Beyond dimension two: A test for higher-order tail risk

by Carsten Bormann, Julia Schaumburg, Melanie Schienle

No. 80 | JANUARY 2016

Beyond dimension two: A test for higher-order tail risk

Carsten Bormann

Karlsruhe Institute of Technology, Germany

Melanie Schienle

Karlsruhe Institute of Technology, Germany

Julia Schaumburg[*]

VU University Amsterdam and Tinbergen Institute, The Netherlands

Abstract

In practice, multivariate dependencies between extreme risks are often only assessed in a pairwise way. We propose a test for detecting situations when such pairwise measures are inadequate and give incomplete results. This occurs when a significant portion of the multivariate dependence structure in the tails is of higher dimension than two. Our test statistic is based on a decomposition of the stable tail dependence function describing multivariate tail dependence. The asymptotic properties of the test are provided and a bootstrap based finite sample version of the test is proposed. A simulation study documents good size and power properties of the test including settings with time-series components and factor models. In an application to stock indices for non-crisis times, pairwise tail models seem appropriate for global markets while the test finds them not admissible for the tightly interconnected European market. From 2007/08 on, however, higher order dependencies generally increase and require a multivariate tail model in all cases.

Keywords: decomposition of multivariate tail dependence, multivariate extreme values, stable tail dependence function, extreme dependence modeling

JEL classification: C01, C46, C58

1   Introduction

Studying extreme co-movements in multidimensional systems is a key concern in finance and insurance. However, tail dependence structures of multivariate distributions are mostly treated in bivariate setups, see for instance Poon et al. (2004) and Klugman & Parsa (1999), but also Straetmans et al. (2008), Li (2013), Rodriguez (2007), among many others. Pairwise simplification is not only standard when analyzing financial systems but is also widely used for studying extreme environmental and weather risks (see de Haan & de Ronde (1998) and Ghosh (2010)). This is due to the fact that in practice, bivariate models are more easily tractable and computationally more appealing. But also from a theoretical point of view, statistical properties of a large group of estimators are only known up to dimension two (Coles et al. (1991), Joe et al. (1991), de Haan et al. (2008), Guillotte et al. (2011)). Yet, for a variety of empirical settings, there are periods in time during which a pairwise approach is too restrictive, as joint extremes occur in cross–sections of dimension three or higher. In particular during the recent financial crisis, markets became increasingly dependent. The financial contagion literature provides a lot of evidence that the major part of this rising interconnectedness was due to complex higher order interdependencies, which could not have been detected by standard pairwise tail dependence

measures (see, e.g., Longstaff (2010), Brunnermeier & Pedersen (2009)). In such situations, the most common bivariate measures for tail dependence, such as the tail dependence coefficient (see Straetmans et al. (2008), Poon et al. (2004), Hartmann et al. (2004)), bivariate copulas (see, e.g. Li (2013), Rodriguez (2007), and references therein), or simple product moment correlation coefficients and correlation matrices fail to explicitly account for a large amount of the complex dependence structure among extreme risks in the system. This leads to severe underestimations of the effects of extreme comovements. For a discussion of the limitations of common bivariate measures of dependence, see also Embrechts (2009) and Mikosch (2006).

We propose a test that indicates whether pairwise modeling of multivariate tail dependence of a d–dimensional random vector $\mathbb{X} = (X^{(1)}, ..., X^{(d)})'$ with $d > 2$ is adequate, or whether it implies significantly different and thus incomplete tail dependence structures. The test is based on the stable tail dependence function (STDF), which was first introduced in Huang (1992) (see also de Haan & Ferreira (2006) and Einmahl et al. (2012)). The STDF maps the univariate tails of a random vector to their joint limit distribution, and therefore completely describes their extremal dependence structure. It is a general and flexible concept of tail dependence and allows for straightforward non-parametric estimation, bearing a smaller risk of model misspecification than alternative parametric approaches. Furthermore, its statistical properties are well understood for $\mathbb{X}$ of dimension beyond two (Einmahl et al. (2012), Bücher et al. (2014)). Moreover, its rather conservative definition of multivariate extreme events fits the needs of (financial) risk management (Segers (2012)).

The main idea of the test is to decompose the STDF for $\mathbb{X}$ into probabilities of univariate extreme events, the STDFs of all possible bivariate pairs within $\mathbb{X}$, and a remainder term capturing extreme events in dimensions three to d. We refer to the latter as higher order tail dependencies (HOTDs), and denote tail events as multivariate when they comprise three or more extremes in the cross-section. If an estimate of the remainder term is not significantly different from zero, we conclude that tail dependence in dimension d can be captured sufficiently well by analyzing only bivariate tails. However, if we reject the null hypothesis that HOTDs have no influence, ignoring high-dimensional joint extreme events leads to underestimation of the actual tail risk dependence, which is then driven by a substantial portion of joint extremes in dimension three and higher. The asymptotic properties of the test statistic are derived and a bootstrap implementation scheme for finite samples is proposed. Simulation studies with standard multivariate risk structures for the *iid* and ARMA–GARCH cases document good size and power properties of the test in finite samples. Moreover, our simulations highlight the need to filter the data for conditional heteroskedasticity before applying the test to financial time series.

Our empirical application deals with the influence of HOTDs in international stock markets. Asset allocation and portfolio diversification, as well as systemic risk assessment require a most accurate picture of tail dependencies between financial markets. Univariate tail losses within a portfolio can be diversified by holding tail independent assets. Bivariate tail dependence eliminates such tail risk diversification opportunities between two assets, as large losses tend to occur simultaneously. The same reasoning applies to higher–dimensional tail risk: Whenever extreme losses of three or more assets coincide, multivariate tail risk cannot be diversified anymore. Ang & Chen (2002), Poon et al. (2004), Chollete et al. (2011) and others estimate bivariate tail measures for indices of international stock markets. The common conclusion is that left bivariate tails, i.e. extreme losses, are dependent, especially intra-continentally. Right tails, however, tend to be independent. We test for HOTDs within two separate sets of stock market indices. In a global portfolio including US, Asian-Pacific and European stock indices, we find no evidence for HOTDs in both left and right tails, until the rise of the financial crisis of 2007. This finding suggests that global tail diversification possibilities are limited ever since, a finding that has also been made by Christoffersen et al. (2012) using a dynamic copula approach. Testing against HOTDs in a multi-country European portfolio, we find strong evidence for HOTDs during the last decades, which can only partly be explained by serial correlation, time variation, and a factor reflecting the development of global markets. Our results therefore contribute to the empirical international finance literature in three points: First, we find that the extent of intra-European tail dependence is more severe than discovered in former contributions. Second, higher-order tail effects in European markets are time-varying, and have increased during the recent financial crisis. Third, multivariate effects in extreme losses on the global level become relevant in the course of the financial crisis, while extreme gains are largely not affected by HOTDs. We conclude our empirical application by quantifying the share of HOTDs in tail dependence. We find time periods in which up to 70% of all bivariate extreme events are in fact multivariate. Also, in recent years, this share has doubled for losses and even tripled for gains on the European portfolio.

The rest of the paper is organized as follows. Section 2.1 discusses necessary concepts from multivariate extreme value theory. Section 2.2 introduces and formalizes test idea, test asymptotics and finite sample implementation. Finite sample properties are studied in Section 3. Section 4 studies HOTDs between international stock indices. Section 5 concludes. The Appendix contains supplementary and theoretical results.

2 Econometric methodology

2.1 Multivariate dependence in extreme tails

For our analysis of extreme risks, we use techniques from multivariate extreme value theory which we introduce and motivate in the following. Denote by $\mathbb{X} := (X^{(1)}, ..., X^{(d)})'$ a d–dimensional random vector with continuous joint cumulative distribution function (CDF) $F_{\mathbb{X}}(\mathbf{x}), \mathbf{x} := (x^{(1)}, ..., x^{(d)})$. Its univariate marginal CDFs are denoted by $F_j(x^{(j)}), j = 1, ..., d$. Suppose we observe a sample of n *iid* draws from the random vector $\mathbb{X}$, collected in the $(n \times d)$ sample matrix $\mathbf{X} = (\mathbf{X}_n^{(1)}, ..., \mathbf{X}_n^{(d)})$ with $\mathbf{X}_n^{(j)} = (X_1^{(j)}, ..., X_n^{(j)})', \ j = 1, ..., d$. We write $\max(\mathbf{X}_n^{(j)}) = \max(X_1^{(j)}, ..., X_n^{(j)})$ for the sample maximum of margin j. For each marginal, we assume that there exist normalizing constants $a_n^{(j)} \in \mathbb{R}_+, b_n^{(j)} \in \mathbb{R}, j = 1, ..., d$, and a limiting distribution $G_{\mathbb{X}}(\mathbf{x})$, such that

$$\lim_{n \to \infty} \mathbb{P}\left(\frac{\max(\mathbf{X}_n^{(1)}) - b_n^{(1)}}{a_n^{(1)}} \leq x^{(1)}, ..., \frac{\max(\mathbf{X}_n^{(d)}) - b_n^{(d)}}{a_n^{(d)}} \leq x^{(d)} \right) = G_{\mathbb{X}}(\mathbf{x}), \tag{1}$$

for all continuity points of $G_{\mathbb{X}}(\mathbf{x})$. Then, $G_{\mathbb{X}}(\mathbf{x})$ is a multivariate extreme value distribution, and $F_{\mathbb{X}}(\mathbf{x})$ is said to be in the domain of attraction of $G_{\mathbb{X}}(\mathbf{x})$, which is denoted by $F_{\mathbb{X}} \in D(G_{\mathbb{X}})$, see de Haan & Ferreira (2006) and Resnick (1987). Necessary and sufficient conditions for $F_{\mathbb{X}} \in D(G_{\mathbb{X}})$ can be found in de Haan & Resnick (1977), Beirlant et al. (2004, p.287), de Haan & Ferreira (2006), and Resnick (1987). Throughout the paper, we assume that they are fulfilled. In general, closed-form expressions for $G_{\mathbb{X}}(\mathbf{x})$ do not exist. Equation (1) can be written as

$$\lim_{n \to \infty} F_{\mathbb{X}}^n(a_n^{(1)} x^{(1)} + b_n^{(1)}, ..., a_n^{(d)} x^{(d)} + b_n^{(d)}) = G_{\mathbb{X}}(\mathbf{x}), \tag{2}$$

implying that the univariate marginals converge individually to one–dimensional extreme value distributions $G_j(x^{(j)})$, which have the standard Fisher–Tippett form

$$\lim_{n \to \infty} F_j^n(a_n^{(j)} x^{(j)} + b_n^{(j)}) = G_j(x^{(j)}) = \exp\left(-(1 + \gamma_j x^{(j)})^{-1/\gamma_j} \right), j = 1, ..., d, \tag{3}$$

where γ_j denotes the tail index of margin j (Fréchet (1927), Fisher & Tippett (1928), Gnedenko (1943)). An equivalent formulation of relation (2), with (3) holding true for all margins, is given by the concept of the stable tail dependence function (STDF) of $\mathbb{X}$, denoted by $\ell_{\mathbb{X}}(\mathbf{x})$ or $\ell(\mathbf{x})$ (Huang (1992), Einmahl et al. (2012)). Equivalent characterizations of $G_{\mathbb{X}}(\mathbf{x})$, and thus $\ell(\mathbf{x})$, can be obtained via the spectral measure and the exponent measure (de Haan & Ferreira 2006, Chapter 6) but are less intuitive in interpretation and decomposition. The STDF $\ell(\mathbf{x}) : \mathbb{R}^d \mapsto \mathbb{R}_+$

is defined as

$$\ell(\mathbf{x}) = -\log G_{\mathbb{X}}\left(\frac{x^{(1)-\gamma_1} - 1}{\gamma_1}, ..., \frac{x^{(d)-\gamma_d} - 1}{\gamma_d}\right).$$

The STDF describes the complete dependence structure of the tails of the univariate marginals. As one can express $\ell(\mathbf{x})$ as

$$\ell(\mathbf{x}) = \lim_{t \to 0} t^{-1}\mathbb{P}\Big(\bigcup_{i=1}^{d}\{F_i^{-1}(1 - tx^{(i)}) \leq X^{(i)}\}\Big), t \in \mathbb{R}_+, \tag{4}$$

the stable tail dependence function (STDF) is an asymptotic measure which can be interpreted as the scaled asymptotic probability that at least one element of $\mathbb{X}$ exceeds an extreme quantile, that is, $X^{(i)}$ exceeds $F_i^{-1}(1 - tx^{(i)})$, as $t \to 0$. From this representation, a direct non-parametric estimate of the STDF can be derived. Also, $\ell(\mathbf{x})$ can be decomposed into component STDFs of dimensions lower than d.

There is a rich statistical literature on general properties of the STDF and its estimators (e.g. Huang (1992), Dietrich et al. (2003), Einmahl et al. (2006), Drees et al. (2006), Einmahl et al. (2012), Bücher et al. (2014)). Importantly, the STDF is a convex function and homogeneous of degree one, i.e. $\ell(\lambda\mathbf{x}) = \lambda\ell(\mathbf{x})$ for $\lambda \in \mathbb{R}$. Moreover, $\ell(\mathbf{x}) \in [\max(\mathbf{x}), \mathbf{x}'\mathbf{1} = \sum_{i=1}^{d} x^{(i)}]$ with $\mathbf{1}$ representing a d-vector of ones. The lower (upper) bound is attained if $\mathbb{X}$ is perfectly tail dependent (independent), that is, extremes of univariate marginals always (never) occur simultaneously (Beirlant et al. (2004), de Haan & Ferreira (2006)). Tail (in)dependence is often also denoted as asymptotic (in)dependence. Numerical values of $\ell(\mathbf{x})$ close to $\max(\mathbf{x})$ indicate that tails of $\mathbb{X}$ are strongly interconnected. Values of $\ell(\mathbf{x})$ close to $\mathbf{x}'\mathbf{1}$ mark the opposite. In practice, perfect tail dependence is rare.

