Multivariate Extremes in Financial Markets: New Statistical Testing Methods and Applications

Zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften
(Dr. rer. pol.)

bei der Fakultät für Wirtschaftswissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte Dissertation

von

M.Sc. Carsten Bormann

Tag der mündlichen Prüfung: 08.02.2017

Referentin: Prof. Dr. Melanie Schienle
Koreferent: Prof. Dr. Christian Conrad

Acknowledgments

I would like to express my deep gratitude to my supervisor Melanie Schienle for enabling my doctoral studies. I enjoyed countless insightful discussions that always refined, and furthered my research ideas. I immensely benefited from the ideal working conditions she provided which came along with great degrees of freedoms in every aspect.

Also, I would like to thank Christian Conrad for agreeing to co-examine this thesis. This is especially delightful to me as he also supervised my Bachelor thesis in Heidelberg which back then manifested my interest in quantitative and econometric research problems.

I would also like to thank Julia Schaumburg for introducing me to the world of extreme values, for joint collaboration on multivariate extremes, and for very helpful comments and joyful discussions.

Finally, I would like to thank my colleagues at the Institute of Econometrics and Statistics as it has always been a pleasure to work in such an enjoyable environment.

Contents

1 Introduction

During the last decades, financial extreme events have substantially gained the attention of academics, financial practitioners and the general public, e.g. Embrechts et al. (1999), Poon et al. (2004), Taleb (2007). Financial regulators have recognized the importance of unforeseeable financial crashes to the stability of the financial system: As is demanded in the Third Basel Accord, banks should "explicitly consider extreme events in stress testing", (BIS 2010, p. 47). Prominent examples of financial extremes are the Black Tuesday 1929, the Black Monday 1987, and, recently, the turmoils of the Subprime Crises starting October 2008. These crises are initiated and characterized by extreme price movements — a first step in understanding, cushioning and even predicting extreme risk in financial markets is a thorough quantification thereof. Statistically, an event is extreme if it is so rare that one cannot expect to have witnessed its occurrence in a given sample. Extreme value statistics provides tools which can, nonetheless, consistently quantify probabilities of such unforeseen events. The usefulness of one–dimensional extreme value statistics in quantitative finance is well documented, in particular for the estimation of extreme risk measures such as the Value at Risk, see McNeil & Frey (2000). Multi–dimensional extreme value methods measure the dependence between occurrences of one–dimensional extremes, and, ultimately, can approximate probabilities of multivariate extreme events. Yet, the scarcity of relevant extremal observations is a statistical challenge that complicates robust estimation of multivariate extremes. Multivariate extreme value methods, however, have only recently found their way in financial risk management, see Poon et al. (2004), Straetmans et al. (2008).

This thesis contributes to the statistical assessment of multivariate (financial) extreme events. We contribute to this venture by proposing four statistical tools that reveal risk patterns that were so far ignored, misjudged or unheard of in quantitative risk management. Our new statistical tests improve the understanding of linkages between at least two extremes. We study their theoretical properties, verify them in practically relevant simulation experiments, and, finally, apply the tests to real financial data. Empirical applications unravel specific data anomalies that were unidentified, or, at least, underestimated as of yet. To attend this matter, we employ and enhance methods from multivariate extreme value statistics. While our applications focus on dependencies of extremes in financial markets, the proposed tests are not limited to finance, but can also be applied in other fields where joint extreme events are of interest, i.e. as hydrology, meteorology, oceanography, or engineering.

The particular contributions of this thesis are as follows. In Chapter 2, we develop a statistical test for the question whether the standard pairwise approach to measure dependence of extreme events is appropriate in high–dimensional settings. Typically, to assess the dependence between several extremes — so–called *tail dependence* — of a cross–section of financial assets, only tail dependencies for all two–dimensional pairs are quantified. Our test investigates whether this approach is valid, or if one underestimates the risks of joint occurrences of three or more extreme events. We term extreme events in dimension three or higher *higher order tail dependencies* (HOTDs). Statistical properties of the test are discussed and validated in a Monte Carlo simulation study. We identify satisfying test properties both for i.i.d. data and serially dependent time series data. In an empirical application with international stock market indices, we find that a solely pairwise approach mostly ignores a substantial part of tail risk, and that this share has steadily increased during the last two decades. While HOTDs occur less frequently on a global level, tail dependence of European financial markets is substantially characterized by HOTDs. This implies fewer diversification opportunities of purely European portfolios and the need for truly multi–dimensional tail models.

The test in Chapter 3 compares dependence between two two–dimensional tails, e.g. tail dependence of joint losses versus joint gains within a two–dimensional portfolio, or joint losses (gains) between two two–dimensional portfolios. Such extreme risk comparisons are of particular importance for efficient asset allocation which heavily depends on optimal risk diversification. While there already exist methods to address this question, we find our test approach to be slightly more powerful in standard situations. For non–standard, asymmetric types of tail dependence, our test is outstandingly more powerful as competing tests oversimplify tail structures, and omit asymmetries. Importantly, our test localizes sample events for which both tails differ the most. Also, our test is tailored to financial data, which is often serially dependent. Test properties are derived and also verified in a Monte Carlo simulation. Studying pairs of S&P500 industries and foreign exchange rates, we establish tail asymmetry is more severe than expected, and that non–standard tail dependence causes competing tests to misjudge the degree of tail inequality. In particular, the most common test type, based on the scalar–value tail dependence coefficient, misses up to 20% of tail dependence differences solely due to non–standard tail events.

The test type in Chapter 4 studies if a single two–dimensional tail is symmetric with respect to its one–dimensional components, i.e. we test against so–called intra–tail (a)symmetry. We propose a non–parametric and a parametric testing approach. Again, test properties are provided and validated in a Monte Carlo simulation study. For the most relevant foreign exchange rate pairs, we find time periods where nearly 20% of bivariate tails are indeed intra–tail asymmetric, i.e. of non–standard type. This implies, that in such cases, standard dependence models, such as symmetric copulas, are inconsistent. As a result, by pre–testing against asymmetries, our test can improve parametric risk modeling.

Lastly, Chapter 5 extends the ideas of Chapter 3 to comparisons of entire dependence structures of bivariate distributions. Theoretical test properties are illustrated in a Monte Carlo Simulation. Due to its computational simplicity, this test is especially suited for massive data sets. We study high–frequency return pairs with respect to the dynamics of their dependence structure. By localizing sample events where dependence typically changes over time, we find that current ways of modeling time variation in dependence structures do not account for the most time sensitive parts of the dependence.

Chapter 2 is joint work with Melanie Schienle and Julia Schaumburg and has been published in the Journal of Financial Econometrics, Bormann et al. (2016). Chapter 3 is joint work with Melanie Schienle. Chapters 4 and 5 are single–authored.

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

This chapter is based on Bormann et al. (2016).

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

2.1 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 & Tawn (1991), Joe et al. (1992), de Haan et al. (2008), Guillotte et al. (2011)). Yet, for a variety of empirical settings, there are periods of 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 co–movements. 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 i.i.d. 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 from conditional heteroscedasticity 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 lower bivariate tails, i.e. bivariate 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 this chapter is organized as follows. Section 2.2 discusses necessary concepts from multivariate extreme value theory. Section 2.3 introduces and formalizes test idea, test asymptotics and finite sample implementation. Finite sample properties are studied in Section 2.4. Section 2.5 studies HOTDs between international stock indices. Section 2.6 concludes. The Appendix contains supplementary and theoretical results.

2.2 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 i.i.d. 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

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\max(\mathbf{X}_n^{(1)}) - b_n^{(1)}}{a_n^{(1)}} \leq x^{(1)}, ..., \frac{\max(\mathbf{X}_n^{(d)}) - b_n^{(d)}}{a_n^{(d)}} \leq x^{(d)}\right) = G_{\mathbb{X}}(\mathbf{x}), \tag{2.1}$$

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, we assume that they are fulfilled. In general, closed–form expressions for $G_{\mathbb{X}}(\mathbf{x})$ do not exist. Equation (2.1) can be written as

$$\lim_{n\to\infty} F_{\mathbb{X}}^n(a_n^{(1)}x^{(1)} + b_n^{(1)}, ..., a_n^{(d)}x^{(d)} + b_n^{(d)}) = G_{\mathbb{X}}(\mathbf{x}), \tag{2.2}$$

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

$$\lim_{n\to\infty} F_j^n(a_n^{(j)}x^{(j)} + b_n^{(j)}) = G_j(x^{(j)}) = \exp\left(-(1+\gamma^{(j)}x^{(j)})^{-1/\gamma^{(j)}}\right), j = 1, ..., d, \tag{2.3}$$

where $\gamma^{(j)}$ denotes the tail index (shape parameter) of margin j (Fréchet (1927), Fisher & Tippett (1928), Gnedenko (1943)). An equivalent formulation of Equation (2.2), with (2.3) 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 $\ell(\mathbf{x})$, see 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. One can express $\ell(\mathbf{x})$ as

$$\ell(\mathbf{x}) = \lim_{t\to 0} t^{-1}\mathbb{P}\Big(\bigcup_{i=1}^{d}\{F_i^{-1}(1-tx^{(i)}) \le X^{(i)}\}\Big), t \in \mathbb{R}_+. \tag{2.4}$$

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. (2002), 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)). 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. Tail (in)dependence is often also denoted as asymptotic (in)dependence. 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 which is a closely related metric for tail dependence. The (upper) tail copula of $\mathbb{X}$ is defined as

$$\Lambda(\mathbf{x}) = \lim_{t\to 0} t^{-1}\mathbb{P}\Big(\bigcap_{i=1}^{d}\{F_i^{-1}(1-tx^{(i)}) \le X^{(i)}\}\Big). \tag{2.5}$$

It only considers joint exceedances to characterize tail dependence, see Schmidt & Stadtmüller (2006). Sibuya (1960), Joe (2001) and Coles et al. (1999) 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 tail copula 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 tail copula for our purpose: Firstly, the tail copula captures only (the most extreme) parts of the multivariate tail dependence in dimensions $d > 2$, while the STDF completely describes it (see Subsection 2.3 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 tail copula 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. Conversely, 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 $\Lambda(\mathbf{x})$ 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.

2.3 A new test for higher order tail dependence

2.3.1 Test idea and asymptotic properties

We aim to detect the share which HOTDs contribute to overall tail dependence. Hence, we decompose the STDF for dimension d into tail copulas for dimensions two to d. In dimension $d = 2$, from Equation (2.4) 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 $\ell(x^{(1)}, x^{(2)}) = x^{(1)} + x^{(2)} - \Lambda(x^{(1)}, x^{(2)})$. 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

$$\ell(\mathbf{x}) = \sum_{i=1}^{d} x^{(i)} - \sum_{i<j\leq d} \Lambda_{ij}(\mathbf{x}^{(i,j)}) + \underbrace{\sum_{h<i<j\leq d} \Lambda_{hij}(\mathbf{x}^{(h,i,j)}) - ... + (-1)^{d+1}\Lambda(\mathbf{x})}_{=:\mathcal{A}}, \tag{2.6}$$

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 (2.6), $\mathcal{A}$ equals zero. In this case, higher dimensional joint extremes are irrelevant. When substituting $\Lambda_{ij}(x^{(i)}, x^{(j)}) = x^{(i)} + x^{(j)} - \ell_{ij}(x^{(i)}, x^{(j)}), i < j \leq d$, Equation (2.6) yields

$$\ell(\mathbf{x}) = (2-d)\sum_{i=1}^{d} x^{(i)} + \sum_{i<j\leq d} \ell_{ij}(\mathbf{x}^{(i,j)}) + \mathcal{A}, \tag{2.7}$$

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

Using Equation (2.7) 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

$$\Delta := \ell(\mathbf{x}) - (2-d)\sum_{i=1}^{d} x^{(i)} - \sum_{i<j\leq d} \ell_{ij}(\mathbf{x}^{(i,j)}) = 0. \tag{2.8}$$

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

$$H_0 : \Delta = 0. \tag{2.9}$$

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,[1] 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 2.4.

[1] I.e. if all bivariate tails of $\mathbb{X}$ are tail independent.

In order to apply the test, we have to estimate the STDF of $\mathbb{X}$, $\ell_{\mathbb{X}}$, and the STDFs for bivariate pairs, ℓ_{ij}. Let $X^{(i)}_{n:m}$ denote the m–th largest order statistic of margin $X^{(i)}$, and let $1(C)$ be the indicator function for event C. In Equation (2.4), replacing the running variable t by k/n and the extreme quantiles $F_i^{-1}(1-tx^{(i)})$ by $X^{(i)}_{n:n+0.5-kx^{(i)}}$ we use the following non–parametric estimator for the STDF (see Huang (1992) and Einmahl et al. (2012))

$$\widehat{\ell}(\mathbf{x}) = \frac{1}{k}\sum_{i=1}^{n} 1\Big\{\bigcup_{j=1}^{d}\{X_i^{(j)} \geq X^{(j)}_{n:n+0.5-kx^{(j)}}\}\Big\}, n \to \infty, k \to \infty, \frac{k}{n} \to 0, \tag{2.10}$$

$\mathbf{x} = (x^{(1)}, ..., x^{(d)})$. Under some technical conditions, the empirical process $\sqrt{k}(\widehat{\ell}(\mathbf{x}) - \ell(\mathbf{x}))$ 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, $\widehat{\ell}(\mathbf{x})$ is still asymptotically normal but with degenerate variance (Hüsler & Li 2009). Note, $\widehat{\ell}(\mathbf{x})$ 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 $X^{(i)}_{n:n+0.5-k}$. The asymptotic distribution of $\widehat{\ell}(\mathbf{1})$ simplifies to

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

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

$$\widehat{\Delta} := \widehat{\ell}(\mathbf{1}) - 2d + d^2 - \sum_{i<j\leq d} \widehat{\ell}_{ij}\left(\mathbf{1}\right). \tag{2.11}$$

These considerations lead us to the asymptotic distribution of the test statistic, which is stated after stating necessary assumptions.

Assumptions 2.1. *Assume $F_{\mathbb{X}} \in D(G_{\mathbb{X}})$. Furthermore, let the following assumptions hold:*

($A1^{\Delta}$) 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 $\mathbb{R}^d$.

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

Proposition 2.2. *Under ($A1^{\Delta}$) and ($A2^{\Delta}$), it holds that*

$$\sqrt{k}(\widehat{\Delta} - \Delta) \xrightarrow{d} N(0, \sigma^2_{\widehat{\Delta}}), \tag{2.12}$$

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

$$\Big(\widehat{\ell}(\mathbf{1}), (\widehat{\ell}_{ij}(1,1))_{i<j\leq d}\Big).$$

The proof can be found in the Appendix. Assumption ($A1^{\Delta}$) 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 ($A1^{\Delta}$) guarantees that an asymptotic bias term for the left hand side of Equation (2.12) 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 and yield convergence of $\widehat{\ell}(\mathbf{x})$ in a wider sense.[2]

In both the simulation study and the empirical application in Sections 2.4 and 2.5, 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 2.4.2 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]In particular, Einmahl et al. (2012) show weak convergence of the empirical process $\sqrt{k}(\widehat{\ell}(\mathbf{x}) - \ell(\mathbf{x}))$ for bounded functions in the sup–norm, while Bücher et al. (2014) show convergence for locally bounded functions in the so–called hypi–semimetric.

2.3.2 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 $\sigma^2_{\widehat{\Delta}}$ works under milder conditions, in particular if $\mathbb{X}$ exhibits asymptotic dependence (Bücher & Dette 2013). In contrast, direct estimation of $\sigma^2_{\widehat{\Delta}}$ 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 2007). 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 $\widehat{\ell}_{\mathbb{X}}$ (see Huang (1992)), might distort the bootstrap distribution. Peng (2010) proposes 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 $\widehat{\theta}$ depends on the threshold parameter k by writing $\widehat{\theta}(k)$. 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 $\widehat{\ell}(k^*)$, and any $\widehat{\ell}_i(k^*), i \in \mathcal{I}^{(d)}_{(2)}$, to determine the full sample test statistic $\widehat{\Delta}(k^*)$ from $\mathbf{X}$.
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 $\sigma^2_{\widehat{\Delta}}$ 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 $0 < \widehat{\Delta}(k^*) + z^{\alpha}\widehat{\sigma}_{\widehat{\Delta}(k^*)}$, where z^{α}

denotes the α–quantile of the standard normal distribution.[3]

A theoretically optimal, data driven choice of the threshold parameter k should balance the bias–variance trade–off that is inherent to 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 work. 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.[4] 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 heteroscedastic financial data (McNeil & Frey (2000), Poon et al. (2004)). See Section 3.3. for details.

2.4 Simulation study

2.4.1 Size and power

In this subsection, we evaluate the empirical size and power of the test in finite samples in an i.i.d. setting. Results for time series data are presented in Subsection 2.4.3. 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 (max) factor models, which are both commonly used in financial risk management (McNeil et al. (2005),

[3]Note, a normal approximation for $\widehat{\Delta}$ is theoretically not justified under tail independence, i.e. if $\ell(\mathbf{x}) = \sum_{i=1}^{d} x^{(i)}$; then, it holds $\sigma^2_{\widehat{\ell}} = \sigma^2_{\widehat{\Delta}} = 0$ and the distributions of both $\widehat{\ell}$ 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.

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

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 2.3.2. 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 detection of HOTDs in moderate dimension is sufficient for concluding that HOTDs are significant.

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

$$C^t_{\nu,P}(\mathbf{x}) = \int_{-\infty}^{t_\nu^{-1}(u^{(1)})} \cdots \int_{-\infty}^{t_\nu^{-1}(u^{(d)})} \frac{\Gamma((\nu+d)/2)}{\Gamma(\nu/2)\sqrt{(\pi\nu)^d}|P|}(1+\nu^{-1}(\mathbf{x}'P^{-1}\mathbf{x}))^{-(\nu+d)/2} d\mathbf{x}, \quad (2.13)$$

where $t_\nu^{-1}(x^{(i)})$ 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 $\nu_C = \nu_m$.[5]

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

[5]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.

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 (ν =1). A d–dimensional max factor model for $\mathbb{X} = (X^{(1)}, ..., X^{(d)})$ is then defined by

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

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 $\Lambda_{123}(1,1,1) = 0$. 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 $d = 3$, 1.4 in $d = 5$, and 1.1 in $d = 7$.[6] 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 $X^{(i)}_{[n:n+0.5-k]}$ 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.

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 (2.1) 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 seven, the

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

Table 2.1: Empirical rejection rates: Max factor models and i.i.d. t–copula ($df = \nu_C$, $\rho = 0.5$) with t–distributed marginals ($df = 5$). Nominal test level is $\alpha = 0.05$.

	$d=3$					$d=5$					$d=7$				
	200	500	1000	1500	2000	200	500	1000	1500	2000	200	500	1000	1500	2000
$t(\nu_C)$															
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

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.

2.4.2 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 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

$$\mathcal{F}_{\mathbb{X},\lambda(n)}(\mathbf{x}) := (1-\lambda(n))F_{\mathbb{X},H0}(\mathbf{x}) + \lambda(n)F_{\mathbb{X},H1}(\mathbf{x}), \tag{2.14}$$

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 $\lambda(n) = \mathcal{O}((\sqrt{k(n)})^{-1})$, we can show that, asymptotically,

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

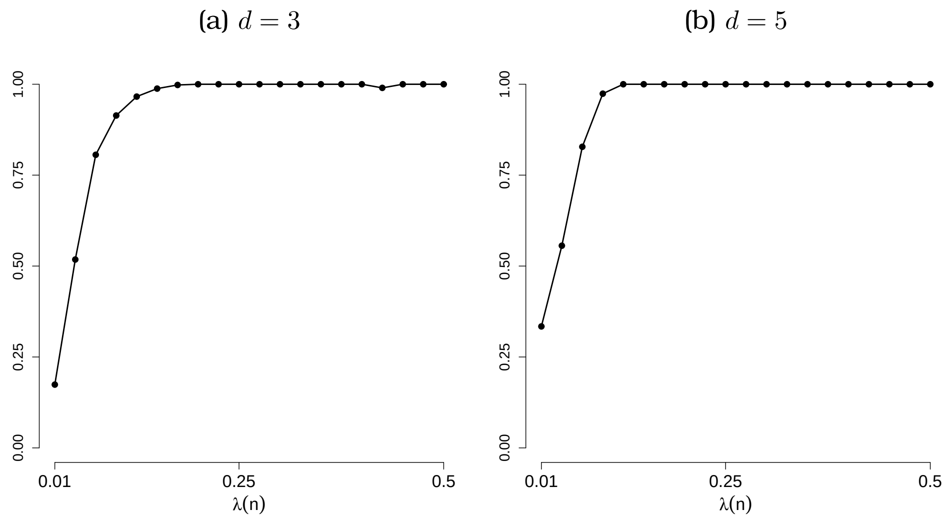
where the asymptotic variance can again be obtained analytically from theorem 4.3 in Einmahl et al. (2012). Thus, the test has power against any local alternatives if and only if these alternatives are at least of order $(\sqrt{k(n)})^{-1}$ apart from the null.

In the following simulations, we illustrate this result. Hence, we are interested in rejection rates of $\Delta = 0$ from mixture distributions defined in Equation (2.14) for $\lambda(n) := \lambda k(n)^{-1/2}$, with $0 < \lambda \leq k(n)^{-1/2}$. 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 $\widehat{p}_n = 1/S\sum_{i=1}^{S} 1\{\widehat{\Delta}_{\lambda(n)} > z^{\alpha}\widehat{\sigma}_{\widehat{\Delta},\lambda(n)}k^{-1/2}\}$ for every $\lambda(n)$. The asymptotic variance $\widehat{\sigma}^2_{\widehat{\Delta},\lambda(n)}$ is estimated by the bootstrap procedure presented in Section 2.3.2. 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 $d = 3$,

$$\mathcal{F}_{\mathbb{X},\lambda(n)}(\mathbf{x}) = (1-\lambda(n))F_{\mathbb{Y}}(\mathbf{y}) + \lambda(n)F_{\mathbb{W}}(\mathbf{w}), \tag{2.15}$$

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. For $d = 5$, we mix the distribution function of B_5^0 and $B_5^{A3} := \left(\begin{smallmatrix} 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \end{smallmatrix}\right)$. If, for example, $\lambda(n) = 0.1$, then 10% of the extreme part of the

Figure 2.1: Empirical test power for the mixture distributions defined in Equation (2.15) with sample size $n = 2000$ at level 5%.



sample should be generated by the $F_{\mathbb{W}}$ which violates the null. Figure (2.1) 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.

2.4.3 Size and power for serially dependent and conditionally heteroscedastic data

While Proposition (2.2) assumes i.i.d. data, financial time series, in particular asset returns, feature small autocorrelations and time–varying conditional volatility, and thus cannot be considered i.i.d. 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 heteroscedasticity (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 i.i.d. We therefore apply the test to these pre–filtered 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.

It is intuitively clear that $\sqrt{n}$–consistent parametric pre–filtering should not impact the consistency of the more slowly converging non–parametric estimator of the STDF. However, 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 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 2.4 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 $\mathcal{C}_\eta$:

$$\begin{aligned}
y_t^{(i)} &= \mu_t^{(i)} + \sigma_t^{(i)}\eta_t^{(i)} = \varphi_0^{(i)} + \varphi_1^{(i)} y_{t-1}^{(i)} + \sigma_t^{(i)}\eta_t^{(i)},\\
\sigma_t^{2,(i)} &= \omega^{(i)} + \alpha^{(i)}\left(y_{t-1}^{(i)} - \mu_{t-1}^{(i)}\right)^2 + \beta^{(i)}\sigma_{t-1}^{2,(i)}\\
\eta &:= (\eta^{(1)},...,\eta^{(d)}) \sim iid\ F_\eta(x^{(1)},...,x^{(d)}) = \mathcal{C}_\eta(F_{\eta,(1)}(\eta^{(1)}),...,F_{\eta,(d)}(\eta^{(d)})),
\end{aligned} \tag{2.16}$$

$t = 1,...,T$. $\theta^{(i)} = (\varphi_0^{(i)} = 0.01, \varphi_1^{(i)} = 0.05, \omega^{(i)} = 0.05, \alpha^{(i)} = 0.1, \beta^{(i)} = 0.8)$ 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 $(\widehat{\eta}_t^{*,(1)},...,\widehat{\eta}_t^{*,(d)})_{t=1}^T, \widehat{\eta}_t^{*,(i)} := (y_t^{(i)} - \widehat{\mu}_t^{(i)})/\widehat{\sigma}_t^{(i)}$. 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 i.i.d. 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.[7] 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 (2.2) reports empirical rejection probabilities for the factor copula with $\Delta = 0$; Table (2.3) 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 i.i.d. 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 the rejection probability of some parametrization that meets the null (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 i.i.d. and the non–i.i.d. 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 results.

Finally, we compare the power results obtained when simulating from the i.i.d. meta

[7] As in the i.i.d. 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 2.2: 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. Nominal test level is $\alpha = 0.05$.

	$d=3$				$d=5$				$d=7$			
	500	1000	1500	2000	500	1000	1500	2000	500	1000	1500	2000
filter 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
no filter 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

t–distribution (Table (2.1)) with those corresponding to correctly filtered, and unfiltered AR(1)–GARCH(1,1) processes connected via the meta t–distribution (Table (2.3)). In case of the correct filter, empirical power never differs significantly from the i.i.d. 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 i.i.d. data.

2.5 Higher order tail dependencies in global and European stock markets

2.5.1 Data description

In the following 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

Table 2.3: 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. Nominal test level is $\alpha = 0.05$.

	$d=3$				$d=5$				$d=7$			
	500	1000	1500	2000	500	1000	1500	2000	500	1000	1500	2000
filter 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
no filter 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

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.[8] 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).[9] The two portfolios are analyzed separately.

The sample period of the global portfolio is 01/30/1970 – 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

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

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

of the MSCI Pacific.[10] After deleting weeks with zero returns, the sample features 2335 observations for each index. The sample period of the European portfolio is 01/04/1988 – 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 in Equation (2.16) in order to eliminate the effects of serial correlation and time–varying volatility. Returns are thus modeled by

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

Standardized residuals $\widehat{\eta}_t^{*,(i)} = (y_t^{(i)} - \widehat{\mu}_t^{(i)})/\widehat{\sigma}_t^{(i)}$ 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–GARCH(1,1) (TGARCH) model, see Glosten et al. (1993).[11] 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 2.3.2 and 2.4 for details). Figures (2.2) and (2.3) show

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

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

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. [12]

2.5.2 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 2006–07, 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 09/11/2001, it is the global financial crisis of 2007–08 that marks the start of global HOTDs to become 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 an additional challenge in asset allocation on top of bivariate tail dependence nowadays. By contrast, 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.[13] Intra–European HOTDs appear to be time–varying and are most of the

[12] 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.

[13] This makes a direct comparison of test results across both data sets difficult.

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, 9/11/2001, 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.

2.5.3 Factor model for the European stock market

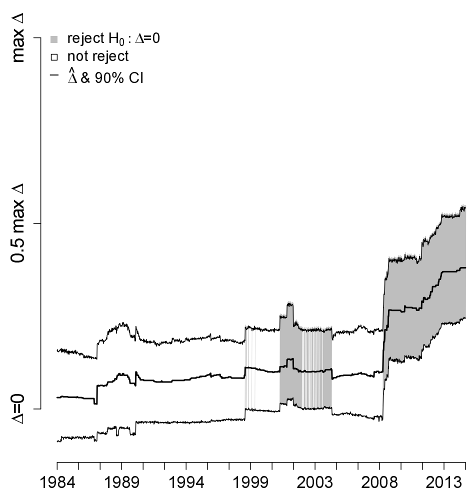
The presence of HOTDs within the European portfolio might be caused by tail events of a common external factor. To distill truly intra–European HOTDs, we now control for effects of global financial markets. Returns $y_t^{(i)}$ are thus modeled by a factor market model

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

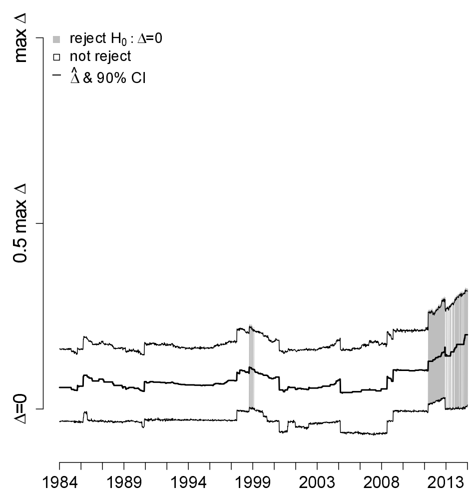
where M_t denotes a common factor for all marginal returns $y_t^{(i)}$. The disturbance $\epsilon_t^{(i)}$ is often interpreted as the idiosyncratic part of $y_t^{(i)}$. An apparent choice for M_t is the

Figure 2.2: Dynamics of the test statistic $\widehat{\Delta}$ (see Equation (2.11)), 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.

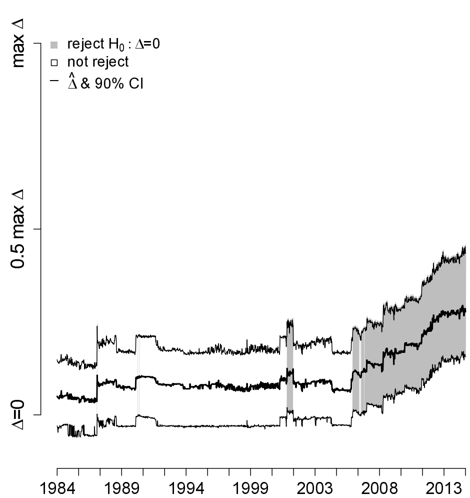
(a) Losses, unfiltered



(b) Gains, unfiltered



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



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

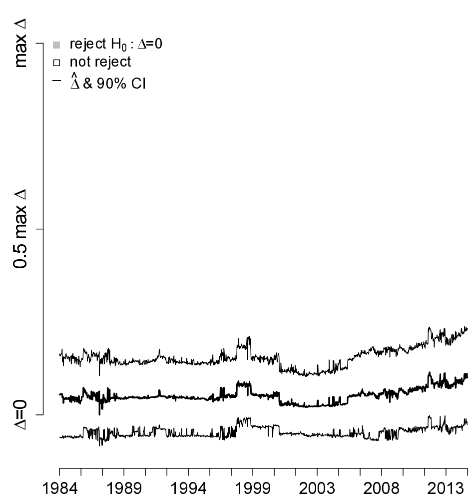
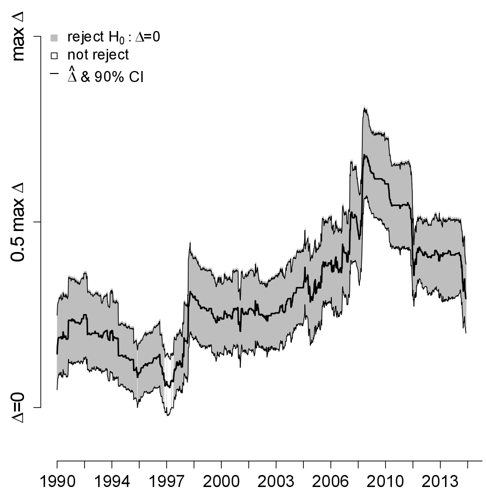
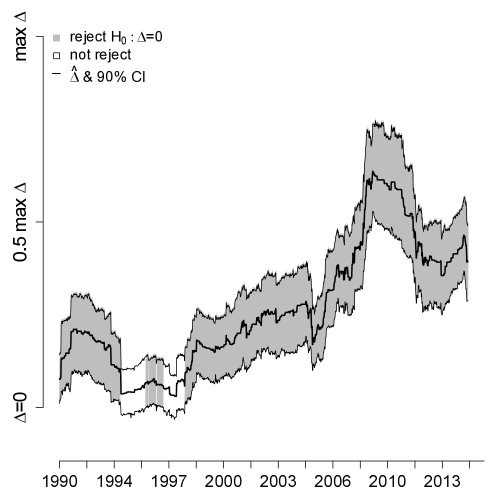


Figure 2.3: Dynamics of the test statistic $\widehat{\Delta}$ (see Equation (2.11)), together with 90% confidence intervals, for the European portfolio, using a rolling window of three to four years. The left panel shows test decisions for portfolio losses, whereas the right panel shows test decisions for portfolio gains. Confidence intervals are colored gray whenever $H_0 : \Delta = 0$ has to be rejected.

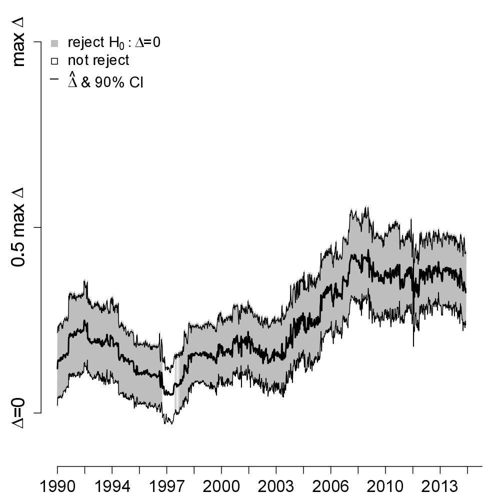
(a) Losses, unfiltered



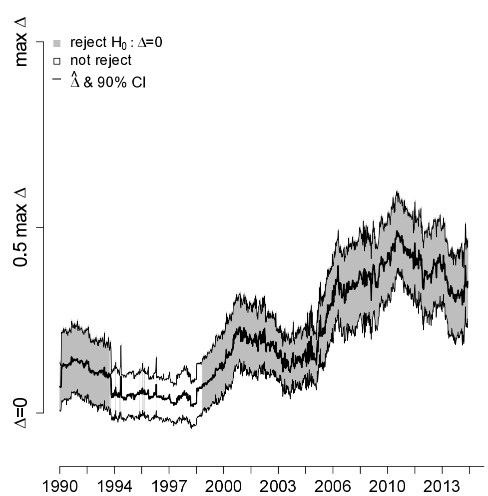
(b) Gains, unfiltered



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



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



return series of the MSCI World Ex Europe[14] 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 $(\widehat{\epsilon}_t^{(1)}, ..., \widehat{\epsilon}_t^{(7)})_{t=1}^n$. Furthermore, we 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 and the orders of the ARMA models are again found with the Schwarz criterion.[15] Test results for both gains and losses of unfiltered data and ARMA–TGARCH(1,1) filtered data are shown in Figure (2.4).

