Universität Hamburg

Nonparametric copula density estimation: testing for independence and other applications

Sebastian Kiwitt & Natalie Neumeyer

Preprint–No. 2008-02
September 2008

DEPARTMENT MATHEMATIK

ARBEITSBEREICH MATHEMATISCHE STATISTIK
UND STOCHASTISCHE PROZESSE

BUNDESSTR. 55, D – 20146 HAMBURG

Nonparametric copula density estimation: testing for independence and other applications

Sebastian Kiwitt and Natalie Neumeyer*

University of Hamburg, Department of Mathematics

Bundesstrasse 55, 20146 Hamburg, Germany

September 8, 2008

Abstract

The structure of dependence between random variables can be modelled by their copula which has uniform marginal distributions. We suggest a new nonparametric estimator for a bivariate copula density, which is based on an orthogonal series expansion and has itself uniform marginals. As application we consider a new consistent asymptotically distribution-free test for independence of the components of bivariate random variables, which applies methods of order-selection tests. We deduce the asymptotic distribution and investigate the small sample performance by means of a simulation study. As further applications of the copula density estimator we discuss the estimation of bivariate densities in situations where informations about the marginals are available. All results can be generalized to the multivariate case.

AMS Classification: 62G10, 62G07

Keywords and Phrases: copula estimation, marginal distributions, test for independence, orthogonal series estimation, order-selection test, null-effect hypothesis

1 Introduction

The concept of modelling dependencies by means of copula functions as introduced by Sklar (1959) has gained much popularity over the last years; see Nelsen (2006) for an overview.

*corresponding author, e-mail: neumeyer@math.uni-hamburg.de

Nonparametric estimators for copula (distribution) functions have been considered by Deheuvels (1979), Fermanian, Radulovic and Wegkamp (2004), and Chen and Huang (2007), among others. In this paper we will propose a new nonparametric estimator $\hat{\gamma}$ for the copula density γ of bivariate random variables (X, Y). The estimator is based on an orthogonal-series and is an alternative for nonparametric copula density estimators as have been proposed by Gijbels and Mielniczuk (1990) and Sancetta and Satchel (2004). The advantage of our new method is that the estimator joins the property of copula densities that the marginals are densities of the uniform distribution on $[0, 1]$. To the authors' best knowledge no nonparametric copula density estimator considered in literature before owns this property. For orthogonal series based density estimators see Schwartz (1967), Watson (1969) and Hall (1981, 1986), for instance.

We give two main applications for the new estimator. Firstly note that the independence of X and Y is equivalent to $\gamma = I_{[0,1]^2}$ a. e., where I_A denotes the indicator function of set A. Based on this we suggest a new nonparametric test for the hypothesis

$$H_0 : X, Y \text{ are independent,} \tag{1.1}$$

which is asymptotically distribution free and consistent. The test does not involve the choice of any smoothing parameter and is similar in spirit to lack-of-fit tests in regression models based on orthogonal series as the order selection test by Eubank and Hart (1992) [see Ledwina (1994), Dette and Munk (1998), Aerts, Claeskens and Hart (1999, 2000) and Eubank (2000) on related topics]. Tests for independence based on copula estimation have been suggested by Deheuvels (1981a, 1981b) and Genest and Rémillard (2004). In a simulation study comparison the new test is shown to have better power properties. Other nonparametric tests for independence were developed by Hoeffding (1948), Blum, Kiefer and Rosenblatt (1961), Rosenblatt (1975), Zheng (1997), and Gretton and Györfi (2008), among others, whereas Fermanian (2005), Genest, Quessy and Rémillard (2006) and Scaillet (2007) propose goodness-of-fit tests for copulas.

The second application we consider is the estimation of the joint density of (X, Y). Our new copula density estimator gives alternative versions of the estimators by Spiegelman and Park (2003) for known parametric models for the marginal distributions, and Hall and Neumeyer (2006) for additional data on the marginal distributions of X or Y.

The paper is organized as follows. In section 2 we describe the orthogonal series estimator for the copula density $\hat{\gamma}$. In section 3 we develop the hypothesis test for independence based on $\hat{\gamma}$ and give its asymptotic distribution. Section 5 demonstrates small sample performance of the suggested test for independence, whereas section 4 describes applications of $\hat{\gamma}$ for estimation of bivariate distributions under additional information and briefly discusses multivariate extensions. All proofs are presented in the appendix.

2 An orthogonal series estimator for the copula density

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent and identically distributed with joint distribution $F_{X,Y}$ and continuous marginal distributions F_X and F_Y, respectively. Denote by γ the copula density (vanishing outside $[0,1]^2$), i. e. the density of $(U_i, V_i) = (F_X(X_i), F_Y(Y_i))$ $(i = 1, \ldots, n)$. As the distributions of U_i and V_i are uniform in $[0, 1]$, an estimator $\hat{\gamma}$ for γ with uniform marginals would be desirable. We assume that γ can be described as orthogonal series

$$\gamma(u, v) \;=\; \sum_{(\ell,k)\in \mathbb{N}_0^2} a_{\ell k}\Phi_{\ell k}(u, v), \quad (u, v) \in [0, 1]^2,$$

with the cosine basis functions ($(\ell, k) \in \mathbb{N}_0^2$)

$$\begin{aligned} \Phi_{00}(u, v) \;&=\; 1, \quad \Phi_{\ell 0}(u, v) \;=\; \sqrt{2}\cos(\pi \ell u), \\ \Phi_{0k}(u, v) \;&=\; \sqrt{2}\cos(\pi k v), \quad \Phi_{\ell k}(u, v) \;=\; 2\cos(\pi \ell u)\cos(\pi k v) \end{aligned}$$

and coefficients

$$a_{\ell k} \;=\; E[\Phi_{\ell k}(U_i, V_i)] \;=\; \int_{[0,1]^2} \Phi_{\ell k}(u, v)\gamma(u, v)\, d(u, v),$$

such that $\sum_{(\ell,k)\in \mathbb{N}_0^2} a_{\ell k}^2 < \infty$. The known marginal densities

$$\int \gamma(u, v)\, du = I_{[0,1]}(v), \qquad \int \gamma(u, v)\, dv = I_{[0,1]}(u)$$

give the constraints $a_{00} = 1$, $a_{\ell 0} = 0$ and $a_{0k} = 0$ for all $\ell, k \in \mathbb{N}$, and hence

$$\gamma(u, v) \;=\; 1 + \sum_{(\ell,k)\in \mathbb{N}^2} a_{\ell k}\Phi_{\ell k}(u, v), \quad (u, v) \in [0, 1]^2.$$

As estimator for γ we propose

$$\hat{\gamma}(u, v) \;=\; 1 + \sum_{\substack{(\ell,k)\in \mathbb{N}^2 \\ g(\ell,k)\le m}} \hat{a}_{\ell k}\Phi_{\ell k}(u, v), \quad (u, v) \in [0, 1]^2,$$

with estimated coefficients

$$\hat{a}_{\ell k} \;=\; \frac{1}{n}\sum_{i=1}^{n} \Phi_{\ell k}(\hat{U}_i, \hat{V}_i), \quad \text{where } \hat{U}_i = F_{X,n}(X_i),\ \hat{V}_i = F_{Y,n}(Y_i),$$

and $F_{X,n}$, $F_{Y,n}$ denote the empirical distribution functions of $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, respectively.