It is important to note the connection, but also the difference, of the STDF to the so–called tail copula (TC) which is a closely related metric for tail dependence. The TC of $\mathbb{X}$ is defined as

$$R(\mathbf{x}) = \lim_{t \to 0} t^{-1}\mathbb{P}\Big(\bigcap_{i=1}^{d}\{F_i^{-1}(1 - tx^{(i)}) \leq X^{(i)}\}\Big).$$

O It only considers joint exceedances to characterize tail dependence, see Schmidt & Stadtmüller (2006). Sibuya (1960), Joe (1997) and Coles et al. (1998) analyze bivariate tail dependence by means of the tail dependence coefficient, which corresponds to the bivariate TC at the point $\mathbf{x} = (1, 1)$. Roughly speaking, it describes the tendency of two random variables to jointly exceed a high threshold. In two dimensions, there is a one-to-one mapping between the TC and the STDF. Due to the lack of natural ordering in higher dimensions, however, the definition of a

multivariate extreme event depends on the research objective. There are several reasons why we prefer the STDF over the TC for our purpose: Firstly, the TC captures only (the most extreme) parts of the multivariate tail dependence in dimensions $d > 2$, while the STDF completely describes it (see Subsection 2.2 for the relationship between the two). Secondly, a practical issue for large d is that joint d–dimensional exceedances are rarely observed in finite samples. Unless a sample contains an observation with all marginals being extreme, the TC indicates tail independence. That is, the TC only considers the most extreme events when all marginals are simultaneously extreme, and disregards more likely tail events. On the other hand, the STDF incorporates events in which a single component of $\mathbb{X}$ becomes extreme, and hence finite samples provide more relevant observations. Segers (2012) interpret $\ell(\mathbf{x})$ as "trouble in the air", whereas $R(\mathbf{x})$ only considers events as extreme when "the sky is falling". The STDF is therefore an important ingredient for a conservative risk monitoring approach, in the sense that not only the "most extreme" extremes are considered.

2.2 A new test for higher-order tail dependence

We aim to detect the share which HOTDs contribute to overall tail dependence. Hence, we decompose the STDF for dimension d into TCs for dimensions two to d. In dimension $d = 2$, from equation (4) we have that $\ell(\mathbf{x})$ is the limiting probability of a union of two events; since $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ for events A and B. Therefore, we have $\ell(x^{(1)}, x^{(2)}) = x^{(1)} + x^{(2)} - R(x^{(1)}, x^{(2)})$. For similar decompositions in arbitrary dimension $2 < d < \infty$, additional notation is required. For $I \subset \{1, ..., d\}$ define the subvectors $\mathbb{X}^{(I)} := (X^{(i)})_{i \in I}, \mathbf{x}^{(I)} := (x^{(i)})_{i \in I}$, and according STDFs as $\ell_I(\mathbf{x}^{(I)})$. Then, in $\mathbb{R}^{2 < d < \infty}$, using the inclusion–exclusion principle, we have

$$\ell(\mathbf{x}) = \sum_{i=1}^{d} x^{(i)} - \sum_{i<j\leq d} R_{ij}(\mathbf{x}^{(i,j)}) + \underbrace{\sum_{h<i<j\leq d} R_{hij}(\mathbf{x}^{(h,i,j)}) - ... + (-1)^{d+1} R(\mathbf{x})}_{=:\mathcal{A}}, \qquad (5)$$

where $\mathcal{A}$ denotes the portion HOTDs contribute to "global" tail dependence in $\mathbb{X}$, that is, the tail dependence of the entire random vector $\mathbb{X}$. Provided that global tail dependence is only caused by bivariate extreme events, i.e. by the first two terms of equation (5), $\mathcal{A}$ equals zero. In this case, higher dimensional joint extremes are irrelevant. When substituting $R_{ij}(x^{(i)}, x^{(j)}) = x^{(i)} + x^{(j)} - \ell_{ij}(x^{(i)}, x^{(j)}), i < j \leq d$, equation (5) yields

$$\ell(\mathbf{x}) = (2 - d) \sum_{i=1}^{d} x^{(i)} + \sum_{i<j\leq d} \ell_{ij}(\mathbf{x}^{(i,j)}) + \mathcal{A}, \qquad (6)$$

which decomposes global tail dependence into asymptotic probabilities for univariate extremes and STDFs for any bivariate combination and HOTDs.

Using equation (6) we can test whether extreme events in dimensions larger than two have a statistically significant impact, that is, if two–dimensional tails explain tail dependence in dimension $d > 2$ sufficiently well. Formally, if $\mathcal{A} = 0$, we have

$$\Delta := \ell(\mathbf{x}) - (2 - d) \sum_{i=1}^{d} x^{(i)} - \sum_{i<j\leq d} \ell_{ij}(\mathbf{x}^{(i,j)}) = 0. \tag{7}$$

In this case, bivariate extreme relations are sufficient for capturing the full global tail dependence. Hence, the null hypothesis that the impact of higher–order tail dependencies is negligible can be formulated as

$$H_0 : \Delta = 0. \tag{8}$$

If Δ substantially deviates from zero, the null is rejected. With $\mathbf{x} = \mathbf{1}$, it is possible to show that $\Delta \in [0, \sum_{i=1}^{d-2} i], d > 2$.

The following proposition clarifies that testing for $\Delta = 0$ is not equivalent to testing whether $\mathbb{X}$ is tail independent. Thus, multivariate distributions exist which are globally tail dependent but have $\Delta = 0$. Hence their global tail dependence is exclusively caused by bivariate tails. A test for tail independence is proposed in Draisma et al. (2004).

Proposition 2.1

If $\mathbb{X}$ is tail independent, that is if all bivariate tails of $\mathbb{X}$ are tail independent, then $\Delta = 0$. The reverse does not hold.

This can, e.g., be easily shown for the family of distributions which we use in the simulation setting in Section 3.

In order to apply the test, we have to estimate the STDF of $\mathbb{X}$, $\ell_{\mathbb{X}}(\mathbf{x})$, and the STDFs for bivariate pairs. Let $X_{n:m}^{(i)}$ denote the m–th largest order statistic of margin $X^{(i)}$, and let $1(C)$ be the indicator function for event C. In equation (4), replacing the running variable t by k/n and the extreme quantiles $F_i^{-1}(1 - tx^{(i)})$ by $X_{n:n+0.5-kx^{(i)}}^{(i)}$ we use the following non-parametric estimator for the STDF (see Huang (1992) and Einmahl et al. (2012))

$$\hat{\ell}(\mathbf{x}) = \frac{1}{k} \sum_{i=1}^{n} 1\Big\{ \bigcup_{j=1}^{d} \{X_i^{(j)} \geq X_{n:n+0.5-kx^{(j)}}^{(j)}\} \Big\}, n \to \infty, k \to \infty, \frac{k}{n} \to 0, \tag{9}$$

$\mathbf{x} = (x^{(1)}, ..., x^{(d)})$. Under some technical conditions, the empirical process $\sqrt{k}(\hat{\ell}(\mathbf{x}) - \ell(\mathbf{x}))$ converges to a sum of a centered Gaussian field and univariate centered Gaussian processes with

given covariance structure (Einmahl et al. (2012), Bücher et al. (2014)). If $\mathbb{X}$ is asymptotically independent, $\hat{\ell}(\mathbf{x})$ is still asymptotically normal but with degenerate variance (Hüsler & Li 2009). Note, $\hat{\ell}(\mathbf{x})$ is invariant against monotone transformations. For simplicity, we fix $\mathbf{x} = \mathbf{1}$, which is standard in the applied extreme value literature, see e.g. Hartmann et al. (2004). In this case, for each marginal, the threshold equals $X_{n:n+1/2-k}^{(i)}$. The asymptotic distribution of $\hat{\ell}(\mathbf{1})$ simplifies to

$$\sqrt{k}\left(\hat{\ell}(\mathbf{1}) - \ell(\mathbf{1})\right) \xrightarrow{d} N(0, \sigma_{\hat{\ell}}^2),$$

where closed form expressions of $\sigma_{\hat{\ell}}^2$ can be reconstructed from theorem 4.6 in Einmahl et al. (2012). Plugging $\hat{\ell}(\mathbf{1})$ into Δ yields the empirical test statistic

$$\widehat{\Delta} := \hat{\ell}(\mathbf{1}) - 2d + d^2 - \sum_{i<j\leq d} \hat{\ell}_{ij}(\mathbf{1}). \tag{10}$$

These considerations lead us to the asymptotic distribution of the test statistic, which is given next.

Proposition 2.2 *Assume $F_{\mathbb{X}} \in D(G_{\mathbb{X}})$. Furthermore, let the following assumptions hold:*

(A1) There exists a constant $\beta > 0$ such that for $t \downarrow 0$ it holds that $t^{-1}\mathbb{P}(\bigcup_{i=1}^d F_i^{-1}(1 - tx^{(i)}) \leq X^{(i)}) = \ell(\mathbf{x}) + \mathcal{O}(t^\beta)$ uniformly on the unit simplex in $\mathbb{R}^d$.

(A2) The threshold parameter $k \to \infty$ for $n \to \infty$ with $k = o(n^{2\beta}/(1 + 2\beta))$ with β from (A1).

Then,

$$\sqrt{k}(\widehat{\Delta} - \Delta) \xrightarrow{d} N(0, \sigma_{\widehat{\Delta}}^2), \tag{11}$$

where $\sigma_{\widehat{\Delta}}^2$ is the sum of all entries of the covariance matrix of

$$\left(\hat{\ell}(\mathbf{1}), (\hat{\ell}_{ij}(1,1))_{i<j\leq d}\right).$$

For the proof, please see the Appendix. Assumption (A1) imposes that $t^{-1}\mathbb{P}(\bigcup_{i=1}^d F_i^{-1}(1 - tx^{(i)}) \leq X^{(i)})$ exists for t small and converges to the STDF at a certain speed. This second-order condition refines the base assumption of max-domain attraction of $F_{\mathbb{X}}$. The second assumption restricts the speed with which k grows to infinity, and in combination with (A1) guarantees that an asymptotic bias term for the left hand side of equation (11) vanishes (see Resnick & de Haan (1996), Einmahl et al. (2008) for details). According to Bücher et al. (2014) a smoothness assumption for the STDF is not required. In particular, we do not need to impose that partial derivatives of ℓ exist for the asymptotic result to hold. Such an assumption might be too rigid, as it would, e.g., exclude factor models, which are practically important in financial applications.

For obtaining Proposition 2.2, we therefore rely on asymptotic results by Bücher et al. (2014), which do not require the existence of partial derivatives of the STDF, but which are also no longer uniform but yield convergence of $\hat{\ell}(\mathbf{x})$ in a weaker sense.[1]

In both the simulation study and the empirical application in Sections 3 and 4, we restrict the test to dimension 7. However, if $\mathbb{X}$ exhibits tail dependence in dimension d larger than 7, it necessarily exhibits tail dependence in dimensions $3 \leq g < d$. Thus, the asymptotic power of the test also increases with larger dimensions. Subsection 3.2 further discusses these details, in the context of the results on the empirical power in the simulation settings. Also, the test can be readily adapted to detect whether joint extremes of dimension $3 < g \leq d$ are significant.

2.3 Finite sample version of the test

Although it is possible to derive the explicit form and calculate empirical versions of the asymptotic variance of the test statistic, a bootstrap version is practically superior. The reason is that bootstrapping $\sigma_{\hat{\Delta}}^2$ works under milder conditions, in particular if $\mathbb{X}$ exhibits asymptotic dependence (Bücher & Dette 2013). In contrast, direct estimation of $\sigma_{\hat{\Delta}}^2$ may require the estimation of partial derivatives of the STDF and of covariances between the different STDFs. In principle, a weighted least squares based estimator for such partial derivatives of the STDF exists, but its statistical properties have only been established for dimension $d = 2$ so far (see Peng & Qi 2006). Furthermore, smoothness assumptions for the STDF might not be met. In such cases, estimating the partial derivatives is not admissible (Bücher & Dette 2013).

As our goal is to bootstrap extremal observations, we do not resample from the full sample, but only from a subsample (Politis & Romano (1994)). Otherwise, an asymptotically vanishing bias term of $\widehat{\Delta}$, inherited from $\hat{\ell}_{\mathbb{X}}$ (see Huang (1992)), might distort the bootstrap distribution. Peng (2010) propose a similar approach and successfully employ a subsample size of $n^{0.95}$. Qi (2008), El-Nouty & Guillou (2000), Danielsson et al. (2001), Geluk & de Haan (2002) generally document the benefits of subsampling for pointwise extreme value statistics. We construct rejection regions for the test from the asymptotic normal distribution of $\widehat{\Delta}$ with the resampled form of the variance. We explicitly mark if an estimator $\hat{\theta}$ depends on the threshold parameter k by writing $\hat{\theta}(k)$. In summary, we proceed along the following six steps for obtaining a test decision:

(1) Choose the threshold parameter, denoted by k^*, for $\widehat{\Delta}$ from the sample $\mathbf{X}$.

(2) Calculate $\hat{\ell}(k^*)$, and any $\hat{\ell}_i(k^*), i \in \mathcal{I}_{(2)}^{(d)}$, to determine the full sample test statistic $\widehat{\Delta}(k^*)$ from $\mathbf{X}$.

