Detecting Structural Differences in Tail Dependence of Financial Time Series

Carsten Bormann
Karlsruhe Institute of Technology
and
Melanie Schienle
Karlsruhe Institute of Technology

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. We decompose the testing problem into piecewise multiple comparisons of Cramér-von Mises distances of tail copulas. In this way, tail regions that cause differences in extreme dependence can be located and consequently be targeted by financial strategies. We derive the asymptotic properties of the test and provide a 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. 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. During the last 90 years in US stock markets, our test detects up to 20% more tail asymmetries than competing tests. This can be attributed to the presence of non-standard tail dependence structures. We also find evidence for diminishing tail asymmetries during every major financial crisis – except for the 2007-09 crisis – reflecting a risk-return trade-off for extreme returns.

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

JEL classification: C12, C53, C58

1 INTRODUCTION

Asymmetric dependence both within and between bivariate extreme returns in different market conditions is not only a key criterion for asset and risk management, but also a main focus of market supervision. During financial crises, financial markets exhibit pronounced cross-sectional co-movements of (lower) tails of return distributions. Thus, the tendency of joint extreme events intensifies, see e.g. Longin and Solnik (2001); Ang and Chen (2002); **?**. 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. Conversely, risk managers and market supervisors might need to set larger capital buffer requirements if the tendency for joint occurrences of extreme losses rises in times of market distress. Specifically aiming at dependence between extreme events, we provide a robust non-parametric statistical test against tail dependence differences. The test accurately detects all types and the full extent of deviations between two tail dependence functions. Our test procedure is based on multivariate extreme value techniques which remain valid during turbulent market periods, e.g. Mikosch (2006). Particular to finance, Ang and Chen (2002), Patton (2006), Chollete et al. (2011) document the economic merits for asset diversification of asymmetric dependence structures, e.g. for optimal portfolio allocation. Under adverse market conditions, standard linear dependence measures are flawed 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)). Furthermore, our test connects concepts of multivariate extreme value theory with multiple testing techniques. Recently, the latter have gained in practical importance in times of abundant data and the risk of data snooping, see e.g. in the finance context Barras et al. (2010) and Bajgrowicz and Scaillet (2012) .

We propose a novel non-parametric test procedure against pairwise differences in tail dependence structures which we measure with tail copulas denoted by $\Lambda(x^{(1)}, x^{(2)})$, $(x^{(1)}, x^{(2)}) \in \mathbb{R}_+^2$. A tail copula is a functional of the complete tail dependence. The flexibility of using empirical tail copulas avoids possible parametric misspecification risk; see e.g. Longin and Solnik (2001); Patton (2013); Jondeau (2016) for parametric approaches. Furthermore, the generality of this approach is in sharp contrast to established approaches, which only estimate and compare scalar 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 over their entire relevant domain in a locally piecewise way. Thus, we study a multiple testing problem of tail copula equality. Piecewise testing allows to pin down specific quantile regions where tail dependence differences are most serious. Such areas then indicate those types of extreme market conditions that typically cause tail asymmetry (inequality). Moreover, our test is still consistent if one (or both) of the two considered tail copulas is non-exchangeable, i.e. $\Lambda(x^{(1)}, x^{(2)}) \neq \Lambda(x^{(2)}, x^{(1)})$. Existing procedures fail to address such intra-tail asymmetric dependence structures. Therefore, for non-exchangeable tail copulas, those tests are inconsistent.

Our test builds on the idea of a two-sample goodness-of-fit test for tail copulas as in Bücher and Dette (2013). However, for increased sensitivity against violations of the null, we compare both tail copulas in a piecewise way on disjoint subintervals of the unit simplex hull. This way, a number of individual tests against tail dependence equality is carried out. For an accurate overall assessment, we use multiple testing principles, such as the familywise error control and the false discovery rate, to jointly control the error rate of all marginal tests. Asymptotic properties of the test are provided. Moreover, a multiplier bootstrap procedure is suggested by extending ideas of Bücher and Dette (2013) to non-i.i.d. data.

A simulation study with widely used factor and Clayton copulas reveals the test's attractive finite sample properties 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 asymmetric copulas. Simulation results strongly suggest that accounting for time series dynamics is essential. This can be achieved by either GARCH pre-filtering or by directly adjusting the bootstrap approximation for serial dependence.

In an empirical application, we establish tail asymmetry dynamics of 49 US stock sectors for the last 90 years, i.e. dynamics of the differences between upper and lower tails of all bivariate industry pairs. We find empirical evidence that tail asymmetries substantially diminish in times of financial distress. The only strong exception is the 2007-2009 financial crisis which apparently was completely different in structure than any other crisis. We conclude dependence between extreme gains increases in crisis. As the danger of joint extreme losses surges during bear markets, this finding documents a type of *extreme* risk-return trade-off as joint extreme gains are more likely compensating for the increased risk of joint extreme losses. This contrasts with other studies that analyze and compare market index pairs. Overall, our test detects up to 20% more tail asymmetries than competing tests. This 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 (2016)), or the tail copula-based test by Bücher and Dette (2013). Thus, our test could serve as a more accurate tool for investors when assessing tail asymmetry in the market, e.g. our test reveals more opportunities for improved tail asymmetry-based portfolio allocation strategies. In the Appendix D of the online supplement material, we also study tail inequalities between foreign exchange rates.

This 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 studies tail asymmetries of US stock sectors. Additionaly, we analyze tail inequalities between major foreign exchange rates in Appendix D of the online material. Finally, Section 6 concludes. All proofs are contained in the Appendix of the paper.

2 TAIL DEPENDENCE AND TAIL COPULAS

To understand the test idea and test statistics, we shortly introduce necessary tools from extreme value statistics. A complete treatment thereof can be found e.g. in de Haan and Ferreira (2006). For a two-dimensional (random) return vector $\mathbb{X} = (X^{(1)}, X^{(2)})$ its marginal components $X^{(i)}$ are assumed to have a continuous distribution function $F_i(x^{(i)})$ and thus a well-defined marginal quantile function F_i^{-1} for $i = 1, 2$.

Our test is based on the full dependence structure in the tails captured by a tail copula. Note, 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, see e.g. Embrechts (2009). For example, the Gaussian copula, which is completely parametrized by the correlation coefficient, is unable to model any tail dependence. That is to say, dependence may vary over different parts of the distribution, and correlation may be unable to measure dependence in the tails.

Measuring the complete tail dependence between $X^{(1)}$ and $X^{(2)}$, the upper and lower tail copula $\Lambda_{\mathbb{X}}^U(x^{(1)}, x^{(2)}), \Lambda_{\mathbb{X}}^L(x^{(1)}, x^{(2)}), \mathbf{x} := (x^{(1)}, x^{(2)}), \mathbf{x} \in \mathbb{R}_+^2$, are defined by

$$\begin{aligned}
\Lambda_{\mathbb{X}}^U(x^{(1)}, x^{(2)}) &:= \lim_{t\to 0} t^{-1}\mathbb{P}(X^{(1)} > F_1^{-1}(1-tx^{(1)}), X^{(2)} > F_2^{-1}(1-tx^{(2)})), \\
\Lambda_{\mathbb{X}}^L(x^{(1)}, x^{(2)}) &:= \lim_{t\to 0} t^{-1}\mathbb{P}(X^{(1)} < F_1^{-1}(tx^{(1)}), X^{(2)} < F_2^{-1}(tx^{(2)})), t \in \mathbb{R}_+ \qquad (1)
\end{aligned}$$

i.e. the tail copula measures how likely both components jointly exceed extreme quantiles, see e.g. de Haan and Ferreira (2006), Schmidt and Stadtmüller (2006), for details. If $\Lambda^U_{\mathbb{X}}(\mathbf{x}) > 0$ ($\Lambda^L_{\mathbb{X}}(\mathbf{x}) > 0$), gains (losses) of $\mathbb{X}$ are said to be tail dependent. For the sake of notational brevity, we omit the superscripts U and L when clear from the context. With $\mathbf{x} = (1, 1)$, the tail copula boils down to the tail dependence coefficient (TDC), $\iota := \Lambda(1, 1)$. The TDC is a standard tool in financial applications to measure tail dependence, e.g. Aloui, Aïssa and Nguyen (2011) or Garcia and Tsafack (2011). However, it 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 $(x^{(1)}, x^{(2)}) \in \mathbb{R}^2_+$, and describes tail association for every possible tail event. It can be shown that $\Lambda_{\mathbb{X}}(x^{(1)}, x^{(2)}) \in [0, \min(x^{(1)}, x^{(2)})]$, and $\Lambda_{\mathbb{X}}(a\mathbf{x}) = a\Lambda_{\mathbb{X}}(\mathbf{x}), a \in \mathbb{R}$. Due to this homogeneity of the tail copula, it is sufficient to analyze $\Lambda_{\mathbb{X}}(\mathbf{x})$ only on the domain $\mathcal{S} \subset \mathbb{R}^2$, where we set wlog $\mathcal{S} := \{(x^{(1)}, x^{(2)}) : x^{(1)}, x^{(2)} \geq 0, ||\mathbf{x}||_1 = 1\}$, as the unit simplex hull. This restriction to the relevant domain of the tail copula reduces computational efforts in practical implementation and is key for our test.

In the following, we require the tail copulas of interest to exist and work in the following setup.

Assumptions 1. *For a bivariate random vector* $\mathbb{X}$*, we assume that*

(A1) $\mathbb{X}_1, \ldots, \mathbb{X}_n$ *are i.i.d. observations of* $\mathbb{X} \sim F_{\mathbb{X}}$*.*

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

Assumption (A1) is standard in extreme value theory, yet restrictive for financial time series. We use it to illustrate our test idea and to formally derive its statistical properties. In Section 4.1, we then show how (A1) can be relaxed to stationarity and strongly mixing making the test applicable to financial data. Assumption (A2) requires that sample tails can be modeled by bivariate extreme value distributions and are asymptotically dependent for nonparametric estimators to be unbiased, see Schmidt and Stadtmüller (2006) for details why this excludes $\Lambda_{\mathbb{X}}(\mathbf{x}) = 0$. Standard distributions with actual tail dependence, such as e.g. the bivariate t-distribution with dispersion parameter $\rho \neq 0$, meet this assumption. The Gaussian copula, however, violates (A2) due to tail independence ($\Lambda = 0$ for $|\rho| < 1$).

