

Submitted to *Bernoulli*

arXiv: 1710.11504

# Goodness-of-Fit Testing for Copulas: A Distribution-Free Approach

SAMI UMUT CAN<sup>1,\*</sup> JOHN H. J. EINMAHL<sup>2</sup> and ROGER J. A. LAEVEN<sup>1,\*\*</sup>

<sup>1</sup>*Department of Quantitative Economics, University of Amsterdam, P.O. Box 15867, 1001 NJ Amsterdam, The Netherlands.*

*E-mail:* \* [s.u.can@uva.nl](mailto:s.u.can@uva.nl); \*\* [r.j.a.laeven@uva.nl](mailto:r.j.a.laeven@uva.nl)

<sup>2</sup>*Department of Econometrics & OR and CentER, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands.*

*E-mail:* [j.h.j.einmahl@tilburguniversity.edu](mailto:j.h.j.einmahl@tilburguniversity.edu)

Consider a random sample from a continuous multivariate distribution function  $F$  with copula  $C$ . In order to test the null hypothesis that  $C$  belongs to a certain parametric family, we construct an empirical process on the unit hypercube that converges weakly to a standard Wiener process under the null hypothesis. This process can therefore serve as a ‘tests generator’ for asymptotically distribution-free goodness-of-fit testing of copula families. We also prove maximal sensitivity of this process to contiguous alternatives. Finally, we demonstrate through a Monte Carlo simulation study that our approach has excellent finite-sample performance, and we illustrate its applicability with a data analysis.

*MSC 2010 subject classifications:* Primary 62H15, 62G10, 62F03; secondary 62G20, 62H10.

*Keywords:* Copula, goodness-of-fit, distribution-free, semi-parametric estimation, Monte Carlo simulation.

## 1. Introduction

Consider a  $d$ -variate ( $d \geq 2$ ) distribution function (df)  $F$  with continuous margins  $F_1, \dots, F_d$ . By the representation theorem of Sklar (1959), there is a unique df  $C$  on the unit hypercube  $[0, 1]^d$  with uniform margins such that

$$F(\mathbf{x}) = C(F_1(x_1), \dots, F_d(x_d)), \quad \mathbf{x} = (x_1, \dots, x_d)^\top \in \mathbb{R}^d. \quad (1)$$

In fact, if  $\mathbf{X} = (X_1, \dots, X_d)^\top$  is a random vector with joint df  $F$ , then it is easily seen that the unique df  $C$  satisfying (1) is the joint df of the component-wise probability integral transforms  $(F_1(X_1), \dots, F_d(X_d))^\top$ , given by

$$C(\mathbf{u}) = F(Q_1(u_1), \dots, Q_d(u_d)), \quad \mathbf{u} = (u_1, \dots, u_d)^\top \in [0, 1]^d, \quad (2)$$

with  $Q_j$  denoting the left-continuous quantile function associated with  $F_j$ , i.e.  $Q_j(\cdot) = \inf\{x \in \mathbb{R} : F_j(x) \geq \cdot\}$ , for  $j = 1, \dots, d$ .

The df  $C$  satisfying (1) or (2) is called the *copula* associated with  $F$ , and it is a representation of the dependence structure between the margins of  $F$ , since  $C$  contains no information about the margins, yet together with the margins it characterizes  $F$ . Thus copulas allow separate modeling of margins and dependence structure in multivariate settings, which has proved to be a useful approach in a wide range of applied fields, from medicine and climate research to finance and insurance. We refer to recent comprehensive monographs such as Nelsen (2006), Joe (2015) and Durante and Sempi (2016) for more background on copula theory and its various applications.

The present work is concerned with goodness-of-fit (GOF) testing for copulas. More specifically, we assume that an i.i.d. sample

$$\mathbf{X}_1 = (X_{11}, \dots, X_{1d})^\top, \dots, \mathbf{X}_n = (X_{n1}, \dots, X_{nd})^\top$$

is observed from an unknown  $d$ -variate df  $F$  with continuous margins  $F_1, \dots, F_d$  and copula  $C$  as above. We are interested in testing the hypothesis  $C \in \mathcal{C}$  against the alternative  $C \notin \mathcal{C}$ , where  $\mathcal{C} = \{C_\lambda : \lambda \in \Lambda\}$  denotes a parametric family of copulas, indexed by a finite-dimensional parameter  $\lambda$ . There is a rich cornucopia of parametric copula families used in various applications (see, e.g., Ch. 4 of Joe (2015) or Ch. 6 of Durante and Sempi (2016) for extensive lists), and new ones are introduced regularly in the literature, so the testing problem just described is clearly very relevant for practitioners making use of copula modeling in their work.

GOF testing for copulas is not a new problem, and several approaches have been proposed in the literature since the early 1990s, each with their advantages and limitations in specific situations. A partial list includes Genest and Rivest (1993), Shih (1998), Wang and Wells (2000), Breymann, Dias and Embrechts (2003), Fermanian (2005), Genest, Quessy and Rémillard (2006) and Dobrić and Schmid (2007). There seems to be no single approach that is universally preferred over others. For a broad overview and comparison of various GOF testing procedures for copulas, we refer to Berg (2009), Genest, Rémillard and Beaudoin (2009) and Fermanian (2013). As pointed out in the latter papers, a common problem with many GOF approaches is that the asymptotic distribution of the test statistic under the null hypothesis  $C \in \mathcal{C}$  depends on the particular family  $\mathcal{C}$  that is tested for, as well as on the unknown true value of the parameter  $\lambda$ . In other words, many of the proposed GOF tests in the literature are not asymptotically *distribution-free*. As a result, the asymptotic distribution of the test statistics under the null hypothesis cannot be tabulated for universal use, and approximate  $p$ -values have to be computed for each model via, e.g., specialized bootstrap procedures such as the ones outlined in Genest, Rémillard and Beaudoin (2009), App. A-D.

In this paper, we develop an approach to construct asymptotically distribution-free GOF tests for any parametric copula family satisfying some rather mild smoothness assumptions. We do not propose a particular test statistic, but instead construct a whole test *process* on the unit hypercube  $[0, 1]^d$ , which converges weakly to a standard  $d$ -variate Wiener process under the null hypothesis. Thus GOF tests can be conducted by comparing the observed path of this test process with the statistical behavior of a standard Wiener process. Various functionals of the test process can be used for this

comparison, such as the absolute maximum over  $[0, 1]^d$  or integral functionals. Since the weak limit of the test process is a standard process independent of the family  $\mathcal{C}$  or the true value of  $\boldsymbol{\lambda}$ , the limiting distributions of these functionals will also be independent of  $\mathcal{C}$  and  $\boldsymbol{\lambda}$ , and they only need to be tabulated once for use in all testing problems. In practice, this is an important advantage. Our results can also be used to test the GOF of fully specified copula models rather than parametric families.

We also show that our approach is optimal, in the sense that the obtained test process does not “lose any information” asymptotically. More precisely: when considering a sequence of contiguous alternatives approaching the null model, the distance in variation between the limiting processes under the null and the alternatives is as large as the limiting distance in variation of the data themselves. As a consequence, for a given sequence of contiguous alternatives, we can find a functional of our test process that yields an asymptotically optimal test and hence outperforms (or matches) competing procedures. Such a test retains the distribution-freeness advantage and thus avoids resampling procedures. Naturally, a variety of high power omnibus tests can be constructed as indicated above; see also Section 6.

Our approach relies on parametric estimation of the marginal distributions  $F_1, \dots, F_d$ . That is, we assume that there is a parametric family of univariate dfs  $\mathcal{F} = \{F_\theta : \theta \in \Theta\}$  such that  $F_j \in \mathcal{F}$  for  $j = 1, \dots, d$ . In fact, this requirement can be relaxed to  $F_j \in \mathcal{F}_j$  for  $j = 1, \dots, d$ , where the parametric families  $\mathcal{F}_j$  may be different, but we will stick with  $F_j \in \mathcal{F}$  for notational simplicity. The parametric structure of the margins might be naturally provided by the specifics of the data-generating process, or it might follow from theoretical considerations such as limit theorems, or it might be assumed based on independent empirical analysis or expert judgement. Alternatively, the family  $\mathcal{F}$  can be chosen to be rich enough to contain or reasonably approximate all distributions relevant for the particular problem at hand. In this paper, we take the parametric structure of the margins as given, and focus on inference about the unknown copula. In the recent work [Lu and Zheng \(2020\)](#), the approach of this paper is extended to accommodate dynamic parametric marginal distributions, such as those provided by GARCH and ARMA processes.

The remainder of this paper is structured as follows. In Section 2, we introduce two estimators for the copula  $C$ , a parametric one that works under the null hypothesis  $C \in \mathcal{C}$  and a semi-parametric one that works in general. We consider the normalized difference  $\eta_n$  between these estimators and determine its weak limit  $\eta$  under the null hypothesis, as the sample size  $n$  tends to infinity. We will see that the distribution of  $\eta$  depends on the family  $\mathcal{C}$ , as well as on the true values of the parameter  $\boldsymbol{\lambda}$  and the marginal parameters, so  $\eta_n$  cannot be used as a basis for distribution-free testing. The crucial step in our approach is introduced in Section 3, where we describe a transformation that turns  $\eta$  into a standard Wiener process on  $[0, 1]^d$ . In Section 4, we apply an empirical version of this transformation to  $\eta_n$ , and show that the resulting process  $W_n$  converges weakly to a standard Wiener process on  $[0, 1]^d$  under the null hypothesis. This is our first main result, and the transformed process  $W_n$  is the test process that was alluded to above. In Section 5, we investigate the behavior of the test process  $W_n$  under a sequence of contiguous alternatives, and we show that transforming the raw data into  $W_n$  does not

lead to any loss of information asymptotically. This is our second main result. In Section 6, we present some simulation results that demonstrate the finite-sample behavior of some functionals of  $W_n$  under the null and alternative hypotheses, and we then apply our approach to a real-world data set and analyze the results. Section 7 and an online appendix contain all the proofs.

## 2. Comparing two copula estimators

As in the Introduction, we assume that  $\mathbf{X}_i = (X_{i1}, \dots, X_{id})^\top$ ,  $i \in \{1, \dots, n\}$ , are i.i.d. random vectors with common df  $F$ , which has continuous marginal dfs  $F_1, \dots, F_d$  and copula  $C$ . We further assume throughout that the marginal dfs are members of some parametric family of univariate dfs,  $\mathcal{F} = \{F_\theta : \theta \in \Theta\}$ , indexed by  $\theta = (\theta_1, \dots, \theta_m)^\top \in \Theta$ , where  $\Theta$  is some open subset of  $\mathbb{R}^m$ . This means that there exist  $\theta_1, \dots, \theta_d \in \Theta$  such that  $F_j = F_{\theta_j}$  for  $j \in \{1, \dots, d\}$ . Our ultimate aim is to test the hypothesis  $C \in \mathcal{C} = \{C_\lambda : \lambda \in \Lambda\}$ , where the parameter  $\lambda = (\lambda_1, \dots, \lambda_p)^\top$  takes values in some open subset  $\Lambda$  of  $\mathbb{R}^p$ . Throughout Sections 2–4, we will assume that the null hypothesis holds, i.e. there is a  $\lambda_0 \in \Lambda$  such that  $C = C_{\lambda_0}$ .

There are various ways of estimating the copula from i.i.d. data, depending on the assumptions one is willing to make about the underlying model, as well as the requirements one chooses to impose on the estimator, such as smoothness. Perhaps the most straightforward and well-known copula estimator is the non-parametric *empirical copula* discussed in [Ruyngaert \(1973\)](#), [Rüschendorf \(1976\)](#), [Deheuvels \(1979, 1981\)](#), [Gaenssler and Stute \(1987\)](#), [Fermanian, Radulović and Wegkamp \(2004\)](#) and [Segers \(2012\)](#), among others. Other commonly used approaches for copula estimation include two-step methods where the first step involves (non-parametric or parametric) estimation of the margins and the second step estimates the copula parametrically from marginal data transformed in accordance with the first step. Such estimators are studied in, e.g., [Genest, Ghoudi and Rivest \(1995\)](#) and [Shih and Louis \(1995\)](#). A broad overview of various copula estimation methods can be found in [Charpentier, Fermanian and Scaillet \(2007\)](#), [Choroś, Ibragimov and Permiakova \(2010\)](#) and Ch. 5 of [Joe \(2015\)](#).

