Detecting structural differences in tail dependence of financial time series

Carsten Bormann

Karlsruhe Institute of Technology, Germany

Melanie Schienle

Karlsruhe Institute of Technology, Germany

Abstract

An accurate assessment of tail inequalities and tail asymmetries of financial returns is key for risk management and portfolio allocation. We propose a new test procedure for detecting the full extent of such structural differences in the dependence of bivariate extreme returns. The test decomposes the testing problem into piecewise multiple comparisons of Cramer-von-Misés distances of tail copulas. In this way, tail regions that drive differences in extreme dependence can be located and consequently be targeted by financial strategies. We derive the asymptotic properties of the test but also provide a multiplier bootstrap approximation for finite samples. Moreover, we account for the multiplicity of the piecewise tail copula comparisons by adjusting individual p-values according to multiple testing techniques. Extensive Monte Carlo simulations demonstrate the test's superior finite-sample properties for common financial tail risk models, both in the i.i.d. and the sequentially dependent case. In a high-dimensional S&P500 industries universe, we compare tail dependence of bivariate lower and upper tails of US Economic sector returns in a rolling window scheme. For the last 90 years, up to 20% more instances of tail asymmetries are detected due to the presence of tail dependence structures that competing standard methods are unable to capture. We also find evidence for diminishing tail asymmetries during every major financial crisis except for the 2007-09 crisis. Finally, for major foreign exchange rates from 2001-2016, we identify EUR-CHF as as the most tail dependent pair in both upper and lower tails, which prevails even after the Swiss National Bank unpegged the Franc from the Euro.

Keywords: Tail dependence, tail copulas, tail asymmetry, tail inequality, extreme values, multiple testing

JEL classification: C12, C53, C58

1 Introduction

Asymmetric dependence structures both within and between bivariate extreme returns under different market conditions are a key criterion for asset and risk management but also a main focus of market supervision. We provide a robust non-parametric statistical test for tail dependence differences. The test accurately detects all and the full extent of deviations between two tail dependence functions. Differences in tail dependence across different asset pairs are denoted as tail inequality, while tail asymmetry refers to different upper and lower tail dependence of a single asset pair. During financial crises, there is large empirical evidence that financial markets exhibit pronounced cross-sectional comovements in the tails of return distributions, thus the occurrence of joint extreme events vastly increases (see e.g. Longin and Solnik (2001); Ang and Chen (2002); Li (2013)). For investment strategies, this should be taken into account by timely and adequate re-allocations of assets e.g. profiting from arbitrage trading opportunities and by appropriate adjustments of hedging decisions. On the other hand, risk managers and market supervisors might need to set larger capital buffer requirements if the tendency for a joint occurrence of extreme losses rises in times of market distress. Our test procedure is based on multivariate extreme value techniques which remain valid during turbulent market periods, e.g. Mikosch (2006). In particular, we use a flexible non-parametric approach avoiding parametric misspecification risk. See e.g. Longin and Solnik (2001); Patton (2013); Jondeau (2015) for parametric approaches. Note in particular, that standard linear dependence measures are flawed under adverse market conditions which demands for alternative statistical models. Most prominently, the Gaussian copula is a convenient tool to model dependence near the mean of multivariate distributions. However, it is not capable of measuring dependence in the far tails (Embrechts (2009)).

This paper provides a novel non-parametric test procedure against pairwise differences in tail dependence structures which we measure with tail copulas (TCs) denoted by $\Lambda(u_1, u_2), (u_1, u_2) \in \mathbb{R}_+^2$. This is in sharp contrast to established approaches, which only estimate and compare summary measures of extreme dependence, such as the tail dependence coefficient (Hartmann et al. (2004), Straetmans et al. (2008)), or the tail index of aggregated tails (Ledford and Tawn (1996)). Specifically, we compare tail copulas on their entire domain in a locally piecewise way. Thus we study a multiple testing problem of tail copula equality. Piecewise testing allows to pin down specific regions where tail dependence differences are most apparent indicating specific types of extreme events that mainly induce the difference between both tail copulas. Moreover, our test assesses all possible and the complete amount of deviations from tail symmetry/equality. In particular, our test is still consistent if one (or both) of the two considered tail copulas is non-exchangeable, i.e. $\Lambda(u_1, u_2) \neq \Lambda(u_2, u_1)$. Existing procedures fail to address such intra-tail asymmetric (ITA) dependence structures. Therefore, for intra-tail asymmetric copulas, such tests suffer from low power. Our test builds on the idea of a two-sample goodness-of-fit test for tail copulas as in Bücher and Dette (2013). For increased sensitivity against violations of the null, however, we compare both tail copulas in a piecewise way on disjoint subintervals of the unit simplex hull. This piecewise testing amounts to a number of individual tests against tail dependence equality in different regions of the sample. For an accurate overall assessment, we use multiple testing principles such as the familywise error control and the False Discovery Rate in order to control the error rate of all marginal tests. Asymptotic properties of the test are provided. Moreover, a multiplier bootstrap procedure is suggested following the ideas of Bücher and Dette (2013) but extending it to non-i.i.d. financial data.

In an extensive simulation study with common tail factor and Clayton copulas, we find good finite sample properties of our test methodology both, for i.i.d. and sequentially dependent time series data. In standard cases, our test is slightly superior to competing tests, while it is much more powerful in case of intra-tail asymmetry. Simulation results strongly suggest that accounting for time series dynamics is essential by either GARCH-pre-filtering or by directly adjusting the bootstrap approximation employing ideas of Bücher and Ruppert (2013).

Finally, we establish tail asymmetry dynamics of 49 S&P500 industry portfolios for the last 90 years. We find empirical evidence that tail asymmetries substantially diminish in times of financial distress with the only strong exception of the 2007-2009 financial crisis which apparently was completely different in structure. Generally, our novel test detects up to 20% more tail asymmetries than competing tests. The latter can specifically be attributed to tail events not detected by standard tail dependence measures as the tail dependence coefficient (TDC) (Hartmann et al. (2004); Jondeau (2015)), or the tail copula based test by Bücher and Dette (2013). Thus, our test could serve as a more accurate tool for investors for assessing tail asymmetry in the market and in this sense could yield opportunities for improved portfolio strategies based on tail asymmetries. We also test pairs of six foreign exchange rates against tail inequalities during 2000-2016. Generally for the entire time period, the Euro-Swiss Franc pairs stands out with the strongest tail dependence. Interestingly, this strong tail dependence even continues after the sudden unpegging of the Franc by the Swiss national Bank on January 2015.

The paper is structured as follows. Section 2 introduces theoretical results on tail dependence necessary for the testing procedures. Section 3 introduces our testing technique. It also provides asymptotic properties and respective finite sample versions of the test procedures. Section 4 studies the finite sample performance in a thorough simulation study, and Section 5 provides detailed applications on subsectors indices of S&P500 and on data of the major foreign exchange rates. Finally, Section 6 concludes.

2 Basic notions and setup

We briefly introduce necessary fundamental definitions and results for tail dependence copulas which are required for the test procedure.

A two-dimensional return vector is denoted by $\mathbf{X} = (X^1, X^2)$. For outlining the basic test idea in this section, marginal returns $X^i, i = 1, 2$, are assumed to be independent and identically distributed (i.i.d.) with continuous distribution $F_i(x), i = 1, 2$, and quantile functions F_i^{-1}. This can be satisfied in practice with pre-whitening of observed returns and by taking the residuals from an appropriate time series fit as $\mathbf{X}$ in the test. Formally, we can also relax the independence assumption directly, as discussed in Section 3.2.

Our test is based on the full dependence distribution in the tails captured by a tail copula. Note that standard dependence measures such as point correlations, quantify the likelihood of aligned return movements of X^1 and X^2. However, if returns of both assets are extreme, i.e. $X^i > F_i^{-1}(1-t)$, or $X^i < F_i^{-1}(t), i = 1, 2$, for $t \to 0$, standard dependence measures are insufficient, and thus measures that focus on the tails should be used, Embrechts (2009). For example, the Gaussian copula, which is completely parametrized by the correlation coefficient, is unable to model tail dependence at all. That is to say, dependence may vary over different parts of the distribution, and correlations may be unable to measure dependence in between tails.

If $\mathbf{X}$ is in the domain of attraction of a two-dimensional extreme value distribution, there exists the so-called tail copula which measures the complete tail dependence between X^1 and X^2. The upper and lower tail copula $\Lambda^U_{\mathbf{X}}(u_1, u_2), \Lambda^L_{\mathbf{X}}(u_1, u_2), \mathbf{u} := (u_1, u_2), \mathbf{u} \in \mathbb{R}^2_+$, are defined by

$$\Lambda^U_{\mathbf{X}}(u_1, u_2) := \lim_{t \to 0} t^{-1} \mathbb{P}(X^1 > F_1^{-1}(1 - tu_1), X^2 > F_2^{-1}(1 - tu_2)), t \in \mathbb{R}_+$$
$$\Lambda^L_{\mathbf{X}}(u_1, u_2) := \lim_{t \to 0} t^{-1} \mathbb{P}(X^1 < F_1^{-1}(tu_1), X^2 < F_2^{-1}(tu_2)),$$

i.e. the tail copula measures the probability of a joint extreme given at least one component is extreme. See, among others, de Haan and Ferreira (2006), Schmidt and Stadtmüller (2006), for further details. If $\Lambda^U_{\mathbf{X}}(\mathbf{u}) > 0$ ($\Lambda^L_{\mathbf{X}}(\mathbf{u}) > 0$), gains (losses) of $\mathbf{X}$ are said to be tail dependent. For the sake of notational brevity we omit the superscripts U and L, unless it becomes important. With $\mathbf{u} = (1, 1)$, the tail copula boils down to the tail dependence coefficient (TDC), $\lambda \equiv \Lambda(1, 1)$. The TDC is a standard tool in financial applications to measure tail dependence, e.g. Frahm et al. (2005), Aloui, Aïssa and Nguyen (2011), Garcia and Tsafack (2011). However, the TDC covers only a fragment of tail dependence, namely dependence between joint quantile exceedances of marginals thresholds along the line $(F_1^{-1}(1-t), F_2^{-1}(1-t)), t \to 0$. In contrast, the tail copula varies marginal thresholds as $u_i \in \mathbb{R}_+, i = 1, 2$ and describes tail association for every possible tail event. It can be shown that $\Lambda_{\mathbf{X}}(u_1, u_2) \in [0, \min(u_1, u_2)]$, and $\Lambda_{\mathbf{X}}(a\mathbf{u}) = a\Lambda_{\mathbf{X}}(\mathbf{u}), a \in \mathbb{R}$. Due to this homogeneity of the tail copula it is sufficient to analyze $\Lambda_{\mathbf{X}}(\mathbf{u})$ with $\mathbf{u} \in \mathcal{S}$, where $\mathcal{S} := \{(u_1, u_2) : u_1, u_2 \geq 0, ||\mathbf{u}|| = c\}$ is, e.g., the unit simplex hull with $|| \cdot || = || \cdot ||_1$ and $c = 1$, or the unit circle hull with $|| \cdot || = || \cdot ||_2$ and $c = 1$. Without loss of generality we choose $\mathcal{S}$ to be the unit simplex hull. The homogeneity property prunes the relevant domain of the tail copula (i.e. from $\mathbb{R}_+^2$ to $\mathcal{S}$) and reduces computational efforts in estimation. The homogeneity property will lay the basis for our test. We assume the tail copulas exist and $\Lambda_{\mathbf{X}}(\mathbf{u}) > 0$, i.e. we assume tail dependent pairs as otherwise it is known non-parametric estimation of $\Lambda_{\mathbf{X}}$ overestimates the degree of tail dependence, Schmidt and Stadtmüller (2006).

We are interested in comparing two tail copulas, i.e. in tail copula differences. To formalize the discussion about tail copula differences and special cases such as tail asymmetry, and tail inequality, we introduce the following definitions and some notation. We say two tail copulas differ if

$$\{\Lambda_{\mathbf{X}}(u_1, u_2) \neq \Lambda_{\mathbf{Y}}(u_1, u_2)\} \quad \text{or} \quad \{\Lambda_{\mathbf{X}}(u_1, u_2) \neq \Lambda_{\mathbf{Y}}(u_2, u_1)\}, (u_1, u_2) \in \mathbb{R}_+^2. \tag{1}$$

For the homogeneity of the tail copula, it is sufficient to consider $(u_1, u_2) \in \mathcal{S}$. We write shorthand $\Lambda_{\mathbf{X}} \neq \Lambda_{\mathbf{Y}}$ for Equation 1. Tail asymmetry is given if two tail copulas of the same return vector differ. To detect tail asymmetry, one should compare $\Lambda_{\mathbf{X}}^U(u_1, u_2)$ with $\Lambda_{\mathbf{X}}^L(u_1, u_2)$ and also with the "flipped" version $\Lambda_{\mathbf{X}}^L(u_2, u_1)$. Tail inequality occurs between two return vectors, i.e. $\Lambda_{\mathbf{X}}^U \neq \Lambda_{\mathbf{Y}}^L$.

Definition 1 (Tail asymmetry (TA)) *A return vector* $\mathbf{X}$ *is tail asymmetric if* $\Lambda_{\mathbf{X}}^L \neq \Lambda_{\mathbf{X}}^U$.

Whenever the likelihood for comovements of extreme losses differs from that of extreme gains, the return vector $\mathbf{X}$ exhibits tail asymmetry. For example, in terms of Value at Risk (VaR) exceedances, $\Lambda_{\mathbf{X}}^L > \Lambda_{\mathbf{X}}^U$ implies joint exceedances of loss VaRs are more likely to occur than those of gain VaRs.