The main focus of this paper is on comparing two tail copulas, in particular in determining if differences of tail copulas exist and where there are located. We formally distinguish between two important cases: tail asymmetry and tail inequality. For their definition we require some notation first. We say two tail copulas $\Lambda_{\mathbb{X}}$ and $\Lambda_{\mathbb{Y}}$ differ if there exists a set I on the unit simplex $\mathcal{S}$ with $I \subseteq \mathcal{S} \subset \mathbb{R}_+^2$ $\mathbb{P}(I) > 0$ such that for all $(x^{(1)}, x^{(2)}) \in I$

$$\left\{\Lambda_{\mathbb{X}}(x^{(1)}, x^{(2)}) \neq \Lambda_{\mathbb{Y}}(x^{(1)}, x^{(2)})\right\} \text{ or } \left\{\{\Lambda_{\mathbb{X}}(x^{(1)}, x^{(2)}) \neq \Lambda_{\mathbb{Y}}(x^{(2)}, x^{(1)})\}\right\}. \tag{2}$$

We write shorthand $\Lambda_{\mathbb{X}} \neq \Lambda_{\mathbb{Y}}$ for Equation (2). Tail asymmetry occurs if upper and lower tail copula of the same return vector $\mathbb{X}$ differ.

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

To detect tail asymmetry, one should compare $\Lambda_{\mathbb{X}}^U(x^{(1)}, x^{(2)})$ not only with $\Lambda_{\mathbb{X}}^L(x^{(1)}, x^{(2)})$ and but also with the flipped components version $\Lambda_{\mathbb{X}}^L(x^{(2)}, x^{(1)})$. In practice, the return vector $\mathbb{X}$ exhibits tail asymmetry whenever the likelihood for co-movements of extreme losses differs from that of extreme gains. For example, in terms of Value at Risk (VaR) exceedances, $\Lambda_{\mathbb{X}}^L > \Lambda_{\mathbb{X}}^U$ implies joint exceedances of loss VaRs are more likely to occur than those of gain VaRs.
If for two different return vectors $\Lambda_{\mathbb{X}} \neq \Lambda_{\mathbb{Y}}$, we call this tail inequality.

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

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

One reason for differences in tail copulas may be intra-tail asymmetry of at least one of the tail copulas considered, where intra-tail asymmetry refers to asymmetry within a single tail copula in the following sense. A return vector $\mathbb{X}$ is intra-tail asymmetric if $\Lambda_{\mathbb{X}}^{W}(x^{(1)}, x^{(2)}) \neq \Lambda_{\mathbb{X}}^{W}(x^{(2)}, x^{(1)}), (x^{(1)}, x^{(2)}) \in \mathcal{S}, W = U, L$. Intra-tail asymmetry refers a single vector $\mathbb{X}$ and a single tail and occurs whenever the corresponding tail copula is not symmetric with respect to its arguments $\mathbf{x} = (x^{(1)}, x^{(2)})$, i.e. if the tail copula is not exchangeable with respect to $X^{(1)}$ and $X^{(2)}$. For example, let $x^{(1)} = 0.2, x^{(2)} = 0.8$ and $t = 0.05$. Then, intra-tail asymmetry is present if the tail event $\{X^{(1)} > VaR_1(0.99)\} \cap \{X^{(2)} > VaR_2(0.96)\}$ is differently likely than the tail event $\{X^{(1)} > VaR_1(0.96)\} \cap \{X^{(2)} > VaR_2(0.99)\}$. The following proposition illustrates the importance of intra-tail asymmetry for comparisons of tail dependence functions.

Proposition 1. *If $\Lambda_{\mathbb{X}}^{W}(x^{(1)}, x^{(2)})$ with $W \in \{U, L\}$ is intra-tail asymmetric, then $\Lambda_{\mathbb{X}}^{W} \neq \Lambda_{\mathbb{Z}}^{H}$ for $(\mathbb{Z}, H) \in \{(\mathbb{X}, \overline{W}), (\mathbb{Y}, U), (\mathbb{Y}, L)\}$, where $\overline{W}$ denotes the complement of W, and $\mathbb{X}, \mathbb{Y}$ are bivariate random vectors with according tail copulas.*

To see this, assume $\Lambda_{\mathbb{X}}^{W}(x^{(1)}, x^{(2)}) = \Lambda_{\mathbb{Z}}^{H}(x^{(1)}, x^{(2)})$. As $\Lambda_{\mathbb{X}}^{W}(x^{(1)}, x^{(2)}) \neq \Lambda_{\mathbb{X}}^{W}(x^{(2)}, x^{(1)})$, it holds $\Lambda_{\mathbb{X}}^{W}(x^{(2)}, x^{(1)}) \neq \Lambda_{\mathbb{Z}}^{H}(x^{(1)}, x^{(2)})$, and Equation (2) applies. If $\Lambda_{\mathbb{X}}^{W}(\mathbf{x}), W = U, L$, is asymmetric with respect to x, any comparison with that tail copula automatically amounts to tail asymmetry (inequality) as there is always a 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 in Einmahl et al. (2012), intra-tail symmetry 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 we detect a considerable amount of intra-tail asymmetries in our comprehensive empirical study for the U.S. stock market in Section 5, and also for foreign exchange rate pairs, similar findings hold, see Appendix D and E.

As the tail copula is the main component for our test, we sketch relevant statistical results for appropriate estimators. To keep notation short, for the remainder of this section we only state the definition, assumptions and results for the estimator in the lower tail case. The upper tail version follows analogously from (1). For estimation of $\Lambda_{\mathbb{X}}(\mathbf{x})$, marginal quantile functions $F_{i,\mathbb{X}}^{-1}, i = 1, 2$, are approximated non-parametrically by the empirical counterpart $\widehat{F}_{i,\mathbb{X}}^{-1}(x) = \frac{1}{n+1}\sum_{j=1}^{n} 1\{X_j^{(i)} \leq x\}$ for each x and $i = 1, 2$. As marginals are typically unknown, empirical distributions yield sufficient flexibility for obtaining consistent estimates in a general setup. The reduced model misspecification risk, however, comes at the price of lower efficiency in comparison to parametric estimates based on the correct but in practice unknown form of the marginals. In the case of the later discussed multiplier bootstrap for tail copulas, however, inference is substantially complicated by accounting for pre-estimated marginals requiring multipliers also in the input marginals for an overall unbiased procedure (Bücher and Dette (2013)).

For an estimator of Λ^L, the limit in the definition of the tail copula (1) is replaced by the value at an arbitrary small point $t = k_{\mathbb{X}}/n = o(n)$ with the sample size $n \to \infty$ and the effective sample size $k_{\mathbb{X}} \to \infty, k_{\mathbb{X}}$. A consistent estimator is then given by

$$\widehat{\Lambda}_{\mathbb{X}}^{L}(x^{(1)}, x^{(2)}) = \frac{1}{k_{\mathbb{X}}}\sum_{m=1}^{n} 1\left\{X_m^{(1)} < \widehat{F}_{1,\mathbb{X}}^{-1}(x^{(1)}k_{\mathbb{X}}/n), X_m^{(2)} < \widehat{F}_{2,\mathbb{X}}^{-1}(x^{(2)}k_{\mathbb{X}}/n)\right\}, \tag{3}$$

$(x^{(1)}, x^{(2)}) \in \mathcal{S}$. By directly defining the quantile threshold $F_{i,\mathbb{X}}^{-1}(x^{(i)}k_{\mathbb{X}}/n), i = 1, 2$, the effective sample size $k_{\mathbb{X}}$ determines which observations are considered extreme. The choice of the tuning parameter $k_{\mathbb{X}}$ is subject to a bias-variance trade-off: For small values of $k_{\mathbb{X}}$ only few observations are used for estimation, which increases the variance while the approximation of empirical tails by extreme value distributions becomes more precise (low bias). On the other hand, using many observations for tail estimation (large $k_{\mathbb{X}}$) amounts to less disperse estimates (low variance), but the tail approximation may not be valid for less extreme observations (large bias).

Under Assumption1 asymptotic results for the empirical tail copula can thus be derived for appropriate assumptions on the tuning parameter $k_{\mathbb{X}}$ and minor smoothness conditions on $\Lambda(x)$, see Bücher and Dette (2013).

Assumptions 2. *For a bivariate random vector $\mathbb{X}$ we assume*

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

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

(A5) The partial derivatives $\frac{\partial\Lambda(x^{(1)},x^{(2)})}{\partial x^{(i)}}$, *exist and are continuous for* $x^{(i)} \in \mathbb{R}_+\backslash\{0\}$*.*

Assumption (A3) requires that the effective sample size $k_{\mathbb{X}}$ increases more slowly than n for $n \to \infty$ for consistency. The second-order regular variation condition (A4) (see Bücher and Dette (2013)) requires the bias of the tail copula approximation to vanish sufficiently fast with rate A which is also key for an appropriate multiplier bootstrap procedure to work. In practice, this only imposes a corresponding slightly tighter condition on the expanding rate of $k_{\mathbb{X}}$. It is satisfied by standard tail distributions such as e.g. the Clayton copula, where $A(t)$ is asymptotically of order $1/t^{\theta}$ with $\theta > 0$. Then $k_{\mathbb{X}}$ should be at most of order $n^{\frac{2\theta}{1+2\theta}} < n$ in order to satisfy the conditions. Note that in Assumption (A5), continuity of the partial derivative of the tail copula is not required at the origin as in Schmidt and Stadtmüller (2006). This is crucial in practice, in order to cover cases of tail dependence such as tail factor models which are excluded otherwise.