In this paper, we will make use of two estimators for  $C$ : a parametric estimator  $C_{\hat{\lambda}}$  and a semi-parametric estimator  $\hat{C}$ . We do not specify the estimator  $\hat{\lambda}$ , but require it to satisfy a rather non-restrictive convergence assumption to be stated below. The semi-parametric estimator  $\hat{C}$  is defined as

$$\hat{C}(\mathbf{u}) = F_n(Q_{\hat{\theta}_1}(u_1), \dots, Q_{\hat{\theta}_d}(u_d)), \quad \mathbf{u} \in [0, 1]^d,$$

where  $\hat{\theta}_1, \dots, \hat{\theta}_d$  denote appropriate estimators for  $\theta_1, \dots, \theta_d$ ,  $Q_\theta$  denotes the quantile function associated with  $F_\theta$ , and  $F_n$  denotes the  $d$ -variate empirical df generated by the sample  $\mathbf{X}_1, \dots, \mathbf{X}_n$ :

$$F_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{\mathbf{X}_i \leq \mathbf{x}\}, \quad \mathbf{x} \in \mathbb{R}^d.$$

Here,  $\mathbf{X}_i \leq \mathbf{x}$  is short-hand notation for “ $X_{ij} \leq x_j$  for all  $j = 1, \dots, d$ ”. Note that in view of the representation (2) of the copula  $C$  and our parametric assumption on the marginal dfs of  $F$ , the estimator  $\widehat{C}$  is a natural one; cf. Segers (2015), Example 3.4.

Under the null hypothesis, both  $\widehat{C}$  and  $C_{\widehat{\lambda}}$  estimate the true copula  $C$ , while only  $\widehat{C}$  correctly estimates  $C$  when the null hypothesis does not hold. Thus the asymptotic discrepancy between the two estimators provides a natural starting point for a GOF test. With that in mind, we define

$$\eta_n(\mathbf{u}) = \sqrt{n}[\widehat{C}(\mathbf{u}) - C_{\widehat{\lambda}}(\mathbf{u})], \quad \mathbf{u} \in [0, 1]^d. \quad (3)$$

Our first result will be a theorem describing the asymptotic behavior of  $\eta_n$ , but first we establish some notation and state the necessary assumptions about the various estimators introduced above, as well as about the parametric families  $\mathcal{C}$  and  $\mathcal{F}$ .

Let  $C_n$  denote the empirical df generated by the (unobserved) copula sample  $(F_1(X_{i1}), \dots, F_d(X_{id}))^\top$ ,  $i \in \{1, \dots, n\}$ . That is,

$$C_n(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{F_1(X_{i1}) \leq u_1, \dots, F_d(X_{id}) \leq u_d\}, \quad \mathbf{u} \in [0, 1]^d.$$

Note that we can then write

$$\widehat{C}(\mathbf{u}) = C_n(F_{\theta_1}(Q_{\widehat{\theta}_1}(u_1)), \dots, F_{\theta_d}(Q_{\widehat{\theta}_d}(u_d))), \quad \mathbf{u} \in [0, 1]^d.$$

We also define

$$\alpha_n(\mathbf{u}) = \sqrt{n}[C_n(\mathbf{u}) - C(\mathbf{u})], \quad \mathbf{u} \in [0, 1]^d, \quad (4)$$

so that  $\alpha_n$  is the classical empirical process associated with the df  $C$ . The asymptotic behavior of  $\alpha_n$  is well-known, see e.g. Neuhaus (1971): we have  $\alpha_n \Rightarrow B_C$  in the Skorohod space  $D([0, 1]^d)$ , where “ $\Rightarrow$ ” denotes weak convergence and  $B_C$  is a  $C$ -Brownian bridge, that is, a mean-zero Gaussian process on  $[0, 1]^d$  with covariance structure

$$E[B_C(\mathbf{u})B_C(\mathbf{u}')] = C(\mathbf{u} \wedge \mathbf{u}') - C(\mathbf{u})C(\mathbf{u}').$$

Here,  $\mathbf{u} \wedge \mathbf{u}' := (u_1 \wedge u'_1, \dots, u_d \wedge u'_d)^\top$ .

The assumptions needed for our first result are listed below, followed by the result itself.

**A1.** There exist a  $p$ -variate random vector  $\zeta_0$  and  $m$ -variate random vectors  $\zeta_1, \dots, \zeta_d$  such that

$$(\alpha_n, \sqrt{n}(\boldsymbol{\lambda}_0 - \widehat{\boldsymbol{\lambda}}), \sqrt{n}(\boldsymbol{\theta}_1 - \widehat{\boldsymbol{\theta}}_1), \dots, \sqrt{n}(\boldsymbol{\theta}_d - \widehat{\boldsymbol{\theta}}_d)) \Rightarrow (B_C, \zeta_0, \zeta_1, \dots, \zeta_d) \quad (5)$$

in  $D([0, 1]^d) \times \mathbb{R}^p \times (\mathbb{R}^m)^d$ .

**A2.** The mappings

$$(\mathbf{u}, \boldsymbol{\lambda}) \mapsto \nabla C_{\boldsymbol{\lambda}}(\mathbf{u}) = (C_{\boldsymbol{\lambda}}^{(1)}(\mathbf{u}), \dots, C_{\boldsymbol{\lambda}}^{(d)}(\mathbf{u}))^\top := \left( \frac{\partial C_{\boldsymbol{\lambda}}(\mathbf{u})}{\partial u_1}, \dots, \frac{\partial C_{\boldsymbol{\lambda}}(\mathbf{u})}{\partial u_d} \right)^\top$$

and

$$(\mathbf{u}, \boldsymbol{\lambda}) \mapsto \dot{C}_{\boldsymbol{\lambda}}(\mathbf{u}) = (\dot{C}_{\boldsymbol{\lambda}}^{(1)}(\mathbf{u}), \dots, \dot{C}_{\boldsymbol{\lambda}}^{(p)}(\mathbf{u}))^{\top} := \left( \frac{\partial C_{\boldsymbol{\lambda}}(\mathbf{u})}{\partial \lambda_1}, \dots, \frac{\partial C_{\boldsymbol{\lambda}}(\mathbf{u})}{\partial \lambda_p} \right)^{\top}$$

are continuous on  $(0, 1)^d \times \Lambda$ .

**A3.** The mapping

$$(x, \boldsymbol{\theta}) \mapsto \dot{F}_{\boldsymbol{\theta}}(x) = (\dot{F}_{\boldsymbol{\theta}}^{(1)}(x), \dots, \dot{F}_{\boldsymbol{\theta}}^{(m)}(x))^{\top} := \left( \frac{\partial F_{\boldsymbol{\theta}}(x)}{\partial \theta_1}, \dots, \frac{\partial F_{\boldsymbol{\theta}}(x)}{\partial \theta_m} \right)^{\top}$$

is continuous on  $\mathbb{R} \times \Theta$ , the mapping  $(u, \boldsymbol{\theta}) \mapsto Q_{\boldsymbol{\theta}}(u)$  is bounded on compact subsets of  $(0, 1) \times \Theta$ , and the mapping  $(u, \boldsymbol{\theta}) \mapsto \dot{F}_{\boldsymbol{\theta}}(Q_{\boldsymbol{\theta}}(u))$  is continuous on  $(0, 1) \times \Theta$ .

**Theorem 2.1.** *Let  $\eta_n$  be the process defined in (3), and let  $0 < \varepsilon < \tau < 1$ . Under Assumptions A1–A3,*

$$\begin{aligned} \eta_n(\mathbf{u}) &\Rightarrow B_C(\mathbf{u}) + \sum_{j=1}^d C^{(j)}(\mathbf{u}) \dot{F}_{\boldsymbol{\theta}_j}(Q_{\boldsymbol{\theta}_j}(u_j))^{\top} \zeta_j + \dot{C}(\mathbf{u})^{\top} \zeta_0 \\ &=: \eta(\mathbf{u}) \end{aligned} \quad (6)$$

in  $D([\varepsilon, \tau]^d)$ , where  $C^{(j)}$  and  $\dot{C}$  are short-hand notation for  $C_{\boldsymbol{\lambda}_0}^{(j)}$  and  $\dot{C}_{\boldsymbol{\lambda}_0}$ .

**Remark 2.2.** Heuristically, the limiting process  $\eta$  consists of a  $C$ -Brownian bridge  $B_C$  plus  $dm$  additive terms “contributed by” the estimation of the  $m$ -dimensional parameters  $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_d$ , plus  $p$  additive terms “contributed by” the estimation of the  $p$ -dimensional parameter  $\boldsymbol{\lambda}_0$ . Since the distribution of  $\eta$  clearly depends on the underlying family  $\mathcal{C}$ , as well as the (unknown) true values of  $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_d$  and  $\boldsymbol{\lambda}_0$ , Theorem 2.1 is far from suitable for distribution-free testing.

**Remark 2.3.** The convergence in (6) does not necessarily hold in the space  $D([0, 1]^d)$  under the stated assumptions. For example, in the case  $m = 1$ , consider the parametric family  $\mathcal{F} = \{F_{\theta} : \theta \in (0, \infty)\}$ , with

$$F_{\theta}(x) = 1 - \sqrt{1 - x/\theta}, \quad x \in (0, \theta).$$

Note that  $F_{\theta}$  is a beta distribution with shape parameters fixed at 1 and 1/2, and a free scale parameter  $\theta > 0$ . Also note that  $F_{\theta}$  satisfies Assumption A3. However, for any  $\theta > 0$ , the expression

$$\dot{F}_{\theta}(Q_{\theta}(u)) = \frac{1}{2\theta} \left( 1 - u - \frac{1}{1-u} \right)$$

is unbounded near  $u = 1$  and hence the process  $\eta$  is in general not well-defined on the closed hypercube  $[0, 1]^d$ , so the convergence in (6) cannot hold in  $D([0, 1]^d)$ .

### 3. Transforming $\eta$ into a standard Wiener process

As we observed in the previous section, the empirical process  $\eta_n$  cannot directly be used as a basis for distribution-free testing, since its limiting process  $\eta$  depends on the underlying family  $\mathcal{C}$  and unknown parameter values. We will remedy this problem by transforming  $\eta$  into a standard  $d$ -variate Wiener process. The transformation itself will depend on  $\mathcal{C}$  and parameter values, but the distribution of the resulting process will not, which will facilitate asymptotically distribution-free testing. We first introduce some notation and assumptions.

Recall that

$$B_C \stackrel{d}{=} V_C - CV_C(\mathbf{1}),$$

with  $V_C$  a  $C$ -Wiener process on  $[0, 1]^d$ , i.e. a mean-zero Gaussian process with covariance  $E[V_C(\mathbf{u})V_C(\mathbf{u}')] = C(\mathbf{u} \wedge \mathbf{u}')$ . We can thus alternatively express the limiting process  $\eta$  in (6) as

$$\eta(\mathbf{u}) = V_C(\mathbf{u}) - C(\mathbf{u})V_C(\mathbf{1}) + \sum_{j=1}^d C^{(j)}(\mathbf{u})\dot{F}_{\theta_j}(Q_{\theta_j}(u_j))^\top \zeta_j + \dot{C}(\mathbf{u})^\top \zeta_0, \quad (7)$$

for  $\mathbf{u} \in (0, 1)^d$ . Hence we see that  $\eta$  is of the form

$$\eta(\mathbf{u}) = V_C(\mathbf{u}) + \sum_{i=1}^{1+dm+p} K_i(\mathbf{u})Z_i, \quad \mathbf{u} \in (0, 1)^d, \quad (8)$$

where the  $Z_i$  are some random variables, and the  $K_i$  are deterministic functions on  $(0, 1)^d$  defined by

$$\begin{aligned} K_1(\mathbf{u}) &= C(\mathbf{u}), \\ K_{1+(j-1)m+i}(\mathbf{u}) &= C^{(j)}(\mathbf{u})\dot{F}_{\theta_j}^{(i)}(Q_{\theta_j}(u_j)), \quad j = 1, \dots, d, \quad i = 1, \dots, m, \\ K_{1+dm+i}(\mathbf{u}) &= \dot{C}^{(i)}(\mathbf{u}), \quad i = 1, \dots, p. \end{aligned} \quad (9)$$

We note that (8) is analogous to the bivariate form (23) in [Can et al. \(2015\)](#). In that paper, a transformation of such processes into a standard bivariate Wiener process was described, which was an application of the ‘‘innovation martingale transform’’ idea developed in [Khmaladze \(1981, 1988, 1993\)](#). This idea has been applied to various statistical problems in the literature over the last couple of decades; see, for example, [McKeague, Nikabadze and Sun \(1995\)](#), [Nikabadze and Stute \(1997\)](#), [Stute, Thies and Zhu \(1998\)](#), [Koenker and Xiao \(2002, 2006\)](#), [Khmaladze and Koul \(2004, 2009\)](#), [Delgado, Hidalgo and Velasco \(2005\)](#) and [Dette and Hetzler \(2009\)](#). We will develop a suitable innovation martingale transform to construct a standard  $d$ -variate Wiener process on  $[0, 1]^d$  from the process  $\eta$  in (8).

