

Telling tales from the tails: high-dimensional tail interdependence

Arnold Polanski^a, Evarist Stoja^b, Frank Windmeijer^b

^a*University of East Anglia, UK*

^b*University of Bristol, UK*

Abstract

We propose a simple and flexible framework that allows for a comprehensive analysis of tail interdependence in high dimensions. We use co-exceedances to capture the structure of the dependence in the tails and, relying on the concept of multi-information, define the coefficient of tail interdependence. Within this framework, we develop statistical tests of (i) independence in the tails, (ii) goodness-of-fit of the tail interdependence structure of a hypothesized model with the one observed in the data, and (iii) dependence symmetry between any two tails. We present an analysis of tail interdependence among 250 constituents of the SP250 index.

Key words: co-exceedance, tail interdependence, high dimensions

JEL: C12, C13, C14, C32, C58

Corresponding author: Evarist Stoja, School of Economics, Finance and Management, University of Bristol, Priory Road, Bristol, BS8 1TU, UK, email: E.Stoja@bristol.ac.uk, tel. +44(0)1173941497

We would like to thank the editor Eric Ghysels and two anonymous referees for their insightful comments and suggestions that have helped to improve the paper greatly. We have also benefited from discussions with Richard Harris, Nick Taylor, Karin Thorburn, Svein-Arne Persson, Tore Leite, Sujit Kapadia, Marcin Kacperczyk, Lauren Cohen, Robin Greenwood, Peter Pope, Fernando Vega Redondo, Raphael Markellos and the seminar and conference participants at University of Bristol, University of East of Anglia, Chinese University of Hong Kong, Cambridge University Isaac Newton Institute, Systemic Risk Centre at LSE, Bank of England, European Central Bank, World Finance Conference (Buenos Aires, 2015) and thank Europlace Institute of Finance (EIF) for research funding. Parts of this paper were written while Evarist Stoja was a Houbлон-Norman Fellow at the Bank of England whose hospitality is gratefully acknowledged. The views expressed here are solely our own and do not necessarily reflect those of the Bank of England.

1. Introduction

Tail interdependence¹ is of paramount importance in economics, finance, insurance and in many other areas of applied probability and statistics. Research has documented that dependence² has a complex nature, is strongly non-normal, with a time-varying strength and shape (e.g. Patton, 2006). To date, simultaneously capturing these characteristics, particularly for high-dimensional systems, has proved difficult.

We propose a simple and flexible framework that allows for a comprehensive analysis of tail interdependence in high dimensions and apply it to the 250 constituents of the Standard & Poor index. The essence of this framework is the concept of tail interdependence of the constituents of a high-dimensional system in a cross-section or intertemporal context. Specifically, we focus on co-exceedances – joint occurrences of extreme events. Under tail interdependence, the co-exceedances of some variables are informative about such events for other variables. Conversely, under independence, tail events in any subset of variables do not convey any information about tail events in other subsets. We use co-exceedances to capture the structure of the tail interdependence and, relying on the concept of multi-information, define the coefficient of tail interdependence. Multi-information is used in many areas of natural sciences and has been also applied in economics and finance (see, e.g. Backus et al., 2014; Segoviano and Goodhart, 2009; Cameron and Windmeijer, 1997).

Our measure of tail dependence is fully non-parametric, easy to interpret and particularly suitable for empirical studies of high-dimensional problems, typically encountered in economics and finance. Importantly, it can be computed along any

¹The pun in the title of the paper has been used before in a different form by Andrade et al. (2012).

²While we often use the terms interdependence and dependence interchangeably, we distinguish between the two concepts as follows. Dependence refers to the relationship between two random variables whereas interdependence refers to the relationship among two or more variables. Hence, the latter concept nests the former.

direction and for any nominal severity level as specified by the user's interest. It can be decomposed into two components. The first component captures the interdependence of the system so we label it *system interdependence*. The second component is the residual interdependence which, for reasons that will become clear below, we label *severity- k interdependence*. Alternatively, our measure of tail dependence can be decomposed into contributions of constituents to the interdependence of a system. These decompositions can, for example, help portfolio managers identify assets that make large contributions to the tail risk of a portfolio or regulators identify systemically important institutions. We offer, therefore, a tool to analyse the tail dependence of large networks without requiring knowledge of or imposing assumptions on their "inner workings", relying instead only on their observed behavior as captured in a series of multidimensional observations.

In the bivariate case, our measure of tail dependence converges to the tail-dependence coefficient (e.g. Joe, 1997; Sibuya, 1960) and can be interpreted as a generalisation of the latter to high dimensions and to any severity level or direction. The tail-dependence coefficient corresponds to the limit of the conditional probability that one variable exceeds a high/low threshold given that the other has exceeded its high/low threshold.

Further, we develop statistical tests of (i) independence in the tails, (ii) goodness-of-fit of the tail interdependence structure of a hypothesized model relative to the structure observed in the data, and (iii) dependence symmetry between any two tails. These tests can be employed unconditionally and, importantly, conditionally to distinguish between different models of conditional dependence such as multivariate GARCH or time-varying copulas (e.g. Dungey and Renault, 2018; McCloud and Hong, 2011). These contributions are provided within a unified framework that relies on the same fundamental concept of multi-information.

The distinction of dependence in the tails relative to other parts of the distrib-

tion is important because financial and economic variables are often dependent in some parts but independent in others. For example, as we show below, the Fama-French-Carhart asset pricing factors are virtually independent in the central part of their joint distribution but dependent in the tails. Indeed, dependence changes considerably with the extremity of returns and this discrepancy in dependence is pervasive. Accounting for this feature is essential for many applications such as hedging and portfolio diversification. In particular, the standard advice to hold a well-diversified portfolio might be questionable if assets are independent in the central part of the joint distribution but tend to fall significantly in value as one of them experiences an extreme drop, thereby further compounding large losses instead of offsetting them. Following Ghysels et al. (2016), we examine the performance of two portfolios that exploit the asymmetry of the joint distribution of returns captured by either co-skewness or the coefficient of tail dependence that we propose. We find that "co-skewness-managed" portfolio has the highest return whereas the "tail-risk-managed" portfolio has the lowest risk relative to the alternative models we use.

