate of $\rho_\theta(\cdot)$ with the corresponding estimate under the null hypothesis. Deviations between between these estimates can easily be interpreted since both measure local dependence. Below we present two types of diagnostic plots, being based on the Gaussian pseudo-observations $\hat{Z}_1, \ldots, \hat{Z}_n$. This means that we are comparing the local dependence properties of an approximate sample from the distribution $C(\Phi(z_1), \Phi(z_2))$ with the local dependence properties of the corresponding distribution $C_\theta(\Phi(z_1), \Phi(z_2))$ under the null hypothesis. Our goal is to identify which regions of the Gaussian pseudo-observations that deviates from H_0, but also to pinpoint *how* the local dependence in these regions deviates from H_0.

Our first approach is a direct comparison between the non-parametric estimate $\rho_{n,b}(\cdot)$ based on the Gaussian pseudo-observations and the estimate under the null hypothesis $\rho_{\theta_n}(\cdot)$. One possibility is to plot $\rho_{n,b}(\cdot)$ and $\rho_{\theta_n}(\cdot)$ separately on a grid on $\mathbb{R}^2$. However, this requires Monte Carlo estimation of $\rho_{\theta_n}(\cdot)$ as in the double bootstrap procedure described in Section 4.3 since $\rho_\theta(\cdot)$ is only valid along the curve $z_1 = z_2$. Thus for a direct comparison we have opted for plotting $\rho_{n,b}(\cdot)$ and $\rho_{\theta_n}(\cdot)$ together along the curve $z_1 = z_2$ as in Figure 6 (a).

To obtain diagnostic plots on $\mathbb{R}^2$ we suggest plotting the results of "local goodness-of-fit" tests performed over a grid on $\mathbb{R}^2$, not limited to a curve. The grid is selected using the methodology of Jones and Koch (2003). First a regular grid is placed over the Gaussian pseudo-observations $\hat{Z}_1, \ldots, \hat{Z}_n$. The regular grid is then screened by selecting the grid points $(z_1, \ldots, z_p)$ satisfying $\hat{f}(z_j) \geq K$, for some constant K and a density estimator $\hat{f}(z_j) = \hat{f}(z_j | \hat{Z}_1, \ldots, \hat{Z}_n)$ (we use an ordinary kernel estimator). In this way we can ensure that the neighbourhood of each grid point contains a sufficient amount of observations for estimating $\rho_{n,b}(z_j)$. We then proceed to estimate $\rho_{n,b}(z_j)$ for each grid point and the null-distribution of $\rho_{n,b}(z_j)$ is subsequently computed using a parametric bootstrap procedure similar to the one described in Section 4.3:

Local goodness-of-fit test

1. Convert the observations $X_1, \ldots, X_n$ into pseudo-observations $\hat{U}_1, \ldots, \hat{U}_n$.

2. Estimate θ under H_0 by a pseudo-observation-based estimator $\theta_n = \theta_n(\hat{U}_1, \ldots, \hat{U}_n)$.

3. Convert the pseudo-observations into Gaussian pseudo observations $\hat{Z}_1 = \Phi^{-1}(\hat{U}_1), \ldots, \hat{Z}_n = \Phi^{-1}(\hat{U}_n)$. Obtain $\rho_{n,b}(z_j)$ by local likelihood using the Gaussian pseudo-observations $\hat{Z}_1, \ldots, \hat{Z}_n$ as observations.