Controlling for changes in global stock markets slightly attenuates European HOTDs, yet results closely resemble the results from the previous subsection (Figure (2.3)). 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 (2.3) (d) and Figure (2.4) (d). However, the effect of the market factor to HOTDs between European gains has increased since 2006 which can be seen by comparing Figures (2.3) (b) and (2.4) (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

[14]This index runs under mnemonic MSWXEU$ in Datastream.

[15]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 040/7/1996 where the optimization of the likelihood for the TGARCH model struggles.

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.

2.5.4 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 joint extremes, κ_2 and κ_3, defined as

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

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

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,

$$\widehat{\kappa}_2 = \frac{\sum_{t=1}^n \sum_{i\neq j} \prod_{g\in\{i,j\}} 1\big\{X_t^{(g)} > X_{n:n+0.5-k}^{(g)}\big\}}{\sum_{t=1}^n 1\{\bigcup_{i=1}^d X_t^{(i)} > X_{n:n+0.5-k}^{(i)}\}}, \tag{2.17}$$

$$\widehat{\kappa}_3 = \frac{\sum_{t=1}^n \sum_{h\neq i\neq j} \prod_{g\in\{h,i,j\}} 1\big\{X_t^{(g)} > X_{n:n+0.5-k}^{(g)}\big\}}{\sum_{t=1}^n 1\{\bigcup_{i=1}^d X_t^{(i)} > X_{n:n+0.5-k}^{(i)}\}}. \tag{2.18}$$

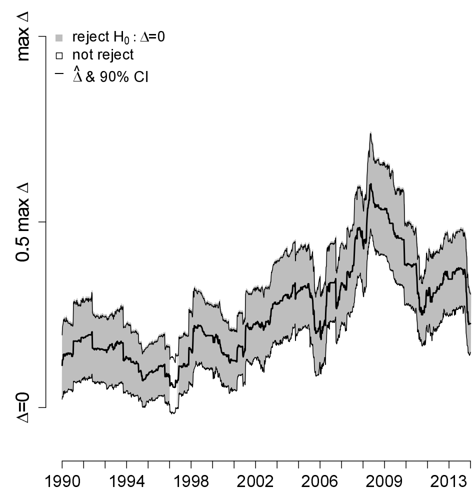
Section 2.7.3 in the Appendix provides a small simulation study that shows κ_3/κ_2 is indeed a reasonable measure for determining the severeness of HOTDs.

Figure (2.5) shows estimates $\widehat{\kappa}_2, \widehat{\kappa}_3$ and $\widehat{\kappa}_3/\widehat{\kappa}_2$ 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 $k \in [0.01n, c\sqrt{n}]$.

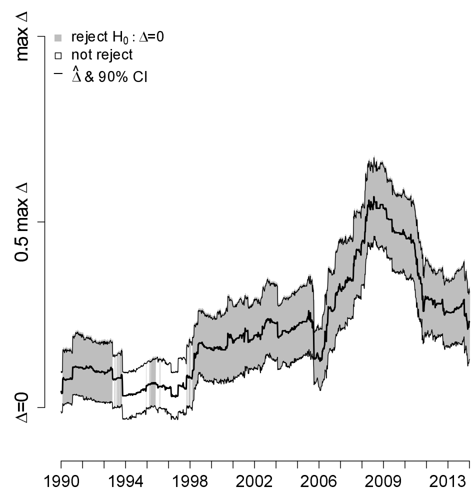
Not surprisingly, trajectories resemble the dynamics of $\widehat{\Delta}$ (Figures (2.3) (c–d)). The

Figure 2.4: Dynamics of the test statistic $\widehat{\Delta}$ (see Equation (2.11)), together with 90% confidence intervals, for the European portfolio, using a rolling window of three to four 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.

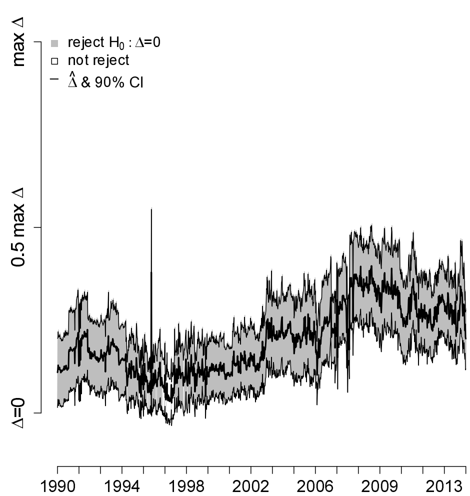
(a) Losses, unfiltered



(b) Gains, unfiltered



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



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

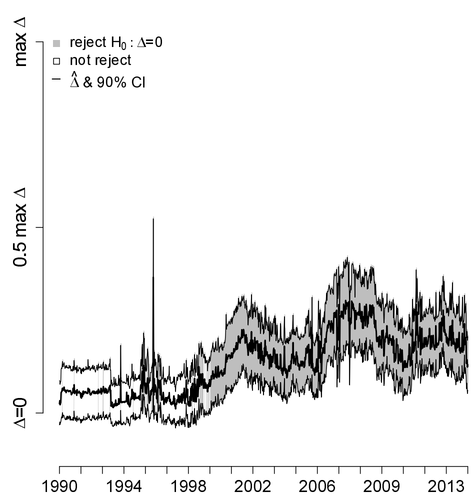
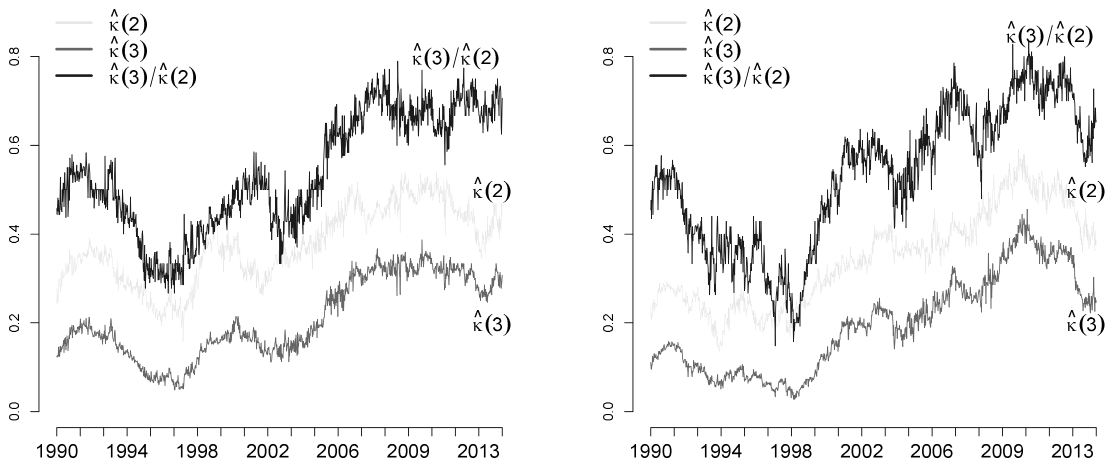


Figure 2.5: Dynamics of $\widehat{\kappa}_2, \widehat{\kappa}_3$, and $\widehat{\kappa}_3/\widehat{\kappa}_2$ (see Equations (2.17) and (2.18)) for the TGARCH filtered European portfolio.

(a) Losses (b) Gains



probability of observing trivariate extremes ($\widehat{\kappa}_3$) 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 2007–09. However, the share of trivariate extremes in bivariate extremes $\widehat{\kappa}_3/\widehat{\kappa}_2$ 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.

2.6 Conclusion

This chapter 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 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.

2.7 Appendix

2.7.1 Model specifications

Table 2.4: Specifications of the max factor models.

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

$$\boldsymbol{B}_{5}^{0} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 \\ 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix} \qquad \boldsymbol{B}_{5}^{A1} = \begin{pmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \end{pmatrix}$$

$$\boldsymbol{B}_{5}^{A2} = \begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 \\ 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$

$$\boldsymbol{B}_{7}^{0} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 1/2 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 1/2 \\ 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix} \qquad \boldsymbol{B}_{7}^{A1} = \begin{pmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 1/3 & 0 & 0 \\ 0 & 1/3 & 0 & 0 & 0 & 1/3 & 1/3 \end{pmatrix}$$

$$\boldsymbol{B}_{7}^{A2} = \begin{pmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix} \boldsymbol{B}_{7}^{A3} = \begin{pmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 1/2 \\ 0 & 1/2 & 0 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 1/2 & 0 & 0 & 0 & 1/2 \end{pmatrix}$$

2.7.2 Proofs

Proof of Proposition (2.1).

If $\mathbb{X}$ is tail independent, $\ell(\mathbf{x}) = \mathbf{x}\mathbf{1} \Leftrightarrow \ell_i(\mathbf{x}^{(i)}) = \mathbf{x}^{(i)}\mathbf{1}$, 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

$$\begin{aligned}
\Delta &= \ell(\mathbf{x}) - 2\sum_{i=1}^{d} x^{(i)} + d\sum_{i=1}^{d} x^{(i)} - \sum_{i<j\leq 2} \ell_i(\mathbf{x}^{(i)}) \\
&= \sum_{i=1}^{d} x^{(i)} - 2\sum_{i=1}^{d} x^{(i)} + d\sum_{i=1}^{d} x^{(i)} - \sum_{i<j\leq 2} \ell_i(\mathbf{x}^{(i)}) \\
&= \sum_{i=1}^{d} x^{(i)} - 2\sum_{i=1}^{d} x^{(i)} + d\sum_{i=1}^{d} x^{(i)} - (d-1)\sum_{i=1}^{d} x^{(i)} \\
&= 0.
\end{aligned}$$

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

$$\begin{aligned}
\ell_{123}(x^{(1)}, x^{(1)}, x^{(3)}) &= \lim_{t\downarrow 0} t\mathbb{P}\Big(\bigcup_{i\in\{1,2,3\}} \{X^{(i)} \geq F_i^{-1}(1-tx^{(i)})\}\Big) \\
&= \lim_{t\downarrow 0} t\mathbb{P}\left(\{X^{(1)} \geq F_1^{-1}(1-tx^{(1)})\} \cup \{X^{(3)} \geq F_3^{-1}(1-tx^{(3)})\}\right) \\
&= x^{(1)} + x^{(3)}.
\end{aligned}$$

Rewriting Δ yields

$$\begin{aligned}
\Delta &= \ell_{123}(x^{(1)}, x^{(1)}, x^{(3)}) - 2(2x^{(1)} + x^{(3)}) + 3(2x^{(1)} + x^{(3)}) \\
&\qquad - 2\ell_{11}(x^{(1)}, x^{(1)}) - \ell_{13}(x^{(1)}, x^{(3)}) \\
&= x^{(1)} + x^{(3)} - 2(2x^{(1)} + x^{(3)}) + 3(2x^{(1)} + x^{(3)}) - x^{(1)} - 2(x^{(1)} + x^{(3)}) \\
&= 0.
\end{aligned}$$

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

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) $\sqrt{k}(\widehat{\ell}(\mathbf{x}) - \ell_{\mathbb{X}}(\mathbf{x}), \mathbf{x} \in [0,1]^d$, 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 $\widehat{\ell}_{\mathbb{X}}(\mathbf{x})$ 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

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

where

$$\begin{aligned}\sigma^2_{\ell} &= \ell(x^{(i)}, x^{(j)}) - 2x^{(i)}\ell_{\partial i}(x^{(i)}, x^{(j)}) - 2x^{(j)}\ell_{\partial j}(x^{(i)}, x^{(j)}) + x^{(i)}\ell^2_{\partial i}(x^{(i)}, x^{(j)}) \\ &\quad + x^{(j)}\ell^2_{\partial j}(x^{(i)}, x^{(j)}) + 2\ell_{\partial i}(x^{(i)}, x^{(j)})\ell_{\partial j}(x^{(i)}, x^{(j)})(x^{(i)} + x^{(j)} - \ell(x^{(i)}, x^{(j)}),\end{aligned}$$

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 (2.6) and (2.7), and setting $\mathbf{x} = \mathbf{1}$, $\widehat{\Lambda}(\mathbf{x})$ is also asymptotic normal. Asymptotic normality of $\widehat{\Delta}$ directly follows from Equation 2.8. Thus,

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

with

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

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). □

2.7.3 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 (2.5) reports averages from 1000 simulation repetitions of all three measures for the distributions considered in the

Table 2.5: Means and standard deviations of simulated $\widehat{\kappa}_3/\widehat{\kappa}_2$, $\widehat{\kappa}_3$, $\widehat{\kappa}_2$ for max factor models and meta t–distributions from Section 2.4 with 1000 repetitions, $d = 7$, $n = 750$ and k as in the empirical application.

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

simulations of Section 2.4. Sample size, dimension and choice of k are as in the empirical application of Section 2.5.

Note, the only distribution in dimension seven that fulfills the null of no HOTDs is the max factor model with loading matrix $\boldsymbol{B_7^0}$. 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, $\widehat{\kappa}_3/\widehat{\kappa}_2$ is indeed capable of reflecting the severeness of HOTDs.

3 Detecting structural differences in tail dependence of financial time series

This chapter is based on Bormann & Schienle (2016).

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. Extensive Monte Carlo simulations demonstrate the test's superior finite–sample properties for common financial tail risk models, both in the i.i.d. and the sequentially dependent case. In a high–dimensional S&P500 industry universe, we compare tail dependence of bivariate lower and upper tails of sector returns in a rolling window scheme. For the last 90 years, 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. Finally, for major foreign exchange rates during 2001–16, we identify EUR–CHF as the most tail dependent pair in both upper and lower tails. This tail dominance prevails even after the Swiss National Bank unpegged the Franc from the Euro.

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

JEL classification: C12, C53, C58

3.1 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 & Solnik (2001), Ang & Chen (2002), Li (2013). For investment strategies, this should be taken into account by timely and adequate re–allocations of assets, e.g. profiting from arbitrage trading opportunities, and by appropriate adjustments of hedging decisions. 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. Particularly 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 & Chen (2002), Patton (2006), Chollete et al. (2011), Li (2013) 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 calls 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)).

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 & 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 & 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 & 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 & 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 S&P500 industry portfolios 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 to 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 & 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.

We also test pairs of six foreign exchange rates against tail inequalities during 2000–2016, i.e. against differences between bivariate tails of different pairs. Generally, for the entire time period, the Euro–Swiss Franc pairs stands out with the strongest tail dependence. Interestingly, this dominance appears to continue after the sudden unpegging of the Franc by the Swiss National Bank on January 2015.

This chapter is structured as follows. Section 3.2 introduces theoretical results on tail dependence necessary for the testing procedures. Section 3.3 introduces our testing technique. It also provides asymptotic properties and respective finite sample versions of the test procedures. Section 3.4 studies the finite sample performance in a thorough simulation study, and Section 3.5 provides detailed applications on sub-sectors indices of S&P500 and on data of the major foreign exchange rates. Finally, Section 3.6 concludes.

3.2 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 in de Haan & Ferreira (2006). A two–dimensional (random) return vector will be denoted by $\mathbb{X} = (X^{(1)}, X^{(2)})$. Marginal returns $X^{(i)}, i = 1, 2$, are assumed to be i.i.d. with continuous distribution $F_i(x^{(i)}), i = 1, 2$, and quantile functions F_i^{-1}. Later, we can relax the independence assumption directly, as will be discussed in Section 3.3.2.

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, 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.

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

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

i.e. the tail copula measures how likely both components jointly exceed extreme quantiles. See, among others, de Haan & Ferreira (2006), Schmidt & Stadtmüller (2006), for further 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 unless it becomes important. 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. Frahm et al. (2005), Aloui, Aïssa & Nguyen (2011), Garcia & Tsafack (2011). However, the TDC covers only a fragment of tail dependence, namely dependence between joint quantile exceedances of marginals thresholds along the line $(F_1^{-1}(1-t), F_2^{-1}(1-t)), t \to 0$. In contrast, the tail copula varies marginal thresholds as $(x^{(1)}, x^{(2)}) \in \mathbb{R}^2_+$, 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 $\Lambda_{\mathbb{X}}(\mathbf{x})$ with $\mathbf{x} \in \mathcal{S}$, where the domain of the tail copula, $\mathcal{S} := \{(x^{(1)}, x^{(2)}) : x^{(1)}, x^{(2)} \geq 0, ||\mathbf{x}|| = c\}$, is, e.g. the unit simplex hull with $|| \cdot || = || \cdot ||_1$ and $c = 1$, or the unit circle hull with $|| \cdot || = || \cdot ||_2$ and $c = 1$. Without loss of generality, we choose $\mathcal{S}$ to be the unit simplex hull. The homogeneity property prunes the relevant domain of the tail copula (i.e. from $\mathbb{R}^2_+$ to $\mathcal{S}$) and reduces computational efforts in practical implementation. The homogeneity property will lay the basis for our test. We assume the tail copulas exist and $\Lambda_{\mathbb{X}}(\mathbf{x}) > 0$, i.e. we assume tail dependent pairs because non–parametric methods are biased for $\Lambda_{\mathbb{X}}(\mathbf{x}) = 0$, see Schmidt & Stadtmüller (2006).

We are interested in comparing two tail copulas, i.e. in differences of tail copulas. To formalize the discussion about tail copula differences and special cases such as tail asymmetry, and tail inequality, we introduce the following definitions and some notation. We say two tail copulas $\Lambda_{\mathbb{X}}$ and $\Lambda_{\mathbb{Y}}, \mathbb{Y} = (Y^{(1)}, Y^{(2)})$, differ if there exists a set $I \subset \mathbb{R}^2_+$ with $\mathbb{P}(I) > 0$ such that

$$\{\Lambda_{\mathbb{X}}(x^{(1)}, x^{(2)}) \neq \Lambda_{\mathbb{Y}}(x^{(1)}, x^{(2)})\} \quad \text{or} \quad \{\Lambda_{\mathbb{X}}(x^{(1)}, x^{(2)}) \neq \Lambda_{\mathbb{Y}}(x^{(2)}, x^{(1)})\}, (x^{(1)}, x^{(2)}) \in I. \tag{3.1}$$

Note, we demand inequality over some set I, and not only at a single point $(x^{(1)}, x^{(2)}) \in \mathbb{R}^2_+$. For the homogeneity of the tail copula, it is sufficient to consider $(x^{(1)}, x^{(2)}) \in \mathcal{S}$. We write shorthand $\Lambda_{\mathbb{X}} \neq \Lambda_{\mathbb{Y}}$ for Equation (3.1). Tail asymmetry is given if two tail copulas of the same return vector differ, e.g. upper and lower tail copulas of $\mathbb{X}$ differ. To detect tail asymmetry, one should compare $\Lambda^U_{\mathbb{X}}(x^{(1)}, x^{(2)})$ with $\Lambda^L_{\mathbb{X}}(x^{(1)}, x^{(2)})$ and also with the *flipped* version $\Lambda^L_{\mathbb{X}}(x^{(2)}, x^{(1)})$. Tail inequality occurs between two return vectors, i.e. $\Lambda_{\mathbb{X}} \neq \Lambda_{\mathbb{Y}}$.

Definition 3.1 (Tail asymmetry). *A return vector $\mathbb{X}$ is tail asymmetric if $\Lambda^L_{\mathbb{X}} \neq \Lambda^U_{\mathbb{X}}$.*

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

Definition 3.2 (Tail inequality). *Return vectors $\mathbb{X}$ and $\mathbb{Y}$ exhibit tail inequality if $\Lambda^W_{\mathbb{X}} \neq \Lambda^Q_{\mathbb{Y}}$ for $W, Q = U, L$.*

The concept of tail inequality can be used to compare competing portfolios with respect to their sensitivity to extreme events. For example, $\Lambda^L_{\mathbb{X}} > \Lambda^L_{\mathbb{Y}}$ 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 $\Lambda^U_{\mathbb{X}} < \Lambda^L_{\mathbb{Y}}$, joint extreme losses in portfolio $\mathbb{Y}$ are more intertwined than joint extreme gains in $\mathbb{X}$.

One reason for tail copula differences may be non–exchangeability of at least one of the tail copulas considered. We term non–exchangeability of a tail copula intra–tail asymmetry as it refers to asymmetry of a single tail copula. A return vector $\mathbb{X}$ is intra–tail asymmetric if $\Lambda^W_{\mathbb{X}}(x^{(1)}, x^{(2)}) \neq \Lambda^W_{\mathbb{X}}(x^{(2)}, x^{(1)}), (x^{(1)}, x^{(2)}) \in \mathcal{S}, W = U, L$. Intra–tail asymmetry refers to one joint tail of $\mathbb{X}$ and occurs whenever the tail copula of that specific tail 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 3.1. *If $\Lambda_{\mathbb{X}}^{W}(x^{(1)}, x^{(2)}), W = U, L,$ 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.*[1]

Figure (3.3) illustrates this idea. If $\Lambda_{\mathbb{X}}^{W}(\mathbf{x}), W = U, L,$ is asymmetric with respect to $\mathbf{x}$, 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, e.g. Bormann (2016) detects a considerable amount of intra–tail asymmetries in foreign exchange rate pairs.

As the tail copula is the main component for our test, we sketch relevant statistical results. Non–parametric estimation of $\Lambda_{\mathbb{X}}(\mathbf{x})$ approximates marginal quantile functions $F_{i,\mathbb{X}}^{-1}, i = 1, 2,$ non–parametrically by the empirical counterpart $\widehat{F}_{i,\mathbb{X}}^{-1}, i = 1, 2$. Further, the running variable t is replaced by k/n with the sample size $n \to \infty$, and the effective sample size $k \to \infty, k \in \mathcal{O}(n)$. A consistent estimator for $\Lambda^{U}(\mathbf{x})$ is

$$\widehat{\Lambda}_{\mathbb{X}}^{U}(x^{(1)}, x^{(2)}) = \frac{1}{k}\sum_{m=1}^{n} 1\left\{X_m^{(1)} > \widehat{F}_{1,\mathbb{X}}^{-1}(1-(k/n)x^{(1)}), X_m^{(2)} > \widehat{F}_{2,\mathbb{X}}^{-1}(1-(k/n)x^{(2)})\right\},$$

$(x^{(1)}, x^{(2)}) \in \mathcal{S}$. An asymptotically equivalent estimator is given by

$$\widehat{\Lambda}_{\mathbb{X}}^{U}(x^{(1)}, x^{(2)}) = \frac{1}{k}\sum_{m=1}^{n} 1\left\{\widehat{F}_{1,\mathbb{X}}(X_m^{(1)}) > 1-(k/n)x^{(1)}, \widehat{F}_{2,\mathbb{X}}(X_m^{(2)}) > 1-(k/n)x^{(2)}\right\},$$

where $\widehat{F}_{i,\mathbb{X}}(x) = \frac{1}{n+1}\sum_{j=1}^{n} 1\{X_j^{(i)} \leq x\}$. Estimators for $\Lambda_{\mathbb{X}}^{L}(\mathbf{x})$ are defined analogously.

Concerning asymptotic results for the empirical tail copula in the standard i.i.d. case, we state both assumptions and results for the tail copula as they are the

[1] 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 (3.1) applies.

backbone of the asymptotic distribution of our test statistic, see Bücher & Dette (2013).

Assumptions 3.1. *For a bivariate random vector $\mathbb{X}$, we assume the following.*

($A1^S$) $\mathbb{X} \sim F_{\mathbb{X}}$, *i.i.d.*

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

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

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

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

Assumption ($A1^S$) is standard, yet restrictive for financial time series. In practice, the necessity of i.i.d. data is bypassed by pre–filtering the data with e.g. GARCH models. We later illustrate how ($A1^S$) may be relaxed to stationarity with a specific mixing rate allowing for a direct application of our test to serially dependent data, see Section 3.3.2. Assumption ($A2^S$) requires sample tails can be modeled by bivariate extreme value distributions and are asymptotically dependent, see Schmidt & Stadtmüller (2006) for details. Standard distributions with actual tail dependence, such as the bivariate t–distribution with dispersion parameter $\rho \neq 0$, meet this assumption. Notably, due to its tail independence ($\Lambda = 0$ for $|\rho| < 1$), the Gaussian copula violates ($A2^S$). Assumption ($A3^S$) imposes that the effective sample size k increases more slowly than n for $n \to \infty$. The second–order condition ($A4^S$) (see Bücher & Dette (2013)) effectively requires that the lower part of the scaled copula can be approximated sufficiently well by the tail copula, i.e. with order A. This, in fact, is a regular variation restriction and in practice imposes a corresponding slightly tighter condition on the expanding rate of k. For example, if $A(t)$ is asymptotically of order $1/t^{\alpha}$ with $\alpha > 0$, then k should be at most of order $n^{\frac{2\alpha}{1+2\alpha}} < n$ in order to satisfy the conditions. For completeness, we state Assumption ($A5^S$). Nevertheless, this smoothness assumption may also be omitted. This results in a more complex limiting behavior of the empirical tail copula, which permits consistent estimation of tail copulas of factor models, see Bücher et al. (2014).

Under Assumptions ($A1^S$)–($A5^S$), the asymptotic distribution for the tail copula can be derived as follows

$$\sqrt{k_{\mathbb{X}}}(\widehat{\Lambda}_{\mathbb{X}}(x^{(1)},x^{(2)}) - \Lambda_{\mathbb{X}}(x^{(1)},x^{(2)})) \overset{w}{\to} \mathbb{G}_{\widehat{\Lambda},\mathbb{X}}(x^{(1)},x^{(2)}),(x^{(1)},x^{(2)}) \in \mathbb{R}^2_+; \tag{3.2}$$

where $\overset{w}{\to}$ denotes weak convergence, $\mathbb{G}_{\widehat{\Lambda},\mathbb{X}}$ is a bivariate Gaussian field of the form

$$\mathbb{G}_{\widehat{\Lambda},\mathbb{X}}(x^{(1)},x^{(2)}) = \mathbb{G}_{\tilde{\Lambda},\mathbb{X}}(x^{(1)},x^{(2)}) - \sum_{i=1}^{2} \Lambda_{\partial_i}(x^{(1)},x^{(2)})\mathbb{G}_{\tilde{\Lambda},\mathbb{X}}(x^{(i)},x_{-i}=\infty),$$

where $\Lambda_{\partial_i}(x^{(1)},x^{(2)}) := \frac{\partial\Lambda(x^{(1)},x^{(2)})}{\partial x^{(i)}}$ denote the partial derivatives of the tail copula, $\mathbb{G}_{\tilde{\Lambda},\mathbb{X}}(x^{(1)},x^{(2)})$ is a *centered* Gaussian field with covariance

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

These results were first established in Schmidt & Stadtmüller (2006); Bücher & Dette (2013) and Bücher et al. (2014) provide related results while also relaxing ($A5^S$), i.e. existence of partial derivatives of the tail copula is generally not needed. This is important in practice, as it covers not only smooth standard models for tail models, but also practically relevant tail dependence model that may arise from (tail) factor models.

3.3 A new testing methodology against tail asymmetry and inequality

3.3.1 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 many disjoint subsets of the relevant support ($\mathcal{S}$) for all subset permutations. 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 as only then null violations in all directions can be found ($\Lambda_{\mathbb{Z}}(x^{(1)}, x^{(2)})$, and also $\Lambda_{\mathbb{Z}}(x^{(2)}, x^{(1)}), \mathbb{Z} = \mathbb{X}, \mathbb{Y}$). 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 their 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 & 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, and thereby we search for differences between tail copulas over distinct, pre–determined subintervals of the unit simplex. Testing for tail equality over many different subintervals amounts to an entire collection of individual tests. If the null of tail dependence equality is rejected within a specific subset, this approach locates those sample regions that cause tail dependence differences. Test power strongly benefits from intra–tail asymmetric tail copulas. Further, in standard cases, i.e. intra–tail symmetric cases, it features similar, yet slightly better test properties as competing tests.

Note, the notation corresponds to the test against tail inequality. However, the test also applies for tail asymmetry by exchanging $\Lambda_{\mathbb{X}}$ by $\Lambda_{\mathbb{X}}^U$ and $\Lambda_{\mathbb{Y}}$ by $\Lambda_{\mathbb{X}}^L$. Due to the homogeneity property of the tail copula, it is sufficient to compare tail copulas only over the unit simplex hull instead of $\mathbb{R}_+^2$. We denote the unit simplex hull by $\mathcal{S} := \{(x^{(1)}, x^{(2)}) : x^{(1)} + x^{(2)} = 1, x^{(i)} \geq 0, i = 1, 2\}$. We apply M Cramér–von Mises tests on $M/2$ disjoint subinterval of $\mathcal{S}$. 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

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

for even numbered M, and with disjoint subintervals $\mathcal{I}_1, ..., \mathcal{I}_{M/2}$ of equal length that fully compose $[0, 1]$. Note that $H_0 : \bigcap_{m=1}^{M} H_{0,m}$, i.e. global tail equality naturally implies tail equality over each subset. Marginal test statistics are given by

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

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 $j \geq (M/2)+1$ guarantees that tail copulas are compared over the entire unit simplex, e.g. in *both directions*. If $H_{0,m}$ is true, $S^m = 0$, while $S^m > 0$ otherwise. Test statistics are estimated by replacing Λ by $\widehat{\Lambda}$. Empirical test statistics will be denoted by $\widehat{S}^m$.

The following proposition provides the marginal test distributions in the i.i.d. case. Section 3.3.2 discusses extensions for time series data.

Proposition 3.2. *Assume that Assumptions ($A1^S$)–($A4^S$) hold for $\mathbb{X},\mathbb{Y}$. Then,*

(a)

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

with

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

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

(b) Under H_0,

$$S^m \xrightarrow{w} 0, m = 1,...,M.$$

(c) Under H_1,

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

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

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 (3.2). Due to the complexity of the limiting stochastic processes, closed forms of the asymptotic distributions do not exist and have to be simulated. We follow Bücher & Dette (2013) and approximate the distribution of $(S^1,...,S^M)$ by a multiplier bootstrap. Further notation is required to construct the bootstrap distribution. The bth bootstrap estimate of $\widehat{S}^m$ is

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

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

$$\begin{aligned}
\widehat{\Lambda}^{(b)}_{\mathbb{Z}}(x^{(1)},x^{(2)}) &= \frac{1}{k_{\mathbb{Z}}}\sum_{i=1}^{n}\tilde{\xi}^{\mathbb{Z}}_i 1\left\{Z^{(1)}_i \geq \tilde{F}^{-1}_{1,\mathbb{Z}}(1-(k_{\mathbb{Z}}/n_{\mathbb{Z}})x^{(1)}), Z^{(2)}_i \geq \tilde{F}^{-1}_{2,\mathbb{Z}}(1-(k_{\mathbb{Z}}/n_{\mathbb{Z}})x^{(2)})\right\},\\
\tilde{\xi}^{\mathbb{Z}}_i &= \xi^{\mathbb{Z}}_i/\overline{\xi^{\mathbb{Z}}}, i=1,...,n_{\mathbb{Z}},\\
\tilde{F}_{j,\mathbb{Z}}(x) &= \frac{1}{n_{\mathbb{Z}}}\sum_{i=1}^{n_{\mathbb{Z}}}\tilde{\xi}^{\mathbb{Z}}_i 1\left\{Z^{(j)}_i \leq x\right\}, j=1,2,
\end{aligned}$$

and $\xi_i, i=1,...,n_{\mathbb{Z}}$, are i.i.d. random variables, called multipliers, with $\mathbb{E}(\xi_i)=\mathbb{V}(\xi_i)=1$. This bootstrap technique guarantees weak convergence of $(\widehat{S}^{m,(b)}-\widehat{S}^m)$ to $(\widehat{S}^m - S^m)$, conditional on the bootstrap samples and conditional on the observed samples of $\mathbb{X}$ and $\mathbb{Y}$. This means the asymptotic distributions of the bootstrap statistics converge to the asymptotic distributions of the empirical test statistics, and can be used to mimic the marginal null distributions in Proposition (3.2). We extend the test assumptions as follows.

Assumptions 3.2 (cont.).

($A6^S$) Multiplier variables $\xi_i, i\in\mathbb{Z}_+$, are i.i.d. random variables; ξ_i are independent of $\mathbb{X},\mathbb{Y}$, and $\mathbb{E}(\xi_i)=\mathbb{V}(\xi_i)=1$.

The following asymptotic result of the bootstrap version of the test ensures test consistency in the i.i.d. case.

Proposition 3.3. *Let ($A1^S$)–($A6^S$) hold. Then*

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

This result provides a feasible bootstrap approximation of the test distribution. For the i.i.d. case, we set $\xi_i \sim Exp(1)$.[2] Finally, a consistent Monte Carlo p–value for

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

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 & 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 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 & 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* $\widehat{S}^{m,(b)}, m = 1, ..., M,$ *for* $b = 1, ..., B$.
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, $\lim_{n,B\to\infty} \mathbb{P}\big(\cup_{m=1}^{M} \{p^m \leq \alpha^m | H_0\}\big) = f \leq \alpha$. We use $B = 1499$ bootstrap repetitions; note the necessary correction of B (1499 instead of 1500) which ensures consistency of the p–value. Unless otherwise stated, we discretize $[0,1]$ as $\mathcal{I}_n = \{0.01j\}_{j=1}^{99}$. We typically apply test Test Algorithm (1) with at most $M = 26$ marginal hypotheses, which discretizes $[0,1]$ into 13 equally sized subintervals.[3] 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.

