

Two-Sample Testing for Tail Copulas with an Application to Equity Indices

Sami Umut Can

Department of Quantitative Economics, University of Amsterdam

John H.J. Einmahl

Department of Econometrics & OR and CentER, Tilburg University

Roger J.A. Laeven

Department of Quantitative Economics, University of Amsterdam

December 1, 2022

Abstract

A novel, general two-sample hypothesis testing procedure is established for testing the equality of tail copulas associated with bivariate data. More precisely, using a martingale transformation of a natural two-sample tail copula process, a test process is constructed, which is shown to converge in distribution to a standard Wiener process. Hence, from this test process a myriad of asymptotically distribution-free two-sample tests can be obtained. The good finite-sample behavior of our procedure is demonstrated through Monte Carlo simulations. Using the new testing procedure, no evidence of a difference in the respective tail copulas is found for pairs of negative daily log-returns of equity indices during and after the global financial crisis.

Keywords: Tail dependence; tail copula; two-sample testing; financial crisis; distribution-free testing; martingale transformation.

1 Introduction

Measuring dependence between economic variables such as asset returns and, in particular, between their tail events is of crucial importance for risk management, asset pricing, and portfolio choice. The tail dependence structure is, for example, a main ingredient for the computation of high quantiles of aggregate loss distributions, and hence for risk capital requirements, and dictates the extent of idiosyncratic tail risk, and hence the diversification potential, within a collection of assets.

A key question in financial econometrics and international finance is whether the dependence structure between economic variables, in particular asset returns, is constant over time and across markets, regions and institutions, or whether it is subject to variation due to e.g., financial market integration, emerging market developments, or financial contagion. Several papers analyze this question using a wide variety of different methodologies; see e.g., King and Wadhvani (1990), Longin and Solnik (1995), Forbes and Rigobon (2002), Bekaert et al. (2005, 2009), Christoffersen et al. (2012), Bücher et al. (2015), Bormann and Schienle (2020), and the references therein.

In this paper, we focus on the tail dependence structure between two risks, which is completely characterized by the bivariate tail copula. In this context, we develop a general procedure to test for the equality of the tail copulas associated with two bivariate samples. The standard one-sample problem, that is, testing for a specific *parametric* family of tail copulas, has been studied in Can et al. (2015). A challenging feature of the two-sample problem is that the—under the null hypothesis—common tail copula is not specified. This complexity arises naturally as it follows from de Haan and Resnick (1977) that the class of all possible tail copulas is so large that it cannot be represented as a finite-dimensional parametric family, rendering the testing problem *non-parametric*. A pivotal property of our procedure is that, through a martingale transformation, it leads to asymptotically

distribution-free test statistics, so that critical values can be tabulated for universal use and no computer-intensive methods have to be developed and validated.

Our testing approach can be summarized as follows. We first construct two semi-parametric estimators of the tail copulas, one for each sample, consider their suitably normalized difference, and prove that it converges weakly to a non-trivial limiting process under the null hypothesis of tail copula equality. Next, we describe the martingale transformation that transforms this limiting process into a standard bivariate Wiener process. Then, we establish that the empirical counterpart of this transformation, when applied to the empirical process given by the normalized difference of the two tail copula estimators, converges weakly to a standard bivariate Wiener process under the null hypothesis. Finally, two-sample tests for tail copulas can now be conducted by comparing the transformed empirical process to a standard Wiener process. Hence, our approach leads to an asymptotically distribution-free test process and thus, by taking functionals, to asymptotically distribution-free test statistics. In other words, our procedure leads to an entire test *process* from which we can generate a myriad of asymptotically distribution-free tests.

We illustrate the finite-sample performance of our approach through Monte Carlo simulations. These confirm that tests based on our test process have good size and power performances, notwithstanding the difficult nature of the testing problem considered in this paper.

We apply our procedure to test for tail copula equality of equity index returns during and after the global financial crisis. In particular, we analyze the tail dependence structure among the two European equity market indices FTSE 100 (UK) and DAX 30 (Germany) and the two transatlantic equity indices FTSE 100 and S&P 500 (US), during 2008–2011 (sample 1, covering the most volatile days of the global financial crisis and the subsequent European debt crisis) and 2012–2015 (sample 2, covering a “post-crisis” period of equal length). We construct synchronized daily observations of negative log-returns on the basis of

high-frequency data using Coordinated Universal Time (UTC). Computing three commonly used test statistics from our test process, we find that the null hypothesis of tail copula equality between the two samples is not rejected, despite a clearly visible change in the marginal distributions of the index returns between the two samples. This finding applies to both the European pair {FTSE 100, DAX 30} and the transatlantic pair {FTSE 100, S&P 500}. That is, although the margins became markedly less volatile from the crisis period to the post-crisis period, no statistical evidence is found for a change in the number of *margin-free joint extremes* for both pairs of equity market indices.

We briefly describe four papers that are related to the testing problem considered in this paper. In Christoffersen et al. (2012), an extensive analysis is conducted of the evolution over time of the regular and tail dependence structures among both developed and emerging financial markets, and the associated diversification potential. Econometrically, a novel dynamic asymmetric copula model of a parametric nature is used to capture asymmetric tail dependence. When assessing tail dependence, attention is restricted to the less informative tail dependence *coefficients*, rather than to tail copula *functions* as in the present paper, and their development over a long period of 20 years. In Bücher and Dette (2013) and Bücher et al. (2015), new statistical methods are developed for testing tail copula equality, tail copula goodness-of-fit testing, and detecting structural breaks in tail dependence. The methods are applied in Bücher et al. (2015) to energy and financial markets data. An important difference between these two papers and the present one is that here convenient distribution-free weak limits of test statistics are obtained, as described above, whereas in Bücher and Dette (2013) and Bücher et al. (2015) computer-intensive multiplier bootstrap procedures are employed. Bormann and Schienle (2020) provides a detailed account of tail dependence asymmetries and inequalities in 90 years of US equity data, using a refined non-parametric testing procedure that compares two tail copulas locally piecewise on disjoint intervals employing multiple testing principles. As the corresponding complex

limiting distributions do not admit closed-form expressions, finite-sample approximations are simulated using multiplier bootstrap procedures, as in Bücher and Dette (2013) and Bücher et al. (2015).

The remainder of this paper is organized as follows. In Section 2 we formalize our testing problem and in Section 3 we introduce our initial testing basis, which compares two semi-parametric tail copula estimators constructed separately from the two samples, and analyze its asymptotic behavior. In Section 4 we describe the martingale transformation and in Section 5 we prove that the empirical counterpart of this transformation applied to our initial testing basis converges weakly to a standard Wiener process. In Section 6 our Monte Carlo simulation analysis is presented. Section 7 describes the results of our empirical analysis of equity indices. Conclusions are given in Section 8. The proofs of our theorems and some additional simulation results are deferred to a supplementary appendix. An R script implementing our testing approach is also provided as a supplementary file.

2 Testing problem

Consider i.i.d. random vectors $(X_1, Y_1), \dots, (X_n, Y_n)$ generated from some bivariate distribution function (df) F and i.i.d. random vectors $(X'_1, Y'_1), \dots, (X'_n, Y'_n)$ independently generated from some bivariate df F' . Let F have marginal dfs F_1, F_2 and let F' have marginal dfs F'_1, F'_2 , so that $X_i \sim F_1, Y_i \sim F_2$ and $X'_i \sim F'_1, Y'_i \sim F'_2$ for $i = 1, \dots, n$.

We assume that the dfs F and F' lie in the respective domains of attraction of bivariate extreme value distributions G and G' . This means that there exist normalizing sequences $a_j(n), a'_j(n) > 0$ and $b_j(n), b'_j(n) \in \mathbb{R}$, for $j = 1, 2$, such that

$$P\left(\frac{\bigvee_{i=1}^n X_i - b_1(n)}{a_1(n)} \leq x, \frac{\bigvee_{i=1}^n Y_i - b_2(n)}{a_2(n)} \leq y\right) \rightarrow G(x, y),$$

$$P\left(\frac{\bigvee_{i=1}^n X'_i - b'_1(n)}{a'_1(n)} \leq x, \frac{\bigvee_{i=1}^n Y'_i - b'_2(n)}{a'_2(n)} \leq y\right) \rightarrow G'(x, y),$$

as $n \rightarrow \infty$, for all continuity points $(x, y) \in \mathbb{R}^2$ of G and G' , respectively. The normalizing sequences are chosen in such a way that the margins G_1, G_2 of G , and the margins G'_1, G'_2 of G' , are in the form of a standard generalized extreme value (GEV) distribution:

$$\begin{aligned} G_j(x) &= \exp\{-(1 + \gamma_j x)^{-1/\gamma_j}\}, \quad 1 + \gamma_j x > 0, \\ G'_j(x) &= \exp\{-(1 + \gamma'_j x)^{-1/\gamma'_j}\}, \quad 1 + \gamma'_j x > 0, \end{aligned} \tag{1}$$

for $j = 1, 2$. The constants $\gamma_1, \gamma_2 \in \mathbb{R}$ and $\gamma'_1, \gamma'_2 \in \mathbb{R}$ are the marginal extreme value (EV) indices associated with F and F' , respectively. Here, and in the rest of the paper, expressions of the form $(1 + \gamma \cdot)^{1/\gamma}$ should be interpreted as $\exp(\cdot)$ when $\gamma = 0$.

The distribution G , hence the asymptotic joint tail behavior of F , is characterized by the marginal EV indices γ_1, γ_2 and the bivariate tail copula R associated with F , which can be defined as

$$\begin{aligned} R(x, y) &:= \lim_{t \rightarrow \infty} tP(F_1(X_1) > 1 - x/t, F_2(Y_1) > 1 - y/t), \\ &(x, y) \in [0, \infty]^2 \setminus \{(\infty, \infty)\}. \end{aligned} \tag{2}$$

The EV indices γ_1, γ_2 specify the margins of G through (1), and the function R specifies the dependence structure between the margins. While R is not a copula itself, it does characterize the copula of G , and hence the tail dependence structure of F . Similarly, the distribution G' is characterized by γ'_1, γ'_2 and the tail copula R' associated with F' , defined analogously to (2). We refer to, e.g., the monographs Kotz and Nadarajah (2000), Beirlant et al. (2004) and de Haan and Ferreira (2006) for further details about multivariate extreme value theory.

Observe that

$$R(1, 1) = \lim_{u \uparrow 1} \frac{P(F_1(X_1) > u, F_2(Y_1) > u)}{1 - u}.$$

For random variables with continuous margins, $R(1, 1)$ is often referred to as the *upper tail dependence coefficient* and denoted by λ_U ; see e.g., Joe (2001), p. 33. It measures the

probability of a “large rank” of X_1 conditionally upon a “large rank” of Y_1 . Indeed,

$$\lambda_U = \lim_{u \uparrow 1} P(F_1(X_1) > u \mid F_2(Y_1) > u) = R(1, 1).$$

We also note here that the function R generates a σ -finite measure R on the Borel subsets of $[0, \infty]^2 \setminus \{(\infty, \infty)\}$ via the identity

$$R([0, x] \times [0, y]) := R(x, y), \quad (x, y) \in [0, \infty]^2 \setminus \{(\infty, \infty)\}.$$

In this paper, we develop a general procedure to construct two-sample tests for the equality of the tail copulas R and R' . In other words, we assume that we have two random samples independently generated from F and F' , and provide a procedure that can be used to test the null hypothesis $R = R'$ against the alternative $R \neq R'$, where R and R' remain unspecified. Note that, while the bivariate extreme value setting is of a semi-parametric nature, both the null and alternative hypotheses we consider in this paper are non-parametric.