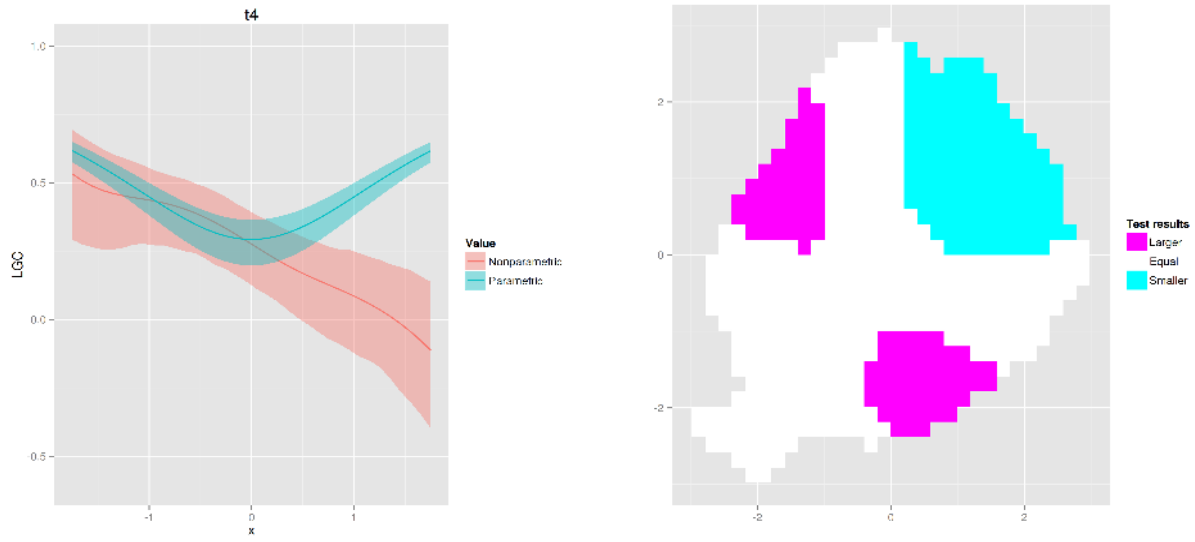
4. For some large integer R, repeat the following steps for every $k \in \{1, \ldots, R\}$:

   (a) Generate a random sample $U^*_{1k}, \ldots, U^*_{nk}$ from the copula C_{θ_n}, and compute the associated pseudo-observations $\hat{U}^*_{1k}, \ldots, \hat{U}^*_{nk}$.

   (b) Compute $\rho^{*,k}_{n,b}(z_j)$ by repeating step 2 and 3 for this sample.

The bootstrap procedure above is carried out simultaneous for all grid points and the second diagnostic plot is constructed as follows. For a given significance level α, grid points z_j with $\rho_{n,b}(z_j)$ larger than the $(1-\alpha/2)\%$ quantile of $\rho^{*,1}_{n,b}(z_j), \ldots, \rho^{*,R}_{n,b}(z_j)$ is assigned the color "magenta"; Grid points z_j with $\rho_{n,b}(z_j)$ smaller than the $\alpha/2\%$ quantile of $\rho^{*,1}_{n,b}(z_j), \ldots, \rho^{*,R}_{n,b}(z_j)$ is assigned the color "cyan"; if neither, the grid point z_j is assigned the color white. Note that this gives us only information about the regions for which $\rho_{n,b}(\cdot)$ is significantly larger ("magenta") or smaller ("cyan") in value than under H_0. To interpret what this means in terms of direction of dependence (positive or negative) one can confer with the previous suggested diagnostic plot.

Figures 6 (a)-(b) illustrates how the diagnostic plots look when the $n = 500$ data comes from $C(\Phi(x_1), \Phi(x_2))$, where C is the Clayton copula with parameter $\theta = 0.5$, but the null hypothesis is that C is the t-copula with 4 degrees of freedom. With a 5% significance level this hypothesis is rejected by the T_n-test (P-value=0.002). Figure 6 (a) clearly shows that the estimated $\rho_{n,b}(\cdot)$ is significantly smaller than that under the null hypothesis in the upper tail. For this plot we have also added standard 95% bootstrap confidence intervals. This is an indication that the estimated t-copula assigns to much dependence in the upper tail compared to the data. This is indeed confirmed by the local goodness-of-fit tests displayed in Figure 6 (b). Here we have used $\alpha = 0.05$ and $R = 1000$ bootstrap samples. Also notice that Figure 6 (b) reveals that the estimated $\rho_{n,b}(\cdot)$ is significantly larger than its estimate under the null hypothesis in points of the 2nd and 4th quadrant. This is a result of the t-copula



(a) (b)

