SFB 823

# A note on bootstrap approximations for the empirical copula process

Axel Bücher, Holger Dette

Nr. 32/2009

Discussion Paper

# A note on bootstrap approximations for the empirical copula process


Axel Bücher
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
e-mail: axel.buecher@ruhr-uni-bochum.de

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
email: holger.dette@ruhr-uni-bochum.de


November 30, 2009


**Abstract**

It is well known that the empirical copula process converges weakly to a centered Gaussian field. Because the covariance structure of the limiting process depends on the partial derivatives of the unknown copula several bootstrap approximations for the empirical copula process have been proposed in the literature. We present a brief review of these procedures. Because some of these procedures also require the estimation of the derivatives of the unknown copula we propose an alternative approach which circumvents this problem. Finally a simulation study is presented in order to compare the different bootstrap approximations for the empirical copula process.


## 1 Introduction

The empirical copula $C_n$ is the most famous and easiest nonparametric estimator for the copula $C$ of a random vector. It is well known that the standardized process $\sqrt{n}(C_n - C)$ converges weakly towards a Gaussian field $\mathbb{G}_C$ with covariance structure depending on the unknown copula and its derivatives, see e.g, Fermanian, Radulovic and Wegkamp (2004). Because these quantities are usually difficult to estimate several authors have suggested to approximate the limit distribution by bootstrap procedures. Fermanian et al. (2004) proposed a bootstrap procedure based on resampling and proved its consistency. A wild bootstrap approach based on the multiplier method was recently proposed by Rémillard and Scaillet (2009) and applied to the problem of testing the equality between two copulas. Recently Kojadinovic and Yan (2009) and Kojadinovic, Yan and Holmes (2009) used the same method to construct a goodness-of-fit test for the parametric form of a copula.

The present paper has two purposes. On the one hand our work is motivated by the fact that the multiplier approach proposed by Rémillard and Scaillet (2009) still requires the estimation of the partial derivatives of the unknown copula. For this reason we propose a modification of this method, which avoids this estimation problem. On the other hand we investigate the finite sample properties of the resampling bootstrap, a slightly modified version of the bootstrap proposed by Rémillard and Scaillet (2009) and the new direct multiplier bootstrap proposed in this paper. In particular it is demonstrated that despite the fact that the new multiplier method has the most attractive theoretical properties and avoids the problem of estimating derivatives, the procedure proposed in Rémillard and Scaillet (2009) yields the best results in most cases.

The remaining part of this note is organized as following. In Section 2 we summarize some basic results on empirical copulas and state the different concepts of the bootstrap. We also introduce the modified multiplier method and prove its consistency. Finally in Section 3 we present a small simulation study, which illustrates the finite sample properties of the different bootstrap approximations.

# 2 The empirical copula process and three bootstrap approximations

For the sake of brevity, we restrict ourselves to the case of bivariate copula, but all results can easily be transferred to higher dimensions. Let $X_1, \dots, X_n$ be independent identically distributed bivariate random vectors with continuous cumulative distribution function (cdf) $F$, marginal distribution functions $F_1$ and $F_2$ and copula $C$. Due to the well known theorem of Sklar [see e.g. Nelsen (1998)] there is the relationship

$$C(u_1, u_2) = F(F_1^-(u_1), F_2^-(u_2)), \tag{1}$$

where $H^-(u) = \inf\{t \in \mathbb{R} | H(t) \geq u\}$ denotes the generalized inverse of a real function $H$. The empirical copula as the simplest nonparametric estimator for $C$ [going back to Deheuvels (1979)] simply replaces the unknown terms in equation (1) by their empirical counterparts, that is

$$C_n(u) = F_n(F_{n1}^-(u_1), F_{n2}^-(u_2))$$

where

$$F_n(x) = F_n(x_1, x_2) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{X_{i1} \le x, X_{i2} \le x_2\},$$

$$F_{np}(x_p) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{X_{ip} \le x_p\} \ , p = 1, 2.$$

denote the corresponding empirical distribution functions. The asymptotic behavior of $C_n$ was studied in several papers, including Gänssler and Stute (1987), Ghoudi and Rémillard (2004) or Tsukahara (2005) among others. For the sake of completeness we will state the result in the form given in a recent paper of Fermanian, Radulovic and Wegkamp (2004). Throughout this paper $\rightsquigarrow$ denotes weak convergence in the metric space $l^\infty([0,1]^2)$ of all uniformly bounded functions on the unit square $[0,1]^2$.