Now the estimator $\hat{\gamma}$ has the desired uniform marginals

$$\int \hat{\gamma}(u,v)\,du = I_{[0,1]}(v), \qquad \int \hat{\gamma}(u,v)\,dv = I_{[0,1]}(u).$$

In this paper for the function $g(\ell,k)$ we will only consider the following choices: $g(\ell,k) = \ell k$ or $g(\ell,k) = \ell + k - 1$ or $g(\ell,k) = \max(\ell,k)$. The truncation point m plays a crucial role in our testing procedure, which is discussed in the next section.

In simulations a threshold approach

$$\tilde{\hat{\gamma}}(u,v) \;=\; 1 + \sum_{\substack{(\ell,k)\in I\!N^2 \\ g(\ell,k)\le m}} \hat{a}_{\ell k}\Phi_{\ell k}(u,v) I\{|\hat{a}_{\ell k}| > \beta_n\}$$

lead to very good approximations of γ, cf. literature on wavelets, for instance Donoho, Johnstone, Kerkyacharian and Picard (1995).

3 A nonparametric test for independence

In the setting of section 2 where we observe independent copies of a bivariate random variable (X,Y), we would like to test the null hypothesis H_0 of independent compontents X and Y [see (1.1)] which is equivalent to $\gamma \equiv 1$ a. e. inside $[0,1]^2$, i. e.

$$H_0 \;:\; \gamma = I_{[0,1]^2} \text{ a. e.}$$

To derive a suitable testing procedure assume for the moment that (U_i,V_i) $(i = 1,\ldots,n)$ were observable and define

$$\tilde{\gamma}(u,v) \;=\; 1 + \sum_{\substack{(\ell,k)\in I\!N^2 \\ g(\ell,k)\le m}} \tilde{a}_{\ell k}\Phi_{\ell k}(u,v)$$

with $\tilde{a}_{\ell k} = \frac{1}{n}\sum_{i=1}^n \Phi_{\ell k}(U_i,V_i)$. Our idea is applying the two-dimensional density estimator to develop a test similar to the null-effect test by Eubank and Hart (1992) for regression functions [for the method compare also Hart (1997, chapter 7)]. Due to the orthonormal properties of the basis functions we have for the integrated squared error

$$ISE \;=\; \int (\tilde{\gamma}-\gamma)^2 \;=\; \sum_{\substack{(\ell,k)\in I\!N^2 \\ g(\ell,k)> m}} a_{\ell k}^2 + \sum_{\substack{(\ell,k)\in I\!N^2 \\ g(\ell,k)\le m}} (\tilde{a}_{\ell k} - a_{\ell k})^2$$

and for the mean integrated squared error

$$MISE \;=\; \int E[(\tilde{\gamma}-\gamma)^2] \;=\; \sum_{\substack{(\ell,k)\in I\!N^2 \\ g(\ell,k)> m}} a_{\ell k}^2 + \frac{1}{n}\sum_{\substack{(\ell,k)\in I\!N^2 \\ g(\ell,k)\le m}} (\tau_{\ell k}^2 - a_{\ell k}^2)$$

where $\tau_{\ell k}^2 = E[\Phi_{\ell k}^2(U_i, V_i)]$ and $E[\Phi_{\ell k}(U_i, V_i)\Phi_{\ell' k'}(U_i, V_i)] = 0$ for $(\ell, k) \neq (\ell', k')$. Please note that under H_0 we have $\tau_{\ell k}^2 = 1$ for all $\ell, k \in I\!N$. It would be desirable to choose m to minimize the MISE. To this end, we replace the coefficients $a_{\ell k}$ by their estimators $\hat{a}_{\ell k}$ and replace $\tau_{\ell k}$ by their value 1 under H_0. Hence, we seek to minimize

$$\sum_{\substack{(\ell,k)\in I\!N^2 \\ g(\ell,k)>m}} (\sqrt{n}\hat{a}_{\ell k})^2 + k(m) - \frac{1}{n} \sum_{\substack{(\ell,k)\in I\!N^2 \\ g(\ell,k)\leq m}} (\sqrt{n}\hat{a}_{\ell k})^2, \tag{3.1}$$

where the number of estimated coefficients is

$$k(m) = |\{(\ell, k) \in I\!N^2 \mid g(\ell, k) \leq m\}|,$$

and the last term in (3.1) is of smaller order than the first two terms and will be neglected in the following. Because the total sum $\sum_{g(\ell,k)>m} + \sum_{g(\ell,k)\leq m} = \sum_{(\ell,k)}$ does not depend on m we can solve the following maximization problem instead of the minimization. We further include a "smoothing parameter" $\lambda > 1$ and finally define as "optimal" truncation point

$$\hat{m}_n(\lambda) \;=\; \arg\sup\Big\{ \sum_{\substack{(\ell,k)\in I\!N^2 \\ g(\ell,k)\leq m}} (\sqrt{n}\hat{a}_{\ell k})^2 - \lambda k(m) \;\Big|\; m \in I\!N_0 : k(m) \leq \kappa_n \Big\}$$

(note that the function g satisfies $\min\{g(\ell, k) \mid \ell, k \in I\!N\} = 1$). We do restrict the number of estimated coefficients from n observations to be less or equal to $\kappa_n \to \infty$, where we assume that $\kappa_n = o(n^{2/3})$ for the choices $g(\ell, k) = \max(\ell, k)$, $g(\ell, k) = \ell + k - 1$ and $\kappa_n = o(n^{1/2})$ for $g(\ell, k) = \ell k$.

Remark 3.1 *Note that the assumption on the rate of κ_n is sufficient for the validity of Theorem 3.3 below, but not neccessary. Higher order Taylor expansions in the proof would lead to less restrictive conditions at the expense of a more technical, less readable proof. In the simulation results in the next section we ignored the technical condition and set $\kappa_n = n$, which worked very well.*

The asymptotic distribution of the coefficient estimators is given in the next lemma.

Lemma 3.2 *Under H_0, $\sqrt{n}\hat{a}_{\ell k}$ converges for $n \to \infty$ in distribution to a random variable $Z_{\ell k}$ with standard normal distribution.*

The truncation point $\hat{m}_n(\lambda)$ converges in distribution to $m(\lambda)$ defined as

$$m(\lambda) \;=\; \arg\sup\Big\{ \sum_{\substack{(\ell,k)\in I\!N^2 \\ g(\ell,k)\leq m}} Z_{\ell k}^2 - \lambda k(m) \;\Big|\; m \in I\!N_0 \Big\},$$

where $Z_{\ell k}$ $(\ell, k \in I\!N)$ denote iid standard normally distributed random variables.

Theorem 3.3 *Under H_0, $P(\hat{m}_n(\lambda) = 0)$ converges for $n \to \infty$ to $P(m(\lambda) = 0)$.*

The asymptotic distribution is completely known and can be evaluated from the next theorem.