Under Assumptions (A1)-(A5), the asymptotic distribution for the tail copula can be derived as

$$\sqrt{k_{\mathbb{X}}}(\widehat{\Lambda}_{\mathbb{X}}(x^{(1)}, x^{(2)}) - \Lambda_{\mathbb{X}}(x^{(1)}, x^{(2)})) \xrightarrow{w} \mathbb{G}_{\widehat{\Lambda},\mathbb{X}}(x^{(1)}, x^{(2)}) \tag{4}$$

where $\xrightarrow{w}$ denotes weak convergence in sup-norm over each compact set in $\mathbb{R}^2_+$ in the sense of Hoffmann-Jørgensen (see, e.g., Bücher and Dette (2013)) , and

$\mathbb{G}_{\widehat{\Lambda},\mathbb{X}}$ is a bivariate Gaussian field of the form

$$\mathbb{G}_{\widehat{\Lambda},\mathbb{X}}(x^{(1)},x^{(2)}) = \mathbb{G}_{\tilde{\Lambda},\mathbb{X}}(x^{(1)},x^{(2)}) - \frac{\partial\Lambda(x^{(1)},x^{(2)})}{\partial x^{(1)}}\mathbb{G}_{\tilde{\Lambda},\mathbb{X}}(x^{(1)},\infty) - \frac{\partial\Lambda(x^{(1)},x^{(2)})}{\partial x^{(2)}}\mathbb{G}_{\tilde{\Lambda},\mathbb{X}}(\infty,x^{(2)})$$

with $\mathbb{G}_{\tilde{\Lambda},\mathbb{X}}(x^{(1)},x^{(2)})$ a *centered* Gaussian field with covariance

$$\mathbb{E}(\mathbb{G}_{\tilde{\Lambda},\mathbb{X}}(x^{(1)},x^{(2)})\mathbb{G}_{\tilde{\Lambda},\mathbb{X}}(v^{(1)},v^{(2)})) = \Lambda(\min(x^{(1)},v^{(1)}),\min(x^{(2)},v^{(2)})),(v^{(1)},v^{(2)}) \in \mathbb{R}^2_+.$$

This result follows directly from Theorem 2.2. in Bücher and Dette (2013). In particular with Assumption (A5), the feasible tail copula estimator (3) leads to a limiting process $\mathbb{G}_{\widehat{\Lambda}}$. Note that direct convergence of a tail copula estimator to the limit process $\mathbb{G}_{\widehat{\Lambda}}$ could only be obtained if the marginals in such an estimator of $\Lambda_{\mathbb{X}}$ were assumed as known (see Lemma 2.1 in Bücher and Dette (2013)).

3 A NEW TESTING METHODOLOGY AGAINST TAIL ASYMMETRY AND INEQUALITY

3.1 Test Idea, Asymptotic Properties, and Implementation

Generally, we test the global null hypothesis of equality between tail copulas by checking for local violations of the null over a collection of disjoint subsets of the relevant support ($\mathcal{S}$). 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, statistical tests are only consistent if all possible permutations of arguments in the tail copulas are considered, i.e. checking both $\Lambda_{\mathbb{Z}}(x^{(1)},x^{(2)})$ and $\Lambda_{\mathbb{Z}}(x^{(2)},x^{(1)}),\mathbb{Z} = \mathbb{X},\mathbb{Y}$). 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 the domain. Yet, we account for possible tail differences within 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 $\mathbb{X}$ with the tail copula of $\mathbb{Y} = (Y^{(1)}, Y^{(2)})$ 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 $S(\mathbb{X}, (Y^{(1)}, Y^{(2)})) \neq S(\mathbb{X}, (Y^{(2)}, Y^{(1)}))$. To fix this, we propose to analyze tail copula differences in *both* directions of the unit simplex hull searching for differences between tail copulas over distinct, pre-determined subintervals of the unit simplex. In this way, test power strongly benefits from intra-tail asymmetric tail copulas, while in standard intra-tail symmetric cases it features similar, yet slightly better test properties as competing tests.

For ease of exposition, in the following we focus in notation on the test against tail inequality. Results for the test against tail asymmetry can be directly obtained by exchanging $\Lambda_{\mathbb{X}}$ by $\Lambda_{\mathbb{X}}^U$ and $\Lambda_{\mathbb{Y}}$ by $\Lambda_{\mathbb{X}}^L$. We apply M Cramér-von Mises tests on $M/2$ disjoint subintervals of the unit simplex $\mathcal{S}$ where the decomposition of $\mathcal{S}$ is complete, i.e. the union of these subsets equals $\mathcal{S}$. The *global* null hypothesis is

$$H_0 : \Lambda_{\mathbb{X}} = \Lambda_{\mathbb{Y}} \text{ over } \mathcal{S}, a.s.,$$

consisting of M individual null hypotheses of the form

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

where $\mathcal{I}_1, ..., \mathcal{I}_{M/2}$ are complete decomposition of $[0,1]$ into disjoint, equidistant subintervals. Note that global tail equality $H_0 : \bigcap_{m=1}^{M} H_{0,m}$ naturally implies tail equality over each subset. Empirical marginal test statistics are given by

$$\widehat{S}^m(\mathbb{X}, \mathbb{Y}) = \begin{cases} \frac{k_{\mathbb{X}} k_{\mathbb{Y}}}{k_{\mathbb{X}} + k_{\mathbb{Y}}} \int_{\mathcal{I}_m} (\widehat{\Lambda}_{\mathbb{X}}(\phi, 1-\phi) - \widehat{\Lambda}_{\mathbb{Y}}(\phi, 1-\phi))^2 \, \mathbf{d}\phi, & m = 1, \dots M/2 \\ \frac{k_{\mathbb{X}} k_{\mathbb{Y}}}{k_{\mathbb{X}} + k_{\mathbb{Y}}} \int_{\mathcal{I}_{m-\frac{M}{2}}} (\widehat{\Lambda}_{\mathbb{X}}(\phi, 1-\phi) - \widehat{\Lambda}_{\mathbb{Y}}(1-\phi, \phi))^2 \, \mathbf{d}\phi, & m = (M/2)+1 \dots M. \end{cases}$$

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 $\Lambda_{\mathbb{Y}}$ for $m \geq (M/2)+1$ guarantees that tail copulas are compared over the entire unit simplex, e.g. in *both* directions. If $H_{0,m}$ is true, $\widehat{S}^m \approx 0$, while $\widehat{S}^m$ is largely different from zero otherwise.

The following proposition provides the marginal test distributions in the i.i.d. case. Subsection 4.1 discusses extensions for time series data.

Proposition 2. *Let Assumptions (A1)-(A4) hold for* $\mathbb{X}, \mathbb{Y}$.

Then under H_0 *for each* $m = 1, ..., M$

$$\widehat{S}^m \xrightarrow{w} S^m \ ,$$

where

$$S^m = \int_{\mathcal{I}_m} \left(\sqrt{1-\lambda}\,\mathbb{G}_{\widehat{\Lambda},\mathbb{X}}(\phi, 1-\phi) - \sqrt{\lambda}\,\mathbb{G}_{\widehat{\Lambda},\mathbb{Y}}(\phi, 1-\phi)\right)^2 \boldsymbol{d}\phi,$$

with $\lambda = \lim_{n\to\infty} \frac{k_{\mathbb{X}}}{k_{\mathbb{X}}+k_{\mathbb{Y}}} \in (0,1)$.

Under H_1*,however,* $\exists m : \widehat{S}^m \xrightarrow{\mathbb{P}} \infty$.

Note, the processes $\mathbb{G}_{\widehat{\Lambda},\mathbb{X}}, \mathbb{G}_{\widehat{\Lambda},\mathbb{Y}}$ correspond to $\mathbb{G}_{\widehat{\Lambda}}(x^{(1)}, x^{(2)})$ from Equation (4). Due to the complexity of the limiting stochastic processes, closed forms of the asymptotic distributions do not exist and have to be simulated. Therefore we follow Bücher and Dette (2013) and approximate the finite sample distribution of $(\widehat{S}^m)$ by a multiplier bootstrap for each $m = 1, \ldots, M$. See also Rémillard and Scaillet (2009) who introduced multiplier techniques for copula inference. For the construction of the bootstrap version of the test statistic, we require the definition of $\mathbb{Z}$-specific multipliers $\xi_i^{\mathbb{Z}}$ where $\mathbb{Z} \in \{\mathbb{X}, \mathbb{Y}\}$ helps to streamline notation.

Assumptions 3 (cont.).

(A6) Multipliers $\xi_i^{\mathbb{Z}}$ *are iid random variables with* $\mathbb{E}(\xi_i^{\mathbb{Z}}) = \mathbb{V}(\xi_i^{\mathbb{Z}}) = 1.$ *and* $E[|\xi_i^{\mathbb{Z}}|^\nu] < \infty$ *for* $\nu > 1$ *which are independent of* $\mathbb{Z}$ *for all* $i = 1, \ldots, n_{\mathbb{Z}}$.