**Theorem 2.1.** *If the Copula $C$ possesses continuous partial derivatives $\partial_p C$ $(p = 1, 2)$ on $[0,1]^2$, then the empirical copula process $\sqrt{n}(C_n - C)$ converges weakly towards a Gaussian field $\mathbb{G}_C$,*

$$\alpha_n = \sqrt{n}(C_n - C) \rightsquigarrow \mathbb{G}_C,$$

*where the limiting process can be represented as*

$$\mathbb{G}_C(u_1, u_2) = \mathbb{B}_C(u_1, u_2) - \partial_1 C(u_1, u_2)\mathbb{B}_C(u_1, 1) - \partial_2 C(u_1, u_2)\mathbb{B}_C(1, u_2) \tag{2}$$

*and $\mathbb{B}_C$ denotes a centered Gaussian field with covariance structure*

$$\tilde{r}(u_1, u_2, v_1, v_2) = \mathrm{Cov}(\mathbb{B}_C(u_1, u_2), \mathbb{B}_C(v_1, v_2)) = C(u_1 \wedge v_1, u_2 \wedge v_2) - C(u_1, u_2)C(v_1, v_2).$$

**Remark 2.2.** The literature provides several similar nonparametric estimators for the copula. For example, Genest et al. (1995) studied the rank-based estimator

$$\bar{C}_n(u) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{F_{n1}(X_{i1}) \le u_1, F_{n2}(X_{i2}) \le u_2\}.$$

In the latter expression the marginal edfs $F_{np}$ are often replaced by their rescaled counterparts $\hat{F}_{np} = \frac{n}{n+1}F_{np}$. Both modifications do not affect the asymptotic behavior, see Fermanian et al. (2004). See also Chen and Huang (2007) or Omelka, Gijbels and Veraverbeke (2009) for a smoothed version of this process.

The limiting Gaussian variable $\mathbb{G}_C(u_1, u_2)$ depends on the unknown copula $C$ and for this reason it is not directly applicable for statistical inference. In the following discussion we will present two known and one new bootstrap approximations for the distribution of the limiting process. We begin with the usual bootstrap based on resampling , which was proposed in Fermanian et al. (2004). To be precise let $W_n = (W_{n1}, \ldots, W_{nn})$ be multinomial distributed random vectors with success probabilities $(1/n, \ldots, 1/n)$ and set

$$C_n^{\#}(u) = F_n^{\#}(F_{n1}^{\#-}(u_1), F_{n2}^{\#-}(u_2)),$$

where

$$\begin{aligned} F_n^{\#}(x) &= \frac{1}{n}\sum_{i=1}^{n} W_{ni}\mathbb{I}\{X_{i1} \le x_1, X_{i2} \le x_2\}, \\ F_{np}^{\#}(x_p) &= \frac{1}{n}\sum_{i=1}^{n} W_{ni}\mathbb{I}\{X_{ip} \le x_p\}, \qquad p = 1, 2. \end{aligned}$$

Finally define

$$\alpha_n^{res} := \sqrt{n}(C_n^{\#} - C_n)$$

as the bootstrap process based on resampling. For a precise statement of the asymptotic properties of this process we denote by $\underset{W}{\overset{\mathbb{P}}{\rightsquigarrow}}$ weak convergence conditional on the data in probability as defined by Kosorok (2008), that is $\alpha_n^{res} \underset{W}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{G}_C$ if

$$(3) \qquad \sup_{h \in BL_1(l^\infty([0,1]^2))} |\mathbb{E}_W h(\alpha_n^{res}) - \mathbb{E}h(\mathbb{G}_C)| \xrightarrow{\mathbb{P}} 0$$

and

$$(4) \qquad \mathbb{E}_\xi h(\alpha_n^{res})^* - \mathbb{E}_\xi h(\alpha_n^{res})_* \xrightarrow{\mathbb{P}^*} 0 \quad \text{for every } h \in BL_1(\mathcal{B}_\infty(\bar{\mathbb{R}}_+^2)),$$