3 Comparing two tail copula estimators

Suppose we have a random sample $(X_1, Y_1), \dots, (X_n, Y_n)$ from F and an independent random sample $(X'_1, Y'_1), \dots, (X'_{n'}, Y'_{n'})$ from F' . Throughout Sections 3–5 we will assume that the null hypothesis holds, i.e., R and R' are the same tail copula, which we will refer to as R . In the present section, we define two semi-parametric estimators for the tail copula R , computed separately from the two available samples, and we describe the asymptotic behavior of the difference between these two estimators as the sample sizes tend to infinity.

From (2), one can verify that $R(x, y) = \lim_{t \rightarrow \infty} R_{(t)}(x, y)$ for $(x, y) \in [0, \infty)^2$, with

$$R_{(t)}(x, y) = tP(\tilde{X}_1(t) \leq x, \tilde{Y}_1(t) \leq y), \quad (x, y) \in [0, \infty)^2, \quad (3)$$

where

$$\begin{aligned}\tilde{X}_i(t) &= \left[\left(1 + \gamma_1 \frac{X_i - b_1(t)}{a_1(t)} \right) \vee 0 \right]^{-1/\gamma_1}, \quad 1 \leq i \leq n, \\ \tilde{Y}_i(t) &= \left[\left(1 + \gamma_2 \frac{Y_i - b_2(t)}{a_2(t)} \right) \vee 0 \right]^{-1/\gamma_2}, \quad 1 \leq i \leq n.\end{aligned}\tag{4}$$

We also define $\tilde{X}'_i(t), \tilde{Y}'_i(t)$ analogously to (4) and $R'_{(t)}$ analogously to (3). We will use the functions $R_{(t)}$ and $R'_{(t)}$ as a basis for estimating R from the two available samples. To that end, we let $k = k(n)$ and $k' = k'(n')$ denote intermediate sequences satisfying $k \uparrow \infty$, $k/n \downarrow 0$ as $n \uparrow \infty$ and $k' \uparrow \infty$, $k'/n' \downarrow 0$ as $n' \uparrow \infty$. For notational brevity, we will write R_n for $R_{(n/k)}$ and $R'_{n'}$ for $R'_{(n'/k')}$.

We estimate R_n and hence R by replacing the unknown quantities $a_j(n/k)$, $b_j(n/k)$ and γ_j in (4) by appropriate estimators $\hat{a}_j(n/k)$, $\hat{b}_j(n/k)$ and $\hat{\gamma}_j$, for $j = 1, 2$, and the probability P by the corresponding empirical measure. We define, therefore,

$$\begin{aligned}\hat{X}_i(n/k) &= \left[\left(1 + \hat{\gamma}_1 \frac{X_i - \hat{b}_1(n/k)}{\hat{a}_1(n/k)} \right) \vee 0 \right]^{-1/\hat{\gamma}_1}, \quad 1 \leq i \leq n, \\ \hat{Y}_i(n/k) &= \left[\left(1 + \hat{\gamma}_2 \frac{Y_i - \hat{b}_2(n/k)}{\hat{a}_2(n/k)} \right) \vee 0 \right]^{-1/\hat{\gamma}_2}, \quad 1 \leq i \leq n,\end{aligned}\tag{5}$$

as an empirical analogue to (4), and we introduce the semi-parametric estimator

$$\hat{R}_n(x, y) = \frac{1}{k} \sum_{i=1}^n \mathbb{1}_{\{\hat{X}_i(n/k) \leq x, \hat{Y}_i(n/k) \leq y\}}, \quad (x, y) \in [0, \infty)^2,\tag{6}$$

for the tail copula R . The random variables $\hat{X}'_i(n'/k')$, $\hat{Y}'_i(n'/k')$ and the estimator $\hat{R}'_{n'}(x, y)$ are defined analogously to (5) and (6), respectively.

Throughout, we fix δ and T such that $0 < \delta < T < \infty$. For later reference, we also introduce the process

$$T_n(x, y) = \frac{1}{k} \sum_{i=1}^n \mathbb{1}_{\{\tilde{X}_i(n/k) \leq x, \tilde{Y}_i(n/k) \leq y\}}, \quad (x, y) \in [0, \infty)^2,\tag{7}$$

and the analogously defined $T'_{n'}$. It is known, by Einmahl et al. (1997, Lemma 3.1), that the weak convergence

$$(\sqrt{k}(T_n - R_n), \sqrt{k'}(T'_{n'} - R'_{n'})) \Rightarrow (V_R, V'_R)$$

holds in the product space $D([\delta/2, T+1]^2) \times D([\delta/2, T+1]^2)$ as $n, n' \rightarrow \infty$. Here, $D([\delta/2, T+1]^2)$ denotes the Skorokhod space of functions with domain $[\delta/2, T+1]^2$, and V_R, V'_R are independent R -Wiener processes on $[0, \infty)^2$, i.e., independent zero-mean Gaussian processes with covariance structure

$$E[V_R(s, t)V_R(x, y)] = E[V'_R(s, t)V'_R(x, y)] = R(s \wedge x, t \wedge y).$$

Henceforth, we will omit the arguments (n/k) and (n'/k') where appropriate, for ease of notation.

Now, the estimators \widehat{R}_n and $\widehat{R}'_{n'}$ estimate the same tail copula R from two different samples, while they would estimate different tail copulas under the alternative hypothesis. So the normalized difference between \widehat{R}_n and $\widehat{R}'_{n'}$ is a natural starting point for a two-sample test. Thus we define $\kappa := kk'/(k+k')$ and

$$\eta_{n,n'}(x, y) := \sqrt{\kappa}[\widehat{R}_n(x, y) - \widehat{R}'_{n'}(x, y)], \quad (x, y) \in [0, \infty)^2.$$

Our first result will establish the asymptotic behavior of $\eta_{n,n'}$ as $n, n' \rightarrow \infty$. We first state the necessary assumptions and definitions.

A1. For some 6-variate random vector $(A_1, A_2, B_1, B_2, \Gamma_1, \Gamma_2)$, we have the joint weak convergence

$$\begin{aligned} \sqrt{k} \left(T_n - R_n, \frac{\widehat{a}_1}{a_1} - 1, \frac{\widehat{a}_2}{a_2} - 1, \frac{\widehat{b}_1 - b_1}{a_1}, \frac{\widehat{b}_2 - b_2}{a_2}, \widehat{\gamma}_1 - \gamma_1, \widehat{\gamma}_2 - \gamma_2 \right) \\ \Rightarrow (V_R, A_1, A_2, B_1, B_2, \Gamma_1, \Gamma_2) \end{aligned} \quad (8)$$

as $n \rightarrow \infty$, in $D([\delta/2, T+1]^2) \times \mathbb{R}^6$. Similarly, for some 6-variate random vector $(A'_1, A'_2, B'_1, B'_2, \Gamma'_1, \Gamma'_2)$, we have the joint weak convergence

$$\begin{aligned} \sqrt{k'} \left(T'_{n'} - R'_{n'}, \frac{\widehat{a}'_1}{a'_1} - 1, \frac{\widehat{a}'_2}{a'_2} - 1, \frac{\widehat{b}'_1 - b'_1}{a'_1}, \frac{\widehat{b}'_2 - b'_2}{a'_2}, \widehat{\gamma}'_1 - \gamma'_1, \widehat{\gamma}'_2 - \gamma'_2 \right) \\ \Rightarrow (V'_R, A'_1, A'_2, B'_1, B'_2, \Gamma'_1, \Gamma'_2) \end{aligned} \quad (9)$$

as $n' \rightarrow \infty$, in $D([\delta/2, T+1]^2) \times \mathbb{R}^6$.

Assumption A1 is formulated such that it allows for flexibility in choosing the estimators. It is fulfilled for, e.g., the moment estimators of γ_j , a_j , b_j , γ'_j , a'_j , and b'_j , provided that k and k' are chosen appropriately; see de Haan and Ferreira (2006), Sections 4.2 and 3.5.

A2. The partial derivatives $R^{(1)}(x, y) := \partial R / \partial x$ and $R^{(2)}(x, y) := \partial R / \partial y$ exist and are continuous on $(0, \infty)^2$.

A3. (i) For some $\phi, \phi' > 0$, as $t \rightarrow \infty$,

$$\begin{aligned} \sup_{(x,y) \in [\delta/2, T+1]^2} |R_{(t)}(x, y) - R(x, y)| &= O(t^{-\phi}), \\ \sup_{(x,y) \in [\delta/2, T+1]^2} |R'_{(t)}(x, y) - R(x, y)| &= O(t^{-\phi'}). \end{aligned}$$

(ii) The sequences k and k' satisfy

$$k = o(n^{2\phi/(2\phi+1)}) \text{ as } n \rightarrow \infty, \quad k' = o((n')^{2\phi'/(2\phi'+1)}) \text{ as } n' \rightarrow \infty,$$

and $k'/(k+k') \rightarrow c$ for some $c \in (0, 1)$ as $n, n' \rightarrow \infty$.

Finally, for $x > 0$ and $\gamma \in \mathbb{R}$, we define the following functions:

$$\begin{aligned} f(x, \gamma) &= \begin{cases} \frac{x(x^\gamma-1)}{\gamma} & \gamma \neq 0 \\ x \log x & \gamma = 0 \end{cases}, & g(x, \gamma) &= -x^{\gamma+1}, \\ h(x, \gamma) &= \begin{cases} \frac{x(1-x^\gamma)}{\gamma^2} + \frac{x \log x}{\gamma} & \gamma \neq 0 \\ -(x \log^2 x)/2 & \gamma = 0 \end{cases}. \end{aligned} \tag{10}$$

We are now prepared to state the convergence result for $\eta_{m, n'}$.

Theorem 3.1. *If Assumptions A1–A3 hold, then*

$$\begin{aligned} \eta_{m, n'}(x, y) &\Rightarrow \sqrt{c} V_R(x, y) - \sqrt{1-c} V'_R(x, y) \\ &\quad + \sqrt{c} R^{(1)}(x, y) [f(x, \gamma_1) A_1 + g(x, \gamma_1) B_1 + h(x, \gamma_1) \Gamma_1] \end{aligned}$$

$$\begin{aligned}
& + \sqrt{c} R^{(2)}(x, y) [f(y, \gamma_2) A_2 + g(y, \gamma_2) B_2 + h(y, \gamma_2) \Gamma_2] \\
& - \sqrt{1-c} R^{(1)}(x, y) [f(x, \gamma'_1) A'_1 + g(x, \gamma'_1) B'_1 + h(x, \gamma'_1) \Gamma'_1] \\
& - \sqrt{1-c} R^{(2)}(x, y) [f(y, \gamma'_2) A'_2 + g(y, \gamma'_2) B'_2 + h(y, \gamma'_2) \Gamma'_2] \\
& =: \eta(x, y)
\end{aligned} \tag{11}$$

in $D([\delta, T]^2)$ as $n, n' \rightarrow \infty$.

Remark 3.2. While the normalized difference between \widehat{R}_n and $\widehat{R}'_{n'}$ provides a natural initial basis for two-sample testing, Theorem 3.1 reveals that it is not easily amenable for this purpose: the distribution of η depends upon the true underlying tail copula R , which is not specified by the null hypothesis, as well as on the true, but unknown, values of the parameters $\gamma_1, \gamma_2, \gamma'_1$ and γ'_2 . To obtain a suitable testing basis, we will in the next section describe a martingale transformation that will turn η into a standard process with a distribution that depends neither on R nor on the marginal extreme value indices.