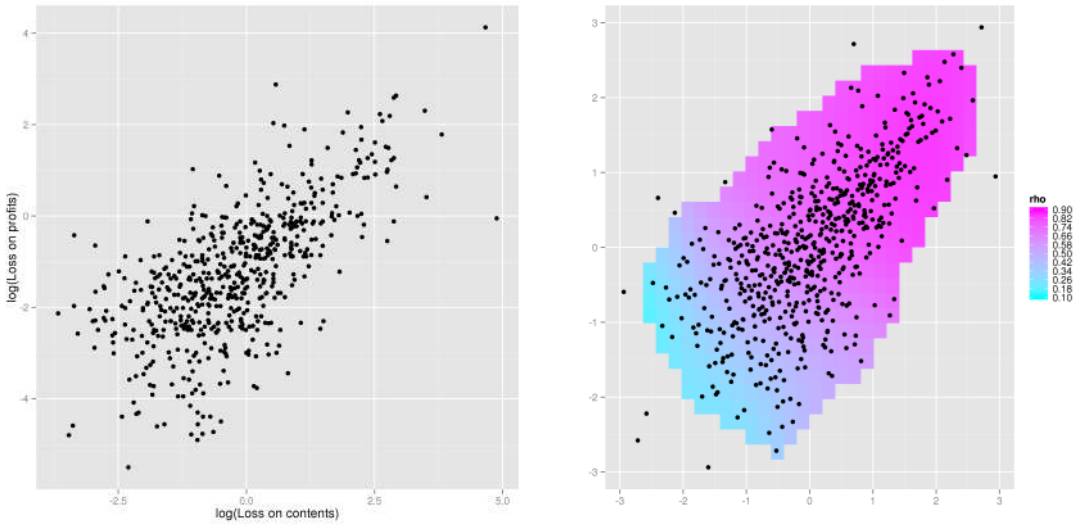
Figure 6: Evaluation of H_0: t-copula when Clayton is the true copula: (a) Diagonal plot with confidence intervals of $\rho_{n,b}$ and ρ_{θ_n} where ρ_{θ_n} is estimated under H_0: t-copula; (b) Pointwise test, H_0 : t-copula

having negative local dependence in these regions (See Berentsen and Tjøstheim (2014) and Tjøstheim and Hufthammer (2013) for a more detailed study of the t-distribution.)