---

[1] In particular, Einmahl et al. (2012) show weak convergence of the empirical process $\sqrt{k}(\hat{\ell}(\mathbf{x}) - \ell(\mathbf{x}))$ for bounded functions in the sup-norm, while Bücher et al. (2014) show convergence for locally bounded functions in the so-called hypi-semimetric.

(3) Draw at least $B = 500$ bootstrap samples with replacement from $\mathbf{X}$ with sample size $n^* = n^{0.95}$ and denote the resulting bootstrap samples by $\mathbf{X}_1^*, ..., \mathbf{X}_B^*$.

(4) For $j = 1, ..., B$, estimate $\widehat{\Delta}(k^*)$ from the bootstrap samples $\mathbf{X}_1^*, ..., \mathbf{X}_B^*$, yielding B bootstrapped estimates $\widehat{\Delta}(k^*)_1, ..., \widehat{\Delta}(k^*)_B$.

(5) Estimate $\sigma_{\widehat{\Delta}}^2$ from the bootstrapped estimates in the previous step by its empirical analogue.

(6) On a $1 - \alpha$ confidence level reject $H_0 : \Delta = 0$ if $0 < \widehat{\Delta}(k^*) + z^\alpha \hat{\sigma}_{\widehat{\Delta}(k^*)}$, where z^α denotes the α quantile of the standard normal distribution.[2]

A theoretically optimal, data driven choice of the threshold parameter k should balance the bias–variance trade–off that is inherent in the estimation of $\ell(\mathbf{x})$. Finding such a solution and deriving its optimality properties is non-standard even in the univariate case and is thus beyond the scope of this paper. In our simulations we choose k randomly from an interval in order to minimize possible distortions from a poorly chosen k. In the application, we estimate Δ over a grid of different values for k and calculate the median over this set of estimates.[3] Further details can be found in the respective sections. For alternative, purely data-driven procedures for determining k in a univariate setup, we refer to Frahm et al. (2005) and Schmidt & Stadtmüller (2006).

For time series data, issues of short-range serial dependence can be addressed by implementing a blocked version of the bootstrap providing appropriate up to second moment adjustments, see, e.g., Straetmans et al. (2008) with an asymptotically optimal choice of block length of order $n^{1/3}$ according to Hall et al. (1995). Instead, however, we use appropriate GARCH-type filtered observations before applying the test. With this we also control for and amend higher order moment effects and volatility clustering of heteroskedastic financial data (McNeil & Frey (2000), Poon et al. (2004)). See Section 3.3. for details.

3 Simulation study

3.1 Size and power

In this subsection, we evaluate the empirical size and power of the test in finite samples in an *iid* setting. Results for time series data are presented in Subsection 3.3. We simulate from two types of distribution families with various subspecifications, for which we know whether the null of no significant HOTDs is true. In particular, we focus on the class of meta t–distributions and

---

[2] Note, a normal approximation for $\widehat{\Delta}$ is theoretically not justified under tail independence, i.e. $\ell(\mathbf{x}) = \sum_{i=1}^d x^{(i)}$; then, it holds $\sigma_{\hat{\ell}}^2 = \sigma_{\widehat{\Delta}}^2 = 0$ and the distributions of both $\hat{\ell}$ and $\widehat{\Delta}$ are degenerate, while the theoretical Δ is zero, i.e. the null is true. However, in such situations, the test typically indicates the correct decision not to reject the null.

[3] Specifically, for a sample size of 750, $k \in \{8, 9, ..., 48\}$ in dimension three and $k \in \{8, 9, ..., 30\}$ in dimension 7.

max factor models, which are both commonly used in financial risk management (McNeil et al. (2006), Fama & French (1992)). The meta t–distribution is a generalization of the multivariate t–distribution and the t–copula, and max factor models have the same tail dependence structure as factor models (Einmahl et al. (2012)). We employ the finite sample version of the test introduced in Subsection 2.3. All simulations are repeated $S = 500$ times.

Model dimensions are $d \in \{3, 5, 7\}$. For a power analysis, considering larger dimensions is often not necessary, as higher order tail dependencies of moderate order are sufficient for concluding that HOTDs are significant.

Let $C_{\nu,P}^t(\mathbf{x})$ denote the t–copula with ν degrees of freedom, and dispersion matrix P. Following Demarta & McNeil (2005),

$$C_{\nu,P}^t(\mathbf{u}) = \int_{-\infty}^{t_\nu^{-1}(u^{(1)})} \cdots \int_{-\infty}^{t_\nu^{-1}(u^{(d)})} \frac{\Gamma((\nu+d)/2)}{\Gamma(\nu/2)\sqrt{(\pi\nu)^d}|P|} (1 + \nu^{-1}(\mathbf{x}'P^{-1}\mathbf{x}))^{-(\nu+d)/2} d\mathbf{x}, \qquad (12)$$

where $t_\nu^{-1}(u^{(i)})$ denotes the quantile transform of a t-distribution with ν degrees of freedom for margin i, and $\Gamma(\cdot)$ is the gamma function. According to Hua & Joe (2011), the t-copula is of second-order regular variation and thus fulfills the assumptions of Proposition 2.2. In contrast to a classical t–copula, meta t–distributions allow the degrees of freedom of marginals $\nu_m^{(i)}$ to differ from the degrees of freedom of the copula, denoted by ν_C. For the simulation, we choose $\nu_C \in \{5, 10, 15, 20\}$, $\nu_m := \nu_m^{(i)} = 5, i = 1, ..., d$, and $P = (0.5)_{i \neq j}, P_{ii} = 1$. Thus, we consider equicorrelated t–distributions with common degrees of freedom ν_m that are linked by the t–copula with ν_C degrees of freedom. Exploiting results from Demarta & McNeil (2005) and Nikoloulopoulos et al. (2009, theorem 2.3), it is possible to show, that for the classical multivariate t–distribution the theoretical values of our test statistic Δ are larger than zero as the t-copula is capable of producing joint extremes in dimension $d > 2$; Δ increases if the degrees of freedom of the copula decreases, and/or if pairwise correlation increases. It equals zero if the correlation parameter equals -1. A meta t–distribution comprises the widely used multivariate t–distribution whenever $\nu_C = \nu_m$.[4]

In finance, often factor models are applied, in which asset returns $X^{(j)}$ depend on common factors $Z^{(i)}$ in a linear fashion. Max factor models assume $X^{(j)}$ can be modeled as the maximum of the factors times a parameter a_{mj}, the so called factor loadings. Both models have the same tail dependence structure (Einmahl et al. (2012)). Let $\mathbb{Z} := (Z^{(1)}, ..., Z^{(r)})'$ be a random vector of independent Fréchet random variables ($\nu = 1$). A d-dimensional max factor model for

---

[4]Theoretically, the dependence structure is only governed by the parametrization of the copula and not by distributional properties of the univariate tails, i.e. ν_m. In additional simulations that are not reported here, we found finite sample properties of the test are robust against changing the marginal degrees of freedom.

$\mathbb{X} = (X^{(1)}, ..., X^{(d)})'$ is then defined by

$$X^{(j)} := \max(a_{1j}Z_i^{(1)}, ..., a_{rj}Z_i^{(r)}), j = 1, ..., d,$$

with $\sum_{j=1}^{d} a_{mj} = 1, a_{mj} \geq 0$. The loading matrix $B_d^A := (a_{mj})$ governs the dependence between the tails of $\mathbb{X}$. Employed calibrations of B_d^A can be found in the Appendix. In the notation of the loading matrix the subscript denotes the dimension d of $\mathbb{X}$ and the superscript denotes whether the model fulfills the null (B^0) or the specific kind of alternative (B^A). The null is fulfilled if at most two entries within a row of the loading matrix are non-zero as then tail dependence is only caused by pairs. For example, given the parametrization

$$B_3^0 = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 0.5 & 0.5 \end{pmatrix},$$

we have that the STDFs for bivariate pairs are $\ell_{12}(1,1) = \ell_{13}(1,1) = \ell_{23}(1,1) = 1.5$, while $R_{123}(1,1,1) = 0$. Einmahl et al. (2012) show that $\ell(\mathbf{x}) = \sum_{i=1}^{r} \max_{j=1,...,d}(a_{ij}/(\sum_{i=1}^{r} a_{ij}))x^{(j)}$, and thus $\ell_{123}(1,1,1) = 1.5$ and $\Delta = 0$. If more than two elements within a row are non-zero, there exist common factors that induce three or more components of $\mathbb{X}$ to become simultaneously extreme. Thus, tail dependence is also caused by higher–dimensional joint extremes, and the null would be false. This is the case for $B_3^{A1}, B_5^{A2}, B_5^{A2}, B_7^{A1}, B_7^{A2}, B_7^{A3}$. Specifically, the number of non-zero entries per row describes the dimension in which joint extremes occur. Model notation is chosen such that with increasing index of A the order of tail events increases, i.e. B_5^{A1} allows for joint extremes of $X^{(1)}, X^{(2)}$ and $X^{(3)}$ (first row) while in case of B_5^{A2} also four-dimensional joint extremes of $X^{(1)}, X^{(2)}, X^{(3)}$ and $X^{(4)}$ can occur (first row).

In extreme value statistics, simulation results are usually sensitive to the choice of the threshold parameter k. Large values of k cause a systematic bias of $\widehat{\Delta}$, whereas a small k induces a large variance. We use a data-driven approach to the threshold choice in our simulation study. Within a reasonable interval, k is chosen randomly within each simulation replication. This interval is defined as $[0.01n, cn^{1/2}], c \in [1, 2]$. By several simulation runs, we found the best choices for c concerning test size are 1.75 in $d = 3$, 1.45 in $d = 5$, and 1.1 in $d = 7$.[5] For comparability of results across increasing dimensions d, we let c decrease with d. With increasing dimensions, the range of $\ell_{\mathbb{X}}$ and the number of possible univariate extremes increase. To achieve comparability across dimensions, higher cut-off values $X_{[n:n+0.5-k]}^{(i)}$ are chosen for higher dimensions. Generally, in our simulation experiments, we find that the power of the test is fairly robust against changes in c.

---

[5]In dimensions $d = 4$, and $d = 6$ we found $c = 1.5$, and $c = 1.2$, respectively, to perform best.

For the simulation in each specification, we employ five sample sizes which are standard for analyzing daily financial data ($n_1 = 200, n_2 = 500, n_3 = 1000, n_4 = 1500, n_5 = 2000$). Table (1) contains the empirical rejection rates of the test in each of the model classes at a nominal significance level of 5%. For max factor models, we find that the empirical power of the test is generally high in all considered dimensions. For models with only a slight impact of HOTDs, however, the test requires sample sizes larger than 1000 in $d = 5$ in order to yield satisfactory power, which appears adequate given the difficulty of the problem in small samples. But empirical power quickly converges to one for larger sample sizes. And empirical sizes appear close to the nominal level and plateaus around 5% for n sufficiently large. Depending on the exact model specification, this can occur already for the smallest sample size of 200. While empirical power is robust against the choice of k, we found that empirical sizes vary substantially when altering the domain of k. Generally, the test rejects too often if k tends to be small, thus empirical sizes are systematically smaller than nominal levels. In financial risk management, however, one would prefer a test with a larger false positive rate over a test that tends to falsely overlook prevalent HOTDs. Still, as we model k as a uniform random variable defined over an interval of reasonable possible values, reported sizes are more robust with respect to k than if k was a fixed value.

For the meta t-distribution, increasing dimensions and decreasing degrees of freedom of the copula imply high empirical rejection rates. This is to be expected given the above discussion of the properties of the meta t–distribution.

For all specifications, empirical power monotonously converges to one as n increases. For perfectly tail dependent DGPs (B_3^{A1}), and meta t–distributions with small ν_C, empirical power is always very high, irrespective of the dimension. Conditional on the choice of k, empirical sizes are also close to α for the DGPs characterized by B_3^0, B_5^0, and B_7^0, again irrespective of the dimension. Hence, up to dimension 7, the usual curse of dimensionality often encountered when employing non-parametric methods appears not to play a role for our test. For small sample sizes, empirical size is slightly larger than the nominal size α. Furthermore, if Δ is close to zero (e.g. for a meta t–distribution with $\nu_C = 20$), larger sample sizes such as $n_3 = 1000$ are required for the test to accurately identify the presence of HOTDs.

3.2 Local power analysis

In this subsection, we study the performance of the test under a series of local deviations from the null hypothesis. In contrast to the fixed alternatives of the subsection before, alternatives here are very close to the null and their distance to the null can shrink with increasing sample size, revealing the power optimality properties of the test. Thus, we evaluate the ability of the test to detect a violation of the null if the nature of the underlying distribution of $\mathbb{X}$ is such

Table 1: Empirical rejection rates: Max factor models and *iid t*–copula ($df = \nu_C$, $\rho = 0.5$) with t– distributed marginals ($df = 5$).