For each bootstrap draw $b = 1, \ldots, B$ of these multipliers $\xi_1^{\mathbb{Z},(b)}, \ldots, \xi_{n_\mathbb{Z}}^{\mathbb{Z},(b)}$, we can construct $\widehat{S}^{m,(b)}$ for $m = 1, \ldots, M$

$$\widehat{S}^{m,(b)}(\mathbb{X}, \mathbb{Y}) = \frac{k_\mathbb{X} k_\mathbb{Y}}{k_\mathbb{X} + k_\mathbb{Y}} \int_{\mathcal{I}_m} \Big((\widehat{\Lambda}_\mathbb{X}^{(b)}(\phi, 1-\phi) - \widehat{\Lambda}_\mathbb{X}(\phi, 1-\phi)) - \\ (\widehat{\Lambda}_\mathbb{Y}^{(b)}(\phi, 1-\phi) - \widehat{\Lambda}_\mathbb{Y}(\phi, 1-\phi)) \Big)^2 \, \mathrm{d}\phi, \tag{5}$$

where $\widehat{\Lambda}_\mathbb{Z}^{(b)}(\mathrm{x})$ for $\mathbb{Z} \in \{\mathbb{X}, \mathbb{Y}\}$ is obtained with standardized multipliers $\widetilde{\xi}_i^{\mathbb{Z},(b)} = \xi_i^{\mathbb{Z},(b)} / \overline{\xi_i^{\mathbb{Z},(b)}}, i = 1, ..., n_\mathbb{Z}$ as

$$\widehat{\Lambda}_\mathbb{Z}^{(b)}(x^{(1)}, x^{(2)}) = \frac{1}{k_\mathbb{Z}} \sum_{i=1}^{n_\mathbb{Z}} \widetilde{\xi}_i^{\mathbb{Z},(b)} 1 \left\{ Z_i^{(1)} \geq \tilde{F}_{1,\mathbb{Z}}^{-1}(1 - x^{(1)} k_\mathbb{Z}/n_\mathbb{Z}), Z_i^{(2)} \geq \tilde{F}_{2,\mathbb{Z}}^{-1}(1 - x^{(2)} k_\mathbb{Z}/n_\mathbb{Z}) \right\}$$

$$\tilde{F}_{j,\mathbb{Z}}(x) = \frac{1}{n_\mathbb{Z}} \sum_{i=1}^{n_\mathbb{Z}} \tilde{\xi}_i^{\mathbb{Z}} 1 \left\{ Z_i^{(j)} \leq x \right\}, j = 1, 2, \tag{6}$$

Note that not only the empirical tail copula, but also the empirical marginal distributions require multiplier bootstrapping for the procedure to yield consistent results. The sample size could theoretically differ for $\mathbb{X}$ and $\mathbb{Y}$ which is marked by the index of $n_\mathbb{Z}$.

Then the bootstrap version $\widehat{S}^{m,\star}$ of the test statistic S^m is obtained as the empirical distribution of $\widehat{S}^{m,(1)}, \ldots S^{m,(B)}$

The following asymptotic result shows the weak convergence of $\widehat{S}^{m,\star}$ to the same asymptotic distribution as $\widehat{S}^m$, conditional on the bootstrap samples. Moreover, it ensures consistency of the multiplier bootstrap version of the test in the i.i.d. case.

Proposition 3. *Let (A1)-(A6) hold. Then under H_0, conditionally on the multipliers,*

$$\widehat{S}^{m,\star} \xrightarrow{w} S^m, m = 1, ..., M,$$

while under $H_1, \exists m : \widehat{S}^{m,\star} \xrightarrow{\mathbb{P}} \infty$.

In practice for the i.i.d. case, we set $\xi_i \sim Exp(1)$. Note, whenever $\mathbb{X}$ and $\mathbb{Y}$ are dependent, one has to use the same multiplier series for both $\mathbb{X}$ and $\mathbb{Y}$. Finally, a consistent Monte Carlo p-value for hypothesis $H_{0,m}$ is given by

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

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).

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.

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

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 are 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 FWER-Bonferroni correction. Also, BH-FDR is better suited for (positively) dependent p-values, which is a natural assumption for our setting. However, as 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).

1. *Determine* $k_{\mathbb{X}}, k_{\mathbb{Y}}$*, and estimate both tail copulas, i.e. calculate* $\widehat{\Lambda}_{\mathbb{X}}(\phi, 1-\phi), \widehat{\Lambda}_{\mathbb{Y}}(\phi, 1-\phi), \phi \in [0,1]$*.*
2. *Set* M*. Decompose* $[0,1]$ *into* $M/2$ *disjoint, equally sized subintervals, i.e.* $\mathcal{I}_1$, ..., $\mathcal{I}_{M/2}$*.*
3. *Calculate* $\widehat{S}^m, m = 1, ..., M$*.*
4. *Set* B*. Calculate* $\widehat{S}^{m,\star}$ *with* $\widehat{S}^{m,(b)}, b = 1, ..., B$ *for* $m = 1, ..., M,$*.*
5. *Calculate* $\widehat{p}^m, m = 1, ..., M$*.*
6. *Fix an error rate* α*. Apply a multiple testing routine on* $\widehat{p}^1, ..., \widehat{p}^M$ *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 $\lim_{n,B\to\infty} FDR = e \leq \alpha$, and in case of FWER, $\lim_{n,B\to\infty} \mathbb{P}\big(\cup_{m=1}^{M} \{p^m \leq \alpha^m | H_0\}\big) = f \leq \alpha$. Unless otherwise stated, in simulations and applications we work with $B = 1499$ bootstrap repetitions; note the necessary correction of B (1499 instead of 1500) which ensures consistency of the p-value.

The choice of M is subject to a trade-off between test power and precision of localization of tail differences. A larger M amounts to lower power as less data fall into finer subintervals, and the multiplicity penalty of the individual p-values increases in M, making rejections even less likely. A larger M also means, the tests very precisely pin down very narrow subintervals with significant tail dependence differences. 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 implode as the harsh p-value adjustment and the decreasing number of observations in small subsets 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, as we do not strive to determine an *optimal* number of subsets we suggest to apply the test several times over a set of grids. Consequently, we combine p-values of the different grids to one embracing test and we refrain from any further multiplicity adjustment.

Test algorithm (2).

1. *For J different grids that increase in grid fineness, individually execute Test Algorithm (1) with M_j subsets, where $M_j = 2j, j = 1, ..., J$.*
2. *For each grid, adjust the p-values for multiplicity: $(\tilde{p}_1^1, \tilde{p}_1^2), ..., (\tilde{p}_J^1, ..., \tilde{p}_J^{2J})$.*
3. *For each grid, pick the minimal adjusted p-value:*

$$(\tilde{p}_1^* = \min(\tilde{p}_1^1, \tilde{p}_1^2), ..., \tilde{p}_J^* = \min(\tilde{p}_J^1, ..., \tilde{p}_J^{2J})).$$

4. *Reject the global H_0 if at least one $\tilde{p}_j^*$ is smaller than α.*

Note, this aggregating test does not adjust the grid-specific p-values a second time. This approach would control exactly for the error rate α, if $\tilde{p}_1^*, ..., \tilde{p}_J^*$ were perfectly dependent. For asymptotic control, however, we can relax this condition to *nearly* perfect dependence, see condition 7 below. This is important, as assuming perfect dependence between grid-minimal p-values is much more rigid than postulating only nearly perfect dependence. For simplicity, we state the following result only for FWER control. We denote α_j as the asymptotic test size of the jth Test (1).

Proposition 4. *For Test (2), if*

$$\mathbb{P}(\cup_{j=1}^J \tilde{p}_j^* \leq \alpha | H_0) \uparrow \max(\alpha_1, ..., \alpha_J), \text{ as } J \to \infty, \tag{7}$$

it holds that

$$\lim_{n,B,\to\infty} \mathbb{P}(\cup_{j=1}^J \tilde{p}_j^* \leq \alpha | H_0) = \alpha.$$

In the proof of Proposition 4 it is key that Condition 7 is fulfilled and that (realized) test sizes of Test (1) converge to zero as $M \to \infty$. Simulation results reported in Appendix B confirm that both conditions are satisfied in standard settings. In particular, we find Test (2) consistently obeys the α-limit due to individual undersizedness of Test (1) and nearly perfect dependence between grid-minimal p-values (see Figure 1 appx). Note that a more explicit lower bound of the strength of dependence between the p-values in (7) can only be achieved by imposing specific parametric forms on the dependence structure (see e.g. Bodnar and Dickhaus (2014)) which would, however, limit the generality of our approach.

3.2 Local Tail Asymmetry

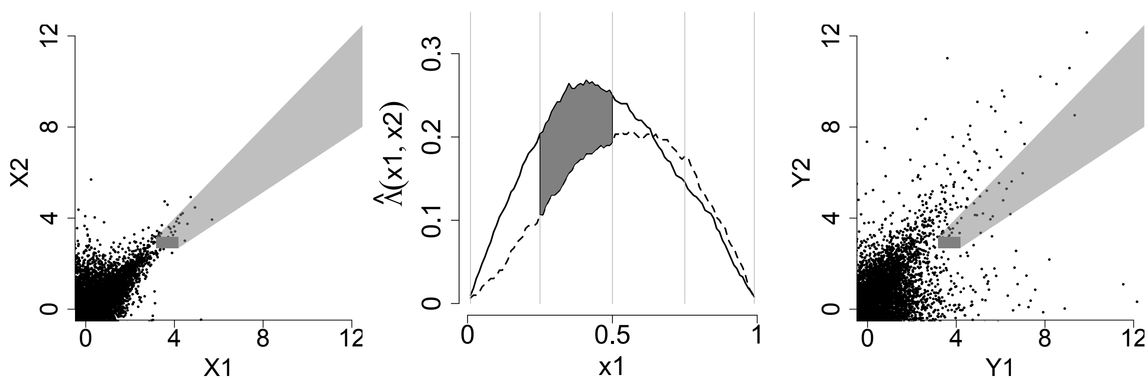
One main feature of our test is that we can *localize* tail dependence differences. This enriches the binary test decision on tail asymmetry/inequality as we can find subspaces in $\mathbb{R}^2_+$ where tail asymmetry/inequality can be expected. If the global null is rejected, significant individual 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.

$$
\begin{aligned}
Q_{\mathbb{X}} &= \Big(F^{-1}_{1,\mathbb{X}}(1-x^{(1)}k/n), F^{-1}_{1,\mathbb{X}}(1-x^{(2)}k/n)\Big) \times \Big(F^{-1}_{2,\mathbb{X}}(1-x^{(1)}k/n)), F^{-1}_{2,\mathbb{X}}(1-x^{(2)}k/n)\Big),\\
Q_{\mathbb{Y}} &= \Big(F^{-1}_{1,\mathbb{Y}}(1-x^{(1)}k/n), F^{-1}_{1,\mathbb{Y}}(1-x^{(2)}k/n)\Big) \times \Big(F^{-1}_{2,\mathbb{Y}}(1-x^{(1)}k/n)), F^{-1}_{2,\mathbb{Y}}(1-x^{(2)}k/n)\Big).
\end{aligned}
$$

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 VaR events, cause tail dependence differences. Conditional on realized returns of $\mathbb{X}$ ($\mathbb{Y}$) falling into $Q_{\mathbb{X}}$ ($Q_{\mathbb{Y}}$), tail dependence of $\mathbb{X}$ and $\mathbb{Y}$ differ; conditional on $\mathbb{X}(\mathbb{Y}) \notin Q_{\mathbb{X}}$ ($Q_{\mathbb{Y}}$), $\Lambda_{\mathbb{X}}$ and $\Lambda_{\mathbb{Y}}$ do not differ significantly.