Remark 3.3. The following relationships between the functions f, g, h defined in (10) can be verified, for any $x > 0$ and $\gamma, \gamma' \in \mathbb{R}$:

$$\begin{aligned}
\gamma f(x, \gamma) + g(x, \gamma) &= \gamma' f(x, \gamma') + g(x, \gamma'), \\
\gamma h(x, \gamma) + f(x, \gamma) &= \gamma' h(x, \gamma') + f(x, \gamma').
\end{aligned}$$

Using these identities, the limiting process η appearing in (11) can also be written as

$$\begin{aligned}
\eta(x, y) &= \sqrt{c} V_R(x, y) - \sqrt{1-c} V'_R(x, y) \\
&+ R^{(1)}(x, y) \left[f(x, \gamma_1) \left(\sqrt{c} A_1 + \sqrt{1-c} A'_1 + (\gamma_1 - \gamma'_1) \sqrt{1-c} B'_1 \right) \right. \\
&\quad + g(x, \gamma_1) \left(\sqrt{c} B_1 + \sqrt{1-c} B'_1 \right) \\
&\quad + h(x, \gamma_1) \left(\sqrt{c} \Gamma_1 + \gamma_1 \sqrt{1-c} A'_1 - \gamma_1 \gamma'_1 \sqrt{1-c} B'_1 \right) \\
&\quad \left. - h(x, \gamma'_1) \left(\sqrt{1-c} \Gamma'_1 - \gamma'_1 \sqrt{1-c} A'_1 + (\gamma'_1)^2 \sqrt{1-c} B'_1 \right) \right] \\
&+ R^{(2)}(x, y) \left[f(y, \gamma_2) \left(\sqrt{c} A_2 + \sqrt{1-c} A'_2 + (\gamma_2 - \gamma'_2) \sqrt{1-c} B'_2 \right) \right.
\end{aligned}$$

$$\begin{aligned}
& + g(y, \gamma_2) \left(\sqrt{c} B_2 + \sqrt{1-c} B'_2 \right) \\
& + h(y, \gamma_2) \left(\sqrt{c} \Gamma_2 + \gamma_1 \sqrt{1-c} A'_2 - \gamma_2 \gamma'_2 \sqrt{1-c} B'_2 \right) \\
& - h(y, \gamma'_2) \left(\sqrt{1-c} \Gamma'_2 - \gamma'_2 \sqrt{1-c} A'_2 + (\gamma'_2)^2 \sqrt{1-c} B'_2 \right) \Big].
\end{aligned}$$

This equivalent, but—importantly—more parsimonious, form of the limiting process η , rather than that in (11), will serve as the basis for the martingale transformation described in the next section, which would otherwise be ill-defined.

Remark 3.4. Although the bivariate case is the most interesting and the most relevant one, the setup and results of this paper can be generalized to the d -variate case, for $d > 2$. We will not pursue this, however, because for $d > 2$ a d -variate tail copula R on the domain $[0, \infty)^d$ yields, in contrast to when $d = 2$, only limited information on the tail dependence structure of F .

4 Martingale transformation

In this section, we describe a martingale transformation that enables us to transform the limiting process η into a standard bivariate Wiener process. The transformation is obtained by a suitable application of the martingale innovation transform approach originally developed in Khmaladze (1981, 1988, 1993), and used e.g., in Koenker and Xiao (2002, 2006) for quantile regression, in Khmaladze and Koul (2004) and Delgado et al. (2005) for goodness-of-fit testing of parametric regression and time series models, and used and extended in Can et al. (2015) for parametric tail copula goodness-of-fit testing. The classical Doob-Meyer decomposition of a Brownian bridge, which turns a Brownian bridge into a standard Brownian motion by subtracting its compensator, occurs as a simple special case of this approach. Heuristically, the present martingale transformation may therefore be understood as the suitable counterpart for η of the Doob-Meyer decomposition for a

Brownian bridge.

Consider the representation we derived in Remark 3.3 for the limiting process η of Theorem 3.1. It is easy to see, by direct computation of the covariance, that the process $\sqrt{c}V_R(x, y) - \sqrt{1-c}V'_R(x, y)$ is an R -Wiener process, which we will call \tilde{V}_R . Thus η is of the form

$$\eta(x, y) = \tilde{V}_R(x, y) + \sum_{i=1}^8 Q_i(x, y)Z_i, \quad (x, y) \in (0, \infty)^2, \quad (12)$$

where the Z_i are random variables and the Q_i are deterministic functions on $(0, \infty)^2$, defined by

$$\begin{aligned} Q_1(x, y) &= R^{(1)}(x, y) f(x, \gamma_1), & Q_2(x, y) &= R^{(2)}(x, y) f(y, \gamma_2), \\ Q_3(x, y) &= R^{(1)}(x, y) g(x, \gamma_1), & Q_4(x, y) &= R^{(2)}(x, y) g(y, \gamma_2), \\ Q_5(x, y) &= R^{(1)}(x, y) h(x, \gamma_1), & Q_6(x, y) &= R^{(2)}(x, y) h(y, \gamma_2), \\ Q_7(x, y) &= R^{(1)}(x, y) h(x, \gamma'_1), & Q_8(x, y) &= R^{(2)}(x, y) h(y, \gamma'_2). \end{aligned} \quad (13)$$

We note that (12) is of the general bivariate form (23) in Can et al. (2015). Hence, η can be transformed into a standard Wiener process by suitably exploiting Theorem 3.1 of that paper. We will formally state this result in Theorem 4.1 below, but we first introduce some assumptions and notation.

A4. The tail copula R has a positive density $r = \partial^2 R / \partial x \partial y$ on $(0, \infty)^2$. The derivatives $r^{(1)} := \partial r / \partial x$ and $r^{(2)} := \partial r / \partial y$ exist and are continuous on $(0, \infty)^2$.

Observe that this condition allows R to have mass on the “axes at infinity” $\{(x, \infty) : x \geq 0\} \cup \{(\infty, y) : y \geq 0\}$, but it excludes (strict) tail independence, i.e., $r(x, y) = 0$ for all $(x, y) \in (0, \infty)^2$. We note that it can be tested whether the Ledford-Tawn coefficient of tail dependence is less than 1, see Draisma et al. (2004). If this is the case, the data are tail independent.

Now, with the functions Q_i as defined in (13), let us denote $q_i = dQ_i/dR$ for $i = 1, \dots, 8$,

so that

$$\begin{aligned}
q_1(x, y) &= f^{(1)}(x, \gamma_1) + f(x, \gamma_1) \frac{r^{(1)}(x, y)}{r(x, y)}, & q_2(x, y) &= f^{(1)}(y, \gamma_2) + f(y, \gamma_2) \frac{r^{(2)}(x, y)}{r(x, y)}, \\
q_3(x, y) &= g^{(1)}(x, \gamma_1) + g(x, \gamma_1) \frac{r^{(1)}(x, y)}{r(x, y)}, & q_4(x, y) &= g^{(1)}(y, \gamma_2) + g(y, \gamma_2) \frac{r^{(2)}(x, y)}{r(x, y)}, \\
q_5(x, y) &= h^{(1)}(x, \gamma_1) + h(x, \gamma_1) \frac{r^{(1)}(x, y)}{r(x, y)}, & q_6(x, y) &= h^{(1)}(y, \gamma_2) + h(y, \gamma_2) \frac{r^{(2)}(x, y)}{r(x, y)}, \\
q_7(x, y) &= h^{(1)}(x, \gamma'_1) + h(x, \gamma'_1) \frac{r^{(1)}(x, y)}{r(x, y)}, & q_8(x, y) &= h^{(1)}(y, \gamma'_2) + h(y, \gamma'_2) \frac{r^{(2)}(x, y)}{r(x, y)},
\end{aligned}$$

where $f^{(1)}(x, \gamma) := \partial f(x, \gamma) / \partial x$ and similarly for g and h . Let $\mathbf{q}(x, y)$ denote the column vector consisting of $q_1(x, y), \dots, q_8(x, y)$. We will also write $\mathbf{q}(x, y, \tilde{\gamma}_1, \tilde{\gamma}_2, \tilde{\gamma}'_1, \tilde{\gamma}'_2, \tilde{r}, \tilde{r}_1, \tilde{r}_2)$ for the vector $\mathbf{q}(x, y)$ with the true parameters $\gamma_1, \gamma_2, \gamma'_1, \gamma'_2$ replaced by arbitrary $\tilde{\gamma}_1, \tilde{\gamma}_2, \tilde{\gamma}'_1, \tilde{\gamma}'_2 \in \mathbb{R}$, the true density r replaced by an arbitrary mapping $\tilde{r} : (0, \infty)^2 \rightarrow (0, \infty)$, and the true partial derivatives $r^{(1)}, r^{(2)}$ replaced by arbitrary mappings \tilde{r}_1, \tilde{r}_2 from $(0, \infty)^2$ to \mathbb{R} .

We also introduce matrices

$$\mathbf{I}_{\delta, T}(t) = \int_{\delta}^T \int_t^T \mathbf{q}(s, t) \mathbf{q}(s, t)^{\top} r(s, t) \, ds \, dt, \quad t \in [\delta, T], \quad (14)$$

and we denote

$$\begin{aligned}
&\mathbf{I}_{\delta, T}(t, \tilde{\gamma}_1, \tilde{\gamma}_2, \tilde{\gamma}'_1, \tilde{\gamma}'_2, \tilde{r}, \tilde{r}_1, \tilde{r}_2) \\
&= \int_{\delta}^T \int_t^T \mathbf{q}(s, t, \tilde{\gamma}_1, \tilde{\gamma}_2, \tilde{\gamma}'_1, \tilde{\gamma}'_2, \tilde{r}, \tilde{r}_1, \tilde{r}_2) \mathbf{q}(s, t, \tilde{\gamma}_1, \tilde{\gamma}_2, \tilde{\gamma}'_1, \tilde{\gamma}'_2, \tilde{r}, \tilde{r}_1, \tilde{r}_2)^{\top} \tilde{r}(s, t) \, ds \, dt,
\end{aligned}$$

for $t \in [\delta, T]$.

The functions q_1, \dots, q_8 bear resemblance to score functions, corresponding to $a_1, a_2, b_1, b_2, \gamma_1, \gamma_2, \gamma'_1$, and γ'_2 , respectively, using terminology from likelihood analysis. Note the asymmetry between the score functions associated to the two samples, which occurs due to Remark 3.3. Furthermore, $\mathbf{I}_{\delta, T}(t)$ can be viewed as a partial Fisher information matrix built from these scores.

Our next assumption is:

A5. The matrices $\mathbf{I}_{\delta,T}(t, \tilde{\gamma}_1, \tilde{\gamma}_2, \tilde{\gamma}'_1, \tilde{\gamma}'_2, \tilde{r}, \tilde{r}_1, \tilde{r}_2)$ are well-defined and invertible for all $t \in [\delta, T)$ and all $\tilde{\gamma}_1, \tilde{\gamma}_2, \tilde{\gamma}'_1, \tilde{\gamma}'_2, \tilde{r}, \tilde{r}_1, \tilde{r}_2$ in some neighborhood of $\gamma_1, \gamma_2, \gamma'_1, \gamma'_2, r, r^{(1)}, r^{(2)}$ w.r.t. the relevant metrics (Euclidean metric for $\tilde{\gamma}_j, \tilde{\gamma}'_j$, the metric generated by the norm $\sup_{[\delta, T]^2} |\cdot|$ for $\tilde{r}, \tilde{r}_1, \tilde{r}_2$).