Definition 2 (Tail inequality (TI)) *Return vectors* $\mathbf{X}$ *and* $\mathbf{Y}$ *exhibit tail inequality if* $\Lambda_{\mathbf{X}}^W \neq \Lambda_{\mathbf{Y}}^Z, W, Z = U, L$.

The concept of tail inequality can be used to compare competing portfolios with respect to their sensitivity to extreme events. For example, $\Lambda_{\mathbf{X}}^L > \Lambda_{\mathbf{Y}}^L$ implies joint exceedances of loss VaRs for portfolio $\mathbf{X}$ are more likely to occur than those portfolio $\mathbf{Y}$, i.e. $\mathbf{X}$ exhibits a stronger tail risk between joint losses. Similarly, if, $\Lambda^U_{\mathbf{X}} < \Lambda^L_{\mathbf{Y}}$, joint extreme losses in portfolio $\mathbf{Y}$ are more intertwined than joint extreme gains in $\mathbf{X}$.

One reason for tail copula differences may be non-exchangeability of at least one of the tail copulas considered. We term non-exchangeability of a tail copula intra-tail asymmetry. A return vector $\mathbf{X}$ is intra-tail asymmetric if $\Lambda^W_{\mathbf{X}}(u_1, u_2) \neq \Lambda^W_{\mathbf{X}}(u_2, u_1), (u_1, u_2) \in \mathcal{S}, W = U, L$. ITA refers to one joint tail of $\mathbf{X}$ and occurs whenever the tail copula of that specific tail is not symmetric with respect to its arguments $\mathbf{u} = (u_1, u_2)$, i.e. if the tail copula is not exchangeable with respect to X^1 and X^2. For example, let $u_1 = 0.2, u_2 = 0.8$ and $t = 0.05$. Then, ITA is at hand if the tail event $\{X^1 > VaR_1(0.99)\} \cap \{X^2 > VaR_2(0.96)\}$ is differently likely from the tail event $\{X^1 > VaR_1(0.96)\} \cap \{X^2 > VaR_2(0.99)\}$. The following proposition illustrates the importance of ITA for comparisons of tail dependence functions.

Proposition 1 *If $\Lambda^W_{\mathbf{X}}(u_1, u_2), W = U, L,$ is intra-tail asymmetric, then $\Lambda^W_{\mathbf{X}} \neq \Lambda^H_{\mathbf{Z}}$, for $Z = X, Y$ and $H = U, L$.*[1]

Figure 2 illustrates this idea. If $\Lambda^W_{\mathbf{X}}(\mathbf{u}), W = U, L,$ is asymmetric with respect to $\mathbf{u}$, any comparison with that tail copula automatically amounts to tail asymmetry/inequality as there is always some point on the unit simplex hull where both tail copulas differ. While parametric models for intra-tail asymmetric tails exist, e.g. the asymmetric logistic copula in Tawn (1988), and factor copulas Einmahl et al. (2012), ITA is implicitly assumed to hold in all standard tests for tail dependence differences. However, we find this phenomenon should not be ruled out ex ante.

As the tail copula is the main component for our test, we sketch relevant statistical results. Non-parametric estimation of $\Lambda_{\mathbf{X}}(\mathbf{u})$ approximates marginal quantile functions $F^{-1}_{i,\mathbf{X}}, i = 1, 2,$ non-parametrically by the empirical counterpart $F^{-1}_{i,\mathbf{X},n}, i = 1, 2$. Further, the running variable t is replaced by $\frac{k}{n}$ with the sample size $n \to \infty$, and the so-called effective sample size $k \to \infty, k \in o(n)$ which has to be determined manually. A consistent estimator for $\Lambda^U(\mathbf{u})$ is

$$\widehat{\Lambda}^U_{\mathbf{X}}(u_1, u_2) = \frac{1}{k} \sum_{m=1}^{n} \mathbf{1} \left\{ X^1_m > F^{-1}_{1,\mathbf{X},n}(1 - (k/n)u_1), X^2_m > F^{-1}_{2,\mathbf{X},n}(1 - (k/n)u_2) \right\}, (u_1, u_2) \in \mathcal{S}.$$

An asymptotically equivalent estimator is given by

$$\widehat{\Lambda}^U_{\mathbf{X}}(u_1, u_2) = \frac{1}{k} \sum_{m=1}^{n} 1 \left\{ F_{1,\mathbf{X},n}(X^1_m) > 1 - (k/n)u_1, F_{2,\mathbf{X},n}(X^2_m) > 1 - (k/n)u_2 \right\},$$

[1]Assume $\Lambda^W_{\mathbf{X}}(u_1, u_2) = \Lambda^H_{\mathbf{Z}}(u_1, u_2)$. As $\Lambda^W_{\mathbf{X}}(u_1, u_2) \neq \Lambda^W_{\mathbf{X}}(u_2, u_1)$, it holds $\Lambda^W_{\mathbf{X}}(u_2, u_1) \neq \Lambda^H_{\mathbf{Z}}(u_1, u_2)$, and Equation 1 applies.

where $F_{i,\mathbf{X},n} = \frac{1}{n+1}\sum_{j=1}^{n} \mathbf{1}\{X_j^i \leq x\}$; dividing by $(n+1)$ improves estimation in finite samples. Estimators for $\Lambda_{\mathbf{X}}^L(\mathbf{u})$ are defined analogously.

Concerning asymptotic results for the empirical tail copula we state both assumptions and results as they are the backbone of the asymptotic distribution of our test statistic.

Assumptions 1 *For a bivariate random vector* $\mathbf{X}$*, assume*

(A1) $\mathbf{X} \sim_{iid} F_{\mathbf{X}}$.

(A2) $F_{\mathbf{X}}$ *is in the max-domain of a bivariate extreme value distribution with tail copula* $\Lambda > 0$.

(A3) $k \to \infty$ *and* $\frac{k}{n} \to 0$ *for* $n \to \infty$.

(A4) It holds that $|\Lambda(u_1, u_2) - tC_{\mathbf{X}}(u_1/t, u_2/t)| = O(A(t))$*, for* $t \to \infty$*, and some function* $A : \mathbb{R}_+ \mapsto \mathbb{R}_+$ *with* $\lim_{t\to\infty} A(t) = 0$ *and* $\sqrt{k}A(n/k) \to 0$ *for* $n \to \infty$*, where* $C_{\mathbf{X}}(u_1, u_2) := P(F_1(X^1) \leq u_1, F_2(X^2) \leq u_2)$ *denotes the copula of* $\mathbf{X}$.

(A5) The partial derivatives $\partial_i \Lambda = \frac{\partial \Lambda(x_1,x_2)}{\partial x_i}, (x_1, x_2) \in \mathbb{R}_+^2$ *exist and are continuous.*

Assumption (A1) is standard, yet restrictive for financial time series. It can, however, be satisfied by residuals from pre-filtering the dynamic structure of financial time series. We also show how (A1) may be relaxed to stationarity with a specific mixing rate allowing for a direct application of our testing procedure without pre-filtering, see Section 3.2. Assumption (A2) requires sample tails can be modelled by bivariate extreme value distributions and are asymptotic dependent, see de Haan and Ferreira (2006) for details. Standard distributions with actual tail dependence, such as the bivariate t-distribution with dispersion parameter $\rho \neq 0$ meet this assumption. Notably, the Gaussian copula violates (A2) due to tail independence ($\Lambda = 0$ for $|\rho| < 1$). Assumption (A3) imposes that the effective sample size k increases slower than n for $n \to \infty$, which is trivially fulfilled. Then the second-order condition (A4) (see Bücher and Dette (2013)) effectively requires a tail approximation order A which is a regular variation restriction and in practice imposes a corresponding slightly tighter condition on the expanding rate of k. For example, if $A(t)$ is asymptotically of order $\frac{1}{t^\alpha}$ with $\alpha > 0$, then k should be at most of order $n^{\frac{2\alpha}{1+2\alpha}} < n$ in order to satisfy the conditions. For completeness, we state assumption (A5); however, this smoothness assumption may also be omitted resulting in a more complex limiting behavior of the empirical tail copula, see Bücher et al. (2014). This permits consistent estimation of tail copulas of factor models, see Section 4. Under Assumptions (A1)-(A5) the asymptotic distribution for the tail copula can be derived as follows where $\xrightarrow{w}$ denotes weak convergence:

$$\sqrt{k}(\widehat{\Lambda}_{\mathbf{X}}(u_1, u_2) - \Lambda_{\mathbf{X}}(u_1, u_2)) \xrightarrow{w} \mathbb{G}_{\widehat{\Lambda},\mathbf{X}}(u_1, u_2), (u_1, u_2) \in \mathbb{R}_+^2, \tag{2}$$

where $\mathbb{G}^{U}_{\hat{\Lambda},\mathbf{X}}$ is a bivariate Gaussian field of the form $\mathbb{G}_{\hat{\Lambda},\mathbf{X}}(u_1,u_2) = \mathbb{G}_{\tilde{\Lambda},\mathbf{X}}(u_1,u_2) - \sum_{i=1}^{2} \partial_i\Lambda(u_1,u_2)\mathbb{G}_{\tilde{\Lambda},\mathbf{X}}(u_i, u_{-i} = \infty)$,, where $\mathbb{G}_{\tilde{\Lambda},\mathbf{X}}(u_1,u_2)$ is a *centered* Gaussian field with covariance $\mathbb{E}[\mathbb{G}_{\tilde{\Lambda},\mathbf{X}}(u_1,u_2)\mathbb{G}_{\tilde{\Lambda},\mathbf{X}}(v_1,v_2)] = \Lambda(\min(u_1,v_1),\min(u_2,v_2))$. These results were first established in Schmidt and Stadtmüller (2006); Bücher and Dette (2013) and Bücher et al. (2014) provide related results but without (A5), i.e. existence of partial derivatives of the tail copula is generally not needed. This is important in practice, for covering many empirically occurring cases of tail dependence.

3 A new testing methodology against tail asymmetry and inequality

3.1 Test idea and asymptotic properties

Generally, we test the global null hypothesis of equality between tail copulas by checking for local violations of the null over many disjoint subsets of the relevant support ($\mathcal{S}$) and in all possible directions. This localization provides additional insights on specific quantile areas which might be a valuable target for adequate risk or portfolio management strategies.

When testing against tail equality, our test takes into account that each of the return vectors could be intra-tail asymmetric. In case of intra-tail asymmetry, tests are only consistent if all possible permutations of arguments in the tail copulas ($\Lambda_{\mathbf{Z}}(u_1,u_2)$, and also $\Lambda_{\mathbf{Z}}(u_2,u_1), \mathbf{Z} = \mathbf{X},\mathbf{Y}$) are considered as only then null violations in all directions can be found. This contrasts sharply with the TDC based test by Hartmann et al. (2004), abbreviated as TDC-test, which only compares tail copulas at a single point of their domain, while we account for possible tail differences along the entire domain of both tail copulas. Our test is closely related to the test by Bücher and Dette (2013), abbreviated as BD13-test, which compares the tail copula of $\mathbf{X}$ with the tail copula of $\mathbf{Y} = (Y^1, Y^2)$ along the unit circle. However, as tail copula differences are only evaluated in one direction, their test statistic is not exchangeable, i.e. for the test statistic S it holds that $S(\mathbf{X},(Y^1,Y^2)) \neq S(\mathbf{X},(Y^2,Y^1))$. To fix this, we propose to analyze tail copula differences in "both directions" of the unit simplex hull, and thereby we search for differences between tail copulas over distinct, pre-determined subintervals of the unit simplex. Testing against tail equality over many different subintervals amounts to an entire collection of individual tests. If the null of equality of tails is rejected within a specific subset, this approach locates those sample regions that induce the tail dependence differences. Test power strongly benefits from intra-tail asymmetric tail copulas. Further, in standard, i.e. intra-tail symmetric cases, it features similar test properties as competing tests.

Note, the notation corresponds to the test against tail inequality. Yet, the test also applies for tail asymmetry by exchanging $\Lambda_{\mathbf{X}}$ by $\Lambda^{U}_{\mathbf{X}}$ and $\Lambda_{\mathbf{Y}}$ by $\Lambda^{L}_{\mathbf{X}}$.