To the best of our knowledge, there is no direct, high-dimensional non-parametric alternative to our framework. An alternative would be the factor-copula class of models (see Oh and Patton, 2017; Lucas et al., 2016 and the references therein for a discussion of the advantages and disadvantages of these models). However, the framework that we propose is different from these models because ours is entirely non-parametric. While this has advantages, it is like other non-parametric techniques data-intensive. However, our system coefficient of tail interdependence addresses this drawback by drastically reducing the number of the relevant joint tails. Monte Carlo studies available in the Appendix show that our measure of tail interdependence has good finite-sample properties even in very high dimensions.

The literature contains several measures of dependence (e.g. Li, 2009; Colangelo

et al., 2005; Heffernan, 2001; Joe, 1997; 1989). The most widely applied dependence measure, the Pearson’s correlation coefficient, is an inadequate measure in many situations as it captures only the linear dependence between pairs of random variables (e.g. Longin and Solnik, 2001). Extreme dependence has been captured by copulas (e.g. Opitz et al., 2017; Nelsen, 2007; Patton, 2006), multivariate quantile regressions (e.g. White et al., 2015) and multivariate extreme-value theory (e.g. Asimit et al., 2016; Bücher et al., 2015; and Jansen and de Vries, 1991). However, these measures of dependence are generally feasible only in low dimensions. Further, there is a large literature that uses co-exceedances to model and test for interdependence (see, for example, Bae et al., 2003; Boyson, et al., 2010; Dungey et al., 2015; and Beine et al., 2015).

The paper proceeds as follows. Section 2 introduces the joint tails and the tail interdependence structure. In Section 3, we define the coefficients of tail interdependence and discuss their properties. Section 4 develops the tail interdependence structure framework for hypothesis testing, while Section 5 applies it, conditionally and unconditionally, to multidimensional data. Section 6 presents the results of optimal portfolios constructed with the coefficient of tail interdependence and Section 7 summarizes the paper. All technical proofs are relegated to the Appendix.

2. Joint tails and the tail interdependence structure

We begin with the simplest example of the bivariate case and generalise the framework to the n dimensional case below. Let $\mathcal{N} = \{1, 2\}$ be a set of two assets with random returns represented by the vector $\mathbf{X} = (X_1, X_2)$ with a continuous joint cdf F (pdf f) with support on R^2 and strictly increasing marginal cdfs F_1, F_2 (pdfs f_1, f_2). Suppose that the realization $\mathbf{x} = (x_1, x_2)$ of \mathbf{X} represents the returns of two assets. An investor who is long on these assets experiences an extreme loss at nominal severity level $\alpha \in (0, 1)$ in asset i when the asset return x_i is smaller than

(exceeds) its relevant quantile $F_i^{-1}(\alpha)$. In this case, we define the indicator function,

$$D_i^\alpha(\mathbf{x}; -1) = I(x_i \leq F_i^{-1}(\alpha)).$$

$D_i^\alpha(\mathbf{x}; -1)$ takes on the value of one when the return x_i lies in the left tail with density mass α and zero otherwise. Extreme gains for this investor (the right tail) at nominal severity level α are represented in the analogous way,

$$D_i^\alpha(\mathbf{x}; 1) = I(x_i \geq F_i^{-1}(1 - \alpha)).$$

Formally, the arguments -1 and 1 of the indicator function $D_i^\alpha(\mathbf{x}; \cdot)$ stand for the direction in which the (unidimensional) tails are measured, specifically the left (negative) and the right (positive) tails respectively.

Next, we generalize unidimensional tails to bivariate joint tails and define the indicator function for the latter. For each subset $C \in \{\{1, 2\}, \{1\}, \{2\}, \emptyset\}$ of $\mathcal{N} = \{1, 2\}$ and the directional vector $\mathbf{d} = (d_1, d_2) \in \{-1, 1\}^2$, the joint tail (JT) $T_C^\alpha(\mathbf{d}) \subseteq R^2$ is the subset of the sample space where x_i falls into the negative ($d_i = -1$) or the positive ($d_i = 1$) tail if $i \in C$. The corresponding indicator function $D_C^\alpha(\mathbf{x}; \mathbf{d})$ is computed as,

$$\begin{aligned} D_{\{1,2\}}^\alpha(\mathbf{x}; \mathbf{d}) &= D_1^\alpha(\mathbf{x}; d_1) \cdot D_2^\alpha(\mathbf{x}; d_2), \\ D_{\{1\}}^\alpha(\mathbf{x}; \mathbf{d}) &= D_1^\alpha(\mathbf{x}; d_1) \cdot (1 - D_2^\alpha(\mathbf{x}; d_2)), \\ D_{\{2\}}^\alpha(\mathbf{x}; \mathbf{d}) &= (1 - D_1^\alpha(\mathbf{x}; d_1)) \cdot D_2^\alpha(\mathbf{x}; d_2), \\ D_{\emptyset}^\alpha(\mathbf{x}; \mathbf{d}) &= (1 - D_1^\alpha(\mathbf{x}; d_1)) \cdot (1 - D_2^\alpha(\mathbf{x}; d_1)). \end{aligned} \tag{1}$$

For example, $D_{\{1\}}^\alpha(\mathbf{x}; (-1, -1)) = 1$ indicates that x_1 falls in the negative tail, i.e., $x_1 \leq F_1^{-1}(\alpha)$ and $D_1^\alpha(\mathbf{x}; -1) = 1$, while x_2 does not fall in its respective negative tail, i.e., $x_2 > F_2^{-1}(\alpha)$ and $D_2^\alpha(\mathbf{x}; -1) = 0$. We can define the JT $T_C^\alpha(\mathbf{d})$ succinctly

using the corresponding indicator function $D_C^\alpha(\mathbf{x}; \mathbf{d})$ as

$$T_C^\alpha(\mathbf{d}) = \{\mathbf{x} \in R^2 : D_C^\alpha(\mathbf{x}; \mathbf{d}) = 1\}. \quad (2)$$

It can be readily verified that the superset $\mathcal{T}^\alpha = \{T_C^\alpha(\mathbf{d}) : C \subseteq \{1, 2\}\}$ partitions the outcome space into 2^2 (the number of all subsets of $\mathcal{N} = \{1, 2\}$) regions. Figure 1 illustrates this partition for $\mathbf{d} = -(1, 1)$.