6 A real data study

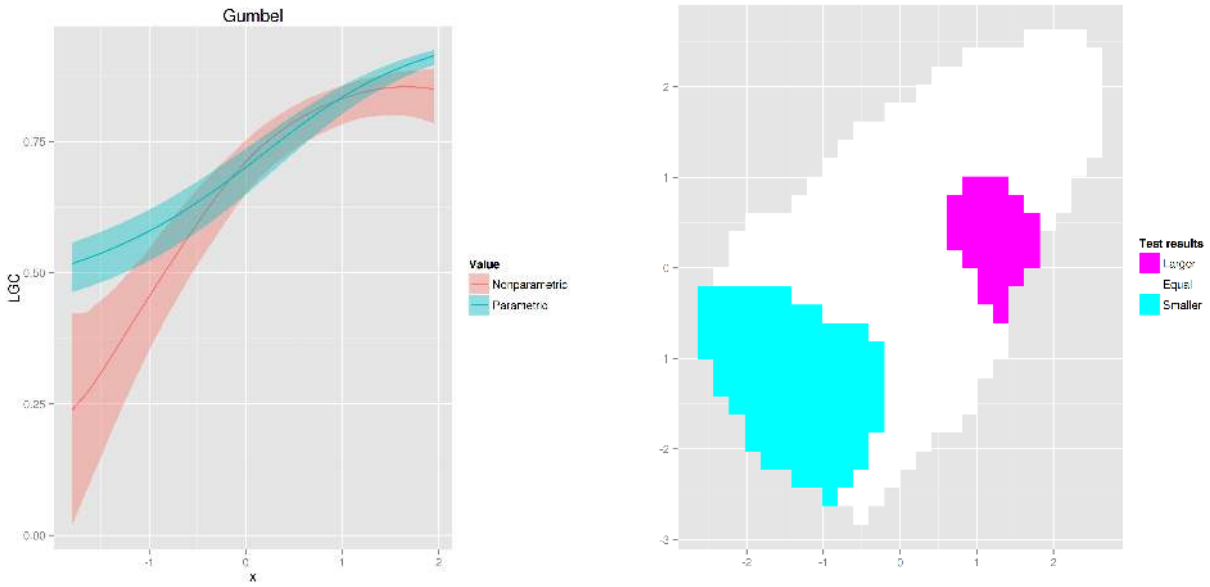
The Danish fire insurance claims data have been studied in actuarial science and extreme value theory (see e.g. McNeil, 1997). The data consist of 2167 losses over one million DKK from the years 1980 to 1990 inclusive. There were registered in total 604 cases where a loss in both contents and profits occurred, and the log-transformed values of these claims can be seen in Figure 7 (a). There were some ties present in the data, so before the analysis the observations where jittered randomly by a small amount to break the ties. The corresponding Gaussian pseudo-observation and the estimated values of $\rho_{n,b}(\cdot)$ based on the Gaussian pseudo-observations can be seen in Figure 7 (b). From the latter Figure we see that the dependence increases towards the upper tail. Indeed, in terms of AIC (Akaike Information Criteria), the best ranked copula amongst Clayton, Frank, Gumbel, Gaussian and Student-t was the Gumbel copula (-379.6), followed by the Gaussian copula (-327.9). Nevertheless, the null hypothesis that the copula of the data is the Gumbel copula was rejected by the goodness-of-fit test proposed in Section 4.3 (P-value ≈ 0) and also by the alternative test discussed in Section 4.5. However, by using the diagnostic plots proposed in Section 5 we now have the possibility to investigate the characteristics of the discrepancy between the data and the null hypothesis. In Figure 8 (a) the parametric estimate of the local Gaussian correlation is plotted together with the nonparametric estimate along the curve $z_1 = z_2$. We see that the (fitted) Gumbel copula assigns too large local correlation in the lower tail compared to the data. This is also supported by Figure 8 (b) which displays the result of the local goodness-of-fit test.

In Section 3.2.1 we investigated the differences between modeling (X_1, X_2) with upper tail dependence by the Gumbel copula with the alternative of modeling $(-X_1, -X_2)$ by the Clayton copula. The conclusion was that $\rho_\theta(\cdot)$ in the "non-dependent" tail approaches zero faster with the Clayton copula than with the Gumbel copula. From the diagnostic plots 8 (a)-(b), we see that this approach could be useful since the Gumbel copula assigns too large local correlation in the lower tail compared to the data. Indeed, the goodness-of-fit test proposed in Section 4.3 did not reject the Clayton copula for the negative observations (P-value = 0.56). Neither did the alternative test discussed in Section 4.5 ((P-value = 0.55)). This illustrates the usefulness of pinpointing the discrepancies between the null-hypothesis and the data: It can help us towards selecting a better model.



(a) (b)

Figure 7: (a) Scatter plot of log-transformed values of loss on contents and loss on profits; (b): Gaussian pseudo-observations overlain the estimated local correlation $\rho_{n,b}(\cdot)$.



(a) (b)

Figure 8: Diagnostic plots: (a) Parametric versus non-parametric estimate of $\rho_\theta(\cdot)$; (b) Local goodness-of-fit test.

7 The multivariate case

Our paper is restricted to the bivariate case. In this section we will briefly discuss extensions to the multivariate case. This is a subject of ongoing research.

7.1 Extension to the multivariate case of the local Gaussian correlation

In principle the local Gaussian approximation of a multivariate density $f(x)$, where $x = (x_1, \ldots, x_d)$, can be set up immediately as

$$\phi(x, \mu(x), \Sigma(x)) = \frac{1}{(2\pi)^{d/2}|\Sigma(x)|^{1/2}} \exp\{-\frac{1}{2}(x - \mu(x))^T \Sigma^{-1}(x)(x - \mu(x))\} \tag{7.1}$$