This additional information might improve tail risk anticipation for regulators, or tail risk-based hedge and trading strategies for investors as those market times are identified which typically induce behavior of bivariate extremes to shift.

Figure 1: Left and right: Upper-right quadrants of scatterplots for $\mathbb{X}, \mathbb{Y}$, 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 $\Lambda(x^{(1)}, x^{(2)}) = x^{(1)} + x^{(2)} - \left[(1 - \psi^{(1)})x^{(1)} + (1 - \psi^{(2)})x^{(2)} + ((\psi^{(1)}x^{(1)})^{-\theta} + (\psi^{(2)}x^{(2)})^{-\theta})^{\theta}\right]$ (see Tawn (1988)), with parameters $(\psi^{(1)}, \psi^{(2)}, \theta) = (0.1, 0.6, 0.1)$, $(\psi^{(1)}, \psi^{(2)}, \theta) = (0.1, 0.5, 0.4)$. The shaded rectangles show the tail asymmetric tail regions; the homogeneity of the tail copula allows to extrapolate this region far into the sample tail. Center: Estimated tail copulas for $x^{(1)} \in \{0.01, 0.02, ..., 0.99\}, k = 500, n = 10000, M = 8$. The shaded area indicates over which subset both tail copulas significantly differ.



4 FINITE SAMPLE STUDY

4.1 Serially Dependent Data

In general for financial time series, the i.i.d. assumption (A1) cannot be fulfilled as financial data typically exhibit strong serial dependence. Though, standard extreme value theory and the multiplier bootstrap rely on the independence assumption. We therefore consider two different approaches to address this problem.

The standard applied approach is to fit an appropriate time series specification, such as e.g. ARMA-GARCH, to the financial raw returns and work with obtained standardized residuals. For a valid time series pre-filter, the latter should roughly resemble an i.i.d. series, and can thus be used for further inference (see, e.g. McNeil and Frey (2000) in the univariate case). It is intuitively clear, that asymptotically such parametric pre-whitening at rate $\sqrt{n}$ should not affect rate and consistency of the slower converging nonparametric tail dependence estimates and thus of the test statistics. In practice, however, the pre-step might still lead in particular to second order effects in the variance for finite samples. In the following subsection we show that such effects are negligible for our test at considered standard sample sizes.

For empirical copulas of dependent data, another remedy is to assume stationarity coupled with appropriate mixing conditions, which consequently allow to directly use unfiltered returns for estimation. Valid statistical inference is ensured by adjusting the bootstrap procedure: 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 ARMA and GARCH models.

We propose to use the dependent data bootstrap methodology also for empirical tail copulas. Thus we define a dependent multiplier sequence as follows.

Assumptions 2.

- ($A6a^*$) *Tapered block multipliers* $(\xi_{j,n})_{j=1,\dots,n}$ *are strictly stationary with* $\mathbb{E}[\xi_{0,n}] = 0$, $E[\xi_{0,n}^2] = 0$ *, and* $\mathbb{E}[|\xi_{0,n}|^\nu] < \infty$ *for all* $\nu \geq 1$ *independent of* $\mathbb{Z}_1, \dots, \mathbb{Z}_n$*.*

- ($A6b^*$) *For any* j*,* $\xi_{j,n}$ *is independent of* $\xi_{j+h,n}$ *for all* $h \geq l(n)$ *where* $l(n)$ *is a strictly positive, deterministic sequence with* $l(n) \to \infty$ *and* $l(n) = o(n)$*.*

- ($A6c^*$) *For any* $h \in \mathbb{Z}$*,* $\exists\, v : \mathbb{R} \to [0, 1]$ *with* $\mathbb{E}(\xi_{0,n}, \xi_{h,n}) = v(h/l(n))$ *where* v *continuous at 0 and symmetric around 0 with* $v(0) = 1$ *and* $v(x) = 0$ *for* $|x| > 1$*.*

Instead of the i.i.d. Assumption (A1) the underlying stochastic process $\mathbb{Z} \in \{\mathbb{X}, \mathbb{Y}\}$ is required to be strictly stationary and $\alpha_{\mathbb{Z}}$-mixing with $\alpha_{\mathbb{Z}}(r) = \alpha_{\mathbb{Z}}(\mathcal{F}_s, \mathcal{F}_{s+r}) = \sup_{A \in \mathcal{F}_s, B \in \mathcal{F}_{s+r}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|$, for the $(\mathbb{Z}_1, \ldots \mathbb{Z}_t)$-induced filtration $\mathcal{F}_t$. The rate of decay $\alpha_{\mathbb{Z}}(r) = \mathcal{O}(r^{-a_{\mathbb{Z}}})$ where $a_{\mathbb{Z}} > 0$ for $r > 0$ marks the degree of admissible serial dependence. In contrast to standard copulas, weak convergence of empirical tail copulas in an α-mixing set-up is a challenging problem which has only been touched upon very recently in special cases so far, see e.g. Bücher and Ruppert (2013) and Bücher and Segers (2017). A general proof is beyond the scope of this paper, however, the later paper suggests, that this is possible under fairly general conditions in our bivariate set-up, in particular allowing for processes of ARMA-GARCH-type which are key to financial applications. For the case $a_{\mathbb{Z}} > 6$, Bücher and Kojadinovic (2016) show a general multiplier bootstrap theorem for dependent data and multipliers of the type as in Assumption 2 for the case of standard copulas. This is key for formally deriving the consistency of the tapered multiplier bootstrap with block length $l(n) \to \infty$, where $l(n) = \mathcal{O}\left(n^{1/2-\epsilon}\right), 0 < \epsilon < 0.5.$, but the theoretical extension to tail copulas is non-trivial and left for future research.

We construct the tapered multiplier bootstrap with a dependent multiplier series as in Assumption 2 entering both empirical copula and marginal empirical distribution in (6). In each bootstrap round b, these yield the tapered version of test statistic $\widehat{S}^{m,(b),tap}$ by plugging them into equation (5), from which the final $\widehat{S}^{m,\star,tap}$ can be constructed for each $m = 1, \ldots, M$. For the choice of multiplier block length l under which the generated multiplier series mimics the resulting dependence structure of $\mathbb{Z}$ we follow Bücher and Ruppert (2013) in their implementation guidelines setting $l(n) = 1.25n^{1/3}$. Moreover, for the tapered block multiplier bootstrap, we employ the 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 from the R-package `npcp`, see Kojadinovic (2015). Our comprehensive simulation study underlines the validity of the tapered multiplier bootstrap for the empirical tail copula, suggesting that $\widehat{S}^{m,\star,tap} \xrightarrow{w} S^m$ for $m = 1, \ldots, M$. With this approach, potential model misspecification from pre-filtering in tail dependence is avoided which may be a problem for large, high-dimensional data sets where automatic GARCH fitting is challenging and computationally expensive.

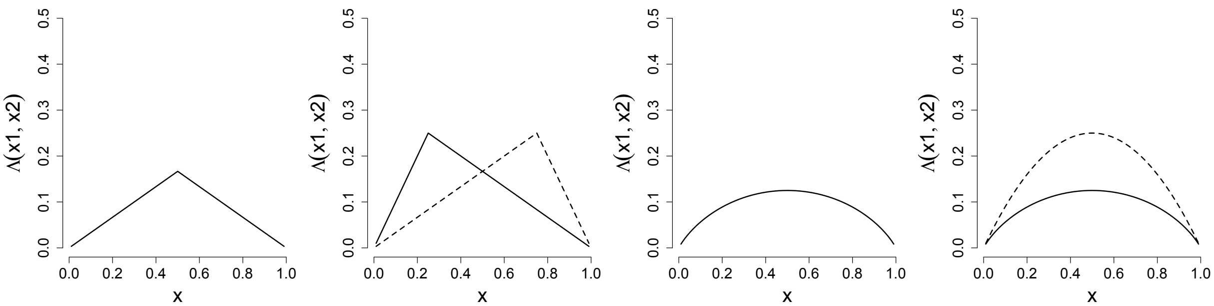
4.2 Simulations

We now compare the finite sample performance of our test with the TDC test, and the BD13 test. We focus on non-parametric tests as in practice parametric specifications may suffer from a model bias, especially if intra-tail asymmetry is not accounted for. We study two types of dependence models that are frequently used in finance. For each copula, we impose one parametrization that fulfills the null, and one that violates the null, leaving us with four DGPs.