[Figure 1]

Our definition of the JTs naturally generalizes to n dimensions. Let $\mathcal{N} = \{1, \dots, n\}$ be a finite set and F a continuous³ joint cdf (pdf f) of a random vector $\mathbf{X} = (X_1, \dots, X_n)$ with the support on R^n and strictly increasing marginal cdfs F_1, \dots, F_n (pdfs f_1, \dots, f_n). The following definition is a direct generalization of (1) and (2) to any (finite) n .

Definition 1. For an n -dimensional observation $\mathbf{x} = (x_1, \dots, x_n)$, the subset $C \subseteq \mathcal{N}$, the direction vector $\mathbf{d} \in \{-1, 1\}^n$ and the nominal level $\alpha \in (0, 1)$, we define the JT indicator,

$$D_C^\alpha(\mathbf{x}; \mathbf{d}) = \prod_{i \in C} D_i^\alpha(\mathbf{x}; d_i) \prod_{j \in \mathcal{N} \setminus C} (1 - D_j^\alpha(\mathbf{x}; d_j)),$$

and the corresponding JT,

$$T_C^\alpha(\mathbf{d}) = \{\mathbf{x} \in R^n : D_C^\alpha(\mathbf{x}; \mathbf{d}) = 1\}.$$

The subset C in this definition contains those variables from $\mathcal{N} = \{1, \dots, n\}$ that have experienced an extreme (positive or negative) event. For a given directional vector \mathbf{d} and nominal severity level α , there are 2^n joint tails as there are 2^n subsets of \mathcal{N} (including the empty set). We refer to upper (or positive) JTs when $\mathbf{d} = \mathbf{1}$

³The assumptions on F are for ease of notation and are not crucial for our framework.

$= (1, \dots, 1)$ and to lower (or negative) JTs when $\mathbf{d} = -\mathbf{1}$. Otherwise, we call the JTs mixed. The directional vector can capture, for instance, the exposure of an investor, institution, economy or, more broadly, system to positive and/or negative extreme events in the underlying risk sources. Although we focus mainly on $\mathbf{d} = -\mathbf{1}$ and $\mathbf{d} = \mathbf{1}$, it is not difficult to imagine a scenario where some elements of \mathbf{d} are positive and some are negative. An intuitive example is a sophisticated investor who is long in some assets and short in others.

Given a joint pdf f , we compute the probability of each JT and collect these probabilities in the *tail interdependence structure (TIS)*.

Definition 2. The tail interdependence structure $\mathbf{p}^\alpha(f, \mathbf{d}) = (p_C^\alpha(f, \mathbf{d}))_{C \subseteq \mathcal{N}}$ is a 2^n -dimensional vector, where

$$p_C^\alpha(f, \mathbf{d}) = \int D_C^\alpha(\mathbf{x}; \mathbf{d}) f(\mathbf{x}) d\mathbf{x}, \quad (3)$$

is the probability mass of the JT $T_C^\alpha(\mathbf{d})$ under f .

For example, an investor might be interested in the probabilities that all ($C = \mathcal{N}$) or none ($C = \emptyset$) of the assets in her portfolio experience an extreme loss,

$$\begin{aligned} p_{\mathcal{N}}^\alpha(f, -\mathbf{1}) &= \int D_{\mathcal{N}}^\alpha(\mathbf{x}; -\mathbf{1}) f(\mathbf{x}) d\mathbf{x} = \Pr\{X_1 \leq F_1^{-1}(\alpha), \dots, X_n \leq F_n^{-1}(\alpha)\}, \\ p_{\emptyset}^\alpha(f, -\mathbf{1}) &= \int D_{\emptyset}^\alpha(\mathbf{x}; -\mathbf{1}) f(\mathbf{x}) d\mathbf{x} = \Pr\{X_1 > F_1^{-1}(\alpha), \dots, X_n > F_n^{-1}(\alpha)\}. \end{aligned}$$

The superset $\mathcal{T}^\alpha(\mathbf{d}) = \{T_C^\alpha(\mathbf{d}) : C \subseteq \mathcal{N}\}$ partitions the outcome space into 2^n (the number of all subsets of \mathcal{N}) regions. The TIS $\mathbf{p}^\alpha(f, \mathbf{d})$ is, therefore, a discrete pdf. By construction, it holds for this pdf that the probabilities $p_C^\alpha(f, \mathbf{d})$ for all subsets $C \subseteq \mathcal{N}$ that contain a particular $i \in \mathcal{N}$ sum up to

$$\sum_{C \subseteq \mathcal{N}: i \in C} p_C^\alpha(f, \mathbf{d}) = \Pr\{X_i \leq F_i^{-1}(\alpha)\} = \Pr\{X_i \geq F_i^{-1}(1 - \alpha)\} = \alpha.$$

When there is no risk of confusion, we write \mathcal{T}^α , T_C^α and \mathbf{p}^α instead of $\mathcal{T}^\alpha(\mathbf{d})$, $T_C^\alpha(\mathbf{d})$ and $\mathbf{p}^\alpha(f, \mathbf{d})$, respectively.