with $\Sigma(x)$ being the local covariance matrix. For d greater than 3-4, the local parameters of (7.1) will be impossible to estimate due to the curse of dimensionality. This means that a simplification has to be found, and in Otneim and Tjøstheim (2014, in progress) we use a simplification where $\mu_i(x) = \mu_i(x_i), \sigma_{ii}(x) = \sigma_i^2(x) = \sigma_i^2(x_i)$ and $\sigma_{ij}(x) = \sigma_{ij}(x_i, x_j)$, that is the i-th component of the local mean and the local variance are just allowed to depend on the i-th coordinate, whereas the i, j-element of the local covariance is just allowed to depend on the coordinates (x_i, x_j). In this way the property of the global multivariate Gaussian, where it is enough to compute all global means and global variances and all pairs of global correlations, is retained for the local Gaussian case. This simplification is akin to the additive approximation in a nonparametric regression analysis of the conditional mean.

In order to do statistical inference and obtain confidence intervals, the theory of Tjøstheim and Hufthammer (2013) will have to be carried over to this case. One then needs to find the asymptotic multivariate normal distribution of the estimated vector $(\hat{\mu}_1(x_1), \ldots, \hat{\mu}_d(x_d); \hat{\sigma}_1^2(x_1), \ldots, \hat{\sigma}_d^2(x_d); \hat{\sigma}_{ij}(x_{ij}),\ \ i, j = 1, ..., d,\ \ i \neq j)$, in particular the asymptotic orders of all covariances between these quantities have to be found. Both in theory and in practice the restricted model has the advantage that first the theory (and practical modeling) can be carried out for the marginals using local means and variances just depending on each marginal variable. In this first step the estimation is simplified, and estimates converge faster. In the next step the local covariance analysis can be carried out pairwise, analysing local correlations between each pair of marginals. In this step, subject

to some regularity conditions, the local means and local variances of the marginal analysis can be taken to be known and equal to the estimated values.

The analysis can be further simplified if each marginal variable is transformed to a standard normal distribution (instead of transforming to the uniform distribution as is done in standard copula theory). This transformation is a main argument of the present article, and it leads to substantial improvements in the copula goodness-of-fit test. It does lead to a local Gaussian correlation that does not depend on the distribution of the marginals, but in some situations this is an advantage. This "canonical local Gaussian correlation" is invariant to one-to one transformations of the marginals, thus giving an analogue to an important copula property. All of this can be carried over to the multivariate case, but a limit theory of the canonical local Gaussian correlation estimates should be derived.

The copula concept has been extended to the time series case, but not without difficulties. One can extend the local Gaussian correlation concept to the time series case quite straightforwardly. Both the ordinary and the canonical local autocorrelation can be used and can be employed to construct serial independence tests. Local autoregressive, AR(p), processes can be defined and along the local diagonal can be compared to additive AR models and possibly AR copula processes.

7.2 Vine copulas and local correlation

The vine copula is a very active field of copula research where, using successive conditioning, the copula modeling in d dimensions is reduced to a series of pairwise conditional copula constructions, see e.g. Aas et al. (2009). Each of these pairwise copulas can of course be modeled by a local Gaussian approximation with ordinary local correlation or canonical Gaussian correlation. In a way the local Gaussian should be especially suited to this conditional modeling task, because if one has a joint Gaussian distribution of d variables, then conditioning on any subset of d_c variables, the conditional distribution of the remaining $d - d_c$ variables given the d_c variables is still Gaussian with explicit formulas for the conditional mean and conditional covariance matrix. This can be exploited in the conditional local Gaussian distributions supposed to be describing conditional copula pairs and possibly recognizing them. The restricted local Gaussian will have to be used to avoid the curse of dimensionality.

In the vine copula constructions it is not obvious in what sequence the conditioning variables should be chosen or when to stop. Possibly the local Gaussian correlation can be of help here. One suggestion (Aas 2014, private communication) is to use local Gaussian approximation in the tail and use the tail properties to select an appropriate next stage and to decide when to stop. Currently a Vuong-type test is used for this, but it is felt that a criterion paying more attention to the tail would be more appropriate at least for variables in econometrics or finance.