The approach here is novel in the sense that direct use of the data naturally leads to processes on the unit hypercube  $[0, 1]^d$ , whereas in other applications of the martingale

transform in multivariate contexts, the data are first *transformed* to  $[0, 1]^d$  by an arbitrary transformation which influences the statistical properties of the procedures (see, for example, [Khmaladze \(1993\)](#) and [Einmahl and Khmaladze \(2001\)](#)).

We first state the necessary assumptions and establish some notation.

**A4.** For each  $\boldsymbol{\lambda} \in \Lambda$ , the copula  $C_{\boldsymbol{\lambda}}$  has a strictly positive density  $c_{\boldsymbol{\lambda}}$  on  $(0, 1)^d$ , and the mappings

$$(\mathbf{u}, \boldsymbol{\lambda}) \mapsto \nabla c_{\boldsymbol{\lambda}}(\mathbf{u}) = (c_{\boldsymbol{\lambda}}^{(1)}(\mathbf{u}), \dots, c_{\boldsymbol{\lambda}}^{(d)}(\mathbf{u}))^{\top} := \left( \frac{\partial c_{\boldsymbol{\lambda}}(\mathbf{u})}{\partial u_1}, \dots, \frac{\partial c_{\boldsymbol{\lambda}}(\mathbf{u})}{\partial u_d} \right)^{\top}$$

and

$$(\mathbf{u}, \boldsymbol{\lambda}) \mapsto \dot{c}_{\boldsymbol{\lambda}}(\mathbf{u}) = (\dot{c}_{\boldsymbol{\lambda}}^{(1)}(\mathbf{u}), \dots, \dot{c}_{\boldsymbol{\lambda}}^{(p)}(\mathbf{u}))^{\top} := \left( \frac{\partial c_{\boldsymbol{\lambda}}(\mathbf{u})}{\partial \lambda_1}, \dots, \frac{\partial c_{\boldsymbol{\lambda}}(\mathbf{u})}{\partial \lambda_p} \right)^{\top}$$

are continuous on  $(0, 1)^d \times \Lambda$ .

Now, with the functions  $K_i$  as defined in (9), let us denote

$$k_i(\mathbf{u}) = dK_i(\mathbf{u})/dC(\mathbf{u}), \quad i = 1, \dots, 1 + dm + p, \quad (10)$$

so that

$$\begin{aligned} k_1(\mathbf{u}) &= 1, \\ k_{1+(j-1)m+i}(\mathbf{u}) &= \dot{F}_{\boldsymbol{\theta}_j}^{(i)}(Q_{\boldsymbol{\theta}_j}(u_j)) \frac{\partial}{\partial u_j} \log c_{\boldsymbol{\lambda}_0}(\mathbf{u}) + \frac{\partial}{\partial u_j} \dot{F}_{\boldsymbol{\theta}_j}^{(i)}(Q_{\boldsymbol{\theta}_j}(u_j)), \\ &\quad j = 1, \dots, d, i = 1, \dots, m, \\ k_{1+dm+i}(\mathbf{u}) &= \frac{\partial}{\partial \lambda_i} \log c_{\boldsymbol{\lambda}}(\mathbf{u}) \Big|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}_0}, \quad i = 1, \dots, p. \end{aligned}$$

Let  $\mathbf{k}(\mathbf{u})$  denote the column vector consisting of  $k_1(\mathbf{u}), \dots, k_{1+dm+p}(\mathbf{u})$ . We will also write  $\mathbf{k}(\mathbf{u}, \boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_d, \boldsymbol{\lambda}')$  for the vector  $\mathbf{k}(\mathbf{u})$  with true parameter values  $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_d, \boldsymbol{\lambda}_0$  replaced by arbitrary values  $\boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_d \in \Theta, \boldsymbol{\lambda}' \in \Lambda$ .

Finally, given  $0 < \delta < 1/2$ , let

$$S_{\delta}(t) = [\delta, 1 - \delta/2]^{d-1} \times [t, 1 - \delta/2], \quad t \in [\delta, 1 - \delta/2],$$

and introduce matrices

$$\mathbf{I}_{\delta}(t) = \int_{S_{\delta}(t)} \mathbf{k}(\mathbf{s})\mathbf{k}(\mathbf{s})^{\top} dC(\mathbf{s}), \quad t \in [\delta, 1 - \delta/2]. \quad (11)$$

We also define

$$\begin{aligned} \mathbf{I}_{\delta}(t, \boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_d, \boldsymbol{\lambda}') &= \int_{S_{\delta}(t)} \mathbf{k}(\mathbf{s}, \boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_d, \boldsymbol{\lambda}') \mathbf{k}(\mathbf{s}, \boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_d, \boldsymbol{\lambda}')^{\top} dC_{\boldsymbol{\lambda}'}(\mathbf{s}), \\ &\quad t \in [\delta, 1 - \delta/2]. \end{aligned} \quad (12)$$

Note that we use  $1 - \delta/2$  rather than  $1 - \delta$  in the definition of  $S_\delta(t)$ , in order to avoid potential problems with the matrices  $\mathbf{I}_\delta(t)$  becoming nearly singular as  $t$  approaches  $1 - \delta$ . Also note that in the nomenclature of likelihood theory,  $\mathbf{k}$  is the vector of *score functions* for the underlying copula model (extended by the constant function 1 in the first component), and  $\mathbf{I}_\delta(t)$  is a partial *Fisher information matrix* constructed from these functions.

Our next assumption is:

**A5.** The matrices  $\mathbf{I}_\delta(t, \boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_d, \boldsymbol{\lambda}')$  in (12) are well-defined and invertible for all  $0 < \delta < 1/2$ ,  $t \in [\delta, 1 - \delta/2)$ ,  $\boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_d \in \Theta$ ,  $\boldsymbol{\lambda}' \in \Lambda$ .

Given  $0 < \delta < 1/2$ , let  $\boldsymbol{\delta}$  denote the point  $(\delta, \dots, \delta)^\top \in (0, 1)^d$ , and given points  $\mathbf{u}, \mathbf{v} \in [0, 1]^d$  with  $u_j \leq v_j$  for  $j = 1, \dots, d$ , let  $[\mathbf{u}, \mathbf{v}]$  denote the hyperrectangle  $[u_1, v_1] \times \dots \times [u_d, v_d]$ . Also, given  $a > 0$ , let  $a\mathbf{u}$  denote the point  $(au_1, \dots, au_d)^\top$ .

We are now ready to state our transformation result. In what follows, the integrals with respect to  $d\eta$  can be interpreted as  $d$ -dimensional Riemann-Stieltjes integrals defined per sample path; see Theorem 1.2(c) of [Towghi \(2002\)](#) for the relevant existence result.

**Theorem 3.1.** *Let  $\eta$  be the limiting process appearing in Theorem 2.1 and let  $0 < \delta < 1/2$ . If Assumptions A4-A5, restricted to the true parameter values  $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_d, \boldsymbol{\lambda}_0$ , hold, then the process*

$$W(\mathbf{u}) = \frac{1}{(1 - 2\delta)^{d/2}} \left[ \int_{[\boldsymbol{\delta}, \boldsymbol{\delta} + (1 - 2\delta)\mathbf{u}]} \frac{1}{\sqrt{c(\mathbf{s})}} d\eta(\mathbf{s}) - \int_{[\boldsymbol{\delta}, \boldsymbol{\delta} + (1 - 2\delta)\mathbf{u}]} \mathbf{k}(\mathbf{s})^\top \left( \mathbf{I}_\delta^{-1}(s_d) \int_{S_\delta(s_d)} \mathbf{k}(\mathbf{s}') d\eta(\mathbf{s}') \right) \sqrt{c(\mathbf{s})} d\mathbf{s} \right] \quad (13)$$

is a standard Wiener process on  $[0, 1]^d$ .

**Remark 3.2.** The transformation in (13) “annihilates” the terms following  $V_C$  in (8) to produce a  $C$ -Wiener process on  $[\delta, 1 - \delta]^d$ , which is then normalized and scaled to the entire hypercube  $[0, 1]^d$ , so that the end result is indeed a standard Wiener process on  $[0, 1]^d$ . In the next section, we will describe how the transformation of Theorem 3.1 facilitates asymptotically distribution-free testing for  $C$ .

## 4. Goodness-of-fit testing: null hypothesis

Recall the empirical process  $\eta_n$  defined in (3). In Theorem 2.1 we have derived its weak limit  $\eta$  as  $n \rightarrow \infty$ , and in Theorem 3.1 we have described a transformation that turns  $\eta$  into a standard Wiener process on  $[0, 1]^d$ . In this section, we will apply the same transformation (or rather, its empirical version, with unknown parameters replaced by

estimators) to  $\eta_n$ , and we will show that the resulting process converges weakly to a standard Wiener process. This is the first main result of this paper.

Applying transformation (13) to  $\eta_n$ , with unknown parameters replaced by estimators, we obtain the following empirical process on  $[0, 1]^d$ :

$$W_n(\mathbf{u}) = \frac{1}{(1-2\delta)^{d/2}} \left[ \int_{[\delta, \delta+(1-2\delta)\mathbf{u}]} \frac{1}{\sqrt{c_{\hat{\lambda}}(\mathbf{s})}} d\eta_n(\mathbf{s}) - \int_{[\delta, \delta+(1-2\delta)\mathbf{u}]} \hat{\mathbf{k}}(\mathbf{s})^\top \left( \hat{\mathbf{I}}_\delta^{-1}(s_d) \int_{S_\delta(s_d)} \hat{\mathbf{k}}(\mathbf{s}') d\eta_n(\mathbf{s}') \right) \sqrt{c_{\hat{\lambda}}(\mathbf{s})} ds \right]. \quad (14)$$

Here,  $\hat{\mathbf{k}}(\cdot)$  and  $\hat{\mathbf{I}}_\delta(\cdot)$  are short-hand notations for  $\mathbf{k}(\cdot, \hat{\boldsymbol{\theta}}_1, \dots, \hat{\boldsymbol{\theta}}_d, \hat{\boldsymbol{\lambda}})$  and  $\mathbf{I}_\delta(\cdot, \hat{\boldsymbol{\theta}}_1, \dots, \hat{\boldsymbol{\theta}}_d, \hat{\boldsymbol{\lambda}})$ , respectively.

Before stating the convergence result on  $W_n$ , we introduce some further notation and assumptions. Given a hyperrectangle  $[\mathbf{a}, \mathbf{b}] \subset \mathbb{R}^d$  and a function  $\varphi : [\mathbf{a}, \mathbf{b}] \rightarrow \mathbb{R}$ , let  $V_{[\mathbf{a}, \mathbf{b}]}(\varphi)$  denote the total variation of  $\varphi$  on  $[\mathbf{a}, \mathbf{b}]$  in the sense of Vitali; see e.g. Owen (2005), Sec. 4 for a definition. Also, given  $I \subset \{1, \dots, d\}$  and  $\mathbf{x} \in \mathbb{R}^d$ , let  $|I|$  denote the cardinality of  $I$ , and let  $\mathbf{x}_I$  denote the point in  $\mathbb{R}^{|I|}$  obtained by discarding all coordinates  $x_j$  of  $\mathbf{x}$  for  $j \notin I$ . Moreover, given disjoint subsets  $I_1, I_2, I_3 \subset \{1, \dots, d\}$  with  $I_1 \cup I_2 \cup I_3 = \{1, \dots, d\}$ , let  $\varphi(\mathbf{x}_{I_1}; \mathbf{a}_{I_2}, \mathbf{b}_{I_3})$  denote the function on  $[\mathbf{a}_{I_1}, \mathbf{b}_{I_1}]$  obtained by fixing the  $j^{\text{th}}$  argument of  $\varphi$  at  $a_j$  for  $j \in I_2$  and at  $b_j$  for  $j \in I_3$ .

We consider an alternative concept of “total variation” on  $[\mathbf{a}, \mathbf{b}]$ , as follows:

$$V_{[\mathbf{a}, \mathbf{b}]}^{\text{HK}}(\varphi) := \sum_{\substack{I_1, I_2, I_3 \subset \{1, \dots, d\}, I_1 \neq \emptyset \\ I_1 + I_2 + I_3 = \{1, \dots, d\}}} V_{[\mathbf{a}, \mathbf{b}]}(\varphi(\mathbf{x}_{I_1}; \mathbf{a}_{I_2}, \mathbf{b}_{I_3})), \quad (15)$$

with  $I_1 + I_2 + I_3$  denoting a disjoint union. In other words,  $V_{[\mathbf{a}, \mathbf{b}]}^{\text{HK}}(\cdot)$  sums the Vitali variations over the hyperrectangle  $[\mathbf{a}, \mathbf{b}]$  and over all of its “faces” where the  $j^{\text{th}}$  coordinate is fixed at  $a_j$  or  $b_j$ , for at least one  $j \in \{1, \dots, d\}$ . Note that  $V_{[\mathbf{a}, \mathbf{b}]}^{\text{HK}}$  is a variant of the so-called *Hardy-Krause variation* for multivariate functions; cf. Owen (2005), Def. 2. For  $0 < \delta < 1/2$  and functions  $\varphi : [\delta, 1-\delta]^d \rightarrow \mathbb{R}$ , we will write  $V_\delta^{\text{HK}}$  instead of  $V_{[\delta, 1-\delta]^d}^{\text{HK}}$ , for brevity.