3. Measurement of tail interdependence

3.1. Multi-information and tail interdependence

The interdependence in the tails as captured by the joint pdf $\mathbf{p}^\alpha = (p_C^\alpha)_{C \subseteq \mathcal{N}}$ is fully defined by the *multi-information* (MI) (Cover and Thomas, 2006),

$$MI(\mathbf{p}^\alpha || \boldsymbol{\pi}^\alpha) = \sum_{C \subseteq \mathcal{N}} p_C^\alpha \ln \frac{p_C^\alpha}{\pi_C^\alpha}, \quad (4)$$

where $\pi_C^\alpha = \alpha^{|C|}(1 - \alpha)^{n - |C|}$ is the probability of the JT T_C^α under tail independence (computed as the product of marginal probabilities of $|C|$ exceedances and $n - |C|$ non-exceedances), $|C|$ is the cardinality of set C and $\boldsymbol{\pi}^\alpha = (\pi_C^\alpha)_{C \subseteq \mathcal{N}}$ is the corresponding TIS under tail independence. Note that $MI(\mathbf{p}^\alpha || \boldsymbol{\pi}^\alpha)$ is well-defined as $\pi_C^\alpha > 0$ for all $\alpha \in (0, 1)$ and $C \subseteq \mathcal{N}$. Although the probabilities p_\emptyset^α and π_\emptyset^α of the no-exceedance event T_\emptyset^α are used in the computation of MI, these probabilities are fully determined by the probabilities of the other joint tails (because all tail probabilities sum up to one). In this sense, they do not contain any independent information and the computation of MI relies exclusively on the information in the probabilities of genuine joint tails with at least one exceedance.

MI is non-negative and equals zero in case of independence only, i.e. if and only if $\mathbf{p}^\alpha = \boldsymbol{\pi}^\alpha$ (Cover and Thomas, 2006). $MI(\mathbf{p}^\alpha || \boldsymbol{\pi}^\alpha)$ is the Kullback-Leibler (KL) divergence between the pdfs \mathbf{p}^α and $\boldsymbol{\pi}^\alpha$ and quantifies the *total* amount of interdependence among random variables that arises from pairwise, triplet or more complex interactions. It is widely used, e.g. in physics (e.g. Schneidman et al., 2003; Chicharro and Ledberg, 2012) and biosciences (e.g. Wennekers and Ay, 2003).

3.2. The coefficient of tail interdependence and its properties

We use MI as defined in (4) to measure tail interdependence. Specifically, we normalize MI to obtain the *coefficient of tail interdependence*.

Definition 3. For the TIS $\mathbf{p}^\alpha = (p_C^\alpha)_{C \subseteq \mathcal{N}}$ and $\boldsymbol{\pi}^\alpha = (\pi_C^\alpha)_{C \subseteq \mathcal{N}}$, where $\pi_C^\alpha = \alpha^{|C|}(1 - \alpha)^{n-|C|}$, the coefficient of tail interdependence (CTI) is computed as,

$$\kappa(\mathbf{p}^\alpha) = \frac{MI(\mathbf{p}^\alpha || \boldsymbol{\pi}^\alpha)}{(1 - n) \ln(\alpha^\alpha(1 - \alpha)^{1-\alpha})}. \quad (5)$$

Proposition 1. *The CTI lies in the unit interval: $\kappa(\mathbf{p}^\alpha) \in [0, 1]$.*

The CTI as defined in (5) has the following properties. Firstly, as shown in Proposition 1, it lies in the unit interval. In particular, $\kappa(\mathbf{p}^\alpha) = 0$ when all exceedances are mutually independent and $\kappa(\mathbf{p}^\alpha) = 1$ in the case of perfect dependence, i.e., when all n variables always exceed together their respective quantiles. Secondly, the CTI is invariant under the permutation of the random variables in \mathbf{X} and scale invariant under strictly increasing transformations of the underlying variables in \mathbf{X} . Specifically, if each $\xi_i(X_i)$ is an increasing and continuous function, then the CTI computed from the transformed variables $\xi(\mathbf{X}) = (\xi_i(X_i))_{i=1, \dots, n}$ is the same as that computed from \mathbf{X} . This property follows by the construction of the TIS from the quantiles of the variables in \mathbf{X} as the same events fall into a JT T_C^α under \mathbf{X} and under $\xi(\mathbf{X})$. Furthermore, sample estimators of TIS/CTI are robust to outliers and asymptotically consistent, i.e. converge to the true TIS/CTI as the sample size grows without bounds.

It is important to note that the empirical computation of the CTI does not suffer from the curse of dimensionality. The computational burden increases only linearly in the product $n \cdot T$ of the dimension n and the sample size T . However, reliable computation of the CTI does require that the sample size is at least of the order 2^n (i.e. the total number of JTs). In the next subsection, we show that the

decomposition of the CTI into the *system* and *severity-k* components circumvents this problem. In higher dimensions, we focus on the system CTI that only needs reliable estimates of the $n + 1$ probabilities of joint exceedances.

Importantly, in the bivariate case, the limit of the CTI as $\alpha \rightarrow 0$ coincides with the lower (upper) tail-dependence coefficient in a definition that goes back to Sibuya (1960; see also Joe, 1997).

Proposition 2. *For the vector $\mathbf{X} = (X_1, X_2)$ with a continuous joint PDF f and continuous, strictly increasing marginal CDFs F_1, F_2 ,*

$$\begin{aligned} \lim_{\alpha \rightarrow 0} \kappa(\mathbf{p}^\alpha(f, -\mathbf{1})) &= \lambda_L \equiv \lim_{\alpha \rightarrow 0} \Pr\{X_1 \leq F_1^{-1}(\alpha) | X_2 \leq F_2^{-1}(\alpha)\}, \\ \lim_{\alpha \rightarrow 0} \kappa(\mathbf{p}^\alpha(f, \mathbf{1})) &= \lambda_U \equiv \lim_{\alpha \rightarrow 0} \Pr\{X_1 \geq F_1^{-1}(1 - \alpha) | X_2 \geq F_2^{-1}(1 - \alpha)\}. \end{aligned}$$

The equivalence of the CTI and the tail-dependence coefficient in the limit is an important finding that connects our framework to the literature on the multivariate extreme value theory (e.g. Asimit et al., 2016; Bücher et al., 2015). We will pursue this avenue in future research.