where

$$BL_1(l^\infty([0,1]^2)) = \big\{f : l^\infty([0,1]^2) \to \mathbb{R} \mid ||f||_\infty \le 1, |f(\beta) - f(\gamma)| \le d(\beta, \gamma)\ \forall\, \gamma, \beta \in l^\infty([0,1]^2)\big\}$$

is the class of all uniformly bounded functions that are Lipschitz continuous with constant smaller one, and $\mathbb{E}_W$ denotes the conditional expectation with respect to the weights $W_n$ given the data $X_1, \ldots X_n$. Moreover, $h(\alpha_n^{res})^*$ and $h(\alpha_n^{res})_*$ denote measurable majorants and minorants with respect to the joint data, including the weights $W_n$. The following result has been established by Fermanian et al. (2004), the proof follows along similar lines as the proof of Theorem 2.6 below.

**Theorem 2.3.** *Under the preceding notations and assumptions the bootstrap approximation $C_n^{\#}$, of the empirical copula yields a valid approximation of the limit variable $\mathbb{G}_C$ in the sense that*

$$\alpha_n^{res} = \sqrt{n}(C_n^{\#} - C_n) \underset{W}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{G}_C$$

*in $l^\infty([0,1]^2)$.*

In a recent paper Rémillard and Scaillet (2009) considered the problem of testing the equality between two copulas [see also Scaillet (2005)] and proposed a multiplier bootstrap to approximate the distribution of the limiting process $\mathbb{G}_C$. To be precise let $Z_1, \ldots, Z_n$ be independent identically distributed centered random variables with variance one, independent of the data $X_1, \ldots, X_n$, which satisfy $||Z||_{2,1} = \int_0^\infty \sqrt{P(|Z| > x)}\, dx < \infty$. Rémillard and Scaillet (2009) defined the bootstrap process

$$C_n^*(u) = \frac{1}{n} \sum_{i=1}^n Z_i \, \mathbb{I}\{F_{n1}(X_{i1}) \le u_1, F_{n2}(X_{i2}) \le u_2\}$$

and showed that $C_n^*$ approximates the Gaussian field $\mathbb{B}_C$, i.e.

$$\big(\sqrt{n}(F_n - C), \sqrt{n}(C_n^* - \bar{Z}_n C_n)\big) \rightsquigarrow (\mathbb{B}_C, \mathbb{B}_C') \tag{5}$$

in $l^\infty([0,1]^2)^2$, where $\bar{Z}_n = n^{-1} \sum_{i=1}^n Z_i$ and $\mathbb{B}_C'$ is an independent copy of $\mathbb{B}_C$. Since one is interested in an approximation of $\mathbb{G}_C$ one is able to utilize identity (2) by estimating the partial derivatives of the copula $C$. As proposed by Rémillard and Scaillet (2009) we use

$$\begin{aligned}
\widehat{\partial_1 C}(u,v) &:= \frac{C_n(u+h,v) - C_n(u-h,v)}{2h}, \\
\widehat{\partial_2 C}(u,v) &:= \frac{C_n(u,v+h) - C_n(u,v-h)}{2h},
\end{aligned}$$

where $h = n^{-1/2} \to 0$ [for a smooth version of these estimators see Scaillet (2005)]. Under continuity assumptions Rémillard and Scaillet (2009) showed that these estimates are uniformly consistent. To approximate the limiting process $\mathbb{G}_C$ set

$$\alpha_n^{pdm}(u_1,u_2) := \beta_n(u_1,u_2) - \widehat{\partial_1 C}(u_1,u_2)\beta_n(u_1,1) - \widehat{\partial_2 C}(u_1,u_2)\beta_n(1,u_2), \tag{6}$$

where the process $\beta_n$ is defined by $\beta_n = \sqrt{n}(C_n^* - \bar{Z}_n C_n)$. The upper index *pdm* in (6) denotes the fact that these authors are using estimates of the partial derivatives and a multiplier concept. By Slutskys Lemma and the continuous mapping theorem one obtains the following result.