We are now equipped to state our martingale transformation result.

Theorem 4.1. *Let η be the limiting process derived in Remark 3.3 . If Assumptions A4–A5, confined to the true $\gamma_1, \gamma_2, \gamma'_1, \gamma'_2, r, r^{(1)}, r^{(2)}$ hold, then the process*

$$\begin{aligned}
W(x, y) = & \int_{\delta}^{\delta+x} \int_{\delta}^{\delta+y} \frac{1}{\sqrt{r(s, t)}} d\eta(s, t) \\
& - \int_{\delta}^{\delta+x} \int_{\delta}^{\delta+y} \mathbf{q}(s, t)^{\top} \left(\mathbf{I}_{\delta, T}^{-1}(t) \int_{\delta}^T \int_t^T \mathbf{q}(s', t') d\eta(s', t') \right) \sqrt{r(s, t)} ds dt
\end{aligned} \tag{15}$$

is a standard Wiener process on $[0, \tau]^2$, for any $0 < \tau < T - \delta$.

Remark 4.2. Recall (12). The martingale transformation induces a nullification of $\eta(x, y) - \tilde{V}_R(x, y) = \sum_{i=1}^8 Q_i(x, y) Z_i$ and normalizes the resulting R -Wiener process to render a standard Wiener process on $[0, \tau]^2$. The double integrals with respect to the process η can be understood pathwise as Riemann-Stieltjes integrals; we refer to e.g., Towghi (2002), Theorem 1.2(a), for an appropriate existence result.

Remark 4.3. We note that in the cases $\gamma_1 = \gamma'_1$ and $\gamma_2 = \gamma'_2$, we will have $q_5 \equiv q_7$ and $q_6 \equiv q_8$, respectively, and the matrix $\mathbf{I}_{\delta,T}(t)$ will be singular. If this information is available, the redundant q -functions can simply be omitted when constructing $\mathbf{I}_{\delta,T}(t)$, and Theorem 4.1 continues to apply *mutatis mutandis*. Otherwise, this case may be included by using the Moore-Penrose pseudoinverse of $\mathbf{I}_{\delta,T}(t)$ in (15) and adapting the proof of Theorem 4.1 accordingly. Besides, as we will show by simulations in Section 6, our testing approach, which uses the regular inverse of the matrix $\mathbf{I}_{\delta,T}(t, \tilde{\gamma}_1, \tilde{\gamma}_2, \tilde{\gamma}'_1, \tilde{\gamma}'_2, \tilde{r}, \tilde{r}_1, \tilde{r}_2)$ defined above, can in practice still be applied without size distortions when $\gamma_1 = \gamma'_1$ or $\gamma_2 = \gamma'_2$.

5 Two-sample testing

Applying the transformation in (15) to $\eta_{n,n'}$, with unknown quantities replaced by estimators, we obtain the following empirical process on $[0, T - \delta]^2$:

$$W_{n,n'}(x, y) = \int_{\delta}^{\delta+x} \int_{\delta}^{\delta+y} \frac{1}{\sqrt{\widehat{r}(s, t)}} d\eta_{n,n'}(s, t) - \int_{\delta}^{\delta+x} \int_{\delta}^{\delta+y} \widehat{\mathbf{q}}(s, t)^{\top} \left(\widehat{\mathbf{I}}_{\delta, T}^{-1}(t) \int_{\delta}^T \int_t^T \widehat{\mathbf{q}}(s', t') d\eta_{n,n'}(s', t') \right) \sqrt{\widehat{r}(s, t)} ds dt, \quad (16)$$

where $\widehat{\mathbf{q}}(\cdot)$ and $\widehat{\mathbf{I}}_{\delta, T}(\cdot)$ are short-hand notations for $\mathbf{q}(\cdot, \widehat{\gamma}_1, \widehat{\gamma}_2, \widehat{\gamma}'_1, \widehat{\gamma}'_2, \widehat{r}, \widehat{r}^{(1)}, \widehat{r}^{(2)})$ and $\mathbf{I}_{\delta, T}(\cdot, \widehat{\gamma}_1, \widehat{\gamma}_2, \widehat{\gamma}'_1, \widehat{\gamma}'_2, \widehat{r}, \widehat{r}^{(1)}, \widehat{r}^{(2)})$, respectively.

Let K be a univariate density serving as a kernel for kernel density estimation. To estimate the tail copula density r , we propose the non-parametric estimator

$$\begin{aligned} \widehat{r}(x, y) &= \frac{k^{2/5}}{k^{2/5} + (k')^{2/5}} \left[\frac{1}{kw^2} \sum_{i=1}^n K \left(\frac{x - \widehat{X}_i(n/k)}{w} \right) K \left(\frac{y - \widehat{Y}_i(n/k)}{w} \right) \right] \\ &\quad + \frac{(k')^{2/5}}{k^{2/5} + (k')^{2/5}} \left[\frac{1}{k'(w')^2} \sum_{i=1}^{n'} K \left(\frac{x - \widehat{X}'_i(n'/k')}{w'} \right) K \left(\frac{y - \widehat{Y}'_i(n'/k')}{w'} \right) \right] \\ &=: \lambda s_w(x, y) + (1 - \lambda) s'_{w'}(x, y), \end{aligned}$$

where $w = \bar{c}_n(x, y)k^{-1/10}$ and $w' = \bar{c}'_{n'}(x, y)(k')^{-1/10}$, with, for some $\bar{c} > 0$, $\bar{c}_n(x, y) \rightarrow \bar{c}$, $\bar{c}'_{n'}(x, y) \rightarrow \bar{c}$, uniformly on $[\delta, T]^2$.

For the partial derivative $r^{(1)}$, we propose the non-parametric estimator

$$\begin{aligned} \widehat{r}^{(1)}(x, y) &= \frac{k^{1/3}}{k^{1/3} + (k')^{1/3}} \left[\frac{1}{kw^3} \sum_{i=1}^n K^{(1)} \left(\frac{x - \widehat{X}_i(n/k)}{w} \right) K \left(\frac{y - \widehat{Y}_i(n/k)}{w} \right) \right] \\ &\quad + \frac{(k')^{1/3}}{k^{1/3} + (k')^{1/3}} \left[\frac{1}{k'(w')^3} \sum_{i=1}^{n'} K^{(1)} \left(\frac{x - \widehat{X}'_i(n'/k')}{w'} \right) K \left(\frac{y - \widehat{Y}'_i(n'/k')}{w'} \right) \right], \end{aligned}$$

where now $w = \widetilde{c}_n(x, y)k^{-1/12}$ and $w' = \widetilde{c}'_{n'}(x, y)(k')^{-1/12}$, with, for some $\widetilde{c} > 0$, $\widetilde{c}_n(x, y) \rightarrow \widetilde{c}$, $\widetilde{c}'_{n'}(x, y) \rightarrow \widetilde{c}$, uniformly on $[\delta, T]^2$, and where $K^{(1)}$ denotes the derivative of K . The partial derivative $r^{(2)}$ is estimated analogously. In the sequel, we choose K to be the well-known triweight kernel $K(u) = \frac{35}{32}(1 - u^2)^3 \mathbf{1}_{[-1, 1]}(u)$, for convenience. We will demonstrate, as part of our main theoretical result that follows, that these estimators are consistent.

In view of Theorems 3.1 and 4.1, one might anticipate the empirical process $W_{n,n'}$ to converge weakly to a standard Wiener process as $n, n' \rightarrow \infty$. This turns out to be true, and we will formally establish this result in Theorem 5.1 below, but first we state one more assumption.

A6. The partial derivatives $r^{(112)} := \partial^3 r / (\partial x^2 \partial y)$, and $r^{(212)} := \partial^3 r / (\partial x \partial y^2)$ exist and are continuous on $(0, \infty)^2$.

Note that this implies that $r^{(11)} := \partial^2 r / \partial x^2$, $r^{(22)} := \partial^2 r / \partial y^2$, and $r^{(12)} := \partial^2 r / (\partial x \partial y)$ exist and are continuous on $(0, \infty)^2$.

Theorem 5.1. *Under Assumptions A1–A6, the process $W_{n,n'}$ in (16) converges weakly to a standard bivariate Wiener process in $D([0, \tau]^2)$ as $n, n' \rightarrow \infty$, for any $0 < \tau < T - \delta$.*

Theorem 5.1, our main theoretical result, states that under the null hypothesis $R = R'$, the test process $W_{n,n'}$ converges weakly to a distribution-free limiting process, and hence $W_{n,n'}$ can be used as a generator of a multitude of asymptotically distribution-free test statistics for testing $R = R'$. We will not investigate how to construct optimal tests against certain specific alternative hypotheses. Instead, we will study in Section 6 three common examples of non-parametric omnibus tests based on $W_{n,n'}$, not only under the null but also under the alternative hypothesis.

6 Simulations

In this section, we study the finite-sample behavior of the test process $W_{n,n'}$ in (16), both under the null hypothesis $R = R'$ and the alternative hypothesis $R \neq R'$, via Monte Carlo simulations. All the computations of the present section, as well as those of the subsequent Section 7, are implemented in the software R. An .R file for the implementation of our testing approach is provided as a supplement.

To improve the finite-sample approximation of Theorem 5.1, we actually consider the process $\widetilde{W}_{n,n'}$, an asymptotically equivalent version of $W_{n,n'}$, obtained by replacing k and k' by $k - \sqrt{k}$ and $k' - \sqrt{k'}$, respectively, in the definition of the normalizing factor $\sqrt{\kappa}$. For each model considered below, we generate 1000 pairs of bivariate samples of size $n = n' = 1500$. We take $\delta = 0.25$ and $T = 2$. In additional simulation results (not reported here), we have seen that the actual choices of the tuning parameters δ and T are relatively unimportant and we have thus fixed them throughout. Also, based on further simulations, we have decided to choose the multiplication factors in the bandwidth to be $\bar{c}_n(x, y) = \bar{c}_{n'}(x, y) = \tilde{c}_n(x, y) = \tilde{c}_{n'}(x, y) = 1.5$. This fixes the bandwidth once k is chosen. The choice of k is known to be important and difficult in extreme value statistics. As a rule of thumb, we consider the relation between k and the sample size n to be $k \approx n^{3/4}$ if n is not very large (say, $n \leq 10,000$). If we are in the luxurious situation that $n > 10,000$, we suggest $k \approx 10n^{1/2}$. In the simulations, for $n = n' = 1500$, we take $k = k' = 250$, but the simulation results under the null hypothesis do not change substantially when k changes within reasonable limits. In particular, we find no clearly visible effect on the PP-plots. Of course, the power is affected when k varies. Indeed, when k becomes larger, the power increases sharply, for two reasons: (i) increased (effective) sample size; and (ii) increased bias. In empirical applications, a further sensitivity analysis can be helpful for a range of k -values, where the largest k is twice the smallest one, with the rule-of-thumb value included; see Section 7. The estimators \hat{a}_j, \hat{a}'_j and $\hat{\gamma}_j, \hat{\gamma}'_j$, $j = 1, 2$, are taken to be the moment estimators (see, e.g., de Haan and Ferreira (2006), Sections 4.2 and 3.5), and we set, as usual, $\hat{b}_1 = X_{n-k:n}$, $\hat{b}_2 = Y_{n-k:n}$ (and analogously for \hat{b}'_1, \hat{b}'_2), with $X_{i:n}, Y_{i:n}$ denoting the marginal order statistics.