3.3. System and additional tail interdependence

The dimension of the TIS grows exponentially in n , which complicates its (empirical) application for larger n . To address this issue, we compress the 2^n -dimensional TIS into an $(n + 1)$ -dimensional vector $\tilde{\mathbf{p}}^\alpha = (\tilde{p}_k^\alpha)_{k=0}^n$, where \tilde{p}_k^α is the probability of observing $k = 0, \dots, n$ tail events.

Definition 4. For the TIS \mathbf{p}^α , the *system TIS* is the $(n + 1)$ -dimensional vector $\tilde{\mathbf{p}}^\alpha = (\tilde{p}_k^\alpha)_{k=0}^n$ where

$$\tilde{p}_k^\alpha = \sum_{C \subseteq \mathcal{N}: |C|=k} p_C^\alpha, \quad \text{for } k = 0, 1, \dots, n.$$

Obviously, $\tilde{\mathbf{p}}^\alpha$ is also a (discrete) pdf. In the special case of mutual independence of tail events, we denote the system TIS by $\tilde{\boldsymbol{\pi}}^\alpha = (\tilde{\pi}_k^\alpha)_{k=0}^n$, where

$$\tilde{\pi}_k^\alpha = \sum_{C \subseteq \mathcal{N}: |C|=k} \alpha^k (1-\alpha)^{n-k} = \binom{n}{k} \alpha^k (1-\alpha)^{n-k}.$$

From the TIS \mathbf{p}^α , we compute also the conditional probabilities $\mathbf{p}^{\alpha,k} = (p_C^\alpha / \tilde{p}_k^\alpha)_{C \subseteq \mathcal{N}: |C|=k}$ given that k exceedances have occurred.⁴ Similarly, we compute the conditional probabilities $\boldsymbol{\pi}^{\alpha,k} = (\pi_C^\alpha / \tilde{\pi}_k^\alpha)_{C \subseteq \mathcal{N}: |C|=k}$ from the pdf $\boldsymbol{\pi}^\alpha$ for each $k = 0, \dots, n$. With this notation in place, we define the system and severity- k MI and CTI.

Definition 5. For the TIS $\mathbf{p}^\alpha = (p_C^\alpha)_{C \subseteq \mathcal{N}}$, $\boldsymbol{\pi}^\alpha = (\pi_C^\alpha)_{C \subseteq \mathcal{N}}$ and the resulting PDFs $\tilde{\mathbf{p}}^\alpha$ and $\mathbf{p}^{\alpha,k}$, we compute the *system* and *the severity- k multi-information*, $MI(\tilde{\mathbf{p}}^\alpha || \tilde{\boldsymbol{\pi}}^\alpha)$ and $MI(\mathbf{p}^{\alpha,k} || \boldsymbol{\pi}^{\alpha,k})$, respectively. After normalization, we obtain the *system* and *the severity- k CTI*,

$$\tilde{\kappa}(\mathbf{p}^\alpha) = \frac{MI(\tilde{\mathbf{p}}^\alpha || \tilde{\boldsymbol{\pi}}^\alpha)}{(1-n) \ln(\alpha^\alpha (1-\alpha)^{1-\alpha})}, \quad \kappa^k(\mathbf{p}^\alpha) = \frac{MI(\mathbf{p}^{\alpha,k} || \boldsymbol{\pi}^{\alpha,k})}{(1-n) \ln(\alpha^\alpha (1-\alpha)^{1-\alpha})}. \quad (6)$$

The measure $\tilde{\kappa}(\mathbf{p}^\alpha)$ quantifies the *system* tail interdependence by the normalized divergence between the distributions $\tilde{\mathbf{p}}^\alpha$ and $\tilde{\boldsymbol{\pi}}^\alpha$ of the *total* number of exceedances under \mathbf{p}^α and under $\boldsymbol{\pi}^\alpha$ (i.e., under tail independence), respectively. On the other hand, each $\kappa^k(\mathbf{p}^\alpha)$ quantifies the interdependence among variables, given that k exceedances have occurred. Thus, while $\tilde{\kappa}(\mathbf{p}^\alpha)$ measures the dependence that is jointly generated by all constituents, $\kappa^k(\mathbf{p}^\alpha)$ is the additional dependence among the constituents given that k of them have exceeded (i.e. a severity- k event has occurred).⁵ Our next result shows that the MI (CTI) can be decomposed into *system*

⁴We set $p^{\alpha,k} = 0$ when $\tilde{p}_k^\alpha = 0$. In the bivariate case, for example, $p_{\{2\}}^{\alpha,1} = p_{\{2\}}^\alpha / \tilde{p}_1^\alpha = p_{\{2\}}^\alpha / (p_{\{1\}}^\alpha + p_{\{2\}}^\alpha)$ is the conditional probability of X_2 exceeding when $k = 1$, i.e., when exactly one exceedance has occurred.

⁵While severity- k dependence is interesting in its own right, in this paper we focus mainly on system dependence. We are pursuing severity- k dependence in a separate project.

and *severity-k MI (CTI)*.