Theorem 3.4 *Let z_l^2 be χ^2-distributed with l degrees of freedom ($l \in \mathbb{N}$). Then*

$$P(m(\lambda) = 0) = \exp\Big(- \sum_{m=1}^{\infty} \frac{P(z_{k(m)}^2 > \lambda k(m))}{m} \Big).$$

For testing hypothesis H_0 with asymptotic level α apply Theorem 3.4 to obtain λ_α such that $P(m(\lambda_\alpha) = 0) = 1 - \alpha$ and reject H_0 whenever $\hat{m}_n(\lambda_\alpha) > 0$. In more typical notation of hypotheses tests we have that $P(\hat{m}_n(\lambda_\alpha) > 0) = P(T_n > \lambda_\alpha)$, $P(m(\lambda_\alpha) > 0) = P(T > \lambda_\alpha) = \alpha$, where the test statistic

$$T_n = \max_{\substack{m \in \mathbb{N} \\ k(m) \leq \kappa_n}} \frac{1}{k(m)} \sum_{\substack{(\ell,k) \in \mathbb{N}^2 \\ g(\ell,k) \leq m}} (\sqrt{n} \hat{a}_{\ell k})^2 \tag{3.2}$$

under H_0 converges in distribution to $T = \max_{m \in \mathbb{N}} \frac{1}{k(m)} \sum_{\substack{(\ell,k) \in \mathbb{N}^2 \\ g(\ell,k) \leq m}} Z_{\ell k}^2$. Table 1 displays some values for λ_α for different functions g and typical levels α.

α	$g(\ell,k) = \ell k$	$g(\ell,k) = \ell + k - 1$	$g(\ell,k) = \max(\ell,k)$
0.01	$\lambda_\alpha = 6.64181$	$\lambda_\alpha = 6.64164$	$\lambda_\alpha = 6.62816$
0.05	$\lambda_\alpha = 3.95092$	$\lambda_\alpha = 3.94231$	$\lambda_\alpha = 3.86348$
0.10	$\lambda_\alpha = 2.96099$	$\lambda_\alpha = 2.93550$	$\lambda_\alpha = 2.82524$

Table 1: The critical value λ_α for different α and different g.

We finish the section by stating a consistency result.

Lemma 3.5 *The test is consistent, i. e. $\lim_{n\to\infty} P(\hat{m}_n(\lambda) = 0) = 0$ if X and Y are dependent.*

Remark 3.6 *Each choice of g results in a consistent asymptotically distribution-free test that does not involve the choice of any smoothing parameter because λ_α is completely determined by the asymptotic level α. For other symmetric, surjective functions $g : \mathbb{N}^2 \to \mathbb{N}$ that are strictly increasing in each component similar asymptotic theory can be developed analogously. As can be seen from the proof in the appendix assumptions on κ_n depend then on the rates of convergence of*

$$k(m|a,b) = \sum_{\substack{(\ell,k) \in \mathbb{N}^2 \\ g(\ell,k) \leq m}} \ell^a k^b \text{ and } \sum_{\substack{m \in \mathbb{N} \\ k(m) \leq \kappa_n}} \frac{m^c}{k(m)},$$

where $k(m) = k(m|0,0)$*. For example for* $g(\ell, k) = \max(\ell, k)$ *we have* $k(m|a,b) = m^{a+b+2}$ $(a, b \in I\!N)$ *and* $\sum_{\substack{m \in I\!N \\ k(m) \leq \kappa_n}} \frac{m^c}{k(m)} = O(\kappa_n^{(c-1)/2})$ $(c \in I\!N,\ c \geq 2)$*.*

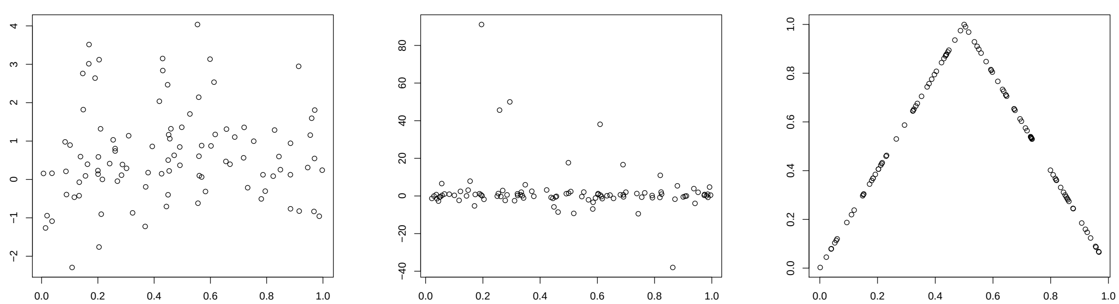
4 Small sample performance

In this section we present a simulation study for the test for independence based on the test statistic T_n with $\kappa_n = n$ and different choices of the function g as described in section 3. In the simulations we also give the results for the copula-based test for independence as proposed by Deheuvels (1981a, 1981b) [see also Genest and Rémillard (2004)]. To be more specific we simulated the Cramér-von Mises test statistic

$$D_n = \int_{[0,1]^2} \Big(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big[I\{\hat{U}_i \leq u\} - U_n(u) \Big] \Big[I\{\hat{V}_i \leq v\} - U_n(v) \Big] \Big)^2 \, d(u,v),$$

where U_n is the distribution function of the uniform distribution on $\{\frac{1}{n+1}, \frac{2}{n+1}, \ldots, \frac{n}{n+1}\}$. We further note the rejection probabilities of tests based on Kendall's tau and Pearson's rho.

In Table 2 a simple example with bivariate normal distributions with standard normal marginals and varying correlation ρ is considered. Here all tests behave very similar. The data for Table 3 were generated with the Gaussian copula combined with Cauchy-distributed marginals. The data for Tables 4 and 5 were generated as follows: Let X be generated by the uniform distribution on $[0,1]$, and let $Z = 2X$ if X is below 0.5, and $Z = -2 \cdot X + 2$ otherwise. Then, although dependent, the correlation between X and Z as well as Kendall's tau are zero. Finally adding standard Normal- or Cauchy-distributed noise to the second component yieldsthe observations (X, Y). Example data sets are shown below with and without the added noise.



The rejection probabilities in Tables 2–5 for tests with nominal level $\alpha = 0.05$ are based on 5000 simulation runs. The sample sizes vary from $n = 50$ to $n = 250$.

As expected tests based on Kendall's tau and Pearson's rho cannot detect the alternatives in Tables 4 and 5. In these examples we further see that the new test's power for the choice

Tables: Rejection probabilities of the different tests for nominal level $\alpha = 0.05$.

$\rho \mid g(\ell,k)$	ℓk	$\ell+k-1$	$\max(\ell,k)$	Pearson	Kendall	D_n
0	0.0588	0.0598	0.0596	0.0556	0.0560	0.0556
0.1	0.0948	0.0948	0.0962	0.1052	0.1034	0.0922
0.25	0.3576	0.3580	0.3640	0.4252	0.3932	0.3512
0.5	0.9196	0.9196	0.9218	0.9664	0.9480	0.9240

Table 2: The (X,Y) are bivariate normal with standard normal marginals and varying correlation ρ, the sample size is $n = 50$.

$\rho \mid g(\ell,k)$	ℓk	$\ell+k-1$	$\max(\ell,k)$	Pearson	Kendall	D_n
0	0.0590	0.0594	0.0582	0.0548	0.0542	0.0554
0.1	0.0966	0.0974	0.0998	0.0724	0.0992	0.0974
0.25	0.3470	0.3472	0.3512	0.1434	0.3758	0.3408
0.5	0.9216	0.9220	0.9240	0.4122	0.9494	0.9206

Table 3: The (X,Y) are bivariate normal with Cauchy marginals and varying correlation ρ, the sample size is $n = 50$.