First for DGP1 and DGP2, we employ the (implicit) factor model copula (see e.g. Einmahl et al. (2012)and Appendix B). DGP1 is intra-tail symmetric with tails between $\mathbb{X}$ and $\mathbb{Y}$ equal, and thus the null is true. DGP2 represents the class of intra-tail asymmetric copulas which violate the null. See Figure 2, first and second from the left, for $\Lambda(x^{(1)}, 1 - x^{(1)}), x^{(1)} \in [0, 1]$. For details on the exact specification, we refer to Appendix B.

Second, representing the broad class of Archimedean copulas, we employ the Clayton copula for DGP3 and DGP4 which is also a popular building block for more complex copula models, such as mixtures of copulas, see e.g. Patton (2006). For the Clayton copula, only the lower left part of the distribution features tail dependence. Thus $\Lambda^U(x^{(1)}, x^{(2)}; \theta) = 0$ but $\Lambda^L(x^{(1)}, x^{(2)}; \theta) = (x^{(1)^{-\theta}} + x^{(2)^{-\theta}})^{-1/\theta}$ with $\Lambda^L(\theta)$ increasing in $\theta \in [0, \infty)$. DGP3 is given by $\mathbb{X}, \mathbb{Y} \sim$ Clayton$(\theta = 0.5)$, thus the null is true, see Figure (2) second from the right. The specific choice of $\theta = 0.5$ implies a TDC of $\iota = 0.25$, which roughly corresponds to a TDC of a bivariate t-distribution with correlation 0.5 and four degrees of freedom (McNeil et al. (2005), p.211). For DGP4, $\mathbb{X} \sim$ Clayton$(\theta = 0.5)$, and $\mathbb{Y} \sim$ Clayton$(\theta = 1)$. Thus, tail equality is violated as the TDC of $\mathbb{Y}$ is $\iota = 0.5$. See Figure (2), first from the right.

Figure 2: Tail copulas of DGPs 1 to 4 from left to right. Note, for DGP2, the solid lines represents $\Lambda(x^{(1)}, x^{(2)}), x^{(2)} = 1 - x^{(1)}$, whereas the dashed line shows $\Lambda(x^{(2)}, x^{(1)})$. For DGP4, two different specifications of the Clayton copula are used for X and Y.



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. We apply the test to *raw* GARCH returns, and to estimated standardized GARCH residuals. 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. and link serially dependent marginals $\mathbb{Z} = (Z^{(1)}, Z^{(2)}), \mathbb{Z} = \mathbb{X}, \mathbb{Y}$, pairwise by the (implicit) copulas of DGPs 1 to 4, allowing us to study the effect of conditional heteroscedasticity on test performance. Thus it holds $Z_t^{(i)} = \sigma_t^{(i)} \eta_{t,\mathbb{Z}}^{(i)}$ with $\sigma_{t,\mathbb{Z}}^{2,(i)} = \omega + \alpha^{(i)} Z_{t-1}^{(i)} + \beta^{(i)} \eta_{t-1,\mathbb{Z}}^{2,(i)}, t = 1, ..., n_{\mathbb{Z}}$ where $\eta_{\mathbb{Z}} = (\eta_{\mathbb{Z}}^{(1)}, \eta_{\mathbb{Z}}^{(2)}) \overset{\text{iid}}{\sim} F_{\eta,\mathbb{Z}}(x^{(1)}, x^{(2)}) = C_{\eta,\mathbb{Z}}(F_{\eta,\mathbb{Z},1}(\eta_{\mathbb{Z}}^{(1)}), F_{\eta,\mathbb{Z},2}(\eta_{\mathbb{Z}}^{(2)}))$. We set $\omega = 0.01, \alpha = 0.15$ and $\beta = 0.8$ such that $\omega + \alpha + \beta$ is close to one. This mimics parameter values often found in financial returns, see e.g. Engle and Sheppard (2001). To impose different tail structures on the time series, we use DGPs 1 to 4 to model the error copula $C_{\eta,\mathbb{Z}}$ generating $\eta_{t,\mathbb{Z}} = (\eta_{t,\mathbb{Z}}^{(1)}, \eta_{t,\mathbb{Z}}^{(2)})$. Please see Appendix B for details.

For sample sizes $n = 750, 1500$, varying values of the effective sample size k, and a nominal test level of $\alpha = 0.05$, 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. Furthermore, we employ Test Algorithm (2) which merges 15 different grids with grid sizes $M_j = 2j, j = 1, ..., 15,$. For some grids, this implies that subintervals are only roughly of equal length. The TDC test is carried out using the multiplier bootstrap at points $x^{(1)} = x^{(2)} = 0.5$. The number of simulations is set to $S = 500$ for each setting.

Table (1) reports empirical rejection frequencies for i.i.d. marginals, filtered GARCH marginals, unfiltered GARCH marginals, GARCH marginals with the block and tapered bootstrap, and the moderate sample size $n = 1500$ while we refer to Appendix B for simulation results with $n = 750$ which is small for the nonparametric estimator. Also, we study the effect of varying effective sample sizes $k \in \{\lfloor 0.1n \rfloor, \lfloor 0.2n \rfloor, \lfloor 0.3n \rfloor\}$. Note, $\Lambda(x^{(1)}, x^{(2)}; k = k^*) = \Lambda(ax^{(1)}, ax^{(2)}; k = ak^*)$. 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 $x^{(1)} = x^{(2)} = 1$. In general, both Test (1) and Test (2) appear to be consistent. For i.i.d. marginals, both obey the nominal test size of $\alpha = 0.05$ (DGP1 and DGP3), irrespective of the choice of k. This is particularly important for Test ((2)) as it points out that grid-specific p-values appear to be sufficiently dependent to keep empirical size below α, although no additional multiplicity penalty is applied. While empirical test size remains untouched by k, the choice of effective sample size notably affects empirical power; for example, for DGP4, 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 $\widehat{\Lambda}_{\mathbb{X}}$ and $\widehat{\Lambda}_{\mathbb{Y}}$ 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 (Test (1)) show larger power than the TDC test, the BD13 test is more powerful in standard cases compared to Test (1). However, combining a multiple of single-grid tests, e.g. Test Algorithm (2), makes our test consistently more powerful than BD13.

Importantly, our test successfully rejects in case of intra-tail asymmetries, as shown by the empirical rejection frequencies for DGP2. Both the TDC test

Table 1: Empirical rejection probabilities for $\alpha = 5\%$, $S = 500$ repetitions and sample size $n = 1500$. Effective sample fraction k/n is evaluated at $(x^{(1)}, x^{(2)}) = (1, 1)$. DGP1: factor model satisfying H_0. DGP2: factor model violating H_0. DGP3: Clayton copula satisfying H_0. DGP4: Clayton copula violating the null. 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	TA(2)			18	6	TA(2)			18	6	TA(2)			18	6	TA(2)
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

and BD13 test fail to reject the null in this case and completely ignore intra-tail asymmetries. If the tail copula is intra-asymmetric, our power of our tests increases in the number of employed subsets. If the tail copula is symmetric, however, power decreases in M. It is thus advisable to apply Test (2).

Also, test results for GARCH filtered returns are in line with i.i.d. series. The estimation step of the GARCH residuals does not downgrade neither test power nor size. However, unfiltered GARCH returns should not be used: In the case of DGP4, test power implodes by roughly $50-75\%$ for all three tests. Empirical sizes for DGP1 are still fine, whereas empirical size of DGP3 generally is too large. The tapered block multiplier bootstrap produces results comparable to the multiplier bootstrap-based on i.i.d. and GARCH filtered marginals. Thus, we prefer a bootstrap adjustment over GARCH-filtering to address serial dependence it can handle serially dependent data and does not require pre-estimation of a parametric model. However, as Table 1 Appx in Appendix B suggests, the tapered block bootstrap should only be applied for larger sample sizes, since for $n = 750$ and GARCH marginals the tapered multiplier block bootstrap appears to be oversized and hence GARCH-filtered data should be used instead.

Finally, we find our aggregating Test (2) is throughout most powerful, while Test (1) with fixed grids is still 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 TAIL ASYMMETRIES IN THE US STOCK MARKET

Related studies, e.g. Ang and Chen (2002), focus on tail asymmetries in pairs of international stock indices, and point out that, especially during financial crises, correlations mainly between extreme losses increase. We are interested whether this finding also applies for sector pairs in the US stock market. Hence, we study possible tail asymmetries between daily returns of 49 US industry sectors. The dataset, available at http://mba.tuck.dartmouth.edu/

pages/faculty/ken.french/data_library.html, accessed on 03/01/2016, contains nearly 90 years of weighted returns of CRSP SIC codes-based industries of NYSE, AMEX, and NASDAQ stocks from 07/1926-01/2016. Fama and French (1994) and Chang et al. (2013) analyze earlier versions of this dataset.