Proposition 3.

$$\begin{aligned}
MI(\mathbf{p}^\alpha || \boldsymbol{\pi}^\alpha) &= MI(\tilde{\mathbf{p}}^\alpha || \tilde{\boldsymbol{\pi}}^\alpha) + \sum_{k=0}^n \tilde{p}_k^\alpha MI(\mathbf{p}^{\alpha,k} || \boldsymbol{\pi}^{\alpha,k}), \\
\kappa(\mathbf{p}^\alpha) &= \tilde{\kappa}(\mathbf{p}^\alpha) + \sum_{k=0}^n \tilde{p}_k^\alpha \kappa^k(\mathbf{p}^\alpha), \\
0 &\leq \tilde{\kappa}(\mathbf{p}^\alpha) \leq \kappa(\mathbf{p}^\alpha) \leq 1,
\end{aligned} \tag{7}$$

with $\tilde{\kappa}(\mathbf{p}^\alpha) = \kappa(\mathbf{p}^\alpha) = 0$ in the case of tail independence and $\tilde{\kappa}(\mathbf{p}^\alpha) = \kappa(\mathbf{p}^\alpha) = 1$ for perfect dependence (i.e., when all exceedances always occur together).

An important corollary of the last proposition shows that the CTI and the system CTI coincide in the bivariate case. Our limit results in Proposition 2 apply, therefore, also to the system tail interdependence.

Corollary 1. If $n = 2$ then $\kappa(\mathbf{p}^\alpha) = \tilde{\kappa}(\mathbf{p}^\alpha)$.

As an example, consider a set of interdependent constituents of a financial or economic system. Then, the TIS \mathbf{p}^α , computed from some measure of performance of these institutions in the negative tails, contains all information about the simultaneous distress of the constituents of this system at the nominal severity level α . The system TIS $\tilde{\mathbf{p}}^\alpha$ of the total number of constituents in distress captures the systemically important risk of multiple extreme events. For example, if the entire system breaks down when k or more constituents are in distress, then the probability of the systemic failure is simply computed as $\tilde{p}_k^\alpha + \dots + \tilde{p}_n^\alpha$. Our measure of tail interdependence $\kappa(\mathbf{p}^\alpha)$ and, particularly, its system part $\tilde{\kappa}(\mathbf{p}^\alpha)$ capture this *systemic distress* in a simple normalized coefficient.

3.4. Dynamic tail interdependence structure

The discussion above is clearly within a static framework. While this approach may be suitable in some applications, it clearly imposes limitations on the usefulness

of the technique, particularly in applications where the dynamics of the dependence are important such as the portfolio application in Section 6. Another drawback is that, like other static frameworks, it suffers from "ghost features" - the large impact of a very influential observation entering and leaving the estimation window. As it is now standard in the literature, this problem can be addressed by weighting the observations so that the most recent observations are more informative about the dependence in the subsequent periods than the observations in the past. For example, we could model the dynamics of the system TIS $\tilde{\mathbf{p}}_{t+1}^\alpha$ at date $t + 1$ by an exponentially weighted moving average (EWMA) process with the parameter $\delta \in [0, 1]$,

$$\tilde{\mathbf{p}}_{t+1}^\alpha = \delta \tilde{\mathbf{p}}_t^\alpha + (1 - \delta) \mathbf{u}_t, \quad (8)$$

where \mathbf{u}_t is a discrete $(n+1)$ -dimensional pdf that puts all probability on the observed number of exceedances in period t and $\tilde{\mathbf{p}}_0^\alpha$ is the sample TIS. This process reflects the autoregressive structure of the number of exceedances. In the empirical section, we estimate (8) by maximum likelihood and apply it to compute the daily forecasts of the system TIS and CTI.

The EWMA model remedies the "ghost features" problem of the static framework and therefore, it is better suited to capturing persistence in risk measures. However, while the EWMA is a relatively simple technique to the weighting of past observations it is not the only approach and we employ it here simply as a first approximation. An alternative to EWMA would be a GARCH-like model. However, EWMA-based tail forecasts have been shown to be superior to those based on GARCH models in many cases (see, for instance, Alexander and Leigh, 1997; Boudoukh et al., 1997; Guermat and Harris, 2002).

4. Hypotheses testing

The (system) TIS/CTI can be directly applied to the testing of statistical hypotheses on dependence. It offers, therefore, a cohesive framework where dependence measures, tests as well as other applications are based on the same fundamental concept of multi-information.

The statistical procedures that we present below are based on the asymptotic likelihood ratio test and are valid, therefore, for sufficiently large samples. Alternatively, we can test them in smaller samples with a generalized version of the Fisher's exact test (Mehta and Hilton, 1993) which we do not discuss here.

4.1. Goodness-of-fit and independence tests

Recall that \mathcal{T}^α is a partition of the sample space of the n -dimensional random vector $\mathbf{X} = (X_1, \dots, X_n)$ into 2^n joint tails and that the TIS \mathbf{p}^α is a pdf over \mathcal{T}^α . Given a sample $\{\mathbf{x}_t\}_{t=1}^T$, an empirical TIS $\hat{\mathbf{p}}^\alpha = (\hat{p}_C^\alpha)_{C \subseteq \mathcal{N}}$ contains the relative frequencies of observations that fall into the JTs $T_C^\alpha \in \mathcal{T}^\alpha$. We use $\hat{\mathbf{p}}^\alpha$ to test whether the observed interdependence structure comes from a hypothesized pdf f , which produces the TIS \mathbf{p}^α . For this purpose, we compute the KL divergence $MI(\hat{\mathbf{p}}^\alpha || \mathbf{p}^\alpha)$,⁶

$$MI(\hat{\mathbf{p}}^\alpha || \mathbf{p}^\alpha) = \sum_{C \subseteq \mathcal{N}} \hat{p}_C^\alpha \ln \frac{\hat{p}_C^\alpha}{p_C^\alpha}. \quad (9)$$

If exceedances are mutually independent under f , this procedure boils down to a test of tail independence. In the latter case, the hypothesized TIS is $\boldsymbol{\pi}^\alpha$ and (9) is proportional to the CTI (5),

$$MI(\hat{\mathbf{p}}^\alpha || \boldsymbol{\pi}^\alpha) = (1 - n) \ln(\alpha^\alpha (1 - \alpha)^{1-\alpha}) \cdot \kappa(\hat{\mathbf{p}}^\alpha). \quad (10)$$

⁶The goodness-of-fit and the interdependence symmetry test below can be conducted only when the test statistic is well-defined, i.e., when all denominators in (9) are strictly positive.