Let us also denote

$$\gamma(\mathbf{u}) = \frac{1}{\sqrt{c(\mathbf{u})}}, \quad \hat{\gamma}(\mathbf{u}) = \frac{1}{\sqrt{c_{\hat{\lambda}}(\mathbf{u})}}, \quad \Delta\gamma(\mathbf{u}) = \hat{\gamma}(\mathbf{u}) - \gamma(\mathbf{u}), \quad \mathbf{u} \in (0, 1)^d.$$

Similarly, we will denote  $\Delta k_i = \hat{k}_i - k_i$  for  $i = 1, \dots, 1 + dm + p$ , with the  $k_i$  as defined in (10). We introduce the final assumption needed for our convergence result:

**A6.** For any  $0 < \delta < 1/2$ , we have  $V_\delta^{\text{HK}}(\gamma) < \infty$  and  $V_\delta^{\text{HK}}(\Delta\gamma) = o_P(1)$ . Also,  $V_\delta^{\text{HK}}(k_i) < \infty$  and  $V_\delta^{\text{HK}}(\Delta k_i) = o_P(1)$  for  $i = 1, \dots, 1 + dm + p$ .

Under assumption A1, a sufficient (but not necessary) set of conditions for A6 is: (i) the existence and continuity of the mixed derivative  $\partial\gamma/\partial u_1 \dots \partial u_d$  on  $[\delta, 1-\delta]^d \times \Lambda$ , (ii) the existence and continuity of the mixed derivative  $\partial k_i/\partial u_1 \dots \partial u_d$  on  $[\delta, 1-\delta]^d \times \Theta^d \times \Lambda$ , for  $i = 1, \dots, 1+dm+p$ , and (iii) the analogous statements to (i) and (ii) for the functions  $\gamma$  and  $k_i$  restricted to each lower-dimensional “face” of the hypercube  $[\delta, 1-\delta]^d$ .

**Theorem 4.1.** *Let  $0 < \delta < 1/2$ . Under Assumptions A1–A6, the process  $W_n$  in (14) converges weakly to a standard Wiener process in  $D([0, 1]^d)$ .*

**Remark 4.2.** Theorem 4.1 is analogous to Theorem 2.1 in that it describes the asymptotic behavior of an empirical process constructed from the data  $\mathbf{X}_1, \dots, \mathbf{X}_n$ . However, unlike the process  $\eta_n$ , the asymptotic behavior of  $W_n$  is distribution-free: it converges to a standard Wiener process. Thus a test for the null hypothesis can now be performed by assessing how the observed path of  $W_n$  compares to the “usual” statistical behavior of a standard Wiener process. Since this comparison can be done through many different functionals of  $W_n$ , we can construct a multitude of asymptotically distribution-free tests. In Section 6, we demonstrate through simulations and a real-world data analysis how such tests can be conducted.

**Remark 4.3.** The statement and proof of Theorem 4.1 also applies, with obvious modifications, in the case  $\mathcal{C} = \{C_0\}$ , where  $C_0$  denotes a fully specified copula. Thus the test process  $W_n$  can also be used for testing null hypotheses of the form  $C = C_0$ . This will also be demonstrated in the simulations of Section 6.

## 5. Goodness-of-fit testing: contiguous alternatives

We now consider testing  $C \in \mathcal{C} = \{C_\lambda : \lambda \in \Lambda\}$  when the true copula of the underlying sample does not lie in  $\mathcal{C}$  but approaches it as the sample size grows.

So let us assume that, for each  $n \geq 1$ , we have an i.i.d. sample

$$\mathbf{X}_{(n)1} = (X_{(n)11}, \dots, X_{(n)1d})^\top, \dots, \mathbf{X}_{(n)n} = (X_{(n)n1}, \dots, X_{(n)nd})^\top \quad (16)$$

generated from a  $d$ -variate df  $F_{(n)}$  with continuous margins  $F_1, \dots, F_d$  and copula  $C_{(n)}$ . We assume that the marginal dfs are independent of  $n$ , and (as in the previous sections) they are all members of some parametric family  $\mathcal{F} = \{F_\theta : \theta \in \Theta\}$ , so that there are  $\theta_1, \dots, \theta_d \in \Theta$  with  $F_j = F_{\theta_j}$  for  $j = 1, \dots, d$ . Regarding the sequence of copulas  $C_{(1)}, C_{(2)}, \dots$ , we assume the following:

**B0.** There exists  $\lambda_0 \in \Lambda$  such that

$$\left[ \frac{dC_{(n)}}{dC_{\lambda_0}} \right]^{1/2} = 1 + \frac{1}{2\sqrt{n}} h_n, \quad n = 1, 2, \dots,$$

for a sequence of functions  $h_1, h_2, \dots$  supported on  $[\delta, 1 - \delta]^d$ , for some  $0 < \delta < 1/2$ . The functions  $h_n$  satisfy

$$\int_{[0,1]^d} (h_n - h)^2 dC_{\lambda_0} \rightarrow 0 \text{ as } n \rightarrow \infty,$$

for some function  $h$  with

$$\int_{[0,1]^d} h^2 dC_{\lambda_0} \in (0, \infty), \quad \int_{[0,1]^d} k_i h dC_{\lambda_0} = 0,$$

where the functions  $k_i, i = 1, \dots, 1 + dm + p$ , are as defined in (10).

Note that for each  $n \geq 1$ , the distribution of the sample in (16) on  $(\mathbb{R}^d)^n$  is given by the  $n$ -fold product measure  $F_{(n)}^n = F_{(n)} \times \dots \times F_{(n)}$ , whereas if the underlying copula was equal to  $C_{\lambda_0}$ , this distribution would of course be  $F_0^n = F_0 \times \dots \times F_0$ , with  $F_0$  denoting the df with margins  $F_1, \dots, F_d$  and copula  $C_{\lambda_0}$ . It follows from Oosterhoff and van Zwet (1979) that condition B0 is sufficient to make the sequence  $\{F_{(n)}^n\}$  contiguous with respect to  $\{F_0^n\}$ , in the sense that  $\lim_{n \rightarrow \infty} F_0^n(A_n) = 0$  implies  $\lim_{n \rightarrow \infty} F_{(n)}^n(A_n) = 0$ , for any sequence of measurable sets  $A_n \subset (\mathbb{R}^d)^n$ .

Our first result in this section will establish the asymptotic behavior of  $\eta_n$  in (3) in the present setting. We define, analogously to (4),

$$\alpha_n(\mathbf{u}) = \sqrt{n}[C_n(\mathbf{u}) - C_{(n)}(\mathbf{u})], \quad \mathbf{u} \in [0, 1]^d,$$

where  $C_n$  is the empirical df generated by the (unobserved) copula sample

$$(F_1(X_{(n)11}), \dots, F_d(X_{(n)1d}))^\top, \dots, (F_1(X_{(n)n1}), \dots, F_d(X_{(n)nd}))^\top,$$

and we state the following analogue of Assumption A1 in Section 2:

**B1.** There exist a  $p$ -variate random vector  $\zeta_a$  and  $m$ -variate random vectors  $\zeta_1, \dots, \zeta_d$  such that

$$\begin{aligned} (\alpha_n, \sqrt{n}(\lambda_0 - \hat{\lambda}), \sqrt{n}(\theta_1 - \hat{\theta}_1), \dots, \sqrt{n}(\theta_d - \hat{\theta}_d)) \\ \Rightarrow (B_{C_{\lambda_0}}, \zeta_a, \zeta_1, \dots, \zeta_d) \end{aligned}$$

in  $D([0, 1]^d) \times \mathbb{R}^p \times (\mathbb{R}^m)^d$ .

The following result is the analogue of Theorem 2.1, and its proof, which we omit, follows along similar lines.

**Theorem 5.1.** *Let  $\eta_n$  be the process defined in (3), and let  $0 < \varepsilon < \tau < 1$ . Under*

Assumptions B0–B1 and A2–A3,

$$\begin{aligned} \eta_n(\mathbf{u}) &\Rightarrow B_{C_{\lambda_0}}(\mathbf{u}) + \sum_{j=1}^d C_{\lambda_0}^{(j)}(\mathbf{u}) \dot{F}_{\theta_j}(Q_{\theta_j}(u_j))^\top \zeta_j + \dot{C}_{\lambda_0}(\mathbf{u})^\top \zeta_a \\ &\quad + \int_{[0, \mathbf{u}]} h(\mathbf{s}) \, dC_{\lambda_0}(\mathbf{s}) \\ &=: \eta_a(\mathbf{u}) + \int_{[0, \mathbf{u}]} h(\mathbf{s}) \, dC_{\lambda_0}(\mathbf{s}) \end{aligned}$$

in  $D([\varepsilon, \tau]^d)$ .

Next, we establish the asymptotic behavior of the test process  $W_n$  in (14) in the present setting. If we let  $W_a$  denote the process  $W$  in (13), with  $\eta$  replaced by  $\eta_a$ , then  $W_a$  is still a standard Wiener process on  $[0, 1]^d$ , since the change of  $\eta$  to  $\eta_a$  does not affect the proof of Theorem 3.1. This, together with Theorem 5.1 above, yields the following analogue of Theorem 4.1, which shows that under the sequence of contiguous alternatives,  $W_n$  converges to a standard Wiener process plus a deterministic shift term.

**Theorem 5.2.** *Under Assumptions B0–B1 and A2–A6, the process  $W_n$  in (14), with  $\delta$  as in Assumption B0, converges weakly to  $\tilde{W} := W + S$  in  $D([0, 1]^d)$ , where*

$$S(\mathbf{u}) = \frac{1}{(1 - 2\delta)^{d/2}} \int_{[\delta, \delta + (1 - 2\delta)\mathbf{u}]} g(\mathbf{s}) \sqrt{c_{\lambda_0}(\mathbf{s})} \, d\mathbf{s},$$

with

$$g(\mathbf{s}) = h(\mathbf{s}) - \mathbf{k}(\mathbf{s})^\top \mathbf{I}_\delta^{-1}(s_d) \int_{S_\delta(s_d)} \mathbf{k}(\mathbf{s}') h(\mathbf{s}') \, dC_{\lambda_0}(\mathbf{s}'), \quad \mathbf{s} \in [\delta, 1 - \delta]^d. \quad (17)$$

In order to judge how “sensitive” the test process  $W_n$  is to the sequence of alternatives  $F_{(1)}, F_{(2)}, \dots$ , we first recall the notion of *distance in variation* for probability measures. Given two such measures  $P$  and  $\tilde{P}$  defined on some sigma-algebra  $\mathcal{B}$ , the distance in variation between  $P$  and  $\tilde{P}$  is defined as

$$d(P, \tilde{P}) = \sup_{B \in \mathcal{B}} |P(B) - \tilde{P}(B)|.$$

If  $L_n$  denotes the log-likelihood ratio  $\log(dF_{(n)}^n/dF_0^n)$ , then

$$d(F_{(n)}^n, F_0^n) = F_{(n)}^n(L_n > 0) - F_0^n(L_n > 0). \quad (18)$$

Moreover, we have

$$L_n \xrightarrow{d} \begin{cases} N(-\frac{1}{2}\|h\|^2, \|h\|^2) & \text{under } F_0^n, \\ N(\frac{1}{2}\|h\|^2, \|h\|^2) & \text{under } F_{(n)}^n, \end{cases} \quad (19)$$

where  $N(\mu, \sigma^2)$  denotes the normal distribution with mean  $\mu$  and variance  $\sigma^2$ , and

$$\|h\| := \left( \int_{[0,1]^d} h^2 dC_{\lambda_0} \right)^{1/2}; \quad (20)$$

see Oosterhoff and van Zwet (1979) for the first statement in (19) and, e.g., Ch. 6 of van der Vaart (1998) for the second statement. Combining (18) and (19), we see that

$$d(F_{(n)}^n, F_0^n) \rightarrow \nu(h) := 2\Phi\left(\frac{1}{2}\|h\|\right) - 1$$

as  $n \rightarrow \infty$ , with  $\Phi$  denoting the standard normal cdf.

The following result is the second main result of the paper. It establishes asymptotic optimality of the test process  $W_n$ .