Due to the homogeneity property of the tail copula, it is sufficient to compare tail copulas only over the unit simplex hull instead of $\mathbb{R}^2_+$ entirely. We denote the unit simplex hull by $\mathcal{S} := \{(u_1, u_2) : u_1 + u_2 = 1, u_i \geq 0, i = 1, 2\}$. We apply M Cramér-von-Mises (CvM) tests on $M/2$ disjoint subinterval of $\mathcal{S}$. The "global" null hypothesis is

$$H_0 : \Lambda_{\mathbf{X}} = \Lambda_{\mathbf{Y}} \text{ over } \mathcal{S},$$

consisting of M individual null hypothesis of the form

$$H_{0,m} : \Lambda_{\mathbf{X}}(u_1, 1-u_1) = \begin{cases} \Lambda_{\mathbf{Y}}(u_1, 1-u_1), & u_1 \in \mathcal{I}_m,\ m = 1, \ldots M/2 \\ \Lambda_{\mathbf{Y}}(1-u_1, u_1), & u_1 \in \mathcal{I}_{m-M/2},\ m = (M/2)+1 \ldots M, \end{cases}$$

where $\mathcal{I}_1, ..., \mathcal{I}_{M/2}$ are disjoint, equally lengthy subintervals of $[0,1]$ whose endpoints are evenly spaced in $[0,1]$. Note that $H_0 : \bigcap_{m=1}^{M} H_{0,m}$, i.e. global tail equality naturally implies tail equality over each subset. Marginal test statistics are given by

$$S^m(\mathbf{X}, \mathbf{Y}) = \begin{cases} \frac{k_X k_Y}{k_X + k_Y} \int_{\mathcal{I}_m} (\Lambda_{\mathbf{X}}(\phi, 1-\phi) - \Lambda_{\mathbf{Y}}(\phi, 1-\phi))^2 \, \mathrm{d}\phi, & j = 1, \ldots M/2 \\ \frac{k_X k_Y}{k_X + k_Y} \int_{\mathcal{I}_{m-M/2}} (\Lambda_{\mathbf{X}}(\phi, 1-\phi) - \Lambda_{\mathbf{Y}}(1-\phi, \phi))^2 \, \mathrm{d}\phi, & j = (M/2)+1 \ldots M. \end{cases}$$

Each marginal test corresponds to a specific subset of $\mathcal{S}$, which can be translated to a subspace of the sample. The switch of arguments in $\Lambda_{\mathbf{Y}}$ at subset $S^{(M/2)+1}$ guarantees that tail copulas are compared over the entire unit simplex, e.g. in "both directions". For the "switched" version of the test statistic, we omit the point $\phi = 1 - \phi$, to avoid comparing $\Lambda_{\mathbf{X}}(0.5, 0.5)$ with $\Lambda_{\mathbf{Y}}(0.5, 0.5)$ twice. If $H_{0,m}$ is true, $S^m = 0$, while $S^m > 0$ otherwise. Test statistics are estimated by replacing Λ by $\widehat{\Lambda}$. Empirical test statistics will be denoted by $\widehat{S}^m$.

The following proposition provides the marginal test distributions in the i.i.d. case. Section 3.2 discusses extensions for time series data.

Proposition 3.1 *Assume that assumptions (A1)-(A5) hold for* $\mathbf{X}, \mathbf{Y}$. *Then,*

(a)

$$(\widehat{S}^1, ..., \widehat{S}^M)' \xrightarrow{w} (S^1, ..., S^M)',$$

with

$$S^m = \int_{\mathcal{I}_m} \left(\sqrt{\frac{k_X}{k_X + k_Y}} \mathbb{G}_{\widehat{\Lambda}, \mathbf{X}}(\phi, 1-\phi) - \sqrt{\frac{k_Y}{k_X + k_Y}} \mathbb{G}_{\widehat{\Lambda}, \mathbf{Y}}(\phi, 1-\phi) \right)^2 d\phi, \phi \in \mathcal{I}_m,$$

$m = 1, ..., M$.

(b) Under H_0*,*

$$S^m \xrightarrow{w} 0, \forall m.$$

(c) Under H_1*,*

$$\exists i : S^i \xrightarrow{w} c,$$

where $c \in \left(0, \int_{\mathcal{I}_i} \min(\phi, 1-\phi)^2 d\phi\right]$ *drives local power.*

Note, the process $\mathbb{G}(\phi, 1-\phi)$ corresponds to $\mathbb{G}_{\widehat{\Lambda}}(u_1, u_2)$ from Equation (2). This guarantees marginal test statistics asymptotically follow well-defined stochastic processes, i.e. this result guarantees test consistency. Due to the complexity of these limiting stochastic processes, closed forms of the asymptotic distributions do not exist and have to be simulated. We follow Bücher and Dette (2013) and approximate the distribution of $(S^1, ..., S^M)'$ by a multiplier bootstrap. Further notation is required to construct the bootstrap distribution. The bth bootstrap estimate of $\widehat{S}^m$ is

$$\widehat{S}^{m,(b)}(\mathbf{X}, \mathbf{Y}) = \frac{k_X k_Y}{k_X + k_Y} \int_{\mathcal{I}_m} \Big((\widehat{\Lambda}^{(b)}_{\mathbf{X}}(\phi, 1-\phi) - \widehat{\Lambda}_{\mathbf{X}}(\phi, 1-\phi)) - \\ (\widehat{\Lambda}^{(b)}_{\mathbf{Y}}(\phi, 1-\phi) - \widehat{\Lambda}_{\mathbf{Y}}(\phi, 1-\phi)) \Big)^2 \, \mathrm{d}\phi,$$

where $\widehat{\Lambda}^{(b)}_{\mathbf{Z}}(\mathbf{u})$ is the multiplier bootstrap version of $\widehat{\Lambda}_{\mathbf{Z}}(\mathbf{u}), \mathbf{Z} = \mathbf{X}, \mathbf{Y}$,

$$\begin{aligned} \widehat{\Lambda}^{(b)}_{\mathbf{Z}}(u_1, u_2) &= \frac{1}{k} \sum_{i=1}^{n} \tilde{\xi}^{\mathbf{Z}}_i \mathbf{1} \left\{ Z^1_i \geq \tilde{F}^{-1}_{1,\mathbf{Z},n}(k u_1/n), Z^2_i \geq \tilde{F}^{-1}_{2,\mathbf{Z},n}(k u_2/n) \right\}, \\ \tilde{\xi}^{\mathbf{Z}}_i &= \xi^{\mathbf{Z}}_i / \overline{\xi^{\mathbf{Z}}}, i = 1, ..., n, \\ \tilde{F}_{j,\mathbf{Z},n}(u) &= \frac{1}{n} \sum_{i=1}^{n} \tilde{\xi}^{\mathbf{Z}}_i \mathbf{1} \left\{ Z^j_i \leq u \right\}, j = 1, 2, \end{aligned}$$

and $\xi_i, i = 1, ..., n,$ are i.i.d. random variables, called multipliers, with $\mathbb{E}[\xi_i] = \mathbb{V}[\xi_i] = 1$. This bootstrap technique guarantees weak convergence of $\widehat{S}^{m,(b)} - \widehat{S}^m$ to $\widehat{S}^m - S^m$, conditional on the bootstrap samples and conditional on observed sample of $\mathbf{X}$ and $\mathbf{Y}$. This means the asymptotic distributions of the bootstrap statistics converge to the asymptotic distributions of the empirical test statistics, and can be used to mimic the marginal null distributions in Proposition 3.1.

Proposition 3.2 *Let (A1)-(A5) hold. Further, chose the multiplier variables* $\xi = (\xi_1, ..., \xi_n)$ *as follows:*

(A6) $\xi_i, i \in \mathbb{Z}_+$*, are i.i.d. random variables,* ξ_i *are independent of* $\mathbf{X}, \mathbf{Y}$*,* $\mathbb{E}[\xi_i] = \mathbb{V}[\xi_i] = 1$.

Then

$$(\widehat{S}^{1,(b)} - \widehat{S}^1, ..., \widehat{S}^{M,(b)} - \widehat{S}^M)' \xrightarrow{w} (S^1 - \widehat{S}^1, ..., S^M - \widehat{S}^M)'.$$

This result provides a feasible bootstrap approximation of the the test distribution. For the i.i.d case, we set $\xi_i \sim Exp(1)$.[2] Finally, a consistent Monte Carlo P-value for hypothesis $H_{0,m}$ is given by

$$\hat{p}^m = \frac{1 + \sum_{b=1}^{B} \mathbf{1}\{\widehat{S}^{m,(b)} \geq \widehat{S}^m\}}{B+1}.$$

Joint testing of M hypothesis requires an adjustment of the individual test level α to control the error rate of the global hypothesis, α^*, say. Common error rates are the familywise error rate (FWER) and the false discovery rate (FDR). We prefer to control the FDR.

In general, for a family of M individual hypotheses $H_{0,1}, H_{0,2}, ..., H_{0,M}$, FDR controls for the expected number of falsely rejected marginal null hypotheses among all rejections, i.e.

$$\mathbb{E}\Big(\frac{\sum_{m=1}^{M} \mathbf{1}\{p^m \leq \alpha^m | H_{0,i}\}}{\sum_{m=1}^{M} \mathbf{1}\{p^m \leq \alpha^m\}}\Big) \leq \alpha^*.$$

The Benjamini-Hochberg algorithm (Benjamini and Hochberg (1995)) sorts all P-values $p^{(1)}, ..., p^{(M)}$, starting with the smallest one, and compares $p^{(i)}$ with $\frac{i}{M}\alpha$ where i denotes the rank of P-value $p^{(i)}$. If $p^{(i)} < \frac{i}{M}\alpha$, marginal hypotheses corresponding to P-values $p^{(1)}, ..., p^{(i)}$ are rejected. Adjusted P-values are $\tilde{p}^{(i)} = p^{(i)}\frac{M}{i}$ and compared with α^*.

The FWER controls for the probability of at least false rejection at a prefixed threshold α, say $\alpha^* = 5\%$, i.e.

$$\mathbb{P}(\cup_{m=1}^{M} \{p^m \leq \alpha^m | H_{0,m}\}) \leq \alpha^*,$$

where p^m denotes the marginal P-value and α^m is determined by the multiple testing method such that the inequality holds. For the well-known Bonferroni control, $\alpha^m = \alpha/M$. Equivalently, individual P-values are adjusted as $\tilde{p}^m = p^m M$ and marginal hypotheses are rejected if $\tilde{p}^m < \alpha^*$.

In general, controlling the BH-FDR control is not as conservative as the Bonferroni correction. Also, BH-FDR is better suited for (positively) dependent P-values, which is a natural assumption for our setting. However, we find in our simulations test performance is only slightly affected by the choice of error rate, and thus we choose BH-FDR with $\alpha^* = 0.05$. See Romano and Wolf (2005) for an overview of multiple testing methods with applications to financial data.

The practical implementation of the basic test works as follows.

Test algorithm 1

[2] Note, whenever $\mathbf{X}$ and $\mathbf{Y}$ are dependent, one has to use the same multiplier series for both $\mathbf{X}$ and $\mathbf{Y}$.

1. *Determine $k_{\mathbf{X}}, k_{\mathbf{Y}}$, and estimate both tail copulas, i.e. calculate $\widehat{\Lambda}_{\mathbf{X}}(\phi, 1-\phi), \widehat{\Lambda}_{\mathbf{Y}}(\phi, 1-\phi), \phi \in [0,1]$.*
2. *Set M. Decompose $[0,1]$ into $M/2$ disjoint, equally sized subintervals, i.e. $\mathcal{I}_1 = [a_1, b_1], ..., \mathcal{I}_{M/2} = [a_{M/2}, b_{M/2}], a_i < b_{i+1}$.*
3. *Calculate $\widehat{S}^m, m = 1, ..., M$.*
4. *Determine B. Calculate $\widehat{S}^{m,(b)}, m = 1, ..., M$, for $b = 1, ..., B$.*
5. *Calculate $\hat{p}^m, m = 1, ..., M$.*
6. *Fix an error rate α. Apply a multiple testing routine on $\hat{p}^1, ..., \hat{p}^M$ and decide on the global null hypothesis.*

This test is, independent of the multiple testing method, asymptotically valid. E.g. for the FDR it holds that $FDR \leq \alpha$, and in case of FWER the test has asymptotic correct size, i.e. $\lim_{n,B\to\infty} \mathbb{P}\big(\cup_{m=1}^{M} \{\hat{\tilde{p}}^m \leq \alpha^* | H_{0,m}\}\big) = \alpha^*$, where $\hat{\tilde{p}}^m$ denotes the estimated adjusted marginal P-value. We use $B = 1499$ bootstrap repetitions; note the necessary correction of B (1499 instead of 1500) which ensures consistency of the P-value. Unless otherwise stated, we discretize $[0,1]$ as $\mathcal{I}_n = \{0.01j\}_{j=1}^{99}$. We typically apply test algorithm 1 with at most $M = 26$ marginal hypotheses, which discretizes $[0,1]$ into 13 (preferably) equally sized subintervals.[3] The choice of M is subject to a trade-off between test power and precision of localization of tail differences (see Section **??**). A larger M amounts lower power as less data fall into finer subintervals, and the multiplicity penalty of the individual P-values increases in M making rejections less likely. A larger M also means, the tests very precisely pins down in subinterval of $[0,1]$ significant tail dependence differences are at hand. In the extreme case where $M \to \infty$ the test algorithm carries out an infinite number of TDC-type tests. While this is a theoretically valid test, test power would decrease as the harsh P-value adjustment would almost never suggest a test rejection due to the strong multiplicity penalty. Simulations suggest a choice of $M = 26$ is reasonable as this also keeps computational effort manageable. However, we do not strive to determine an "optimal" number of subsets but suggest to apply the test for a reasonable grid of M, say, $M = 2, 4, 6, ..., 26$. Note, if $M = 1$, we arrive at a generalized BD13-test that compares both TCs along both directions of $\mathcal{S}$. It will turn out to be fruitful to combine P-values of different grids to one embracing test as follows.

Test algorithm 2

1. *Execute test algorithm 1 for J grids which increase in grid fineness, i.e. $M = 2, 4, ..., 2(J+1)$.*

[3] For $M = 26$, $\mathcal{I}_1 = \{0.01, 0.02, ..., 0.09\}, \mathcal{I}_2 = \{0.10, 0.11, ..., 0.17\}, ..., \mathcal{I}_{13} = \{0.93, ..., 0.99\}$. Note, subsets may consist of a varying number of points.