[3] For $M = 26$, $\mathcal{I}_1 = \{0.01, 0.02, ..., 0.09\}, \mathcal{I}_2 = \{0.10, 0.11, ..., 0.17\}, ..., \mathcal{I}_{13} = \{0.93, ..., 0.99\}$. Note, subsets may not exactly be of equal length.

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 (3.3) 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 3.4. *For Test* (2)*, if*

$$\mathbb{P}(\cup_{j=1}^J \tilde{p}_j^* \leq \alpha | H_0) \uparrow \max(\alpha_1, ..., \alpha_J), \text{ as } J \to \infty, \tag{3.3}$$

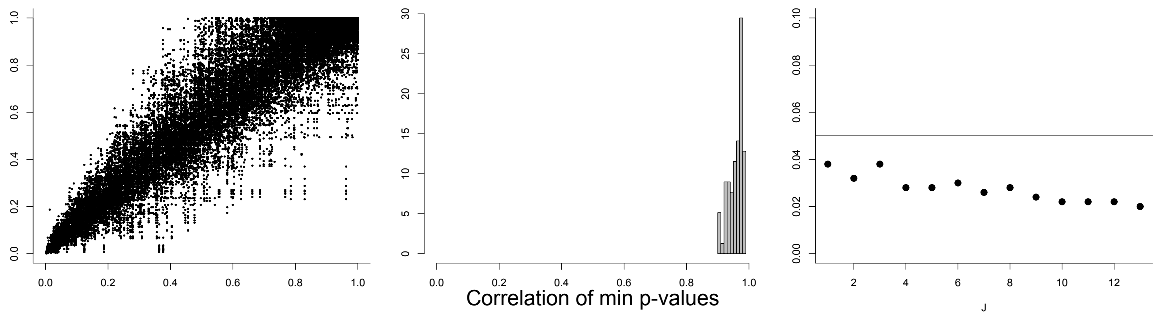
it holds that

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

The formal proof of Proposition (3.4) is contained in the Appendix. For the result, Condition (3.3) is key, and also that (realized) test sizes of Test (1) converge to zero as $M \to \infty$.

Simulation results from Section 3.4 confirm that Condition (3.3) appears to be satisfied in standard settings. 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. Figure (3.1), which shows p–values of one specific setting, illustrates that both these points hold; results of other settings are in line, but not reported. We see that individual test sizes are consistently below α, and decrease in the number of marginal hypotheses. Furthermore, correlation between minimal p–values of different grids is close to one, indicating nearly perfect (linear) dependence. Hence, we find Test (2) is appropriate.

Figure 3.1: Exemplary p–values from the simulation study for Test (1) with $j = 1, ..., 13$ (GARCH marginals equipped with a Factor model, $k = 0.1n, n = 1500$, tapered bootstrap). In this case, test size is estimated with 500 repetitions. Left: Scatterplots of p–values for all grid pairs. Middle: Histogram of estimated correlations between all pairs of grid–minimal p–values. Right: J, the fineness of the grids, is plotted against estimated test sizes according to Test (1).



Generally, it would be desirable to provide a lower bound of the strength of dependence between the p–values, i.e. a sufficient convergence rate in Condition (3.3). Convergence rates of individual test sizes and the unknown p–value dependence structure determine this lower bound. Unfortunately, to explicitly state this bound in our setting, we would have to assume specific closed–form distributions for the test statistics (Proschan & Shaw (2011)), or specific parametric dependence model for the p–values, see Stange et al. (2015) and Bodnar & Dickhaus (2014). Yet, the precise dependence structure between the p–values is unknown, whereas tails of the test distributions may be approximated by χ^2 distributions, see Beran (1975).

3.3.2 Inference for serially dependent data

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

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

For empirical copulas of dependent data, another remedy is to assume stationarity coupled with some mixing conditions, which consequently allows to use unfiltered returns for estimation. Valid statistical inference is ensured by adjusting the bootstrap 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 & Ruppert (2013). Necessary assumptions are met for a wide class of time series models, such as ARMA and GARCH models. We suggest to use the dependent data bootstrap methodology also for empirical tail copulas. Yet, we do not prove the validity of this approach as this difficult task is beyond the scope of this chapter. However, Assumption ($A4^S$) puts the tail copula process close to the scaled copula process in the respective tail — in finite samples, where the running variable t has to be replaced by k/n, and both $k, n > 0$ are fixed, differences between $\Lambda(x^{(1)}, x^{(2)})$ and $tC(x^{(1)}/t, x^{(2)}/t)$ might be neglectable. This suggests that results of the empirical copula process $(\sqrt{n_\mathbb{Z}}(\widehat{C}_\mathbb{Z}(x^{(1)}, x^{(2)}) - C_\mathbb{Z}(x^{(1)}, x^{(2)})))$ carry over to the empirical tail copula process $(\sqrt{k}_\mathbb{Z}(\widehat{\Lambda}_\mathbb{Z}(x^{(1)}, x^{(2)}) - \Lambda_\mathbb{Z}(x^{(1)}, x^{(2)}))$. We employ the tapered block multiplier and the block bootstrap for tail copula estimation. For completeness, results of the previous section are adopted for the tapered block multiplier bootstrap. The i.i.d. Assumption ($A1^S$) is replaced by the following assumption, see Bücher & Ruppert (2013).

Assumptions 3.2 (cont.).

($A1^{S*}$) $\mathbb{X}, \mathbb{Y}$ *are realizations of a strictly stationary process that is strongly mixing with rate* $\alpha_\mathbb{Z} = \mathcal{O}(r^{-a_\mathbb{Z}}), r > 0, a_\mathbb{Z} > 1, \mathbb{Z} = \mathbb{X}, \mathbb{Y}$.

The mixing coefficient is defined as $\alpha_\mathbb{Z}(r) = \alpha_\mathbb{Z}(\mathcal{F}_s, \mathcal{F}_{s+r}) = \sup_{A\in\mathcal{F}_s, B\in\mathcal{F}_{s+r}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|$, where $\mathcal{F}_t$ denotes the filtration of the underlying stochastic process up to time point t, and $\mathbb{Z}$ is strongly mixing if $\alpha_\mathbb{Z}(r) \to 0$ for $r \to \infty$, i.e. serial dependence vanishes as the interval length between two events increases.

This assumption specifies the rate at which serial dependence has to vanish. Consequently, under ($A1^{S*}$) & ($A2^S$)–($A4^S$), the empirical tail copula should converge to some centered Gaussian process $G_\alpha(x^{(1)}, x^{(2)})$ that is governed by the mixing rate $\alpha_\mathbb{Z}$, i.e.

$$\sqrt{k_\mathbb{Z}}(\widehat{\Lambda}_\mathbb{Z}(x^{(1)}, x^{(2)}) - \Lambda_\mathbb{Z}(x^{(1)}, x^{(2)})) \xrightarrow{w} \mathbb{G}_{\alpha,\mathbb{Z}}(x^{(1)}, x^{(2)}), (x^{(1)}, x^{(2)}) \in \mathbb{R}^2_+.$$

The functional delta theorem ensures convergence of $\widehat{\Lambda}_{\mathbb{Z}}, \mathbb{Z} = \mathbb{X}, \mathbb{Y}$, carry over to the test statistics $(\widehat{S}^1, ..., \widehat{S}^M)$. To approximate the limiting behavior of the test statistics, now the tapered block multiplier bootstrap has to be applied. The tapered block multiplier bootstrap generates series of block–dependent multipliers that replace the i.i.d. multipliers. The following conditions have to be met for the consistency of the tapered block multiplier bootstrap in case of the empirical copula process, see Theorem 3 in Bücher & Ruppert (2013).

Assumptions 3.2 (cont.).

($A7^{S*}$) *The underlying stochastic process of* $\mathbb{Z}$ *is strongly mixing with* $\sum_{r=1}^{\infty}(r+1)^c\sqrt{\alpha_{\mathbb{Z}}(r)} < \infty, c = \max(28, \lfloor 2/\epsilon \rfloor + 1)$.

($A8^{S*}$) *The tapered block multiplier process* $(\xi_{j,n})_{j=1,...,n}$ *is strictly stationary, has bounded moments, is independent of* $\mathbb{Z}$*, and positively* $cl(n)$*–near epoch dependent,*[4] *where* c *is some constant and* $l(n) \to_{n\to\infty} \infty, l(n) = o(n)$*, and for all positive valued integers* j, h *assume* $\mathbb{E}(\xi_{j,n}) = \mu > 0, \mathbb{V}(\xi_{j,n}, \xi_{j+h,n}) = \mu^2 v(h/l(n))$ *and* v *is a bounded function symmetric around zero, and w.l.o.g.* $\mu = 1, v(0) = 1$.

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

($A7^{S*}$) demands the serial dependence in $\mathbb{Z}$ must vanish sufficiently fast. For example, AR and GARCH processes fulfill this condition. ($A8^{S*}$) and ($A9^{S*}$) give conditions on the (dependent) multiplier process and the multiplier block length (l) under which the generated multiplier series consistently mimics the resulting dependence structure of $\mathbb{Z}$. Bücher & Ruppert (2013) provide detailed advice on implementation strategies. The authors suggest to fix a block length of $l(n) = 1.25n^{1/3}$ for the block bootstrap. Moreover, for the tapered block multiplier bootstrap, we employ 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 the R–package `npcp`, see Kojadinovic (2015).

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

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

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

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

3.3.3 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.

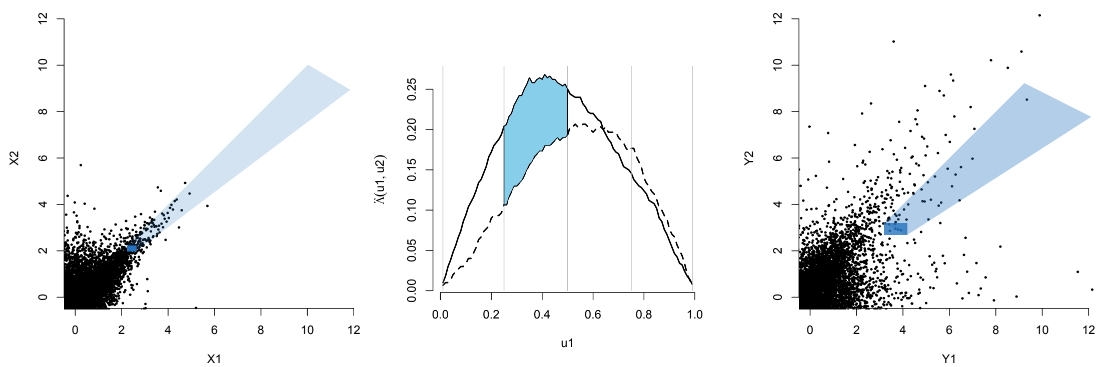
$$Q_{\mathbb{X}} = \left(F^{-1}_{1,\mathbb{X}}(1-k/nx^{(1)}), F^{-1}_{1,\mathbb{X}}(1-k/nx^{(2)})\right) \times \left(F^{-1}_{2,\mathbb{X}}(1-k/nx^{(1)})), F^{-1}_{2,\mathbb{X}}(1-k/nx^{(2)})\right),$$
$$Q_{\mathbb{Y}} = \left(F^{-1}_{1,\mathbb{Y}}(1-k/nx^{(1)}), F^{-1}_{1,\mathbb{Y}}(1-k/nx^{(2)})\right) \times \left(F^{-1}_{2,\mathbb{Y}}(1-k/nx^{(1)})), F^{-1}_{2,\mathbb{Y}}(1-k/nx^{(2)})\right).$$

Due to the homogeneity of the tail copulas, these extreme sets can be extrapolated arbitrarily far into the tail, given the extreme value conditions hold. In particular, Figure (3.2) 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 3.2: 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 $\Lambda(x^{(1)}, x^{(2)}) = x^{(1)} + x^{(2)} - \big[(1 - \psi^{(1)})x^{(1)} + (1 - \psi^{(2)})x^{(2)} + ((\psi^{(1)}x^{(1)})^{-\theta} + (\psi^{(2)}x^{(2)})^{-\theta})^{\theta}\big]$ (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.



3.4 Simulation study

We now compare the finite sample performance of our test with the TDC test, and the BD13 test.[5] For this, we study two types of dependence models that are frequently used in finance. First, we employ the (implicit) factor model copula. See Fama & French (1992), Einmahl et al. (2012), and Oh & Patton (2015) for factor models in finance, tail dependence of factor models, and tail dependence of factor copulas in finance, respectively. Second, representing the broad class of Archimedean copulas, we employ the Clayton copula, which models solely lower tail dependence. Its lean parametric form makes the Clayton copula a popular building block for more complex copula models, such as mixtures of copulas, see Rodriguez (2007) and Patton (2006). For each copula, we impose one parametrization that fulfills the null, and one that violates the null, leaving us with four DGPs.

[5]We focus on non–parametric tests only as in practice parametric specifications may suffer from a model bias, especially if intra–tail asymmetry is not accounted for.

DGP1 and DGP2 are based on the tail factor model. Bivariate return vectors $\mathbb{Z} = (Z^{(1)}, Z^{(2)}), \mathbb{Z} = \mathbb{X}, \mathbb{Y}$, follow a bivariate factor model with r factors $V^{(j)}, j = 1, ..., r$, and loadings $a_{ij}, i = 1, 2, j = 1, ..., r$, when

$$Z^{(i)} = \sum_{j=1}^{r} a_{ij} V^{(j)} + \varepsilon^{(i)}, i = 1, 2, \tag{3.4}$$

where factors are i.i.d. Fréchet with $\nu = 1$, independent of the error term $\varepsilon^{(i)}$ which feature thinner tails than $V^{(j)}$; we set $\varepsilon^{(i)}$ as Fréchet with $\nu_\varepsilon = 2$. In this way, the matrix of factor loadings $A = (a_{ij})$ directly determines the tail copula of $\mathbb{Z}$. In particular, the (upper) tail copula of $\mathbb{Z}$ is equivalent to the tail copula of the max factor model $\bar{Z}^{(i)} = \max_{j=1,...,r}(a_{ij} V^{(j)})$, which is

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

see Einmahl et al. (2012) for further details. DGP1 consists of $\mathbb{X}, \mathbb{Y}$ both resulting from a factor model as in Equation (3.4), but with loading matrix

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

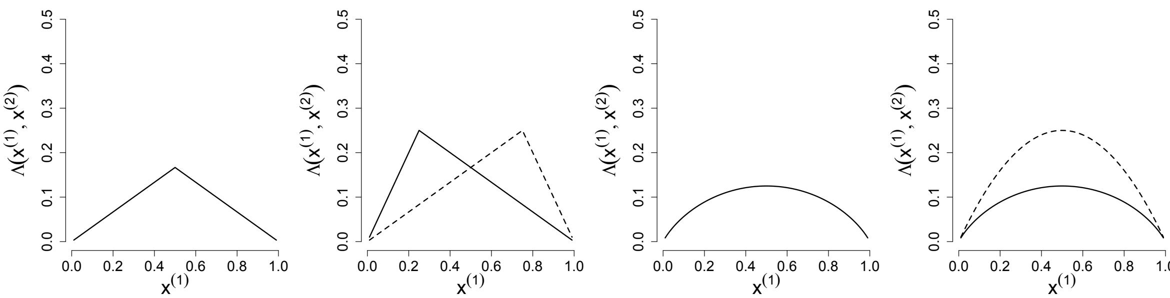
Here, the first factor only influences $X^{(1)}$ $(Y^{(1)})$, the second factor influences both $X^{(1)}$ $(Y^{(1)})$ and $X^{(2)}$ $(Y^{(2)})$, and the third factor only influences $X^{(2)}$ $(Y^{(2)})$. That is, A_1 amounts to intra–tail symmetry and to tail equality between $\mathbb{X}$ and $\mathbb{Y}$, and thus the null is true. See Figure (3.3), first from the left, for $\Lambda(x^{(1)}, 1 - x^{(1)}), x^{(1)} \in [0, 1]$. For DGP2, both $\mathbb{X}$ and $\mathbb{Y}$ stem from a factor model as in Equation (3.4) with

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

where the second factor only influences $X^{(2)}$ $(Y^{(2)})$, causing the tail copula to become intra–tail asymmetric, $\Lambda(x^{(1)}, x^{(2)}) \neq \Lambda(x^{(2)}, x^{(1)})$, and consequently tail copulas of $\mathbb{X}$ and $\mathbb{Y}$ coincide only when $x^{(1)} = x^{(2)}$, see Figure (3.3), second from the left. DGP2 thus represents the class of intra–tail asymmetric copulas which violate the null according to Proposition (3.1).

For the Clayton copula, only the lower left part of the distribution features tail de-

Figure 3.3: 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 $\mathbb{X}$ and $\mathbb{Y}$.



pendence,

$$\Lambda^L(x^{(1)}, x^{(2)}; \theta) = ({x^{(1)}}^{-\theta} + {x^{(2)}}^{-\theta})^{-1/\theta},$$
$$\Lambda^U(x^{(1)}, x^{(2)}; \theta) = 0,$$

where (lower) tail dependence increases in the parameter $\theta \in [0, \infty)$. DGP3 is given by $\mathbb{X}, \mathbb{Y} \sim$ Clayton($\theta = 0.5$); this specific choice of θ 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 DGP3, the null is true. See Figure (3.3), second from the right. 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 (3.3), first from the right.

To check whether the test also works for financial time series data, we combine all DGPs with i.i.d. as well as GARCH marginals. We apply the test to *raw* GARCH returns, and to standardized GARCH residuals as it is important to analyze whether using *estimated* residuals affects test performance. Moreover, we study the test performance for unfiltered returns using the block bootstrap and the tapered block multiplier bootstrap. In particular, we employ GARCH(1,1) dynamics for any marginal return process. We follow Oh & Patton (2013) and employ bivariate AR–GARCH models. We can link serially dependent marginals by the (implicit) copulas of DGPs 1 to 4, allowing us to study the effect of conditional heteroscedasticity on test performance.

For both bivariate return series $\mathbb{Z} = (Z^{(1)}, Z^{(2)}), \mathbb{Z} = \mathbb{X}, \mathbb{Y}$, it holds

$$\begin{aligned} Z_t^{(i)} &= \sigma_t^{(i)} \eta_{t,\mathbb{Z}}^{(i)}, \\ \sigma_{t,\mathbb{Z}}^{2,(i)} &= \omega + \alpha^{(i)} Z_{t-1}^{2,(i)} + \beta^{(i)} \sigma_{t-1,\mathbb{Z}}^{2,(i)}, \\ \eta_{\mathbb{Z}} &:= (\eta_{\mathbb{Z}}^{(1)}, \eta_{\mathbb{Z}}^{(2)}) \sim iid\ F_{\eta,\mathbb{Z}}(x^{(1)}, x^{(2)}) = C_{\eta,\mathbb{Z}}(F_{\eta,\mathbb{Z},1}(\eta_{\mathbb{Z}}^{(1)}), F_{\eta,\mathbb{Z},2}(\eta_{\mathbb{Z}}^{(2)})), t = 1, ..., n_{\mathbb{Z}}, \end{aligned}$$

where we set $\omega = 0.01, \alpha = 0.15$ and $\beta = 0.8$ such that $\alpha + \beta$ is close to one. This mimics parameter values often found in financial returns, see for example Engle & Sheppard (2001). To impose the tail structures of DGPs 1 to 4 on the time series, we use DGPs 1 to 4 to model the error copula $C_{\eta,\mathbb{Z}}(F_{\eta,\mathbb{Z},1}(\eta^{(1)}), F_{\eta,\mathbb{Z},2}(\eta^{(2)}))$ and to generate $\eta_{t,\mathbb{Z}} = (\eta_{t,\mathbb{Z}}^{(1)}, \eta_{t,\mathbb{Z}}^{(2)})$: In a first step, we simulate observations $\eta_{t,\mathbb{Z}}$ according to DGPs 1 to 4. Consequently, we transform simulated errors to pseudo–observations by means of the marginal empirical cumulative distribution, $\widehat{F}_{\eta,\mathbb{Z},i}(\eta_{t,\mathbb{Z}}^{(i)}), i = 1, 2$. Finally, we apply the quantile function of the t–distribution function with 10 degrees of freedom to the pseudo–observations. Thus, the final errors are linked by the copulas of DGPs 1 to 4 with fat–tailed t–marginals.[6] Those are used to generate the GARCH series for $\mathbb{X}$ and $\mathbb{Y}$. We obtain standardized residuals from estimation by quasi maximum likelihood.

For sample sizes $n = 750, 1500$, varying values of the effective sample size k, and a nominal test level 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 $M_j = 2j, j = 1, ..., 15,$.[7] The TDC test is carried out using the multiplier bootstrap at points $x^{(1)} = x^{(2)} = 0.5$. The number of simulations is $S = 500$ for each setting.

Table (3.1) reports empirical rejection frequencies for i.i.d. marginals, filtered GARCH marginals, unfiltered GARCH marginals, GARCH marginals with the block and tapered bootstrap, and sample size $n = 1500$. Also, we study the effect of varying the effective sample size, $k \in \{\lfloor 0.1n \rfloor, \lfloor 0.2n \rfloor, \lfloor 0.3n \rfloor\}$. 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 $x^{(1)} = x^{(2)} = 1$. Table (3.2) contains

[6] Monotone transformations, such as the quantile transformation, do not alter the tail dependence structure. However, t–transformed error distributions are a more realistic approximation of asset returns.

[7] Note, for some grids, this implies subintervals are only roughly of equal length.

empirical rejection frequencies for $n = 750$. As non–parametric methods for tail dependence are often criticized for unsatisfactory small sample performance, it is worth studying test behavior for small and moderate sample sizes.

In general, 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 & 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 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 (3.2) 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 (Test (2)) is throughout most powerful, while the test with fixed grids (Test (1)) is consistently more powerful than the TDC test, slightly less powerful than the BD13 test, and more powerful than the latter in case of intra–tail asymmetry.

3.5 Empirical application

3.5.1 Tail asymmetries within S&P500 industry portfolios

Related studies, e.g. Ang & 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 S&P500 industry portfolios. The data set[8] contains nearly 90 years of weighted returns of CRSP SIC codes–based industries.[9]

We proceed as follows. We aim to detect tail asymmetry dynamics within the complete S&P500 universe. Applying a rolling window analysis with window length of $n = 1500$, i.e. nearly six years, and a step size of 300 trading days, i.e. roughly 14 months, we arrive at 74 (overlapping) time periods. In each period, we build all possible bivariate industry combinations , $\mathbb{X} = (X^{(i)}, X^{(j)})$, and test the nulls

$$H_0 : \Lambda^U_{\mathbb{X}} = \Lambda^L_{\mathbb{X}}.$$

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 85000 times. To avoid possible model risk by pre–filtering the returns, we throughout analyze raw returns using the tapered block multiplier bootstrap; Section 3.3.2 and the results

[8]Available at `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`, accessed on 03/01/2016.

[9]For detailed information on industry composition we refer to the website just mentioned.

Table 3.1: 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	TA2			18	6	TA2			18	6	TA2			18	6	TA2
iid																				
5%	4.0	3.2	3.2	2.4	4.8	3.2	4.2	100	100	100	5.0	4.8	3.2	4.2	6.8	73.8	86.2	78.2	82.2	88.2
10%	2.0	3.8	2.0	2.4	5.4	4.0	4.4	100	100	100	2.2	3.6	3.4	2.8	4.8	91.8	97.6	94.8	95.8	98.2
15%	4.4	3.2	2.8	2.6	6.0	5.2	5.8	100	100	100	3.0	3.0	3.0	2.4	7.0	96.6	99.8	98.4	98.6	100
fil.																				
5%	3.4	4.4	2.8	3.4	5.8	5.4	7.6	100	100	100	3.6	4.0	3.0	2.8	5.8	73.8	86.2	78.2	82.2	87.2
10%	4.0	4.4	2.4	3.8	5.8	5.0	7.6	100	100	100	4.4	3.8	3.4	2.6	6.6	92.6	97.4	95.6	96.0	97.8
15%	5.2	4.0	3.0	3.0	5.4	9.2	8.8	100	100	100	3.0	3.0	3.0	2.4	7.0	97.2	98.8	97.6	98.4	98.8
unfil.																				
5%	6.0	6.6	4.2	4.6	8.0	8.6	12.6	83.2	52.0	86.8	9.6	12.4	8.6	9.8	14.2	17.8	21.2	18.0	19.2	24.6
10%	4.6	5.8	4.0	4.6	7.4	6.6	8.8	100	100	100	7.0	11.4	9.0	9.8	14.8	22.0	31.0	25.4	26.2	34.6
15%	4.8	4.2	3.0	4.2	6.4	5.6	7.8	100	100	100	6.8	7.4	6.4	6.0	10.2	33.2	44.2	35.8	39.8	48.0
blo.																				
5%	6.6	5.0	3.6	3.4	6.6	8.0	8.0	73.4	40.2	81.4	7.4	11.0	8.8	9.4	14.6	37.0	44.6	39.0	42.6	49.0
10%	6.0	4.8	3.4	4.0	5.4	6.6	6.6	100	99.8	100	5.6	8.0	8.4	7.8	13.0	70.4	80.0	70.2	76.0	82.8
15%	6.0	5.0	3.2	3.4	5.8	5.6	6.4	100	100	100	4.0	7.2	8.0	6.6	12.8	88.0	94.0	90.8	92.2	95.0
tap.																				
5%	3.8	4.8	2.6	3.8	6.0	5.4	6.8	100	100	100	3.8	4.2	2.6	2.2	6.0	75.8	85.6	77.8	82.6	87.8
10%	3.8	4.6	2.6	3.2	5.6	5.4	7.4	100	100	100	4.2	3.4	4.0	2.6	6.4	92.8	97.6	95.2	96.8	97.8
15%	5.2	4.0	2.8	3.0	4.8	4.4	4.6	100	100	100	4.4	5.0	3.8	3.4	6.8	97.0	99.0	97.4	98.4	99.0

Table 3.2: Empirical rejection probabilities as in Table (3.1), but with a sample size of $n = 750$.

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

of the simulation study justify this approach.[10] Also, we fix the effective sample size to $k = 0.2n$,[11] which, 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.

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 (3.6) shows the trajectory of the mean and q–quantiles, $q \in \{0.01, ..., 0.99\}$, for both upper and lower tails covering 1931–2015.

The null hypothesis of tail equality is tested by the TDC test, the BD13 test and

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

[11] This corresponds to $k = 0.1n$ in TDC studies which evaluate the tail copula at $(x^{(1)}, x^{(2)}) = (1, 1)$.

Test (2), which aggregates over 15 grids in the spirit of the simulation study. Figure (3.4) displays trajectories of the share of rejections for each test, i.e. the share of tail asymmetric pairs according to each test. Figure (3.5) 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 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 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 coincide strikingly with during times 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–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. 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 (3.6)) 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

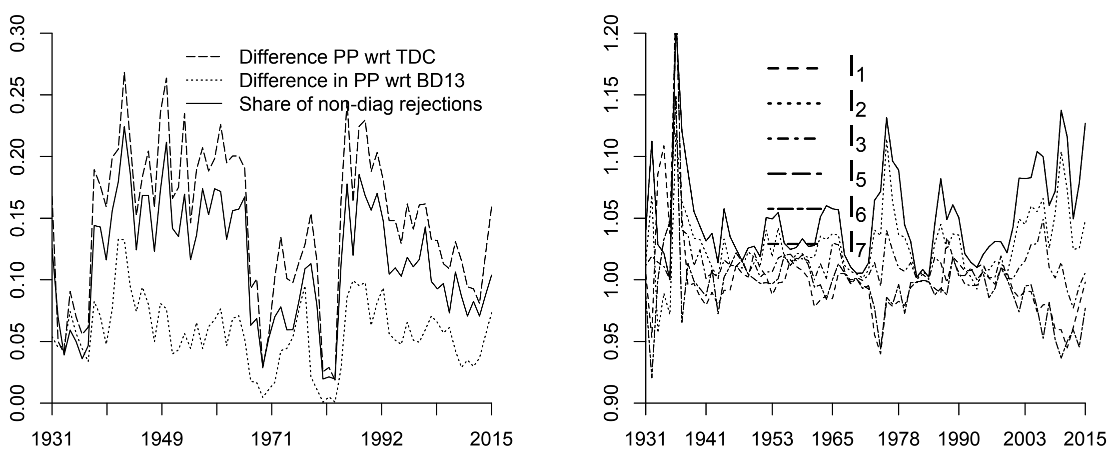
Figure 3.4: Dynamics of the percentage of detected tail asymmetries among all pairs using a rolling window of size $n = 1500$, and a step size of 300 trading days for the TDC test (dashed), BD13 test (dotted) and our test (solid), respectively.

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 (3.5) (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 (3.5)). This line runs nearly parallel to the graph of the differential in found tail asymmetries between the TDC test and our test.

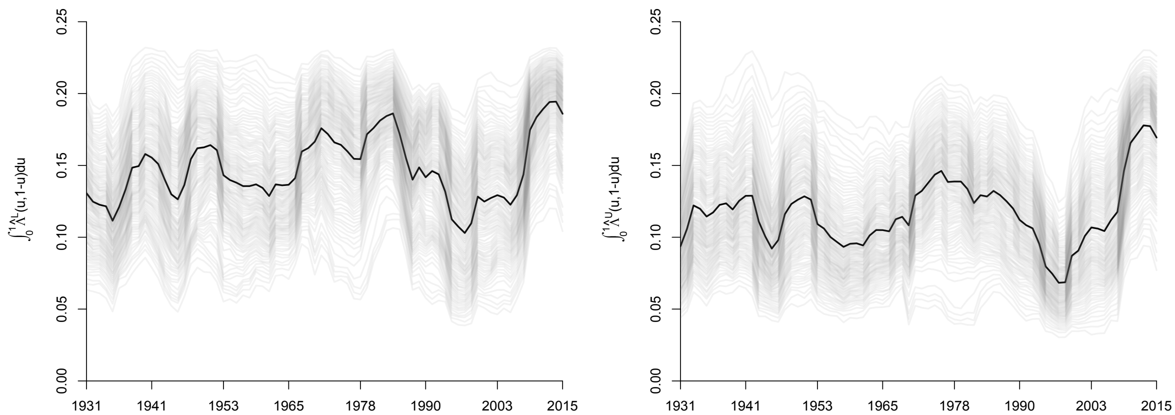
To further underline the importance of non–standard tail dependence structures, we quantify the number of tail asymmetric pairs that scalar approaches would miss due to *off–diagonal* tail asymmetries. In Figure (3.5) (right), for each period, we compare the number of rejections of *non–central* subintervals 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.

Figure 3.5: (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 $M = 14$.



Furthermore, the difference in found asymmetries between our test and BD13 suggests some degree of intra–tail asymmetry among all pairs. The simulation study demonstrated both tests' power differs mainly in intra–tail asymmetric cases. However, quantifying the effect of intra–tail asymmetries on test rejection rates is beyond the scope of this chapter as independent tests against intra–tail asymmetries have been developed, see Kojadinovic & Yan (2012), Bormann (2016).

Figure 3.6: $\int_0^1 \widehat{\Lambda}(x, 1-x)du$ 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.



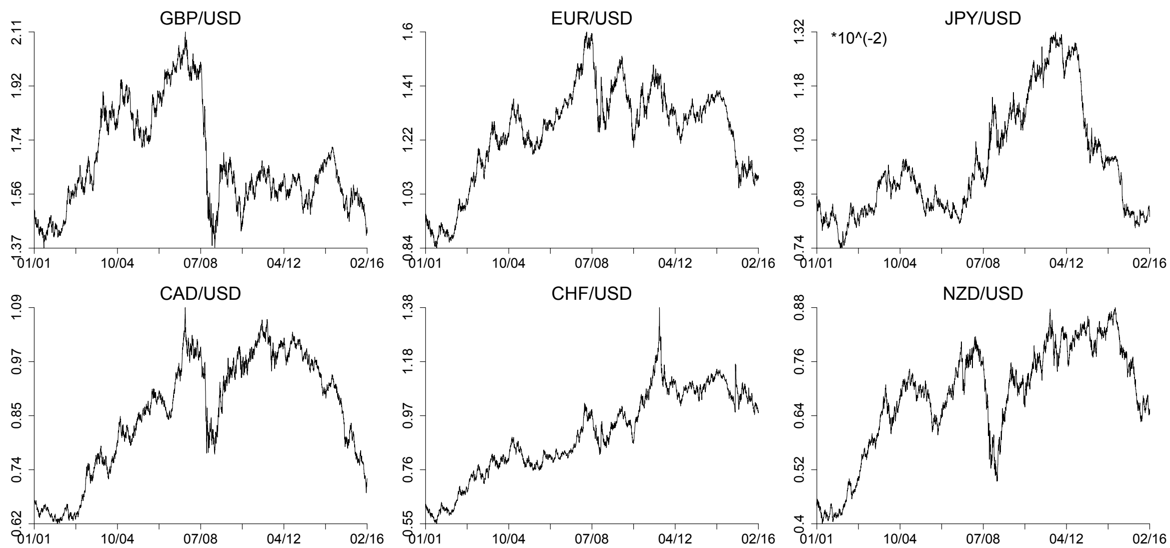
3.5.2 Tail inequalities of foreign exchange rates

We now analyze tail equality in pairs of six main foreign exchange rates, namely Euro (EUR), British Pound (GBP), Canadian Dollar (CAD), Japanese Yen (JPY), New Zealand Dollar (NZD) and Swiss Franc (CHF), all nominated in USD.[12] The sample consists of returns of daily closing prices covering the period 01/05/2001 to 02/01/2016.