$n \mid g(\ell,k)$	ℓk	$\ell+k-1$	$\max(\ell,k)$	Pearson	Kendall	D_n
50	0.1190	0.1202	0.0878	0.0518	0.0542	0.1052
100	0.3068	0.3064	0.2104	0.0480	0.0516	0.2064
250	0.3170	0.3172	0.2196	0.0350	0.0502	0.2156

Table 4: The (X,Y) are generated by adding standard normal noise to one component of random variables generated from a distribution with Kendall's τ=0 and correlation $\rho = 0$, the sample sizes are $n = 50$, 100 and 250.

$n \mid g(\ell,k)$	ℓk	$\ell+k-1$	$\max(\ell,k)$	Pearson	Kendall	D_n
50	0.0682	0.0678	0.0616	0.0346	0.0498	0.0644
100	0.1064	0.1064	0.0792	0.0302	0.0480	0.0858
250	0.8380	0.8392	0.7296	0.0464	0.0544	0.6738

Table 5: The (X,Y) are generated by adding Cauchy noise to one component of random variables generated from a distribution with Kendall's τ=0 and correlation $\rho = 0$, the sample sizes are $n = 50$, 100 and 250.

$g(\ell, k) = \max(\ell, k)$ is much less than for the other choices of g [power here is also less for Deheuvels' test]. Our observation that the choice of $g(\ell, k) = \max(\ell, k)$ in some cases leads to less power is in accordance to the observations by Claeskens (1999, p. 178–180), where for orthogonal series based bivariate regression estimation it is explained that with this choice of g the number of model parameters grows too quickly at each step, which in general leads to poor power properties.

5 More applications

Multivariate case and goodness-of fit. All the presented results can be generalized to the multivariate context, where the copula density of a random vector $(X^{(1)}, \ldots, X^{(k)})$ $(k \geq 3)$ shall be estimated and the hypothesis

$$H_0: \quad X^{(1)}, \ldots, X^{(k)} \text{ are independent}$$

is to be tested.

The parametric form of γ could be tested if this was given in form of the orthogonal series expansion.

Serial dependence. It is a future research project to generalize the results to test for serial dependence in the context of stationary time series in the same way as Genest and Rémillard (2004) considered for Deheuvel's (1981a, 1981b) test.

Density estimation under information on the marginals. In the remainder of the section we consider a different application of the new copula density estimator in more detail. Let again $(X_1, Y_1), \ldots, (X_n, Y_n)$ be iid paired data with distribution function $F_{X,Y}$, density $f_{X,Y}$ and continuous marginal distributions F_X, F_Y. We consider semi- resp. non-parametric estimators $\hat{f}_{X,Y}$ for $f_{X,Y}$ that can incorporate additional marginal information in two cases.

In the first case, **(i)**, parametric models for the marginal distributions are assumed: $F_X \in \{F_{X,\vartheta} \mid \vartheta \in \Theta\}$, $F_Y \in \{F_{Y,\psi} \mid \psi \in \Psi\}$.

In the second case, **(ii)**, there are additional (not paired) observations on the marginals available: $X_{n+1}, \ldots, X_{n+m_1}$ iid $\sim F_X$, $Y_{n+1}, \ldots, Y_{n+m_2}$ iid $\sim F_Y$.

Now let Γ denote the copula of (X_i, Y_i) and γ its density. Then

$$\begin{aligned} F_{X,Y}(x, y) &= \Gamma(F_X(x), F_Y(y)) \\ f_{X,Y}(x, y) &= f_X(x) f_Y(y)\, \gamma(F_X(x), F_Y(y)) \end{aligned}$$

and a suitable estimator is

$$\hat{f}_{X,Y}(x,y) \;=\; \hat{f}_X(x)\hat{f}_Y(y)\,\hat{\gamma}(\hat{F}_X(x),\hat{F}_Y(y)),$$

[compare Liebscher (2005) and Faugeras (2008)] where we apply the orthogonal series estimator $\hat{\gamma}$ suggested in section 2 and estimators $\hat{F}_X$, $\hat{F}_Y$ for the marginals that are parametric in the first case, (i) $\hat{F}_X = F_{X,\hat{\vartheta}}$, $\hat{F}_Y = F_{Y,\hat{\psi}}$ (with densities $f_{X,\hat{\vartheta}}$, $f_{Y,\hat{\psi}}$) and nonparametric in the second case (ii), but based on all marginal data. In the latter case for example kernel density estimators $\hat{f}_{X,n+m_1}$, $\hat{f}_{Y,n+m_2}$ based on $X_1,\dots,X_{n+m_1}$ and $Y_1,\dots,Y_{n+m_2}$, respectively, could be applied with $\hat{F}_X = \hat{F}_{X,n+m_1}$, $\hat{F}_Y = \hat{F}_{Y,n+m_2}$ their corresponding distribution function.

As $\hat{\gamma}$ has uniform marginals we obtain that $\hat{f}_{X,Y}$ has exactly the desired marginals, namely the estimators of the marginals under the additional information, which are considered to be very accurate (as they are based on parametric models or on a larger number of observations, respectively).

Our model (i) is the situation considered by Spiegelman and Park (2003), who estimated the marginal quantiles parametrically and forced the bivariate density estimator to have similar marginal quantiles. As a real data example they consider air pollution measurements on two air pollutants, obtained from Clinton drive in Houston, TX. The assumed marginals are lognormal distributions. With the new estimator we achieve Spiegelman and Park's (2003) aim to develop a bivariate density estimator, which marginals coincide with parametrically pre-estimated marginals.

Our estimator in model (ii) is the same as was considered by Hall and Neumeyer (2006), when their wavelet-based copula density estimator is replaced by the orthogonal series estimator considered here. Their theoretical result can be shown to hold under similar regularity conditions for the new estimator, i. e. that for very smooth copula densities γ one can achieve univariate convergence rates for the bivariate density estimator, cf. Kiwitt (2007). Hall and Neumeyer (2006) consider a real data example of arrival times of two air planes each day on the same route where the extra data correspond to dates where only one flight took place.

A Proofs

A.1 Proof of Lemma 3.2

For this and the following proof we define $\varphi_\ell(u) = \sqrt{2}\cos(\pi\ell u)$, such that $\Phi_{\ell k}(u,v) = \varphi_\ell(u)\varphi_k(v)$. We apply Taylor's expansion for $\varphi_\ell(\hat{U}_i)$,

$$\varphi_\ell(\hat{U}_i) \;=\; \varphi_\ell(U_i) + (\hat{U}_i - U_i)\varphi_\ell'(U_i) + R_\ell^u(U_i), \tag{A.1}$$

where for the remainder there exists some constant c (not depending on n or ℓ), such that

$$|R_\ell^u(U_i)| \;\le\; (\hat{U}_i - U_i)^2 \sup_{t\in[0,1]} |\varphi_\ell''(t)| \tag{A.2}$$

$$\leq c\ell^2 Z_u^2 \tag{A.3}$$

with a random variable $Z_u = \sup_{x\in\mathbb{R}} |\hat{F}_X(x) - F_X(x)| = O_p(n^{-1/2})$. With the analogous representation for $\varphi_k(\hat{V}_i)$ we obtain