2. *For each grid adjust the P-values for multiplicity:* $(\tilde{p}_1^2, \tilde{p}_2^2), ..., (\tilde{p}_1^{2(J+1)}, ..., \tilde{p}_J^{2(J+1)})$.
3. *For each grid pick the minimal adjusted P-value:* $(p_2^* = \min(\tilde{p}_1^2, \tilde{p}_2^2), ..., p_{2(J+1)}^* = \min(\tilde{p}_1^{2(J+1)}, ..., \tilde{p}_J^{2(J+1)}))$.
4. *Reject the global* H_0 *if* $\cup_{j=1}^J \{p_j^* \leq \alpha\}$.

Note, this aggregating test does not adjust the grid-specific P-values a second time. This approach controls for the error rate α, if $\tilde{p}_2^*, ..., \tilde{p}_{2(J+1)}^*$ are tail comonotonic. All P-values stem from the identical bootstrap sample, and correspond to identical nulls which naturally implies very strong dependence. However, reconstructing the joint distribution of the minimal P-values is hardly possible. Imposing a parametric model for the joint distribution of P-values is a standard procedure in multiple testing theory, see Dickhaus (2014); to ensure test 2 obeys the desired error rate we have to assume a certain degree of dependence in the lower part of the joint null distribution. With increasing dependence, test algorithm 2 asymptotically approaches the nominal error rate while being systematically biased if dependence is too weak. More formally, we write δ for the difference between the lower-tail copula of the P-values and the upper (Fréchet-Hoeffing) copula bound in case of comonotonicity, i.e. $\delta =: P(\tilde{p}_2^* \leq x_2, ..., \tilde{p}_{2(J+1)}^* \leq x_{2(J+1)} | x_i \leq \alpha, i = 2, 4, ..., 2(J+1)) - \min(x_2, ..., x_{2(J+1)})$. It holds that

$$\lim_{\delta \searrow 0} \mathbb{P}\Big(\bigcup_{i=2}^{2(J+1)} \{\tilde{p}_i^* \leq \alpha | H_{0,i}\} \Big) = \alpha.$$

The simulation study confirms test algorithm 2 abides the imposed error rate in finite samples implying the assumption of tail monotonicity is reasonable.

3.2 Inference for serially dependent data

The i.i.d. assumption is unreasonable for financial time series as financial data typically exhibit serial dependence. However, standard extreme value theory and the multiplier bootstrap rely on the independence assumption. We know of two approaches to address the problem of dependent data.

The standard approach is to fit financial returns to an appropriate time series model, such as a ARMA-GARCH model, to compute standardized residuals. The latter should roughly resemble an i.i.d. series, and can be used for further inference. See McNeil and Frey (2000) who proposed this method in a univariate setting. However, we do not know of any results that provide rigorous proof for convergence when using estimated residuals.

For empirical copulas of dependent data another remedy is to assume stationarity coupled with some mixing conditions which consequently allows to use unfiltered returns for estimation. Valid statistical inference is ensured by adjusting the bootstrap: For strongly mixing time series, convergence of the block bootstrap and the so-called tapered block multiplier bootstrap has been shown for the empirical copula process (Bücher and Ruppert (2013)). Necessary assumptions are met for a wide class of time series models, such as AR and GARCH models. We suggest to use the dependent data bootstrap methodology also for empirical tail copulas. Yet, we do not prove the validity of this approach as this a difficult task is beyond the scope of this paper. However, assumption (A3) puts the tail copula process close to the scaled copula process in the respective tail – in finite samples, where the running variable t has to be replaced by $\frac{k}{n}$, and both $k, n > 0$ are fixed, differences between $\Lambda(u_1, u_2)$ and $tC(u_1/t, u_2/t)$ might be neglectable. This suggests that results of the empirical copula process $(C_{\mathbf{X},n}(u_1, u_2))$ carry over to the empirical tail copula process $(\widehat{\Lambda}(u_1, u_2))$. We employ the tapered block multiplier and the block bootstrap for tail copula estimation. For completeness results of the previous section are adopted for the tapered block multiplier bootstrap. The i.i.d. assumption (A1) is replaced by

(A1*) $\mathbf{X}, \mathbf{Y}$ are realizations of a strictly stationary process that is strongly mixing with rate $\alpha_{\mathbf{Z}} = O(r^{-a}), a > 1, \mathbf{Z} = \mathbf{X}, \mathbf{Y}$,[4]

Consequently, under (A1*,A2-A5), the empirical tail copula should converge to some centered Gaussian process $G_\alpha(u_1, u_2)$ that is governed by the mixing rate α, i.e.

$$\sqrt{k}(\widehat{\Lambda}_{\mathbf{X}}(u_1, u_2) - \Lambda_{\mathbf{X}}(u_1, u_2)) \xrightarrow{w} \mathbb{G}_\alpha(u_1, u_2), (u_1, u_2) \in \mathbb{R}_+^2.$$

Bücher and Ruppert (2013) provide detailed advise on implementation strategies for empirical copula processes.[5] The functional delta theorem ensures convergence of $\widehat{\Lambda}_{\mathbf{Z}}, \mathbf{Z} = \mathbf{X}, \mathbf{Y}$ carry over to the test statistics $(\widehat{S}^1, ..., \widehat{S}^M)'$. To approximate the limiting behavior of the test statistics, now the tapered block multiplier bootstrap has to be applied. The tapered block multiplier bootstrap generates series of block-dependent multipliers instead of using i.i.d. multipliers. The following conditions have to be met for the consistency of the tapered block multiplier bootstrap in case of the empirical copula process, see Theorem 3 in Bücher and Ruppert (2013).

(A7) The underlying stochastic process of $\mathbf{Z}$ is strongly mixing with $\sum_{r=1}^{\infty}(r+1)^c\sqrt{\alpha_{\mathbf{Z}}(r)} < \infty, c = \max(8d + 12, \lfloor 2/\epsilon \rfloor + 1)$.

[4]The mixing coefficient is defined as $\alpha_{\mathbf{Z}}(r) = \alpha_{\mathbf{Z}}(\mathcal{F}_s, \mathcal{F}_{s+r}) = \sup_{A\in\mathcal{F}_s, B\in\mathcal{F}_{s+r}} |P(A \cap B) - P(A)P(B)|$, where $\mathcal{F}_t$ denotes the filtration of the underlying stochastic process up to time point t, and $\mathbf{Z}$ is strongly mixing if $\alpha_{\mathbf{Z}}(r) \to 0$ for $r \to \infty$.
[5]The authors suggest to fix a block length of $l(n) = 1.25n^{1/3}$ for the block bootstrap. Moreover, for the tapered block multiplier bootstrap, we employ their uniform kernel κ_1, and use $\Gamma(q, q)$ distributed base multipliers, with $q = 1/(2l(n) - 1)$, where $l(n)$ is the multiplier block length, which can be automatically determined using the R-package `npcp`, Kojadinovic (2015).

(A8) The tapered block multiplier process $(\xi_{j,n})_{j=1,\dots,n}$ is strictly stationary, has bounded moments, is independent of $\mathbf{Z}$, and positively $cl(n)$-near epoch dependent[6], where c is some constant and $l(n) \to_{n\to\infty} \infty, l(n) = o(n)$, and for all $j, h \in \mathbb{Z}$, assume $\mathbb{E}[\xi_{j,n}] = \mu > 0, \mathbb{V}[\xi_{j,n}, \xi_{j+h,n}] = \mu^2 v(h/l(n))$ and v is a bounded function symmetric around zero, and w.l.o.g. $\mu = 1, v(0) = 1$.

(A9) For the tapered block length $l(n) \to \infty$, where $l(n) = O(n^{1/2-\epsilon}), 0 < \epsilon < 0.5$.

Now, under (A1*,A2-A9), the tapered block multiplier bootstrap version of the test statistics, $\widehat{S}^{i,(b),tap}$, converge weakly to the counterpart of the original sample, i.e.

$$(\widehat{S}^{1,(b),tap} - \widehat{S}^1, ..., \widehat{S}^{M,(b),tap} - \widehat{S}^M)' \overset{w}{\to} (S^1 - \widehat{S}^1, ..., S^M - \widehat{S}^M)'.$$

The simulation study underlines the validity of the tapered multiplier bootstrap for the empirical tail copula. An advantage of this approach is the tail dependence structure is not polluted due to model misspecification which may be a problem for large, high-dimensional data sets where automatic GARCH fitting is challenging and computationally expensive.

3.3 Local tail asymmetry

One main feature of our test is we can *localize* tail dependence differences. This enriches the binary test decision on tail asymmetry/inequality as we can find subspaces in $\mathbb{R}^2_+$ where tail asymmetry/inequality can be expected. If the global null is rejected, significant individual P-values trace the subsets of the unit simplex hull where both tail copulas differ. The boundary points of the significant subsets amount to empirical quantile threshold vectors which span a tail asymmetric subspace in the sample space, i.e.

$$Q_{\mathbf{X}} = \left(F^{-1}_{1,\mathbf{X},n}(1 - k/nu_1), F^{-1}_{1,\mathbf{X},n}(1 - k/nu_2)\right) \times \left(F^{-1}_{2,\mathbf{X},n}(1 - k/nu_1)), F^{-1}_{2,\mathbf{X},n}(1 - k/nu_2)\right),$$
$$Q_{\mathbf{Y}} = \left(F^{-1}_{1,\mathbf{Y},n}(1 - k/nu_1), F^{-1}_{1,\mathbf{Y},n}(1 - k/nu_2)\right) \times \left(F^{-1}_{2,\mathbf{Y},n}(1 - k/nu_1)), F^{-1}_{2,\mathbf{Y},n}(1 - k/nu_2)\right).$$

Due to the homogeneity of the tail copulas, these extreme sets can be extrapolated arbitrarily far into the tail, given the extreme value conditions hold. In particular, Figure 1 illustrates how to trace tail asymmetry.

Thus, when comparing tail dependencies of return vectors, our test provides precise information on which specific tail events, or Value-at-Risk events, induce tail dependence differences. Conditional on realized returns of $\mathbf{X}$ falling into $Q_{\mathbf{X}}$, tail dependence of $\mathbf{X}$ and $\mathbf{Y}$ differ; conditional on $\mathbf{X} \notin Q_{\mathbf{X}}$, $\Lambda_{\mathbf{X}}$ and $\Lambda_{\mathbf{Y}}$ do not differ significantly.

[6]I.e. for fixed j, $\xi_{j,n}$ is independent of $\xi_{j+h,n}$ for all $|h| \geq cl(n)$

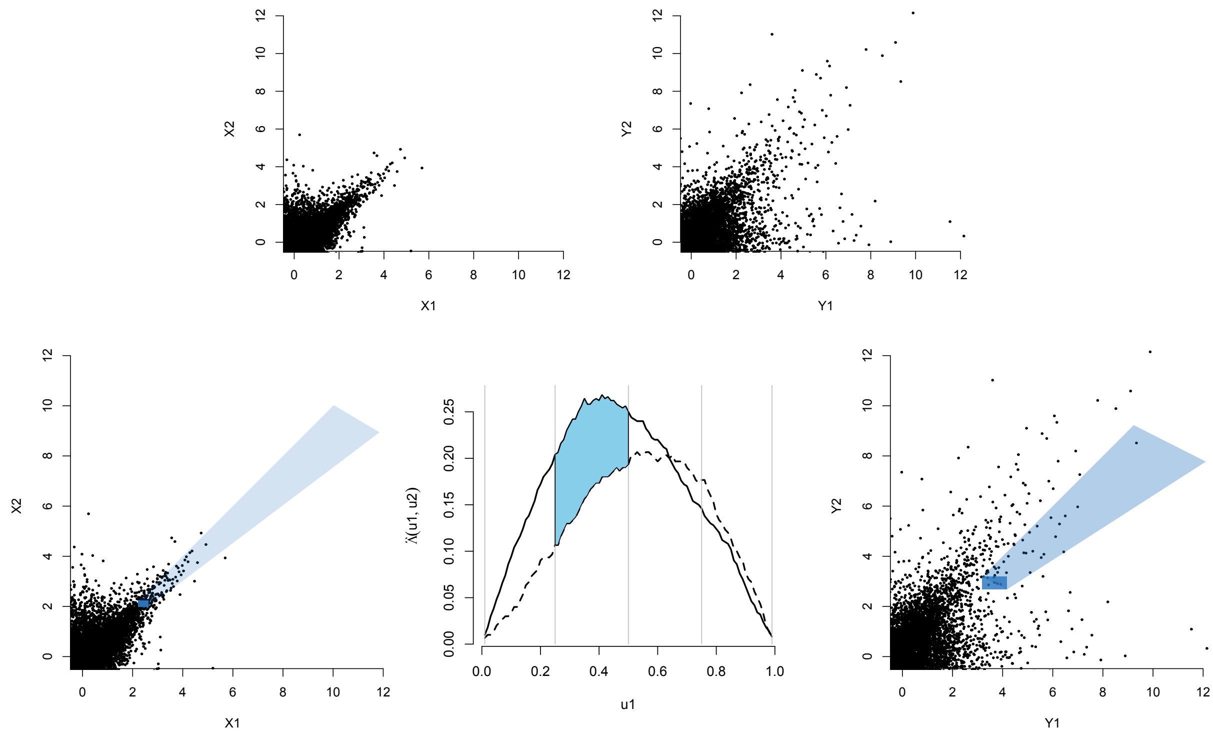

Figure 1: Top: Upper-right quadrants of scatterplots for $\mathbf{X}, \mathbf{Y}$, both equipped with an asymmetric logistic copula and marginal distributions $X_i \sim t(df = 3)$, $Y_i \sim t(df = 10), i = 1, 2$. The corresponding tail copula is $\Lambda(u_1, u_2) = u_1 + u_2 - \left[(1-\psi_1)u_1 + (1-\psi_2)u_2 + ((\psi_1 u_1)^{-\theta_1} + (\psi_2 u_2)^{-\theta_2})^{\theta}\right]$ (see Tawn (1988)), with parameters $(\psi_{\mathbf{X},1}, \psi_{\mathbf{X},2}, \theta_{\mathbf{X}}) = (0.1, 0.6, 0.1)$, $(\psi_{\mathbf{Y},1}, \psi_{\mathbf{Y},2}, \theta_{\mathbf{Y}}) = (0.1, 0.5, 0.4)$. Bottom, center: Estimated tail copulas for $u_1 \in \{0.01, 0.02, ..., 0.99\}, k = 500, n = 10.000, M = 8$. The blue shaded area indicates over which subset both tail copulas significantly differ. Bottom left and right: The blue rectangles show the tail-asymmetric tail regions; the homogeneity of the tail copula allows to extrapolate this region far into the sample tail.