**Theorem 5.3.** *Let  $Q$  denote the distribution of a standard Wiener process  $W$  on  $D([0, 1]^d)$ , and let  $\tilde{Q}$  denote the distribution of the process  $\tilde{W}$  defined in Theorem 5.2. Then,*

$$\log\left(\frac{d\tilde{Q}}{dQ}\right) \sim \begin{cases} N\left(-\frac{1}{2}\|h\|^2, \|h\|^2\right) & \text{under } Q, \\ N\left(\frac{1}{2}\|h\|^2, \|h\|^2\right) & \text{under } \tilde{Q}. \end{cases}$$

Hence  $d(\tilde{Q}, Q) = \nu(h)$ .

**Remark 5.4.** The result shows that the limiting distance in variation of the processes  $W_n$  under the null and the contiguous alternatives is the same as that of the samples:  $\nu(h)$ . In fact, the respective distributions of the log-likelihood ratio  $\log(d\tilde{Q}/dQ)$  under the two measures are identical to the limiting distributions of  $\log(dF_{(n)}^n/dF_0^n)$ . Hence, the process  $W_n$  is asymptotically as good as the data themselves for testing purposes. Indeed, for a given sequence of alternatives satisfying Assumption B0, consider the test that rejects  $H_0$  if

$$(1 - 2\delta)^{d/2} \int_{[0,1]^d} \hat{g}(\boldsymbol{\delta} + (1 - 2\delta)\mathbf{s}) \sqrt{c_{\hat{\lambda}}(\boldsymbol{\delta} + (1 - 2\delta)\mathbf{s})} dW_n(\mathbf{s}) \geq \widehat{\|h\|} \Phi^{-1}(1 - \alpha),$$

with  $\hat{g}$  and  $\widehat{\|h\|}$  denoting the obvious estimators of  $g$  in (17) and  $\|h\|$  in (20). Then under regularity assumptions the probability of a type I error converges to  $\alpha$  as  $n \rightarrow \infty$ , and the power converges to  $1 - \Phi(\Phi^{-1}(1 - \alpha) - \|h\|)$ . According to the Neyman-Pearson Lemma, this limiting power is equal to that of the most powerful level- $\alpha$  tests for a simple null (picked from our  $H_0$ ) against the simple  $h_n$ -alternatives. Hence, for the more general problem of testing the composite null hypothesis  $C \in \mathcal{C}$  against the composite  $h_n$ -alternatives, we have an asymptotically uniformly most powerful test. This optimality shows that our approach can favorably compete in terms of power with any other approach in the literature.

## 6. Simulations and data analysis

In this section we present the results of a simulation study and a data analysis in order to illustrate the applicability of our approach in finite samples. All computations are performed in **R**. The code to implement the simulations and the data analysis is available from the authors upon request.

### 6.1. Simulation study

We consider four widely used bivariate copula models: Clayton, Gumbel, Normal and Student's  $t$ . We demonstrate how one might test for the goodness-of-fit of these models, both in parametric and fully specified form, using our approach.

More specifically, the models we consider are the following:

- Clayton:  $C_\lambda(u, v) = (u^{-\lambda} + v^{-\lambda} - 1)^{-1/\lambda}$ ,  $\lambda \in (0, \infty)$
- Gumbel:  $C_\lambda(u, v) = \exp\{-[(-\log u)^\lambda + (-\log v)^\lambda]^{1/\lambda}\}$ ,  $\lambda \in (1, \infty)$
- Normal:

$$C_\lambda(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\lambda^2}} \exp\left(-\frac{x^2 - 2\lambda xy + y^2}{2(1-\lambda^2)}\right) du dv, \\ \lambda \in (-1, 1)$$

- $t_3$ :

$$C_\lambda(u, v) = \int_{-\infty}^{T_3^{-1}(u)} \int_{-\infty}^{T_3^{-1}(v)} \frac{1}{2\pi\sqrt{1-\lambda^2}} \left(1 + \frac{x^2 - 2\lambda xy + y^2}{3(1-\lambda^2)}\right)^{-5/2} du dv, \\ \lambda \in (-1, 1)$$

Here,  $\Phi^{-1}$  denotes the quantile function of the standard normal distribution, and  $T_3^{-1}$  denotes the quantile function of the univariate  $t$  distribution with 3 degrees of freedom. For computational ease, the degrees of freedom parameter of the  $t$  copula is fixed at 3, which leaves only one parameter  $\lambda$  to be estimated. We will also consider testing for the fully specified copula models Clayton(2), Gumbel(2), Normal(0.5) and  $t_3(0.5)$ .

To test for Clayton and Clayton(2) models under the null hypothesis, we generate 1000 samples of size  $n = 200$  from the bivariate distribution with Exponential(1) margins and Clayton(2) copula. To test for Gumbel and Gumbel(2) under the null hypothesis, we generate 1000 samples of size  $n = 200$  from the bivariate distribution with Lomax(3,1) margins and Gumbel(2) copula. Recall that for  $\alpha > 0$  and  $\sigma > 0$ , the Lomax( $\alpha, \sigma$ ) distribution has the cdf

$$F(x) = 1 - \left(1 + \frac{x}{\sigma}\right)^{-\alpha}, \quad x > 0,$$

so it is a shifted Pareto distribution with tail parameter  $\alpha$  and scale parameter  $\sigma$ . To test for Normal and Normal(0.5) models under the null hypothesis, we generate 1000 samples of size  $n = 200$  from the bivariate normal distribution with standard normal margins and correlation coefficient 0.5. Finally, to test for  $t_3$  and  $t_3(0.5)$  models under the null hypothesis, we generate 1000 samples of size  $n = 200$  from the bivariate distribution with standard normal margins and  $t_3(0.5)$  copula.

From each simulated sample, we compute the test process  $W_n$  in (14) on a  $100 \times 100$  grid  $\mathcal{G}$  of equally spaced points covering  $(0, 1)^2$ , taking  $\delta = 0.001$ . Parameter estimates are computed through maximum likelihood (ML) estimation. For the parametric copula models, ML estimates are computed for the entire bivariate distribution, while for the fully specified copula models, ML estimates are computed separately for the two marginal parameter sets. Note that Assumption A1 of Section 2 holds for these estimators, which follows from arguments similar to those for asymptotic normality of MLEs. Assumptions A2-A6 are smoothness assumptions that are straightforward to verify for the considered models.

To compare the observed paths of  $W_n$  to a standard Wiener process, two functionals are computed from each path of  $W_n$ , namely:

$$\begin{aligned} \kappa_n &= \max_{(x,y) \in \mathcal{G}} |W_n(x,y)|, && \text{(Kolmogorov-Smirnov type statistic)} \\ \omega_n^2 &= \|\mathcal{G}\|^2 \sum_{(x,y) \in \mathcal{G}} W_n(x,y)^2, && \text{(Cramér-von Mises type statistic)} \end{aligned}$$

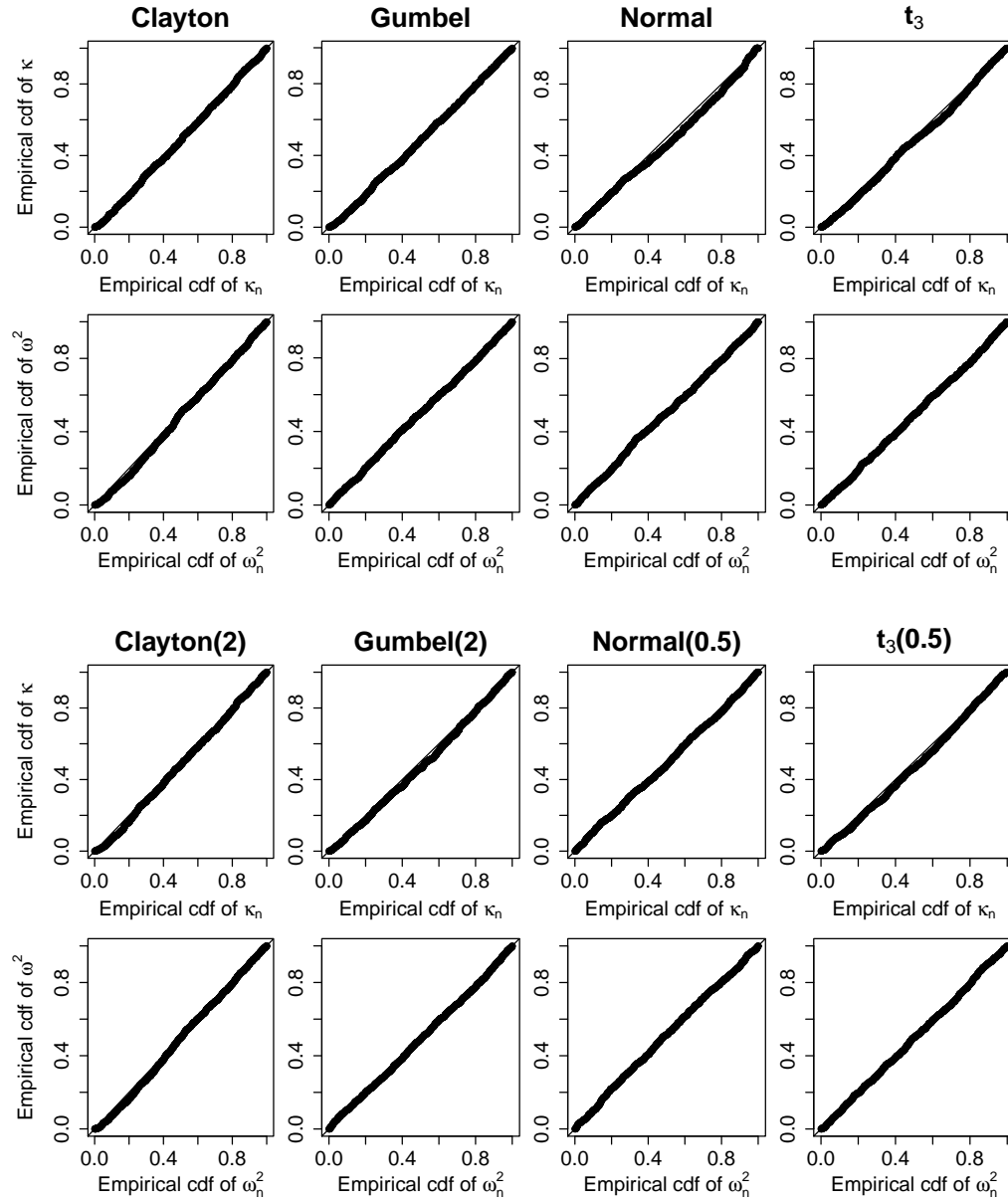
where  $\|\mathcal{G}\|$  denotes the mesh length of the grid  $\mathcal{G}$ , i.e.  $1/100$ . To create benchmark distribution tables for these statistics, we also simulate 10,000 true standard Wiener process paths on the grid  $\mathcal{G}$  and compute the same functionals for each path. We denote these functionals by  $\kappa$  and  $\omega^2$ .

For each model, we construct PP-plots to compare the empirical distributions of  $\kappa_n$  and  $\omega_n^2$  with the theoretical distributions of  $\kappa$  and  $\omega^2$  (as inferred from the 10,000 simulated Wiener process paths). The results are shown in Fig. 1 below. We observe a very good match of empirical and limiting distributions for both statistics, especially in the upper right corners of the plots, which are important for testing. These results suggest that Theorem 4.1 yields good finite-sample approximations. This is confirmed by the observed fractions of Type I errors (i.e. empirical sizes) at the 5% significance level, given in Table 1. Note that all rejection counts are consistent with draws from a Binomial(1000, 0.05) distribution.

**Table 1.** Number of rejections for 1000 samples at the 5% significance level under the various null hypotheses

	Clayton	Gumbel	Normal	$t_3$	Clayton(2)	Gumbel(2)	Normal(0.5)	$t_3(0.5)$
$\kappa_n$	41	46	51	52	45	46	46	58
$\omega_n^2$	53	43	43	49	53	47	57	48

We emphasize here that due to the distribution-free nature of our approach, the critical values of the test statistics need to be computed only once. The critical values of  $\kappa_n$  and  $\omega_n^2$  at commonly used significance levels are given in Table 2 below.



**Figure 1.** PP-plots for the Kolmogorov-Smirnov and Cramér-von Mises type test statistics computed for the four parametric models (top) and the four fully specified models (bottom)

**Table 2.** Critical values of  $\kappa_n$  and  $\omega_n^2$  at various significance levels

	10%	5%	1%
$\kappa_n$	2.100	2.362	2.865
$\omega_n^2$	0.526	0.708	1.186

To observe the behavior of the test statistics under the alternative hypothesis, we focus on the Clayton and Gumbel models. To test for the Clayton and Clayton(2) models under the alternative hypothesis, we generate 1000 samples of size  $n = 200$  from the bivariate distribution with Exponential(1) margins and Gumbel(2) copula. Note that the Gumbel(2) copula has the same Kendall's tau as Clayton(2), namely  $1/2$ . To test for the Gumbel and Gumbel(2) models under the alternative hypothesis, we generate 1000 samples of size  $n = 200$  from the bivariate distribution with Lomax(3,1) margins and Clayton(2) copula. From each sample, we construct the test process  $W_n$  on the grid  $\mathcal{G}$  and compute the two test statistics  $\kappa_n$  and  $\omega_n^2$  as before. The resulting rejection frequencies at 5% significance level are shown in Table 3 below. These numbers confirm that tests based on our approach have high power, even with a moderate sample size of 200.