$$\begin{aligned}
\hat{a}_{\ell k} &= \frac{1}{n}\sum_{i=1}^{n}\Big(\varphi_\ell(U_i) + (\hat{U}_i - U_i)\varphi_\ell'(U_i) + R_\ell^u(U_i)\Big)\Big(\varphi_k(V_i) + (\hat{V}_i - V_i)\varphi_k'(V_i) + R_k^v(V_i)\Big) \\
&= \tilde{a}_{\ell k} + R_n,
\end{aligned}$$

where

$$\sqrt{n}\tilde{a}_{\ell k} = n^{-1/2}\sum_{i=1}^{n}\varphi_\ell(U_i)\varphi_k(V_i)$$

converges to a standard normal distribution by the central limit theorem. To show that the remainder R_n is of order $o_p(n^{-1/2})$ one applies the independence of U_i and V_i (under H_0) and the fact that $E[\varphi_\ell(U_i)] = E[\varphi_k(V_i)] = 0$. Consider, for instance, the term

$$b_{\ell k} = \frac{1}{n}\sum_{i=1}^{n}(\hat{U}_i - U_i)\varphi_\ell'(U_i)\varphi_k(V_i) = \Upsilon_n + \Lambda_n, \tag{A.4}$$

where

$$\Upsilon_n = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}(I\{U_j \leq U_i\} - U_i)\varphi_\ell'(U_i)\varphi_k(V_i)$$

is a degenerate, mean zero U-statistic of order $O_p(n^{-1})$, and the diagonal terms

$$\Lambda_n = \frac{1}{n^2}\sum_{i=1}^{n}(1 - U_i)\varphi_\ell'(U_i)\varphi_k(V_i)$$

form a sequence of mean zero random variables of order $O_p(n^{-3/2})$ by the central limit theorem. □

A.2 Proof of Theorem 3.3

The proof is similar to the proof by Eubank and Hart (1992). Our situation is different because we consider a two-dimensional density estimator instead of an one-dimensional regression estimator, and moreover, our estimator depends not on iid-data, but on "estimated data" $(\hat{U}_i, \hat{V}_i)$ $(i = 1, \ldots, n)$.

For the test statistic T_n defined in (3.2) we have the expansion $T_n = \tilde{T}_n + \tilde{R}_n$ for

$$\tilde{T}_n = \max_{\substack{m\in\mathbb{N}\\ k(m)\leq\kappa_n}} \frac{1}{k(m)} \sum_{\substack{(\ell,k)\in\mathbb{N}^2\\ g(\ell,k)\leq m}} (\sqrt{n}\tilde{a}_{\ell k})^2 \tag{A.5}$$

based on the true unknown data (U_i, V_i) $(i = 1, \dots, n)$, and

$$\tilde{R}_n \le \max_{\substack{m\in I\!N \\ k(m)\le\kappa_n}} \frac{1}{k(m)} \Big| \sum_{\substack{(\ell,k)\in I\!N^2 \\ g(\ell,k)\le m}} \Big((\sqrt{n}\tilde{a}_{\ell k})^2 - (\sqrt{n}\tilde{a}_{\ell k})^2 \Big) \Big|. \tag{A.6}$$

We will first prove the result for $\tilde{T}_n$ and then show that $\tilde{R}_n = o_p(1)$.

I. In the following let $E_n = \{\tilde{T}_n \le \lambda\}$. Then we have

$$E_n = \{\hat{m}_n(\lambda) = 0\} = \Big\{ \max_{\substack{m>0 \\ k(m)\le\kappa_n}} \sum_{\substack{(\ell,k)\in I\!N^2 \\ 0<g(\ell,k)\le m}} (\sqrt{n}\tilde{a}_{\ell k})^2 \le \lambda k(m) \Big\} = A_n \cap B_n,$$

where

$$A_n := \{ \max_{0<m\le M_n} V_{n,m} \le \lambda \}, \quad B_n := \{ \max_{\substack{m>M_n \\ k(m)\le\kappa_n}} V_{n,m} \le \lambda \}$$

for

$$V_{n,m} = \frac{1}{k(m)} \sum_{\substack{(\ell,k)\in I\!N^2 \\ 0<g(\ell,k)\le m}} (\sqrt{n}\tilde{a}_{\ell k})^2.$$

The sequence M_n in the definitions of A_n and B_n converges to infinity for $n \to \infty$ at a rate which will be specified later. Further define V_m analogous to $V_{n,m}$, but with $\sqrt{n}\tilde{a}_{\ell k}$ replaced by their limits $Z_{\ell k}$, and let

$$\tilde{A}_n := \{ \max_{0<m\le M_n} V_m \le \lambda \}.$$

We will show in the following

$$\begin{aligned} (i) &\quad \lim_{n\to\infty} P(B_n) = 1 \\ (ii) &\quad \lim_{n\to\infty} P(A_n) - P(\tilde{A}_n) = 0 \\ (iii) &\quad \lim_{n\to\infty} P(\tilde{A}_n) = P(m(\lambda) = 0) \end{aligned}$$

from which the assertion follows.

Proof of (i). For the complement we will show that $P(B_n^c) \to 0$. We have

$$P(B_n^c) \le P\Big(\bigcup_{\substack{m\ge M_n+1 \\ k(m)\le\kappa_n}} \Big\{ \Big| \sum_{\substack{(\ell,k)\in I\!N^2 \\ 0<g(\ell,k)\le m}} \big((\sqrt{n}\tilde{a}_{\ell k})^2 - 1 \big) \Big| \frac{1}{k(m)} > \lambda - 1 \Big\} \Big). \tag{A.7}$$

We follow Eubank and Hart (1992) to show that (A.7) converges to zero. To this let end $n_j = j^2$, and define $j(1)$ as the largest integer j such that $n_j \le M_n$ and $j(2)$ the largest integer j such that $k(n_j) \le \kappa_n$. The assertion follows when (A.8) and (A.9) below can be proven for $n \to \infty$,

$$P\Big(\bigcap_{j=j(1)}^{j(2)} \Big\{\Big| \sum_{\substack{(\ell,k)\in I\!N^2\\ 0<g(\ell,k)\le n_j}} \frac{((\sqrt{n}\tilde{a}_{\ell k})^2-1)}{k(n_j)}\Big| \le \lambda - 1\Big\}\Big) \to 1 \qquad \text{(A.8)}$$

$$P\Big(\bigcap_{j=j(1)}^{j(2)} \Big\{\max_{1<r\le n_{j+1}-n_j}\Big| \sum_{\substack{(\ell,k)\in I\!N^2\\ n_j+1<g(\ell,k)\le n_j+r}} \frac{((\sqrt{n}\tilde{a}_{\ell k})^2-1)}{k(n_j)}\Big| \le \lambda - 1\Big\}\Big) \to 1. \qquad \text{(A.9)}$$