As foreign exchange rates are the most frequently traded financial security with an average daily trading volume of more than five trillion in April 2013,[13] investors and regulators have a natural interest in a comparison of extreme co–movements of foreign exchange rates. We again apply a rolling window analysis, now with a window size of $n = 1000$ and step size of 50 days to draw a finer picture of the tail (in)equality dynamics. For any pair comparison trading days with missing data or zero returns

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

Figure 3.7: Foreign exchange rates nominated in US Dollars during 01/2001 – 02/2016.



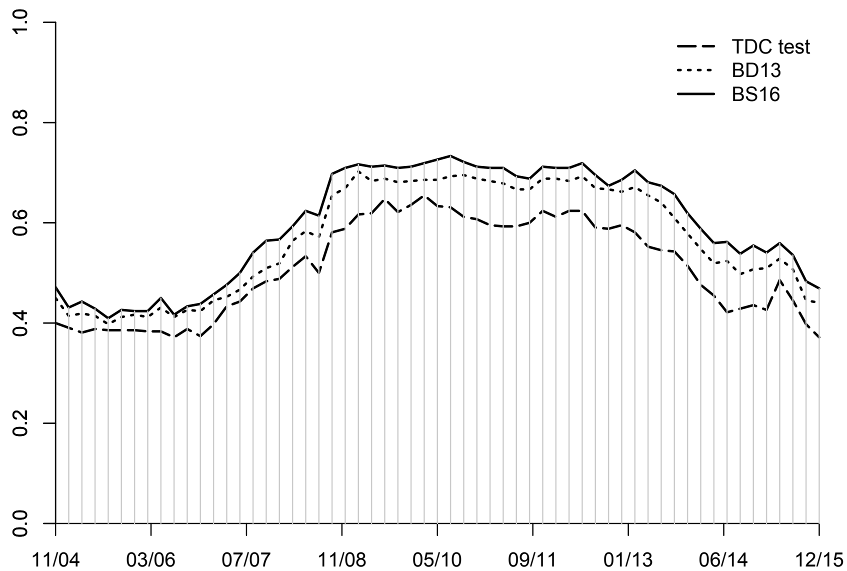
are discarded. The effective sample size is fixed $k = 0.2n$ which is backed by the results of the simulation study. We conduct the following tail pair comparisons

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

for all 15 bivariate pairs, amounting to $4 \cdot \binom{6}{2} = 420$ tests in each period. Figure (3.8) shows the share of tail inequalities among all possible comparisons. The fraction of rejected tail equalities, ranging from 45% to 75%, suggests bivariate tail dependence of foreign exchange rates systematically differ. We observe a steady increase of tail inequalities from 2006 to 2008 which coincides with a major depreciation of the USD with respect to the EUR. This evolution is reversed when the USD appreciates during the European Sovereign Crisis (2013 onwards). Thus, in the last decade, a strong (weak) USD (EUR) came along with more (less) tail equality within the foreign exchange rates market.

Figure (3.9) displays a dynamic ranking for all 15 pairs based on the TDC and the summary statistic $\int_0^1 \Lambda(\phi, 1-\phi)\mathrm{d}\phi$ which was introduced in the last subsection. A careful inspection of all four plots shows there is only little difference between the TDC–based and the tail copula–based ranking. Tail dependence of appreciations and

Figure 3.8: Dynamics of the percentage of detected tail inequalities among all pairs, comparing the following tails: Upper–upper, upper–lower, lower–lower. The window size is $n = 1000$ with a step size of 50 trading days, and rejections based on the TDC test (BD13 test, our test) correspond to the dashed (dotted, solid) line.



depreciations of EUR and CHF with respect to the USD tends to be the strongest throughout the sample. While the pair GBP–EUR exhibits strong tail dependence for joint upper tails (depreciations), the lower tails show a strong tail link only in the last five years (as well as until 2007). Also, JPY–CAD (upper tail) and CAD–NZD (both tails) feature comparably strongly connected tails. The pairs JPY–NZD, JPY–CAD and GBP–JPY feature the weakest tail dependence in both tails.

The pair EUR–CHF dominates tail comparisons throughout, which is probably due to the fixed exchange rate regime until 01/2015 with a EUR:CHF minimum rate of $1{:}1.20$. Also, the tight economic linkage between both parties may attribute to the relatively strong tail dependence. On 01/15/2015, the Swiss Central Bank unpegged its currency from the Euro, intending to avoid a continued depreciation of the Swiss currency as the EUR had steadily devaluated since 2008/2009. This policy change caused the CHF to appreciate by 20% with regards to the EUR within a single day.

We now test whether the break of the CHF–EUR currency peg had a significant impact on the tail dependence between both currencies. This would be the case if the TC had changed after 15/01/2015. Unfortunately, the sample contains only 273 observations after the policy change and we thus compare TCs for overlapping time periods, that is 01/01/2006 – 14/01/2015 (Λ^{T1}) and 01/01/2006 – 16/01/2016 ($\Lambda^{T1,T2}$). The null is

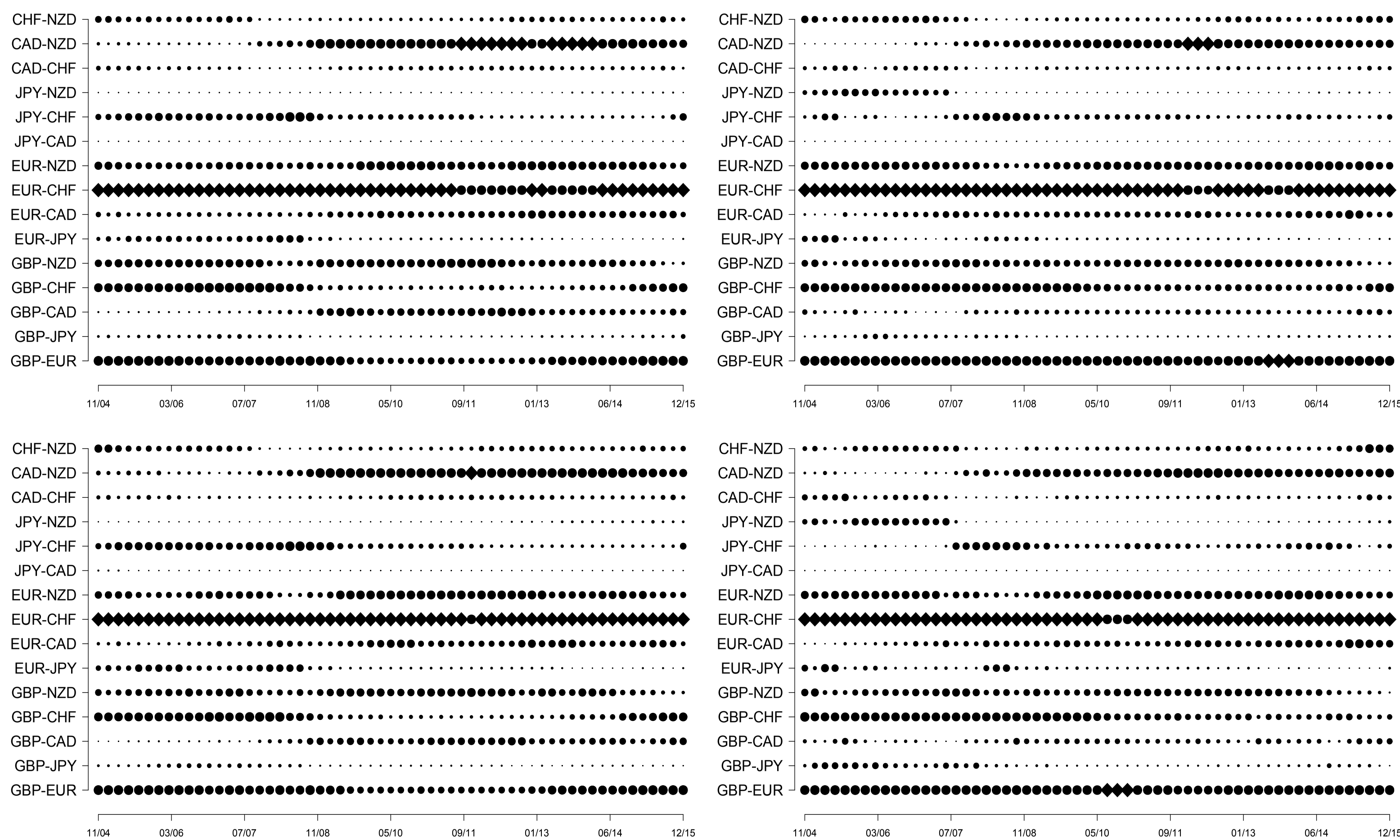
$$H_0 : \Lambda^{W,T1}_{CHF-EUR} = \Lambda^{W,T1+T2}_{CHF-EUR}, W = U, L \tag{3.5}$$

However, the tapered block multiplier bootstrap has to be adjusted to account for the dependence of both samples. For the tail copula of the entire period ($T_1 + T_2$), we use the multiplier vector $\xi^{T1+T2} = (\xi_1, ..., \xi_{T1}, \xi_{T1+1}, ..., \xi_{T1+T2})$; for the tail copula of the first subperiod, we only use the first T_1 entries of ξ^{T1+T2}. We execute the test for 15 different values of the effective sample size, namely $k_{Ti} = 0.02n_{Ti}, 0.04n_{Ti}, ..., 0.3n_{Ti}, i = 1, 2$, where n_{Ti} denotes the sample size of the first subperiod (T_1) and the entire period (T_2), respectively. Table (3.3) contains p–values of Test (2). To this date, there is no evidence for a structural change in neither the lower nor the right tail.

Table 3.3: p–values corresponding to the null hypothesis of constant tail dependence between EUR and CHF (see Equation (3.5)) for varying effective sample sizes.

tails	k/n														
	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.18	0.20	0.22	0.24	0.26	0.28	0.30
L-L	4.4	10.6	37.2	30.8	17.9	19.9	22.4	35.6	52.0	61.5	46.2	33.6	46.0	56.1	63.7
U-U	99.2	7.4	98.0	87.2	99.9	87.8	44.7	80.2	98.6	99.6	95.7	96.1	99.8	99.6	96.6

Figure 3.9: Tail dependence ranking of all 15 pairs for the lower (left column) and upper (right column) tails according to the summary statistic $\int_0^1 \Lambda(\phi, 1-\phi)\mathrm{d}\phi$ (top row) and the TDC (bottom row), respectively. The size of each dot reflects the rank. For each period, the diamond shaped dots mark the strongest pair.



3.6 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 S&P500 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.

3.7 Appendix

3.7.1 Proofs

Proof of Proposition (3.2).

Equation (3.2) guarantees convergence of the empirical tail copula $\sqrt{k_{\mathbb{Z}}}\widehat{\Lambda}_{\mathbb{Z}}(x^{(1)}, x^{(2)}), \mathbb{Z} = \mathbb{X}, \mathbb{Y}$, for $(x^{(1)}, x^{(2)}), (v^{(1)}, v^{(2)}) \in \mathbb{R}_+^2$. Define

$$\widehat{\Delta}(x^{(1)}, x^{(2)}, v^{(1)}, v^{(2)}) =: \sqrt{k_{\mathbb{Y}}/(k_{\mathbb{X}} + k_{\mathbb{Y}})}\widehat{\Lambda}_{\mathbb{X}}(x^{(1)}, x^{(2)}) - \sqrt{k_{\mathbb{X}}/(k_{\mathbb{X}} + k_{\mathbb{Y}})}\widehat{\Lambda}_{\mathbb{Y}}(v^{(1)}, v^{(2)}),$$

which is a sum of rescaled tail copula processes with $\mathbb{G}_{\widehat{\Lambda},\mathbb{Z}}, \mathbb{Z} = \mathbb{X}, \mathbb{Y}$, is a bivariate Gaussian process. It directly follows from Equation (3.2) that

$$\begin{aligned}\widehat{\Delta}(x^{(1)}, x^{(2)}, v^{(1)}, v^{(2)}) &\overset{w}{\to} \Delta(x^{(1)}, x^{(2)}, v^{(1)}, v^{(2)}) \\ &:= \sqrt{k_{\mathbb{Y}}/(k_{\mathbb{X}} + k_{\mathbb{Y}})}\mathbb{G}_{\widehat{\Lambda},\mathbb{X}}(x^{(1)}, x^{(2)}) - \sqrt{k_{\mathbb{X}}/(k_{\mathbb{X}} + k_{\mathbb{Y}})}\mathbb{G}_{\widehat{\Lambda},\mathbb{Y}}(v^{(1)}, v^{(2)}).\end{aligned}$$

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

Proof of Proposition (3.4).

We show that individual tests are asymptotically undersized. Due to this, grid–specific p–values need not to be perfectly dependent.

For Test (1) with M_j subsets, denote the test statistic corresponding to the minimal p–value by S_j^*, and the denote the factor of S_j^* by $\upsilon_j := (k_{\mathbb{X},j} + k_{\mathbb{Y},j})/(k_{\mathbb{X},j}k_{\mathbb{Y},j})$, where $k_{\mathbb{X},j}, k_{\mathbb{Y},j}$ denote the *realized* effective sample sizes of $\mathbb{X}$ and $\mathbb{Y}$ in the subinterval corresponding to p_j^*. Obviously,

$$\upsilon_j \leq \upsilon := (k_{\mathbb{X}} + k_{\mathbb{Y}})/(k_{\mathbb{X}}k_{\mathbb{Y}}),$$

and υ_j decreases in both $k_{\mathbb{X},j}$ and $k_{\mathbb{Y},j}$, while $k_{\mathbb{Z},j}, k_{\mathbb{Y},j}$ both decrease in the fineness of the grid ($j \to \infty$): The finer the grid, the smaller $k_{\mathbb{Z},j}$, i.e. less observations are in each subinterval. For υ, a test against copula equality would be asymptotically exact, i.e. $\mathbb{P}(p \leq \alpha | H_0) \to \alpha$, under ($A1^S$)–($A4^S$).

Realize that — under the null — the test statistic integrates over squared differences of centralized normal variables. We may approximate the right tail of the null distribution by a centered χ^2 distribution with, say, $\varpi_j > 0$, degrees of freedom; see Beran (1975). Hence, for x large enough, test size can be approximated as $\alpha_j := \mathbb{P}(\widetilde{S}_j^* > x | H_0) \sim \chi^2(\varpi_j), j = 1, ..., J$, where $\widetilde{S}_j^*$ denotes the theoretical test statistic corresponding to the adjusted p–value $\tilde{p}_j^*$. Also, for the variance of the test statistic, it holds that $\mathbb{V}(\widetilde{S}_j) = \mathcal{O}(\upsilon_j^2)$, i.e. the variance increases as grid fineness increases ($j \downarrow$) and less observations enter the estimation ($k_{\mathbb{X},j}, k_{\mathbb{Y},j} \downarrow, \upsilon_i \uparrow$). According to the Markov inequality, with fixed critical values x_j,

$$\alpha_j := \mathbb{P}(\widetilde{S}_j^* \geq x_j | H_0) \leq \frac{\mathbb{E}(\widetilde{S}_j^*)}{x_j} = \frac{\mathbb{V}(\widetilde{S}_j^*)/2}{x_j}) = \mathcal{O}(\mathbb{V}(\widetilde{S}_j^*)),$$

i.e. under the null, realized test sizes α_j decrease with rate υ_j^2. Furthermore, grid–specific p–values are continuous and uniformly distributed. Now, Sklar's Theorem implies their dependence under the null can be characterized by a copula, C_α, say, i.e. $C_\alpha(\boldsymbol{u}) = \mathbb{P}(\tilde{p}_1^* \leq u^{(1)}, ..., \tilde{p}_J^* \leq u^{(J)} | H_0)$. Under the null, the FWER in terms of the copula C_α, is given by

$$\mathbb{P}(\cup_{j=1}^J \tilde{p}_j^* \leq \alpha | H_0) = 1 - C_\alpha(1 - \alpha_1, ..., 1 - \alpha_J), . \tag{3.6}$$

For illustration, let nearly any observations at all fall in relevant subintervals, i.e. $\forall j : \upsilon_j \approx 0$,

$$1 - C_\alpha(1 - \alpha_1(\upsilon_1), ..., 1 - \alpha_J(\upsilon_J)) \downarrow 1 - C_\alpha(1, ..., 1) = 0,$$

and Test (2) naturally obeys the α–limit in this unrealistic case. In all other cases, as $J \to \infty$, for FWER control $\mathbb{P}(\cup_{j=1}^J p_j^* \leq \alpha | H_0) \leq \alpha$, it must hold that

$$1 - C_\alpha(1 - \alpha_1(\upsilon_1), ..., 1 - \alpha_J(\upsilon_J)) \nearrow \alpha_\star(\upsilon_\star),$$

where $\alpha_\star := \max(\alpha_1(\upsilon_1), ..., \alpha_J(\upsilon_J)) \to 0$. This means, for FWER control, the copula C_α must approach its upper bound — $(\alpha_1, ..., \alpha_J)$ must be nearly perfectly dependent — but the upper bound does not need to be exactly obtained due to $\alpha_j \to 0, j = 1, ..., J$. □

4 Testing against intra–tail asymmetries in financial time series

This chapter is based on Bormann (2016).

Abstract

When univariate tails contribute asymmetrically to the tail dependence function, tail dependence is non–exchangeable. In quantitative risk management, such intra–tail asymmetries rule out popular parametric dependence models, like elliptical and Archimedean (tail) copulas, as they would induce a model error. We propose a simulation–based Cramér–von Mises test and a maximum likelihood–based test against intra–tail asymmetry for financial time series. A simulation study for sequentially dependent copula–based Markov chains documents the tests' satisfactory finite sample properties. For foreign exchange rate pairs, during the last 15 years, we estimate intra–tail asymmetry dynamics and reveal that up to 20% of the pairs under study exhibit non–exchangeable tails. This finding renders standard (tail) copula models inappropriate for these instances.

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

JEL classification: C12, C53, C58

4.1 Introduction

Tail copulas and related functions are a proven statistical tool to assess dependence between tail events, e.g. Poon et al. (2004), Garcia & Tsafack (2011). Non–exchangeability of tail dependence arises when tail dependence is not symmetric with respect to the ordering of marginal components. We denote this phenomenon intra–tail asymmetry as it addresses possible skewness of tail dependence functions within a bivariate tail. (Tail) copulas are not only used to measure contemporaneous dependence, but can also be used to evaluate serial dependence of a univariate time series (Chen & Fan (2006)).

For serial dependence in univariate time series, intra–tail asymmetry may also be called *non–reversibility* of tail events, see Beare (2010), Beare & Seo (2014), who focus on the entire copula instead of its asymptotic (tail) regions. Hong et al. (2009) connect intra–tail asymmetry with tail Granger causality as intra–tail asymmetry may improve predictive power of tail measures. Ultimately, this can be used to improve trading strategies during extreme market conditions, i.e. this specific type of tail dependence can be used to hedge against certain extreme events. As mentioned in Beare (2010) and Beare & Seo (2014), for copula–based Markov chains, copula non–exchangeability occurs when many (few) small (large) increases are followed by few (many) large (small) decreases in the time series. Examples are business cycles, and oligopolistic price settings where a steady, monotonous time series evolution is erupted by a short–term shock. Correspondingly, tail copula non–reversibility, i.e. intra–tail asymmetry, occurs when one extremely large (small) extreme is typically followed by a smaller (larger) extreme.

For bivariate, i.e. cross–sectional, time series, intra–tail asymmetry implies that extreme events of marginal $X^{(1)}$ may tend to drag marginal $X^{(2)}$ to its tail, while vice versa this effect may be less pronounced. In other words, one component $X^{(i)}$ is more important to joint tail dependence than $X^{(j)}, i \neq j$. Importantly, for statistical modeling of tails, symmetric tail copula models, such as elliptical and Archimedean tail copulas, are inadequate when intra–tail asymmetry is on hand. For example, Schmidt (2002) exploits elliptical tail structures to improve estimation of the tail dependence coefficient. Hence, pre–testing for intra–tail asymmetry may rule out many popular symmetric tail copula models, eliminating one possible source of model misspecification. Also, tail dependence non–exchangeability implies the tail copula of interest is *unique* in that there is no tail copula with an identical tail structure, see

Bormann & Schienle (2016), Proposition (1).

These arguments substantiate a practical need for statistical tests that identify intra–tail asymmetry for cross–sectional and serial data. This chapter proposes and compares two types of tests against intra–tail asymmetry: A computationally intensive simulation–based non–parametric test and a parametric test based on the maximum likelihood machinery that is computationally less demanding. The non–parametric test exploits recent empirical tail copula results and performs a Cramér–von Mises test, integrating squared differences between the tail copula and the tail copula with switched components. Test distributions can readily be approximated by the (tapered) multiplier bootstrap, see Bücher & Dette (2013), Bücher & Ruppert (2013). Besides weak standard extreme value assumptions, the test is flexible in that it has power against any form of intra–tail asymmetry, i.e. it is independent of the parametric form of the underlying theoretical tail copula, and even works if the tail copula is not smooth. Furthermore, the maximum likelihood–based test fits preselected asymmetric tail copula models and tests against equality of asymmetry parameters. We employ the peaks over threshold–type maximum likelihood approach by Stephenson & Tawn (2005); test distributions directly follow from standard maximum likelihood arguments. In contrast to the non–parametric approach, the maximum likelihood test requires smoothness of the tail copula, which is violated for e.g. factor models.

Relatedly, Kojadinovic & Yan (2012) propose a non–parametric Cramér–von Mises test based on Pickands dependence function for i.i.d. data. However, the independence assumption is inappropriate for possibly dependent financial data. Yet, their test is closely related to our non–parametric test, but we specifically adjust the test to address eventual serial dependence in the data. Also, Beare & Seo (2014) propose a Kolmogorov–Smirnov–type test for copula non–exchangeability in copula–based Markov chains. Berg (2009) finds that, specifically for copula models, Kolmogorov–Smirnov tests are less powerful than Cramér–von Mises tests. Furthermore, their test focuses on the entire copula and not specifically on the tail copula; in a tail setting, much less observations are available for estimation which would decrease the power of the Kolmogorov–Smirnov test relative to the Cramér–von Mises test even more. Our tests hence complement and significantly extend current tests for financial data and for tail dependence, respectively.

In a small simulation study with copula–based Markov chains, we find satisfying finite–sample properties for both tests. Moreover, for the considered types of data generating processes, the impact of model risk of the maximum likelihood approach appears to be small, rendering it even more powerful than the non–parametric test in small samples. In an empirical application, we can identify intra–tail asymmetry for up to 20% of pairs of the most important foreign exchange rates. This implies standard, i.e. symmetric, tail copula models may be inadequate, and a more accurate modeling of joint extremes can be achieved by allowing for intra–tail asymmetry.

This chapter is structured as follows. Section 4.3 introduces our tests including test distributions and implementations. Section 4.4 provides a simulation study, and Section 4.5 studies intra–tail asymmetry of foreign exchange rates while Section 4.6 concludes. The Appendix provides additional details to maximum likelihood estimation for tail dependence functions.

4.2 Cross–sectional and intertemporal intra–tail asymmetry

We denote a bivariate (random) return vector by $\mathbb{X} = (X^{(1)}, X^{(2)})$, and assume its joint distribution function $F_{\mathbb{X}}$ is in the domain of attraction of a bivariate extreme value distribution $G(x^{(1)}, x^{(2)})$. Its continuous marginal distributions $F_i, i = 1, 2$, are consequently in the max–domain of univariate extreme value distributions, i.e.

$$
\begin{aligned}
(\max(X_1^{(i)}, ..., X_n^{(i)}) - b_n^{(i)})/a_n^{(i)} \xrightarrow{d} W, \\
W \sim G_i, \\
G_i(x) = \exp\left(-(1 + \gamma^{(i)}(x - \mu^{(i)})/\sigma^{(i)})_+^{-1/\gamma^{(i)}}\right),
\end{aligned}
\tag{4.1}
$$

where $\gamma^{(i)}, \mu^{(i)}, \sigma^{(i)}$ denote the shape, location and scale parameter of margin i, and $a_n^{(i)}, b_n^{(i)}$ are appropriately chosen normalizing constants, see de Haan & Ferreira (2006) for details. Recall the definition of the theoretical upper tail copula of $\mathbb{X}$, and its empirical version, respectively,

$$
\Lambda_{\mathbb{X}}^U(x^{(1)}, x^{(2)}) = \lim_{t \to 0} t^{-1} \mathbb{P}(X^{(1)} > F_1^{-1}(1 - tx^{(1)}), X^{(2)} > F_2^{-1}(1 - tx^{(2)})), (x^{(1)}, x^{(2)}) \in \mathbb{R}_+^2,
$$

and

$$
\widehat{\Lambda}_{\mathbb{X}}^U(x^{(1)}, x^{(2)}) = \frac{1}{k} \sum_{m=1}^{n} 1 \left\{ X_m^{(1)} > \widehat{F}_1^{-1}(1 - (k/n)x^{(1)}), X_m^{(2)} > \widehat{F}_2^{-1}(1 - (k/n)x^{(2)}) \right\},
$$

with sample size n, effective sample size k, and empirical marginal distribution functions $\widehat{F}_i(x) = \frac{1}{n+1}\sum_{j=1}^{n} 1\{X_j^{(i)} \leq x\}$. The (empirical) lower tail copula, Λ^L, is defined accordingly,

$$\widehat{\Lambda}_{\mathbb{X}}^{L}(x^{(1)}, x^{(2)}) = \frac{1}{k}\sum_{m=1}^{n} 1\left\{X_m^{(1)} < \widehat{F}_1^{-1}((k/n)x^{(1)}), X_m^{(2)} < \widehat{F}_2^{-1}((k/n)x^{(2)})\right\}.$$

All results are also valid for the upper and lower tail copula and thus we omit the upper indices U and L, respectively. The tail copula is homogeneous, $\Lambda(a\boldsymbol{x}) = \Lambda(\boldsymbol{x}), a \in \mathbb{R}$, and it is hence sufficient to study the tail copula only on the unit simplex, denoted by $\mathcal{S}$, see, among others, Huang (1992). We are interested in the following type of tail dependence.

Definition 4.1 (Intra–tail asymmetry). *A bivariate return vector* $\mathbb{X}$ *is intra–tail asymmetric if there exists a set* $I \subset \mathbb{R}_+^2$ *with* $\mathbb{P}(I) > 0$ *such that*

$$\Lambda_{\mathbb{X}}(x^{(1)}, x^{(2)}) \neq \Lambda_{\mathbb{X}}(x^{(2)}, x^{(1)}), \text{ for } (x^{(1)}, x^{(2)}) \in I.$$

That is to say, there exists an entire set I, with non–zero (probability) measure, where the tail copula is not symmetric with respect to its arguments. Intra–tail asymmetry is present if extremes of $X^{(1)}$ have a different impact on extremes of $X^{(2)}$ than vice versa. For example, let $x^{(1)} = 0.2, x^{(2)} = 0.8$ and $t \approx 0.05$: In terms of Value at Risk (VaR) events, intra–tail asymmetry is present if $\{X^{(1)} > VaR_{X1}(0.01)\} \cap \{X^{(2)} > VaR_{X2}(0.04)\}$ is differently likely from $\{X^{(1)} > VaR_{X1}(0.04)\} \cap \{X^{(2)} > VaR_{X2}(0.01)\}$. Testing against tail copula non–exchangeability boils down to the null hypothesis of

$$H_0 : \Lambda(x^{(1)}, x^{(2)}) = \Lambda(x^{(2)}, x^{(1)}), \text{ a. s.} \forall (x^{(1)}, x^{(2)}) \in \mathcal{S}.$$

In a univariate time series, bivariate tail copulas can also measure partial tail dependence between X_t and $X_{t-h}, h \in \mathbb{N}$. Following Chen & Fan (2006), we model a one–dimensional time series $\{X_t\}$ as a stationary first–order Markov process by capturing first order dependence with a copula C, i.e. by a copula for the joint distribution of $(X_{t-h}, X_t), h = 1$. We denote this type of copula as intertemporal copula of first order. Assume strong stationarity for X_t, i.e. for all t, the distribution function of X, F, remains constant. To be precise, the intertemporal copula of first order is defined as $C_{h=1}(x^{(1)}, x^{(2)}) := \mathbb{P}(F(X_{t-1}) \leq x^{(1)}, F(X_t) \leq x^{(2)})$. Intertemporal tail dependence between X_t and X_{t-1} is completely determined by the tail copula that evolves from $C_{h=1}$. The first order tail copula directly follows as

$$\Lambda_{h=1}(x^{(1)}, x^{(2)}) := \lim_{v \to 0} v^{-1}\big(x^{(1)}/v + x^{(2)}/v - 1 + C_{h=1}(1 - x^{(1)}/v, 1 - x^{(2)}/v)\big). \tag{4.2}$$

Extensions to higher order dependence ($\Lambda_{h>1}$) are possible. Yet, in standard financial time series, first order tail dependence should cover the most important part of autoregressive extreme dependence. Beare (2010) shows that such stochastic processes are ρ–mixing under mild conditions, e.g. also allowing for tail dependent and asymmetric copulas. This time series model is attractive as modeling the temporal extreme dependence, $\Lambda_{h=1}$, and modeling the marginal distribution of $\{X_t\}$, F, are completely separated. Furthermore, and similar to GARCH–type models, clusters of extremes can be modeled if $\Lambda_{h=1} > 0$. Intertemporal intra–tail asymmetry is similarly defined as in the cross–sectional case.

Definition 4.2 (Intertemporal intra–tail asymmetry)**.** *A univariate stochastic process X_t is intertemporally intra–tail asymmetric if there exists a set $I \subset \mathbb{R}_+^2$ with $\mathbb{P}(I) > 0$ such that*

$$\Lambda_{h=1}(x^{(1)}, x^{(2)}) \neq \Lambda_{h=1}(x^{(2)}, x^{(1)}),\ \textit{for}\ (x^{(1)}, x^{(2)}) \in I.$$

We only study intra–tail asymmetry for one lag, while extensions of intra–tail asymmetry to higher lags are immediate; see Hong et al. (2009) for Granger causality of tail events for arbitrary lag length. Intertemporal intra–tail asymmetry (of first order) is on hand if the tail copula of (X_{t-1}, X_t) is non–exchangeable. This occurs when extremes in $t-1$ tend to disproportionately often trigger *more extreme* extreme events in t ($\Lambda_{h=1}(x^{(1)}, x^{(2)}) < \Lambda_{h=1}(x^{(2)}, x^{(1)})$), i.e. extremes behave progressively. On the other hand, when $\Lambda_{h=1}(x^{(1)}, x^{(2)}) > \Lambda_{h=1}(x^{(2)}, x^{(1)})$, extremes tend to be followed by *less extreme* extreme events , i.e. extremes behave regressively. Recently, Hong et al. (2009) extend the concept of Granger causality to extreme spillovers in a bivariate time series setup. Granger causality in the tail, in a cross–sectional context, is defined as extreme events of $X^{(1)}$ can improve predicting contemporaneous extreme events of $X^{(2)}$. Similarly, Granger causality in a copula–based Markov chain setting means that extremes of X_{t-1} can improve predicting extreme events of X_t.

4.3 Testing procedures

4.3.1 Non–parametric testing

The null hypothesis of intra–tail symmetry can be translated to

$$H_0 : \Lambda(x^{(1)}, 1 - x^{(1)}) = \Lambda(1 - x^{(1)}, x^{(1)}), x^{(1)} \in [0, 1], \tag{4.3}$$

which, in the copula–based Markov chain setting, reads as

$$H_0 : \Lambda_{h=1}(x^{(1)}, 1 - x^{(1)}) = \Lambda_{h=1}(1 - x^{(1)}, x^{(1)}), x^{(1)} \in [0, 1]. \tag{4.4}$$

For the intertemporal setting, we propose the following tail copula–based test statistic

$$D = \frac{k}{2} \int_{[0,0.5)} (\Lambda_{h=1}(\phi, 1 - \phi) - \Lambda_{h=1}(1 - \phi, \phi))^2 d\phi,$$

which evaluates squared differences of the tail copula and its flipped version over $[0, 1]$. In the copula–based Markov chain setting, the upper empirical tail copula is

$$\widehat{\Lambda}_{h=1}(x^{(1)}, x^{(2)}) = \frac{1}{k} \sum_{m=2}^{n} 1 \left\{ \widehat{F}(X_{m-1}) > 1 - (k/n)x^{(1)}, \widehat{F}(X_m) > 1 - (k/n)x^{(2)} \right\}.$$

Setting $\Lambda_{h=1} = \Lambda_{\mathbb{X}}, X_{m-1} = X_m^{(1)}, X_m = X_m^{(2)}, \widehat{F}_i, i = 1, 2$, covers the cross–sectional case. We impose the following assumptions, that ensure existence of the tail copula and, ultimately, consistency of the test distribution even in the serially dependent case. Assumptions are as in Bücher & Ruppert (2013), Bücher & Dette (2013), Bormann & Schienle (2016). Notation is kept loose to cover both cross–sectional (bivariate) and the intertemporal (univariate) case; results are valid for both the cross–sectional and the univariate case.

Assumptions 4.1.

($A1^{ITA}$) $\mathbb{X}$ *exhibits tail dependence,* $\Lambda > 0$.

($A2^{ITA}$) $k \to \infty$ *and* $\frac{k}{n} \to 0$ *for* $n \to \infty$.