**Table 3.** Number of rejections for 1000 samples at the 5% significance level under the various alternative hypotheses

testing for: sampled from:	Clayton Gumbel(2)	Gumbel Clayton(2)	Clayton(2) Gumbel(2)	Gumbel(2) Clayton(2)
$\kappa_n$	787	956	1000	768
$\omega_n^2$	857	938	999	761

## 6.2. Data analysis

We consider a data set consisting of log-concentrations of seven metallic elements (uranium [U], lithium [Li], cobalt [Co], potassium [K], caesium [Cs], scandium [Sc], titanium [Ti]) in 655 water samples collected near Grand Junction, Colorado in the late 1970s. In [Cook and Johnson \(1986\)](#), the pairwise dependence structures of U-Cs, Co-Ti and Cs-Sc log-concentrations were investigated, and it was found that the Clayton copula, or a two-parameter extension of it, provides a better fit (in terms of likelihood values) to each of these pairs than the normal copula, under the assumption of normal marginal distributions. This two-parameter extension of the Clayton copula is defined as:

$$\begin{aligned}
 C_{\lambda_1, \lambda_2}(u, v) &= (1 + \lambda_2)(u^{-\lambda_1} + v^{-\lambda_1} - 1)^{-1/\lambda_1} + \lambda_2(2u^{-\lambda_1} + 2v^{-\lambda_1} - 3)^{-1/\lambda_1} \\
 &\quad - \lambda_2(2u^{-\lambda_1} + v^{-\lambda_1} - 2)^{-1/\lambda_1} - \lambda_2(u^{-\lambda_1} + 2v^{-\lambda_1} - 2)^{-1/\lambda_1},
 \end{aligned} \tag{21}$$

for parameters  $\lambda_1 > 0$  and  $\lambda_2 \in [0, 1]$ . Note that when  $\lambda_2 = 0$ , the expression (21) reduces to the usual Clayton copula with parameter  $\lambda_1$ .

For our analysis, we focus on the pair Co-Sc (which was not investigated in [Cook and Johnson \(1986\)](#)), since the assumption of normal margins seems most plausible for

the Co and Sc log-concentrations; see normal QQ-plots in Fig. 2 below. Also see Fig. 3 for a scatter plot of the Co-Sc log-concentrations as well as a scatter plot of the rank-transformed data.

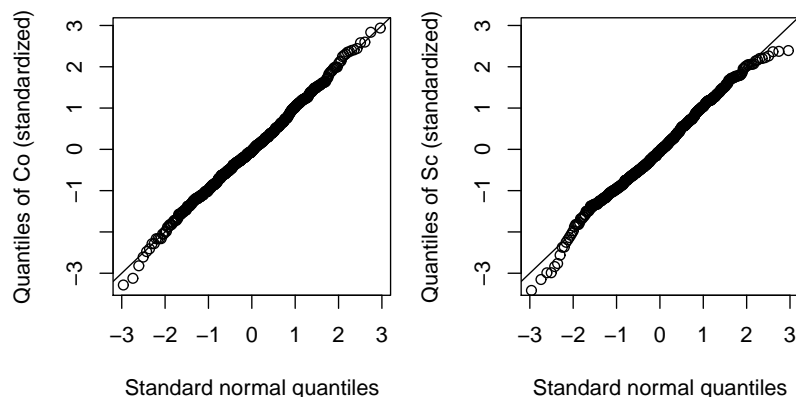


Figure 2. Normal QQ-plots for Co and Sc log-concentrations

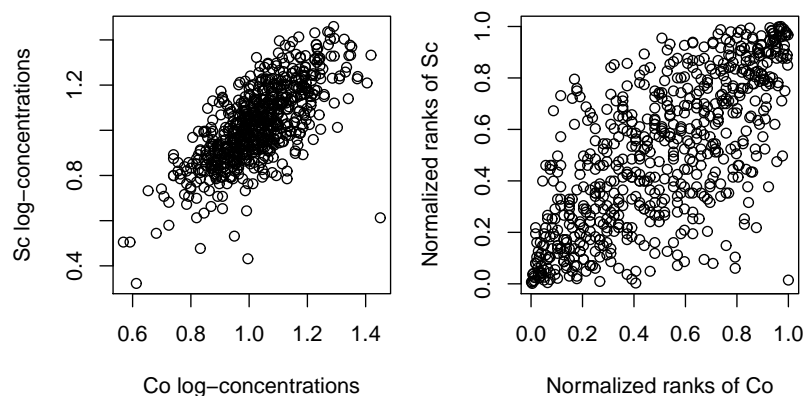


Figure 3. Scatter plots for Co and Sc log-concentrations (left) and normalized ranks of Co and Sc log-concentrations (right)

Under the assumption of normal margins, we test for four parametric copula families: Clayton, Frank, Gumbel, and the two-parameter family described in (21), which we will call the *Cook-Johnson copula*. Recall that the bivariate Frank family of copulas is given by

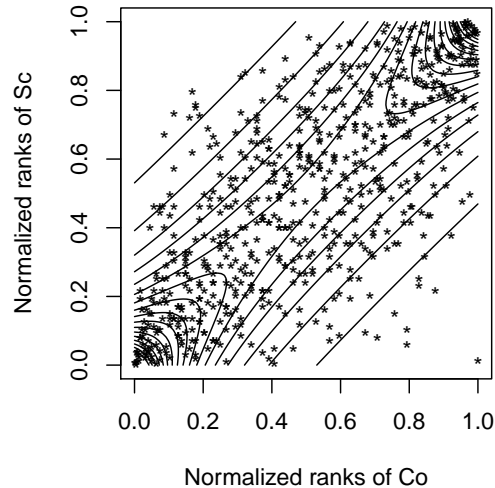
$$C_\lambda(u, v) = -\frac{1}{\lambda} \log \left( 1 + \frac{(e^{-\lambda u} - 1)(e^{-\lambda v} - 1)}{e^{-\lambda} - 1} \right), \quad \lambda \in \mathbb{R},$$

with  $C_0(u, v) = uv$ . Following the methodology described in Subsection 6.1, we compute the test process  $W_n$  and the corresponding test statistics  $\kappa_n$  and  $\omega_n^2$  for each of these four models. For the observed values of the test statistics, we compute  $p$ -values from the benchmark distribution tables constructed from the true bivariate standard Wiener process. The results are given in Table 4. We observe that all models except Frank are rejected at 5% significance level, and Clayton and Cook-Johnson models in particular are rejected very strongly.

**Table 4.**  $p$ -values for various copula models for the Co-Sc log-concentrations, under assumption of marginal normality

	Clayton	Frank	Gumbel	C-J
$\kappa_n$	0.0000	0.1664	0.0318	0.0000
$\omega_n^2$	0.0000	0.1281	0.0278	0.0000

The model with normal margins and Frank copula yields maximum likelihood estimates of  $\hat{\mu}_{Co} = 1.025$  and  $\hat{\sigma}_{Co} = 0.136$  for the mean and standard deviation of Co log-concentrations,  $\hat{\mu}_{Sc} = 1.021$  and  $\hat{\sigma}_{Sc} = 0.178$  for the mean and standard deviation of Sc log-concentrations, and  $\hat{\lambda} = 6.589$  for the Frank copula parameter. The estimated value of  $\lambda$  suggests moderate positive dependence, corresponding to a Kendall's  $\tau$  of 0.544 and Spearman's  $\rho$  of 0.743. Direct sample estimates for these coefficients are 0.535 and 0.718, respectively. The contour plot of the fitted Frank copula density, together with the scatter plot of the rank-transformed data, can be seen in Fig. 4.



**Figure 4.** Contour plot of the fitted Frank copula density, superimposed on the scatter plot of the rank-transformed Co and Sc log-concentrations

## 7. Proofs

The proofs of Theorems 2.1 and 3.1 can be found in an online appendix. We present the proofs of Theorems 4.1 and 5.3 below.

### Proof of Theorem 4.1

By Theorem 2.1 and Skorohod's representation theorem (Billingsley (1999), Theorem 6.7), there is a probability space where probabilistically equivalent versions of  $\eta_n$  and  $\eta$  are defined, and these satisfy  $\|\eta_n - \eta\|_{[\delta, 1-\delta]^d} \rightarrow 0$  a.s., with  $\|\cdot\|_S := \sup_S |\cdot|$  for  $S \subset [0, 1]^d$ . We will show that in this probability space,

$$\|W_n - W\|_{[0, 1]^d} \xrightarrow{P} 0, \quad (22)$$

with  $W$  as defined in (13). In view of Theorem 3.1, this will suffice for the proof.

Throughout the proof, we will let  $A_\delta(\mathbf{u})$  denote the set  $[\boldsymbol{\delta}, \boldsymbol{\delta} + (1 - 2\delta)\mathbf{u}]$  for  $\mathbf{u} \in [0, 1]^d$ . Note that (22) will follow from

$$\left\| \int_{A_\delta(\mathbf{u})} \frac{1}{\sqrt{c_{\hat{\chi}}(\mathbf{s})}} d\eta_n(\mathbf{s}) - \int_{A_\delta(\mathbf{u})} \frac{1}{\sqrt{c(\mathbf{s})}} d\eta(\mathbf{s}) \right\|_{[0, 1]^d} \xrightarrow{P} 0 \quad (23)$$

and

$$\begin{aligned} & \left\| \int_{A_\delta(\mathbf{u})} \hat{\mathbf{k}}(\mathbf{s})^\top \left( \hat{\mathbf{I}}_\delta^{-1}(s_d) \int_{S_\delta(s_d)} \hat{\mathbf{k}}(\mathbf{s}') d\eta_n(\mathbf{s}') \right) \sqrt{c_{\hat{\chi}}(\mathbf{s})} ds \right. \\ & \quad \left. - \int_{A_\delta(\mathbf{u})} \mathbf{k}(\mathbf{s})^\top \left( \mathbf{I}_\delta^{-1}(s_d) \int_{S_\delta(s_d)} \mathbf{k}(\mathbf{s}') d\eta(\mathbf{s}') \right) \sqrt{c(\mathbf{s})} ds \right\|_{[0, 1]^d} \xrightarrow{P} 0. \end{aligned} \quad (24)$$

We will prove (23) first. Let  $\Delta_n := \eta_n - \eta$ . Then (23) will follow from

$$\left\| \int_{A_\delta(\mathbf{u})} \Delta\gamma(\mathbf{s}) d\eta(\mathbf{s}) \right\|_{[0, 1]^d} \xrightarrow{P} 0, \quad \left\| \int_{A_\delta(\mathbf{u})} \hat{\gamma}(\mathbf{s}) d\Delta_n(\mathbf{s}) \right\|_{[0, 1]^d} \xrightarrow{P} 0. \quad (25)$$

Applying integration by parts (Henstock (1973), Theorem 3) to the first integral term in (25), we obtain the following bound:

$$\begin{aligned} \left| \int_{A_\delta(\mathbf{u})} \Delta\gamma(\mathbf{s}) d\eta(\mathbf{s}) \right| & \leq \sum_{\mathbf{v} \in \mathcal{V}_{A_\delta(\mathbf{u})}} |\Delta\gamma(\mathbf{v})\eta(\mathbf{v})| + \|\eta\|_{A_\delta(\mathbf{u})} V_{A_\delta(\mathbf{u})}^{\text{HK}}(\Delta\gamma) \\ & \leq \|\eta\|_\delta (2^d \|\Delta\gamma\|_\delta + V_\delta^{\text{HK}}(\Delta\gamma)), \end{aligned} \quad (26)$$

where  $\mathcal{V}_{A_\delta(\mathbf{u})}$  denotes the set of the  $2^d$  vertices of the hyperrectangle  $A_\delta(\mathbf{u})$ , and  $\|\cdot\|_\delta$  is short-hand notation for  $\|\cdot\|_{[\delta, 1-\delta]^d}$ . Now, Assumptions A2-A3 ensure that  $\eta$  is continuous (hence bounded) on  $[\delta, 1-\delta]^d$ , A4 ensures that  $|\Delta\gamma|$  is  $o_P(1)$  uniformly over  $[\delta, 1-\delta]^d$ ,

and A6 ensures that  $V_\delta^{\text{HK}}(\Delta\gamma)$  is  $o_P(1)$  as well. It follows that the far right-hand side of (26) vanishes in probability, and the first convergence in (25) is proved. The second convergence in (25) follows from a similar integration by parts argument:

$$\left| \int_{A_\delta(\mathbf{u})} \widehat{\gamma}(\mathbf{s}) d\Delta_n(\mathbf{s}) \right| \leq \|\Delta_n\|_\delta (2^d \|\widehat{\gamma}\|_\delta + V_\delta^{\text{HK}}(\widehat{\gamma})),$$

where the right-hand side is  $o_P(1)$  since  $\|\Delta_n\|_\delta$  is  $o_P(1)$  and  $\|\widehat{\gamma}\|_\delta$  as well as  $V_\delta^{\text{HK}}(\widehat{\gamma})$  are  $O_P(1)$  terms.