**Theorem 2.4.** *Under the preceding notations and assumptions we have*

$$\left(\sqrt{n}(C_n - C), \alpha_n^{pdm}\right) \rightsquigarrow (\mathbb{G}_C, \mathbb{G}'_C)$$

*in* $l^\infty([0,1]^2)^2$, *i.e.* $\alpha_n^{pdm}$ *approximates the limit distribution unconditionally.*

**Remark 2.5.** In the finite sample study presented in the following section we use a slightly modified version of this bootstrap procedure. To be precise let $\xi_1, \ldots, \xi_n$ denote independent identically distributed nonnegative random variables, independent of the data $X_1, \ldots, X_n$, with expectation $\mu$ and finite variance $\tau^2 > 0$ such that

$$||\xi||_{2,1} = \int_0^\infty \sqrt{P(|\xi| > x)}\, dx < \infty.$$

We define $\bar{\xi}_n = n^{-1} \sum_{i=1}^n \xi_i$ as the mean of $\xi_1, \ldots, \xi_n$ and consider the multiplier statistics

$$\tilde{C}_n^*(u) \;=\; F_n^*(F_{n1}^-(u_1), F_{n2}^-(u_2)),$$

where

$$F_n^*(x) \;=\; \frac{1}{n} \sum_{i=1}^n \frac{\xi_i}{\bar{\xi}_n} \mathbb{I}\{X_{i1} \le x_1, X_{i2} \le x_2\}$$

If we standardize the $\xi_i$ to $Z_i = (\xi_i - \mu)\tau^{-1}$ we observe that both approaches are indeed closely related by

$$\sqrt{n}\, \frac{\mu}{\tau}\, (\tilde{C}_n^* - C_n) \approx \sqrt{n}\, \frac{\mu}{\bar{\xi}_n} (C_n^* - \bar{Z}_n C_n).$$

$\tilde{C}_n^*$ approximates the Gaussian field $\mathbb{B}_C$ conditionally on the data in the sense that

$$\tilde{\beta}_n(u_1, u_2) = \sqrt{n}\, \frac{\mu}{\tau}\, (\tilde{C}_n^*(u_1, u_2) - C_n(u_1, u_2)) \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{B}_C(u_1, u_2)$$

in $l^\infty([0,1]^2)$. Estimating the partial derivatives we can now consider a multiplier bootstrap approximation

$$\tilde{\alpha}_n^{pdm}(u_1, u_2) := \tilde{\beta}_n(u_1, u_2) - \widehat{\partial_1 C}(u_1, u_2) \tilde{\beta}_n(u_1, 1) - \widehat{\partial_2 C}(u_1, u_2) \tilde{\beta}_n(1, u_2) \tag{7}$$

similar to the one in (6) that yields a conditional approximation of $\mathbb{G}_C$.

The final resampling concept considered in this section is new and combines both approaches in order to avoid the estimation of the derivatives. On the one hand it makes use of multipliers and on the other hand it is also based on identity (1) and the functional delta method. To be precise we consider multipliers as defined as in Remark 2.5 and define the statistic

$$C_n^+(u) \;=\; F_n^*(F_{n1}^{*-}(u_1), F_{n2}^{*-}(u_2)),$$

where

$$\begin{aligned} F_n^*(x) \;&=\; \frac{1}{n}\sum_{i=1}^n \frac{\xi_i}{\bar{\xi}_n}\mathbb{I}\{X_{i1} \le x_1, X_{i2} \le x_2\}, \\ F_{np}^*(x_p) \;&=\; \frac{1}{n}\sum_{i=1}^n \frac{\xi_i}{\bar{\xi}_n}\mathbb{I}\{X_{ip} \le x_p\}. \end{aligned}$$

As before set

$$\alpha_n^{dm} = \sqrt{n}\,\frac{\mu}{\tau}\,(C_n^+ - C_n),$$

We call this bootstrap the direct multiplier method, which is reflected by the the superscript $dm$ in its definition. The following result shows that the process $\alpha_n^{dm}$ yields a consistent bootstrap approximation of the empirical copula process. Note that this approach avoids the estimation of the partial derivatives of the copula.

**Theorem 2.6.** *Under the preceding notations and assumptions we have*

$$\alpha_n^{dm} = \sqrt{n}\,\frac{\mu}{\tau}(C_n^+ - C_n) \overset{\mathbb{P}}{\underset{\xi}{\rightsquigarrow}} \mathbb{G}_C$$

*in* $l^\infty([0,1]^2)$.