	$d = 3$					$d = 5$					$d = 7$				
	200	500	1000	1500	2000	200	500	1000	1500	2000	200	500	1000	1500	2000
t-distr. ν_C															
5	35.2	51.2	66.2	74.2	78.6	60.2	79.4	91.6	96.2	99.6	70.4	82.2	97.0	100	100
10	29.8	39.0	45.4	55.4	63.2	54.8	68.0	78.2	88.2	93.4	59.2	77.0	91.0	97.6	99.8
15	25.6	32.8	42.0	45.8	57.8	54.2	57.2	73.6	83.0	91.8	58.4	72.2	85.0	94.0	98.8
20	24.0	30.0	41.6	40.0	51.6	60.4	62.2	73.2	81.2	87.0	56.6	69.0	86.0	92.0	96.6
max factor															
B_3^0	5.2	4.2	4.2	5.0	4.8	–	–	–	–	–	–	–	–	–	–
B_3^{A1}	100	100	100	100	100	–	–	–	–	–	–	–	–	–	–
B_5^0	–	–	–	–	–	7.2	6.8	7.6	5.6	5.2	–	–	–	–	–
B_5^{A1}	–	–	–	–	–	20.4	34.4	48.2	60.0	70.2	–	–	–	–	–
B_5^{A2}	–	–	–	–	–	59.2	76.8	94.4	97.4	99.4	–	–	–	–	–
B_5^{A3}	–	–	–	–	–	100	100	100	100	100	–	–	–	–	–
B_7^0	–	–	–	–	–	–	–	–	–	–	4.0	4.4	2.4	5.2	5.8
B_7^{A1}	–	–	–	–	–	–	–	–	–	–	49.2	73.8	95.4	100	100
B_7^{A2}	–	–	–	–	–	–	–	–	–	–	59.6	80.6	98.6	100	100
B_7^{A3}	–	–	–	–	–	–	–	–	–	–	68.8	89.4	99.0	99.8	100

that fewer and fewer joint extremes in dimension ≥ 3 occur in finite samples. Following Berg & Quessy (2009) and Kojadinovic & Yan (2010), such distributions are generated by mixing distributions that violate the null, denoted by $F_{\mathbb{X},H1}$, with distributions that comply with the null, denoted by $F_{\mathbb{X},H0}$. We define the mixture distribution by

$$\mathcal{F}_{\mathbb{X},\lambda(n)}(\mathbf{x}) := (1 - \lambda(n))F_{\mathbb{X},H0}(\mathbf{x}) + \lambda(n)F_{\mathbb{X},H1}(\mathbf{x}), \tag{13}$$

where $\lambda(n)$ decreases to zero for increasing sample size n and $F_{\mathbb{X},H0}(\mathbf{x})$ satisfying $\Delta = 0$, $F_{\mathbb{X},H1}(\mathbf{x})$ satisfying $\Delta > 0$, and $F_{\mathbb{X},H0}(\mathbf{x}) \leq F_{\mathbb{X},H1}(\mathbf{x}), \forall \mathbf{x}$, ensuring realizations from $F_{\mathbb{X},H1}$ enter the extreme part of the sample. Denote the test statistic resulting from the mixture distribution $\mathcal{F}_{\mathbb{X},\lambda(n)}(\mathbf{x})$ by $\Delta_{\lambda(n)}$. For $\lambda(n) = O((\sqrt{k(n)})^{-1})$, we can show that, asymptotically,

$$\sqrt{k}(\widehat{\Delta}_{\lambda(n)} - \Delta_{\lambda(n)}) \xrightarrow{d} N(0, \sigma^2_{\widehat{\Delta}_{\lambda(n)}}),$$

where the asymptotic variance can again be obtained analytically from theorem 4.3 in Einmahl et al. (2012). Thus, the test has power against any local alternatives if and only if these alternatives are at least of order $(\sqrt{k(n)})^{-1}$ apart from the null.

In the following simulations, we illustrate this result. Hence, we are interested in rejection rates of $\Delta = 0$ from mixture distributions defined in equation (13) for $\lambda(n) := \lambda k(n)^{-1/2}$, with

Figure 1: Empirical test power for the mixture distributions defined in equation (14) with sample size $n = 2000$ at level 5%.

(a) $d = 3$ (b) $d = 5$
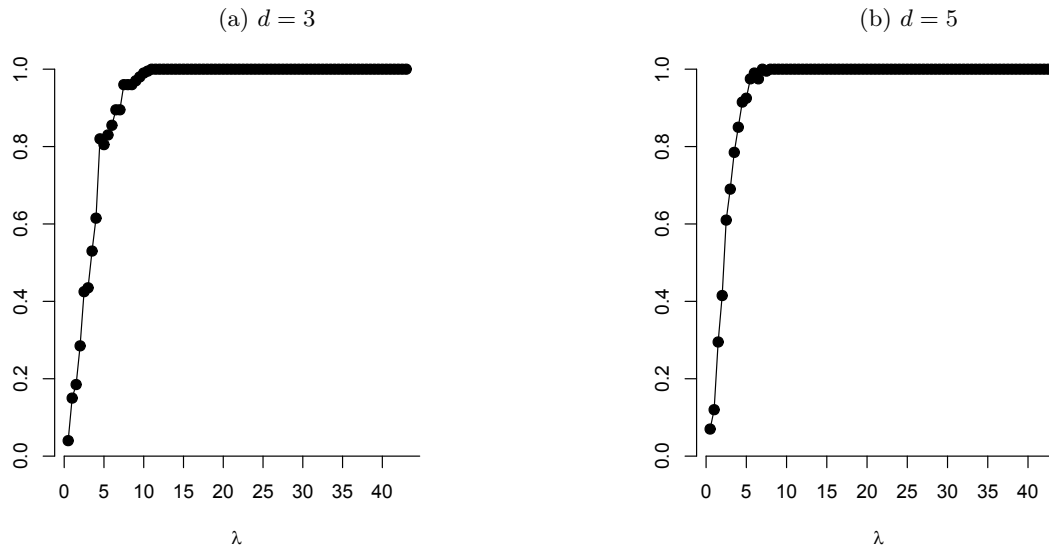
$0 < \lambda \leq k(n)^{-1/2}$. We determine k as in the simulations before. In order to calculate local powers p_n, we generate $S = 1000$ samples from a DGP of mixture distribution type, with fixed sample size and increasing λ. Local power is estimated by $\hat{p}_n = 1/S \sum_{i=1}^{S} 1\{\widehat{\Delta}_{\lambda(n)} > z^{\alpha} \hat{\sigma}_{\widehat{\Delta},\lambda(n)} k^{-1/2}\}$ for every $\lambda(n)$. The asymptotic variance $\hat{\sigma}^2_{\widehat{\Delta},\lambda(n)}$ is estimated by the bootstrap procedure presented in Section (2.3). Berg & Quessy (2009), Kojadinovic & Yan (2010) carry out similar analyses for goodness–of–fit tests of parametric (extreme value) copulas. For the sake of brevity, we concentrate on dimensions $d \in \{3, 5\}$, sample size $n = 2000$, and we let λ increase. For $d = 3$,

$$\mathcal{F}_{\mathbb{X},\lambda(n)}(\mathbf{x}) = (1 - \lambda(n))F_{\mathbb{Y}}(\mathbf{y}) + \lambda(n)F_{\mathbb{W}}(\mathbf{w}), \tag{14}$$

where $F_{\mathbb{Y}}(\mathbf{y})$ and $F_{\mathbb{W}}(\mathbf{w})$ are the distribution functions of the max factor model B_3^0 and B_3^{A1}, respectively. To ensure that realizations of B_3^{A1} actually enter the extremal part of the sample, factors Z are first used to generate data from $F_{\mathbb{Y}}(\mathbf{y})$ and then multiplied by a constant larger than one when generating data from $F_{\mathbb{W}}(\mathbf{w})$. For $d = 5$, we mix the distribution function of B_5^0 and $B_5^{A3} := (1/5 \ 1/5 \ 1/5 \ 1/5 \ 1/5)$. If, for example, $\lambda = 2$, we have $\lambda(n) \approx 0.045$. Thus 4.5% of the extreme part of the sample is generated by the $F_{\mathbb{W}}$ which violates the null. This share increases in λ. Figure 1 shows estimated local powers with $\alpha = 0.05$. The test successfully detects minor violations from the null. Even for small λ, when the impact of the perturbating DGP is small, rejection rates quickly converge to one. Increasing the dimension d accelerates the convergence speed of empirical power.

3.3 Size and power for serially dependent and conditionally heteroskedastic components

While Proposition 2.2 assumes *iid* data, financial time series, in particular asset returns, feature small autocorrelations and time–varying conditional volatility, and thus cannot be considered *iid*. In order to apply our test to detect the cross-sectional tail dependence structure of financial time series, the data have to be pre-filtered. We use autoregressive moving average (ARMA) models for the mean, and the class of generalized autoregressive conditional heteroskedasticity (GARCH) models for the variance, (Bollerslev (1986)). After filtering, we expect that the resulting standardized residuals are largely free of serial dependence in first and second moments, and are thus nearly *iid*. We therefore apply the test to these pre-filter residuals instead of the raw observations. In the applied extreme value literature, this approach is common when dealing with time-series effects, see e.g. McNeil & Frey (2000) in a univariate setting for extreme quantile estimation, and Poon et al. (2004) for estimating bivariate tail dependencies between financial time series.

While it is intuitively clear that $\sqrt{n}$-consistent parametric pre-filtering should not impact the consistency of the slower converging non–parametric estimator of the STDF, there are no formal theoretical results on the asymptotic properties of such dependence estimates for pre-estimated residuals available yet. In comparison to semi-parametric and non-parametric distribution copula estimation (see e.g. Chen & Fan (2006), Rémillard (2010), Oh & Patton (2013)) such results for non-parametric tail dependence estimation would require completely different empirical process techniques for respective rank statistics which do not exist and are extremely challenging to develop. In what follows, we therefore focus on the finite sample performance of the test in such settings. In particular, we explore if and how empirical size and power of the DGPs from Section 3 change when introducing autocorrelation and time–varying conditional volatility.

We follow Oh & Patton (2013) and generate random draws from the following AR(1)–GARCH(1,1) processes, which are linked by the error term copula $\mathcal{C}_\eta$:

$$y_t^{(i)} = \mu_t^{(i)} + \sigma_t^{(i)}\eta_t^{(i)} = \varphi_0^{(i)} + \varphi_1^{(i)}y_{t-1}^{(i)} + \sigma_t^{(i)}\eta_t^{(i)},$$

$$\sigma_t^{2,(i)} = \omega^{(i)} + \alpha^{(i)}\left(y_{t-1}^{(i)} - \mu_{t-1}^{(i)}\right)^2 + \beta^{(i)}\sigma_{t-1}^{2,(i)}$$

$$\eta := (\eta^{(1)}, ..., \eta^{(d)}) \sim iid \ F_\eta(x^{(1)}, ..., x^{(d)}) = \mathcal{C}_\eta(F_{\eta,(1)}(\eta^{(1)}), ..., F_{\eta,(d)}(\eta^{(d)})), \qquad (15)$$

$t = 1, ..., T$. $\theta^{(i)} = (\varphi_0^{(i)} = 0.01, \varphi_1^{(i)} = 0.05, \omega^{(i)} = 0.05, \alpha^{(i)} = 0.1, \beta^{(i)} = 0.85)'$ denotes the vector of AR–GARCH parameters for marginal i, F_η is the continuous joint distribution function of the vector of error terms $\eta = (\eta^{(1)}, ..., \eta^{(d)})$, and $F_{\eta,(i)}(\eta^{(i)})$ are the marginal distributions of the error terms linked by error term copula $\mathcal{C}_\eta$. Hence the dependence structure of η is the "true" but

unobserved dependence we are interested in, and from which the observed dependence structure between the realizations $y_t^{(i)}$ might differ due to autocorrelation and GARCH effects. See Oh & Patton (2013) for details on such DGPs.

We test for HOTDs in the observed, unfiltered realizations $(y_t^{(1)}, ..., y_t^{(d)})_{t=1}^T$, and in correctly standardized residuals $(\hat{\eta}_t^{*,(1)}, ..., \hat{\eta}_t^{*,(d)})_{t=1}^T, \hat{\eta}_t^{*,(i)} := (y_t^{(i)} - \hat{\mu}_t^{(i)})/\hat{\sigma}_t^{(i)}$. To evaluate the size of the test, we choose the max factor model of type B_3^0, B_5^0, and B_7^0 as models for the error term copula $\mathcal{C}_\eta$. Thus, the test is again applied in dimensions $3, 5$, and 7. In contrast to the *iid* setting, we do not employ a Fréchet(1) distribution, which would produce very extreme observations such that numerical GARCH–estimation may fail to converge. As marginal error distributions $F_{\eta,(i)}(\eta^{(i)})$ we choose t–distributions with degrees of freedom $\nu_m \in \{5, 10, 15, 20\}$ for size analysis. For power analysis, $\mathcal{C}_\eta$ is the t–copula with degree of freedom $\nu_C \in \{5, 10, 15, 20\}$, and fixed marginal degrees of freedom $\nu_m = 5$, i.e. η follows a meta t–distribution.[6] Thereby we can observe how quickly the test reacts to a steadily diminishing degree of HOTDs.

Simulations are repeated $S = 500$ times with sample sizes $n_2 = 500, n_3 = 1000, n_4 = 1500, n_5 = 2000$. We do not include $n_1 = 200$ in this section as GARCH estimates for a sample size of 200 may be unreliable. The parameter vector $\theta = (\theta^{(1)}, ..., \theta^{(d)})'$ is estimated by maximum likelihood, assuming marginal t–distributions with estimated degrees of freedom. Table (2) reports empirical rejection probabilities for the factor copula with $\Delta = 0$; Table (3) reports empirical rejection probabilities in case of the t-copula as error term copula for filtered and unfiltered data, respectively. We find that disregarding serial correlation and time–varying volatility worsens size and power properties, and a correct filter leads to similar results as in the *iid* case. Empirical rejection rates for the max factor copula indicate that the test is slightly undersized. Yet empirical sizes are still satisfactorily close to 5%.