The proofs of (A.8) and (A.9) are similar to the proofs in Hart (1997, p. 170/171) and we will only discuss the first one in more detail. We have

$$\begin{aligned}
&P\Big(\bigcap_{j=j(1)}^{j(2)} \Big\{\Big| \sum_{\substack{(\ell,k)\in I\!N^2\\ 0<g(\ell,k)\le n_j}} \frac{((\sqrt{n}\tilde{a}_{\ell k})^2-1)}{k(n_j)}\Big| \le \lambda - 1\Big\}\Big)\\
&\ge 1 - \sum_{j=j(1)}^{j(2)} \frac{1}{(\lambda-1)^2 k(n_j)^2}\mathrm{Var}\Big(\sum_{\substack{(\ell,k)\in I\!N^2\\ 0<g(\ell,k)\le n_j}} (\sqrt{n}\tilde{a}_{\ell k})^2 - 1\Big)
\end{aligned}$$

and applying properties of the basis functions and independence of X and Y one obtains for example $E[\varphi_\ell(U_i)\varphi_\ell(U_\nu)\varphi_k(V_i)\varphi_k(V_\nu)] = \delta_{i,\nu}$, and similarly

$$\begin{aligned}
&\mathrm{Var}\Big(\sum_{\substack{(\ell,k)\in I\!N^2\\ 0<g(\ell,k)\le n_j}} (\sqrt{n}\tilde{a}_{\ell k})^2 - 1\Big)\\
&= \sum_{\substack{(\ell,k)\in I\!N^2\\ 0<g(\ell,k)\le n_j}} \sum_{\substack{(\ell',k')\in I\!N^2\\ 0<g(\ell',k')\le n_j}} \frac{1}{n^2}\sum_{i=1}^{n}\sum_{\nu=1}^{n}\sum_{i'=1}^{n}\sum_{\nu'=1}^{n}\Big\{E[\varphi_\ell(U_i)\varphi_\ell(U_\nu)\varphi_{\ell'}(U_{i'})\varphi_{\ell'}(U_{\nu'})]\\
&\qquad\qquad \times E[\varphi_k(V_i)\varphi_k(V_\nu)\varphi_{k'}(V_{i'})\varphi_{k'}(V_{\nu'})]\\
&\quad - \frac{1}{n}E[\varphi_\ell(U_i)\varphi_\ell(U_\nu)]E[\varphi_k(V_i)\varphi_k(V_\nu)] - \frac{1}{n}E[\varphi_{\ell'}(U_{i'})\varphi_{\ell'}(U_{\nu'})]E[\varphi_{k'}(V_{i'})\varphi_{k'}(V_{\nu'})] + \frac{1}{n^2}\Big\}\\
&= O(\frac{1}{n})(k(n_j))^2 + O(1)k(n_j).
\end{aligned}$$

Hence for some $c, c' > 0$ the probability in (A.8) is larger than

$$1 - c\sum_{j=j(1)}^{j(2)} \frac{1}{(\lambda-1)^2 k(n_j)}(1 + \frac{k(n_j)}{n}) \ \ge\ 1 - \frac{c'}{(\lambda-1)^2}\sum_{j=j(1)}^{j(2)}\frac{1}{j^2} \qquad \text{(A.10)}$$

(note that for $j \le j(2)$ one has $k(n_j) \le \kappa_n = o(n)$, and $k(n_j) \ge n_j = j^2$ for all our choices of g). Now (A.10) converges to 1 because with $M_n \to \infty$ we have that $j(1) \to \infty$.

Proof of (ii). Applying the multivariate central limit theorem for $k(M_n)$-dimensional vectors with components $\sqrt{n}\tilde{a}_{\ell k}$ one can use the Berry-Esseen theorem by Bhattacharya and

Rao (1976, p. 118) to obtain that $|P(A_n) - P(\tilde{A}_n)|$ can be bounded by $a(M_n)k(M_n)^2/n^{1/2}$ for some increasing sequence $a(M_n)$. M_n is chosen increasing to infinity so slowly that $a(M_n)k(M_n)^2/n^{1/2}$ converges to zero and $M_n \leq \kappa_n$.

Proof of (iii). This follows from the definition of the decreasing sequence $\tilde{A}_n$.

II. We finish the proof by showing that $\tilde{R}_n = o_p(1)$ [see (A.6)]. To this end, we apply the definitions and arguments given in the proof of Lemma 3.2. With (A.1) we obtain

$$\begin{aligned}\hat{a}_{\ell k}^2 - \tilde{a}_{\ell k}^2 \;=\; & \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\Big[\Big(\varphi_\ell(U_i) + (\hat{U}_i - U_i)\varphi_\ell'(U_i) + R_\ell^u(U_i)\Big) \\ & \times \Big(\varphi_\ell(U_j) + (\hat{U}_j - U_j)\varphi_\ell'(U_j) + R_\ell^u(U_j)\Big)\Big(\varphi_k(V_i) + (\hat{V}_i - V_i)\varphi_k'(V_i) + R_k^v(V_i)\Big) \\ & \times \Big(\varphi_k(V_j) + (\hat{V}_j - V_j)\varphi_k'(V_j) + R_k^v(V_j)\Big) - \varphi_\ell(U_i)\varphi_\ell(U_j)\varphi_k(V_i)\varphi_k(V_j)\Big].\end{aligned} \tag{A.11}$$

Two of the resulting terms we will consider in detail in the following. The first one is

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\hat{U}_i - U_i)\varphi_\ell'(U_i)\varphi_\ell(U_j)\varphi_k(V_i)\varphi_k(V_j) \;=\; \tilde{a}_{\ell k}b_{\ell k},$$

where $b_{\ell k}$ is defined in (A.4). We apply Cauchy-Schwarz' inequality to the sum $\sum_{(\ell,k)} n\tilde{a}_{\ell k}b_{\ell k}$ to see that

$$\max_{\substack{m\in\mathbb{N}\\ k(m)\leq\kappa_n}} \frac{1}{k(m)}\Big|\sum_{\substack{(\ell,k)\in\mathbb{N}^2\\ g(\ell,k)\leq m}} n\tilde{a}_{\ell k}b_{\ell k}\Big| \;\leq\; (\tilde{T}_n)^{1/2}\Big(\max_{\substack{m\in\mathbb{N}\\ k(m)\leq\kappa_n}} \frac{1}{k(m)}\Big|\sum_{\substack{(\ell,k)\in\mathbb{N}^2\\ g(\ell,k)\leq m}} (\sqrt{n}b_{\ell k})^2\Big|\Big)^{1/2}.$$

From the above proof it follows that $\tilde{T}_n = O_p(1)$ for $\tilde{T}_n$ from (A.5) and it remains to show that

$$\lim_{n\to\infty} P\Big(\bigcup_{\substack{m\in\mathbb{N}\\ k(m)\leq\kappa_n}}\Big\{\sum_{\substack{(\ell,k)\in\mathbb{N}^2\\ 0<g(\ell,k)\leq m}}(\sqrt{n}b_{\ell k})^2\frac{1}{k(m)} > \epsilon\Big\}\Big) \;=\; 0 \tag{A.12}$$

for all $\epsilon > 0$. Set $g(\ell,k) = \max(\ell,k)$ and, hence, $k(m) = m^2$ (the proof is similar for the two other choices of g). By Markov's inequality we bound (A.12) by

$$\lim_{n\to\infty}\sum_{\substack{m\in\mathbb{N}\\ k(m)\leq\kappa_n}}\;\sum_{\substack{(\ell,k)\in\mathbb{N}^2\\ 0<g(\ell,k)\leq m}}\;\sum_{\substack{(\ell',k')\in\mathbb{N}^2\\ 0<g(\ell',k')\leq m}}\frac{n^2E[b_{\ell k}^2 b_{\ell' k'}^2]}{k(m)^2\epsilon^2} \;\leq\; \frac{c}{\epsilon^2}\lim_{n\to\infty}\sum_{\substack{m\in\mathbb{N}\\ m^2\leq\kappa_n}}\;\sum_{\substack{\ell,k,\ell',k'\in\\ \{1,\ldots,m\}}}\frac{\ell^2\ell'^2}{m^4n^2}$$

which is obtained by a simple calculation of expectations, and where the constant c does not depend on ℓ, k or n. We hence obtain the limit

$$\frac{c}{\epsilon^2}\lim_{n\to\infty}\sum_{\substack{m\in\mathbb{N}\\ m^2\leq\kappa_n}}\frac{m^4}{n^2} = \frac{c}{\epsilon^2}\lim_{n\to\infty}\frac{\kappa_n^{5/2}}{n^2} = 0$$

because $\kappa_n = o(n^{2/3})$ by assumption.