**Proof.** First note that it is sufficient to consider only the case of independent identically distributed random vectors with $\mathcal{U}[0,1]$-marginals and copula $C$. Indeed, let $\mathcal{U}_1, \ldots, \mathcal{U}_n$ be independent identically distributed random vectors with cdf $C$ and set

$$\begin{aligned} G_n(x) \;&=\; \frac{1}{n}\sum_{i=1}^n \mathbb{I}\{\mathcal{U}_{i1} \le x, \mathcal{U}_{i2} \le x_2\}, \\ G_{np}(x_p) \;&=\; \frac{1}{n}\sum_{i=1}^n \mathbb{I}\{\mathcal{U}_{ip} \le x_p\}\;, \qquad p = 1, 2. \end{aligned}$$

Clearly,

$$\begin{aligned} F_n(x) \;&\overset{\mathcal{D}}{=}\; G_n(F_1(x_1), F_2(x_2)), \\ F_{np}(x_p) \;&\overset{\mathcal{D}}{=}\; G_{np}(F_p(x_p))\;, \qquad p = 1, 2 \end{aligned}$$

and from the definition of the generalized inverse we conclude

$$F_{np}^{-}(u_p) \stackrel{\mathcal{D}}{=} F_p^{-}(G_{np}^{-}(u_p)) \ , \qquad p = 1, 2,$$

so that

$$C_n(u) \stackrel{\mathcal{D}}{=} G_n(G_{n1}^{-}(u_1), G_{n2}^{-}(u_2))$$

as asserted. An analogue result holds for $C_n^{+}$ and for this reasoning we may assume in the following that $X_1, \ldots, X_n$ are independent identically distributed distributed according to the cdf $C$. Next, note that Theorem 2.6 in Kosorok (2008) yields

$$\begin{aligned} \sqrt{n}(F_n - C) &\rightsquigarrow \mathbb{B}_C, \\ \sqrt{n}\,\frac{\mu}{\tau}\,(F_n^{*} - F_n) &\underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{B}_C. \end{aligned}$$

For a distribution function $H$ on $[0,1]^2$ let $H_1(x_1) = H(x_1, \infty)$ and $H_2(x_2) = H(\infty, x_2)$, denote the marginal distributions, then the mapping

$$\Phi : H \mapsto H(H_1^{-}, H_2^{-}) \tag{8}$$

is Hadamard differentiable [see Lemma 2 in Fermanian et al. (2009)]. Moreover, $C_n = \Phi(F_n)$ and $C = \Phi(C)$, and consequently the functional delta method [see Kosorok (2008)] yields

$$\sqrt{n}\,\frac{\mu}{\tau}\,(C_n^{*} - C_n) = \sqrt{n}\,\frac{\mu}{\tau}\,(\Phi(F_n^{*}) - \Phi(F_n)) \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} \Phi_C'(\mathbb{B}_C) = \mathbb{G}_C,$$

where the derivative of the map $\Phi$ at the point $C$ is given by

$$\Phi_C'(H)(u_1, u_2) = H(u_1, u_2) - \partial_1 C(u_1, u_2) H(u_1, \infty) - \partial_2 C(u_1, u_2) H(\infty, u_2).$$

This proves the assertion of Theorem 2.6. $\square$

# 3 Finite sample properties

In this section we present a small comparison of the finite sample properties of the three bootstrap approximations given in the previous section. For the sake of brevity we consider the Clayton copula with parameter $\theta = 1$ (corresponding to Kendall's-$\tau = 1/3$), but other copulas yield similar results. The sample size in our study is either $n = 100$ or $n = 200$.