This additional information might improve tail risk anticipation for regulators, or tail risk based hedge and trading strategies for investors.

4 Monte Carlo simulation

We compare finite sample performance of our test with the TDC-test, and the BD13-test.[7] For this we study two types of dependence models that are frequently used in finance. First, we employ the (implicit) copula of factor models. See Fama and French (1992), Einmahl et al. (2012), and Oh and Patton (2015) for factor models in finance, tail dependence of factor models, and tail dependence of factor copulas in finance, respectively. Secondly, representing the broad

[7]We focus on non-parametric tests only as in practice parametric specifications often suffer from a substantial model bias. Only if the chosen model class is appropriate, parametric tests might provide performance improvements. But even then, common parametric tests would fail to detect tail dependence differences in case of intra-tail asymmetry which, however, could be uncovered by our test procedure.

class of Archimedean copulas, we employ the Clayton copula, which models solely lower tail dependence. Its lean parametric form makes the Clayton copula a popular building block of more complex copula models, such as mixtures of copulas, Rodriguez (2007), Patton (2006). For each copula, we impose one parametrization that fulfils the null, and one that violates the null, leaving us with four DGPs.

DGPs 1 and 2 are based on the tail factor model for $\mathbf{X} = (X^1, X^2)$ with r factors $Z^j, j = 1, ..., r$, and loadings $a_{ij}, i = 1, 2, j = 1, ..., r$ given by

$$X^i = \sum_{j=1}^{r} \mathbf{a}_{ij} Z^j + \varepsilon^i, i = 1, 2. \tag{3}$$

where factors are i.i.d. Frèchet with $\nu = 1$ independent of the error term ε^i with less fat tails than Z^j; we set ε^i as Frèchet with $\nu_\varepsilon = 2$. In this way, the matrix of factor loadings $A = (a_{ij})$ directly determines the degree of tail dependence of $\mathbf{X}$ on the factors and the corresponding tail dependence function can be easily derived. In particular, the (upper) tail copula of (X^1, X^2) is equivalent to the tail copula of the max-factor model $\bar{X}^i = \max_{j=1,...,r}(a_{ij} Z^j)$ which is

$$\Lambda^U(u_1, u_2) = u_1 + u_2 - \sum_{j=1}^{r} \max\left(\frac{a_{1j}}{\sum_{j=1}^{r} a_{1j}} u_1, \frac{a_{2j}}{\sum_{j=1}^{r} a_{2j}} u_2 \right),$$

see Einmahl et al. (2012) for further details. DGP 1 consists of $\mathbf{X}, \mathbf{Y}$ both resulting from a factor model (3) with loading matrix

$$A_1 = \begin{bmatrix} 2 & 0 \\ 1 & 1 \\ 0 & 2 \end{bmatrix}$$

Here, the first factor only influences X^1 (Y^1), the second factor influences both X^1 (Y^1) and X^2 (Y^2), and the third factor only influences X^2 (Y^2). That is, A_1 amounts to intra-tail symmetry and to tail equality between $\mathbf{X}$ and $\mathbf{Y}$, and thus the null is true. See Figure 2, first from the left, for $\Lambda(u_1, 1 - u_1), u_1 \in [0, 1]$. For DGP 2, both $\mathbf{X}$ and $\mathbf{Y}$ stem from a factor model (3) with

$$A_2 = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix},$$

where the second factor only influences X^2 (Y^2), causing the TC to become intra-tail asymmetric, $\Lambda(u_1, u_2) \neq \Lambda(u_2, u_1)$, and consequently TCs of $\mathbf{X}$ and $\mathbf{Y}$ differ on some parts of $\mathcal{S}$, see Figure 2, second from the left. DGP 2 thus represents the class of intra-tail asymmetric copulas which violate the null according to Proposition 1.

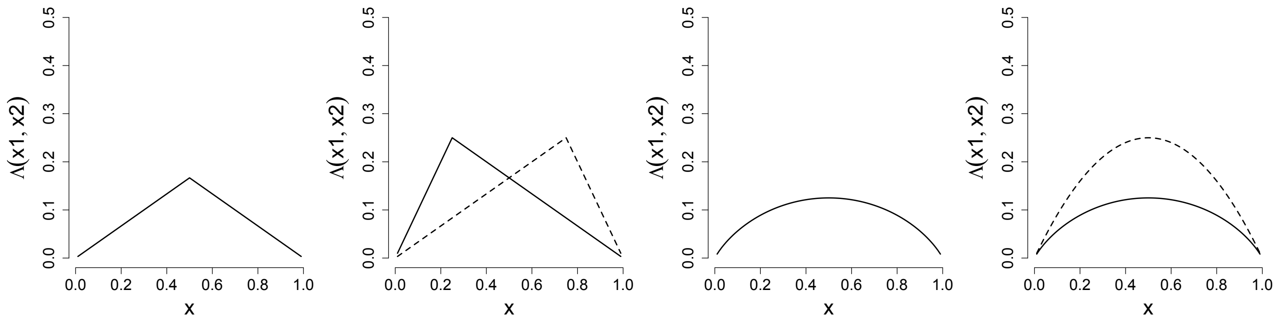

Figure 2: Tail copulas of DGPs 1 to 4 from left to right. Note, for DGP 2, the solid lines represents $\Lambda(u_1, 1-u_1)$, whereas the dashed line shows $\Lambda(1-u_1, u_1)$. For DGP 4, two different specifications of the Clayton copula are used for X and Y.

For the Clayton copula, only the lower TC features tail dependence,

$$\Lambda^L(u_1, u_2; \theta) = (u_1^{-\theta} + u_2^{-\theta})^{-1/\theta},$$
$$\Lambda^U(u_1, u_2; \theta) = 0,$$

where (lower) tail dependence increases in the parameter $\theta \in [0, \infty)$. DGP 3 is given by $\mathbf{X}, \mathbf{Y} \sim$ Clayton$(\cdot; \theta = 0.5)$; this specific choice of θ implies a TDC of $\lambda = 0.25$, which roughly corresponds to a TDC of a bivariate t-distribution with correlation 0.5 and 4 degrees of freedom (McNeil et al. (2005), p.211). For DGP 3, the null is true. See Figure 2, second from the right. For DGP 4, $\mathbf{X} \sim$ Clayton$(\cdot; \theta = 0.5)$, and $\mathbf{Y} \sim$ Clayton$(\cdot; \theta = 1)$. Thus, tail equality is violated as the TDC of $\mathbf{Y}$ is $\lambda = 0.5$. See Figure 2, first from the right.

To check whether the test also works for financial time series data, we combine all DGPs with i.i.d. as well as GARCH marginals. For the latter case, apply the test to "raw" GARCH returns, and to standardized GARCH residuals as it is important to analyze whether using *estimated* residuals affect test performance. Moreover, we study the test performance for unfiltered returns using the block bootstrap and the tapered block multiplier bootstrap. In particular, we employ GARCH(1,1) dynamics for any marginal return process. For both bivariate return series $\mathbf{r} = (r_1, r_2)$ it holds

$$r_{i,t} = \sigma_{i,t}\epsilon_{i,t},$$
$$\sigma_{i,t}^2 = \omega + \alpha_1 \sigma_{i,t-1}^2 + \beta_1 \epsilon_{i,t-1}^2,$$
$$\epsilon_{i,t} | \mathcal{F}_{t-1} \sim (0, \sigma_{i,\epsilon}^2), \qquad i = 1, 2, \quad t = 1, ..., n,$$

where we set $\omega = 0.01, \alpha_1 = 0.15$ and $\beta = 0.8$ such that $\omega + \alpha_1 + \beta$ is close to one to mimic parameter values often found in financial returns, see for example Engle and Sheppard (2001). Imposing the tail structures of DGPs 1 to 4 on the test series, we use DGPs 1 to 4 to generate $\epsilon_{\mathbf{Z},t} = (\epsilon_{\mathbf{Z}1,t}, \epsilon_{\mathbf{Z}2,t})'$: In a first step, we simulate observations $(\tilde{\epsilon}_{\mathbf{X},t}, \tilde{\epsilon}_{\mathbf{Y},t})_{t=1}^{n}$ according to DGPs 1 to 4. Consequently, we transform the errors to pseudo-observations by means of the marginal empirical cumulative distribution, $(F_{n\tilde{\epsilon}\mathbf{Z}1}(\tilde{\epsilon}_{\mathbf{Z}1,t}), F_{n\tilde{\epsilon}\mathbf{Z}2}(\tilde{\epsilon}_{\mathbf{Z}2,t}))_{t=1}^{n}, \mathbf{Z} = \mathbf{X}, \mathbf{Y}$. Finally, we apply the quantile function of the t-distribution function with 10 degrees of freedom to the pseudo-observations. Thus, the final errors are linked by the copulas of DGPs 1 to 4 with fat-tailed t-marginals.[8] Those are used to generate the GARCH series for $\mathbf{X}$ and $\mathbf{Y}$. We obtain standardized residuals from estimation by quasi maximum likelihood.

For sample sizes $n = 750, 1500$, varying values of the effective sample size k, and a nominal test level $\alpha = 0.05$, we compare empirical rejection frequencies. Also, for test algorithm 1, we employ two subset discretizations ($M = 6, 18$) to evaluate the sensitivity of the test performance with regard to the user-dependent test calibration. Further, we employ test algorithm 2 which merges 14 different grids. The TDC-test is carried out using the multiplier bootstrap at points $u_1 = u_2 = 0.5$, i.e. for a given value of k estimates correspond to standard TDC estimates at point $u_1 = u_2 = 1$ with $k^* = k/2$. The number of simulations is $S = 500$ for each setting.

Table 1 reports empirical rejection frequencies for i.i.d. marginals and GARCH filtered marginals and sample size $n = 1500$. Also, we study the effect of varying $k = \lfloor 0.1n \rfloor, \lfloor 0.2n \rfloor, \lfloor 0.3n \rfloor$. Note, $\widehat{\Lambda}(u_1, u_2; k = k^*) = \widehat{\Lambda}(au_1, au_2; k = ak^*)$. Hence, these values for k correspond to $\lfloor 0.05n \rfloor, \lfloor 0.1n \rfloor, \lfloor 0.15n \rfloor$ in the standard case of TDC estimation with $u_1 = u_2 = 1$. Table 2 contains empirical rejection frequencies s $n = 750$. As non-parametric methods for tail dependence are often criticized for unsatisfactory small sample performance, it is worth studying test behavior for small and moderate sample sizes.

In general, our test appears to be asymptotically consistent. For i.i.d. marginals our test obeys the nominal test size of $\alpha = 0.05$ (DGP 1 and 3), irrespective of the choice of k. While empirical test size remains untouched by k, the choice of effective sample size notably affects empirical power; for example, for DGP 4, power increases by up to 25% both for $M = 6, 18$. Hence, this suggests a larger choice of k is favorable. As noted in Bücher and Dette (2013), for a large k, bias terms in $\widehat{\Lambda}_{\mathbf{X}}$ and $\widehat{\Lambda}_{\mathbf{Y}}$ cancel out. This suggests the choice of k, which in essence is a bias-variance problem for $\widehat{\Lambda}$, is slightly facilitated compared to other extreme value based peaks-over-threshold problems. Thus, $k \approx 0.1n$ seems a reasonable rule of thumb.

While single-grid tests show larger power than the TDC-test, the BD13-test is more powerful in standard cases whenever our test is based on a single discretization. However, combining a multiple of single-grid tests, e.g. test algorithm 2, makes our test consistently more powerful than BD13.