The effect of serial correlation and GARCH effects becomes clear when comparing the number of test rejections of the binomial test $H_0 : p_i = 0.05$, where p_i denotes a rejection probability (test level 5%). That is, for each setting we compare all 48 empirical rejection rates of filtered and unfiltered data with the nominal size of 5%. The correctly specified AR(1)-GARCH(1,1) filter leads to 18.2% of all cases in which the empirical rejection probability significantly differs from the nominal size. Not filtering the data amounts to 31.3% significant deviations. With a binomial test one can also compare empirical powers of the *iid* and the non-*iid* settings. In 95.8% of comparisons, applying the test to the residuals of the correctly specified GARCH process produces significantly higher power than testing in the unfiltered returns. Hence, disregarding the time series properties of the data worsens size and power

---

[6]As in the *iid* case, empirical power is robust against varying the marginal degrees of freedom. Yet, we report empirical sizes for different ν_m in order to have more data points for a more accurate comparison between test performances for unfiltered and filtered time series.

Table 2: Empirical rejection rates under H_0: Max factor copula as error term copula, $t(df = \nu_m)$–distributed errors, and GARCH(1,1) volatility model.

	$d = 3$				$d = 5$				$d = 7$			
	500	1000	1500	2000	500	1000	1500	2000	500	1000	1500	2000
filtered												
ν_m												
5	4.6	3.0	3.8	5.2	4.4	6.0	6.4	6.2	2.8	2.4	4.2	3.4
10	3.4	2.0	2.8	3.8	6.0	4.0	5.6	5.2	2.8	3.6	3.2	2.4
15	5.2	5.0	5.4	4.4	4.4	3.6	4.8	4.2	3.0	2.0	3.4	5.0
20	3.8	2.6	5.4	4.4	5.6	5.2	5.4	4.4	3.6	1.8	3.6	4.8
unfiltered												
ν_m												
5	3.4	2.6	4.4	4.6	4.2	5.4	4.6	2.8	1.4	3.4	1.8	1.6
10	4.6	2.2	4.2	4.2	4.8	4.4	5.2	4.6	1.6	3.8	2.8	2.0
15	4.6	4.6	4.6	5.8	5.4	4.2	4.8	5.8	3.4	2.0	2.8	1.8
20	5.2	2.6	5.6	4.4	6.4	6.0	4.2	7.0	3.0	2.0	2.2	2.8

Table 3: Empirical rejection rates under H_1: Correctly filtered and unfiltered GARCH processes with t–copula $(df = \nu_C, \rho = 0.5)$ as error copula, $t(\nu_m = 5)$–distributed errors.

	$d = 3$				$d = 5$				$d = 7$			
	500	1000	1500	2000	500	1000	1500	2000	500	1000	1500	2000
filtered												
ν_C												
5	45.0	63.2	71.8	79.4	77.8	90.8	96.8	99.2	83.6	97.2	99.4	99.8
10	49.4	62.2	74.2	83.0	67.0	81.8	86.4	92.4	72.4	91.2	97.4	99.8
15	32.2	42.0	45.4	52.8	60.4	74.6	83.0	90.6	70.8	86.0	93.8	100
20	26.0	39.6	42.2	49.6	56.4	70.6	79.4	86.8	66.6	82.2	93.0	97.0
unfiltered												
ν_C												
5	24.6	27.2	30.2	31.6	51.2	53.8	60.0	61.6	61.8	68.2	76.0	79.4
10	17.0	19.2	21.0	23.4	41.6	42.0	42.4	43.4	54.0	55.8	61.4	60.4
15	17.4	15.0	14.2	15.4	37.8	39.0	38.4	38.8	45.6	50.8	55.2	56.6
20	16.4	12.4	18.0	13.2	36.8	35.2	35.2	38.0	43.0	48.2	51.2	53.2

results.

Finally, we compare the power results obtained when simulating from the *iid* meta t–distribution (Table (1)) with those corresponding to correctly filtered, and unfiltered AR(1)-GARCH(1,1) processes connected via the meta t–distribution (Table (3)). In case of the correct filter, empirical power never differs significantly from the *iid* case (test level 5%). In the unfiltered series, however, empirical power is significantly lower in 91.6% of all cases (test level 5%). Hence, disregarding time–varying volatility amounts to lower power and lower test size, and testing in correctly filtered series produces nearly identical results as for *iid* data.

4 Higher order tail dependencies in global and European stock markets

4.1 Data description

In our empirical application, we study extreme gains and losses of two different sets of stock indices. First, we test for HOTDs in left and right tails of the weekly stock return distributions on a global level, while in a second study we focus exclusively on daily European stock returns. The "global portfolio" consists of three stock indices of the USA, Europe and the Asian Pacific region, namely the MSCI USA, MSCI Pacific, and MSCI Europe.[7] The "European portfolio" consists of seven individual European MSCI indices, including the largest European economies (United Kingdom, Germany, France, Italy, Spain), as well as smaller economies that played a role during the recent European sovereign debt crisis (Greece, Portugal).[8] The two portfolios are analyzed separately.

The sample period of the global portfolio ranges from 01/30/1970 to 10/29/2014. To overcome problems arising from different time zones, we use weekly returns. As observations of MSCI Pacific are only available on a monthly frequency until 12/30/1983, a weekly proxy for MSCI Pacific during that time period is created by averaging over weekly observations of the MSCI Japan and MSCI Australia, with weights equal to 2/3 and 1/3, respectively. This weighting scheme resembles the current composition of the MSCI Pacific.[9] After deleting weeks with zero returns, the sample features 2335 observations for each index. The sample period of the European portfolio ranges from 01/04/1988 to 10/29/2014. In this second portfolio time zone effects do not matter, so we can use daily returns. After discarding days with zero returns, the sample has 6889 observations for each index.

Both samples are tested against HOTDs with rolling windows containing $n = 750$ observations, corresponding to roughly 15 years in the global portfolio, and roughly three and a half years in the European portfolio. Simulation studies document appropriate test performance for such a sample size, and we aim to keep the window length as short as reasonably possible. The test is applied bi–weekly for the global portfolio and every fifth day for the European portfolio. We test against HOTDs in raw observations, and in standardized ARMA–GARCH residuals, along the lines of model (15) in order to eliminate the effects of serial correlation and time–varying volatility. Returns are thus modeled by

$$y_t^{(i)} = \mu_t^{(i)} + \sigma_t^{(i)}\eta_t^{(i)}, \ i = 1,...,7, \eta_t^{(i)} \overset{iid}{\sim} t(\nu_m).$$

---

[7]Data are available on Datastream with mnemonics MSUSAML, MSPACF\$ and MSEROP\$.

[8]Data are available on Datastream with mnemonics MSUTDKL, MSGERML, MSFRNCL, MSITALL, MSGREEL, MSSPANL and MSPORDL.

[9]Data are available on Datastream with mnemonics MSJPANL and MSAUSTL.

Standardized residuals $\hat{\eta}_t^{*,(i)} = (y_t^{(i)} - \hat{\mu}_t^{(i)})/\hat{\sigma}_t^{(i)}$ are re–estimated in each window to address potential parameter changes. In every window, each time series is fitted to an ARMA($p \leq 2, q \leq 2$) model with automatic order choice according to the Schwarz information criterion. Subsequently, the data are fitted to a threshold- (T)GARCH(1,1) (see Glosten et al. (1993)) model.[10] A TGARCH(1,1) model is given by

$$\sigma_t^2 = \omega + \alpha\sigma_{t-1}^2\eta_{t-1}^2 + \delta 1\{\eta_{t-1} < 0\}\sigma_{t-1}^2\eta_{t-1}^2 + \beta\sigma_{t-1}^2, \quad t = 1, ..., T.$$

Notably, a TGARCH model is able to capture asymmetric impacts of positive and negative shocks. Hence, in each window and for both return losses and return gains, we test against HOTDs in raw returns and in standardized ARMA(p,q)–TGARCH(1,1) residuals.

In order to calculate $\widehat{\Delta}$, we have to choose the number of upper order statistics k. The rolling window scheme complicates a manual choice for each window. Thus, for each $k \in [0.01n, c\sqrt{n}]$ we compute $\widehat{\Delta}(k)$ and take the median thereof as the final estimator for Δ (see Sections 2.3 and 3 for details). Figures 2 and 3 show the evolution of $\widehat{\Delta}$, 90% confidence intervals for Δ and test decisions for the global and the European portfolio at each time point; a confidence interval is colored gray whenever $H_0 : \Delta = 0$ has to be rejected, i.e. whenever multivariate tail risk is not only bivariate. [11]

4.2 Results and economic implications

Regarding global portfolio gains, the TGARCH(1,1) filtered series never allows rejecting the null hypothesis of no HOTDs, while for the unfiltered series the null has to be rejected after 2010 with P–values close to 5%. Still, the absolute amount of HOTDs is also small after 2010. For losses, we detect an accentuated increase in HOTDs after 2006/2007, whereas no significant HOTDs up to 2006 can be found. The gradually increasing HOTDs appear to be still on the rise at the end of the sample. Although the sample covers major historical events such as the 1970s oil crises, the Black Monday 1987, the dissolution of the Soviet Union, the Gulf War 1990–91, the Asian financial crisis in 1997, the introduction of the Euro, the burst of the dot–com bubble, and 09/11/01, it is the global financial crisis of 2007–08 that marks the start of global HOTDs becoming significant. Thus the latter is the only event within the sample, that is capable of herding global high–dimensional extreme losses. Before 2007–08, investors holding a globally diversified portfolio did not have to pay attention to HOTDs while this has apparently become

---

[10]Results obtained by using a standard GARCH model were qualitatively very similar and are therefore not reported.

[11]Note, as $\Delta \geq 0$, theoretically, confidence interval lower bounds should not become negative yet this bound decides whether Δ is significantly larger than zero. Furthermore, as we are conducting one-sided tests, the shaded areas within the 90% confidence intervals refer to test rejections on 5% significance level.

an additional challenge in asset allocation on top of bivariate tail dependence nowadays. On the other hand, throughout the considered time span, investors cannot expect to benefit from HOTDs between gains: The financial turmoils during 2007–08 caused univariate extreme losses to trigger joint global extreme losses, whereas univariate extreme gains still do not spread out (unfiltered), or at least not as strongly as extreme losses (filtered).
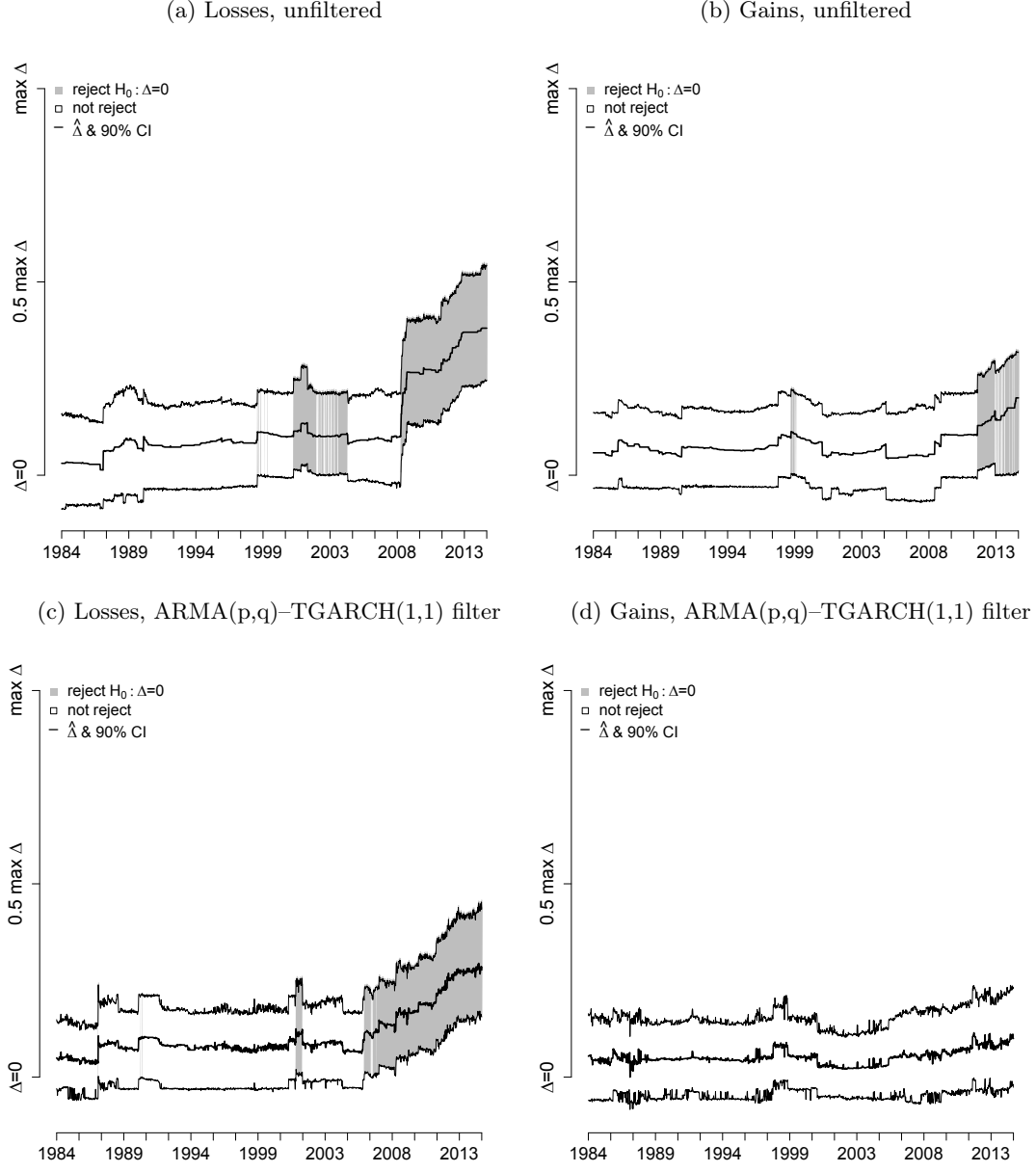
Losses and gains within the European portfolio, on the other hand, are more prone to HOTDs. This may be explained by closer economic connections, but also by the fact that now seven indices are considered, implying that extreme connections between three or more components are more likely than within a three–dimensional portfolio.[12] Intra–European HOTDs appear to be time–varying and are most of the time significant. The TGARCH filter smoothes the evolution of $\widehat{\Delta}$, suggesting that the unstable behavior of $\widehat{\Delta}$ for the unfiltered series can be partly explained by serial correlation and time–varying conditional volatility. However, the results do not differ qualitatively with respect to whether the filter is employed or not, as the test decisions on a significance level of 5% are mostly alike for both specifications. Overall, the empirical variance of $\widehat{\Delta}$ appears to be constant for both the filtered and the raw data. For losses, one observes a decrease of HOTDs from the sample beginning until the mid to mid/end–1990s; also, HOTDs are not significant between 1994 and 1998. Afterwards, the importance of HOTDs increases until the beginning of the 2007–08 crisis, remaining on a stable, high level ever since. Interestingly, this movement is continuous and the major political events that fall in this period (dot–com crisis, 9/11/01, introduction of the Euro) do not cause discontinuities of the trajectory of $\widehat{\Delta}$.