In our first example we show a comparison of the different resampling methods studying their covariances. More precisely, we chose four points $\left\{\left(\frac{i}{3}, \frac{j}{3}\right), i, j = 1, 2\right\}$ in the unit square and show in the first row of Table 1 and 2 the true covariances of the limiting process. The second rows in the two tables show the simulated covariances of the the process $\sqrt{n}(C_n - C)$ on the basis of $10^6$ simulation runs (note that this distribution cannot be used in applications because the "true" copula is usually not known). We observe a rather good approximation of the covariances of the limiting process by the empirical copula process $\alpha_n$. Rows 3 - 5 of Table 1 and 2 show the covariances obtained by the bootstrap approximation. These covariances are based on the average of 1000 simulation runs, where in each run the covariance is estimated on the basis of $B = 1000$ bootstrap replications. The corresponding results for the mean squared errors are shown in Table 3 and 4. The multipliers for the partial derivative and the direct multiplier bootstrap are simulated from two-point distributions with variance 1. We have also investigated other multipliers but it turns out that the two-point distributions with variance 1 yield the best results (the other results are not presented here for the sake of brevity).

The results of Table 1 - 4 show that the partial derivative multiplier method yields the best approximations in almost all cases, despite the fact that it requires the estimation of the partial derivatives of the copula. The advantages of this approach are particulary visible in the estimation of the variances. The approximations based on the resampling bootstrap and the multiplier bootstrap are similar but less accurate than the results obtained by the partial derivative method. It is also worthwhile to mention that the *pdm*-method needs slightly more computational time to simulate a bootstrap sample, since it requires evaluation of the multiplier process in the boundary-points.

In our second example we investigate the approximation of the 90% and 95% quantile of the Kolmogorov-Smirnov statistic

$$K_n = \sup_{x \in [0,1]^2} |f_n(x)| \tag{9}$$

and the Crámer van Mises statistic

$$L_n = \int_{[0,1]^2} f_n^2(x)\, dx. \tag{10}$$

The corresponding results are presented in Table 5, where the first and fifth row show the quantiles of the "true" process $f_n = \alpha_n$, which are calculated by $10^6$ simulation runs. For the bootstrap methods the quantiles are estimated by 1000 simulation runs with $B = 1000$ Bootstrap-replications in each scenario. We observe again that the partial derivatives multiplier method yields the best approximation of the quantiles, while the resampling bootstrap and the direct multiplier bootstrap usually give too large quantiles, in particular for sample size $n = 100$. A similar observation for

| | | (1/3,1/3) | (1/3,2/3) | (2/3,1/3) | (2/3,2/3) |
|---|---|---|---|---|---|
| True | (1/3,1/3) | 0.0486 | 0.0202 | 0.0202 | 0.0100 |
| | (1/3,2/3) | | 0.0338 | 0.0093 | 0.0185 |
| | (2/3,1/3) | | | 0.0338 | 0.0185 |
| | (2/3,2/3) | | | | 0.0508 |
| $\alpha_n$ | (1/3,1/3) | 0.0489 | 0.0198 | 0.0198 | 0.0097 |
| | (1/3,2/3) | | 0.0334 | 0.0089 | 0.0181 |
| | (2/3,1/3) | | | 0.0333 | 0.0180 |
| | (2/3,2/3) | | | | 0.0510 |
| $\alpha_n^{pdm}$ | (1/3,1/3) | 0.0527 | 0.0205 | 0.0205 | 0.0093 |
| | (1/3,2/3) | | 0.0361 | 0.0092 | 0.0188 |
| | (2/3,1/3) | | | 0.0360 | 0.0188 |
| | (2/3,2/3) | | | | 0.0554 |
| $\alpha_n^{res}$ | (1/3,1/3) | 0.0619 | 0.0244 | 0.0236 | 0.0094 |
| | (1/3,2/3) | | 0.0460 | 0.0091 | 0.0211 |
| | (2/3,1/3) | | | 0.0450 | 0.0208 |
| | (2/3,2/3) | | | | 0.0694 |
| $\alpha_n^{dm}$ | (1/3,1/3) | 0.0627 | 0.0251 | 0.0248 | 0.0112 |
| | (1/3,2/3) | | 0.0456 | 0.0119 | 0.0213 |
| | (2/3,1/3) | | | 0.0451 | 0.0233 |
| | (2/3,2/3) | | | | 0.0711 |

Table 1: *Sample covariances for the Clayton Copula with $\theta = 1$ and sample size $n = 100$. The first and second rows show the true covariances and the covariances of the empirical copula process, while rows 3 - 5 show the corresponding results for the bootstrap approximations.*

the partial multiplier derivative and the resampling method has been made by Scaillet (2005) in the context of testing hypothesis regarding the copula.