($A3^{ITA}$) *It holds that* $|\Lambda(x^{(1)}, x^{(2)}) - (1 - tC_{\mathbb{X}}(1 - x^{(1)}/t, 1 - x^{(2)}/t))| = \mathcal{O}(A(t))$, *for* $t \to \infty$, *and some function* $A : \mathbb{R}_+ \mapsto \mathbb{R}_+$ *with* $\lim_{t\to\infty} A(t) = 0$ *and* $\sqrt{k}A(n/k) \to 0$ *for* $n \to \infty$.

($A4^{ITA}$) $\mathbb{X}$ *is strongly mixing with* $\sum_{r=1}^{\infty}(r + 1)^c\sqrt{\alpha_{\mathbb{X}}(r)} < \infty, r > 0, c = \max(28, \lfloor 2/\epsilon \rfloor + 1)$, *where* $\alpha_{\mathbb{X}}^r$ *denotes the mixing coefficient of* $\mathbb{X}$.

For consistency, the tail copula must not necessarily be continuous, see Bücher et al. (2014), i.e. this would also cover factor models for tail dependence. If ($A1^{ITA}$) is violated, non–parametric tail estimation is biased, see Schmidt & Stadtmüller (2006); ($A2^{ITA}$) requires the sample size (n) and the effective sample size (k) to increase, yet n must increase much faster. This ensures that k stays relatively small, and only the truly extreme observations are used for estimation. The second–order condition ($A3^{ITA}$) particularizes the rates of k and n that are needed to put the tail copula sufficiently close to the extreme part of the (scaled) copula. ($A4^{ITA}$) allows for serially dependent data whose serial dependence wears out sufficiently fast over time. For cross–sectional dependence models, this covers AR and GARCH processes (Bücher & Ruppert (2013)). For intertemporal dependence models, first–order copula–Markov chains, that exhibit tail dependence and also intra–tail symmetry, are ρ–mixing, which implies strong mixing, and thus such models are also covered as well, see Beare (2010).

The empirical test statistic is

$$\widehat{D} = \frac{k}{2} \int_{[0,0.5)} (\widehat{\Lambda}_{h=1}(\phi, 1-\phi) - \widehat{\Lambda}_{h=1}(1-\phi, \phi))^2 d\phi.$$

Under the null, when $\Lambda_{h=1}(x^{(1)}, x^{(2)})$ and $\Lambda_{h=1}(x^{(2)}, x^{(1)})$ are identical, it should hold that $\widehat{D} \approx 0$, and the null has to be rejected if the test statistic is too large. The asymptotic test distribution directly follows from the asymptotic distribution of $\widehat{\Lambda}_{h=1}$ for α–mixing time series (see Bücher & Ruppert (2013), Bücher & Dette (2013), Bormann & Schienle (2016)) and the continuous mapping theorem.

Proposition 4.1. *Under ($A1^{ITA}$)–($A4^{ITA}$),*

$$\widehat{D} \overset{H_0}{\to} \mathbb{Q} := \int_{[0,0.5)} (\mathbb{G}(\phi, 1-\phi) - \mathbb{G}(1-\phi, \phi))^2 d\phi,$$

where $\mathbb{G}(x^{(1)}, x^{(2)})$ denotes a centered Gaussian process with covariance

$$\mathbb{E}\left(G(x^{(1)}, x^{(2)})\mathbb{G}(y^{(1)}, y^{(2)})\right) = \Lambda(\min(x^{(1)}, y^{(1)}), \min(x^{(2)}, y^{(2)})).$$

The proof is straightforward using the functional delta theorem, see Kojadinovic & Yan (2012). Kojadinovic & Yan (2012) introduce a Cramér–von Mises test against tail dependence non–exchangeability based on Pickands dependence function for i.i.d. data. In contrast, we formulate the test in terms of the tail copula which allows for serially dependent data as follows. Due to the complexity of the limiting process, the null distribution has to be simulated. We employ the tapered multiplier bootstrap, which simulates a series of so–called multipliers $(\xi_{j,n})_{j=1,...,n}$. In the i.i.d. case, bootstrap estimates of the tail copula are generated by

$$\widehat{\Lambda}^{(b)}_{h=1}(x^{(1)}, x^{(2)}) = \frac{1}{k}\sum_{m=2}^{n} \tilde{\xi}_i 1\left\{X_{m-1} \geq \tilde{F}^{-1}(1-(k/n)x^{(1)}), X_m \geq \tilde{F}^{-1}(1-(k/n)x^{(2)})\right\},$$
$$\tilde{\xi}_i = \xi_i/\overline{\xi}, i = 1, ..., n,$$
$$\tilde{F}(x) = \frac{1}{n}\sum_{m=1}^{n} \tilde{\xi}_i 1\left\{X_m \leq x\right\}.$$

In the mixing case, $(\xi_{j,n})_{j=1,...,n}$ has to be adjusted as follows.

Assumptions 4.3.

($A5^{ITA}$) *The tapered block multiplier process $(\xi_{j,n})_{j=1,...,n}$ is strictly stationary, has bounded moments, is independent of $\mathbb{X}$, and positively $cl(n)$–near epoch dependent, where c is some constant and $l(n) \to_{n\to\infty} \infty, l(n) = o(n)$, and for all positive valued integers j, h assume $\mathbb{E}(\xi_{j,n}) = \mu > 0, \mathbb{V}(\xi_{j,n}, \xi_{j+h,n}) = \mu^2 v(h/l(n))$ and v is a bounded function symmetric around zero, and w.l.o.g. $\mu = 1, v(0) = 1$.*

($A6^{ITA}$) *For the tapered block length $l(n) \to \infty$, where $l(n) = \mathcal{O}(n^{1/2-\epsilon}), 0 < \epsilon < 0.5$.*

($A5^{ITA}$) and ($A6^{ITA}$) give conditions on the multiplier process to achieve consistency of the simulated null distribution. The independent (multiplier) bootstrap is not capable to reconstruct the distribution of non–i.i.d. data. Thus, the (multiplier) bootstrap must be adjusted to also capture serial dependence. This is achieved by simulating serially dependent multipliers.

Now, the limiting null distribution of $\widehat{D}$ can be consistently approximated by the tapered multiplier bootstrap techniques in Bücher & Dette (2013), Bücher & Ruppert (2013). We adopt implementation details from Kojadinovic (2015) and Bormann & Schienle (2016).

4.3.2 Parametric testing

A parametric test against intra–tail asymmetry can be performed by maximum likelihood imposing some parametric model for $\Lambda_{h=1}$ ($\Lambda_{\mathbb{X}}$). This model must allow for both intra–tail asymmetry and intra–tail symmetry. Then, fitted asymmetry parameters are compared with the case of symmetry. Popular models for intra–tail asymmetry are the asymmetric logistic tail copula as it generalizes the logistic tail copula (Tawn (1988), Coles & Tawn (1994)), and the negative logistic tail copula (Joe (1990)). In two dimensions, both models remain parsimonious with only three parameters.

Stephenson & Tawn (2005) extend standard maximum likelihood estimation of tail dependence to include occurrence time information of joint extremes. This allows to approximate only the tail and not the entire distribution by an extreme value distribution. Consequently, this approach also depends on the choice of effective sample size k as k determines which observations to consider extreme. Introducing the choice of effective sample size makes this maximum likelihood approach directly comparable to the non–parametric tail copula test as both methods use the same part of the sample. In the Appendix, we provide details of maximum likelihood estimation of tail dependence.

On top of standard maximum likelihood regularity conditions (e.g. Amemiya (1985)), we briefly state further assumptions that are needed in an extreme value scenario to ensure asymptotic normality of the maximum likelihood estimators and, consequently, consistency of the maximum likelihood test.

Assumptions 4.4.

($A1^{ML}$) $\mathbb{X}$ *is an i.i.d. series.*

($A2^{ML}$) For all marginal decay parameters, it holds that $\gamma^{(i)} > -0.5$*, as in Equation* (4.1).

($A3^{ML}$) $\mathbb{X}$ *exhibits tail dependence, but not perfect tail dependence,* $0 < \Lambda(x^{(1)}, x^{(2)}) < \min(x^{(1)}, x^{(2)})$.

($A4^{ML}$) Λ *belongs to a parametric tail copula class that is characterized by parameter vector* $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^q, q > 0$*, and* Θ *is compact and convex.*

($A5^{ML}$) Second derivatives of Λ *exist and are continuous.*

($A1^{ML}$) excludes many financial time series, but, in general, asymptotic maximum likelihood properties still carry over as long as the data generating process is β–mixing with polynomial rate, see Joe (2001). For first order copula–Markov chains, however, Beare (2010) points out that $\Lambda_{h=1} > 0$ rules out even β–mixing, i.e. this assumption is violated. Nonetheless, simulations with non–β–mixing data document consistency of the maximum likelihood–based test in finite samples. ($A2^{ML}$) ensures maximum likelihood estimation is regular, which is needed to establish standard maximum likelihood consistency result; see Smith (1985). This assumption allows for Gumbel– and Fréchet–type tails in F, while restricting F to have no finite end-points. The latter, however, is only of minor importance for financial (log) returns which are typically modeled by the Gaussian or heavier–tailed distributions, e.g. Cont (2001). ($A3^{ML}$), in many parametric models, excludes dependence parameters of $\Lambda(;\boldsymbol{\theta})$ living on the boundary which typically causes perfect tail dependence, or no tail dependence. This ensures identifiability of marginal parameters, see Beirlant et al. (2004). Also, ($A5^{ML}$) is needed to guarantee a regular likelihood function. For the non–parametric test, we need no assumptions an the smoothness of the tail copula at all, see Bücher et al. (2014). In the Appendix, we provide details on maximum likelihood estimation of tail copulas.

For the admissible parameter space Θ, denote by $\Theta_0 \subset \Theta$ the set of parameter constellations for which the null of intra–tail symmetry (Equation (4.3)) holds. The null and the alternative hypothesis can thus be rewritten as

$$H_0 : \boldsymbol{\theta} \in \Theta_0, \qquad H_1 : \boldsymbol{\theta} \in \Theta_1, \tag{4.5}$$

where Θ_1 denotes the complement of Θ_0. To exemplify the maximum likelihood test, we shortly discuss two popular asymmetric models for the tail dependence function. Notably, both nest the symmetric case.

The asymmetric logistic copula is defined as

$$C_{h=1}^{AL}(x^{(1)}, x^{(2)}) = \exp\left(-\left((1-\psi^{(1)})x^{(1)} + (1-\psi^{(2)})x^{(2)} + \left[(\psi^{(1)}x^{(1)})^{\theta} + (\psi^{(2)}x^{(2)})^{1/\theta}\right]^{\theta}\right)\right),$$

and the tail copula directly follows as

$$\begin{aligned}\Lambda_{h=1}^{AL}(x^{(1)}, x^{(2)}) = x^{(1)} + x^{(2)} - \Big(&(1-\psi^{(1)})x^{(1)} + (1-\psi^{(2)})x^{(2)} \\ &+ \big[(\psi^{(1)}x^{(1)})^{1/\theta} + (\psi^{(2)}x^{(2)})^{1/\theta}\big]^{\theta}\Big),\end{aligned} \tag{4.6}$$

with parameter vector $\boldsymbol{\theta} = (\psi^{(1)}, \psi^{(2)}, \theta)$. The asymmetry parameters $\psi^{(1)}$ and $\psi^{(2)}$ govern the impact of X_{t-1} and X_t, respectively, on the symmetry of the tail copula. Within the asymmetric logistic model, the null hypothesis (Equations (4.4) and (4.5)) is equivalent to

$$H_0 : \psi^{(1)} - \psi^{(2)} = 0,$$

as only then the tail copula is symmetric. Under ($A1^{ML}$)–($A5^{ML}$), test statistic and asymptotic test distribution follow as

$$\widehat{W} = \frac{\widehat{\psi}^{(1)} - \widehat{\psi}^{(2)}}{\sigma^2_{\widehat{\psi}^{(1)} - \widehat{\psi}^{(2)}}} \overset{H_0}{\to} N(0,1), k, n \to \infty, k/n \to 0.$$

The bivariate asymmetric negative logistic copula has a tail copula of the form

$$\Lambda^{ANL}_{h=1}(x^{(1)}, x^{(2)}) = ((\psi^{(1)}x^{(1)})^{-\theta} + (\psi^{(2)}x^{(2)})^{-\theta})^{-1/\theta}, \theta \in [0, \infty), \psi^{(i)} \in [0,1], i = 1, 2,$$

where $\psi^{(i)}$ governs the degree of asymmetry of margin i; see Joe (1990) for further details. Maximum likelihood estimation of both the asymmetric logistic and the asymmetric negative logistic model is implemented in the R–package `evd`, Stephenson (2002). Importantly, maximum likelihood estimation depends on the marginal distributional properties, i.e. estimates may drastically vary when F_i changes. Hence, for maximum likelihood estimation we throughout operate with Gumbel marginals.

4.4 Simulation study

In this section, for first–order copula–based Markov chains, we compare finite sample properties of the non–parametric test with the parametric maximum likelihood–based test. To model intertemporal tail dependence, we simulate $\{X_t\}$ according to Model (4.2) with $\Lambda_{h=1} = \Lambda^{ANL}_{h=1}(\mathbf{x}; \psi^{(1)}, \psi^{(2)}, \theta)$, and choose F, the marginal distribution of X_t, as the standard Gumbel distribution function. The Gumbel distribution attracts (standardized) maxima of the Normal distribution and has tail index $\gamma = 0 > -0.5$. Hence, it fulfills ($A2^{ML}$), while e.g. the uniform distribution violates this assumption and renders the parametric test inconsistent. Note, the choice of (marginal) distribution model is only important for the parametric case, and we do not investigate how fat tails influence test performance of the maximum likelihood approach.

Importantly, to assess the impact of model misspecification for the maximum likelihood test, we assume the asymmetric logistic copula is the true data generating process, i.e. we deliberately misspecify the parametric test to analyze whether it is still consistent under a violation of a standard maximum likelihood assumption.

We vary asymmetry ($\psi^{(1)}, \psi^{(2)}$) and dependence (θ) parameters to study test performances for various types of nulls and violations thereof. In particular, for studying test size, we set $\theta \in \{0.301, 0.43, 0.575, 0.756, 1\}$, while fixing $\psi^{(1)} = \psi^{(2)} = 1$, amounting to tail dependence coefficient of $\iota := \Lambda(1,1) \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$. The null is violated whenever $\psi^{(1)} \neq \psi^{(2)}$. Concerning statistical power, we fix $\theta = 10$, and iterate the asymmetry vector $\psi = (\psi^{(1)}, \psi^{(2)})$ over six values that amount to tail dependence coefficients of $\iota \in \{0.125, 0.25, 0.50, 0.85, 0.125, 0.25\}$; for details see Table (4.2). We set $n \in 500, 1000, 1500, k/n \in \{0.1, 0.15, 0.25, 0.3\}$, nominal test level $\alpha = 0.05$, and simulations are repeated 500 times. Table (4.1) shows empirical test sizes, and Table (4.2) contains empirical test power. We find both tests perform reasonably well.

Remarkably, for the parametric test, the choice of a wrong tail copula model has little impact on empirical test power and size. Simulated power results, assuming the *true* tail copula $\Lambda^{ANL}_{h=1}$ (not reported), are very similar to the misspecified scenario here. We believe, for asymmetric and differentiable tail copulas, such as the asymmetric and the negative logistic tail copula, the model choice is only minor. In contrast, an incorrectly specified model might severely distort test results when the true tail copula is not differentiable, for example, if the tail copula stems from a max factor model. Then, the non–parametric test is still consistent (Einmahl et al. (2012)), and thus works without any constraints. Conversely, for non–differentiability of the tail copula a likelihood function does not exist, and approximations may only yield inconsistent results.

Both the tail copula and the maximum likelihood test typically obey the α–limit in case of a true null, independent of the strength of tail dependence. Interestingly, the employed Markov processes are neither i.i.d. nor β–mixing, violating one core maximum likelihood assumption, yet we do not observe major violations of the α–limit. For both tests, empirical rejection probabilities are closest to α for $k/n = 0.3$ – else, the non–parametric test appears slightly undersized, while the parametric test overrejects for small values of k/n, i.e. $k/n < 0.2$.

Concerning power, both tests appear to be consistent, while the parametric tests typically outperforms the non–parametric test. Excluding four cases where numerical optimization fails too often, the maximum likelihood test is more powerful especially for small sample sizes ($n = 500$). That is, the maximum likelihood test is to be preferred although numerical problems might request careful manual implementation. For $n > 1000$, both tests are equally powerful.

Table 4.1: Empirical test size for the negative logistic tail copula both for the non–parametric and the parametric test with varying dependence parameter, $\theta \in \{0.301, 0.43, 0.575, 0.756, 1\}$, and asymmetry parameters fixed to one, $(\psi^{(1)}, \psi^{(2)}) = (1, 1)$. The significance level is $\alpha = 0.05$.

	n	$\theta = 0.301$					$\theta = 0.430$					$\theta = 0.575$				
		k/n					k/n					k/n				
		0.1	0.15	0.2	0.25	0.3	0.1	0.15	0.2	0.25	0.3	0.1	0.15	0.2	0.25	0.3
$\widehat{\Lambda}$																
	500	0.8	0.8	0.8	3.2	2.4	0.4	2.0	0.8	0.4	2.8	0.4	0.4	1.2	1.2	2.8
	1000	1.6	2.8	4.0	2.0	4.0	0.8	0.8	0.8	1.6	3.2	1.2	0.8	1.6	0.8	2.8
	1500	1.6	2.4	3.2	3.6	3.2	1.6	2.8	4.0	1.6	3.6	3.2	2.0	2.8	2.8	4.4
ML																
	500	1.6	6.8	1.2	0.8	0.4	6.2	6.4	6.8	0.0	1.6	10.0	7.6	6.4	4.8	4.0
	1000	0.5	0.0	0.4	4.0	2.8	5.1	0.0	2.4	3.6	3.2	5.3	1.6	6.0	3.6	4.0
	1500	0.4	1.6	2.4	2.8	1.6	0.4	4.4	4.0	6.8	5.2	1.6	3.2	5.2	4.4	5.6

	n	$\theta = 0.756$					$\theta = 1$				
		k/n					k/n				
		0.1	0.15	0.2	0.25	0.3	0.1	0.15	0.2	0.25	0.3
$\widehat{\Lambda}$											
	500	0.8	0.4	0.0	1.2	1.2	1.2	0.0	0.4	0.8	0.4
	1000	0.8	2.0	4.4	2.0	3.2	0.4	0.8	2.4	2.0	1.6
	1500	0.8	0.8	1.6	1.6	2.0	0.8	1.6	2.4	1.6	1.6
ML											
	500	4.0	8.0	5.6	7.2	6.0	8.0	5.6	8.5	3.3	2.0
	1000	7.2	4.0	5.6	2.4	2.8	2.8	4.0	4.8	2.0	5.6
	1500	4.0	5.6	5.2	3.2	3.6	4.8	5.2	3.2	1.6	5.6

Table 4.2: Empirical test power of the non–parametric and the parametric test with fixed dependence parameter ($\theta = 10$), and varying asymmetry parameters. For the parametric test, the asymmetric logistic tail copula is assumed, while results worsen by at most 5% when assuming a negative logistic tail copula. When maximum likelihood estimation fails in at least 50% of all cases, we treat that specific test level as missing (NA). The significance level is $\alpha = 0.05$.

		$(\psi^{(1)}, \psi^{(2)}) = (1/8, 1)$					$(\psi^{(1)}, \psi^{(2)}) = (1/4, 1)$					$(\psi^{(1)}, \psi^{(1)}) = (1/2, 1)$				
	n	k/n					k/n					k/n				
		0.1	0.15	0.2	0.25	0.3	0.1	0.15	0.2	0.25	0.3	0.1	0.15	0.2	0.25	0.3
$\widehat{\Lambda}$																
	500	35.6	64.8	66.0	77.2	84.0	82.4	97.6	98.4	98.4	100	96.4	100	100	100	100
	1000	82.4	92.8	96.0	99.2	100	99.2	100	100	100	100	100	100	100	100	100
	1500	98.0	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ML																
	500	92.1	92.8	84.6	83.9	100	97.8	100	99.5	99.1	98.7	97.3	99.5	99.7	100	100
	1000	NA	93.5	100	100	100	98.9	99.1	100	100	100	99.5	100	99.5	100	100
	1500	97.4	100	100	100	100	99.4	100	100	100	100	99.5	98.7	100	100	100

		$(\psi^{(1)}, \psi^{(2)}) = (7/8, 1)$					$(\psi^{(1)}, \psi^{(2)}) = (1/8, 1/2)$					$(\psi^{(1)}, \psi^{(2)}) = (1/4, 1/2)$				
		k/n					k/n					k/n				
		0.1	0.15	0.2	0.25	0.3	0.1	0.15	0.2	0.25	0.3	0.1	0.15	0.2	0.25	0.3
$\widehat{\Lambda}$																
	500	12.8	31.2	47.2	70.8	80.8	23.6	39.6	48.0	54.8	63.6	27.6	46.0	55.2	62.8	71.6
	1000	54.4	86.4	96.0	99.2	100	62.8	81.6	82.8	87.2	91.2	76.4	88.0	94.0	96.5	99.2
	1500	86.8	98.8	100	100	100	84.8	93.6	98.4	99.6	98.4	90.4	98.0	99.6	100	98.8
ML																
	500	NA	90.9	90.2	99.4	98.3	48.2	47.6	61.9	51.6	57.5	43.8	75.9	82.7	89.2	91.6
	1000	90.5	98.3	96.4	99.0	100	64.3	NA	NA	96.4	99.2	95.5	96.3	99.2	100	100
	1500	97.8	96.6	100	100	100	NA	70.0	99.2	100	100	90.0	97.9	100	100	100

4.5 Intra–tail asymmetries in foreign exchange rates

Returns on foreign investments change as domestic currencies change. *Extreme* exchange rates induce high volatility in foreign investments, and can cause severe losses. Hence, investors strive to minimize exposure to exchange rate extreme risks, which necessitates an appropriate quantitative modeling of dependencies between exchange rates. By detecting intra–tail asymmetries, we can rule out complete classes of popular tail copula models, such as elliptical and Archimedean models. Hence, our test can be used to improve existing statistical models for extreme dependence within foreign exchange rates.

The data set, stemming from Bloomberg, consists of daily returns of the Euro (EUR), British Pound (GBP), Canadian Dollar (CAD), Japanese Yen (JPY), New Zealand Dollar (NZD) and the Swiss Franc (CHF), all nominated in USD for the period 01/05/2001 – 02/01/ 2016. We use the same data set as in Bormann & Schienle (2016). They compare tail dependence of pairs in exchange rates and indicate some tail copula differences are due to intra–tail asymmetry. This section re–examines their conclusion by applying of our non–parametric test in order to quantify the amount of intra–tail asymmetries in the given data set.

As foreign exchange rates typically exhibit serial dependence violating ($A1^{ML}$), we only employ the non–parametric test of Section 4.3.1. Furthermore, it is unclear whether tail copulas of foreign exchange rate pairs are sufficiently smooth to justify a direct application of the parametric test, i.e. ($A5^{ML}$), too, may be violated. Also, the maximum–likelihood test must pre–estimate marginal distributions which induces additional estimation error; for details, see the Appendix. In contrast, the non–parametric test, does not need any assumption on the smoothness of the tail copulas at all, and is margin–free. In the simulations, for larger sample sizes ($n > 1000$), both tests exhibit approximately the same power of roughly 100%. Under non–ideal conditions, we believe the non–parametric test is more robust as it requires weaker assumptions. Hence, it should be sufficient to study the data only with the non–parametric test.

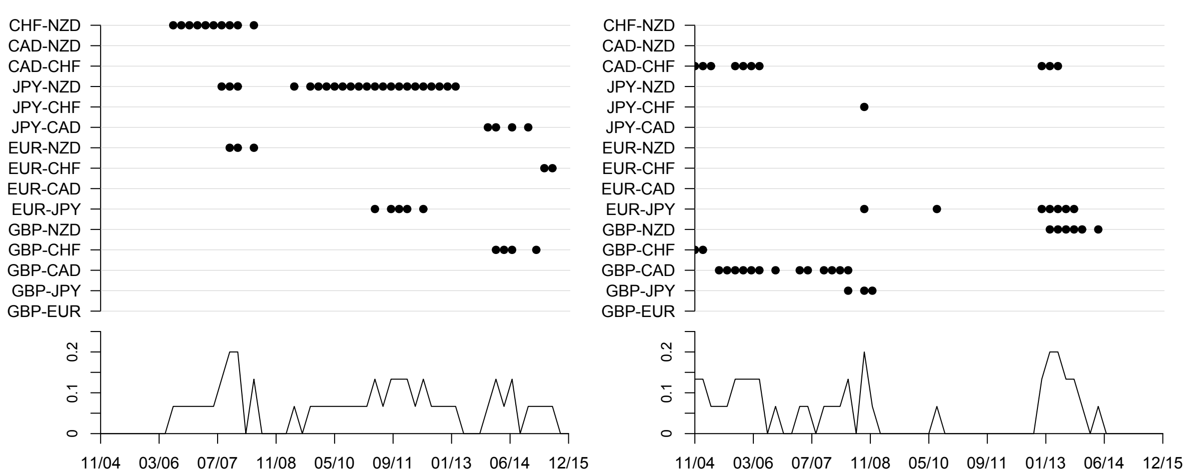
We apply a rolling window analysis with a window size of $n = 1000$ and step size of 50 to draw a fine picture of the tail (in)equality dynamics. This corresponds to a window length of approximately four years. For any pair comparison, trading days with missing data or zero returns are discarded. The effective sample size is set to $k = 0.1n$; we analyze unfiltered data and use the tapered block multiplier bootstrap to simulate the test distributions. We test against intra–tail asymmetry in each pair, within each period, and separately for upper and lower tails, i.e.

$$H_0 : \Lambda^W_{\mathbb{X}}(x^{(1)}, 1 - x^{(1)}) = \Lambda^W_{\mathbb{X}}(1 - x^{(1)}, x^{(1)}), a.s., W = U, L, x^{(1)} \in [0, 1],$$

for exchange rate pairs $\mathbb{X} = (X^{(i)}, X^{(j)})$. Figure (4.1) displays the share of intra–tail asymmetries with the 15 bivariate pairs and also shows which pairs exhibit intra–tail asymmetry.

The test reveals that only a small, yet non–neglectable share of pairs is intra–tail asymmetric, which backs the presumption in Bormann & Schienle (2016) of intra–tail asymmetries standard financial returns. The pairs JPY–NZD (lower tail) and GBP–CAD (upper tail) feature a pronounced phase of intra–tail asymmetry during 2008–13. Pairs for which our test frequently rejects (five or more times) contain an *exotic* currency, such as NZD or CAD, which are less frequently traded. This empirical phenomenon indicates that opportunities for trading strategies based on intra–asymmetric tails can mainly be found in pairs with one prominent currency (e.g. EUR, GDP, JPY) and one exotic currency (e.g. NZD, CAD). Interestingly, intra–tail asymmetry is mostly found in either lower tails or upper tails.

Figure 4.1: Intra–tail asymmetric pairs for lower (left) and upper (right) tails. Significant intra–tail asymmetries are marked by black dots (upper panel). We also plot the dynamics of the share of intra–tail asymmetric pairs among all pairs (lower panel).



Intra–tail asymmetry in at least one tail excludes entire tail copula classes. This finding provides another argument against a hasty usage of basic model classes, i.e. it excludes elliptical or Archimedean models. Those not only require symmetry within each tail, but also symmetry between the lower and the upper tail. Elliptical distributional properties are exploited in Schmidt (2002) to improve the estimation of the tail dependence coefficient. Such approaches would not remain valid in instances of intra–tail asymmetric upper and lower tails.

4.6 Conclusion

We propose two tests against tail copula non–exchangeability based on the empirical tail copula and on the maximum likelihood method, respectively. Test asymptotics are provided and implementation details are discussed. A simulation study with dependent data reveals the maximum likelihood test is more powerful, yet might suffer from numerical problems. For foreign exchange rate pairs, we find a noteworthy share of intra–tail asymmetries. This finding can improve parametric modeling, and is of interest for tail risk–based trading strategies.

4.7 Appendix

4.7.1 Maximum likelihood estimation of tail dependence

For a more complete treatment of maximum likelihood estimation of tail dependence, we refer to Huser et al (2016). The limiting distribution of the componentwise maxima of (appropriately standardized) $\mathbb{X}$, $G_{\mathbb{X}}(x^{(1)}, x^{(2)})$, is a bivariate extreme value distribution

$$G_{\mathbb{X}}(x^{(1)}, x^{(2)}) = \exp(-V_{\mathbb{X}}(x^{(1)}, x^{(2)})), x^{(1)}, x^{(2)} > 0,$$

where $V_{\mathbb{X}}$ is the so–called exponent measure that, equivalently to the tail copula, completely describes the tail dependence within $\mathbb{X}$. In particular,

$$V(x^{(1)}, x^{(2)}) = \frac{x^{(1)} + x^{(2)} - \Lambda(x^{(1)}, x^{(2)})}{x^{(1)} + x^{(2)}}.$$

The joint density of $G_{\mathbb{X}}$, $g_{\mathbb{X}}(x^{(1)}, x^{(2)}) = \frac{\partial G_{\mathbb{X}}}{\partial \mathbf{x}}$, is

$$g_{\mathbb{X}}(x^{(1)}, x^{(2)}) = \left(\frac{\partial V_{\mathbb{X}}}{\partial x^{(1)}} \frac{\partial V_{\mathbb{X}}}{\partial x^{(2)}} - \frac{\partial^2 V_{\mathbb{X}}}{\partial x^{(1)} \partial x^{(2)}} \right) G_{\mathbb{X}}(x^{(1)}, x^{(2)}),$$

requiring the existence of second derivatives of the tail dependence function.

For maximum likelihood estimation, let $V_{\mathbb{X}}$ be member of a parametric family, $V(x^{(1)}, x^{(2)}; \boldsymbol{\theta})$, with parameters $\boldsymbol{\theta} \in \mathbb{R}^{q<\infty}$. Note, the parameter vector $\boldsymbol{\theta}$ also directly determines the tail copula. The joint density $g(x^{(1)}, x^{(2)}; \boldsymbol{\theta})$ directly allows for maximum likelihood estimation for $\boldsymbol{\theta}$ when marginals are simultaneously estimated by

$$\widehat{x}^{(i)} = (-(1 + \widehat{\gamma}^{(i)}(x^{(i)} - \widehat{\mu}^{(i)})/\widehat{\sigma}^{(i)}))_{+}^{-1/\widehat{\gamma}^{(i)}},$$

according to Equation (4.1). Note that the marginal parameters $\widehat{\gamma}^{(i)}, \widehat{\mu}^{(i)}, \widehat{\sigma}^{(i)}$ directly enter the density $g(x^{(1)}, x^{(2)}; \boldsymbol{\theta})$, i.e. density and likelihood alter in the distributional properties of the marginals. Stephenson & Tawn (2005) show that the asymptotic joint density collapses to

$$g(\widehat{x}^{(1)}, \widehat{x}^{(2)}; \boldsymbol{\theta}) = -\frac{\partial^2 V(\widehat{x}^{(1)}, \widehat{x}^{(2)}; \boldsymbol{\theta})}{\partial \widehat{x}^{(1)} \partial \widehat{x}^{(2)}} G(\widehat{x}^{(1)}, \widehat{x}^{(2)}; \boldsymbol{\theta}),$$

when extremes occur simultaneously, which simplifies the log likelihood to

$$l(\widehat{x}^{(1)}, \widehat{x}^{(2)}; \boldsymbol{\theta}) = \sum_{i \in R} \log(-\partial^2 V_i(\widehat{x}^{(1)}, \widehat{x}^{(2)}; \boldsymbol{\theta})/\partial \widehat{x}^{(1)} \partial \widehat{x}^{(2)} \ G_i(\widehat{x}^{(1)}, \widehat{x}^{(2)}; \boldsymbol{\theta})),$$

where R denotes the index of all observations with a joint extreme of $X^{(1)}$ and $X^{(2)}$, i.e. $R := \{i : \{X_i^{(1)} > F_1^{-1}(1 - k/n)\} \cap \{X_i^{(2)} > F_2^{-1}(1 - k/n)\}\}$. Final estimates are found by numerical optimization of $l(\widehat{x}^{(1)}, \widehat{x}^{(2)}; \boldsymbol{\theta})$ with respect to θ, and asymptotic properties (consistency, asymptotic normality, asymptotic efficiency) follow from standard arguments, e.g. Amemiya (1985).

5 Bivariate copula comparisons with multiple testing techniques

This chapter is based on Bormann (2016).

Abstract

Comparisons of bivariate dependence structures are vital for financial risk management and form a basis for quantitative trading strategies. We propose a simple non–parametric test for comparing two bivariate copulas of financial time series. We partition the copula domain into cells and aggregate comparisons between empirical cell probabilities over a variety of grid configurations by multiple testing techniques. While existing cell–based copula tests are not consistent in scenarios when at least one copula is non–exchangeable, we study all cell combinations, rendering our test consistent for any copula family. Further, in contrast to standard tests, our test does not require a specific optimal cell choice. The test allows to pin down sample regions that cause copula inequality providing precise information on what market conditions induce copula changes. In contrast to simulation–based tests, it is computationally efficient and allows for analyses of massive data sets. Also, the test is suitable for many financial time series, such as ARMA–GARCH processes. A copula–GARCH simulation study confirms the satisfactory finite sample properties of the test. An empirical study of high–frequency return pairs shows that serial copula structures exhibit distinct periods of time variation. Our test suggests to extend dynamic models for bivariate copulas by accounting for time variation in all four tails instead of only joint upper and lower tail.