We proceed as follows. We aim to detect tail asymmetry dynamics within the US stock market. Applying a rolling window analysis with window length of $n = 1500$, i.e. nearly six years, and a step size of 250 trading days, i.e. roughly 12 months, we arrive at 85 (overlapping) time periods. In each period, we build all possible bivariate industry combinations , $\mathbb{X} = (X^{(i)}, X^{(j)})$, and test the nulls

$$H_0 : \Lambda_{\mathbb{X}}^U = \Lambda_{\mathbb{X}}^L.$$

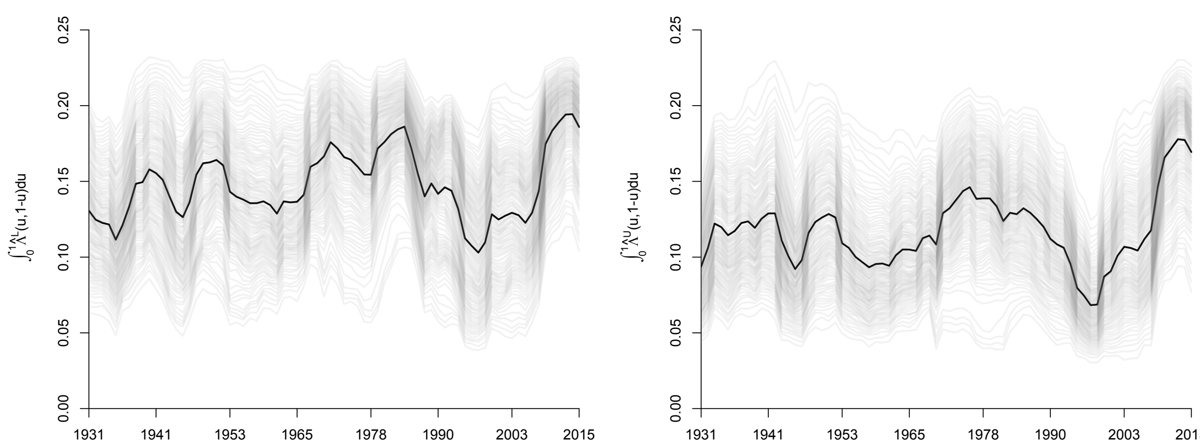
Discarding pairs with missing data, in each period, there are at most 1176 pairs to test against tail asymmetry. In total, we apply the test approximately $85,000$ times. To avoid possible model risk by pre-filtering the returns, we throughout analyze raw returns using the tapered block multiplier bootstrap; Section 4.1 and the results of the simulation study justify this approach. For completeness, however, we also computed results from GARCH pre-filtering. As there are only minor differences to the results from tapered bootstrap we only provide them in Appendix C of the Web-Appendix. We set the window parameter of the tapered block multiplier bootstrap to $l = 8$. Yet, we find no change of results worth mentioning when altering l. Also, we fix the effective sample size to $k = 0.2n$. This, too, is inspired by the findings in the simulation study. We are not interested in particular industry pairs as our focus is on tail asymmetry of the general market. Hence, a fixed k for all pairs is an operable solution to the question of number of extremes as over- and underestimation might eventually balance out when aggregating test decisions over all 1176 pairs. Note, this section studies tail asymmetries. In the online appendix, we also provide an empirical study on tail inequalities between foreign exchange rates.

To grasp the general evolution of lower and upper bivariate tails, we introduce a descriptive measure for upper and lower market tail dependence. In period t, for each pair i, we integrate the empirical tail copula $\widehat{\Lambda}_i(\phi, 1-\phi)$ over $[0,1]$ and provide empirical location statistics across all pairs, e.g. the mean and empirical quantiles. For the mean,

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

where n_t is the number of sectors in period t, and empirical quantiles are computed accordingly. It is easy to see that $\int_0^1 \Lambda(\phi, 1-\phi)\mathrm{d}\phi \in [0, 0.25]$. The lower (upper) bound is attained if pair i has no (perfect) tail dependence. Figure (3) shows the trajectory of the mean and q-quantiles, $q \in \{0.01, ..., 0.99\}$, for both upper and lower tails covering 1931 - 2015.

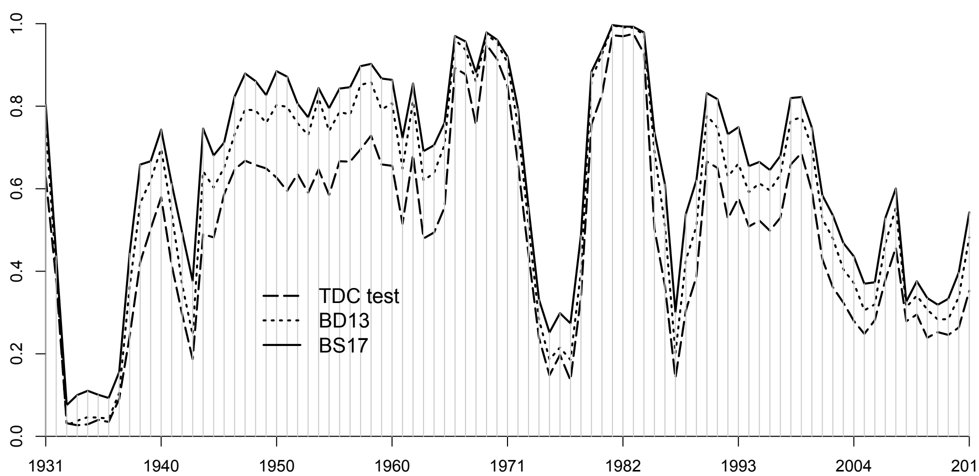
Figure 3: $\int_0^1 \widehat{\Lambda}(x, 1-x)du$ for all possible pairs (up to 1176) in each period; dark line: empirical mean; gray lines: empirical quantiles: $0.01i, i = 1, ..., 99$. Left: losses. Right: gains.



The H_0 of tail equality is tested by the TDC test, the BD13 test and Test (2), which aggregates over 15 grids in the spirit of the simulation study. Figure 4 displays trajectories of the share of rejections for each test, i.e. the share of tail asymmetric pairs by each test. Figure 5 documents the importance of non-standard tail events, i.e. *non-TDC* events that occur off the diagonal $(x^{(1)} = x^{(2)})$.

All tests indicate that most of the time, a substantial amount of tail asymmetries exists in the market. We find that our test reveals more tail asymmetries

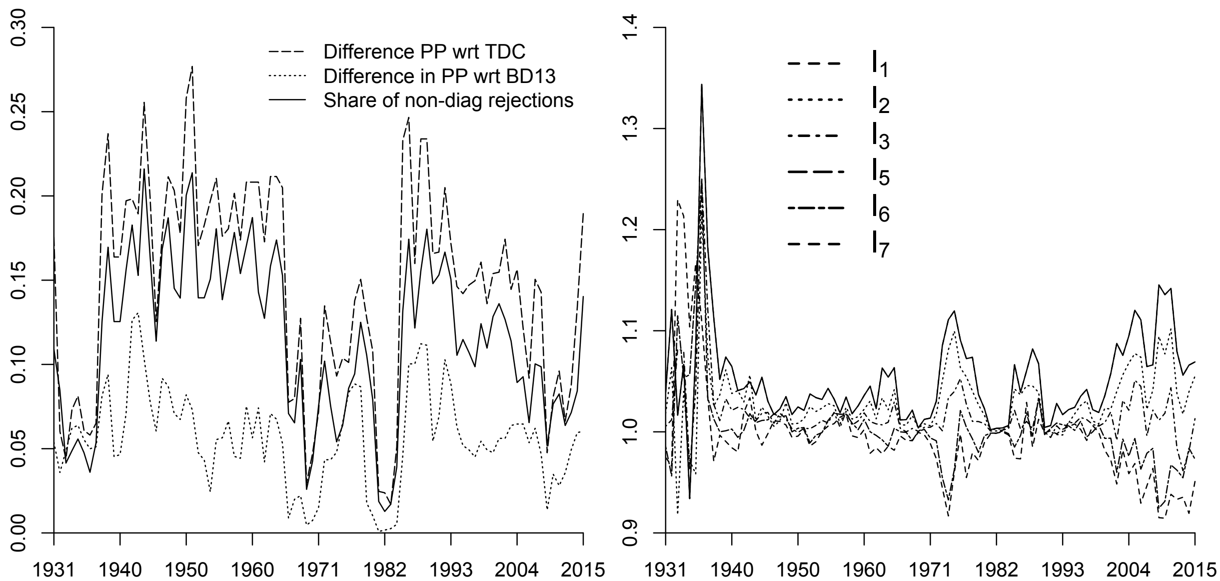
Figure 4: Dynamics of the percentage of detected tail asymmetries among all pairs using a rolling window of size $n = 1500$, and a step size of 250 trading days for the TDC test (dashed), BD13 test (dotted) and our test (solid), respectively.



than competing tests which we attribute to non-diagonal tail dependence and intra-tail asymmetry. Furthermore, we find tail asymmetry typically vanishes during financial crises, expect for the subprime crisis when tail asymmetries occurred more frequently than shortly before and afterwards. This finding may reflect the classical risk-return trade-off with a new livery: As lower tail dependence, i.e. the risk of joint extreme losses, spikes during financial distress, opportunities for joint extreme gains must counteractively increase as we detect more tail asymmetries during bear markets.

On average, our test finds that 64% (sd=0.25) of all pairs exhibit tail asymmetry. We can identify a long lasting phase of pronounced tail asymmetries between 1940-70 where on average 80% (sd=0.10) of all pairs are tail asymmetric. Collapses of the number of tail asymmetries strikingly coincide with of financial crises, such as the beginning of the Great Depression (1932-37), the Oil Crisis (1968-74 until 1972-78), Black Monday (1987) and the Asian and millennium crisis accumulating into the Dot-Com crisis (1995 - 2003). It is empirically documented that in crises losses increasingly move in extreme ways. We can only conclude that, during crises, the tendency of extreme gains to co-

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 $\mathcal{I}_i, i = 1, 2, ..., 7$, i.e. off-diagonally, compared to number of rejections in subsets $\mathcal{I}_4$, i.e. around $(x^{(1)} = 0.5, x^{(2)} = 0.5)$, based on a grid with $M = 14$.