Once a pair of bivariate samples is generated, we construct the process $\widetilde{W}_{n,n'} =: W_n$ on a 100×100 finite grid of equidistant points spanning $[0, 1]^2$. To compare this observed

path with a standard Wiener process, we compute three test statistics, namely

$$\begin{aligned}\kappa_n &= \max_{(x,y) \in \mathcal{G}} |W_n(x,y)|, && \text{(Kolmogorov-Smirnov type)} \\ \omega_n^2 &= \|\mathcal{G}\|^2 \sum_{(x,y) \in \mathcal{G}} W_n(x,y)^2, && \text{(Cramér-von Mises type)} \\ A_n^2 &= \|\mathcal{G}\|^2 \sum_{(x,y) \in \mathcal{G}} \frac{W_n(x,y)^2}{xy}, && \text{(Anderson-Darling type)}\end{aligned}\tag{17}$$

where $\|\mathcal{G}\|$ is the mesh length, i.e., $1/100$.

The same statistics are also computed from 10,000 paths of the true standard Wiener process generated on the same grid \mathcal{G} , to create benchmark distribution tables. Recall that these tables have to be made only once as a consequence of obtaining a distribution-free limit. We denote the statistics, computed from the true Wiener process, by κ , ω^2 and A^2 . For each model, we present PP-plots for a visual comparison of the empirical distributions of κ_n vs. κ , ω_n^2 vs. ω^2 and A_n^2 vs. A^2 . We also present rejection frequencies at 5% and 1% significance levels.

6.1 Simulations under the null hypothesis

We consider the following four models:

Model I. F has Gumbel(2) copula and Pareto(3) margins, F' has Gumbel(2) copula and Pareto(4) margins.

Model II. Both F and F' have Gumbel(2) copula and Pareto(3) margins.

Model III. F has Clayton(1) survival copula and Pareto(3) margins, F' has Clayton(1) survival copula and Pareto(4) margins.

Model IV. F has Gumbel(2)-independence mixture copula with equal weights and Pareto(3) margins, F' has Gumbel(2)-Clayton(1) mixture copula with equal weights and Pareto(4) margins.

Note that we specify each bivariate distribution F (and F') in terms of a copula C and marginal dfs F_1, F_2 , which characterize F via $F(x, y) = C(F_1(x), F_2(y))$. Also, Pareto(α) with $\alpha > 0$ refers to the df $1 - 1/x^\alpha$, $x > 1$; Gumbel(θ) with $\theta \geq 1$ refers to the copula

$$\exp\{-[(-\log u)^\theta + (-\log v)^\theta]^{1/\theta}\}, \quad (u, v) \in [0, 1]^2;$$

Clayton(θ) with $\theta > 0$ refers to the copula

$$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, \quad (u, v) \in [0, 1]^2;$$

and therefore the Clayton(θ) *survival* copula is given by

$$((1 - u)^{-\theta} + (1 - v)^{-\theta} - 1)^{-1/\theta} + u + v - 1, \quad (u, v) \in [0, 1]^2.$$

Since the tail copula of a bivariate df is entirely determined by the copula, the null hypothesis $R = R'$ clearly holds in Models I-III, where F and F' have identical copulas. In Model IV, the copulas of F and F' are different, but $R = R'$ still holds, as both the independence copula and the Clayton(1) copula have upper tail independence (i.e., zero tail copula).

The PP-plots in Fig. 6.1 indicate that the process W_n indeed behaves like a standard Wiener process, providing an empirical confirmation of Theorem 5.1. Moreover, in Table 6.1 we present the rejection counts at 5% significance level. For each particular model and test statistic, the rejection count is simply the observed number of times (out of 1000) the test statistic exceeds the 95th percentile of the corresponding distribution obtained from the true Wiener process. In all cases, the observed counts are consistent with the Binomial(1000, 0.05) distribution, substantiating a good size performance. Note that the results for Model II show that our approach works well when F and F' have identical marginal distributions; see Remark 4.3.

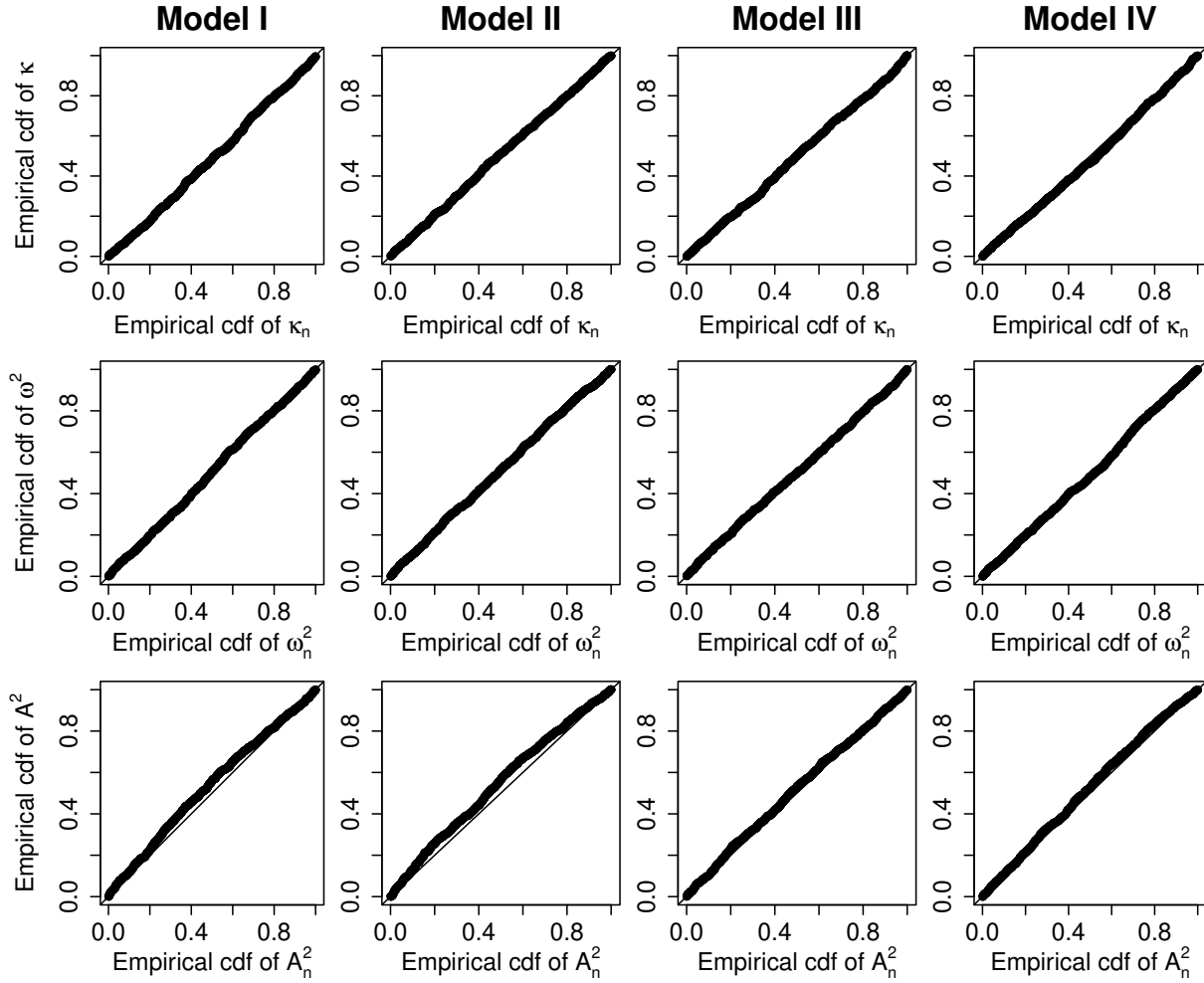


Figure 6.1: *PP-plots for the three test statistics constructed from 1000 simulated sample pairs per model*

	Model I	Model II	Model III	Model IV
κ_n	36	49	41	38
ω_n^2	47	50	41	53
A_n^2	47	63	49	60

Table 6.1: *Number of rejections in 1000 repetitions at 5% significance level*

6.2 Simulations under the alternative hypothesis

Before describing the simulation models under the alternative, we first note that the estimation of $R(x, y)$ is statistically difficult in the sense that generally the corresponding

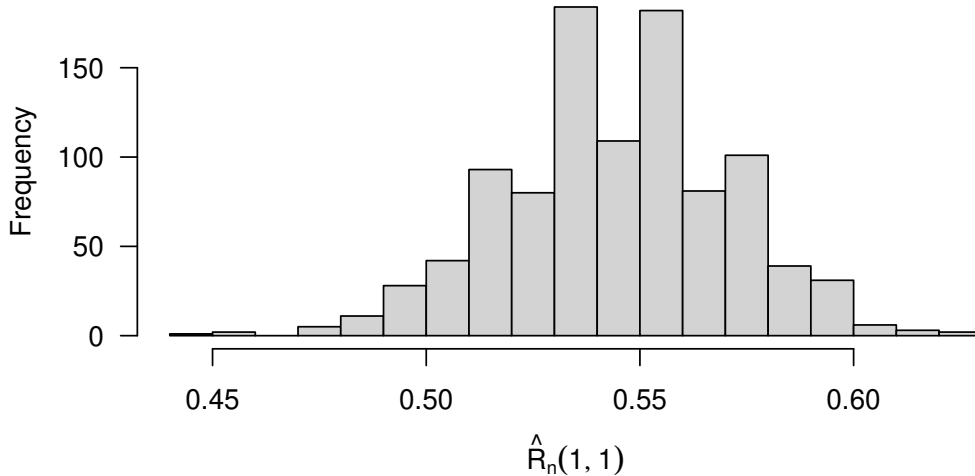


Figure 6.2: *Histogram of $\widehat{R}_n(1,1)$ values computed from 1000 bivariate samples with Clayton(1) survival copula and Pareto(3) margins*

estimator has considerable bias and variance. For simplicity, consider $R(1,1)$, the upper tail dependence coefficient. For the sample sizes used in this simulation section, it is hard to distinguish between the values $R(1,1) = 0.5$ and $R'(1,1) = 0.6$, say, so a test cannot achieve a high power in such a case. In order to substantiate this, we reuse the simulation results from the df F with the Clayton(1) survival copula and Pareto(3) margins in Model III above. In this case $R(1,1) = 0.5$. The histogram in Figure 6.2 depicts the distribution of the 1000 estimates $\widehat{R}_n(1,1)$. The (empirical) bias of these estimates is as high as 0.046 and the sample standard deviation is 0.026. We see from the figure that the central 95% of the estimates range from 0.492 to 0.596; the true value is barely included in this interval and the range is rather wide. Note that when performing a two-sample test, the bias and variance of the second sample also have to be taken into account.

In this subsection, we consider the four models listed below. The *linear factor model* of Model IVa refers to the bivariate random vector

$$(\lambda Z_1 + (1 - \lambda)Z_2, \mu Z_1 + (1 - \mu)Z_2), \quad (18)$$

with Z_1, Z_2 independent and $\lambda, \mu \in (0, 1)$ deterministic parameters.

Model Ia. F has Gumbel(2) copula and Pareto(3) margins, F' has Gumbel(6) copula and Pareto(4) margins.

Model IIa. F has Gumbel(4) copula and Pareto(3) margins, F' has Gumbel(1) (i.e., independence) copula and Pareto(4) margins.

Model IIIa. F has Clayton(1) survival copula and Pareto(3) margins, F' has Clayton(5) survival copula and Pareto(4) margins.

Model IVa. F is the df of the linear factor model (18), with $Z_1, Z_2 \sim \text{Pareto}(1)$, $\lambda = 0.9$, $\mu = 0.8$, and F' is the df of the same model (18) with $Z_1, Z_2 \sim \text{Pareto}(2)$, $\lambda = 0.9$, $\mu = 0.2$.