The other terms resulting from the expansion (A.11) are treated similarly as they are two–factor products of terms $\tilde{a}_{\ell,k}$, $b_{\ell,k}$, $c_{\ell,k}$, $d_{\ell,k}$, $e_{\ell,k}$, $f_{\ell,k}$ (and similar terms reversing the roles of U and V) with definitions

$$\begin{aligned}
c_{\ell,k} &= \frac{1}{n}\sum_{i=1}^{n} R_\ell^u(U_i) R_k^v(V_i) \\
d_{\ell,k} &= \frac{1}{n}\sum_{i=1}^{n} \varphi_\ell(U_i) R_k^v(V_i) \\
e_{\ell,k} &= \frac{1}{n}\sum_{i=1}^{n} (\hat{U}_i - U_i)\varphi_\ell'(U_i)(\hat{V}_i - V_i)\varphi_k'(V_i) \\
f_{\ell,k} &= \frac{1}{n}\sum_{i=1}^{n} (\hat{U}_i - U_i)\varphi_\ell'(U_i) R_k^v(V_i).
\end{aligned}$$

For some terms the calculation of higher order moments is necessary, but this does not constitute any problem as there exist arbitrary many moments because the random variables U, V as well as the trigonometric functions are bounded. We also apply the bound given by (A.2) for the remainder terms.

The second term we would like to consider in detail is $f_{\ell,k}$, which leads to the slowest convergence rate. It is bounded by

$$\begin{aligned}
&\sup_{x\in\mathbb{R}} |\hat{F}_X(x) - F_X(x)| \sup_{u\in[0,1]} |\varphi_\ell'(u)| \; (\sup_{y\in\mathbb{R}} |\hat{F}_Y(y) - F_Y(y)|\,)^2 \sup_{v\in[0,1]} |\varphi_k''(v)| \\
&\leq c\cdot \ell k^2 Z_u Z_v^2
\end{aligned}$$

for some constant c, see (A.3), where Z_u and Z_v are of order $O_p(n^{-1/2})$. As before for some $\epsilon > 0$ we have to consider the probability

$$\begin{aligned}
&P\Big(\bigcup_{\substack{m\in\mathbb{N} \\ k(m)\le\kappa_n}} \Big\{ \sum_{\substack{(\ell,k)\in\mathbb{N}^2 \\ 0<g(\ell,k)\le m}} (\sqrt{n} f_{\ell k})^2 \frac{1}{k(m)} > \epsilon \Big\} \Big) \\
\leq\ &P\Big(\bigcup_{\substack{m\in\mathbb{N} \\ k(m)\le\kappa_n}} \Big\{ \sum_{\substack{(\ell,k)\in\mathbb{N}^2 \\ 0<g(\ell,k)\le m}} n c^2 \ell^2 k^4 Z_u^2 Z_v^4 \frac{1}{k(m)} > \epsilon \Big\} \Big).
\end{aligned}$$

We first consider the case $g(\ell,k) = \max(\ell,k)$. Then $k(m) = m^2$ and $\sum_{\substack{(\ell,k)\in\mathbb{N}^2 \\ 0<g(\ell,k)\le m}} \ell^2 k^4$ is of order m^8. This yields a probability of the form

$$P\Big(\bigcup_{\substack{m\in\mathbb{N} \\ m^2\le\kappa_n}} \Big\{ n Z_u^2 Z_v^4 m^6 > \eta \Big\} \Big) = P\Big(n Z_u^2 Z_v^4 \kappa_n^3 > \eta \Big),$$

which converges to zero because $Z_u^2 Z_v^4 = O_p(n^{-3})$ and $\kappa_n = o(n^{2/3})$ by assumption. The case $g(\ell,k) = \ell + k - 1$ is treated similarly. For the choice $g(\ell,k) = \ell k$ one obtains $k(m) = O(m \cdot \log(m))$ and $\sum_{\substack{(\ell,k)\in I\!N^2 \\ 0<g(\ell,k)\leq m}} \ell^2 k^4$ is of order m^5. This yields a probability of the form

$$P\Big(\bigcup_{\substack{m\in I\!N \\ k(m)\leq\kappa_n}} \Big\{ nZ_u^2 Z_v^4 \frac{m^4}{\log(m)} > \eta \Big\} \Big) \leq P\Big(\bigcup_{\substack{m\in I\!N \\ m\leq\kappa_n}} \Big\{ nZ_u^2 Z_v^4 \frac{m^4}{\log(m)} > \eta \Big\} \Big) \leq P\Big(nZ_u^2 Z_v^4 \kappa_n^4 > \eta \Big),$$

which converges to zero because $Z_u^2 Z_v^4 = O_p(n^{-3})$ and $\kappa_n = o(n^{1/2})$ by assumption.

□

A.3 Proof of Theorem 3.4

Similarly to Eubank and Hart (1992) we apply the theory by Spitzer (1956) to prove the asserted representation of the asymptotic distribution. To this end let $Z_1, Z_2, \ldots$ be iid standard normally distributed random variables and $\xi_\ell = Z_\ell^2 - \lambda$. Then $\xi_1, \xi_2, \ldots$ are iid and

$$P\Big(m(\lambda) = 0\Big) = P\Big(\max_{m\in I\!N_0} \sum_{\ell=1}^{k(m)} \xi_\ell = 0 \Big) = P\Big(\max_{m\in I\!N} S_m \leq 0 \Big) = P\Big(\max_{m\in I\!N} S_m^+ = 0 \Big),$$

where $S_m = \sum_{\ell=1}^{k(m)} \xi_\ell$ and $S_m^+ = \max(0, S_m)$. The sets $\{\max_{m\in\{1,\ldots,n\}} S_m^+ = 0\}$, $n \in I\!N$, are descending, and hence with continuity from above we have

$$P\Big(\max_{m\in I\!N} S_m^+ = 0 \Big) = \lim_{n\to\infty} P\Big(\max_{m\in\{1,\ldots,n\}} S_m^+ = 0 \Big) = \lim_{n\to\infty} q_n$$

in Spitzer's (1956) notation (p. 331) with $q_n = P(\max_{m\in\{1,\ldots,n\}} S_m \leq 0)$.