Our goodness-of-fit test with the mutual independence test as a special case, is conditional on sufficient statistics estimated from the data (e.g. on the estimates of quantiles in the sample). For the conditional test, the asymptotic distribution of the test statistic $2 \cdot T \cdot MI(\widehat{\mathbf{p}}^\alpha || \mathbf{p}^\alpha)$, where T is the sample size, follows the χ^2 -distribution with df degrees of freedom (e.g. McCullagh, 1986). For the degrees of freedom, we observe that we have 2^n outcomes (JTs) and $n + 1$ restrictions on probabilities or frequencies of these outcomes: these probabilities sum up to one and, moreover,

$$\sum_{C \subseteq \mathcal{N}: i \in C} p_C^\alpha = \sum_{C \subseteq \mathcal{N}: i \in C} \widehat{p}_C^\alpha = \alpha, \quad \forall i = 1, \dots, n.$$

Therefore, we apply $df = 2^n - n - 1$ degrees of freedom in our goodness-of-fit tests.

Alternatively, we can use the observed system TIS $\widehat{\mathbf{p}}^\alpha$ and the theoretical system TIS $\widetilde{\mathbf{p}}^\alpha$ to compute the KL divergence $D(\widehat{\mathbf{p}}^\alpha || \widetilde{\mathbf{p}}^\alpha)$. In this case, $2 \cdot T \cdot D(\widehat{\mathbf{p}}^\alpha || \widetilde{\mathbf{p}}^\alpha)$ is distributed approximately as χ^2 with $df = n - 1$ degrees of freedom as there are $n + 1$ outcomes and two restrictions on probabilities of these outcomes,

$$\sum_{k=0}^n \widehat{p}_k^\alpha = 1, \quad \text{and} \quad \sum_{k=0}^n k \widehat{p}_k^\alpha = n\alpha.$$

4.2. Interdependence symmetry test

Another interesting question is whether two tail interdependence structures along any two directional vectors \mathbf{d}^+ and \mathbf{d}^- (e.g. negative and positive tails) are symmetric. Specifically, let $\widehat{\mathbf{p}}^{\alpha+} = (\widehat{p}_C^{\alpha+})_{C \subseteq \mathcal{N}}$ and $\widehat{\mathbf{p}}^{\alpha-} = (\widehat{p}_C^{\alpha-})_{C \subseteq \mathcal{N}}$ be two empirical TISs computed for \mathbf{d}^+ and \mathbf{d}^- , respectively. Our objective is to test whether $\widehat{\mathbf{p}}^{\alpha+}$ and $\widehat{\mathbf{p}}^{\alpha-}$ are generated by a process with an identical tail interdependence structure. In order to test the null $\mathbf{p}^{\alpha+} = \mathbf{p}^{\alpha-}$, we apply the Kullback–Leibler test statistic,

$$KL^\pm = \sum_{C \subseteq \mathcal{N}} T^+ \widehat{p}_k^{\alpha+} \ln \frac{\widehat{p}_C^{\alpha+}}{\widehat{p}_C^\alpha} + \sum_{C \subseteq \mathcal{N}} T^- \widehat{p}_k^{\alpha-} \ln \frac{\widehat{p}_C^{\alpha-}}{\widehat{p}_C^\alpha},$$

where,

$$\widehat{p}_C^\alpha = \frac{(T^+ \widehat{p}_C^{\alpha+} + T^- \widehat{p}_C^{\alpha-})}{T^+ + T^-},$$

and T^+ (T^-) is the size of the sample from which $\widehat{\mathbf{p}}^{\alpha+}$ ($\widehat{\mathbf{p}}^{\alpha-}$) have been computed. The asymptotic distribution of $2 \cdot KL^\pm$ follows the χ^2 -distribution with $2^n - 1$ degrees of freedom (e.g. Quine and Robinson, 1985). We refer to this procedure as the interdependence symmetry test. Alternatively, the statistic KL^\pm can be computed from the system TIS, in which case $2 \cdot KL^\pm$ follows the χ^2 -distribution with n degrees of freedom.

5. Application: constituents of the S&P250 index

We apply our tail interdependence framework to the daily returns of the constituents of a major U.S. index. We focus on the 250 constituents of the Standard & Poor 250 index (S&P250) due to their importance for the U.S. and global economy. Prices and capitalization of each constituent have been obtained from Datastream between 02 January 1990 and 29 December 2017 (7,057 observations). The data has mean returns close to zero and displays volatility clustering consistent with the literature. We use a number of return series for the analysis ranging from raw returns to AR-GARCH-standardised returns to account for heteroskedasticity which can distort dependence measures, as well as the residuals of the regression of raw returns on the Fama-French-Carhart factors. The latter factors are the dominant pricing factors in the literature but little is known about their mutual relationship or ability to account for tail risk.

In all statistical tests that follow, we say that the null is strongly rejected if the p-value of the relevant test does not exceed 0.01. A simple rejection occurs with a p-value below 0.1. If we (do not) reject the null for any tail probability α , this implies that we tested the null for the nominal levels of extremity $\alpha \in \{0.1, 0.15, \dots, 0.9\}$.

5.1. Daily returns of S&P250 constituents

We investigate the interdependence of the constituents of the S&P250 index. We compute the daily returns between 02 January 1990 and 29 December 2017 ($T = 7,056$ synchronized observations). Summary statistics for the 250 constituents, not presented here, are in line with established evidence and show that the returns are leptokurtic and negatively skewed.