As in Subsection 6.1 above, we construct PP-plots for each model and each of the three test statistics; see Fig. 6.3. The PP-plots show dramatic deviations from the behavior under a standard Wiener process. We also present rejection counts at 5% significance level in Table 6.2. As suggested by the PP-plots, the three omnibus tests have high power with, for the present alternatives, Cramér-von Mises outperforming Kolmogorov-Smirnov, and Anderson-Darling performing the best. Based on this, we recommend to apply both the Cramér-von Mises and the Anderson-Darling type test statistics.

	Model Ia	Model IIa	Model IIIa	Model IVa
κ_n	836	704	778	994
ω_n^2	922	896	891	1000
A_n^2	977	963	974	1000

Table 6.2: *Number of rejections in 1000 repetitions at 5% significance level*

In Fig. 6.4 below, we also provide empirical power curves at 5% significance level for the three test statistics, computed from a sequence of eight models similar to Model Ia.

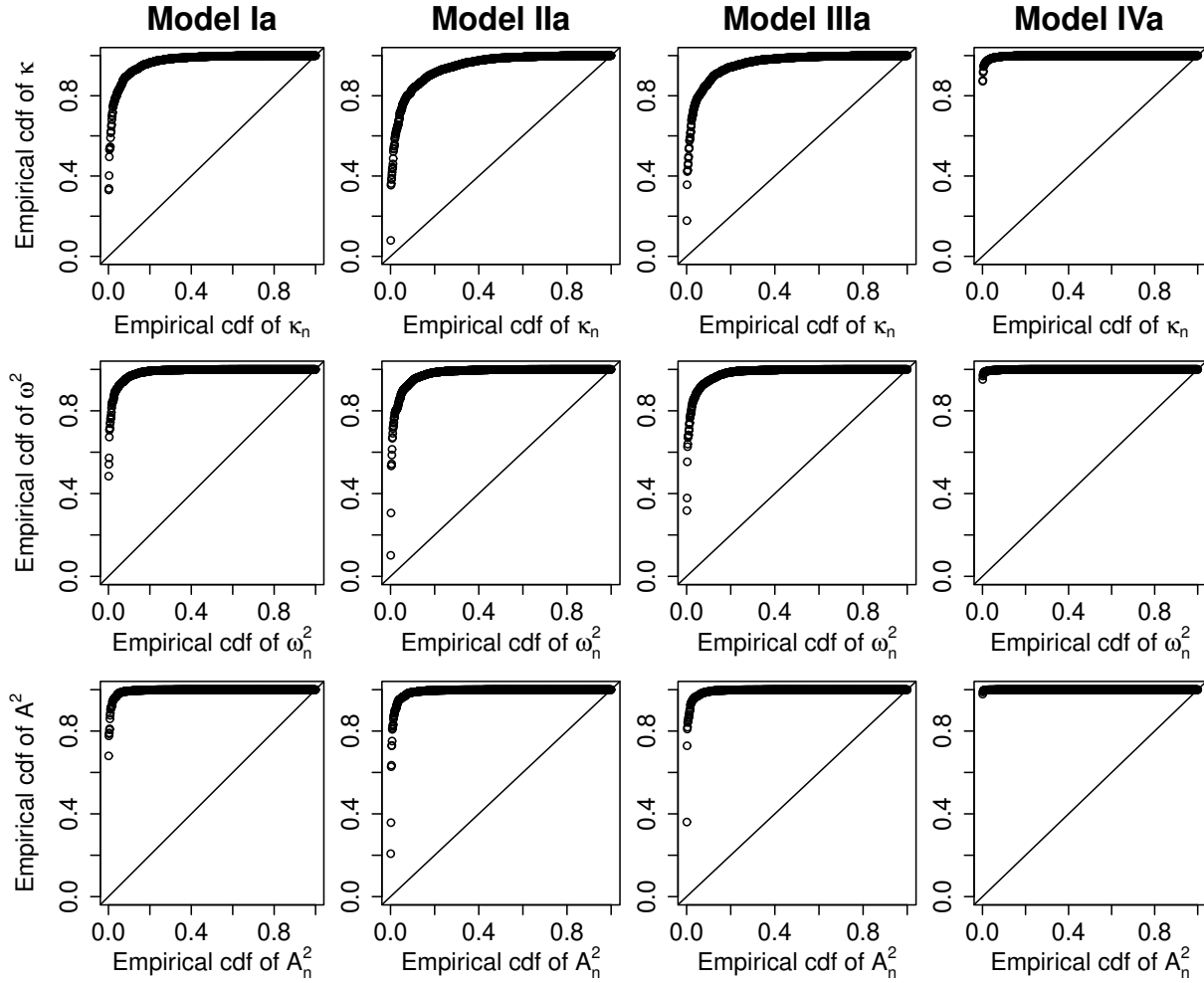


Figure 6.3: *PP-plots for the three test statistics constructed from 1000 simulated sample pairs per model*

In each model, F has Gumbel(θ) copula and Pareto(3) margins, F' has Gumbel(6) copula and Pareto(4) margins, and θ takes a sequence of values in the interval (1, 6) so that the difference of the tail dependence coefficients $R'(1, 1) - R(1, 1)$ takes a sequence of values ranging from 0.1 to 0.8. Observe that the power grows quickly and is high as soon as the difference is 0.2. The exact values of θ that were used, and the corresponding differences $R'(1, 1) - R(1, 1)$, can be seen in Table 6.3 below. Each point in a given power curve represents the observed rejection rate at the 5% level among 200 sample pairs.

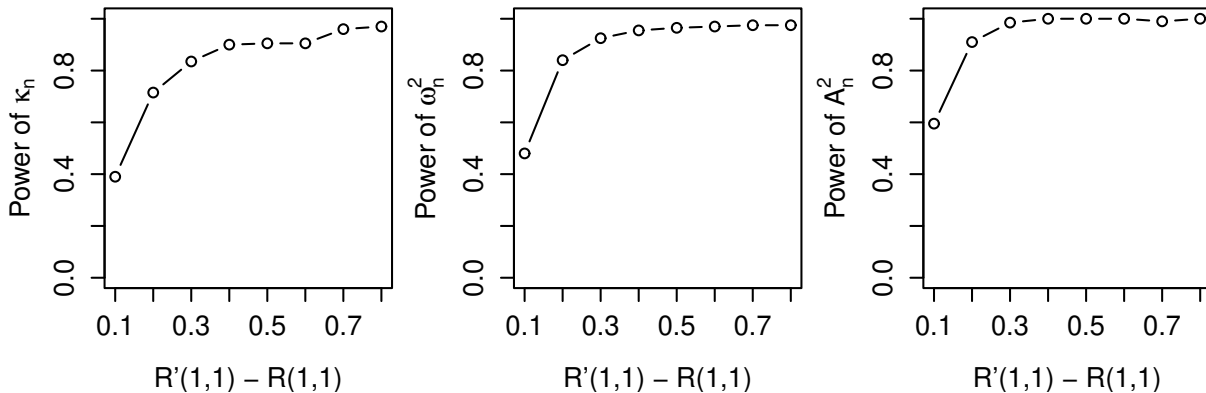


Figure 6.4: Empirical power curves for the three test statistics, constructed from a sequence of eight models similar to Model Ia

θ	3.45	2.48	1.97	1.65	1.43	1.27	1.15	1.06
$R'(1,1) - R(1,1)$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8

Table 6.3: Gumbel parameter values used for the empirical power curves, and the corresponding differences $R'(1,1) - R(1,1)$

7 Empirical analysis

Emboldened by the Monte Carlo simulation results of the previous section, demonstrating good size and power performances, we now take our approach to real equity index data of three financial markets. We take $\delta = 0.25$, $T = 2$ and $\bar{c}_n(x, y) = \bar{c}'_n(x, y) = \tilde{c}_n(x, y) = \tilde{c}'_n(x, y) = 1.5$ as in Section 6, and consider a range of k -values (see below).

An important concern when analyzing multiple financial markets operating in different time zones is the synchronicity of the data, which requires a careful treatment. We use intra-daily data to construct three synchronized time series of daily equity index returns: for FTSE 100 (UK), DAX 30 (Germany), and S&P 500 (US). Data are obtained from the Tick History Database of Thomson Reuters. Our full sample spans the period May 2008 to August 2015. Using Coordinated Universal Time (UTC), care has been taken to account

for the change to daylight savings time (DST), which does not occur on the same day around the world.

We focus on the pair {FTSE 100, DAX 30} for a detailed exposition. We construct two bivariate samples from the daily negative log-returns of these two indices: the first sample is generated from the 900 trading days preceding January 1, 2012 (spanning the period from May 23, 2008 to December 31, 2011), and the second sample is generated from the 900 trading days following January 1, 2012 (spanning the period from January 1, 2012 to August 13, 2015). To be more precise, the daily negative log-returns for a given index are the numbers $\log(p_t/p_{t+1})$, with p_t denoting the index price recorded at 14:30 UTC (13:30 UTC during European DST) on trading day t . Note that the first sample covers some of the most volatile trading days of the global financial crisis and the subsequent European debt crisis, while the second sample can be considered a relatively calm “post-crisis” period. We aim to test whether or not the tail dependence structure in the two samples, as characterized by the tail copula, shows statistical evidence of a change. As an interesting alternative analysis, which we will not pursue here, one could compare the tail copulas associated with the upper and lower tails of the daily bivariate log-returns within the *same* sampling period. Indeed, these bivariate extremal gains can be seen to be asymptotically independent of the bivariate extremal losses.

Figure 7.1 displays plots of the daily negative log-returns for the two indices during the two time periods, and Table 7.1 provides some sample statistics for these time series. It is visibly clear that the marginal structure changes for both indices from period 1 to period 2, but a possible change in the tail dependence structure is much more difficult to detect by visual inspection. The change in the marginal structure can also be observed in Figure 7.2, which plots the estimated extreme value index γ for each marginal data set over a set of k values ranging from 100 to 200. It is clear that the (right) tail of the daily negative log-returns is heavier in the first period (with EV index estimates firmly positive) than in

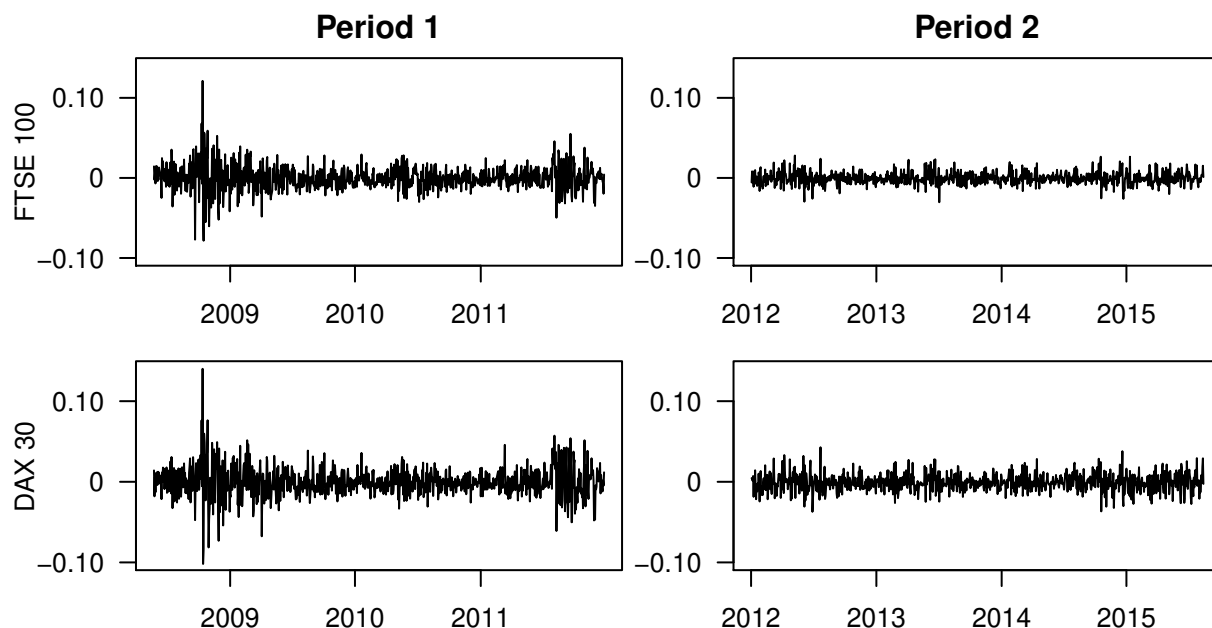


Figure 7.1: *Daily negative log-returns of FTSE 100 (top) and DAX 30 (bottom) in the two periods. Tick marks on the horizontal axes indicate the first trading day of each year.*

the second period (with EV index estimates mostly hovering below 0).