For the function f defined by

$$f(\xi_1, \ldots, \xi_{k(n)}) = \exp\Big(i\lambda \max_{m\in\{1,\ldots,n\}} S_m^+ \Big)$$

assumption (1.1) by Spitzer (1956) is valid, namely $E[f(\xi_1, \ldots, \xi_{k(n)})]$ is invariant under permutation of the iid components of $(\xi_1, \ldots, \xi_{k(n)})$ [compare also (3.1) in that reference]. Hence, Theorem 3.1 by Spitzer (1956) is applicable and with the same argumentation as on p. 331 in that reference we obtain

$$\lim_{n\to\infty} q_n = \exp\Big(-\sum_{m=1}^{\infty} \frac{P(S_m > 0)}{m} \Big),$$

where

$$P(S_m > 0) = P\Big(\sum_{\ell=1}^{k(m)} \xi_\ell > 0 \Big) = P\Big(\sum_{\ell=1}^{k(m)} Z_\ell^2 > \lambda k(m) \Big)$$

and $z_{k(m)}^2 := \sum_{\ell=1}^{k(m)} Z_\ell^2$ is χ^2-distributed with $k(m)$ degrees of freedom. □

A.4 Proof of Lemma 3.5

Under dependence of X and Y at least one coefficient $a_{\ell k}$ is not zero. Because $\hat{a}_{\ell k}$ consistently estimates $a_{\ell k}$, $(\sqrt{n}\hat{a}_{\ell k})^2 - \lambda$ will converge to infinity in probability, and the supremum of

$$\sum_{\substack{(\ell,k)\in\mathbb{N}^2 \\ g(\ell,k)\leq m}} \left((\sqrt{n}\hat{a}_{\ell k})^2 - \lambda \right)$$

will (for large n) not be assumed for $m = 0$, for which the sum vanishes. □

Acknowledgements. The idea to estimate the copula density by an orthogonal series is due to Peter Hall and was discussed with the second author during her stay at the Australian National University financed by the Deutsche Forschungsgemeinschaft. The authors are also grateful to Gerda Claeskens for helpful suggestions and Holger Dette and László Györfi for comments on related literature.

B References

M. Aerts, G. Claeskens and J. D. Hart (1999). *Testing the fit of a parametric function.* J. Am. Stat. Assoc. 94, 869–879.

M. Aerts, G. Claeskens and J. D. Hart (2000). *Testing lack of fit in multiple regression.* Biometrika 87, 405–424.

R. N. Bhattacharya and R. R. Rao (1976). *Normal approximation and asymptotic expansions.* Wiley, New York.

J. R. Blum, J. Kiefer, M. Rosenblatt (1961). *Distribution free tests of independence based on the sample distribution functions.* Ann. Math. Stat. 32, 485–498.

S. Chen and T. Huang (2007). *Nonparametric Estimation of Copula Functions for Dependence Modelling.* Canad. J. Statist. 35, 265–282.

G. Claeskens (1999). *Smoothing techniques and bootstrap methods for multiparameter likelihood models.* PhD-thesis, available online at http://www.econ.kuleuven.ac.be/public/ndbaf45/publicationsGC.html

P. Deheuvels (1979). *La fonction de dépendance empirique et ses propriétés: un test non paramétrique d'indépendance.* Acad. R. Belg. Bull. Cl. Sci., 5 Sér. 65, 274–292.

P. Deheuvels (1981a). *A nonparametric test for independence.* Public. de l'Institut de Statistic de l'Université de Paris 26, 29–50.

P. Deheuvels (1981b). *A Kolmogorov-Smirnov type test for independence in multivariate samples.* Rev. Roum. Math. Pures et Appl. 26, 213–226.

H. Dette and A. Munk (1998). *Validation of linear regression models.* Ann. Statist. 26, 778–800.

D. Donoho, I. M. Johnstone, G. Kerkyacharian and D. Picard (1995). *Wavelet shrinkage: asymptopia?* (with Discussion.) J. R. Statist. Soc. B 57, 301–369.

R. L. Eubank (2000). *Testing for No Effect by Cosine Series Methods.* Scand. J. Stat. 27, 747–763.

R. L. Eubank and J. D. Hart (1992). *Testing goodness-of-fit in regression via order selection criteria.* Ann. Stat. 20, 1412–1425.

O. P. Faugeras (2008). *A quantile-copula approach to conditional density estimation.* Preprint.

J. Fermanian (2005). *Goodness-of-Fit Tests for Copulas.* J. Mult. Anal. 95. 119–152.

J. Fermanian, D. Radulovic and M. Wegkamp (2004). *Weak convergence of empirical copula processes.* Bernoulli 10, 847–860.

C. Genest and B. Rémillard (2004). *Tests of independence and randomness based on the empirical copula process.* Test 13, 335–370.

C. Genest, J.-F. Quessy and B. Rémillard (2006). *Goodness-of-fit procedures for copula models based on the probability integral transformation.* Scand. J. Statist. 33, 337–366.

I. Gijbels and J. Mielniczuk (1990). *Estimating the Density of a Copula Function.* Commun. Stat. Theory Meth. 19, 445–464.

A. Gretton and L. Györfi (2008). *Consistent nonparametric tests of independence.* Preprint.

P. Hall (1981). *On Trigonometric Series Estimates of Densities.* Ann. Statist. 9, 683–685.

P. Hall (1986). *On the Rate of Convergence of Orthogonal Series Density Estimators.* J. R. Stat. Soc. B 48, 115–122.

P. Hall and N. Neumeyer (2006). *Estimating a Bivariate Density when there are Extra Data on One or Both Components.* Biometrika 93, 439–450.

J. D. Hart (1997). *Nonparametric Smoothing and Lack-of-Fit Tests.* Springer Series in Statistics.

W. Hoeffding (1948). *A nonparametric test of independence.* Ann. Math. Statist. 19, 546–557.

S. Kiwitt (2007). *Zweidimensionale Dichteschätzung mittels nichtparametrischer Copula-Schätzer.* Diploma thesis, Ruhr-University Bochum (in German).

T. Ledwina (1994). *Data-driven version of Neyman's smooth test of fit.* J. Am. Stat. Assoc. 89, 1000–1005.

E. Liebscher (2005). *Semiparametric Density Estimators using Copulas.* Commun. Statist. Theory Meth. 34, 59–71.

R. B. Nelsen (2006). *An Introduction to Copulas.* Springer Series in Statistics.

M. Rosenblatt (1975). *A quadratic measure of deviation of two-dimensional density estimates and a test of independence.* Ann. Statist. 3, 1–14.

A. Sancetta and S. Satchell (2004). *The Bernstein Copula and its Applications to Modelling and Approximations of Multivariate Distributions.* Econom. Theory 20, 535–562.

O. Scaillet (2007). *Kernel-based goodness-of-fit tests for copulas with fixed smoothing parameters.* J. Mult. Anal. 98, 533–543.

S. C. Schwartz (1967). *Estimation of Probability Density by an Orthogonal Series.* Ann. Math. Statist. 38, 1261–1265.

A. Sklar (1959). *Fonctions de répartition à n dimensions et leurs marges.* Publ. Inst. Stat. Univ. Paris 8, 229–231.

C. Spiegelman and E. S. Park (2003). *Nearly Nonparametric Multivariate Density Estimates that Incorporate Marginal Parametric Information.* Amer. Statist. 30, 183–188.

F. Spitzer (1956). *A combinatorial lemma and its application to probability theory.* Trans. Amer. Math. Soc. 82, 323–339.

G. S. Watson (1969). *Density Estimation by Orthogonal Series.* Ann. Math. Statist. 40, 1496–1498.

J. X. Zheng (1997). *A consistent specification test of independence.* J. Nonparam. Stat. 7, 297–306.