Acknowledgements

This work has been partly supported by SNF (Institute for Research in Economics and Business Administration, project no. 1330). We thank a referee and the Editors for suggestions leading to an improved presentation.

References

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009), "Pair-copula constructions of multiple dependence," *Insurance: Mathematics and Economics*, 44 (2), 182–198.

Berentsen, G. D., Kleppe, T., and Tjøstheim, D. (2014), "Introducing localgauss, an R package for estimating and visualizing local Gaussian correlation," *Journal of Statistical Software*, 56(12).

Berentsen, G. D. and Tjøstheim, D. (2014), "Recognizing and visualizing departures from independence in bivariate data using local Gaussian correlation," To appear in *Statistics and Computing*.

Berg, D. (2009), "Copula goodness-of-fit testing: an overview and power comparison," *The European Journal of Finance*, 15, 675–701.

Chen, X., Fan, Y., and Tsyrennikov, V. (2006), "Efficient estimation of semiparametric multivariate copula models," *Journal of the American Statistical Association*, 101, 1228–1240.

Deheuvels, P. (1979), "La fonction de dependance empirique et ses proprietes: Un test non parametric d'independence," *Bulletin de la Classe des Sciences, 5e Series*, 274–292.

Demarta, S. and McNeil, A. J. (2005), "The t copula and related copulas," *International Statistical Review*, 73, 111–129.

Gao, J., King, M., Lu, Z., and Tjøstheim, D. (2009), "Specification testing in nonlinear and nonstationary time series autoregression," *Annals of Statistics*, 37, 3893–3928.

Genest, C. and Rémillard, B. (2008), "Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models," *Annales Henri Poincaré*, 44, 1096—-1127.

Genest, C., Remillard, B., and Beaudoin, D. (2009), "Goodness-of-fit tests for copulas: A review and a power study," *Insurance: Mathematics and Economics*, 44, 199–213.

Härdle, W. and Mammen, E. (1993), "Comparing nonparametric versus parametric regression fits," *Annals of Statistics*, 21, 1926–1947.

Hjort, N. L. and Jones, M. C. (1996), "Locally parametric nonparametric density estimation," *Annals of Statistics*, 24, 1619–1647.

Joe, H. (1997), *Multivariate models and dependence concepts*, Chapman and Hall, London.

Jones, M. C. and Koch, I. (2003), "Dependence maps: local dependence in practice," *Statistics and Computing*, 13, 241–255.

McNeil, A. J. (1997), "Estimating the tails of loss severity distributions using extreme value theory," *ASTIN bulletin*, 27, 117–132.

Nelsen, R. B. (2006), *An Introduction to Copulas*, Springer, New York.

Romano, C. (2002), "Calibrating and simulating copula functions: an application to the italian stock market," *Risk Management*, 23.

Rosenblatt, M. (1952), "Remarks on a multivariate transformation," *The Annals of Mathematical Statistics*, 23, 470–472.

Sklar, A. (1959), "Fonctions de repartition a n dimensions et leurs marges," *Publications de l'Institute de Statistique de l'Universite de Paris 8*, 229–231.

Støve, B. and Tjøstheim, D. (2014), "Measuring asymmetries in financial returns: An empirical investigation using local Gaussian correlation," To appear in Essays in Nonlinear Time Series Econometrics, Oxford University Press, N. Haldrup, M. Meitz and P. Saikkonen eds.

Støve, B., Tjøstheim, D., and Hufthammer, K. (2014), "Using local Gaussian correlation in a nonlinear re-examination of financial contagion," *Journal of Empirical Finance*, 25, 62–82.

Stute, W., Manteiga, W., and Quindimil, M. (1993), "Bootstrap based goodness-of-fit-tests," *Metrika*, 40, 243–256, 10.1007/BF02613687.

Terasvirta, T., Tjøstheim, D., and Granger, C. W. J. (2010), *Modelling Nonlinear Economic Time Series*, Oxford University Press.

Tjøstheim, D. and Hufthammer, K. (2013), "Local Gaussian correlation: A new measure of dependence," *Journal of Econometrics*, 172, 33–48.

Tsukahara, H. (2005), "Semiparametric estimation in copula models," *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 33, 357–375.