We have thus established (23), and it remains to prove (24). For ease of notation, we let

$$\begin{aligned} H(\mathbf{s}) &= \mathbf{k}(\mathbf{s})^\top \mathbf{I}_\delta^{-1}(s_d) \int_{S_\delta(s_d)} \mathbf{k}(\mathbf{s}') d\eta(\mathbf{s}'), \\ H_n(\mathbf{s}) &= \mathbf{k}(\mathbf{s})^\top \mathbf{I}_\delta^{-1}(s_d) \int_{S_\delta(s_d)} \mathbf{k}(\mathbf{s}') d\eta_n(\mathbf{s}'), \\ \widehat{H}(\mathbf{s}) &= \widehat{\mathbf{k}}(\mathbf{s})^\top \widehat{\mathbf{I}}_\delta^{-1}(s_d) \int_{S_\delta(s_d)} \widehat{\mathbf{k}}(\mathbf{s}') d\eta(\mathbf{s}'), \\ \widehat{H}_n(\mathbf{s}) &= \widehat{\mathbf{k}}(\mathbf{s})^\top \widehat{\mathbf{I}}_\delta^{-1}(s_d) \int_{S_\delta(s_d)} \widehat{\mathbf{k}}(\mathbf{s}') d\eta_n(\mathbf{s}'). \end{aligned}$$

Then (24) can be written succinctly as

$$\left\| \int_{A_\delta(\mathbf{u})} \left( \widehat{H}_n(\mathbf{s}) \sqrt{c_{\widehat{\boldsymbol{\lambda}}}(\mathbf{s})} - H(\mathbf{s}) \sqrt{c(\mathbf{s})} \right) d\mathbf{s} \right\|_{[0,1]^d} \xrightarrow{P} 0,$$

which can be proved by showing

$$\|H(\sqrt{c_{\widehat{\boldsymbol{\lambda}}}(\mathbf{s})} - \sqrt{c(\mathbf{s})})\|_\delta \xrightarrow{P} 0, \quad \|(\widehat{H}_n - H) \sqrt{c_{\widehat{\boldsymbol{\lambda}}}(\mathbf{s})}\|_\delta \xrightarrow{P} 0. \quad (27)$$

The first convergence in (27) follows easily from the continuity (hence boundedness) of  $H$  over  $[\delta, 1-\delta]^d$  and the continuity of  $\sqrt{c_{\widehat{\boldsymbol{\lambda}}}(\mathbf{u})}$  over  $(\mathbf{u}, \boldsymbol{\lambda}) \in [\delta, 1-\delta]^d \times \Lambda$ . As for the second convergence in (27), since  $\|\sqrt{c_{\widehat{\boldsymbol{\lambda}}}(\mathbf{s})}\|_\delta = O_P(1)$ , we need to show that  $\|\widehat{H}_n - H\|_\delta \xrightarrow{P} 0$ . We will do this by proving

$$\|H_n - H\|_\delta \xrightarrow{P} 0, \quad \|\widehat{H}_n - H_n\|_\delta \xrightarrow{P} 0. \quad (28)$$

Consider the first convergence in (28). We have

$$\|H_n - H\|_\delta = \left\| \mathbf{k}(\mathbf{s})^\top \mathbf{I}_\delta^{-1}(s_d) \int_{S_\delta(s_d)} \mathbf{k}(\mathbf{s}') d\Delta_n(\mathbf{s}') \right\|_\delta,$$

with  $\Delta_n = \eta_n - \eta$ , as before. The term  $|\mathbf{k}(\mathbf{s})^\top \mathbf{I}_\delta^{-1}(s_d)|$  is component-wise bounded on  $[\delta, 1-\delta]^d$  by continuity, so we need to show that

$$\sup_{t \in [\delta, 1-\delta]} \left| \int_{S_\delta(t)} k_i(\mathbf{s}') d\Delta_n(\mathbf{s}') \right| \xrightarrow{P} 0, \quad i = 1, \dots, 1 + dm + p. \quad (29)$$

Applying integration by parts as before, we obtain

$$\left| \int_{S_\delta(t)} k_i(\mathbf{s}') d\Delta_n(\mathbf{s}') \right| \leq \|\Delta_n\|_\delta (2^d \|k_i\|_\delta + V_\delta^{\text{HK}}(k_i)),$$

where the right-hand side is  $o_P(1)$  since  $\|k_i\|_\delta < \infty$ ,  $V_\delta^{\text{HK}}(k_i) < \infty$  and  $\|\Delta_n\|_\delta = o_P(1)$ . Hence (29) is established and it remains to prove the second convergence in (28).

By virtue of the first convergence in (28), and an analogous result for  $\widehat{H}_n$  and  $\widehat{H}$ , it will suffice to prove  $\|\widehat{H} - H\|_\delta \xrightarrow{P} 0$ . Note that

$$\begin{aligned} |\widehat{H}(\mathbf{s}) - H(\mathbf{s})| &\leq \left| \widehat{\mathbf{k}}(\mathbf{s})^\top \widehat{\mathbf{I}}_\delta^{-1}(s_d) - \mathbf{k}(\mathbf{s})^\top \mathbf{I}_\delta^{-1}(s_d) \right| \cdot \left| \int_{S_\delta(s_d)} \mathbf{k}(\mathbf{s}') d\eta(\mathbf{s}') \right| \\ &\quad + \left| \widehat{\mathbf{k}}(\mathbf{s})^\top \widehat{\mathbf{I}}_\delta^{-1}(s_d) \right| \cdot \left| \int_{S_\delta(s_d)} (\widehat{\mathbf{k}}(\mathbf{s}') - \mathbf{k}(\mathbf{s}')) d\eta(\mathbf{s}') \right|, \end{aligned} \quad (30)$$

where absolute values should be interpreted component-wise, as usual. Consider the first term on the right-hand side of (30). Since the mapping

$$(\mathbf{s}, \boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_d, \boldsymbol{\lambda}') \mapsto \mathbf{k}(\mathbf{s}, \boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_d, \boldsymbol{\lambda}')$$

is continuous over  $[\delta, 1 - \delta]^d \times \Theta^d \times \Lambda$ , the difference

$$\left| \widehat{\mathbf{k}}(\mathbf{s})^\top \widehat{\mathbf{I}}_\delta^{-1}(s_d) - \mathbf{k}(\mathbf{s})^\top \mathbf{I}_\delta^{-1}(s_d) \right|$$

is  $o_P(1)$  uniformly over  $\mathbf{s} \in [\delta, 1 - \delta]^d$ . Moreover, an integration by parts argument as before yields that

$$\left| \int_{S_\delta(s_d)} k_i(\mathbf{s}') d\eta(\mathbf{s}') \right| \leq \|\eta\|_\delta (2^d \|k_i\|_\delta + V_\delta^{\text{HK}}(k_i)),$$

for  $i = 1, \dots, 1 + dm + p$ , where the right-hand side is  $O_P(1)$ . So the first summand on the right-hand side of (30) is  $o_P(1)$  uniformly over  $\mathbf{s} \in [\delta, 1 - \delta]^d$ . The second summand there can be handled similarly: the term  $\left| \widehat{\mathbf{k}}(\mathbf{s})^\top \widehat{\mathbf{I}}_\delta^{-1}(s_d) \right|$  is  $O_P(1)$ , and integration by parts yields

$$\left| \int_{S_\delta(s_d)} \Delta k_i(\mathbf{s}') d\eta(\mathbf{s}') \right| \leq \|\eta\|_\delta (2^d \|\Delta k_i\|_\delta + V_\delta^{\text{HK}}(\Delta k_i))$$

for  $i = 1, \dots, 1 + dm + p$ , where the right-hand side is  $o_P(1)$ .

Both convergences in (27) are thereby established, which in turn proves (24).  $\square$

### Proof of Theorem 5.3

We have, from the Cameron-Martin theorem (see, e.g., Theorem 5.1 of Lifshits (2012)), that

$$\begin{aligned} \log\left(\frac{d\tilde{Q}}{dQ}\right) &= (1-2\delta)^{d/2} \int_{[0,1]^d} g(\boldsymbol{\delta} + (1-2\delta)\mathbf{u}) \sqrt{c_{\lambda_0}(\boldsymbol{\delta} + (1-2\delta)\mathbf{u})} dV(\mathbf{u}) \\ &\quad - \frac{(1-2\delta)^d}{2} \int_{[0,1]^d} g^2(\boldsymbol{\delta} + (1-2\delta)\mathbf{u}) c_{\lambda_0}(\boldsymbol{\delta} + (1-2\delta)\mathbf{u}) d\mathbf{u}, \end{aligned}$$

with  $V = W$  under  $Q$  and  $V = \tilde{W}$  under  $\tilde{Q}$ , and  $g$  as defined in (17). This immediately yields

$$\log\left(\frac{d\tilde{Q}}{dQ}\right) \sim \begin{cases} N(-\frac{1}{2}\|g\|^2, \|g\|^2) & \text{under } Q, \\ N(\frac{1}{2}\|g\|^2, \|g\|^2) & \text{under } \tilde{Q}, \end{cases}$$

with  $\|g\|$  as in (20), with the convention that  $g = 0$  outside of  $[\delta, 1-\delta]^d$ . Using the fact that

$$d(\tilde{Q}, Q) = \tilde{Q}[\log(d\tilde{Q}/dQ) > 0] - Q[\log(d\tilde{Q}/dQ) > 0],$$

we also obtain  $d(\tilde{Q}, Q) = \nu(g) = 2\Phi(\frac{1}{2}\|g\|) - 1$ .

Thus it remains to show that  $\|g\| = \|h\|$ , which is equivalent to

$$\begin{aligned} &2 \int_{[\delta, 1-\delta]^d} \mathbf{k}(\mathbf{s})^\top \left( \mathbf{I}_\delta^{-1}(s_d) \int_{S_\delta(s_d)} \mathbf{k}(\mathbf{s}') h(\mathbf{s}') dC_{\lambda_0}(\mathbf{s}') \right) h(\mathbf{s}) dC_{\lambda_0}(\mathbf{s}) \\ &= \int_{[\delta, 1-\delta]^d} \left[ \mathbf{k}(\mathbf{s})^\top \left( \mathbf{I}_\delta^{-1}(s_d) \int_{S_\delta(s_d)} \mathbf{k}(\mathbf{s}') h(\mathbf{s}') dC_{\lambda_0}(\mathbf{s}') \right) \right]^2 dC_{\lambda_0}(\mathbf{s}). \end{aligned} \tag{31}$$

For ease of notation, let  $E_1$  and  $E_2$  denote the left- and right-hand sides of (31), respectively. Also let  $S_\delta^-(t) = [\delta, 1-\delta/2]^d \setminus S_\delta(t)$  for  $t \in [\delta, 1-\delta/2)$  and define

$$\mathbf{H}(t) = \int_{S_\delta^-(t)} \mathbf{k}(\mathbf{s}) h(\mathbf{s}) dC_{\lambda_0}(\mathbf{s}), \quad t \in [\delta, 1-\delta/2).$$