[8]Note that monotone transformations, such as the quantile transformation, do not alter the dependence structure but the transformed error distributions are closer to empirical residual distributions found in real data.

k/n	DGP1					DGP2					DGP3					DGP4				
	TDC	BD13	BS16			TDC	BD13	BS16			TDC	BD13	BS16			TDC	BD13	BS16		
			18	6	TA2			18	6	TA2			18	6	TA2			18	6	TA2
iid																				
5%	4.0	3.2	3.2	2.4	4.8	3.2	4.2	100	100	100	5.0	4.8	3.2	4.2	6.8	73.8	86.2	78.2	82.2	88.2
10%	2.0	3.8	2.0	2.4	5.4	4.0	4.4	100	100	100	2.2	3.6	3.4	2.8	4.8	91.8	97.6	94.8	95.8	98.2
15%	4.4	3.2	2.8	2.6	6.0	5.2	5.8	100	100	100	3.0	3.0	3.0	2.4	7.0	96.6	99.8	98.4	98.6	100
fil.																				
5%	3.4	4.4	2.8	3.4	5.8	5.4	7.6	100	100	100	3.6	4.0	3.0	2.8	5.8	73.8	86.2	78.2	82.2	87.2
10%	4.0	4.4	2.4	3.8	5.8	5.0	7.6	100	100	100	4.4	3.8	3.4	2.6	6.6	92.6	97.4	95.6	96.0	97.8
15%	5.2	4.0	3.0	3.0	5.4	9.2	8.8	100	100	100	3.0	3.0	3.0	2.4	7.0	97.2	98.8	97.6	98.4	98.8
unfil.																				
5%	6.0	6.6	4.2	4.6	8.0	8.6	12.6	83.2	52.0	86.8	9.6	12.4	8.6	9.8	14.2	17.8	21.2	18.0	19.2	24.6
10%	4.6	5.8	4.0	4.6	7.4	6.6	8.8	100	100	100	7.0	11.4	9.0	9.8	14.8	22.0	31.0	25.4	26.2	34.6
15%	4.8	4.2	3.0	4.2	6.4	5.6	7.8	100	100	100	6.8	7.4	6.4	6.0	10.2	33.2	44.2	35.8	39.8	48.0
blo.																				
5%	6.6	5.0	3.6	3.4	6.6	8.0	8.0	73.4	40.2	81.4	7.4	11.0	8.8	9.4	14.6	37.0	44.6	39.0	42.6	49.0
10%	6.0	4.8	3.4	4.0	5.4	6.6	6.6	100	99.8	100	5.6	8.0	8.4	7.8	13.0	70.4	80.0	70.2	76.0	82.8
15%	6.0	5.0	3.2	3.4	5.8	5.6	6.4	100	100	100	4.0	7.2	8.0	6.6	12.8	88.0	94.0	90.8	92.2	95.0
tap.																				
5%	3.8	4.8	2.6	3.8	6.0	5.4	6.8	100	100	100	3.8	4.2	2.6	2.2	6.0	75.8	85.6	77.8	82.6	87.8
10%	3.8	4.6	2.6	3.2	5.6	5.4	7.4	100	100	100	4.2	3.4	4.0	2.6	6.4	92.8	97.6	95.2	96.8	97.8
15%	5.2	4.0	2.8	3.0	4.8	4.4	4.6	100	100	100	4.4	5.0	3.8	3.4	6.8	97.0	99.0	97.4	98.4	99.0

Table 1: Empirical rejection probabilities for $\alpha = 5\%$, $S = 500$ repetitions and sample size $n = 1500$. Effective sample fraction k/n is evaluated at $(u_1, u_2) = (1, 1)$. DGP 1: H_0 factor model, DGP 2: H_1 factor model, DGP 3: H_0 Clayton copula, DGP 4: H_1 Clayton copula. Rejection frequencies are shown for a varying effective sample size, i.i.d. marginals and GARCH marginals for which the tests are applied to raw observations ("unfiltered") and also to standardized residuals ("filtered"). For the latter, estimation was carried out by quasi maximum likelihood.

k/n	DGP1					DGP2					DGP3					DGP4				
	TDC	BD13	BS16			TDC	BD13	BS16			TDC	BD13	BS16			TDC	BD13	BS16		
			18	6	TA2			18	6	TA2			18	6	TA2			18	6	TA2
iid																				
5%	4.6	4.8	3.6	3.8	5.6	2.6	4.4	97.8	100	99.8	3.4	4.0	3.6	3.2	6.2	43.2	57.8	44.6	52.2	60.8
10%	3.2	2.8	2.6	2.2	5.0	5.4	5.8	100	100	100	3.4	3.6	2.8	3.2	6.0	65.2	79.6	69.6	75.8	82.4
15%	4.0	4.2	3.2	2.8	7.2	4.2	5.6	100	100	100	4.0	5.8	2.8	3.4	8.0	76.4	86.2	81.0	83.6	88.6
tap.																				
5%	5.0	6.2	4.2	4.6	6.8	4.4	3.8	98.0	78.6	99.8	11.6	15.2	13.8	13.8	20.4	26.6	39.4	33.4	37.0	44.8
10%	4.4	5.4	3.0	4.4	7.2	4.8	6.2	100	100	100	8.0	12.6	12.4	12.2	19.4	48.2	61.4	55.8	57.8	66.2
15%	2.4	4.2	3.2	3.6	5.6	6.2	6.4	100	100	100	8.2	9.8	7.6	8.0	13.8	62.0	75.4	69.6	73.2	79.6

Table 2: Empirical rejection probabilities as in Table 1, but with a sample size of $n = 750$.

Importantly, our test successfully detects intra-tail asymmetries, as shown by the empirical rejection frequencies for DGP 2. Both the TDC-test and BD13-test fail to reject the null in this case and treat intra-tail asymmetries as tail symmetric. If the tail copula is intra-asymmetric, it is beneficial to apply our test with more subsets. If the tail copula is symmetric, however, power decreases in the number of subsets. It is thus advisable to apply the test for a variety of M, say $M \in \{2, 4, 6, ..., 26\}$, to compare both TCs over many different subset specifications.

Also, test results for GARCH filtered returns are in line with i.i.d. series. Thus, the estimation step of the GARCH residuals does not downgrade test power nor size. However, unfiltered GARCH returns should not be used as in the case of DGP 4, test power implodes by roughly 50-75% for all three tests, whereas empirical sizes for DGP 1 are still fine, empirical size of DGP 3 generally is too large. As the tapered block multiplier bootstrap produces comparable results as the multiplier bootstrap based on i.i.d. and GARCH filtered marginals, we prefer this adjusted bootstrap for empirical applications work as it can handle non-i.i.d. data and it does not require pre-estimation of a parametric model. However, as Table 2 suggests, the tapered block bootstrap should only be applied for generous sample sizes as for $n = 750$ and GARCH marginals the tapered multiplier block bootstrap appears to be oversized and standardized GARCH residuals as approximation of i.i.d. observations should be used instead.

Finally, we find our aggregating test (test algorithm 2) is throughout most powerful, while the test with fixed grids (test algorithm 1) is consistently more powerful than the TDC-test, slightly less powerful than the BD13-test, and more powerful than the latter in case of intra-tail asymmetry.

5 Empirical application

5.1 Tail asymmetries within S&P500 industry portfolios

We study possible tail asymmetries between daily returns 49 S&P500 industry portfolios. The data set, publicly available at `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`,[9] contains nearly 90 years of weighted returns of CRSP SIC codes based industries. For detailed information on industry composition we refer to the website just mentioned.

With this exhaustive data set we aim to detect tail asymmetry dynamics of the complete S&P500 universe. Applying a rolling window analysis with window length of $n = 1500$, i.e. nearly 6 years, and a step size of 300 trading days, i.e., roughly 14 months, we arrive at 74 (overlapping) time periods. In each period, we build all possible bivariate combinations and test the nulls $H_0 : \Lambda^U_{\mathbf{X}} = \Lambda^L_{\mathbf{X}}$. Discarding pairs with missing data, in each period there are at most 1176 pairs to test against tail asymmetry. To avoid possible model risk by pre-filtering the returns we throughout analyze raw returns using the tapered block multiplier bootstrap; Section 3.2 and the results of the simulation study justify this approach.[10] Also, we fix the effective sample size to $k = 0.1n$ (evaluated at $(u_1, u_2) = (1, 1), \Lambda(u_1, u_2)$) which as well is inspired from

[9] Accessed 1 March 2016.

[10] For simplicity, we fix the window parameter of the tapered block multiplier bootstrap as $l = 8$. Yet, we find no change of results worth mentioning when altering l.

the simulation study. We are not interested in particular industry pairs but our focus is on tail asymmetry of the general market. Hence, a fixed k for all pairs is an operable solution to the choice of number of extremes as over- and underestimation might eventually balance out when counting the number of test rejections over all 1176 pairs.

To grasp the general evolution of lower and upper bivariate tails, we introduce a descriptive measure for market tail dependence. In period t, for each pair i, we integrate the TC over $[0,1]$ and provide empirical location statistics of all pairs, e.g. for the mean,

$$\overline{\Lambda}_t := \frac{1}{\binom{n_t}{2}} \sum_{i=1}^{\binom{n_t}{2}} \int_0^1 \widehat{\Lambda}_i(u, 1-u)\ du,$$

where n_t is the number of sectors in period t. We also consider empirical quantiles for integrated TCs. It is easy to see that $\Lambda_i \in [0, 0.25]$: For pair i there is no (perfect) bivariate tail dependence if $\overline{\Lambda}_t = 0$ (0.25). Figure 3 shows the trajectory of the mean and q-quantiles, $q = 0.01, ..., 0.99$, for both upper and lower tails covering 1931-2015.

The null is tested by the TDC-test, the BD13-test and test 2, aggregation over 14 grids in the spirit of the simulation study. Figure 4 displays trajectories of the share of tail asymmetric pairs according to each test. Figures 3 and 5 are supplementary.

All tests indicate, most of the time, substantial amount of tail asymmetries in the market. On average our test finds 64% (sd=0.25) of all pairs exhibit TA. We can identify an enduring phase of pronounced share of tail asymmetries between 1940-1970 where on average 80% (sd=0.10) of all pairs are tail asymmetric. Collapses of the number of tail asymmetries fall strikingly into times of financial crises, such as the beginning of the Great Depression (1932-1937), the Oil Crisis (1968-74 until 1972-1978), Black Monday (1987) and the Asian and millennium crisis accumulating into the Dot-Com crisis (1995-2003). In contrast, the recent financial crisis 2007-2009 is characterized by a temporary bump which opposes a phase of steady decline of tail asymmetries since the the mid 1990s. One might argue that in contrast to former financial crises only dependence between extreme losses was affected while tail dependence between gains did not experience such change. However, Figure 3 refutes this hypothesis as apparently tail dependence across losses and gains co-moved nearly in a parallel way as in previous crises. While the summary measures for market tail dependence suggest left and right tails are connected equally strongly during the 2000s, all three tests report otherwise and reveal a pattern not captured by descriptive statistics.

In comparison to the two competing tests our tests consistently detects more asymmetries, see Figure 5 (left), which we attribute to the fact that competing tests overlook non-central tail dependence structures (TDC-test) or intra-tail asymmetry (TDC-test, BD13-test). Hence, our

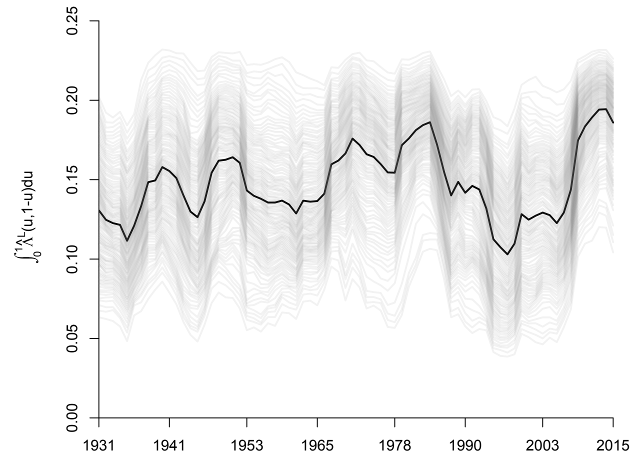

Figure 3: $\int_0^1 \widehat{\Lambda}(u, 1-u)du$ for all possible pairs (up to 1176) in each period; dark line: empirical mean; gray lines: empirical quantiles: $0.01i, i = 1, ..., 99$. Left: losses. Right: gains.

test provides a more accurate assessment of extreme risk within the market; our test suggests the phenomenon of tail asymmetry is more common. With respect to the TDC-test (BD13-test), we find 2.5%-27% (0%-12%) more tail asymmetric pairs. We also plot the trajectory of the percentage of rejections where the adjusted P-value of the central subinterval does not suggest a rejection, while at least one non-central P-value does (solid line). This line runs nearly parallel to the graph of the differential in found tail asymmetries between the TDC-test and our test. Furthermore, the difference in found asymmetries between our test and BD13 suggests some degree of intra-tail asymmetry among all pairs as the simulation study demonstrated both tests' power mainly differ only in intra-tail asymmetric case. With our test we can also quantify the number of tail asymmetric pairs that such approaches would miss due to "off-diagonal" tail asymmetries. Figure 5 (right) compares the number of rejections of "non-central" subintervals with the number of rejections found in the central subinterval. We find that our test, when restricted to non-diagonal subintervals, finds up to 20% more asymmetries than a TDC-based analysis that solely focusses on the central subinterval. Throughout the sample, there exists at least one non-central subinterval with more test rejections than the central subinterval. Furthermore, there are periods of time, which match the major financial crises, when not considering off-diagonal parts of the TC is especially serious. Yet, in the finance literature, it is common practice to analyze tail dependence solely by the tail dependence coefficient $\Lambda(u, u)$, i.e. the tail copula along the diagonal $u_1 = u_2 = u \in \mathbb{R}_+$. We show that this approach might overlook non-standard types of tail dependence leading to a substantial misconception of tail asymmetry.