We conclude that HOTDs in the European portfolio are not driven by one–time events but rather mirror established, mid- to long-term processes due to the European financial and economic integration. This also gives an explanation for why gains HOTDs of the European portfolio prevail throughout the sample, which stands in contrast to nearly non–existent HOTDs in gains within the global portfolio. Diversification opportunities of cross–sectional extreme losses are limited within Europe, as it was also found in Christoffersen et al. (2012). Our test results for right tails indicate, however, that there is potential to benefit from cross–sectional extreme gains. This generalizes the results in Poon et al. (2004) as the presence of HOTDs implies their results based on pairwise bivariate analysis. Moreover, we observe that tail dependence, at least within European stock markets, is more severe than assumed so far.

---

[12]The latter makes a comparison of test results across both data sets difficult.

Figure 2: Dynamics of the test statistic $\widehat{\Delta}$ (see equation (10)), together with 90% confidence intervals, the global portfolio, using a rolling window of roughly 15 years. The left panel shows test decisions for portfolio losses, whereas the right panel shows test decisions for portfolio gains. Confidence intervals are colored gray whenever $H_0 : \Delta = 0$ has to be rejected.
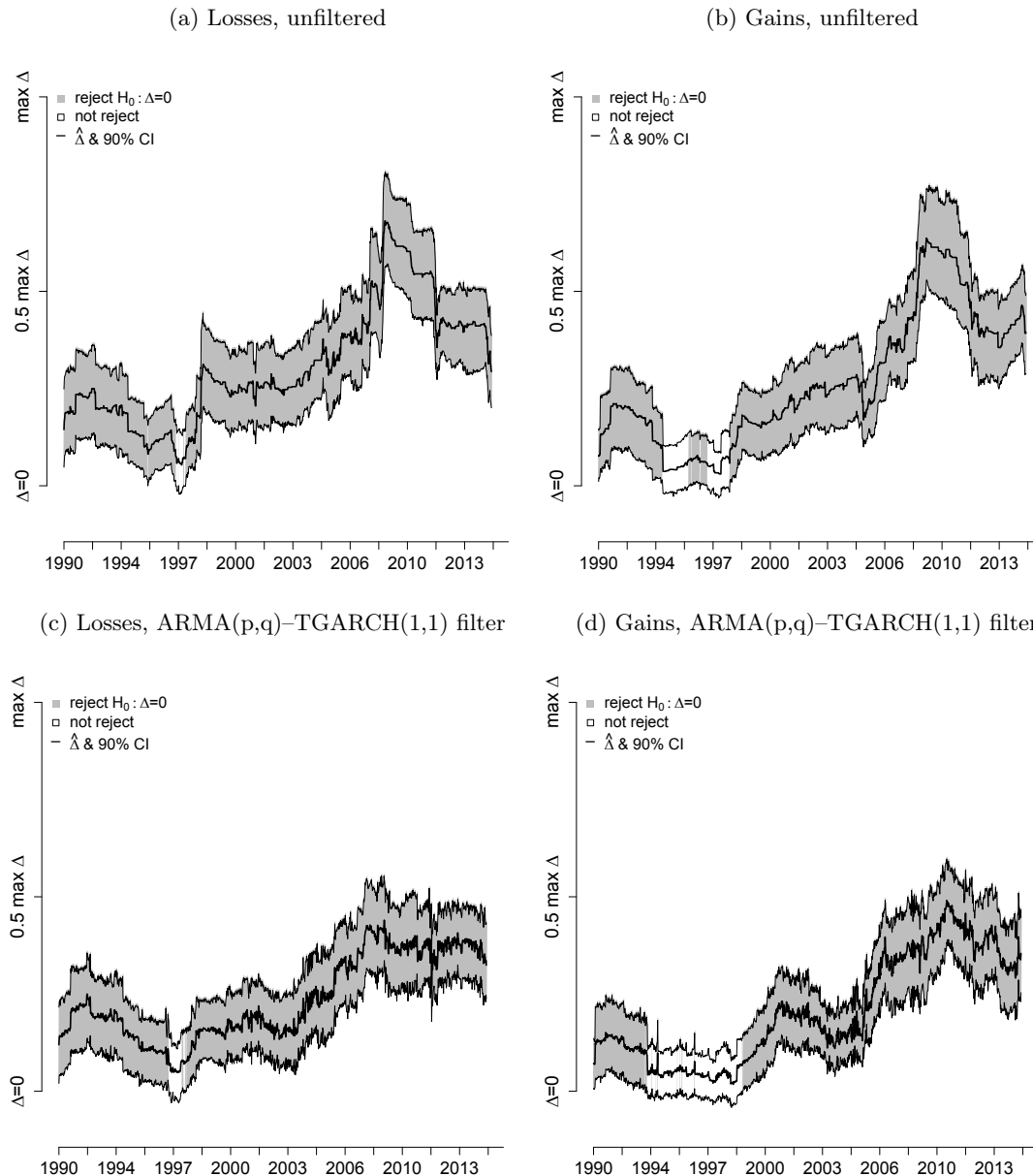
(a) Losses, unfiltered

(b) Gains, unfiltered

(c) Losses, ARMA(p,q)–TGARCH(1,1) filter

(d) Gains, ARMA(p,q)–TGARCH(1,1) filter

4.3   Factor model for the European stock market

The presence of HOTDs within the European portfolio might be caused by tail events of a common external factor. To distill truly intra–European HOTDs, we now control for effects of global financial markets. Returns $y_t^{(i)}$ are thus modeled by a factor market model

$$y_t^{(i)} = \zeta^{(i)} M_t + \epsilon_t^{(i)}, \; i = 1, ..., 7,$$

Figure 3: Dynamics of the test statistic $\widehat{\Delta}$ (see equation (10)), together with 90% confidence intervals, for the European portfolio, using a rolling window of 3 to 4 years. The left panel shows test decisions for portfolio losses, whereas the right panel shows test decisions for portfolio gains. Confidence intervals are colored gray whenever $H_0 : \Delta = 0$ has to be rejected.

(a) Losses, unfiltered

(b) Gains, unfiltered

(c) Losses, ARMA(p,q)–TGARCH(1,1) filter

(d) Gains, ARMA(p,q)–TGARCH(1,1) filter

where M_t denotes a common factor for all marginal returns $y_t^{(i)}$. The disturbance $\epsilon_t^{(i)}$ is often interpreted as the idiosyncratic part of $y_t^{(i)}$. An apparent choice for M_t is the return series of the MSCI World Ex Europe[13] as it is an index of all relevant stock markets except for European ones.

We repeat the rolling window analysis of the previous section for the European portfolio and test for HOTDs between (unfiltered) factor model residuals $(\hat{\epsilon}_t^{(1)}, ..., \hat{\epsilon}_t^{(7)})_{t=1}^n$. Furthermore, we

---

[13]This index runs under mnemonic MSWXEU\$âĂŃ in Datastream.

obtain standardized residuals from a ARMA(p,q)–TGARCH(1,1) model for the factor model residuals. Test decisions for the latter two models thus account for serial correlation, time–varying volatility and the effect of the common risk driver. The remaining dependence structure can be considered as idiosyncratic to the European stock market system. For all seven indices, the return model is re–estimated in each window[14] and the orders of the ARMA models are again found with the Schwarz criterion. Test results for both gains and losses of unfiltered data and ARMA-TGARCH(1,1) filtered data are shown in Figure 4.

Controlling for changes in global stock markets slightly attenuates European HOTDs, yet results closely resemble the results from the previous subsection (Figure 3). The only major exception where controlling for the world index alters the test decision, in the sense that it causes HOTDs to be significant, is for gains during 1990–94, Figure 3 (d) and Figure 4 (d). However, the effect of the market factor to HOTDs between European gains has increased since 2006 which can be seen by comparing Figures 3 (b) and 4 (b). Both do not account for ARMA–(T)GARCH effects and the only possible source for a difference is the accounting for the common factor. Overall, HOTDs between the idiosyncratic risks of European stock markets have increased since 2000. Thus, we can reveal that joint extremes are truly due to intra–European HOTDs. For a practitioner, this provides econometric evidence that losses on portfolios with different European based Exchange-traded funds, or with different single European stocks, are likely to add up in times of crisis, and diversification effects may fade away in case of tail events for solely stock-based portfolios. As multivariate extreme losses of European stock markets are apparently only slightly affected by events of the market factor, there exist tail diversification opportunities between both. These opportunities slightly diminish for extreme gains. Besides the importance for asset allocation, significance of HOTDs also seems to mark periods of distress in the markets, i.e. when stock indices tend to jointly experience large losses.

4.4   Importance and share of higher order tail dependencies in practice

To show the importance of testing for HOTDs, we provide some simple descriptive screening tools in this subsection. In particular, we assess the share of bivariate tail events that cannot be captured by tail correlations. For this, we use the asymptotic probabilities of two or three

[14]Whenever numerical optimization of the likelihood function failed for the given setting, we first changed the conditional distribution from a t to a normal distribution. In seven out of 8596 estimated models we then only came across convergence problems for 8 TGARCH models. In these cases we used residuals from the GARCH(1,1) model as substitute. There appears to be one outlier of $\widehat{\Delta}$ for TGARCH residuals at 4/7/96 where the optimization of the likelihood for the TGARCH model struggles.

Figure 4: Dynamics of the test statistic $\widehat{\Delta}$ (see equation (10)), together with 90% confidence intervals, for the European portfolio, using a rolling window of 3 to 4 years, after controlling for a market factor. The left panel shows test decisions for portfolio losses, whereas the right panel shows test decisions for portfolio gains. Confidence intervals are colored gray whenever $H_0 : \Delta = 0$ has to be rejected.

(a) Losses, unfiltered   (b) Gains, unfiltered

(c) Losses, ARMA(p,q)–TGARCH(1,1) filter   (d) Gains, ARMA(p,q)–TGARCH(1,1) filter
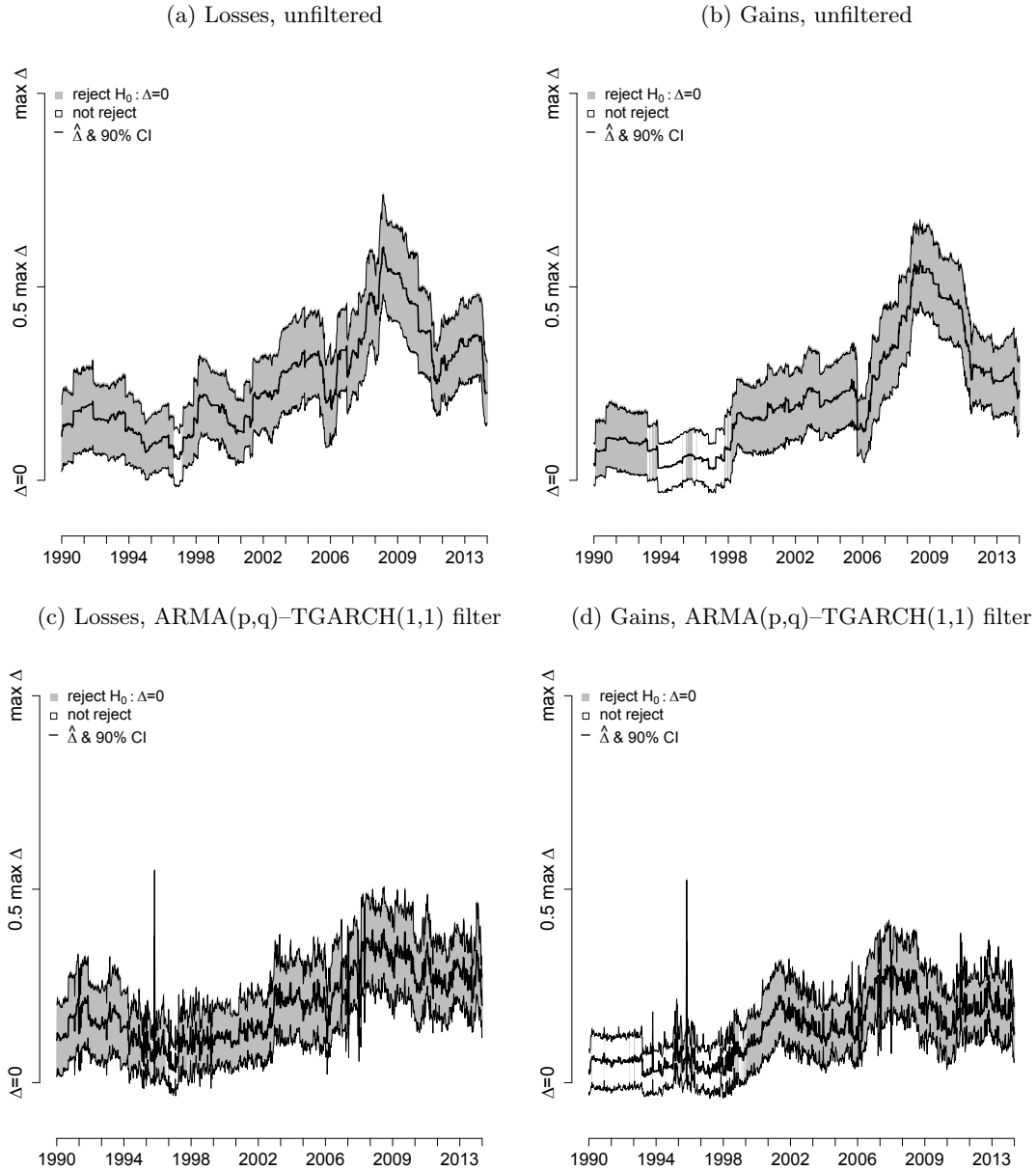
joint extremes, κ_2 and κ_3, defined as

$$\kappa_2 = \lim_{t \to 0} t^{-1} P\Big(\bigcup_{i \neq j} \big\{ \{X^{(i)} \geq F_i^{-1}(1 - tx^{(i)})\} \cap \{X^{(j)} \geq F_j^{-1}(1 - tx^{(j)})\} \big\} \Big),$$

$$\kappa_3 = \lim_{t \to 0} t^{-1} P\Big(\bigcup_{h \neq i \neq j} \big\{ \{X^{(h)} \geq F_h^{-1}(1 - tx^{(h)})\} \cap \{X^{(i)} \geq F_i^{-1}(1 - tx^{(i)})\} \cap \{X^{(j)} \geq F_j^{-1}(1 - tx^{(j)})\} \big\} \Big).$$