Keywords: Two–sample Goodness–of–fit, copula modeling, dependence modeling, high–frequency data,

JEL classification: C12, C53, C58

5.1 Introduction

Assessing and comparing dependence between two financial assets is crucial for financial investors, banks, and regulating agencies. Ranking pairs of financial securities according to their tendency to co–move is essential for investment decisions. For example, when hedging one's position with respect to the market, investors aim to find securities that are market neutral, i.e. as weakly dependent with the market as possible. In contrast, statistical arbitrage trading typically aims to find stock pairs that are strongly dependent in order to profit from temporary drifting apart and mean–reversion, e.g. Gatev et al. (2006). In the simplest form of statistical arbitrage, investors concentrate on only two assets ignoring higher–dimensional risk structures. Yet, even in the bivariate case, comparing entire dependence structures is a difficult task. It is well known that standard correlation–based comparisons are insufficient to model non–linear and asymmetric dependencies, see Longin & Solnik (2001), Ang & Chen (2002). The most complete measure of bivariate dependence is a bivariate copula as it models the entire joint distribution. Hence, bivariate copulas are ideally suited for detecting significant dependence inequalities by two–sample copula goodness–of–fit tests.

We propose a simple non–parametric test for the equality of two copulas that is valid for strongly mixing processes, including many standard econometric time series models such as ARMA and GARCH processes. The test partitions the domain of the copulas into cells and jointly compares corresponding empirical probabilities by means of multiple testing techniques. Thereby we can account for possible copula non–exchangeability, which is a specific type of asymmetric dependence. Non–exchangeability describes asymmetric contributions of marginal (risk) components to the joint distribution often occurring in credit risk, see McNeil et al. (2005). Moreover, we can localize where in the copula domain differences are significant, i.e. under which market conditions co–movements of return vector $\mathbb{X}$ are more likely than co–movements of return vector $\mathbb{Y}$. This additional information can be exploited by investors to refine existing trading strategies by extracting more trading signals, or by improving parametric copula models.

In contrast to standard χ^2–tests, our test is also consistent if one or both copulas are non–exchangeable. Besides financial applications, the test finds applications in any multiple pairs risk scenario where one needs to prioritize safety measures, such as hydrology, e.g. Chen et al. (2012). Our method is tailored to the bivariate case, but can be readily extended to higher dimensions. Yet, for higher dimensions, we expect finite sample properties to worsen significantly due to the curse of dimensionality of non–parametric methods. Also, our rank–based test is invariant against monotone transformations.

Comparisons of copula structures are typically based on comparing fitted parametric copula models, i.e. one–sample goodness of fit testing. Non–parametric tests compare some copula measures without imposing any structure beforehand. Overviews of existing goodness–of–fit tests can be found in Genest et al (2009), Berg (2009), Patton (2012), Jaworksi et al (2013). Throughout, the Cramér–von Mises test is typically found to be most powerful. Avoiding the detour of fitting parametric copulas, Rémillard & Scaillet (2009) directly propose a two–sample Cramér–von Mises test that is based on empirical copulas of two samples. Theoretically, this approach is most appealing. Unfortunately, it comes with practical limitations. Firstly, this type of test says little about which return vector is more dependent, and, secondly, in which market conditions copula inequality is on hand, i.e. in bear, bull or average markets, corresponding to lower, upper, and mid parts of the return distributions. That is, it remains unclear which return pair exhibits a higher idiosyncratic risk during specific market periods, and also which market conditions cause copula inequalities. Thirdly, test distributions do not exist in closed form and must be simulated which is computationally burdensome. For ultra large data sets, such as financial high–frequency data, daily comparisons of many return pairs become infeasible due to computational limitations.

Closely related to our test, many heuristic approaches define fixed grids of the copula domain and compare empirical cell probabilities by a standard χ^2–test, as in Dobrić & Schmid (2005), Hu (2006), Jondeau and Rockinger (2006), Hong et al (2007), Patton (2013). This type of test also exhibits three problems. Firstly, the choice of cells is rather arbitrary and subject to a trade–off between test precision and robustness; empirical probabilities over too few cells may represent the copula structure too coarsely while estimates are stable since sufficiently many observations fall in each cell. In case of many cells, variance increases due to the lack of observations in some cells, yet estimates are less biased. Secondly, the χ^2–test is also unable to trace market conditions with significant copula differences for the χ^2–test statistic sums up all squared cell probability differences. Thirdly, these tests typically do not account for non–exchangeability of the copula, and are inconsistent in such cases.

Our approach builds on the subset idea, but treats cells individually and robustifies test findings across differently sized grids. It compares individual empirical cell probabilities by simple Wald tests for each reasonable combination of cells. Hence, we are able to identify cells that induce copula inequality, i.e. we trace market conditions where dependence differs. Further, by comparing every reasonable combination of cell probabilities, we address possible copula non–exchangeability. However, joint testing demands an adjustment of individual p–values to restrain the overall test decision from becoming oversized. Multiple testing techniques, such as the Bonferroni adjustment, are applied to each cell p–value. We reject the null of copula equality if a single adjusted cell p–value undercuts the α–threshold. Finally, we iteratively apply this method for different cell grids. For each grid, we pick the smallest adjusted p–value, and reject the null of copula equality if at least one of the p–values is smaller than α. In such a setting, this appears to be valid: As minimum p–values are highly dependent, and the stand–alone tests become severely undersized for increasingly many cell comparisons, small effective cell sample sizes cause the variance of the test statistics to grow. Similarly, Bormann & Schienle (2016) study differences between tail dependence functions — we generalize their approach to copulas. Monte Carlo simulations confirm our approach. Our test improves on all problems of the standard χ^2–test while remaining computationally attractive.

In a rolling window framework, we apply the test to reveal intertemporal copula changes of high–frequency returns of five financial stocks during 2007–2015. We find copula structures tend to be extremely time–varying in times of economic turmoil. During slack periods, copulas remain constant. We find that exclusively univariate extreme events drive short-term copula changes. This sharply contrasts with the standard of parametric copula modeling of bivariate returns which mainly focuses on capturing (time–varying) joint tail dependence, see Patton (2006). We outline how parametric models can be adjusted to account for our finding.

This chapter is structured as follows. Section 5.2 introduces notation, presents test idea, asymptotics. and discusses implementation issues. Section 5.4 studies finite sample properties. Section 5.5 studies copula dependence of high–frequency returns, Section 5.6 concludes.

5.2 Dependence and copulas

We briefly outline core concepts of bivariate copula theory. Denote the distribution function of a bivariate random vector $\mathbb{Z} = (Z^{(1)}, Z^{(2)})$ by $F_{\mathbb{Z}}(x_1, x_2) = \mathbb{P}(Z^{(1)} \leq x^{(1)}, Z^{(2)} \leq x^{(2)}), (x^{(1)}, x^{(2)}) \in \mathbb{R}^2$. Univariate distribution functions are denoted as $F_i(x) = \mathbb{P}(Z^{(i)} \leq x), i = 1, 2$. Sklar's theorem states that for continuous $\mathbb{Z}$ there exists a unique bivariate distribution function with uniform marginals, the copula, which expresses the joint distribution $F_{\mathbb{Z}}$ in terms of the marginal distributions $F_i, i = 1, 2$. To be precise, there exists a function $C_{\mathbb{Z}} : [0,1]^2 \mapsto [0,1]$ such that

$$
\begin{aligned}
C_{\mathbb{Z}}(x^{(1)}, x^{(2)}) &:= \mathbb{P}(F_1(Z^{(1)}) \leq x^{(1)}, F_2(Z^{(2)}) \leq x^{(2)}) && (5.1)\\
&= \mathbb{P}(Z^{(1)} \leq F_1^{-1}(x^{(1)}), Z^{(2)} \leq F_2^{-1}(x^{(2)})), (x^{(1)}, x^{(2)}) \in [0,1]^2, && (5.2)
\end{aligned}
$$

and $C_{\mathbb{Z}}$ is called the copula of $\mathbb{Z}$. Since $C_{\mathbb{Z}}$ characterizes the complete joint probability distribution of $\mathbb{Z}$, $C_{\mathbb{Z}}$ captures the complete dependence between $Z^{(1)}$ and $Z^{(2)}$. This separates modeling of joint and marginal distributions. Furthermore, a copula is called non–exchangeable if $C_{\mathbb{Z}}(x^{(1)}, x^{(2)}) \neq C_{\mathbb{Z}}(x^{(2)}, x^{(1)}), (x^{(1)}, x^{(2)}) \in I, \mathbb{P}(I) > 0$, i.e. over a set I that has positive measure, $C_{\mathbb{Z}}$ is not symmetric with respect to the order of $(Z^{(1)}, Z^{(2)})$. There exist many parametric models for $C_{\mathbb{Z}}$, for example the Gauss copula, t–copula, Clayton copula, Gumbel copula or the Frank copula — see Section 5.4 for some analytical expressions and short discussions.

Estimation is typically performed in two stages. First, F_1, F_2 are estimated either parametrically or non–parametrically by the empirical distribution function. Then, the copula function can be estimated by maximum likelihood according to the parametric model. The choice of copula family is critical, and a vast literature on specification and goodness of fit tests has emerged, see Berg (2009).

If one is unwilling to restrict oneself to a specific parametric copula model, the copula can also be estimated non–parametrically using ranks. This avoids copula model risk but results in slower convergence rates, i.e. larger sample sizes are needed. We shortly present main results of non–parametric copula statistics which rely on the so–called empirical copula process.

Marginal distributions are estimated by the empirical distribution function,

$$
\widehat{F}_i(x) = \frac{1}{n+1} \sum_{j=1}^{n} 1\{Z_j^{(i)} \leq x\},
$$

with observations $Z_1^{(i)}, ..., Z_n^{(i)}, i = 1, 2$. The copula $C_{\mathbb{Z}}$ is estimated by the empirical copula $\widehat{C}_{\mathbb{Z}}$:

$$\widehat{C}_{\mathbb{Z}}(x^{(1)}, x^{(2)}) = \frac{1}{n+1} \sum_{j=1}^{n} 1 \left\{ \widehat{F}_1(Z_j^{(1)}) \leq x^{(1)}, \widehat{F}_2(Z_j^{(2)}) \leq x^{(2)} \right\}.$$

The transformation $\widehat{F}_i(Z^{(i)_j}) =: \widetilde{Z}_j^{(i)} \in [0,1], j = 1, ..., n$, yields so–called pseudo observations which are approximately uniform. Marginals are mapped into the unit square which dissolves marginal distributional properties but preserves rank–related dependencies between $Z^{(1)}$ and $Z^{(2)}$, since F_i^{-1} is a monotone transformation. In practice, $F_i, i = 1, 2$, is typically unknown and must be estimated, which induces additional variability of the empirical copula.

Asymptotics of the empirical copula process $\sqrt{n}(\widehat{C}_{\mathbb{Z}}(\mathbf{x}) - C_{\mathbb{Z}}(\mathbf{x}))$ are well–known; see, among others, Rüschendorf (1976), Fermanian et al (2004), Segers (2012), Bücher et al. (2014), who establish consistency and asymptotic normality under various regularity conditions concerning the existence of partial copula derivatives, denoted by $C_{\partial i} := \frac{\partial C_{\mathbb{Z}}(\mathbf{x})}{\partial x^{(i)}}$. Importantly, Bücher & Ruppert (2013) transfer consistency results to the serially dependent, non–i.i.d. case. See the Appendix for details.

5.3 Test idea and asymptotic properties

We impose the following assumptions for test consistency. The test allows for dynamic data with serial dependence; this also covers the i.i.d. case as a special case, which is vital for applications with financial data.

Assumptions 5.1.

($A1^C$) $\mathbb{X}_t := \{(X_t^{(1)}, X_t^{(2)})\}_{t=1}^{n_{\mathbb{X}}}, \mathbb{Y}_t := \{(Y_t^{(1)}, Y_t^{(2)})\}_{t=1}^{n_{\mathbb{Y}}}$, *are samples of ergodic and covariance–stationary stochastic processes* $\mathbb{X}, \mathbb{Y} \in \mathbb{R}^2$, *respectively.*

($A2^C$) Cross–sectional dependence within $\mathbb{X}$ *and* $\mathbb{Y}$ *is characterized by copulas* $C_{\mathbb{X}}$ *and* $C_{\mathbb{Y}}$, *respectively.*

($A3^C$) $\mathbb{X}$ *and* $\mathbb{Y}$ *are mutually independent.*

($A4^C$) For α*–mixing* $\mathbb{Z} = \mathbb{X}, \mathbb{Y}$, *it holds that*

$$\sup_t ||Z_t^{(i)} - \mathbb{E}(Z^{(i)})||_{2+\delta} < \infty, \qquad \sum_{t=1}^{\infty} \alpha_{\mathbb{Z},i}(t)^{\delta/(2+\delta)} < \infty, \qquad \delta > 0,$$

and $\mathbb{E}\left(\sum_{t=1}^{n_{\mathbb{Z}}}(Z_t^{(i)}-\mathbb{E}(Z^{(i)}))^2\right)/n_{\mathbb{Z}} \to \sigma^2_{\mathbb{Z},i} < \infty.$

Assumptions ($A3^C$)–($A4^C$) entail many linear processes, such as ARMA and GARCH models, Francq & Zakoïan (2010). Assumption ($A3^C$) facilitates estimation of the test statistic as no covariance term appears this way. This assumption may be dropped whenever estimation of the test statistic is adjusted accordingly. Assumptions ($A1^C$) and ($A4^C$) ensure consistency of unconditional empirical moments. Together with ($A4^C$), this provides $\sqrt{n}$–central limit results for serially dependent data, see Herrndorf (1984). ($A4^C$), which is directly from Herrndorf (1984), postulates a specific mixing rate, a finite variance, and that the process remains reasonably close to its mean. Notably, our test asymptotics will not depend on (non–)differentiability of the underlying copulas.

We are interested in whether joint dependence in $\mathbb{X}$ and $\mathbb{Y}$ is equal, i.e. in the null hypothesis of

$$H_0 : C_{\mathbb{X}} = C_{\mathbb{Y}}, a.s. \tag{5.3}$$

Note, this implies comparisons of the type $C_{\mathbb{X}}(x^{(1)}, x^{(2)}) = C_{\mathbb{Y}}(x^{(1)}, x^{(2)})$ as well as $C_{\mathbb{X}}(x^{(1)}, x^{(2)}) = C_{\mathbb{Y}}(x^{(2)}, x^{(1)})$, which addresses possible non–exchangeability of $C_{\mathbb{X}}$ or $C_{\mathbb{Y}}$. We divide the copula domain $[0,1]^2$ into $J^2 \in \mathbb{N}$ equally large square cells, denoted by $b_{11}, ..., b_{J1}, b_{12}, b_{22}, ..., b_{JJ}$, i.e. $b_{ij} = \{(x^{(1)}, x^{(2)}) : (i-1)/J \leq x^{(1)} < i/J, (j-1)/J \leq x^{(2)} < j/J\}, i, j = 1, 2, ..., J.$

After pseudo–transforming both samples, $\widetilde{\mathbb{X}} = (F_{1,\mathbb{X}}(X^{(1)}), F_{2,\mathbb{X}}(X^{(2)}))$, $\widetilde{\mathbb{Y}} = (F_{1,\mathbb{Y}}(Y^{(1)}), F_{2,\mathbb{Y}}(Y^{(2)}))$, we compare cell probabilities $p_{ij}^{\widetilde{\mathbb{X}}} = \mathbb{P}(\widetilde{\mathbb{X}} \in b_{ij}), p_{ij}^{\widetilde{\mathbb{Y}}} = \mathbb{P}(\widetilde{\mathbb{Y}} \in b_{ij})$ for all cell combinations that are in line with Equation (5.3). Rewriting $p_{ij}^{\widehat{\mathbb{Z}}}$ in terms of the copula yields

$$p_{ij}^{\widehat{\mathbb{Z}}} = \int_{b_{ij}} \mathrm{d}C_{\mathbb{Z}}(\mathbf{v}), \tag{5.4}$$

i.e. obviously cell probabilities can be expressed by the copula, and may also be approximated by a sum of Binomial variables. The null implies equality of all corresponding cell probabilities. With marginal null hypotheses,

$$H_{0,ij}^{\rightarrow} : p_{ij}^{\mathbb{X}} = p_{ij}^{\mathbb{Y}},$$

where $i, j = 1, ..., J$, and

$$H^{\leftarrow}_{0,ij} : p^{\mathbb{X}}_{ij} = p^{\mathbb{Y}}_{ji},$$

where $i, j = 1, ..., J$ and $i \neq j$, or, equivalently,

$$\begin{aligned} &H_{0,1} = H^{\rightarrow}_{0,11}, && H_{0,2} = H^{\rightarrow}_{0,21}, && ..., && H_{0,J^2} = H^{\rightarrow}_{0,JJ}, \\ &H_{0,J^2+1} = H^{\leftarrow}_{0,21}, && H_{0,J^2+2} = H^{\leftarrow}_{0,31}, && ..., && H_{0,2J^2+J} = H^{\leftarrow}_{0,J(J-1)}, \end{aligned}$$

the global hypothesis (Equation (5.3)) can be written as

$$H_0 : \bigcap_{i=1}^{2J^2+J} H_{0,i}. \tag{5.5}$$

This covers all *reasonable* comparisons between cells, thus capturing possible non–exchangeabilities. However, probabilities of diagonal cells $p^{\widetilde{\mathbb{X}}}_{ii}, p^{\widetilde{\mathbb{Y}}}_{ii}$ are compared only once.[1] The null is violated if at least one marginal cell comparison has to be rejected. The empirical Wald type test statistic for the marginal hypothesis $H^{\rightarrow}_{0,ij}$ is

$$\widehat{Q}^{\rightarrow}_{ij} = \frac{\left(\widehat{p}^{\widetilde{\mathbb{X}}}_{ij} - \widehat{p}^{\widetilde{\mathbb{Y}}}_{ij}\right)^2}{((\widehat{p}^{\widetilde{\mathbb{X}}}_{ij}(1 - \widehat{p}^{\widetilde{\mathbb{X}}}_{ij}))/n_{\mathbb{X}} + (\widehat{p}^{\widetilde{\mathbb{Y}}}_{ij}(1 - \widehat{p}^{\widetilde{\mathbb{Y}}}_{ij}))/n_{\mathbb{Y}})},$$

where cell probabilities are estimated by

$$\widehat{p}^{\widetilde{\mathbb{Z}}}_{ij} = \frac{1}{n_{\mathbb{Z}}} \sum_{i=1}^{n_{\mathbb{Z}}} 1\{\widetilde{\mathbb{Z}}_i \in b_{ij}\}.$$

Test statistics for flipped versions of the marginal hypotheses $\widehat{Q}^{\leftarrow}_{ij}$ are defined accordingly. The asymptotic test distribution immediately follows.

Proposition 5.1. *Under the null hypothesis and under Assumptions ($A1^C$)–($A4^C$),*

$$\widehat{Q}^{\rightarrow}_{ij} \xrightarrow{d} F_{1,\min(n_{\mathbb{X}}, n_{\mathbb{Y}})-5}, n_{\mathbb{X}}, n_{\mathbb{Y}} \to \infty.$$

By ($A1^C$), $\widehat{p}^{\widetilde{\mathbb{Z}}}_{ij}$ is consistent, and, due to ($A4^C$), also asymptotically normally distributed, see Herrndorf (1984). Assumption ($A3^C$) allows to estimate $\mathbb{V}(\widehat{p}^{\mathbb{Z}}_{ij} - p^{\mathbb{Z}}_{ij}) =: \sigma^2_{\widehat{p}^{\mathbb{Z}}_{ij}}$ with the standard empirical variance estimator, without addressing any covariances.

[1] Otherwise, the multiplicity penalty would penalize comparisons along the diagonal, i.e. $p^{\mathbb{X}}_{ii} = p^{\mathbb{Y}}_{ii}, i = 1, ..., J$, twice.

However, if Assumption ($A3^C$) is violated, covariances must be estimated, e.g. by bootstrap.

The degrees of freedom of the basic t–statistic have to be adjusted as five additional nuisance parameters occur; four pre–estimated marginal distributions (see Dobrić & Schmid (2005)), and the estimated variance. This adjustment directly enters the asymptotic distribution of $\widehat{Q}_{ij}^{\rightarrow}$ and appears also in the second term on the right hand side of Equation (5.7), the asymptotic limit of the empirical copula process, which only occurs when marginal distributions are unknown.

Clearly, under the null of copula equality, test statistics should be close to zero. Violations of the null amount to at least one of the Wald statistic $\widehat{Q}_{ij}^{\rightarrow}$ being too large, i.e. $\widehat{Q}_{ij}^{\rightarrow} > F^{1-\alpha}_{1,\min(n_{\mathbb{X}},n_{\mathbb{Y}})-5}$. This is equivalent to a one–sided test decision with $\alpha^* = \alpha/2$. Consequently, the test directly reveals if one pair is significantly more dependent within a specific cell b_{ij}.

Standard χ^2–tests compare the sum of squared differences between empirical probabilities with quantiles of the χ^2 distribution (Hu (2006), Jondeau and Rockinger (2006), Hong et al (2007), Patton (2013)). However, this does not allow to pin down boxes where copulas significantly differ. By reformulating the copula comparison as a multiple testing problem as in Equation (5.5), we see which marginal hypothesis is rejected. This directly links copula inequality to specific cells in the copula domain $[0,1]^2$, or, equivalently, links copula inequality to quantile events. Note again, we compare all (reasonable) cell combinations to account for possible non–exchangeability of the copulas which standard χ^2 box tests ignore. As we jointly test $M := 2J^2 - J$ marginal hypotheses, we have to reduce the marginal significance level to achieve overall α–control. Denote the individual p–value for test statistic $H_{0,ij}^{\rightarrow}$ by $P_{ij}^{\rightarrow}$. The simplest method accounting for the multiplicity of this testing problem is the Bonferroni adjustment, which multiplies p–values with the number of marginal hypotheses, i.e $\widetilde{P}_{ij}^{\rightarrow} = MP_{ij}^{\rightarrow}$.

This is the baseline procedure of the test. However, we find aggregating many baseline tests to one single test decision substantially improves power while α–control is still held when p–values across grids are sufficiently dependent and grid–specific baseline tests tend to be undersized. This aggregation step has first been proposed in Bormann & Schienle (2016) for tail copula comparisons. As is often criticized, the choice of the number of cells for χ^2–tests is arbitrary. Naturally, more cells imply fewer observations in each cell inducing lower power. In contrast, fewer cells imply that non–standard dependence structures might be overlooked, also inducing lower power. We propose to apply the test over many different grids and aggregate grid–specific p–values by simply picking the smallest adjusted p–value over all grids. This not only robustifies the test, but also increases test power.

Aggregating p–values in this manner comes along with the danger of the test size being too large. However, the test appears to be undersized for each single grid individually due to multiplicity penalty, and the fact that with smaller cells, less observations fall into cells and realized convergence rates decelerate. Estimation error rises, reducing the tendency to reject. Moreover, boxes overlap and exhibit strong to nearly perfect dependence. All these points justify refraining from further multiplicity penalties. Simulations indicate, irrespective of sample size or underlying copula model, that (i) the test obeys nominal α–control, and (ii) test results are very similar when aggregating over a number of baseline tests with different grids.

Notably, the test is very easy to implement, and has no computational costs in contrast to simulation–based tests. Also, by comparing cell probabilities individually, we can pin down cells, i.e. quantile events, where dependence differs. Furthermore, we can directly conclude which random vector is more dependent in a specific cell. In contrast to other cell–based tests, we explicitly account for possible non–exchangeability of both copulas.

5.4 Simulation study

This section investigates the finite sample properties of the test. In particular, we compare two versions of our aggregation test with the standard two–sample χ^2–test. The first version of our test aggregates baseline tests with $J^2 = 2^2, ..., 10^2$. We denote this test by M_{10}. The second version aggregates baseline tests with $J^2 = 2^2, ..., 15^2$, which we denote as M_{15}. Note, a baseline test with $J^2 = 4^2$ consists of $2J^2 - J = 28$ individual cell comparisons. We also compute test levels for the standard two–sample χ^2–test with 81 cells, denoted by χ^2_9. However, we do not include the Cramér–von Mises test by Rémillard & Scaillet (2009) due to its high computational burden even for moderately large samples. For example, for $n > 250$ (parallelized) computation takes minutes on a standard laptop while our test, even unparallelized, instantaneously provides a test result even for large sample sizes ($n > 5000$). Yet, we expect that test to be more powerful than the tests presented here.

We will find that, in a copula–GARCH setting, our test is consistent, typically holds the nominal α–level, and is more powerful than the competing standard χ^2–test. Moreover, the latter is inconsistent in case of copula non–exchangeability, while our text exhibits excellent power properties for non–exchangeable copulas.

Specifically, we compare rejection probabilities when testing equality of cross–sectional dependence structures between two bivariate, serially dependent processes, $\mathbb{X}$ and $\mathbb{Y}$. As GARCH–processes are successfully used in modeling financial returns, we employ GARCH(1,1) processes for univariate dynamics. Cross–sectional dependence between GARCH–processes are governed by imposing pre–specified copula models for GARCH innovations. This approach is in line with Oh & Patton (2013). Formally, for both bivariate return processes $Z = (Z^{(1)}, Z^{(2)}), \mathbb{Z} = \mathbb{X}, \mathbb{Y}$, such copula–GARCH models can be written as,

$$
\begin{aligned}
Z_t^{(i)} &= \sigma_t^{(i)} \eta_t^{(i)}, \\
\sigma_t^{2,(i)} &= \omega + \alpha Z_{t-1}^{2,(i)} + \beta \sigma_{t-1}^{2,(i)}, \\
\eta := (\eta^{(1)}, \eta^{(2)}) &\sim iid\ F_{\eta,\mathbb{Z}}(x^{(1)}, x^{(2)}) = C_{\eta,\mathbb{Z}}(F_{\eta,1}(\eta^{(1)}), F_{\eta,2}(\eta^{(2)})), t = 1, ..., n_{\mathbb{Z}},
\end{aligned}
$$

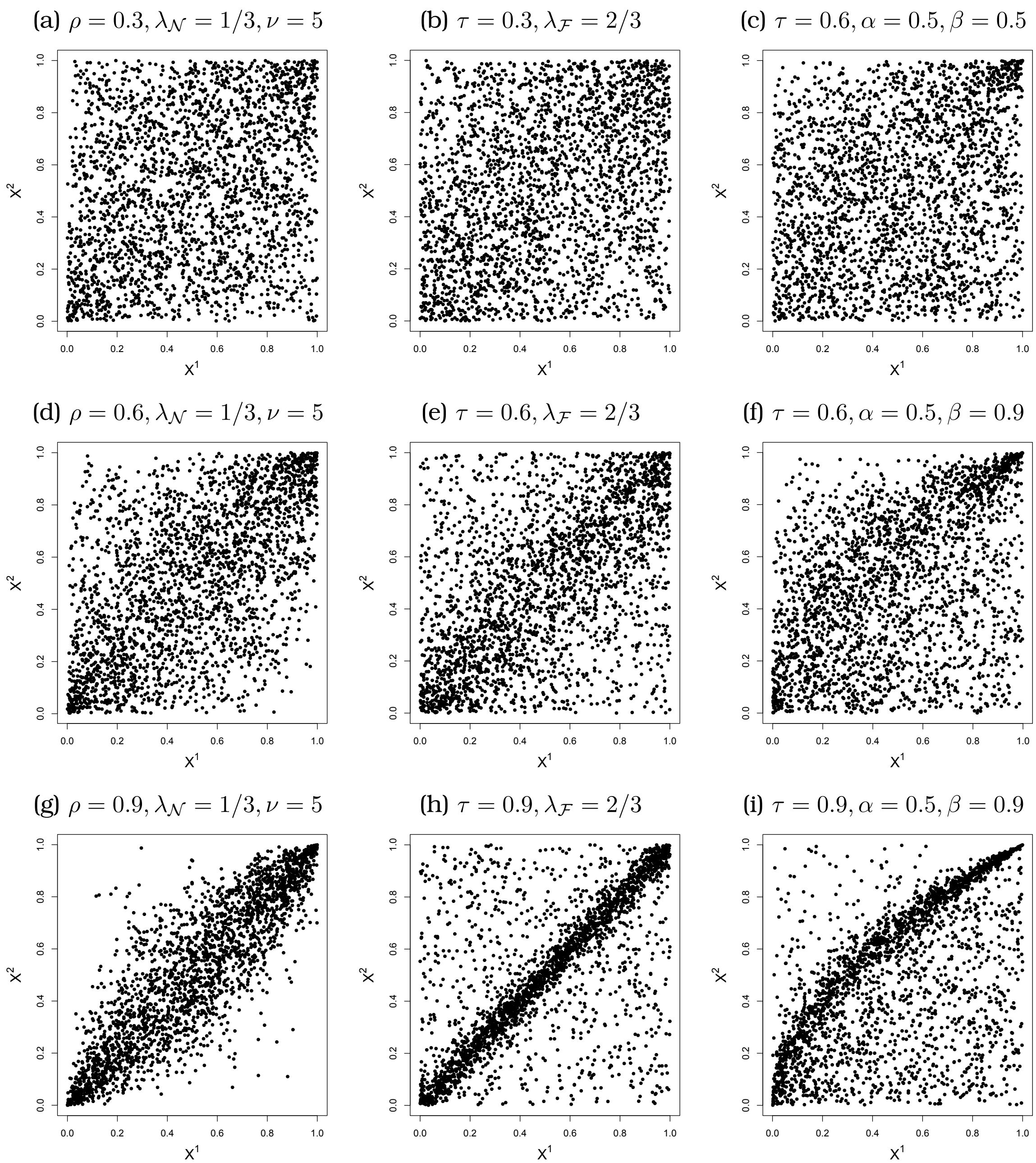
where we set $\omega = 0.01, \alpha = 0.15$ and $\beta = 0.8$ (see Engle & Sheppard (2001) for typical empirical values), and $C_{\eta,\mathbb{Z}}$ denotes the error term copula for process $\{Z_t^{(1)}, Z_t^{(2)}\}_{t=1}^{n_{\mathbb{Z}}}$, $F_{\eta,\mathbb{Z}}$ is the joint error term distribution function for process $\mathbb{Z}$, and $F_{\mathbb{Z},i}, i = 1, 2$, are marginal distribution functions of $Z^{(1)}$ and $Z^{(2)}$. Note that spillover effects are solely governed by the error term copula. Errors are marginally t–distributed five degrees of freedom to account for fat tails.[2]

In the following, we estimate test rejection probabilities when both processes, $\mathbb{X}$ and $\mathbb{Y}$, follow copula–GARCH processes with (i) the same error term copula, and with (ii) differently parametrized error term copulas, i.e. parametrizations of $C_{\eta,\mathbb{X}}$ and $C_{\eta,\mathbb{Y}}$ vary. The former approximates test size, the latter approximates test power. One draw of simulated data of all copula families considered are shown in Figure (5.1). We choose sample sizes as $n \in \{250, 500, 1000, 5000, 10000\}$. Simulations are repeated 1000 times.

The first choice for $C_{\eta,\mathbb{Z}}$ (DGP1) is a mixture of a Gauss and a t–copula. Both copulas are popular in financial practice for their simplicity and easy implementation. Further, the t–copula also captures tail dependence, and a mixture of both extends

[2] Marginal error distributions are standardized, i.e. $\mathbb{V}(\eta^{(i)}) = 1, i = 1, 2$.

Figure 5.1: Scatterplots of simulated samples from DGPs 1 to 3. Left column: DGP1, mixtures of Gauss copulas and t–copulas. Mid column: DGP2, mixtures of Frank copulas and the independence copula. Right column: DGP3, the non–exchangeable copula by Khoudraji (1995); note how the degree of non–exchangeability increases from top to bottom.



the linear dependence modeling of the Gauss distribution by accounting for possible joint tail events. Denote the Gaussian copula by $C_{\mathcal{N}}(\mathbf{x};\rho_{\mathcal{N}})$ and the t–copula by $C_{\mathcal{T}}(\mathbf{x};\rho_{\mathcal{T}},\nu)$ with ν degrees of freedom. For the Gauss copula, ρ denotes the correlation parameter, and similarly, it denotes the dispersion parameter in case of the t–copula. The mixture copula is given by

$$C_{\mathcal{N},\mathcal{T}}(\mathbf{x};\nu,\lambda_{\mathcal{N}},\rho_{\mathcal{N}},\rho_{\mathcal{T}}) := \lambda_{\mathcal{N}}\ C_{\mathcal{N}}(\mathbf{x};\rho_{\mathcal{N}}) + (1-\lambda_{\mathcal{N}})\ C_{\mathcal{T}}(\mathbf{x};\rho_{\mathcal{T}},\nu), \lambda_{\mathcal{N}} \in [0,1],$$

where $\lambda_{\mathcal{N}}$ measures the share the Gauss copula contributes to the entire copula, i.e. how strongly the Gauss copula enters the mixture copula. If $\lambda_{\mathcal{N}} = 1$, $C_{\mathcal{N},\mathcal{T}}$ is a pure Gauss copula. If $\lambda_{\mathcal{N}} < 1$, the model allows for tail dependence due to the influence of the t–copula $(1-\lambda_{\mathcal{N}})$.