	Period 1				Period 2			
	Mean	SD	Min	Max	Mean	SD	Min	Max
FTSE 100	0.000	0.016	-0.078	0.121	0.000	0.008	-0.030	0.028
DAX 30	0.000	0.019	-0.102	0.140	0.001	0.011	-0.037	0.043

Table 7.1: *Sample statistics of the daily negative log-returns of FTSE 100 and DAX 30 in the two periods*

The next step in our analysis is to transform the observed negative log-returns (X_i, Y_i) in period 1 and (X'_i, Y'_i) in period 2 into $(\widehat{X}_i, \widehat{Y}_i)$ and $(\widehat{X}'_i, \widehat{Y}'_i)$ as described in (5), which leads to estimates \widehat{R}_n and \widehat{R}'_n of the tail copulas R and R' in the two periods. Figure 7.3 displays scatter plots of the transformed negative log-returns in the two periods, for $k = 150$. Since the transformation (5) maps joint extremes (X_i, Y_i) to points near the origin, the tail

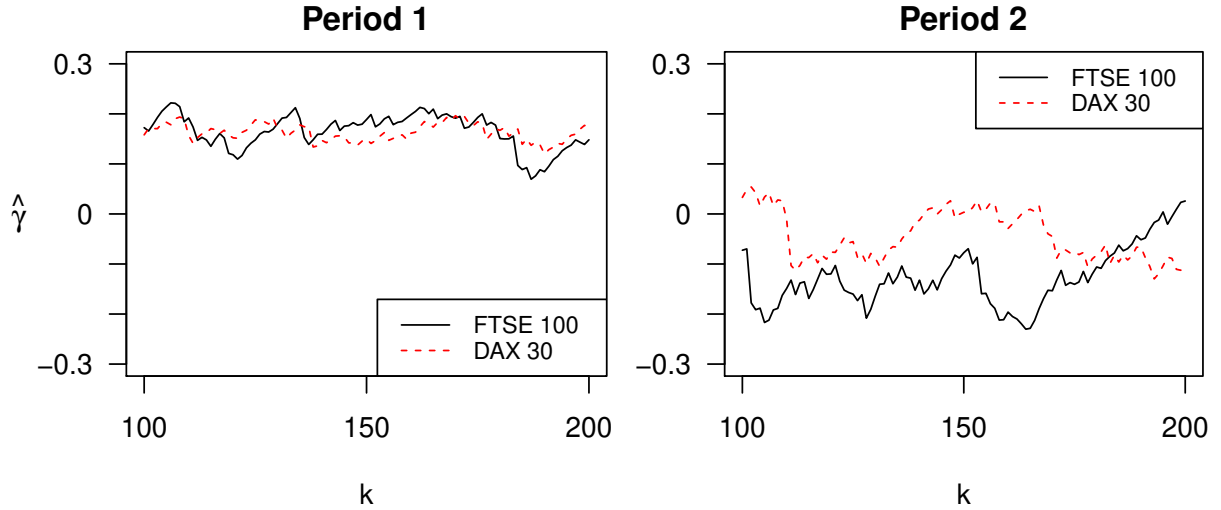


Figure 7.2: *Estimated marginal extreme value indices of the daily negative log-returns of FTSE 100 and DAX 30 in the two periods*

copula estimate $\widehat{R}_n(x, y)$ is proportional to the number of points $(\widehat{X}_i, \widehat{Y}_i)$ inside the rectangle $[0, x] \times [0, y]$ (and analogously for \widehat{R}'_n). In particular, the upper tail dependence coefficients $R(1, 1)$ of period 1 and $R'(1, 1)$ of period 2 can be estimated by dividing the number of points $(\widehat{X}_i, \widehat{Y}_i)$ or $(\widehat{X}'_i, \widehat{Y}'_i)$ in $[0, 1]^2$ by k . This leads to estimates $\widehat{R}_n(1, 1) = 115/150 = 0.77$ and $\widehat{R}'_n(1, 1) = 104/150 = 0.69$. The difference is small, and, also in light of the estimator's substantial bias and variance illustrated in Figure 6.2 for simulated data, it is not at all clear if it indicates a difference in the true tail copulas R and R' . Also note that $R(1, 1)$ does not tell the whole story about tail dependence, which can only be properly described by the entire tail copula R .

It remains to construct the test process W_n and compute the test statistics in (17), as in Section 6. We do this again for $k = 100, \dots, 200$. Plots of the resulting test statistics can be seen in Figure 7.4, together with the 95th and the 99th percentiles of their distributions under the null hypothesis. Each test statistic stays below the 95th percentile of the corresponding null distribution essentially for the entire range of k values, leading to non-rejection of the null hypothesis. Thus, the conclusion based on these three test statistics is that the tail

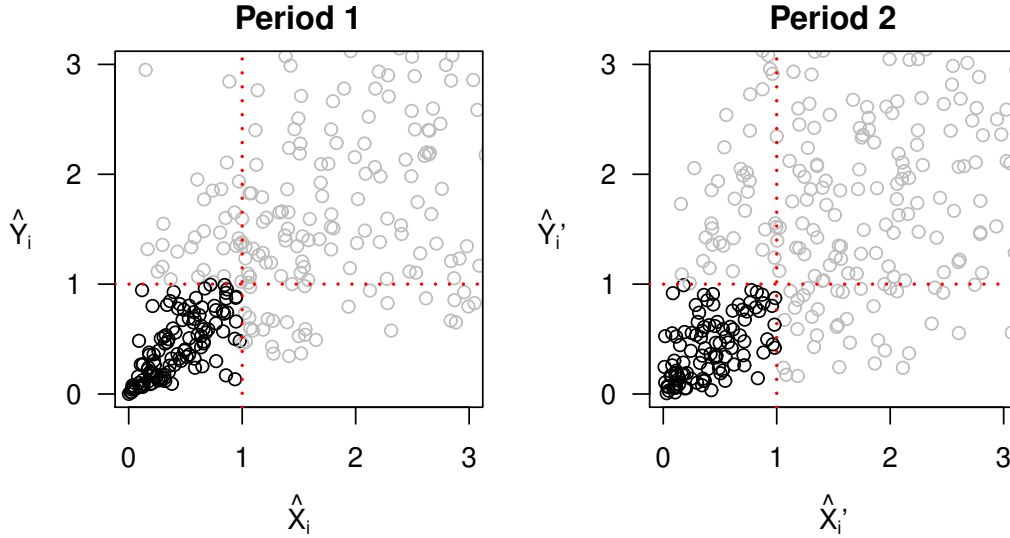


Figure 7.3: *Transformed daily negative log-returns* (\hat{X}_i, \hat{Y}_i) of period 1 and (\hat{X}'_i, \hat{Y}'_i) of period 2, with $k = 150$. The number of points in the square $[0, 1]^2$, divided by k , gives the estimates $\hat{R}_n(1, 1) = 115/150 = 0.77$ and $\hat{R}'_n(1, 1) = 104/150 = 0.69$.

dependence structure between the daily negative log-returns of FTSE 100 and DAX 30 does *not* change from period 1 to period 2, despite the visible evidence of change in the marginal (tail) behaviors of both equity indices.

For an independent verification, we also apply the multiplier bootstrap test for the equality of tail copulas described in Bücher and Dette (2013) to our data set. An implementation of the partial derivative multiplier (pdm) approach, as outlined in Section 4.1 of that paper, with $k = 150$ and 500 bootstrap replications, leads to an estimated p -value of 0.42 for the null hypothesis of equal tail copulas, which is consistent with our conclusion.

A possible objection against this application could be that, per period, the data may be serially dependent. First note that pairwise dependence in the tails can only be “positive”; for example, for the countermonotonic copula, the case of perfect negative dependence, the corresponding tail copula is just the 0-function, the same as for the independence copula. Loosely speaking, in case of (positive) serial dependence at high levels, the data contain

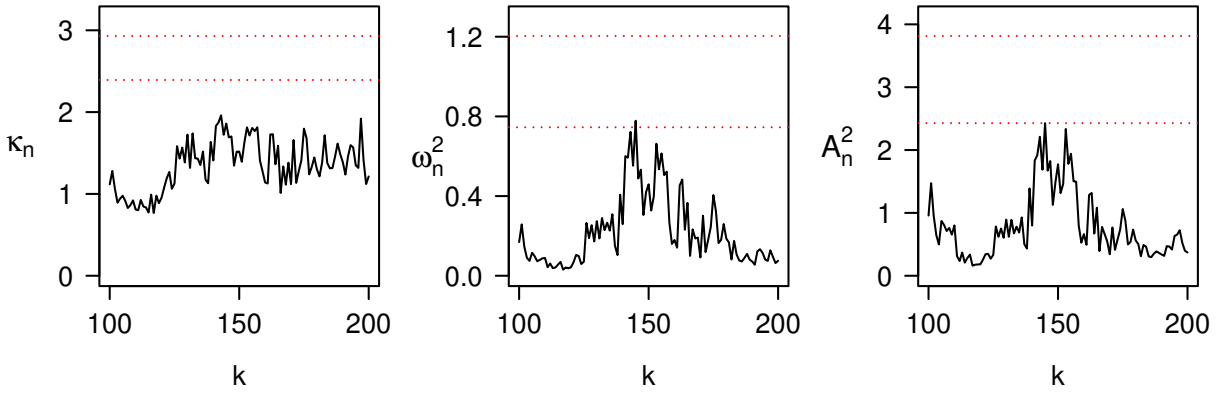


Figure 7.4: *The three test statistics in (17) computed from the daily negative log-returns of FTSE 100 and DAX 30 in periods 1 and 2. The dotted lines indicate the 95th and 99th percentiles of the corresponding null distributions.*

less information about the underlying df than in the i.i.d. case. Hence, estimation errors, that is, variances, become larger and so do therefore the critical values of the tests. As a result, our tests can, in fact, be considered as anticonservative (that is, they overreject): if the null hypothesis is not rejected under the i.i.d. assumption, it will not be rejected when serial dependence is taken into account. In the Appendix, we provide some empirical support for this heuristic argument via simulations: re-running Model III of Section 6.1 with component-wise serial dependence in both samples leads to higher rejection rates as expected. So, our conclusion of no change in tail copulas remains.