We have

$$\begin{aligned} E_1 &= 2 \int_{[\delta, 1-\delta]^d} \left[ \int_{S_\delta(s_d)} \mathbf{k}(\mathbf{s}')^\top h(\mathbf{s}') dC_{\lambda_0}(\mathbf{s}') \right] \mathbf{I}_\delta^{-1}(s_d) \mathbf{k}(\mathbf{s}) h(\mathbf{s}) dC_{\lambda_0}(\mathbf{s}) \\ &= -2 \int_{[\delta, 1-\delta]^d} \left[ \int_{S_\delta^-(s_d)} \mathbf{k}(\mathbf{s}')^\top h(\mathbf{s}') dC_{\lambda_0}(\mathbf{s}') \right] \mathbf{I}_\delta^{-1}(s_d) \mathbf{k}(\mathbf{s}) h(\mathbf{s}) dC_{\lambda_0}(\mathbf{s}) \\ &= -2 \int_\delta^{1-\delta} \left[ \int_{S_\delta^-(s_d)} \mathbf{k}(\mathbf{s}')^\top h(\mathbf{s}') dC_{\lambda_0}(\mathbf{s}') \right] \mathbf{I}_\delta^{-1}(s_d) \\ &\quad \times \left[ \int_{[\delta, 1-\delta]^{d-1}} \mathbf{k}(\mathbf{s}) h(\mathbf{s}) c_{\lambda_0}(\mathbf{s}) ds_1 \dots ds_{d-1} \right] ds_d \end{aligned}$$

$$= -2 \int_{\delta}^{1-\delta} \mathbf{H}(s_d)^\top \mathbf{I}_{\delta}^{-1}(s_d) d\mathbf{H}(s_d),$$

where the second equality above follows from

$$\int_{[\delta, 1-\delta/2]^d} \mathbf{k}(\mathbf{s}) h(\mathbf{s}) dC_{\lambda_0}(\mathbf{s}) = \mathbf{0},$$

which is a consequence of Assumption B0. Now, denoting  $\mathbf{G}(t) = \mathbf{I}_{\delta}^{-1}(t)\mathbf{H}(t)$  and applying integration by parts, we obtain

$$E_1 = -2 \int_{\delta}^{1-\delta} \mathbf{G}(s_d)^\top d\mathbf{H}(s_d) = 2 \int_{\delta}^{1-\delta} \mathbf{H}(s_d)^\top d\mathbf{G}(s_d).$$

By the product rule of differentiation, we can write

$$d\mathbf{G}(s_d) = [(\mathbf{I}_{\delta}^{-1})'(s_d)\mathbf{H}(s_d) + \mathbf{I}_{\delta}^{-1}(s_d)\mathbf{H}'(s_d)] ds_d,$$

where derivatives should be interpreted component-wise. Using the identity

$$(\mathbf{I}_{\delta}^{-1})'(s_d) = -\mathbf{I}_{\delta}^{-1}(s_d)\mathbf{I}'_{\delta}(s_d)\mathbf{I}_{\delta}^{-1}(s_d),$$

we obtain

$$\begin{aligned} E_1 &= -2 \int_{\delta}^{1-\delta} \mathbf{H}(s_d)^\top \mathbf{I}_{\delta}^{-1}(s_d)\mathbf{I}'_{\delta}(s_d)\mathbf{I}_{\delta}^{-1}(s_d)\mathbf{H}(s_d) ds_d \\ &\quad + 2 \int_{\delta}^{1-\delta} \mathbf{H}(s_d)^\top \mathbf{I}_{\delta}^{-1}(s_d)\mathbf{H}'(s_d) ds_d \\ &= -2 \int_{\delta}^{1-\delta} \mathbf{H}(s_d)^\top \mathbf{I}_{\delta}^{-1}(s_d)\mathbf{I}'_{\delta}(s_d)\mathbf{I}_{\delta}^{-1}(s_d)\mathbf{H}(s_d) ds_d - E_1. \end{aligned}$$

It follows that

$$\begin{aligned} E_1 &= - \int_{\delta}^{1-\delta} \mathbf{H}(s_d)^\top \mathbf{I}_{\delta}^{-1}(s_d)\mathbf{I}'_{\delta}(s_d)\mathbf{I}_{\delta}^{-1}(s_d)\mathbf{H}(s_d) ds_d \\ &= \int_{\delta}^{1-\delta} \mathbf{H}(s_d)^\top \mathbf{I}_{\delta}^{-1}(s_d) \left[ \int_{[\delta, 1-\delta/2]^{d-1}} \mathbf{k}(\mathbf{s})\mathbf{k}(\mathbf{s})^\top c_{\lambda_0}(\mathbf{s}) ds_1 \dots ds_{d-1} \right] \\ &\quad \cdot \mathbf{I}_{\delta}^{-1}(s_d)\mathbf{H}(s_d) ds_d \\ &= \int_{[\delta, 1-\delta]^d} \mathbf{H}(s_d)^\top \mathbf{I}_{\delta}^{-1}(s_d)\mathbf{k}(\mathbf{s})\mathbf{k}(\mathbf{s})^\top \mathbf{I}_{\delta}^{-1}(s_d)\mathbf{H}(s_d) dC_{\lambda_0}(\mathbf{s}) \\ &= \int_{[\delta, 1-\delta]^d} \left[ \mathbf{k}(\mathbf{s})^\top \mathbf{I}_{\delta}^{-1}(s_d)\mathbf{H}(s_d) \right]^2 dC_{\lambda_0}(\mathbf{s}) \\ &= E_2, \end{aligned}$$

as desired. Thus (31) is established.  $\square$

## Acknowledgements

We are very grateful to an editor and an anonymous referee for the insightful comments and suggestions, and to Prof. Estate V. Khmaladze for stimulating discussions about this paper. We are also grateful to the participants of the Eurandom-ISI Workshop on Actuarial and Financial Statistics, the 10<sup>th</sup> Extreme Value Analysis Conference, the 4<sup>th</sup> Conference of the International Society for Nonparametric Statistics, as well as seminar participants at Tilburg and Bocconi Universities, for their comments and suggestions. Can and Laeven were supported in part by the Netherlands Organization for Scientific Research under an NWO-Vidi grant. Einmahl holds the Arie Kapteyn Chair 2019-2022 and gratefully acknowledges the corresponding research support.

## References

- BERG, D. (2009). Copula goodness-of-fit testing: an overview and power comparison. *Eur. J. Finance* **15** 675–701.
- BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, second ed. Wiley.
- BREYMAN, W., DIAS, A. and EMBRECHTS, P. (2003). Dependence structures for multivariate high-frequency data in finance. *Quant. Finance* **3** 1–14.
- CAN, S. U., EINMAHL, J. H. J., KHMALADZE, E. V. and LAEVEN, R. J. A. (2015). Asymptotically distribution-free goodness-of-fit testing for tail copulas. *Ann. Statist.* **43** 878–902.
- CHARPENTIER, A., FERMANIAN, J.-D. and SCAILLET, O. (2007). The estimation of copulas: theory and practice. In *Copulas: From Theory to Application in Finance* (J. Rank, ed.) 35–60. Risk Books, London.
- CHOROŚ, B., IBRAGIMOV, R. and PERMIAKOVA, E. (2010). Copula estimation. In *Copula Theory and Its Applications* (P. Jaworski, F. Durante, W. K. Härdle and T. Rychlik, eds.) 77–91. Springer, Heidelberg.
- COOK, R. D. and JOHNSON, M. E. (1986). Generalized Burr-Pareto-logistic distributions with applications to a uranium exploration data set. *Technometrics* **28** 123–131.
- DEHEUVELS, P. (1979). La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance. *Acad. Roy. Belg. Bull. Cl. Sci. (5)* **65** 274–292.
- DEHEUVELS, P. (1981). A Kolmogorov-Smirnov type test for independence and multivariate samples. *Rev. Roumaine Math. Pures Appl.* **26** 213–226.
- DELGADO, M. A., HIDALGO, J. and VELASCO, C. (2005). Distribution free goodness-of-fit tests for linear processes. *Ann. Statist.* **33** 2568–2609.
- DETTE, H. and HETZLER, B. (2009). Khmaladze transformation of integrated variance processes with applications to goodness-of-fit testing. *Math. Methods Statist.* **18** 98–116.
- DOBRIĆ, J. and SCHMID, F. (2007). A goodness of fit test for copulas based on Rosenblatt's transformation. *Comput. Statist. Data Anal.* **51** 4633–4642.
- DURANTE, F. and SEMPI, C. (2016). *Principles of Copula Theory*. CRC Press, Boca Raton, FL.

- EINMAHL, J. H. J. and KHMALADZE, E. V. (2001). *The two-sample problem in  $\mathbb{R}^m$  and measure-valued martingales*. In *State of the Art in Probability and Statistics. Lecture Notes–Monograph Series* **36** 434–463. Institute of Mathematical Statistics.
- FERMANIAN, J.-D. (2005). Goodness-of-fit tests for copulas. *J. Multivariate Anal.* **95** 119–152.
- FERMANIAN, J.-D. (2013). An overview of the goodness-of-fit test problem for copulas. In *Copulae in Mathematical and Quantitative Finance* (P. Jaworski, F. Durante and W. K. Härdle, eds.) 61–89. Springer, Heidelberg.
- FERMANIAN, J.-D., RADULOVIĆ, D. and WEGKAMP, M. (2004). Weak convergence of empirical copula processes. *Bernoulli* **10** 847–860.
- GAENSSLER, P. and STUTE, W. (1987). *Seminar on empirical processes. DMV Seminar* **9**. Birkhäuser Verlag, Basel.
- GENEST, C., GHOUDI, K. and RIVEST, L. P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82** 543–552.
- GENEST, C., QUESSY, J.-F. and RÉMILLARD, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scand. J. Statist.* **33** 337–366.
- GENEST, C., RÉMILLARD, B. and BEAUDOIN, D. (2009). Goodness-of-fit tests for copulas: a review and a power study. *Insurance Math. Econom.* **44** 199–213.
- GENEST, C. and RIVEST, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *J. Amer. Statist. Assoc.* **88** 1034–1043.
- HENSTOCK, R. (1973). Integration by parts. *Aeq. Math.* **9** 1–18.
- JOE, H. (2015). *Dependence Modeling with Copulas*. CRC Press, Boca Raton, FL.
- KHMALADZE, E. V. (1981). Martingale approach in the theory of goodness-of-fit tests. *Theory Probab. Appl.* **26** 240–257.
- KHMALADZE, E. V. (1988). An innovation approach to goodness-of-fit tests in  $R^m$ . *Ann. Statist.* **16** 1503–1516.
- KHMALADZE, E. V. (1993). Goodness of fit problem and scanning innovation martingales. *Ann. Statist.* **21** 798–829.
- KHMALADZE, E. V. and KOUL, H. L. (2004). Martingale transforms goodness-of-fit tests in regression models. *Ann. Statist.* **32** 995–1034.
- KHMALADZE, E. V. and KOUL, H. L. (2009). Goodness-of-fit problem for errors in nonparametric regression: Distribution free approach. *Ann. Statist.* **37** 3165–3185.
- KOENKER, R. and XIAO, Z. (2002). Inference on the quantile regression process. *Econometrica* **70** 1583–1612.
- KOENKER, R. and XIAO, Z. (2006). Quantile autoregression. *J. Amer. Statist. Assoc.* **101** 980–990.
- LIFSHTS, M. (2012). *Lectures on Gaussian Processes*. Springer, Heidelberg.
- LU, X. and ZHENG, X. (2020). A goodness-of-fit test for copulas based on martingale transformation. *J. Econom.* **215** 84–117.
- MCKEAGUE, I. W., NIKABADZE, A. M. and SUN, Y. Q. (1995). An omnibus test for independence of a survival time from a covariate. *Ann. Statist.* **23** 450–475.
- NELSEN, R. B. (2006). *An Introduction to Copulas*. Springer-Verlag New York, Inc.

- NEUHAUS, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. *Ann. Math. Statist.* **42** 1285–1295.
- NIKABADZE, A. M. and STUTE, W. (1997). Model checks under random censorship. *Statist. Probab. Lett.* **32** 249–259.
- OOSTERHOFF, J. and VAN ZWET, W. R. (1979). A note on contiguity and Hellinger distance. In *Contributions to Statistics* 157–166. Reidel, Dordrecht-Boston, Mass.-London.
- OWEN, A. B. (2005). Multidimensional variation for quasi-Monte Carlo. In *Contemporary Multivariate Analysis and Design of Experiments* (J. Fan and G. Li, eds.) 49–74. World Scientific Publishing.
- RÜSCHENDORF, L. (1976). Asymptotic distributions of multivariate rank order statistics. *Ann. Statist.* **4** 912–923.
- RUYMGAART, F. H. (1973). *Asymptotic Theory of Rank Tests for Independence*. Mathematisch Centrum, Amsterdam Mathematical Centre Tracts, 43.
- SEGBERS, J. (2012). Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli* **18** 764–782.
- SEGBERS, J. (2015). Hybrid copula estimators. *J. Statist. Plann. Inference* **160** 23–34.
- SHIH, J. H. (1998). A goodness-of-fit test for association in a bivariate survival model. *Biometrika* **85** 189–200.
- SHIH, J. H. and LOUIS, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics* **51** 1384–1399.
- SKLAR, M. (1959). Fonctions de répartition à  $n$  dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **8** 229–231.
- STUTE, W., THIES, S. and ZHU, L. (1998). Model checks for regression: an innovation process approach. *Ann. Statist.* **26** 1916–1934.
- TOWGHI, N. (2002). Multidimensional extension of L. C. Young’s inequality. *J. Inequal. Pure Appl. Math.* **3** Art. 22.
- VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge University Press, Cambridge.
- WANG, W. and WELLS, M. T. (2000). Model selection and semiparametric inference for bivariate failure-time data. *J. Amer. Statist. Assoc.* **95** 62–76.