We model the dependence between the innovations of $\mathbb{X}$ and $\mathbb{Y}$, respectively, by $C_{\mathcal{N},\mathcal{T}}$, and vary the mixture and copula parameters to achieve situations in which the null is either fulfilled or violated. To keep the analysis manageable, we set $\rho_{\mathcal{N}} = \rho_{\mathcal{T}}, \nu^{\mathbb{X}} = \nu^{\mathbb{Y}}$, i.e. within each random vector, correlation and dispersion of the Gauss and the t–copula are always identical. Furthermore, the impact of varying tail dependence, correlation and dispersion are investigated by setting $\nu^{\mathbb{X}} = \nu^{\mathbb{Y}} \in \{5, 20\}, \lambda_{\mathcal{N}} \in \{1/3, 2/3, 1\}, \rho^{\mathbb{X}}, \rho^{\mathbb{Y}} \in \{0.3, 0.6, 0.9\}$. The null is true whenever $\rho^{\mathbb{X}} = \rho^{\mathbb{Y}}$ and $\nu^{\mathbb{X}} = \nu^{\mathbb{Y}}$.

Table (5.1) contains empirical size and power of our tests M_{10} and M_{15} and the fixed cells χ^2–test. As a comparison with the serially dependent case, Table (5.5) contains complementary results with i.i.d. marginals (t–distributed, five degrees of freedom).

Our test features attractive power results for sample sizes larger than 500. For $n \geq 5000$, false nulls are rejected almost surely. However, for $n \leq 500$, the test struggles to detect copula inequalities if $|\rho^{\mathbb{X}} - \rho^{\mathbb{Y}}| = 0.3$, i.e. when the discrepancy between cross–sectional dependencies is moderate. Most empirical test sizes are close to the nominal level of $\alpha = 5\%$. Empirical test sizes of both of our tests are close to each other, while for M_{15}, empirical power is slightly larger. Nevertheless, we also observe empirical rejection probabilities that exactly coincide for M_{10} and M_{15}. Both tests hence produce similar results, while M_{15} is slightly more appealing. For large sample sizes, empirical size appears to be slightly too large. This might result from the fact, that for large sample sizes, estimated GARCH coefficients may indicate almost non–stationarity and persistent serial dependence ($\widehat{\alpha} + \widehat{\beta} \approx 1$), see Mikosch & Stărică (2004). This might explain test size distortion for very large sample sizes. Figure

(5.8) visualizes this phenomenon in our setting. In the i.i.d. case, this phenomenon is less pronounced, see Table (5.5). Consequently, data anomalies may explain why empirical test levels are too large for large sample sizes.

In comparison, the χ^2–test also obeys the nominal test size, but is undersized irrespective of sample size and copula parametrization. Empirical power is typically lower. For some settings, the χ^2–test is not able to produce reliable results; as a rule of thumb, the χ^2–test demands at least five observations in each cell to work properly. In all other cases, *NA* indicates the test fails in all of the 1000 simulation runs.

For DGP2, we model $C_{\mathbb{X}}$ and $C_{\mathbb{Y}}$ as a mixture of the Frank and the independence copula. The Frank copula, denoted by $C_{\mathcal{F}}(\mathbf{x};\theta)$, is given by

$$C_{\mathcal{F}}(\mathbf{x};\theta) = -\frac{1}{\theta}\log\left(1+\frac{(\exp(-\theta x^{(1)})-1)(\exp(-\theta x^{(2)})-1)}{\exp(-\theta)-1}\right), \theta \in \mathbb{R},$$

where we choose θ such that Kendalls tau[3] $\tau \in \{0.3, 0.6, 0.9\}$, i.e. $\theta \in \{2.9174,\ 7.9296,\ 38.29121\}$. The independence copula is given by

$$C_{\mathcal{I}}(\mathbf{x}) = x^{(1)}x^{(2)}.$$

The mixture copula for DGP2 follows as

$$C_{\mathcal{F},\mathcal{I}}(\mathbf{x};\theta,\lambda_{\mathcal{F}}) = \lambda_{\mathcal{F}} C_{\mathcal{F}}(\mathbf{x};\theta) + (1-\lambda_{\mathcal{F}})C_{\mathcal{I}}(\mathbf{x}),$$

where $\lambda_{\mathcal{F}} \in \{0, 1/3, 2/3\}$. The mixture parameter reflects the share of the Frank copula to the mixture copula. To keep results manageable, we fix $\lambda_{\mathcal{F}}^{\mathbb{X}} = \lambda_{\mathcal{F}}^{\mathbb{Y}}$. The null holds whenever $\theta^{\mathbb{X}} = \theta^{\mathbb{Y}}$. Table (5.2) shows results of the copula–GARCH model.

Results are similar to DGP1. However, for large sample sizes, our test is oversized whenever the independence copula is involved, and the Frank copula exhibits strong correlation. In theory, the aggregating step should render the test oversized. However, if only few or even no data fall in some cells, oversizedness is absorbed as individual cell p–values are undersized. With the independence copula, data are evenly distributed over $[0,1]^2$, and the absorbing effect of (some) undersized cell p–values vanishes. The test should hence be only used when at least some dependence

[3]Kendall's tau is another dependence measure, which is completely determined by the copula function as $\tau^{\mathbb{X}} = 4\int_{[0,1]^2} C_{\mathbb{X}}(\mathbf{u})\mathrm{d}C_{\mathbb{X}}(\mathbf{x}) - 1$, see McNeil et al. (2005).

in both return vectors is on hand.

Finally, we use asymmetrized mixtures of Gumbel copulas to show χ^2_9 is not consistent for non–exchangeable copulas, while our test is extremely powerful in such cases. Non–exchangeable copulas can be constructed following Khoudraji (1995), who combines two exchangeable copulas C_1 and C_2 by

$$C_{\mathcal{K}}(\mathbf{x}) = C_1\left({x^{(1)}}^{1-a}, {x^{(2)}}^{1-b}\right) C_2\left({x^{(1)}}^{a}, {x^{(2)}}^{b}\right),$$

with asymmetry parameters $a, b \in [0,1]$. If $a \neq b$, $C_{\mathcal{K}}$ is non–exchangeable, see Figure (5.1), right panel, for simulated data. We choose C_1, C_2 to be Gumbel copulas with identical parametrization for both C_1 and C_2. The Gumbel copula is given by

$$C_{\mathcal{G}}(\mathbf{x};\theta) = \exp\left(-((-\log x^{(1)})^{\theta} + (-\log x^{(2)})^{\theta})^{1/\theta}\right), 1 \leq \theta < \infty. \tag{5.6}$$

For the error term copulas of $\mathbb{X}$ and $\mathbb{Y}$, set $C_{\eta,\mathbb{X}}(\mathbf{x}) = C_{\mathcal{K}}(\mathbf{x};\theta^{\mathbb{X}}, a^{\mathbb{X}}, b^{\mathbb{X}}), C_{\eta,\mathbb{Y}}(\mathbf{x}) = C_{\mathcal{K}}(\mathbf{x};\theta^{\mathbb{Y}}, a^{\mathbb{Y}}, b^{\mathbb{Y}})$, and choose $\theta^{\mathbb{X}}, \theta^{\mathbb{Y}}$ such that $\tau \in \{0.3, \quad 0.6, \quad 0.9\}$, i.e. $\theta^{\mathbb{X}}, \theta^{\mathbb{Y}} \in \{1.428571, 2.5, 10\}$. Concerning the asymmetry parameters, we set $a^{\mathbb{X}} \in \{0.5, \quad 0.7, 0.9\}$, $a^{\mathbb{Y}} = 0.5$, $b^{\mathbb{X}} = b^{\mathbb{Y}} = 0.9$, i.e. at least one copula is always non–exchangeable, and the tests ideally always reject the null. Note that $a^{\mathbb{Y}}$ is fixed, while $a^{\mathbb{X}}$ varies; for the case of $a^{\mathbb{X}} = a^{\mathbb{Y}} = 0.5$ the null is only seemingly true as parameters for both copulas are equal. However, due to the non–exchangeability, the null is not true. Table (5.3) contains test results for the copula–GARCH model.

Our test exhibits ideal power properties, also for moderate sample sizes. In contrast, when null rejection is only seemingly true, the χ^2–test is not able to reject the false null, rendering it inconsistent in these cases. However, for our test, care must be taken if at least one copula under consideration is similar to the independence copula.

Table 5.1: Empirical rejection frequencies for mixtures of the Gauss and the t–copula, where marginals follow GARCH(1,1) processes.

		M_{10}	M_{15}	χ^2_9	M_{10}	M_{15}	χ^2_9	M_{10}	M_{15}	χ^2_9		M_{10}	M_{15}	χ^2_9	M_{10}	M_{15}	χ^2_9	M_{10}	M_{15}	χ^2_9
n	$(\lambda_{\mathcal{N}}, \rho_{\mathbb{X}})$	$\rho^{\mathbb{Y}} = 0.3$			$\rho^{\mathbb{Y}} = 0.6$			$\rho^{\mathbb{Y}} = 0.9$			$(\lambda_{\mathcal{N}}, \rho_{\mathbb{X}})$	$\rho^{\mathbb{Y}} = 0.3$			$\rho^{\mathbb{Y}} = 0.6$			$\rho^{\mathbb{Y}} = 0.9$		
	$(1/3, 0.3)$										$(1/3, 0.6)$									
250		0.5	0.5	1.0	17.9	17.9	8.2	100	100	100		15.1	15.1	7.8	0.3	0.3	0.0	69.1	69.1	NA
500		0.3	0.3	0.6	63.0	63.0	43.3	100	100	100		62.6	62.6	41.8	0.2	0.2	0.7	100	100	100
1000		0.7	0.7	0.6	98.9	98.9	94.4	100	100	100		98.1	98.1	93.9	1.5	1.5	0.3	100	100	100
5000		2.9	4.1	1.7	100	100	100	100	100	100		100	100	100	3.5	5.7	1.4	100	100	100
10000		4.1	8.1	1.1	100	100	100	100	100	100		100	100	100	4.8	7.4	0.5	100	100	100
	$(1/3, 0.9)$										$(2/3, 0.3)$									
250		100	100	100	67.8	67.8	NA	0.0	0.0	NA		0.0	0.0	0.3	17.4	17.4	10.1	99.9	99.9	100
500		100	100	100	99.8	99.8	100	0.1	0.1	NA		0.8	0.8	0.5	64.6	64.6	42.1	100	100	100
1000		100	100	100	100	100	100	0.4	0.4	NA		1.3	1.4	0.6	98.3	98.3	99.9	100	100	100
5000		100	100	100	100	100	100	4.3	5.5	0.0		3.0	4.2	0.5	100	100	100	100	100	100
10000		100	100	100	100	100	100	5.5	7.2	1.4		3.9	7.1	0.8	100	100	100	100	100	100
	$(2/3, 0.6)$										$(2/3, 0.9)$									
250		16.5	16.5	7.7	0.4	0.4	0.0	67.7	67.7	NA		100	100	100	67.0	67.0	NA	0.0	0.0	NA
500		64.6	64.6	39.5	0.6	0.6	0.6	99.8	99.8	100		100	100	100	100	100	100	0.1	0.1	NA
1000		98.4	98.4	91.1	1.2	1.2	0.7	100	100	100		100	100	100	100	100	100	0.5	0.5	NA
5000		100	100	100	3.8	5.8	1.5	100	100	100		100	100	100	100	100	100	3.0	3.6	3.0
10000		100	100	100	5.1	8.4	1.4	100	100	100		100	100	100	100	100	100	4.5	6.4	1.2
	$(1, 0.3)$										$(1, 0.6)$									
250		0.1	0.1	0.6	16.9	16.9	10.3	100	100	100		13.9	13.9	8.8	0.0	0.0	7.1	66.4	66.4	NA
500		0.5	0.5	0.9	57.1	57.1	32.7	100	100	100		61.1	61.2	35.3	0.8	0.8	0.3	99.6	99.6	100
1000		1.3	1.3	0.3	97.9	97.9	89.1	100	100	100		97.5	97.5	89.6	0.8	0.8	0.4	100	100	100
5000		3.6	6.5	0.8	100	100	100	100	100	100		100	100	100	3.6	6.0	0.9	100	100	100
10000		3.5	6.5	0.7	100	100	100	100	100	100		100	100	100	4.9	7.4	1.0	100	100	100
	$(1, 0.9)$																			
250		99.9	99.9	100	65.5	65.5	NA	0.0	0.0	NA										
500		100	100	100	99.6	99.6	100	0.0	0.1	NA										
1000		100	100	100	100	100	100	0.8	0.8	NA										
5000		100	100	100	100	100	100	5.4	6.0	0.7										
10000		100	100	100	100	100	100	5.3	6.8	0.9										

Table 5.2: Empirical rejection frequencies for mixtures of the Frank and the independent copula, where marginals follow GARCH(1,1) processes.

n	$(\lambda_{\mathcal{F}}, \tau^{\mathbb{X}})$	M_{10}	M_{15}	χ^2_9	M_{10}	M_{15}	χ^2_9	M_{10}	M_{15}	χ^2_9
		$\tau_{\mathbb{Y}} = 0.3$			$\tau^{\mathbb{Y}} = 0.6$			$\tau^{\mathbb{Y}} = 0.9$		
	$(0, 0.3)$									
250		0.2	0.2	2.1	74.4	74.4	83.3	100	100	100
500		0.8	0.8	0.8	99.9	99.9	98.8	100	100	100
1000		1.8	1.8	0.9	100	100	100	100	100	100
5000		2.9	5.6	0.7	100	100	100	100	100	100
10000		4.1	7.2	0.7	100	100	100	100	100	100
	$(0, 0.9)$									
250		100	100	**NA**	96.3	96.3	**NA**	0.0	0.0	**NA**
500		100	100	100	100	100	**NA**	0.0	0.0	**NA**
1000		100	100	100	100	100	**NA**	0.0	0.0	**NA**
5000		100	100	100	100	100	100	0.4	1.2	**NA**
10000		100	100	100	100	100	100	0.7	2.7	**NA**
	$(1/3, 0.6)$									
250		10.2	10.2	10.0	0.4	0.4	0.0	24.1	24.1	71.4
500		42.8	42.9	41.8	0.6	0.6	0.7	89.4	89.4	99.5
1000		94.6	94.6	91.6	1.4	1.4	1.7	100	100	100
5000		100	100	100	4.9	6.5	1.1	100	100	100
10000		100	100	100	5.9	9.6	1.1	100	100	100
	$(2/3, 0.3)$									
250		0.1	0.1	0.8	0.7	0.7	1.7	8.3	8.3	23.3
500		1.0	1.0	1.1	3.1	3.1	4.3	51.2	51.2	68.3
1000		1.0	1.0	0.5	15.6	15.6	12.2	98.1	98.2	99.4
5000		3.0	6.1	1.1	99.0	99.1	97.9	100	100	100
10000		3.5	6.8	0.6	100	100	100	100	100	100
	$(2/3, 0.9)$									
250		8.0	8.0	22.1	1.2	1.2	5.8	0.4	0.4	2.3
500		52.1	52.3	70.2	10.4	10.5	24.9	1.6	1.6	3.3
1000		99.0	99.1	99.6	49.9	51.0	68.4	3.5	4.1	3.7
5000		100	100	100	100	100	100	6.8	11.3	3.7
10000		100	100	100	100	100	100	6.9	12.8	4.2

n	$(\lambda_{\mathcal{F}}, \rho^{\mathbb{X}})$	M_{10}	M_{15}	χ^2_9	M_{10}	M_{15}	χ^2_9	M_{10}	M_{15}	χ^2_9
		$\tau^{\mathbb{Y}} = 0.3$			$\tau^{\mathbb{Y}} = 0.6$			$\tau^{\mathbb{Y}} = 0.9$		
	$(0, 0.6)$									
250		73.7	73.7	57.1	0.0	0.0	**NA**	96.7	96.7	**NA**
500		99.8	99.8	99.3	0.1	0.1	**NA**	100	100	**NA**
1000		100	100	100	1.6	1.6	0.0	100	100	**NA**
5000		100	100	100	4.1	5.1	0.5	100	100	100
10000		100	100	100	4.2	6.5	1.1	100	100	100
	$(1/3, 0.3)$									
250		0.2	0.2	0.1	8.3	8.3	10.7	94.2	94.2	99.6
500		0.4	0.4	1.0	45.6	45.6	40.5	100	100	100
1000		1.0	1.1	0.7	95.5	95.5	93.4	100	100	100
5000		3.3	5.2	0.7	100	100	100	100	100	100
10000		2.9	6.7	0.6	100	100	100	100	100	100
	$(1/3, 0.9)$									
250		94.3	94.3	98.8	25.0	25.0	60.0	1.0	1.0	0.0
500		100	100	100	85.8	86.4	99.3	3.2	3.3	5.8
1000		100	100	100	100	100	100	6.9	7.2	8.3
5000		100	100	100	100	100	100	12.1	15.2	6.0
10000		100	100	100	100	100	100	14.6	19.2	8.2
	$(2/3, 0.6)$									
250		1.3	1.3	2.9	0.4	0.4	0.6	1.0	1.0	5.4
500		2.5	2.5	2.5	0.6	0.6	0.9	8.6	8.7	22.0
1000		12.6	12.9	12.2	1.3	1.5	0.3	50.8	52.2	71.8
5000		99.0	99.0	98.3	4.5	7.1	0.7	100	100	100
10000		100	100	100	3.1	6.3	0.8	100	100	100

Table 5.3: Empirical rejection frequencies for mixtures of the asymmetrized Gumbel copula with asymmetry parameters $a^{\mathbb{X}} = a^{\mathbb{Y}} = 0.5, b^{\mathbb{X}} = b^{\mathbb{Y}} = 0.9$. Note, in all cases, the null is false — pseudo test size is bold. Marginals follow GARCH(1,1) processes.

		M_{10}	M_{15}	χ^2_9	M_{10}	M_{15}	χ^2_9	M_{10}	M_{15}	χ^2_9		M_{10}	M_{15}	χ^2_9	M_{10}	M_{15}	χ^2_9	M_{10}	M_{15}	χ^2_9
n	$(a^{\mathbb{X}}, \tau_{\mathbb{X}})$	$\tau_{\mathbb{Y}} = 0.3$			$\tau_{\mathbb{Y}} = 0.6$			$\tau_{\mathbb{Y}} = 0.9$			$(a^{\mathbb{X}}, \tau_{\mathbb{X}})$	$\tau_{\mathbb{Y}} = 0.3$			$\tau_{\mathbb{Y}} = 0.6$			$\tau_{\mathbb{Y}} = 0.9$		
	$(0.5, 0.3)$										$(0.5, 0.6)$									
250		**0.2**	**0.2**	**0.8**	13.0	13.0	5.2	82.7	82.7	79.2		13.4	13.4	5.7	**6.4**	**6.4**	**1.7**	57.5	57.5	27.3
500		**1.4**	**1.4**	**0.7**	57.6	57.6	22.9	100	100	99.9		55.8	55.8	21.4	**35.5**	**35.5**	**1.1**	99.2	99.2	69.4
1000		**3.0**	**3.1**	**1.0**	97.0	97.1	74.5	100	100	100		96.5	96.5	73.1	**92.5**	**92.5**	**0.9**	100	100	98.8
5000		**50.6**	**54.0**	**0.9**	100	100	100	100	100	100		100	100	100	**100**	**100**	**0.4**	100	100	100
10000		**94.7**	**95.1**	**0.9**	100	100	100	100	100	100		100	100	100	**100**	**100**	**0.4**	100	100	100
	$(0.5, 0.9)$										$(0.7, 0.3)$									
250		82.5	82.5	81.2	56.8	51.2	16.7	**92.5**	**92.5**	**NA**		0.3	0.3	1.4	34.7	86.7	23.9	98.3	98.3	100
500		100	100	99.7	99.4	85.8	65.2	**100**	**100**	**1.8**		1.5	1.5	1.6	89.6	99.7	77.7	100	100	100
1000		100	100	100	100	99.7	99.0	**100**	**100**	**2.2**		4.9	4.9	1.8	100	100	99.9	100	100	100
5000		100	100	100	100	100	100	**100**	**100**	**4.9**		67.0	68.7	10.5	100	100	100	100	100	100
10000		100	100	100	100	100	100	**100**	**100**	**4.9**		97.8	98.1	37.7	100	100	100	100	100	100
	$(0.7, 0.6)$										$(0.7, 0.9)$									
250		4.8	4.8	4.7	6.7	21.0	0.0	78.8	78.8	100		70.1	70.1	73.5	25.9	49.2	100	95.3	95.3	NA
500		30.2	30.2	12.9	37.2	42.6	6.7	99.9	99.9	100		99.8	99.8	99.5	91.1	83.9	69.8	100	100	94.7
1000		83.5	83.7	51.0	92.2	76.5	23.8	100	100	100		100	100	100	100	99.0	98.7	100	100	100
5000		100	100	100	100	100	99.8	100	100	100		100	100	100	100	100	100	100	100	100
10000		100	100	100	100	100	100	100	100	100		100	100	100	100	100	100	100	100	100
	$(0.9, 0.3)$										$(0.9, 0.6)$									
250		1.1	1.1	1.2	71.0	99.2	63.6	100	100	100		3.2	3.2	2.9	18.1	68.4	11.8	99.7	99.7	100
500		3.3	3.3	2.2	99.3	100	99.5	100	100	100		13.3	13.4	10.8	69.7	96.7	59.5	100	100	100
1000		13.5	13.7	8.0	100	100	100	100	100	100		57.0	57.3	45.1	99.5	100	97.6	100	100	100
5000		97.5	97.7	80.1	100	100	100	100	100	100		100	100	100	100	100	100	100	100	100
10000		100	100	100	100	100	100	100	100	100		100	100	100	100	100	100	100	100	100
	$(0.9, 0.9)$																			
250		58.6	58.6	75.2	23.9	81.6	NA	99.3	99.3	NA										
500		99.2	99.2	99.8	86.9	99.1	97.5	100	100	100										
1000		100	100	100	100	100	100	100	100	100										
5000		100	100	100	100	100	100	100	100	100										
10000		100	100	100	100	100	100	100	100	100										

5.5 Copula dynamics in high-frequency returns

5.5.1 Data description

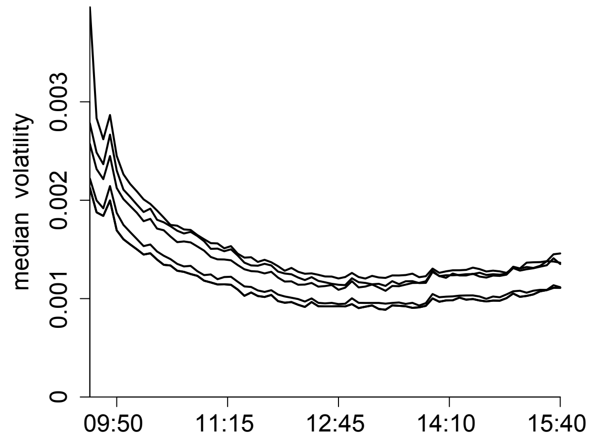
We analyze high-frequency returns of five financial stocks with respect to intertemporal copula changes during 06/27/2007 – 10/13/2015, i.e. JP Morgan (JPM), Bank of America (BAC), Goldman Sachs (GS), Wells Fargo & Co (WFC), Morgan Stanley (MS). Returns are computed from NASDAQ limit order book data obtained from `https://lobsterdata.com/`. NASDAQ trading hours are from 9:30 to 16:00. We are interested in (i) whether copulas change at all, (ii) when copulas change, and also (iii) *where* copulas change, i.e. what parts of the bivariate return distribution are time–varying.

We build all ten bivariate pairs, and apply the test to compare each pair's copula structure in two disjoint, successive time periods, i.e. we test $H_0 : C_{T_1}^{(i,j)} = C_{T_2}^{(i,j)}$, where $C_{T_1}^{(i,j)}$ denotes the copula of the return vector $(r_t^{(i)}, r_t^{(j)})_{i=1}^{T_1}$, and $C_{T_2}^{(i,j)}$ denotes the copula of $(r_t^{(i)}, r_t^{(j)})_{i=T_1+1}^{T_2}$. We repeat testing in a rolling window scheme, and also consider different sample frequencies. To be precise, for 60 second returns we compare two trading weeks of observations ($n \leq 1800$); for 150 second returns we compare six trading weeks ($n \leq 2160$); for 300 second returns we compare ten trading weeks ($n \leq 1800$) for 300 second returns we compare 24 trading weeks ($n \leq 2160$). Our test allows for a computationally rapid analysis, whereas for simulation–based tests, this test setup would be an immense computational burden.

High–frequency data should be cleaned with respect to data errors, which might arise due to erroneous recording, see Hautsch (2012), Chapter 3. We proceed by computing mid–quotes by averaging first level bid and ask prices; only executed trades are employed. Further, to minimize effects of market opening and closing, we only consider observations during 9:45 – 15:45. Observations where the midquote is larger (smaller) than 1.3 (0.7) of the daily midquote median are discarded, excluding implausibly extreme price movements. Next, we compute $60, 150, 300$ and 600 second log returns which regularizes the data to an evenly spaced time grid, reducing market microstructure noise.

One major concern with high–frequency data is intra–day seasonality. Figure (5.2) shows estimated intra–day volatility medians. Figure (5.3) shows empirical autocorrelation and partial autocorrelation in case of BAC; Figure (5.7) shows similar results

Figure 5.2: Median intra–day realized volatility of all stocks' five minute returns. For each time point, we take the median over all days.



for all other stocks. All of these figures support the well known empirical facts of uncorrelatedness of first moments, and seasonality and persistence in higher moments. Highly persistent serial dependence violates the assumption of a α–mixing time series and, in this setting, also $(A3^C)$. This would render our test inconsistent unless the data is properly filtered. Martens et al. (2002) summarize appropriate deseasonalization procedures, which are all based on realized volatility measures. In the same way, we impose a standard volatility seasonality model for returns $r_t^{(i)}$, given by

$$r_t^{(i)} = \mu_t^{(i)} + \sigma_t^{(i)} r_t^{*,(i)},$$

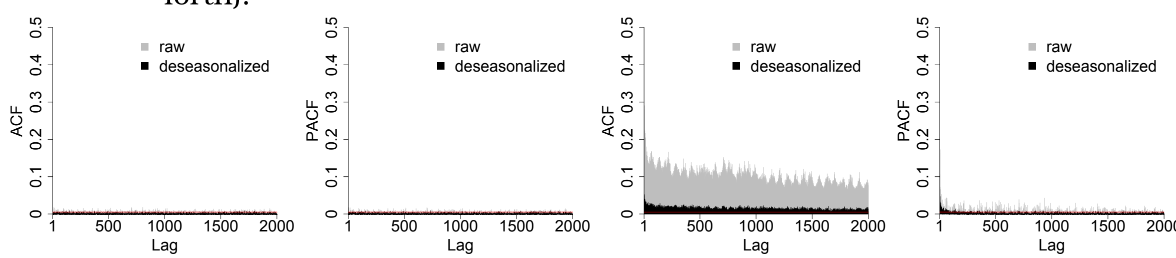
where $\mu_t^{(i)}$ is conditional mean, $\sigma_t^{(i)}$ is conditional volatility, and $r_t^{*,(i)}$ is the deseasonalized component of the observed return $r_t^{(i)}$. The conditional mean $\mu_t^{(i)}$ is zero and we estimate $\sigma_t^{(i)}$ to achieve deseasonalized returns $\widehat{r}_t^{*,(i)} := \frac{r_t^{(i)}}{\widehat{\sigma}_t^{(i)}}$. We estimate the volatility in each intra–day time window $[t_{j-1}, t_j]$ simply by the square root of the average over all squared returns in $[t_{j-1}, t_j]$ — excluding the right end point. Figure (5.7) shows this simple filter successfully wipes out nearly any serial dependence in first and second moments, and also any seasonality.[4] Table (5.4) shows basic summary statistics of raw and deseasonalized returns.

[4]Note, there exist more sophisticated realized volatility measures, such as the two–time–scale by Zhang et al. (2005). However, our aim is to achieve an approximately α–mixing time series which is also possible by this simple cleaning strategy.

Table 5.4: Summary statistics of all stocks' five minute returns, covering 2086 days, and $n = 148106$ observations.

	BAC	GS	JPM	MS	WFC		BAC	GS	JPM	MS	WFC
raw ($*1000$)						desea.					
$q_{0.01}$	-8.824	-6.719	-7.296	-9.400	-7.967		-2.449	-2.193	-2.716	-2.828	-2.724
$\overline{x}$	-0.022	-0.002	-0.002	-0.008	-0.004		-0.007	0.006	0.002	-0.001	-0.003
$\widehat{\sigma}$	2.975	2.385	2.452	3.601	2.647		1.070	0.876	1.152	1.202	1.161
$q_{0.99}$	8.786	6.733	7.298	9.490	8.123		2.449	2.217	2.725	2.829	2.746

Figure 5.3: Exemplary ACF and PACF for BAC. Raw and deseasonalized five minute returns (first and second), and of squared versions thereof (third and forth).

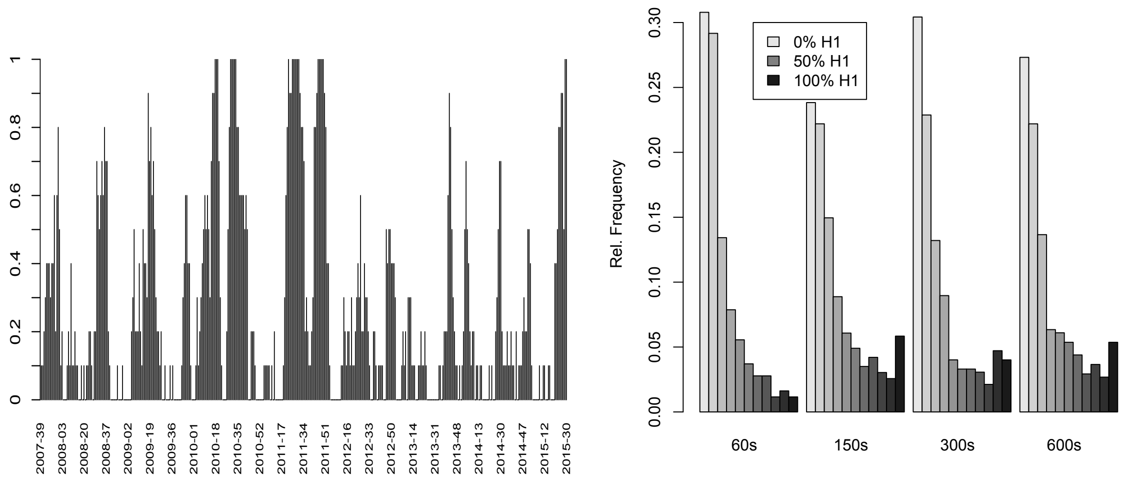


5.5.2 Detecting and localizing of copula time variation

With the high–frequency returns now cleaned, we apply our test to investigate copula dynamics, i.e. how the copula of a single pair evolves over time. Figure (5.4) (right) displays the share of test rejections across all ten pairs for ten minute returns, and relative rejection frequencies over the entire time period for all frequencies. Results appear similar for all frequencies while results of higher frequencies appear a bit more noisy. We thus mostly present results for ten minute returns.

More than 70% of time points feature at least one copula change (out of ten possible changes), rendering high–frequency return copulas time–varying. Hence, copula stationarity, needed for many statistical applications in quantitative finance, can be considered a more than questionable assumption. However, time points where more than 50% of the copulas change, i.e. copula changes in large cross–sections, are rare (10–20%). Test rejections in nearly every pair would indicate a systematic shift in cross–sectional dependence of the entire market. Figure (5.4) (left) precisely documents when copulas change. Wide changes of copulas occurred more often before 2012 than after. Moreover, periods with many test rejections more frequently fall into

Figure 5.4: (Left) Dynamics of rejection frequencies over all pairs for 600 second returns. (Right) Within the cross–section of all ten pairs, we compute the number of test rejections in each period, and plot the relative frequencies over all time windows. For 600 second returns, we observe copula changes in all pairs in nearly 5% of all cases (black bar furthest right).



time periods with economic tension (default of Lehman Brothers in 09/2008, Flash Crash in 05/2010, beginning of the Euro Crisis 08/2011, Chinese stock market crash 06/2015).

Additionally, our test is capable of localizing where copulas change. Figures (5.5) and (5.6) highlight each cell that led to test rejections. Iterating through the entire sample, we repeatedly plot such cells according to their tendency to induce a test rejection. Darkly shaded cells correspond to cells where copulas most often change, allowing us to understand which market conditions cause time variation of copulas. Dependence between average price movements, i.e. the *core* of the copula, surprisingly remains relatively stable over time as the null is only rarely rejected in the middle of the unit square. Interestingly, copula changes most often appear in the upper left and the lower right quadrant, while changes in the upper right and lower left quadrant are scarce. The latter are the joint tail regions, and thus the copula appears to remain stable for joint extremes. However, probabilities for exclusive one–dimensional extreme events constantly change for all pairs and for all frequencies.

5.5.3 Implications for parametric models of time–varying copulas

The finding that copulas most often change in upper left and lower right tails is remarkable in that parametric copula modeling typically aims to dynamically model joint tails to account for time–varying tail dependence, e.g. Patton (2006).