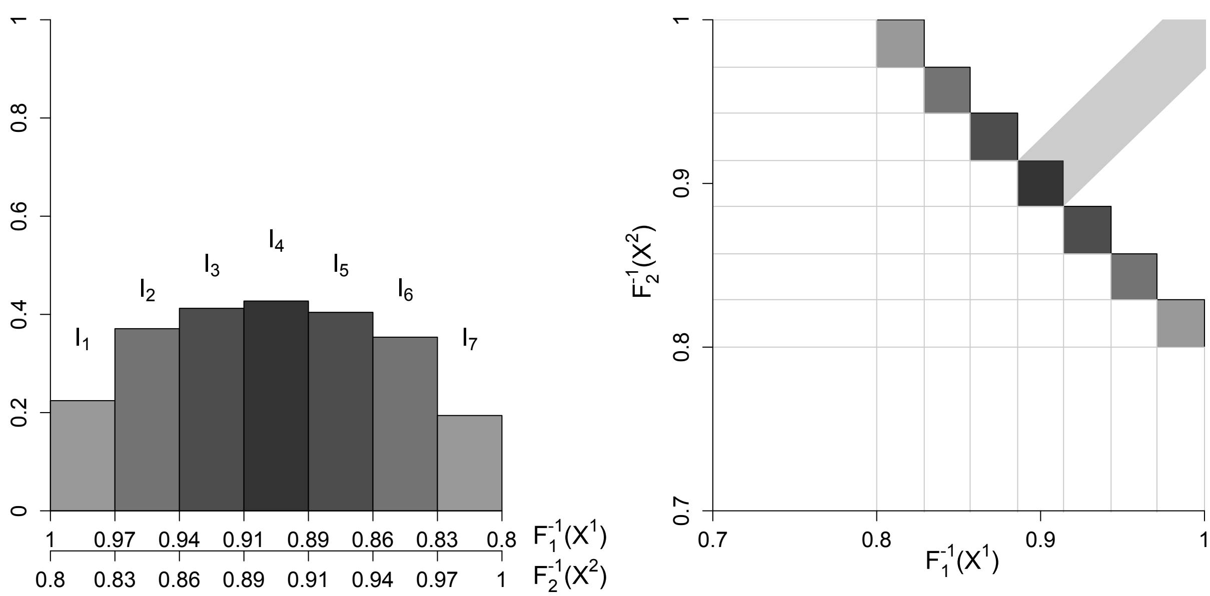


move also increases. The latter might compensate investors for facing extreme downside risk in large cross-sections. That is to say, when bivariate losses occur more frequently, one can also expect more bivariate extreme gains. In contrast, the recent financial crisis 2007-09 is characterized by a temporary bump in tail asymmetries which subtends a phase of steady decline of tail asymmetries since the mid 1990s. This increase in tail asymmetries is even more pronounced when implementing the test with GARCH pre-filtering and the standard multiplier bootstrap (see Appendix C for details). One might argue that, in contrast to former financial crises, only tail dependence between losses was affected. But tail dependence between gains did not experience such change. This makes the subprime crises particularly disastrous as investors did not encounter much extreme upside potential. However, aggregated tails of the market (Figure 3) hardly back this conclusion as we observe a nearly parallel progression of both upper and lower tail measures. Thus, by aggregating bivariate tails to an index measure, much information on the tail dependence between tails of the index' constituents is lost. 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. This implies tail measures for indices do not tell the same story their constituents can.

In comparison to the two competing tests, our test 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 test provides a more accurate assessment of tail asymmetry within the market and suggests tail asymmetry is more common than expected. 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, for Test (1) with $M = 14$, the adjusted p-value of the central subinterval does not suggest a rejection, while at least one non-central p-value does (solid line, Figure 5). This line is nearly parallel to the difference in found tail asymmetries between the TDC test and our test.

To further underline the importance of non-standard tail dependence structures, we quantify the number of tail asymmetric pairs that scalar approaches would miss due to *off-diagonal* tail asymmetries. In Figure 5 (right), for each period, we compare 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 focuses 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 – where not considering off-diagonal parts of the TC is especially serious. Yet, in the finance literature, e.g. Jondeau (2016), it is common practice to analyze tail dependence solely by the tail dependence coefficient ι, i.e. the tail copula along the diagonal where $x^{(1)} = x^{(2)}$. We document that this approach might overlook non-standard types of tail dependence leading to a substantial misconception of tail asymmetry. A more detailed picture of the local impact on tail asymmetry is provided in Figure 6 which marks rejection frequencies for specific quantile regions as discussed in Subsection 3.2. This could directly translated into investment and hedging strategies. Furthermore, the difference in found asymmetries between our test and BD13 suggests some degree of intra-tail asymmetry (ITA) among all pairs. The simulation study demonstrated that the power of the two tests differs mainly in intra-tail asymmetric cases. In Appendix D, we conduct formal tests for ITA. They confirm and quantify that periods of strong differences in the two test indeed coincide with the presence if ITA. Please Appendix D for details.

Figure 6: Localization of test rejections. Both figures show applications the localization idea of Figure 1. (Left) Histogram of rejection frequencies of specific quantile regions for $J = 14$, aggregated over every tail comparison over all periods. (Right) Test rejection frequencies translated to quantile regions in $[0, 1]^2$.



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 slightly higher power else. The test idea may also be applied for general copula comparisons, and also for tail dependence comparisons in higher dimensions. An empirical study of US stock market sectors and foreign exchange rates shows our test typically finds more asymmetries/inequalities than competing tests; we find time periods where our test clearly benefits from respecting non-diagonal TC differences, meaning our test detects substantially more opportunities to hedge tail risks.

ACKNOWLEDGMENTS

We would like to thank the Editor, the Associate Editor and two anonymous referees for many helpful comments that substantially improved the article. We are also grateful to Christian Conrad and Julia Schaumburg, participants of the HeiKaMEtrics Workshop 2015, CFE-ERCIM 2016, VfS 2017, and of seminars in Aarhus, CREATES and Essex.

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 and Finance*, 35(1), 130-141.

Ang, A., and J. Chen (2002), “Asymmetric correlations of equity portfolios,” *Journal of Financial Economics*, 63(3), 443-494.

Barras, L., O. Scaillet, and R. Wermers (2010), “False discoveries in mutual fund performance: Measuring luck in estimated alphas,” *Journal of Finance*, 65(1), 179-216.

Bajgrowicz, P., and O. Scaillet (2012), “Technical trading revisited: False discoveries,persistence tests, and transaction costs,” *Journal of Fin. Econ.*, 106, 473-491.

Benjamini, Y., 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.

Bodnar, T., and T. Dickhaus (2014), “False discovery rate control under Archimedean copula,” *Electronic Journal Statistics*, 8(2), 2207-2241.

Brunnermeier, M. K., and L. H. Pedersen (2009) “Market liquidity and funding liquidity,” *Review of Financial Studies*, 22(6), 2201-2238.

Bücher, A., and H. Dette (2013), “Multiplier bootstrap of tail copulas with applications,” *Bernoulli*, 19(5A), 1655-1687.

Bücher, A., and I. Kojadinovic (2016)“ A dependent multiplier bootstrap for the sequential empirical copula process under strong mixing”, *Bernoulli*, 22(2), 927-968.

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.

Bücher, A., J. Segers, and S. Volgushev (2017), “Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series” *Bernoulli.*

Chang, B. Y., P. Christoffersen, and K. Jacobs (2013), “Market skewness risk and the cross-section of stock returns,” *Journal of Financial Economics*, 107(1), 46-68.

Chollete, L., V. de la Peña, and C.-C. Lu) (2011), “International diversification: A copula approach,” *Journal of Banking and Finance*, 35, 403-417.

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

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

Drees, H. (2003), “Extreme Quantile Estimation for Dependent Data with Applications to Finance”. *Bernoulli* 9, 617-657

Einmahl, J. H. J., A. Krajina, and J. Segers 2012), "An M-estimator for tail dependence in arbitrary dimensions," *The Annals of Statistics*, 40(3), 1764-1793.

Embrechts, P. (2009), "Linear correlation and EVT: Properties and caveats," *Journal of Financial Econometrics*, 7(1), 30-39.

Engle, R. F., and K. Sheppard (2001), "Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH," NBER Working paper, No 8554.

Fama, E. F., and K. R. French (1994), "Industry costs of equity," *Journal of Financial Economics*, 43, 153-193.

Garcia, R., and G. Tsafack (2011), "Dependence structure and extreme comovements in international equity and bond markets," *Journal of Banking and Finance*, 35(8), 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. (2016), "Asymmetry in tail dependence in equity portfolios," *Computational Statistics and Data Analysis*, 100, 351-368.

Kojadinovic, I. (2015), "npcp: Some nonparametric tests for change-point detection in possibly multivariate observations," 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(1), 169-187.

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

McNeil, A. J., and R. Frey (2000), "Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach," *Journal of Empirical Finance*, 7(3-4), 271-300.

McNeil, A. J., R. Frey, and P. Embrechts (2005) *Quantitative risk management: Concepts, techniques and tools*, Princeton, NJ: Princeton University Press.

Mikosch, T. (2006), "Copulas: Tales and facts," *Extremes*, 9(1), 3-20.

Oh, D. H., and A. J. Patton (2017), "Modelling dependence in high dimensions with factor copulas," *Journal of Business & Economic Statistics*, 35(1), 139-154.

Patton, A. J. (2006), "Modelling asymmetric exchange rate dependence," *International Economic Review*, 47(2), 527-556.

Patton, A. J. (2013), "Copula methods for forecasting multivariate time series," In *Handbook of Economic Forecasting*, 2(B): 899-960.

Rémillard, B. and O. Scaillet (2009), "Testing for equality between two copulas," *Journal of Multivariate Analysis*, 100, 377-386.

Romano, J. P., and M. Wolf (2005), "Stepwise multiple testing as formalized data snooping," *Econometrica*, 73(4), 1237-1282.

Schmidt, R. (2002), "Tail Dependence for Elliptically Contoured Distributions," *Mathematical Methods of Operations Research*, 55(2), 301-327.

Schmidt, R., and U. Stadtmüller (2006), "Nonparametric estimation of tail dependence," *The Scandinavian Journal of Statistics*, 33(2), 307-335.

Stephenson, J., and J. Tawn (2005), "Exploiting occurrence times in likelihood inference for componentwise maxima," *Biometrika* 92(1), 213-227.

Straetmans, S., W. Verschoor, and C. Wolff (2008), “Extreme US stock market fluctuations in the wake of 9/11,” *Journal of Applied Econometrics*, 23(1),17-42.

Tawn, J. A. (1988), "Bivariate extreme value theory: Models and estimation," *Biometrika*, 75(3), 397-415.