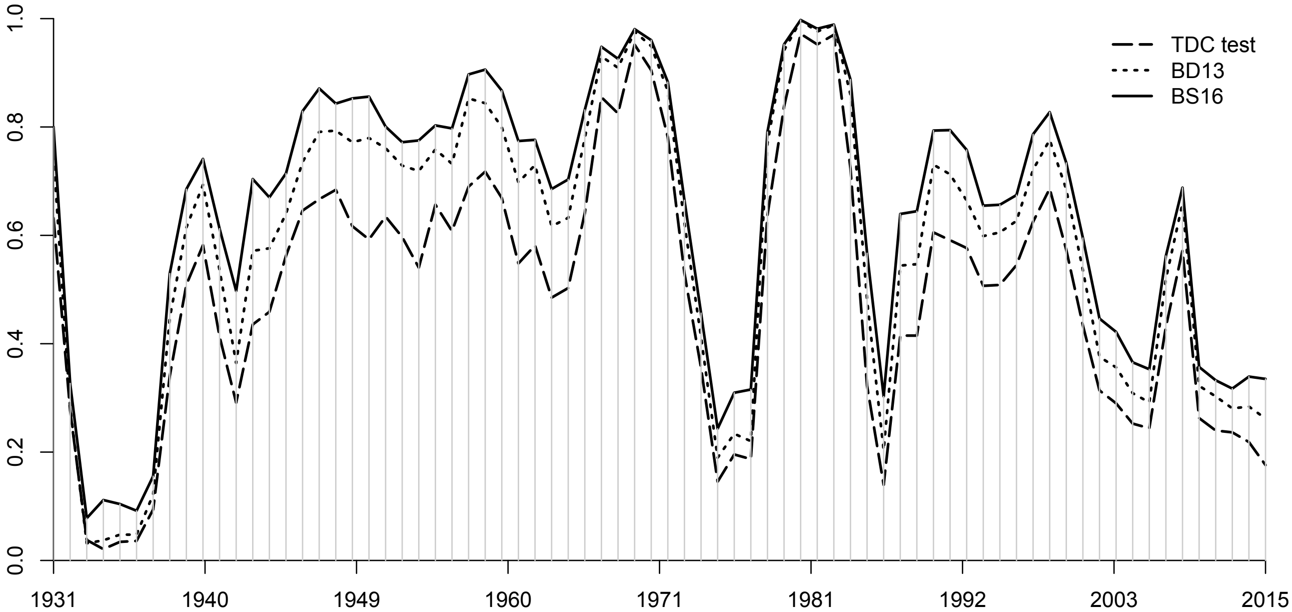

Figure 4: Dynamics of the percentage of detected tail asymmetries among all pairs using a rolling window of size $n = 1500$, and a step size of 300 trading days for the TDC-test (dashed), BD13-test (dotted) and our test (solid), respectively.

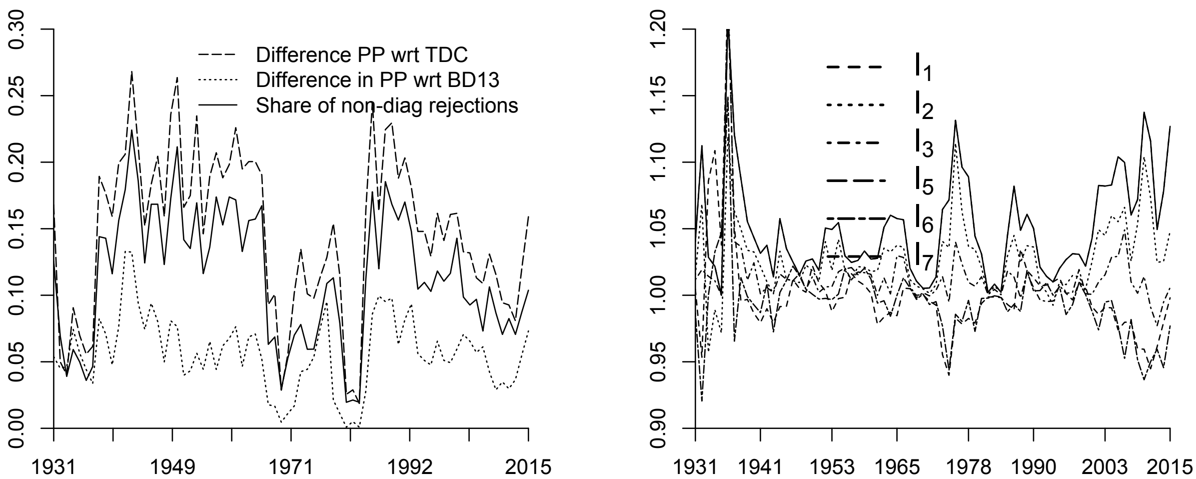

Figure 5: (Left) Difference of detected asymmetries in percentage points with respect to the TDC-test (dashed) and BD13-test (dotted), and percentage of our test's rejections that are induced by subintervals off the diagonal, based on a grid with $M = 14$. (Right) Number of rejections in subsets $\mathcal{I}_i, i = 1, 2, 3, 5, 6, 7$, i.e. off-diagonally, compared to number of rejections in subsets $\mathcal{I}_4$, i.e. around (0.5,0.5), based on a grid with $M = 14$.

5.2 Tail inequalities of foreign exchange rates

We now analyze tail equality in pairs of six main foreign exchange rates, namely Euro (EUR), British Pound (GBP), Canadian Dollar (CAD), Japanese Yen (JPY), New Zealand Dollar

(NZD) and Swiss Franc (CHF), all nominated in USD.[11] The sample consists of returns of daily closing prices ranging from 01/05/2001 to 02/01/2016. As foreign exchange rates are the most frequently traded financial security with an average daily trading volume of more than 5 trillion in April 2013,[12], investors and regulators have a natural interest in a comparison of extreme foreign exchange rates co-movements. We again apply a rolling window analysis, now with a window size of $n = 1000$ and step size of 50 days to draw a finer picture of the tail (in)equality dynamics. For any pair comparison trading days with missing data or zero returns are discarded. The effective sample size is set to $k = 0.2n$ at $\Lambda(1,1)$; we analyze unfiltered data and use the tapered block multiplier bootstrap. We conduct the following pair tail comparisons

$$H_0^{(L-L)} : \Lambda_{\mathbf{X}}^L = \Lambda_{\mathbf{Y}}^L, \quad H_0^{(L-U)} : \Lambda_{\mathbf{X}}^L = \Lambda_{\mathbf{Y}}^U, \quad H_0^{(U-U)} : \Lambda_{\mathbf{X}}^U = \Lambda_{\mathbf{Y}}^U,$$

for all 15 bivariate pairs, amounting to 420 tests in each period. Figure 8 shows the share of tail inequalities among all possible comparisons.

The fraction of rejected tail equalities, ranging from 45% to 75%, suggests bivariate tails of foreign exchange rates systematically differ. We observe a steady increase of tail inequalities from 2006 to 2008 which coincides with the depreciation of the USD with respect to the Euro. This evolution is reversed when the USD appreciates during European Sovereign Crisis (2013 onwards). Thus, in the last decade, a strong (weak) USD (Euro) came along with more (less) tail equality within the foreign exchange rates market.

Figure 7 displays a time-varying tail dependence ranking for all 15 pairs based on the TDC and the summary statistic $\int_0^1 \Lambda(u, 1-u)\mathrm{d}u$ which was introduced in the last subsection. A careful inspection of all four plots shows there is only little difference between the TDC based and the TC based ranking. Tail dependence of appreciations and depreciations of EUR and CHF with respect to the USD tends to be the strongest throughout the sample. While the pair GBP-EUR exhibits strong tail dependence for joint upper tails (depreciations), the lower tail shows a strong tail link only in the last five years (as well as until 2007). Also, JPY-CAD (upper tail) and CAD-NZD (both tails) feature comparably strongly connected tails. The pairs JPY-NZD, JPY-CAD and GBP-JPY feature the weakest tail dependence in both tails.

The pair EUR-CHF dominates tail comparisons throughout which is due to the fixed exchange rate until 01/2015 with a Euro minimum rate of 1.20 CHF. Also, the tight economic linkage between both parties may attribute to the relatively strong tail dependence. On 15 Jan 2015 the Swiss Central Bank unpegged its currency from the Euro intending to avoid a

[11] Time series data are standard exchange rates from Bloomberg.
[12] See Rime and Schrimpf (2013)

$k/n:$	0.02	0.04	0.06	0.08	0.1	0.12	0.14	0.16	0.18	0.2	0.22	0.24	0.26	0.28	0.3
L-L	4.4	10.6	37.2	30.8	17.9	19.9	22.4	35.6	52.0	61.5	46.2	33.6	46.0	56.1	63.7
U-U	99.2	7.4	98.0	87.2	99.9	87.8	44.7	80.2	98.6	99.6	95.7	96.1	99.8	99.6	96.6

Table 3: P-values corresponding to the null hypothesis of constant tail dependence between EUR and CHF (see Equation 4) for varying effective sample sizes.

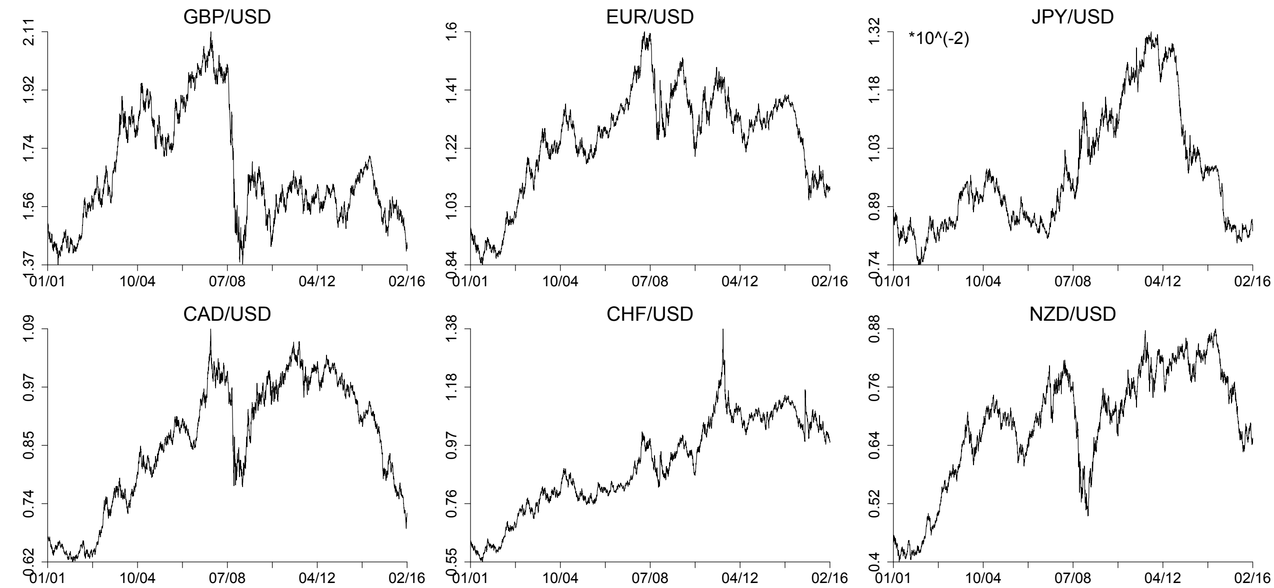

Figure 6: Foreign exchange rates nominated in US Dollars. The time span is 01/2001 to 02/2016.

continued depreciation of the Swiss currency as the Euro devaluated since 2008/2009. This policy change caused the CHF to appreciate by 20% with regards to the Euro within a single day. We now test whether the break of the CHF-EUR currency peg had a significant impact on the tail dependence between both currencies. This would be the case if the TC had changed after 15/01/2015. Unfortunately, the sample contains only 273 observations after the policy change and we thus compare TCs for overlapping time periods, that is 01/01/2006-14/01/2015 (Λ^{T1}) and 01/01/2006-16/01/2016 ($\Lambda^{T1,T2}$). The null is

$$H_0 : \Lambda^{W,T1}_{CHF-EUR} = \Lambda^{W,T1+T2}_{CHF-EUR}, W = U, L \tag{4}$$

However, the tapered block multiplier bootstrap has to be adjusted to account for the dependence of both samples. For the TC of the entire period ($T_1 + T_2$) we use the multiplier vector $\xi^{T1+T2} = (\xi_1, ..., \xi_{T1}, \xi_{T1+1}, ..., \xi_{T1+T2})$; for the TC of the first subperiod, we only use the first T_1 entries of ξ^{T1+T2}. We execute the test for 15 different values of the effective sample size, namely $k_T = 0.02n_T, 0.04n_T, ..., 0.3n_T$, where n_T denotes the sample size of the first subperiod or the entire period, respectively. Table 3 contains p-values of test 2. To this date, there is no evidence for a structural change both in the left and right tail.