On the basis of the results presented in this study and further simulations (which are not shown for the sake of brevity) we conclude our investigation with the statement that, despite the fact that the partial derivatives multiplier bootstrap requires the estimation of the partial derivatives, it outperforms the resampling and the direct multiplier bootstrap.

**Acknowledgements.** This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation (DFG).

| | | (1/3,1/3) | (1/3,2/3) | (2/3,1/3) | (2/3,2/3) |
|---|---|---|---|---|---|
| True | (1/3,1/3) | 0.0486 | 0.0202 | 0.0202 | 0.0100 |
| | (1/3,2/3) | | 0.0338 | 0.0093 | 0.0185 |
| | (2/3,1/3) | | | 0.0338 | 0.0185 |
| | (2/3,2/3) | | | | 0.0508 |
| $\alpha_n$ | (1/3,1/3) | 0.0492 | 0.0203 | 0.0203 | 0.0100 |
| | (1/3,2/3) | | 0.0339 | 0.0093 | 0.0185 |
| | (2/3,1/3) | | | 0.0339 | 0.0185 |
| | (2/3,2/3) | | | | 0.0508 |
| $\alpha_n^{pdm}$ | (1/3,1/3) | 0.0513 | 0.0203 | 0.0201 | 0.0092 |
| | (1/3,2/3) | | 0.0356 | 0.0087 | 0.0184 |
| | (2/3,1/3) | | | 0.0355 | 0.0185 |
| | (2/3,2/3) | | | | 0.0537 |
| $\alpha_n^{res}$ | (1/3,1/3) | 0.0583 | 0.0228 | 0.0228 | 0.0098 |
| | (1/3,2/3) | | 0.0413 | 0.0092 | 0.0199 |
| | (2/3,1/3) | | | 0.0417 | 0.0202 |
| | (2/3,2/3) | | | | 0.0609 |
| $\alpha_n^{dm}$ | (1/3,1/3) | 0.0577 | 0.0226 | 0.0227 | 0.0104 |
| | (1/3,2/3) | | 0.0408 | 0.0103 | 0.0210 |
| | (2/3,1/3) | | | 0.0412 | 0.0213 |
| | (2/3,2/3) | | | | 0.0634 |

Table 2: *Sample covariances for the Clayton Copula with $\theta = 1$ and sample size $n = 200$. The first and second rows show the true covariances and the covariances of the empirical copula process, while rows 3 - 5 show the corresponding results for the bootstrap approximations.*

| | | (1/3,1/3) | (1/3,2/3) | (2/3,1/3) | (2/3,2/3) |
|---|---|---|---|---|---|
| $\alpha_n^{pdm}$ | (1/3,1/3) | 0.8887 | 0.5210 | 0.5222 | 0.3716 |
| | (1/3,2/3) | | 1.0112 | 0.1799 | 0.2988 |
| | (2/3,1/3) | | | 0.9899 | 0.2818 |
| | (2/3,2/3) | | | | 0.6250 |
| $\alpha_n^{res}$ | (1/3,1/3) | 2.2612 | 0.6640 | 0.5424 | 0.3447 |
| | (1/3,2/3) | | 2.3702 | 0.1781 | 0.3554 |
| | (2/3,1/3) | | | 2.1336 | 0.3554 |
| | (2/3,2/3) | | | | 3.9469 |
| $\alpha_n^{dm}$ | (1/3,1/3) | 2.6734 | 0.7566 | 0.7067 | 0.3037 |
| | (1/3,2/3) | | 2.3636 | 0.2461 | 0.5189 |
| | (2/3,1/3) | | | 2.2544 | 0.5324 |
| | (2/3,2/3) | | | | 4.6142 |