They describe the likelihood of at least two or respectively three assets becoming extreme at once. Their ratio κ_3/κ_2 quantifies the share of bivariate extremes that also amount to a trivariate

extreme event. Similar to the estimation of the STDF, this magnitude can be estimated by its empirical counterpart. We compare the days featuring a bivariate (κ_2) or trivariate (κ_3) extreme with the number of days with at least one univariate extreme,

$$\hat{\kappa}_2 = \frac{\sum_{t=1}^{n} \sum_{i \neq j} \prod_{g \in \{i,j\}} 1\{X_t^{(g)} > X_{n:n+0.5-k}^{(g)}\}}{\sum_{t=1}^{n} 1\{\bigcup_{i=1}^{d} X_t^{(i)} > X_{n:n+0.5-k}^{(i)}\}}, \tag{16}$$

$$\hat{\kappa}_3 = \frac{\sum_{t=1}^{n} \sum_{h \neq i \neq j} \prod_{g \in \{h,i,j\}} 1\{X_t^{(g)} > X_{n:n+0.5-k}^{(g)}\}}{\sum_{t=1}^{n} 1\{\bigcup_{i=1}^{d} X_t^{(i)} > X_{n:n+0.5-k}^{(i)}\}}. \tag{17}$$

Section 6.3 in the Appendix provides a small simulation study that shows κ_3/κ_2 is indeed a reasonable measure for determining the severeness of HOTDs.

Figure 5 shows estimates $\hat{\kappa}_2, \hat{\kappa}_3$ and $\hat{\kappa}_3/\hat{\kappa}_2$ for the TGARCH filtered European portfolio without controlling for a common factor. As before, final estimates in each window were found by taking the estimates' medians for $k \in [0.01n, c\sqrt{n}]$.

Not surprisingly, trajectories resemble the dynamics of $\widehat{\Delta}$ (Figures 3 c-d). The probability of observing trivariate extremes ($\hat{\kappa}_3$) has steadily increased from 10–20% for losses, and 5–10% for gains, respectively, during the 1990s up to 20–30% for losses, and 30-40% for gains, respectively, at the peak of the recent financial crisis 2007–2009. However, the share of trivariate extremes in bivariate extremes $\hat{\kappa}_3/\hat{\kappa}_2$ steadily declined both for losses and gains during the 1990s (from 60% to 35% for losses, and from 50% to 20% for gains) and has consequently ascended for both tails until the end of the 2010s (up to 70-80% for both losses and gains). Thus, for losses, the probability that multivariate extremes occur in larger cross-sections has doubled during the 2000s, while it has even tripled for gains in that time span. This highlights that extremes more than ever occur not only in bivariate pairs, but also in larger cross-sections.

5 Summary

This paper proposes a test that reveals situations in which common bivariate measures for tail dependence underdiagnose the potential for higher-dimensional extreme events. Test asymptotics are derived and simulations show the bootstrap implementation routine features attractive finite sample properties, despite the challenging threshold choice, inherent to extreme value statistics, which occasionally affects test size. In the case of data that exhibit serial correlation and GARCH effects, we recommend studying estimated residuals instead observed realizations, to maintain the good size and power properties.

On global stock markets, we find that cross–sectional extremes become relevant in the course of the financial crisis of 2007–08. Multivariate extremes on European stock markets are historically more intertwined, as the impact of high–dimensional extremes is significant

Figure 5: Dynamics of $\hat{\kappa}_2$, $\hat{\kappa}_3$, and $\hat{\kappa}_3/\hat{\kappa}_2$ (see equations (16) and (17)) for the TGARCH filtered European portfolio.
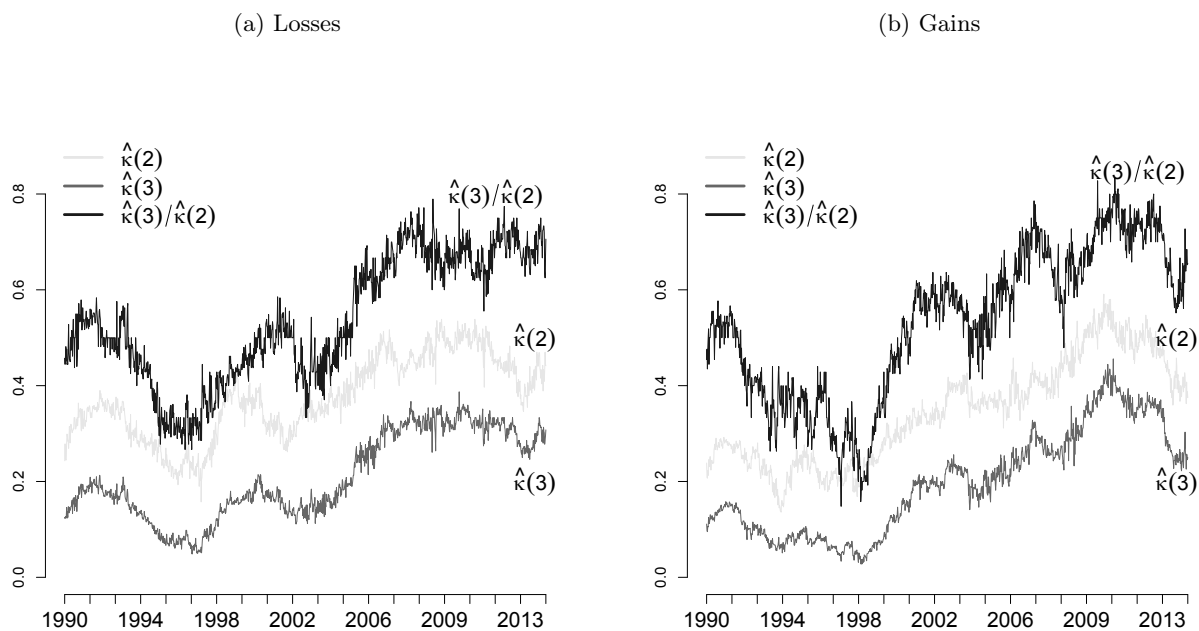
(a) Losses

(b) Gains

throughout the considered sample. There appears to be diversification potential of multivariate extreme losses between European and non–European stock markets, while extreme gains do not share this feature. Within the European system, left tail events feature no potential for diversification. We find time periods when up to 80% of extremes are truly multivariate.

6 Appendix

6.1 Model specifications

Table 4: Specifications of the max factor models.

$$B_3^0 = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix} \qquad B_3^{A1} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \end{pmatrix}$$

$$B_5^0 = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 \\ 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix} \qquad B_5^{A1} = \begin{pmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \end{pmatrix} \qquad B_5^{A2} = \begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 \\ 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$

$$B_7^0 = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 1/2 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 1/2 \\ 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix} \qquad B_7^{A1} = \begin{pmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 1/3 & 0 & 0 \\ 0 & 1/3 & 0 & 0 & 0 & 1/3 & 1/3 \end{pmatrix} B_7^{A2} = \begin{pmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}$$

$$B_7^{A3} = \begin{pmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 1/2 \\ 0 & 1/2 & 0 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 1/2 & 0 & 0 & 0 & 1/2 \end{pmatrix}$$

6.2 Proofs

Proof of proposition 2.1

If $\mathbb{X}$ is tail independent, $\ell(\mathbf{x}) = \mathbf{x1} \Leftrightarrow \ell_i(\mathbf{x}^{(i)}) = \mathbf{x}^{(i)}\mathbf{1}$, for all possible bivariate combinations i. Plugging this into the general form of Δ, and realizing that in this case $\sum_{i<j\leq 2} \ell_i(\mathbf{x}^{(i)}) =$

$(d-1)\sum_{i=1}^{d} x^{(i)}$, it follows that

$$\Delta = \ell(\mathbf{x}) - 2\sum_{i=1}^{d} x^{(i)} + d\sum_{i=1}^{d} x^{(i)} - \sum_{i<j\leq 2} \ell_i(\mathbf{x}^{(i)})$$

$$= \sum_{i=1}^{d} x^{(i)} - 2\sum_{i=1}^{d} x^{(i)} + d\sum_{i=1}^{d} x^{(i)} - \sum_{i<j\leq 2} \ell_i(\mathbf{x}^{(i)})$$

$$= \sum_{i=1}^{d} x^{(i)} - 2\sum_{i=1}^{d} x^{(i)} + d\sum_{i=1}^{d} x^{(i)} - (d-1)\sum_{i=1}^{d} x^{(i)}$$

$$= 0.$$

The reverse does not hold true. E.g. let $\mathbb{X} := (X^{(1)}, X^{(2)}, X^{(3)})$, with $\mathbb{X}^{(3)}$ being independent of $X^{(1)}$, and let $X^{(1)} \overset{a.s.}{=} X^{(2)}$, i.e. $X^{(1)}$ and $X^{(2)}$ are perfectly tail dependent. Thus, $\ell_{12}(x^{(1)}, x^{(2)}) \equiv \ell_{11}(x^{(1)}, x^{(1)}) = x^{(1)}, \ell_{13}(x^{(1)}, x^{(3)}) = x^{(1)} + x^{(3)}$, and

$$\ell_{123}(x^{(1)}, x^{(1)}, x^{(3)}) = \lim_{t\downarrow 0} t\mathbb{P}\Big(\bigcup_{i\in\{1,2,3\}} \{X^{(i)} \geq F_i^{-1}(1-tx^{(i)})\}\Big)$$

$$= \lim_{t\downarrow 0} t\mathbb{P}\left(\{X^{(1)} \geq F_1^{-1}(1-tx^{(1)})\} \cup \{X^{(3)} \geq F_3^{-1}(1-tx^{(3)})\}\right)$$

$$= x^{(1)} + x^{(3)}.$$

Rewriting Δ yields

$$\Delta = \ell_{123}(x^{(1)}, x^{(1)}, x^{(3)}) - 2(2x^{(1)} + x^{(3)}) + 3(2x^{(1)} + x^{(3)})$$

$$- 2\ell_{11}(x^{(1)}, x^{(1)}) - \ell_{13}(x^{(1)}, x^{(3)})$$

$$= x^{(1)} + x^{(3)} - 2(2x^{(1)} + x^{(3)}) + 3(2x^{(1)} + x^{(3)}) - x^{(1)} - 2(x^{(1)} + x^{(3)})$$

$$= 0.$$

Hence, we have tail dependence in $\mathbb{X}$ and Δ is zero as extreme events in dimension three do not matter.