We find that, to fully comprehend copula dynamics, one should also explicitly model exclusive one–dimensional extreme events. This can be achieved by extending the approach by Patton (2006) who employs a time–varying symmetrized Joe–Clayton (sJC) copula that only accounts for dynamics in upper and lower tail dependence. The symmetrized Joe–Clayton (sJC) copula is defined by

$$\begin{aligned} C_{\mathcal{JC}}(x^{(1)}, x^{(2)}; \tau^U, \tau^L) = 1 - (1 - \{[1-(1-x^{(1)})^{\kappa}]^{-\gamma} + [1-(1-x^{(2)})^{\kappa}]^{-\gamma} - 1\}^{-1/\gamma})^{1/\kappa}, \\ \kappa = 1/\log(2-\tau^U), \gamma = -1/\log(\tau^L), \\ \tau^U, \tau^L \in (0,1), \end{aligned}$$

where copula parameters τ^U, τ^L measure upper and lower tail dependence. A copula that explicitly models all quadrants is given by a mixture of a standard sJC copula, $C_{\mathcal{JC}}(x^{(1)}, \; x^{(2)}; \tau^U, \; \tau^L)$, and a sJC copula for $(x^{(1)}, 1-x^{(2)})$, denoted by $C_{\mathcal{JC}}^{1-x^{(2)}}(x^{(1)}, x^{(2)}; \tau^U, \tau^L)$, i.e.

$$\begin{aligned} C^*_{\mathcal{JC}}(x^{(1)}, x^{(2)}; \tau^{UR}, \tau^{LL}, \tau^{UL}, \tau^{LR}, \lambda) :=& \lambda C_{\mathcal{JC}}(x^{(1)}, x^{(2)}; \tau^{UR}, \tau^{LL}) \\ &+ (1-\lambda) C_{\mathcal{JC}}^{1-x^{(2)}}(x^{(1)}, x^{(2)}; \tau^{UL}, \tau^{LR}), \\ &\lambda \in [0,1], \end{aligned}$$

where UR, LL, UL and LR are short for upper right, lower left, upper left and lower right quadrant, respectively, and λ denotes the mixture coefficient, which has to be estimated from the data.

Figure 5.5: Pinning down copula changes (part I). We overlay cells that suggest test rejections across the entire time period. The darker a subregion is, the more often this area induces copula inequality. Note, for clarity, we omit results for the test with $J = 4$.

(a) BAC-GS (b) BAC-JPM (c) BAC-MS (d) GS-JPM (e) GS-MS

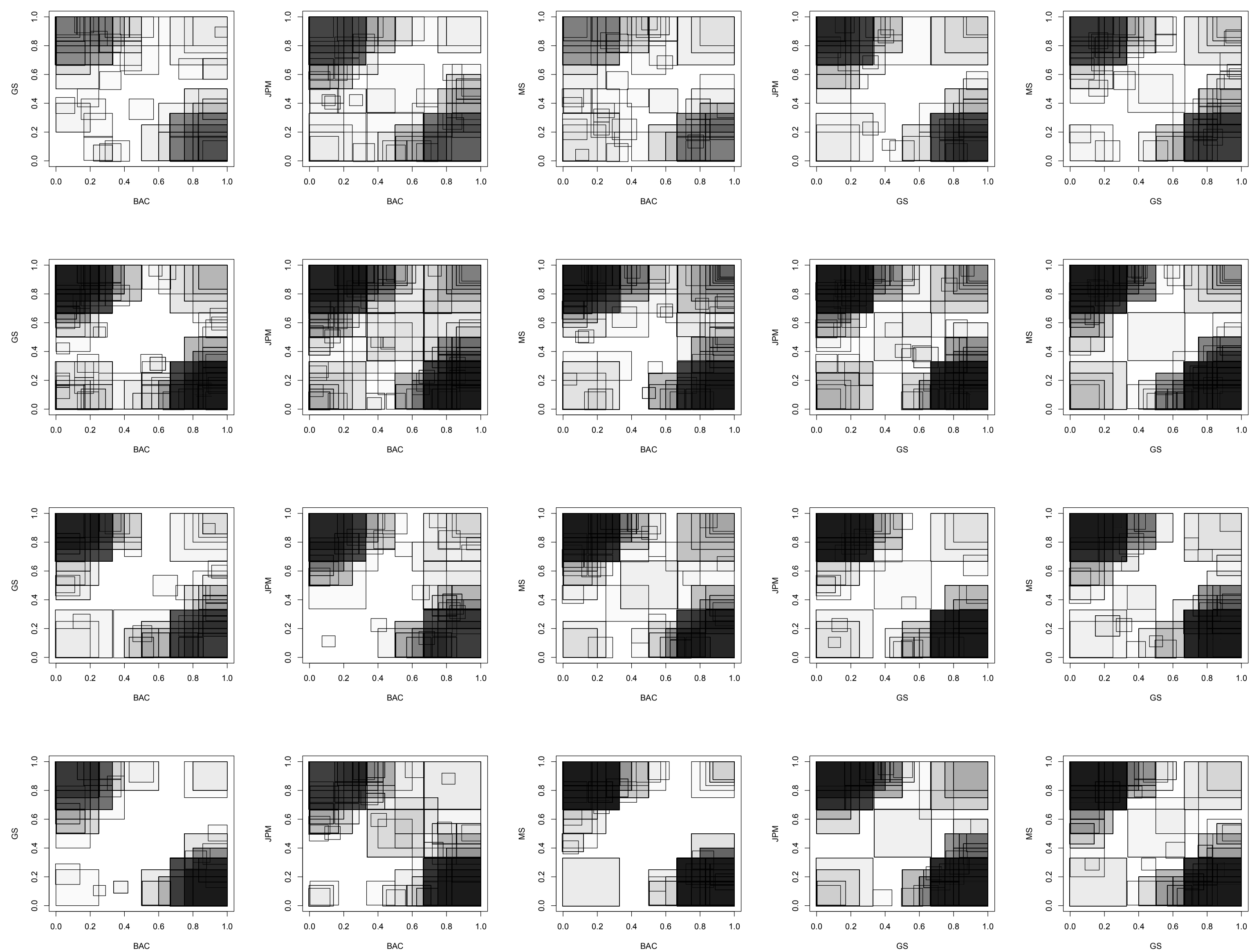
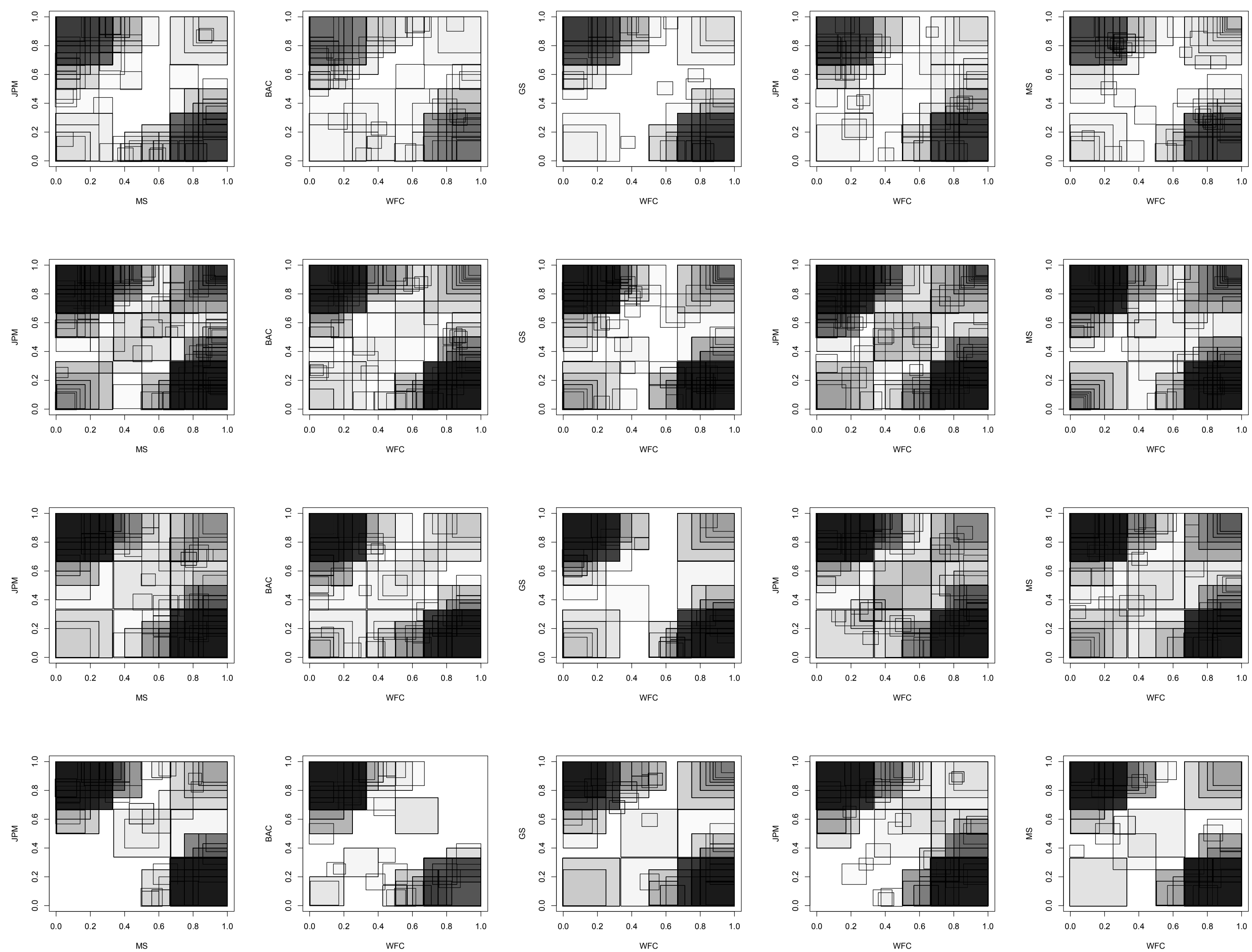


Figure 5.6: Pinning down copula changes (part II). We overlay cells that suggest test rejections across the entire time period. The darker a subregion is, the more often this area indicates copula inequality. Note, for clarity, we omit results for the test with $J = 4$.

(a) MS–JPM (b) WFC–BAC (c) WFC–GS (d) WFC–JPM (e) WFC–MS



5.6 Conclusion

We propose a computationally attractive test against copula equality in a time series framework based on multiple testing. The test accounts for possible copula non–exchangeability, features good finite sample properties and is easy to implement. Importantly, we provide information on which areas of the copula domains induce copula inequality. This improves understanding of time variation of copula structures.

For high–frequency data of five financial stocks, we provide evidence for (serial) time variation in copula structures. We find that copulas typically alter in the upper left and lower right parts of their domain, which is often not modeled in parametric approaches. These findings suggest to extend standard copula models to also explicitly model these often overseen parts of the copula.

5.7 Appendix

5.7.1 Asymptotic properties of the empirical copula

For completeness, we briefly outline the asymptotic properties of the empirical copula. To be precise,

$$\sup_{\mathbf{x}} |\widehat{C}(\mathbf{x}) - C(\mathbf{x})| \xrightarrow{p} 0, n \to \infty,$$

and

$$\sqrt{n}(\widehat{C}(x^{(1)}, x^{(2)}) - C(x^{(1)}, x^{(2)})) \underset{n\to\infty}{\xrightarrow{w}} \mathbb{B}(x^{(1)}, x^{(2)}) - \sum_{k=1}^{2} \partial C_i \mathbb{B}(\mathbf{x}^{(i)})), \forall \mathbf{x} \in [0,1]^2,$$

where $\mathbb{B}$ is a centered bivariate Gaussian process with covariance structure

$$\mathbb{E}(\mathbb{B}(\mathbf{x})\mathbb{B}(\mathbf{v})) = C(\min(x^{(1)}, v^{(1)}), \min(x^{(2)}, v^{(2)})) - C(\mathbf{x})C(\mathbf{v})),$$

where $\mathbf{x}^{(i)} := (x^{(i)}, u_{-i} = 0)$. Weak asymptotic normality of the empirical copula process with estimated marginals can also be established for sequentially dependent data. Assume observations $\mathbb{X}_1, ..., \mathbb{X}_n$ stem from a stationary process X which is strongly mixing with rate $\alpha(r) = \mathcal{O}(r^{-a}), a > 1$. Then

$$\sqrt{n}(\widehat{C}(\mathbf{x}) - C(\mathbf{x})) \xrightarrow{w} \mathbb{B}(\mathbf{x}) - \sum_{i=1}^{2} \partial C_i(\mathbf{x}) \mathbb{B}(\mathbf{x}^{(i)}), \forall \mathbf{x} \in [0,1]^2, \tag{5.7}$$

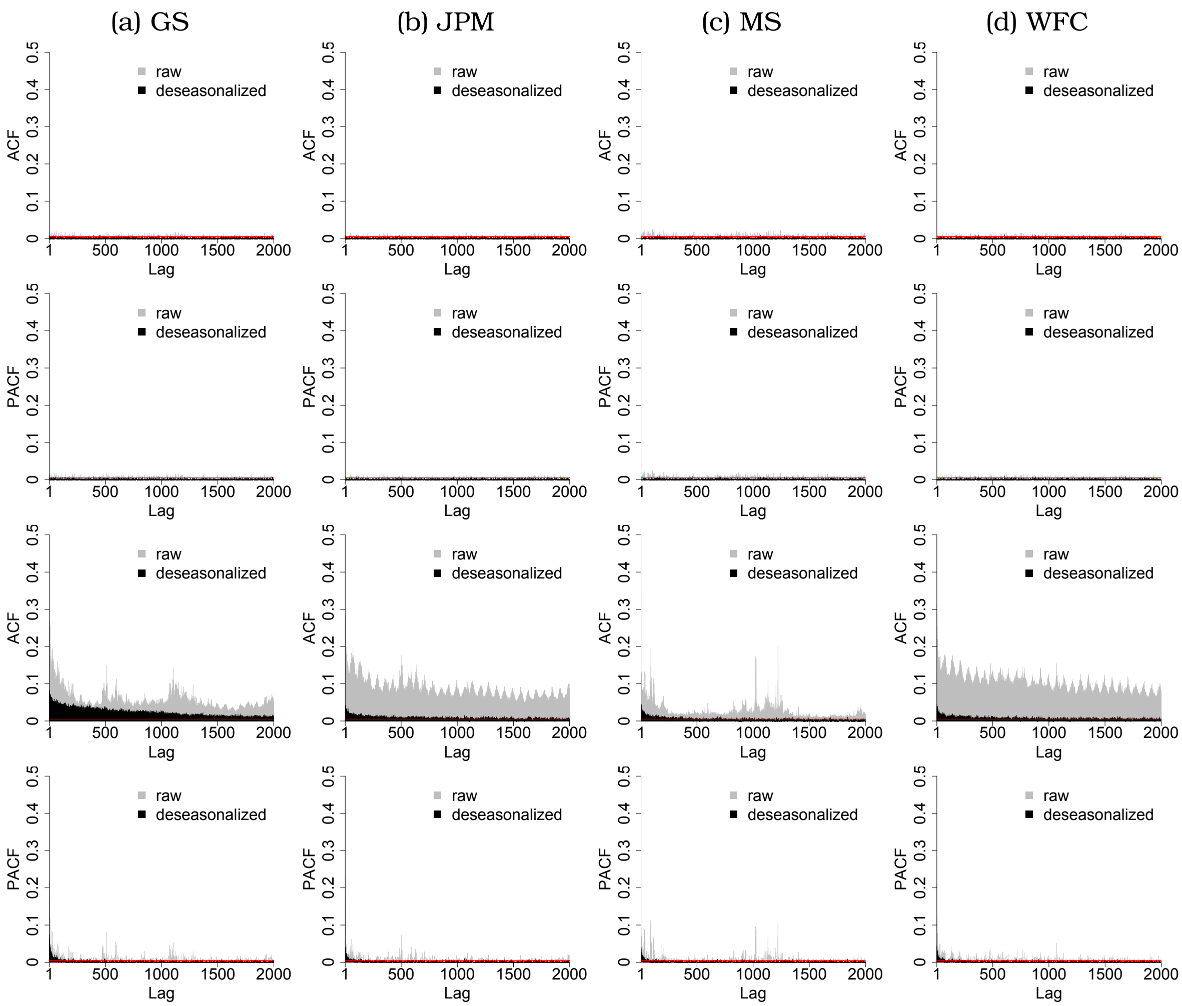
and $\mathbb{B}$ is a bivariate centered tight Gaussian field with covariance

$$\gamma(\mathbf{x}, \mathbf{v}) = \sum_{j \in \mathbb{Z}} \text{Cov}(1\{\tilde{\mathbf{U}}_0 \leq \mathbf{x}\}, 1\{\tilde{\mathbf{U}}_j \leq \mathbf{v}\}, \forall \mathbf{x}, \mathbf{v} \in [0,1]^2,$$

with $\mathbf{U}_t = (F_1(Z_t^{(1)}), F_2(Z_t^{(2)}))$, see Bücher & Ruppert (2013).

5.7.2 Additional empirical results

Figure 5.7: ACF and PACF of raw and deseasonalized five min returns (upper two panels), and of squared versions thereof (lower two panels).



5.7.3 Additional simulation results

Figure 5.8: Density of the sum of estimated GARCH parameters $\hat{\alpha}$ and $\hat{\beta}$ for $n = 10000$ based on 40000 repetitions. Dashed line: True value. This indicates large sample sizes induce long memory, violating Assumption ($A4^C$), and possibly distorting the test size.

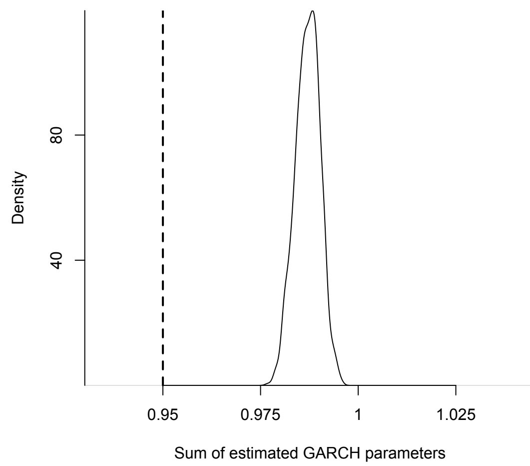


Table 5.5: Empirical rejection frequencies for mixtures of the Gauss and the t–copula, where marginals are i.i.d.

		M_{10}	M_{15}	χ^2_9	M_{10}	M_{15}	χ^2_9	M_{10}	M_{15}	χ^2_9
n	$(\lambda_{\mathbb{X}}, \rho_{\mathbb{X}})$	$\tau_{\mathbb{Y}} = 0.3$			$\tau_{\mathbb{Y}} = 0.6$			$\tau_{\mathbb{Y}} = 0.9$		
	$(1/3, 0.3)$									
250		0.0	0.0	0.7	18.6	18.6	9.3	100	100	100
500		0.4	0.4	0.5	69.2	69.2	43.5	100	100	100
1000		1.4	1.4	0.8	99.0	99.0	96.5	100	100	100
5000		2.4	4.4	0.4	100	100	100	100	100	100
10000		3.5	6.2	0.6	100	100	100	100	100	100
	$(1/3, 0.9)$									
250		100	100	100	66.7	66.7	**NA**	0.0	0.0	**NA**
500		100	100	100	99.9	99.9	100	0.0	0.0	**NA**
1000		100	100	100	100	100	100	0.3	0.3	**NA**
5000		100	100	100	100	100	100	3.6	4.5	**NA**
10000		100	100	100	100	100	100	4.4	5.8	0.0
	$(2/3, 0.6)$									
250		17.0	17.0	8.8	0.2	0.2	0.0	67.6	67.6	**NA**
500		65.1	65.1	38.5	0.4	0.4	0.4	99.8	99.8	100
1000		98.8	98.8	93	1.3	1.3	1.0	100	100	100
5000		100	100	100	4.4	5.8	1.0	100	100	100
10000		100	100	100	5.4	8.9	1.4	100	100	100
	$(1, 0.3)$									
250		0.2	0.2	1.0	15.5	15.5	10.2	100	100	100
500		0.6	0.6	0.7	66.7	66.7	39.2	100	100	100
1000		1.2	1.2	0.9	97.6	97.6	91.3	100	100	100
5000		3.2	5.2	1.0	100	100	100	100	100	100
10000		3.4	6.3	0.8	100	100	100	100	100	100
	$(1, 0.9)$									
250		100	100	100	66.3	66.3	**NA**	0.1	0.1	**NA**
500		100	100	100	99.9	99.9	100	0.2	0.2	**NA**
1000		100	100	100	100	100	100	0.3	0.3	**NA**
5000		100	100	100	100	100	100	2.5	3.1	0.0
10000		100	100	100	100	100	100	5.6	7.2	0.4

		M_{10}	M_{15}	χ^2_9	M_{10}	M_{15}	χ^2_9	M_{10}	M_{15}	χ^2_9
n	$(\lambda_{\mathbb{X}}, \rho_{\mathbb{X}})$	$\tau_{\mathbb{Y}} = 0.3$			$\tau_{\mathbb{Y}} = 0.6$			$\tau_{\mathbb{Y}} = 0.9$		
	$(1/3, 0.6)$									
250		19.6	19.6	7.5	0.2	0.2	0.0	66.4	66.4	**NA**
500		68.8	68.8	42.5	0.5	0.5	1.3	100	100	100
1000		99.3	99.4	95.3	1.3	1.3	0.5	100	100	100
5000		100	100	100	4.1	6.2	0.7	100	100	100
10000		100	100	100	5.3	8.2	0.6	100	100	100
	$(2/3, 0.3)$									
250		0.3	0.3	1.2	16.0	16.0	8.9	100	100	100
500		0.6	0.6	0.5	62.0	62.0	35.9	100	100	100
1000		1.4	1.4	0.6	97.8	97.9	92.7	100	100	100
5000		3.9	6.4	0.6	100	100	100	100	100	100
10000		2.5	5.8	0.4	100	100	100	100	100	100
	$(2/3, 0.9)$									
250		100	100	100	66.5	66.5	**NA**	0.0	0.0	**NA**
500		100	100	100	99.9	99.9	100	0.0	0.0	**NA**
1000		100	100	100	100	100	100	0.4	0.4	**NA**
5000		100	100	100	100	100	100	2.9	3.7	0.0
10000		100	100	100	100	100	100	5.3	6.0	1.7
	$(1, 0.6)$									
250		16.2	16.2	7.8	0.2	0.2	0.0	67.5	67.5	**NA**
500		64.2	64.2	36.3	0.4	0.4	1.0	99.8	99.8	100
1000		98.7	98.7	92.0	0.5	0.5	0.5	100	100	100
5000		100	100	100	4.5	6.8	0.6	100	100	100
10000		100	100	100	3.7	7.7	0.7	100	100	100

Bibliography

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

Amemiya, T. 1985. *Advanced econometrics*. Cambridge, Massachusetts: Harvard University Press.

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

Bank of International Settlement. 2010. *Basel III: A global regulatory framework for more resilient banks and banking systems*, December.

Beare, B. K. 2010. Copulas and temporal dependence. *Econometrica* 78(1): 395-410.

Beare, B. K., and J. Seo. 2014. Time irreversible copula-based Markov models. *Econometric Theory* 30: 923-960.

Beirlant, J., Y. Goegebeur, J. Segers and J. Teugels. 2004. *Statistics of extremes: Theory and applications*. Chichester: Wiley.

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.

Beran, R. 1975. Tail probabilities of noncentral quadratic forms. *Annals of Statistics* 3(4): 969-974.

Berg, D. 2009. Copula goodness-of-fit testing: An overview and power comparison. *European Journal of Finance* 15(7-8): 675-701.

Berg, D., and J. F. Quessy. 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.

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

Bormann, C. 2016. Testing against intra-tail asymmetries in financial time series. Working Paper.

Bormann, C. 2016. Bivariate copula comparisons with multiple testing techniques. Working Paper.

Bormann, C., J. Schaumburg and M. Schienle. 2016. Beyond dimension two: A test for higher-order tail risk. *Journal of Financial Econometrics* 14(3): 552-580.

Bormann, C., and M. Schienle. 2016. Detecting structural differences in tail dependence of financial time series. Working Paper.

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 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.

Chen, X., and Y. Fan. 2006. Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics* 135(1-2): 125-154.

Chen, L., V. Singh , G. Shenglian, and T. Li. 2012. Flood coincidence risk analysis using multivariate copula functions. *Journal of Hydrologic Engineering* 17(6): 742-755.

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

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

Coles, S. G., and J. A. Tawn. 1991. Modelling extreme multivariate events. *Journal of the Royal Statistical Society: Series B* 53(2): 377-392.

Coles, S. G., J. Heffernan, and J. A. Tawn. 1999. Dependence measures for extreme value analyses. *Extremes* 2(4): 339-365.

Coles, S. G., and J. Tawn. 1994. Statistical methods for multivariate extremes: An application to structural design. *Applied Statistics* 43: 1-48.

Cont, R. 2001. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance* 1(2): 223-236.

Danielsson, J., L. de Haan, L. Peng, and C. G. de Vries. 2001. Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis* 76(2): 226-248.

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

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

de Haan, L., C. Neves, and L. Peng. 2008. Parametric tail copula estimation and model testing. *Journal of Multivariate Analysis* 99(6): 1260-1275.

de Haan, L., and S. I. Resnick. 1977. Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 40(4): 317-337.

Demarta, S., and A. J. McNeil. 2005. The t copula and related copulas. *International Statistical Review* 73(1): 111-129.

Dietrich, D., L. de Haan, and J. Hüsler. 2003. Testing extreme value conditions. *Extremes* 5(1): 71-85.

Dobrić, J., and F. Schmid. 2005. Testing goodness of fit for parametric families of copulas - Application to financial data. *Communications in Statistics - Simulation and Computation* 34(4): 1053-1068.

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

Drees, H., L. de Haan, and D. Li. 2006. Approximations to the tail empirical distribution function with application to testing extreme value conditions. *Journal of Statistical Planning and Inference* 136(10): 3498-3538.

Einmahl, J. H. J., L. de Haan, and D. Li. 2006. Weighted approximations of tail copula processes with application to testing the bivariate extreme value condition. *The Annals of Statistics* 34(4): 1987-2014.

Einmahl, J. H. J., A. Krajina, and J. Segers. 2008. A method of moments estimator of tail dependence. *Bernoulli* 14(4): 1003-1026.

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.

El-Nouty, C., and A. Guillou. 2000. On the bootstrap accuracy of the pareto index. *Statistics and Decisions* 18: 275-289.

Embrechts, P., S. Resnick, and G. Samorodnitsky. 1999. Extreme value theory as a risk management tool. *North American Actuarial Journal* 3: 30-41.

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. National Bureau Economic Research, working paper, No 8554.

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

Fermanian, J.-D. Radulović, D., and M. Wegkamp. 2004. Weak convergence of empirical copula processes. *Bernoulli* 10(5): 847-860.

Fisher, R. A., and L. H. C. Tippett. 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society* 24(2): 180-190.

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

Francq, C. and J. M. Zakoïan. 2010. *GARCH models: Structure, statistical inference and financial applications*. Chichester, Wiley.

Fréchet, M. 1927. Sur la loi de probabilité de l' écart maximum. *Annales de la Société Polonaise de Mathématique* 6: 93-116.

Garcia, R., and G. Tsafack. 2011. Dependence structure and extreme comovements in international equity and bond markets. *Journal of Banking & Finance* 35(8): 1954-1970.

Gatev, E., W. N. Goetzmann, and K. G. Rouwenhorst. 2006. Pairs trading: Performance of a relative-value arbitrage rule. *Review of Financial Studies* 19(3): 797-827.

Geluk, J. L., and L. de Haan. 2002. On bootstrap sample size in extreme value theory. *Publications de l'Institut Mathématique Nouvelle série* 71(85): 21-25.

Genest, C., B. Rémillard, R. Garcia, and D. Beaudoin. 2009. Goodness-of-fit tests for copulas: a review and a power study. *Insurance: Mathematics and Economics* 44(2): 199-213.

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

Gnedenko, B. V. 1943. Sur la distribution limite du terme maximum d' une série aléatoire. *Annals of Mathematics* 44(3): 423-453.

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

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

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.

Hautsch, N. 2012. *Econometrics of financial high-frequency data*. Berlin Heidelberg, Springer.

Herrndorf, N. 1984. A functional central limit theorem for weakly dependent sequences of random variables. *The Annals of Probability* 12(1): 141-153.

Hong, Y., J. Tu, and G. Zhou. 2007. Asymmetries in stock returns: Statistical tests and economic evaluation. *The Review of Financial Studies* 20(5): 1547-1581.

Hong, Y., Y. Liu, and S. Wang. 2009. Granger causality in risk and detection of extreme risk spillover between financial markets. *Journal of Econometrics* 150(2): 271-287.

Hu, L. 2006. Dependence patterns across financial markets: A mixed copula approach. *Applied Financial Economics* 16(10): 717-729.

Hua, L., and H. Joe. 2011. Second order regular variation and conditional tail expectation of multiple risks. *Insurance: Mathematics and Economics* 49(3): 537-546.

Huang, X. 1992. *Statistics of bivariate extreme values* (PhD thesis). Erasmus University Rotterdam, Tinbergen Institute Research Series 22.

Huser, R., A. C. Davison, and M. G. Genton. 2016. Likelihood estimators for multivariate extremes. *Extremes* 19(1): 79-103.

Hüsler, J., and D. Li. 2009. Testing asymptotic independence in bivariate extremes. *Journal of Statistical Planning and Inference* 139(3): 990-998.

Jaworski, F., F. Durante, and W. K. Härdle (Eds.). 2013. *Copulae in mathematical and quantitative finance*. Lecture Notes in Statistics - Proceedings, Berlin Heidelberg, Springer.

Joe, H. 1990. Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statistics and Probability Letters* 9(1): 75-81.

Joe, H. 2001. *Multivariate models and dependence concepts*. Monographs on Statistics and Applied Probability, 73. London, Chapman and Hall.

Joe, H., R. L. Smith, and I. Weissman. 1992. Bivariate threshold methods for extremes. *Journal of the Royal Statistical Society: Series B* 54(1): 171-183.

Jondeau, E. 2016. Asymmetry in tail dependence in equity portfolios. *Computational Statistics and Data Analysis* 100:351-368.

Jondeau, E., and M. Rockinger. 2006. The Copula-GARCH model of conditional dependencies: An international stock market application. *Journal of International Money and Finance* 25(5): 827-853.

Khoudraji, A. 1995. Contributions à l' étude des copules et à la modélisation des copules de valeurs extrémes bivariées (PhD thesis), Université de Laval, Québec, Canada.

Klugman, S., and R. Parsa. 1999. Fitting bivariate loss distributions with copulas. *Insurance: Mathematics and Economics* 24(1-2): 139-148.

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

Kojadinovic, I., and J. Yan. 2010. Nonparametric rank-based tests of bivariate extreme-value dependence. *Journal of Multivariate Analysis* 101(9): 2234-2249.

Kojadinovic, I. and J. Yan. 2012. A non-parametric test of exchangeability for extreme-value and left-tail decreasing bivariate copulas. *Scandinavian Journal of Statistics* 39: 480-496. .

Ledford, A. W., and J. A. Tawn. 1996. Statistics for near independence in multivariate extreme values. *Biometrika* 83(1): 169-187.

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

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

Longstaff, F. A. 2010. The subprime credit crisis and contagion in financial markets. *Journal of Financial Economics* 97(3): 436-450.

Martens, M., Y.-C. Chang, and S. J. Taylor. 2002. A comparison of seasonal adjustment methods when forecasting intraday volatility. *The Journal of Financial Research* 25(2): 283-299.

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, Princeton University Press.

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

Mikosch, T, and C. Stărică. 2004. Nonstationarities in financial times series, the long range dependence, and the IGARCH effects. *The Review of Economics and Statistics* 86: 378-390.

Nikoloulopoulos, A. K., H. Joe, and H. Li. 2009. Extreme value properties of multivariate t copulas. *Extremes* 12(2): 129-148.

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

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

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

Patton, A. J. 2012. A review of copula models for economic time series. *Journal of Multivariate Analysis* 110: 4-18.

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

Peng, L. 2010. A practical way for estimating tail dependence functions. *Statistica Sinica* 20(1): 365-378.

Peng, L., and Y. Qi. 2007. Partial derivatives and confidence intervals of bivariate tail dependence functions. *Journal of Statistical Planning and Inference* 137(7): 2089-2101.

Proschan, M. A., and P. A. Shaw. 2011. Asymptotics of Bonferroni for dependent normal test statistics. *Statistics & Probability letters* 81(7): 739-748.

Politis, D. N., and J. P. Romano. 1994. Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics* 22(4): 2031-2050.

Poon, S.-H., M. Rockinger, and J. A. Tawn. 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(1): 81-97.

Rémillard, B. 2010. "Goodness-of-fit tests for copulas of multivariate time series." Working Paper; Available at SSRN: http://ssrn.com/abstract=1729982 .

Rémillard, B., and O. Scaillet. 2009. Testing for equality between two copulas. *Journal of Multivariate Analysis* 100(3): 377-386.

Resnick, S. I., and L. de Haan. 1996. Second-order regular variation and rates of convergence in extreme-value theory. *The Annals of Probability* 24(1): 97-124.

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

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

Rodriguez, J. C. 2007. Measuring financial contagion: A copula approach. *Journal of Empirical Finance* 14(3): 401-423.

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

Rüschendorf, L. 1976. Asymptotic distributions of multivariate rank order statistics. *The Annals of Statistics* 4(5): 912-923.

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.

Segers, J. 2012. Max-stable models for multivariate extremes. *Revstat - Statistical Journal* 10(1): 61-82.

Segers, J. 2012. Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli* 18(3): 764-782.

Sibuya, M. 1960. Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics* 11(3): 195-210.

Smith, R. L. 1985. Maximum likelihood estimation in a class of nonregular cases. *Biometrika* 72(1): 67-90.

Stange, J., T. Bodnar, T. Dickhaus. 2015. Uncertainty quantification for the family-wise error rate in multivariate copula models. *AStA Advances in Statistical Analysis* 99(3)(1): 281–310.

Stephenson, A. 2002. evd: Extreme value distributions. *R News* 2(2): 31-32. `http://CRAN.R-project.org/doc/Rnews/`

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

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

Taleb, N. N. 2007. *The black swan: the impact of the highly improbable.* New York: Random House.

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

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia. 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100(472): 1394-1411.

List of Figures

List of Tables