Table 3: *Mean squared error (multiplied with $10^5$) of the different estimates for the covariance. The underlying copula is the Clayton copula with $\theta = 1$ and the sample size is $n = 100$.*

| | | (1/3,1/3) | (1/3,2/3) | (2/3,1/3) | (2/3,2/3) |
|---|---|---|---|---|---|
| $\alpha_n^{pdm}$ | (1/3,1/3) | 0.4595 | 0.2673 | 0.2798 | 0.1961 |
| | (1/3,2/3) | | 0.5211 | 0.1069 | 0.1577 |
| | (2/3,1/3) | | | 0.5092 | 0.1681 |
| | (2/3,2/3) | | | | 0.2992 |
| $\alpha_n^{res}$ | (1/3,1/3) | 1.3820 | 0.3476 | 0.3715 | 0.2102 |
| | (1/3,2/3) | | 1.0414 | 0.1133 | 0.1940 |
| | (2/3,1/3) | | | 1.2112 | 0.1993 |
| | (2/3,2/3) | | | | 1.614 |
| $\alpha_n^{dm}$ | (1/3,1/3) | 1.2682 | 0.3602 | 0.3471 | 0.2083 |
| | (1/3,2/3) | | 1.0394 | 0.1101 | 0.2484 |
| | (2/3,1/3) | | | 1.0544 | 0.2642 |
| | (2/3,2/3) | | | | 1.9483 |

Table 4: *Mean squared error (multiplied with $10^5$) of the different estimates for the covariance. The underlying copula is the Clayton copula with $\theta = 1$ and the sample size is $n = 200$.*

| $n$ | $f_n$ | 90% ($L^2$) | 95% ($L^2$) | 90% ($KS$) | 95% ($KS$) |
|---|---|---|---|---|---|
| 100 | $\alpha_n$ | 0.04593 | 0.05722 | 0.59254 | 0.65000 |
| | $\alpha_n^{pdm}$ | 0.04870 | 0.06086 | 0.62042 | 0.68611 |
| | $\alpha_n^{res}$ | 0.07060 | 0.08700 | 0.80000 | 0.80000 |
| | $\alpha_n^{dm}$ | 0.07402 | 0.09241 | 0.76154 | 0.83721 |
| 200 | $\alpha_n$ | 0.04544 | 0.05660 | 0.58925 | 0.64829 |
| | $\alpha_n^{pdm}$ | 0.04715 | 0.05867 | 0.61236 | 0.67528 |
| | $\alpha_n^{res}$ | 0.06030 | 0.07425 | 0.70711 | 0.77782 |
| | $\alpha_n^{dm}$ | 0.06066 | 0.07507 | 0.70192 | 0.77030 |

Table 5: *Sample quantiles of the Crámer van Mises statistic (10) and the Kolmogorov-Smirnov statistic (9) for the Clayton copula with parameter $\theta = 1$.*

## References


Chen, S. X. and Huang, T.-M. (2007). Nonparametric estimation of copula functions for dependence modelling. Canadian Journal of Statistics 35, 265-282.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. Bulletin de la Classe des Sciences, Académie Royale de Belgique 65, 274 - 292.

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

Gänssler, P. and Stute, W. (1987). Seminar on Empirical Processes. Birkhäuser, Basel.

Ghoudi, K. and Rémillard, B. (2004). Empirical processes based on pseudo-observations II: the multivariate case. In *Asymptotic Methods in Stochastics: Festschrift for Miklós Csörgö* (L. Horváth and B. Szyszkowicz, eds.), Fields Institute Communications 44, American Mathematical Society, 381-406.

Kojadinovic, I. and Yan, J. (2009). A goodness-of-fit test for multivariate copulas based on multiplier central limit theorems. Statistics and Computing (to appear).

Kojadinovic, I., Yan, J. and Homes, M. (2009). Fast large sample goodness-of-fit tests for copulas, Statistica Sinica (to appear).

Kosorok, M. R. (2008). Introduction to Empirical Processes and Semiparametric Inference. Springer Series in Statistics, New York.

Nelsen, R. B. (1998). An introduction to copulas. Springer-Verlag, New York.

Omelka, M., Gijbels, I. and Veraverbeke, N. (2009). Improved kernel estimation of copulas: Weak convergence and goodness-of-fit testing. Annals of Statistics 37, Number 5B, 3023-3058.

Rémillard, B. and Scaillet, O. (2009). Testing for equality between two copulas. J. Multiv. Anal. 100, 377 - 386 .

Scaillet, O. (2005). A Kolmogorov-Smirnov type test for positive quadrant dependence. Canadian Journal of Statistics 33, 415-427.

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