Having considered the tail dependence between a pair of European equity indices, we now turn our attention to the tail dependence between the transatlantic pair {FTSE 100, S&P 500}. We compute, like above, the daily negative log-returns of these two indices during the same periods 1 and 2 defined earlier, using daily synchronous index price data recorded at 14:30 UTC (13:30 UTC during European DST). Figure 7.5 displays plots of the daily negative log-returns for the two indices during the two time periods, and Table 7.2 provides sample statistics. As before, we see clear evidence of a change in the marginal behavior of the negative log-returns in period 1 vs. in period 2.

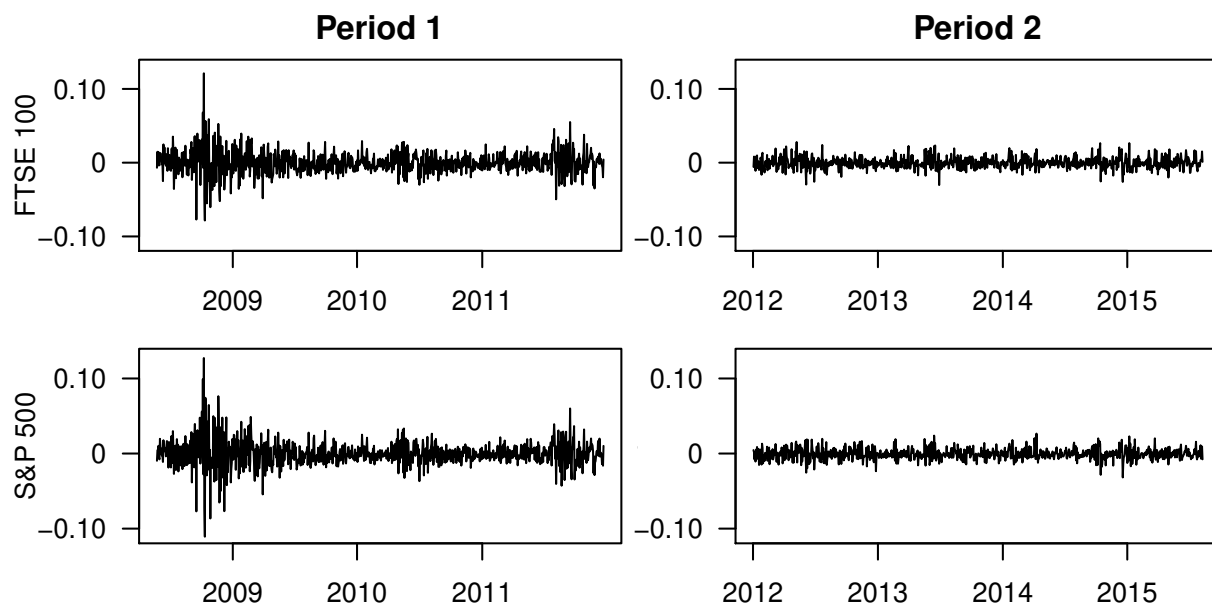


Figure 7.5: *Daily negative log-returns of FTSE 100 (top) and S&P 500 (bottom) in the two periods. Tick marks on the horizontal axes indicate the first trading day of each year.*

	Period 1				Period 2			
	Mean	SD	Min	Max	Mean	SD	Min	Max
FTSE 100	0.000	0.016	-0.078	0.121	0.000	0.008	-0.030	0.028
S&P 500	0.000	0.018	-0.111	0.127	0.001	0.007	-0.032	0.027

Table 7.2: *Sample statistics of the daily negative log-returns of FTSE 100 and S&P 500 in the two periods*

To test if there is a change in the tail dependence structure, we again construct our test process W_n from the two bivariate samples generated from the two periods, and compute the three test statistics in (17) for $k = 100, \dots, 200$; see Figure 7.6. We find that each test statistic mostly stays below the 95th percentile of the corresponding null distribution, with a few exceptions at particular k values. Keeping in mind that (i) looking at several k values simultaneously amounts to a multiple testing problem, and (ii) our data sets likely feature some serial dependence, both of which have an *upward* effect on critical values for the test

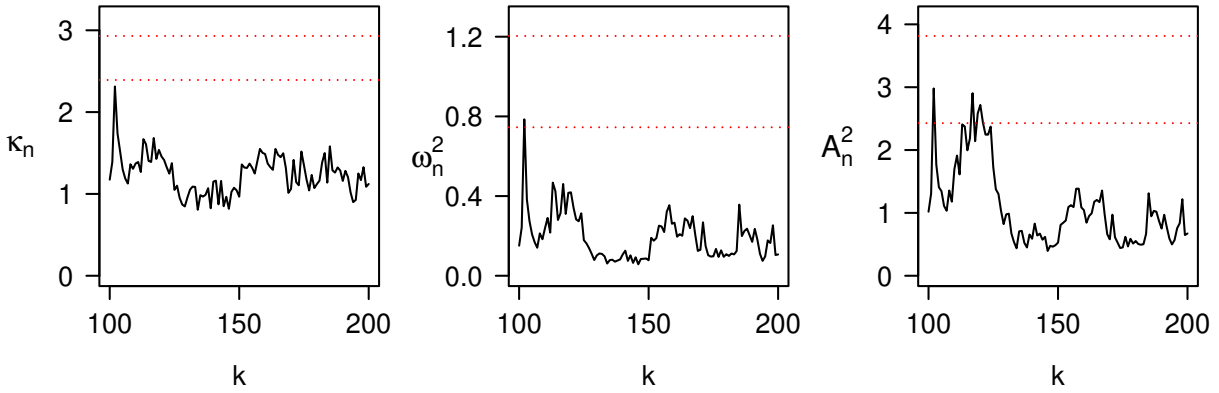


Figure 7.6: *The three test statistics in (17) computed from the daily negative log-returns of FTSE 100 and S&P 500 in periods 1 and 2. The dotted lines indicate the 95th and 99th percentiles of the corresponding null distributions.*

statistics, we conclude that our testing approach again supports the null hypothesis of no change in the tail dependence structure of the negative log-returns in period 1 vs. in period 2.

As with the earlier data set of {FTSE 100, DAX 30} negative log-returns, we also apply the *pdm* bootstrap test of Bücher and Dette (2013) to the present data set, with $k = 150$ and 500 bootstrap runs. This leads to an estimated p -value of 0.38 for the null hypothesis of equal tail copulas, which is again consistent with our conclusion.

8 Conclusions

We have developed a novel, general procedure to test for the equality of the tail copulas associated with two samples of bivariate data. A natural but complex feature of this problem is that the tail copulas, of which equality is tested for, are not specified. Deploying a martingale transformation of the suitably normalized difference between two semi-parametric estimators of the tail copulas, we have constructed a two-sample hypothesis testing process and established its weak convergence to a standard Wiener process under the null hypothe-

sis. Applying our hypothesis testing procedure to samples of negative log-returns of equity indices, during and after the global financial crisis, we find no evidence to reject the null hypothesis of tail copula equality. That is, although large negative returns occur more frequently and severely during crisis vs. post-crisis for the pairs of equity indices we have analyzed, our testing procedure reveals that there is no statistical evidence of a change in their tail dependence structure. These findings suggest that, for the highly developed equity markets we consider, whereas inference about marginal (tail) behavior should account for changes over time or across “regimes”, the tail dependence structure appears to be more stable and can thus be inferred from longer time periods.

Acknowledgements

We are very grateful to the Editor Professor Ivan Canay, the Associate Editor, and two referees for many useful comments that greatly improved the paper. We also thank Estate V. Khmaladze for stimulating discussions about this paper. John Einmahl holds the Arie Kapteyn Chair 2019–2022 and gratefully acknowledges the corresponding research support. Roger Laeven is supported in part by the Netherlands Organization for Scientific Research under an NWO-Vici grant 2020–2025. The authors report that there are no competing interests to declare.

SUPPLEMENTARY MATERIAL

Appendix: Proofs of Theorems 3.1, 4.1, 5.1; simulation results under serial dependence.

(.pdf file)

Code: R code that implements our testing procedure. (.R file)

References

- Beirlant, J., Goegebeur, Y., Teugels, J., and Segers, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley.
- Bekaert, G., Harvey, C. R., and Ng, A. (2005). Market integration and contagion. *J. Bus.*, 78(1):39–69.
- Bekaert, G., Hodrick, R. J., and Zhang, X. (2009). International stock return comovements. *J. Finance*, 64(6):2591–2626.
- Bormann, C. and Schienle, M. (2020). Detecting structural differences in tail dependence of financial time series. *J. Bus. Econ. Stat.*, 38(2):380–392.
- Bücher, A. and Dette, H. (2013). Multiplier bootstrap of tail copulas with applications. *Bernoulli*, 19(5A):1655–1687.
- Bücher, A., Jäschke, S., and Wied, D. (2015). Nonparametric tests for constant tail dependence with an application to energy and finance. *J. Econom.*, 187(1):154–168.
- 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(2):878–902.
- Christoffersen, P., Errunza, V. R., Jacobs, K., and Langlois, H. (2012). Is the potential for international diversification disappearing? A dynamic copula approach. *Rev. Financ. Stud.*, 25(12):3711–3751.
- de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer.
- de Haan, L. and Resnick, S. I. (1977). Limit theory for multivariate sample extremes. *Z. Wahrsch. verw. Gebiete*, 40(4):317–337.

- Delgado, M. A., Hidalgo, J., and Velasco, C. (2005). Distribution free goodness-of-fit tests for linear processes. *Ann. Statist.*, 33(6):2568–2609.
- Draisma, G., Drees, H., Ferreira, A., and de Haan, L. (2004). Bivariate tail estimation: Dependence in asymptotic independence. *Bernoulli*, 10(2):251–280.
- Einmahl, J., de Haan, L., and Sinha, A. (1997). Estimating the spectral measure of an extreme value distribution. *Stochastic Process. Appl.*, 70(2):143–171.
- Forbes, K. J. and Rigobon, R. (2002). No contagion, only interdependence: Measuring stock market comovements. *J. Finance*, 57(5):2223–2261.
- Joe, H. (2001). *Multivariate Models and Dependence Concepts*. Chapman & Hall/CRC, second edition.
- Khmaladze, E. V. (1981). Martingale approach in the theory of goodness-of-fit tests. *Theory Probab. Appl.*, 26(2):240–257.
- Khmaladze, E. V. (1988). An innovation approach to goodness-of-fit tests in R^m . *Ann. Statist.*, 16(4):1503–1516.
- Khmaladze, E. V. (1993). Goodness of fit problem and scanning innovation martingales. *Ann. Statist.*, 21(2):798–829.
- Khmaladze, E. V. and Koul, H. L. (2004). Martingale transforms goodness-of-fit tests in regression models. *Ann. Statist.*, 32(3):995–1034.
- King, M. A. and Wadhvani, S. (1990). Transmission of volatility between stock markets. *Rev. Financ. Stud.*, 3(1):5–33.
- Koenker, R. and Xiao, Z. (2002). Inference on the quantile regression process. *Econometrica*, 70(4):1583–1612.

- Koenker, R. and Xiao, Z. (2006). Quantile autoregression. *J. Amer. Statist. Assoc.*, 101(475):980–990.
- Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications*. Imperial College Press.
- Longin, F. and Solnik, B. (1995). Is the correlation in international equity returns constant: 1960–1990? *J. Int. Money Finance*, 14(1):3–26.
- Towghi, N. (2002). Multidimensional extension of L.C. Young’s inequality. *J. Inequal. Pure Appl. Math.*, 3(2).