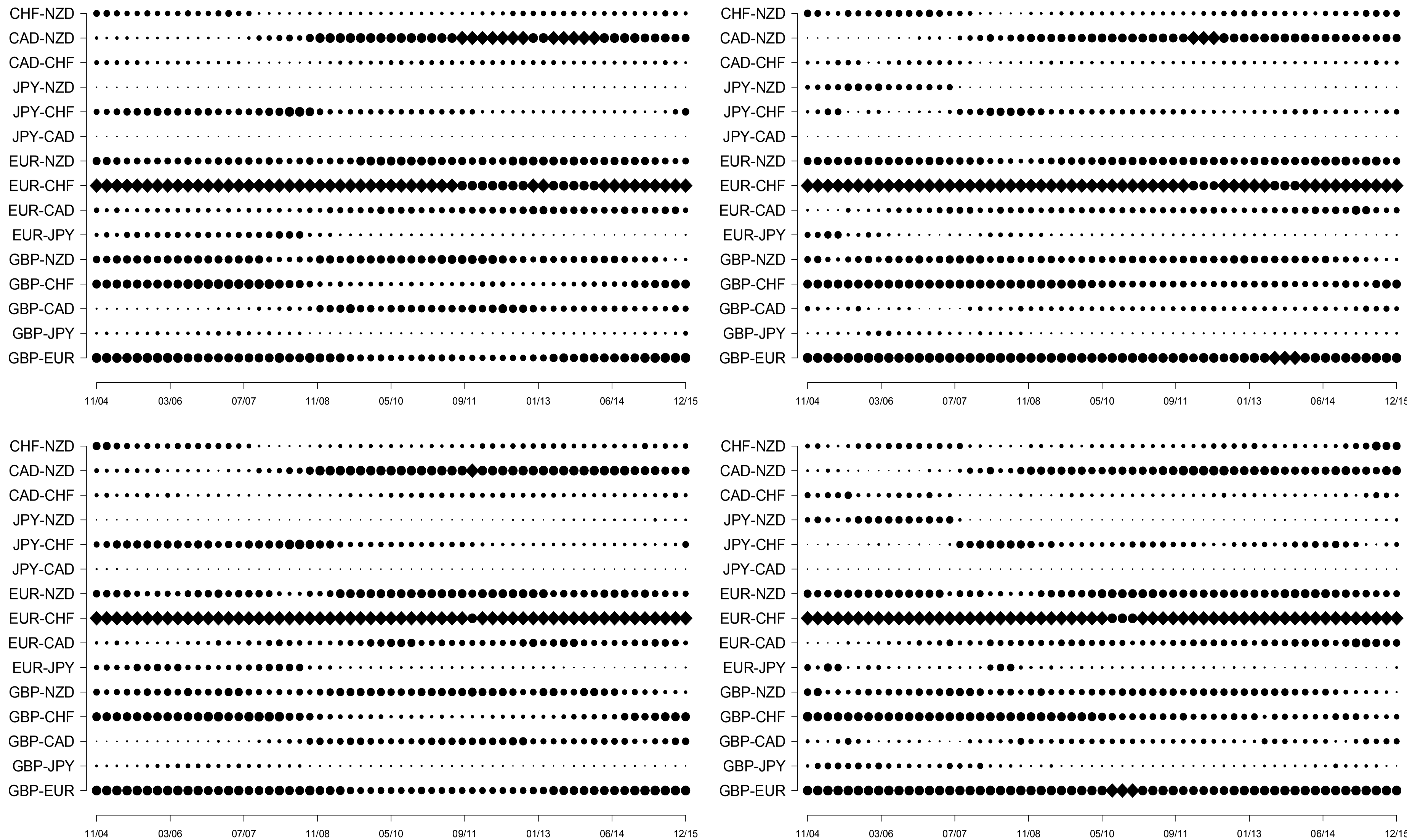

Figure 7: Tail dependence ranking of all 15 pairs for the lower (left column) and upper (right column) tails according to the summary statistic $\int_0^1 \Lambda(u, 1-u)\mathrm{d}u$ (top row) and the TDC (bottom row), respectively. The size of each dot reflects the rank. The diamond shaped dot marks the strongest pair.

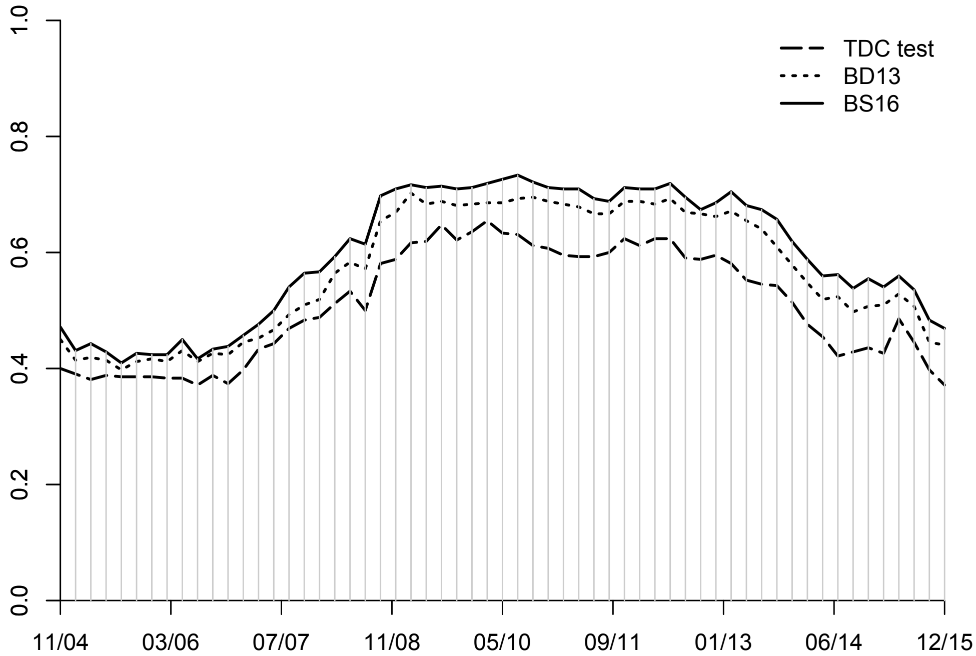

Figure 8: Dynamics of the percentage of detected tail inequalities among all pairs, comparing the following tails: Upper-upper, upper-lower, lower-lower. The window size is $n = 1000$ with a step size of 50 trading days, and rejections based on the TDC-test (BD13-test, our test) correspond to the dashed (dotted, solid) line.

6 Conclusion

We propose a novel test against asymmetries/inequalities between tail dependence functions. The test is based on the empirical tail copula and conducts piecewise comparisons between tail copulas. Importantly, our test considers intra-tail asymmetries and achieves higher power in intra-tail asymmetric cases, and comparable power else. The test idea might also be applied for general copula comparisons, and also for tail copula comparisons in higher dimensions. An empirical study of S&P500 and foreign exchange rates shows our test typically finds more asymmetries/inequalities than competing tests; we find time periods when our test clearly benefits from respecting non-diagonal TC differences and meaning our test detects substantially more opportunities to hedge tail risks.

References

Aloui, R., M. S. B. Aïssa, and D. K. Nguyen. 2011. Global financial crisis, extreme interdependences, and contagion effects: The role of economic structure? *Journal of Banking & Finance* 35(1): 130-141.

Ang, A., and J. Chen. 2002. Asymmetric correlations of equity portfolios. *Journal of Financial Economics* 63: 443–494.

Y. Benjamini, and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B* 57(1): 289–300.

Bücher, A., and H. Dette. 2013. Multiplier bootstrap of tail copulas with applications. *Bernoulli*. 5: 1655–1687.

Bücher, A., and M. Ruppert. 2013. Consistent testing for a constant copula under strong mixing based on the tapered block multiplier technique. *Journal of Multivariate Analysis*. 116: 208–229.

Bücher, A., J. Segers, and S. Volgushev. 2014. When uniform weak convergence fails: Empirical processes for dependence functions and residuals via epi- and hypographs. *The Annals of Statistics* 42(4): 1598–1634.

Dickhaus, T. 2014. *Simultaneous statistical inference with applications in the life sciences*. Heidelberg, Springer.

de Haan, L., and A. Ferreira. 2006. *Extreme value theory: An introduction*. New York, Springer.

Einmahl, J. H. J., Kraihna, A. , and Segers, J. 2012. An M–estimator for tail dependence in arbitrary dimensions. *The Annals of Statistics* 40(3): 1754–1793.

Embrechts, P. 2009. Linear correlation and EVT: Properties and caveats. *Journal of Financial Econometrics* 7: 30-39.

Engle, R. F., and K. Sheppard. 2001. Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. National Bureau Economic Research, working paper, No 8554.

Fama, E. F., and K. R. French. 1992. The cross–section of expected stock returns. *Journal of Finance* 47: 427–486.

Frahm, G., M. Junker, and R. Schmidt. 2005. Estimating the tail dependence coefficient: Properties and pitfalls. *Insurance: Mathematics & Economics* 37(1): 80–100.

Garcia, R., and G. Tsafack. 2011. Dependence structure and extreme comovements in international equity and bond markets. *Journal of Banking & Finance* 35: 1954-1970.

Hartmann, P., S. Straetmans, and C. G. de Vries. 2004. Asset market linkages in crisis periods. *The Review of Economics and Statistics* 1: 313–326.

Jondeau, E.. 2015. Asymmetry in tail dependence in equity portfolios. *Computational Statistics and Data Analysis* (forthcoming).

Kojadinovic, I. 2015. npcp: Some nonparametric tests for change-point detection in possibly multivariate observations. `R package version 0.1-6`, http://CRAN.R-project.org/package=npcp.

Ledford, A. W., and J. A. Tawn. 1996. Statistics for near independence in multivariate extreme values. *Biometrika* 83: 169–187.

Li, F. 2013. Identifying asymmetric comovements of international stock market returns. *Journal of Financial Econometrics* 12(3): 507–543.

Longin, F. and B. Solnik. 2001. Extreme correlation of international equity markets. *The Journal of Finance*. 56(2): 649-676.

Frey, R. and A. J. McNeil. 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance*. 7: 271-300.

McNeil, A., and R. Frey, P. Embrechts. 2005. *Quantitative Risk Management: Concepts, Techniques*. New Jersey, Princeton University Press.

Mikosch, T. 2006. Copulas: Tales and facts. *Extremes* 9: 3–20.

Oh, D. H., and A. Patton. 2015. Modelling dependence in high dimensions with factor copulas. *Journal of Business and Economic Statistics* (forthcoming).

Patton, A. 2006. Modelling asymmetric exchange rate dependence. *International Economic Review* 47: 527–556.

Patton, A. 2013. Copula Methods for Forecasting Multivariate Time Series. In *Handbook of Economic Forecasting* 2(B).

Rime, D. and A. Schrimpf. 2013. The anatomy of the global FX market through the lens of the 2013 Triennial Survey. *BIS Quarterly Review* December: 27-44.

Rodriguez, J. C. 2007. Measuring financial contagion: A copula approach. *Journal of Empirical Finance* 14: 401–423.

Romando, J. P., and M. Wolf. 2005. Stepwise multiple testing as formalized data snooping. *Econometrics* 73(4): 1237-1282.

Schmidt, R., and U. Stadtmüller. 2006. Nonparametric estimation of tail dependence. *The Scandinavian Journal of Statistics* 2: 307–335.

Straetmans, S. T. M., W. F. C. Verschoor, and C. C. P. Wolff. 2008. Extreme US stock market fluctuations in the wake of 9/11. *Journal of Applied Econometrics* 23: 17–42.

Tawn, A. J. 1988. Bivariate extreme value theory: Models and estimation. *Biometrika* 75: 397–415.

7 Appendix

7.1 Proofs

Proof [Proposition 3.1] Equation 2 guarantees convergence of the empirical TC $\sqrt{k_{\mathbf{Z}}}\widehat{\Lambda}_{\mathbf{Z}}(u_1,u_2), \mathbf{Z}=\mathbf{X},\mathbf{Y}$, for $(u_1,u_2),(v_1,v_2)\in\mathbb{R}^2_+$. Define

$$\widehat{\Delta}(u_1,u_2,v_1,v_2) =: \sqrt{k_{\mathbf{Y}}/(k_{\mathbf{X}}+k_{\mathbf{Y}})}\widehat{\Lambda}_{\mathbf{X}}(u_1,u_2) - \sqrt{k_{\mathbf{X}}/(k_{\mathbf{X}}+k_{\mathbf{Y}})}\widehat{\Lambda}_{\mathbf{Y}}(v_1,v_2),$$

which is a sum of rescaled tail copula processes with $\mathbb{G}_{\widehat{\Lambda},\mathbf{Z}}, \mathbf{Z}=\mathbf{X},\mathbf{Y}$ is bivariate Gaussian process. It directly follows from Equation 2 that

$$\widehat{\Delta}(u_1,u_2,v_1,v_2) \xrightarrow{w} \Delta(u_1,u_2,v_1,v_2) := \sqrt{k_{\mathbf{Y}}/(k_{\mathbf{X}}+k_{\mathbf{Y}})}\mathbb{G}_{\widehat{\Lambda},\mathbf{X}}(u_1,u_2) - \sqrt{k_{\mathbf{X}}/(k_{\mathbf{X}}+k_{\mathbf{Y}})}\mathbb{G}_{\widehat{\Lambda},\mathbf{Y}}(v_1,v_2).$$

Only under the null and $|\mathbf{u}|=|\mathbf{v}|$, $\mathbb{E}[\Delta(u_1,u_2,v_1,v_2)]=0$. By the continuous mapping theorem $\widehat{\Delta}^2(u_1,u_2)\xrightarrow{w}\Delta^2(u_1,u_2)$. For a fixed grid $\mathcal{I}_{(i)}$, and some subinterval $[a,b]\subset\mathcal{I}_{(i)}, 0<a<b<\infty$, consider the test statistic corresponding to the ith null $H_{0,i}$ that integrates over $[a,b]$, i.e. $\widehat{S}^{i,[a,b]}$. Then it directly follows $\widehat{\Delta}_i^2(u_1,1-u_1)\xrightarrow{w}\Delta_i^2(u_1,1-u_1), u_1\in[a,b]$. Under the null of $H_0:\Lambda_{\mathbf{X}}=\Lambda_{\mathbf{Y}}$, for all i, $\widehat{\Delta}_i^2\xrightarrow{w}0$ as $\Delta_i^2=0$. Under the alternative, there naturally is at least one subinterval where the test statistic does not converge to zero.