Proof of proposition 2.2 The result directly follows from Einmahl et al. (2012), theorem 4.6, and Bücher & Dette (2013), Bücher et al. (2014) $\sqrt{k}(\hat{\ell}(\mathbf{x}) - \ell_{\mathbb{X}}(\mathbf{x}), \mathbf{x} \in [0,1]^d$, is asymptotic normal with zero mean and covariance matrix equal to a sum of a centered Gaussian field and Gaussian processes. It is assumed that $\ell_{\mathbb{X}}(\mathbf{x}) < \mathbf{x}'\mathbf{1}$ to ensure the asymptotic variance of $\hat{\ell}_{\mathbb{X}}(\mathbf{x})$ is non–zero. This holds if at least one bivariate pair $(X^{(i)}, X^{(j)})$ is asymptotic dependent. In $\mathbb{R}^2$,

where $\mathbf{x} = (x^{(i)}, x^{(j)})$, it holds that

$$\sqrt{k}\hat{\ell}_{ij}(x^{(i)}, x^{(j)}) \xrightarrow{d} N(\ell(x^{(i)}, x^{(j)}), \sigma_\ell^2), x^{(i)}, x^{(j)} > 0,$$

where

$$\sigma_\ell^2 = \ell(x^{(i)}, x^{(j)}) - 2x^{(i)}\ell_{\partial i}(x^{(i)}, x^{(j)}) - 2x^{(j)}\ell_{\partial j}(x^{(i)}, x^{(j)}) + x^{(i)}\ell_{\partial i}^2(x^{(i)}, x^{(j)})$$
$$+ x^{(j)}\ell_{\partial j}^2(x^{(i)}, x^{(j)}) + 2\ell_{\partial i}(x^{(i)}, x^{(j)})\ell_{\partial j}(x^{(i)}, x^{(j)})(x^{(i)} + x^{(j)} - \ell(x^{(i)}, x^{(j)})),$$

with $\ell_{\partial j}(\mathbf{x}) := (\partial\ell/\partial x^{(j)})(\mathbf{x})$ denoting the partial derivative of the STDF with respect to argument $x^{(j)}$. According to equations (5) and (6), and setting $\mathbf{x} = \mathbf{1}$, $\widehat{R}(\mathbf{x})$ is also asymptotic normal. Asymptotic normality of $\widehat{\Delta}$ directly follows from equation (7). Thus,

$$\widehat{\Delta} \xrightarrow{d} N(\Delta, \sigma_{\widehat{\Delta}}^2),$$

with

$$\sigma_{\widehat{\Delta}}^2 = k^{-1}\sigma_{\hat{\ell}}^2 + k^{-1}\sum_{i<j\leq 2}\sigma_{\hat{\ell}_i}^2 + 2\Big(\sum_{i<j\leq 2}Cov(\hat{\ell}_i, \hat{\ell}) + \sum_{i<j\leq 2; g<h\leq 2; i\neq g}Cov(\hat{\ell}_{ij}, \hat{\ell}_{gh})\Big) \in (0, \infty).$$

Whenever partial derivatives of the STDF do not exist, the same reasoning for the limit law of $\sqrt{k}\widehat{\Delta}$ applies using asymptotic results in Bücher et al. (2014).

6.3 Auxiliary simulations

The ratio κ_3/κ_2 gives the share of bivariate extremes that are also extremes in dimension three or larger; as this ratio conditions on the occurrence of bivariate extremes, the magnitude is driven by multivariate ($d > 2$) tails and is not driven by the number of bivariate extremes, as is the case for κ_3. Table 5 reports averages from 1000 simulation repetitions of all three measures for the distributions considered in the simulations of Section (3). Sample size, dimension and choice of k are as in the empirical application of Chapter (4).

Note, the only distribution in dimension 7 that fulfills the null of no HOTDs is the max factor model with loading matrix B_7^0. In this case, both κ_3 and κ_3/κ_2 are close to zero. Theoretically, they should be exactly zero, however, for a sample size of $n = 750$ this distortion can be be interpreted as finite sample bias. Yet in this case, a simple t-test would not indicate a statistical significance ($\alpha = 0.05$). For the meta t-distribution, κ_3/κ_2 grows with decreasing degree of freedom of the copula, which governs the strength of bivariate and multivariate extremes. Thus, $\hat{\kappa}_3/\hat{\kappa}_2$ is indeed capable of reflecting the severeness of HOTDs.

Table 5: Means and standard deviations of simulated $\hat{\kappa}_3/\hat{\kappa}_2$, $\hat{\kappa}_3$, $\hat{\kappa}_2$ for max factor models and meta t-distributions from Section 3 with 1000 repetitions, $d = 7$, $n = 750$ and k as in the empirical application.

		$\overline{\hat{\kappa}_3/\hat{\kappa}_2}$	$\overline{\hat{\kappa}_3}$	$\overline{\hat{\kappa}_2}$
t-distr.				
	ν_C			
	5	0.507	0.212	0.419
		(0.070)	(0.040)	(0.046)
	10	0.453	0.168	0.371
		(0.071)	(0.036)	(0.042)
	15	0.431	0.151	0.351
		(0.072)	(0.033)	(0.041)
	20	0.425	0.145	0.344
		(0.072)	(0.032)	(0.040)
max factor				
	B_7^0	0.050	0.040	0.795
		(0.027)	(0.022)	(0.048)
	B_7^{A1}	0.590	0.500	0.852
		(0.063)	(0.035)	(0.033)
	B_7^{A2}	0.558	0.534	0.958
		(0.062)	(0.053)	(0.028)
	B_7^{A3}	0.517	0.328	0.633
		(0.019)	(0.028)	(0.048)

References

Ang, A. & Chen, J. (2002), 'Asymmetric correlations of equity portfolios', *Journal of Financial Economics* **63**, 443–494.

Beirlant, J., Goegebeur, J., Teugels, Y. & Segers, J. (2004), *Statistics of extremes*, Wiley, Chichester.

Berg, D. & Quessy, J. F. (2009), 'Local sensitivity analyses of goodness–of–fit tests for copulas', *Scandinavian Journal of Statistics* **36**, 389–412.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **31**, 307–327.

Brunnermeier, M. & Pedersen, L. (2009), 'Market liquidity and funding liquidity', *Review of Financial Studies* **22**, 2201ï£¡2238.

Bücher, A. & Dette, H. (2013), 'Multiplier bootstrap of tail copulas with applications', *Bernoulli* **5**, 1655–1687.

Bücher, A., Segers, J. & Volgushevi, S. (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.

Chen, X. & Fan, Y. (2006), 'Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification', *Journal of Econometrics* **135 (1)**, 125–154.

Chollete, L., de la Peña, V. & Ching-Chih, L. (2011), 'International diversification: A copula approach', *Journal of Banking and Finance* **35**, 403–417.

Christoffersen, P., Errunza, V., Jacobs, K. & Langlois, H. (2012), 'Is the potential for international diversification disappearing? a dynamic copula approach', *The Review of Financial Studies* **25**(12), 3711–3751.

Coles, S., Heffernan, J. & Tawn, J. (1991), 'Modelling extreme multivariate events', *Journal of the Royal Statistical Society: Series B* **53**, 377–392.

Coles, S., Heffernan, J. & Tawn, J. (1998), 'Dependence measures for extreme value analyses', *Extremes* **1**, 7–45.

Danielsson, J., de Haan, L., Peng, L. & de Vries, C. (2001), 'Using a bootstrap method to choose the sample fraction in tail index estimation', *Journal of Multivariate Analysis* **76**, 226–248.

de Haan, L. & de Ronde, J. (1998), 'Sea and wind: Multivariate extremes at work', *Extremes* **1**, 7–45.

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

de Haan, L., Neves, C. & Peng, L. (2008), 'Parametric tail copula estimation and model testing', *Journal of Multivariate Analysis* **99**, 1260–1275.

de Haan, L. & Resnick, S. I. (1977), 'Limit theory for multivariate sample extremes', *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **40**, 317–337.

Demarta, S. & McNeil, A. J. (2005), 'The *t* copula and related copulas', *International Statistic Review* **73**, 111–129.

Dietrich, D., de Haan, L. & Hüsler, J. (2003), 'Testing extreme value conditions', *Extremes* **5**, 71–85.

Draisma, G., Drees, H., Ferreira, A. & de Haan, L. (2004), 'Bivariate tail estimation: Dependence in asymptotic independence', *Bernoulli* **10(2)**, 251–280.

Drees, H., de Haan, L. & Li, D. (2006), 'Approximations to the tail empirical distribution function with application to testing extreme value conditions', *Journal of Statistical Planning and Inference* **136**, 3498–3538.

Einmahl, J. H. J., de Haan, L. & Li, D. (2006), 'Weighted approximations of tail copula processes with application to testing the bivariate extreme value condition', *The Annals of Statistics* **34**, 1987–2014.

Einmahl, J. H. J., Krajina, A. & Segers, J. (2008), 'A method of moments estimator of tail dependence', *Bernoulli* **14(4)**, 1003–1026.

Einmahl, J. H. J., Krajina, A. & Segers, J. (2012), 'An m–estimator for tail dependence in arbitrary dimensions', *The Annals of Statistics* **40(3)**, 1754–1793.

El-Nouty, C. & Guillou, A. (2000), 'On the bootstrap accuracy of the pareto index', *Statistics and Decisions* **18**, 275–289.

Embrechts, P. (2009), 'Linear correlation and evt: Properties and caveats', *Journal of Financial Econometrics* **7**, 30–39.

Fama, E. F. & French, K. R. (1992), 'The cross–section of expected stock returns', *Journal of Finance* **47**, 427–486.

Fisher, R. A. & Tippett, L. H. C. (1928), *Limiting forms of the frequency distribution of the largest or smallest member of a sample*, Proceedings of the Cambridge Philosophical Society. 180–190.

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

Fréchet, M. (1927), 'Sur la loi de probabilité de l' écart maximum', *Annales de la Société Polonaise de Mathématique* **6**, 92–116.

Geluk, J. & de Haan, L. (2002), 'On bootstrap sample size in extreme value theory', *Publ. de l Inst. Math. Nouvelle serie 71* **71(85)**. 21–25.

Ghosh, S. (2010), 'Modelling bivariate rainfall distribution and generating bivariate correlated rainfall data in neighbouring meteorological subdivisions using copula', *Hydrological Processes* **24**, 3558–3567.

Glosten, L. R., Jagannathan, R. & Runkle, D. E. (1993), 'On the relation between the expected value and the volatility of the nominal excess return on stocks', *The Journal of Finance* **48(5)**(5), 1779–1801.

Gnedenko, B. V. (1943), 'Sur la distribution limite du terme maximum une série aléatoire', *Annals of Mathematics* **44**, 423–453.

Guillotte, S., F. Perron, F. & Segers, J. (2011), 'Non-parametric bayesian inference on bivariate extremes', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**(3), 377–406.

Hall, P., Horowitz, J. L. & Jing, B. Y. (1995), 'On blocking rules for the bootstrap with dependent data', *Biometrika* **82(3)**, 561–574.

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

Hua, L. & Joe, H. (2011), 'Second order regular variation and conditional tail expectation of multiple risks', *Insurance: Mathematics and Economics* **49(3)**, 537âĂŞ546.

Huang, X. (1992), Statistics of bivariate extreme values, PhD thesis, Erasmus University Rotterdam, The Netherlands.

Hüsler, J. & Li, D. (2009), 'Testing asymptotic independence in bivariate extremes', *Journal of Statistical Planning and Inference* **139**, 990–998.

Joe, H. (1997), *Multivariate models and dependence concepts*, Chapman and Hall, London.

Joe, H., Smith, R. L. & Weissman, I. (1991), 'Bivariate threshold methods for extremes', *Journal of the Royal Statistical Society: Series B* **54**, 171–183.

Klugman, S. & Parsa, R. (1999), 'Fitting bivariate loss distributions with copulas', *Insurance: Mathematics and Economics* **24**, 139–148.

Kojadinovic, I. & Yan, J. (2010), 'Nonparametric rank–based tests of bivariate extreme–value dependence', *Journal of Multivariate Analysis* **101**, 2234–2249.

Li, F. (2013), 'Idenifying asymmetric comovements of international stock market returns', *Journal of Financial Econometrics, forthcoming* .

Longstaff, F. A. (2010), 'The subprime credit crisis and contagion in financial markets', *Journal of Financial Economics* **97**, 436ï£¡450.

McNeil, A. & Frey, R. (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., Frey, R. & Embrechts, P. (2006), *Quantitative risk management: Concepts, techniques and tools*, Princeton University Press, Princeton.

Mikosch, T. (2006), 'Copulas: Tales and facts', *Extremes* **9**, 3–20.

Nikoloulopoulos, A. K., Joe, H. & Li, H. (2009), 'Extreme value properties of multivariate t copulas', *Extremes* **12**, 129–148.

Oh, D. H. & Patton, A. (2013), 'Simulated method of moments estimation for copula-based multivariate models', *Journal of the American Statistical Association* **108(502)**, 689–700.

Peng, L. (2010), 'A practical way for estimating tail dependence functions', *Statistica Sinica* **20**, 365–378.

Peng, L. & Qi, Y. (2006), 'Partial derivatives and confidence intervals of bivariate tail dependence functions', *Journal of Statistical Planning and Inference* **137**, 2089–2101.

Politis, D. N. & Romano, J. P. (1994), 'Large sample confidence regions based on subsamples under minimal assumptions', *Annals of Statistics* **22**, 2031–2050.

Poon, S. H., Rockinger, M. & Tawn, J. (2004), 'Extreme value dependence in financial markets: Diagnostics, models and financial implications', *The Review of Financial Studies* **17(2)**, 581–610.

Qi, Y. (2008), 'Bootstrap and empirical likelihood methods in extremes', *Extremes* **11**, 81–97.

Rémillard, B. (2010), 'Goodness-of-fit tests for copulas of multivariate time series', *Available at SSRN: http://ssrn.com/abstract=1729982* .

Resnick, S. & de Haan, L. (1996), 'Second-order regular variation and rates of convergence in extreme-value theory', *The Annals of Probability* **24**, 97–124.

Resnick, S. I. (1987), *Extreme values, regular variation, and point processes*, Springer, Berlin Heidelberg New York.

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

Schmidt, R. & Stadtmüller, U. (2006), 'Nonparametric estimation of tail dependence', *The Scandinavian Journal of Statistics* **2**, 307–335.

Segers, J. (2012), 'Max–stable models for multivariate extremes', *Revstat – Statistical Journal* **10**, 61–82.

Sibuya, M. (1960), 'Bivariate extreme statistics', *Annals of the Institute of Statistical Mathematics* **11**, 195–210.

Straetmans, S. T. M., Verschoor, W. F. C. & Wolff, C. C. P. (2008), 'Extreme us stock market fluctuations in the wake of 9/11', *Journal of Applied Econometrics* **23**, 17–42.

Working Paper Series in Economics