Since we have 2^{250} joint tails but only 7,056 observations we resort to system CTI $\tilde{\kappa}(\alpha)$ which requires only 251 empirical probabilities estimated from the sample. Figure 2 shows system CTI $\tilde{\kappa}(\alpha)$ computed in the negative and the positive JTs for the empirical distribution of the raw daily returns and of AR(1)-GARCH(1,1)-standardised daily returns. The values of $\tilde{\kappa}(\alpha)$ for $\alpha \in [0.1, 0.5]$ correspond to the negative joint tails $\mathcal{T}^\alpha(-\mathbf{1})$ and for $\alpha \in (0.5, 0.9]$ to the positive joint tails $\mathcal{T}^{1-\alpha}(\mathbf{1})$. Focusing first on the dependence in raw returns, there is an asymmetry between the negative and the positive tails in the sample: the interdependence in the negative tails is higher relative to the positive tails. Considering tail interdependence in the negative tails as a manifestation of system distress, we note that the latter is higher at more extreme levels, i.e., the risk of system breakdown increases in the severity of the distress. In other words, the more extreme the scenario, the more likely the system breakdown. Our interdependence symmetry test strongly rejects the null of the same interdependence structure for the system CTI for $\alpha \in (0, 0.3)$ and $\alpha \in (0.7, 1)$. Therefore, negative extreme returns appear indeed more closely tied together than their positive counterparts.

[Figure 2]

While the interdependence of extreme equity raw returns is strongly asymmetric (see also, Longin and Solnik, 2001; Patton, 2009), the literature suggests this could be driven by GARCH effects. Therefore, we repeat the analysis employing AR(1)-

GARCH(1,1)-standardised daily returns. The results suggest that GARCH effects do account for a considerable part of dependence although this holds only for dependence in the tails (low and high α) but not in the central part of the distribution. Indeed for $\alpha \in [0.3, 0.7]$, the dependence is almost identical to that of the raw returns. However, for $\alpha \notin [0.3, 0.7]$ the dependence of GARCH-standardised returns is considerably lower. While this dependence appears to increase slightly in α in the negative tails, the opposite happens in the positive tails. Our symmetry test rejects the null of identical interdependence structure in these tails.

5.2. Integration of S&P250 companies

In this subsection, we address questions pertaining to the integration of S&P250 companies by examining the evolution of their tail interdependence over time. Although these companies are headquartered in the U.S., they are truly global and operate in markedly different industries such as materials, healthcare and retail. In light of the findings above and to be consistent with the literature, we conduct this analysis employing AR(1)-GARCH(1,1)-standardised returns. Figure 3 shows the system CTI $\tilde{\kappa}(\alpha)$ computed by (8) in the negative tail ($\alpha = 0.1$), the central part of the distribution ($\alpha = 0.5$) and positive tails ($\alpha = 0.9$)

[Figure 3]

The dynamics of the interdependence in the central part of the distribution and positive tails largely have the same pattern until around 2005. However, since then, a noticeable dependence asymmetry between the returns in the central part and the positive tails of the distribution has developed, where in particular the dependence is higher in the positive tails than the central part of the distribution. The dependence in the negative tail has always been higher than in other parts of the support of the distribution and this discrepancy has increased since 2007, which coincides with the beginning of the financial crisis. In particular, it is clear from Figure 3 that the gap

between $\tilde{\kappa}(0.1)$ and $\tilde{\kappa}(0.5)$ has increased markedly since then (although it seems to be narrowing recently). Therefore, there are strong asymmetries among the negative tail, where systemic distress materializes, the central part of the distribution which reflects small day-to-day price moves, and positive tails capturing extreme gains. In particular, the negative tail displays the strongest dependence, followed by the positive tail. This is a confirmation of the rotated-J shape dependence depicted in Figure 2. Further, while dependence increased steadily until the financial crisis, it has started decreasing although it still remains historically high for all three parts of the distribution.

5.3. Contribution to tail dependence

It is important for the study of spillovers and contagion to isolate the impact or contribution of an individual component to the overall systemic distress. As the (system) CTI can be computed for different subsets of variables, we can find the marginal contribution of each variable to the interdependence in the subsets of other variables and, then, apply the Shapley Value to decompose the CTI into individual contributions (see, e.g. Tarashev et al., 2016). Here, however, we simply compute the contribution of a single variable as the difference of the system CTIs that include and exclude that particular variable.

Figure 4 depicts the contributions (sorted in decreasing order) of the constituents of the S&P250 index to the system interdependence in this index. The contribution of stock i is computed as the difference $\widetilde{MI}(\alpha) - \widetilde{MI}_{-i}(\alpha)$, where $\widetilde{MI}(\alpha)$ is the system multi-information computed for all 250 stocks in the index after controlling for conditional volatility and $\widetilde{MI}_{-i}(\alpha)$ excludes stock $i \in \{1, 2, \dots, 250\}$. We apply the system MI in this subsection in order to avoid the confounding effect of the normalization factor used in the computation of the system CTI. The latter factor depends on n (the dimension of the random vector) and would be different for $\widetilde{MI}(\alpha)$ and for $\widetilde{MI}_{-i}(\alpha)$.

Interestingly, we observe for each $\alpha \in \{0.1, 0.5, 0.9\}$ a distinct group of ca. 30 firms that make the largest positive contributions and a group of roughly the same size making negative contributions to the system MI (except for $\alpha = 0.5$ where the latter group contains ca. 50 companies). In particular, the five largest positive contributors turn out to be (electric) utility companies (Ameren Corporation, American Electric Power, Consolidated Edison, Pinnacle West Capital Corporation and Xcel Energy), while the largest negative contributions are Apple, Cisco Systems, Microsoft, Oracle and Intel. This finding suggests that the dependence of the S&P250 index constituents is mainly due to Old Economy utility stocks. The leaders of the New Economy, on the other hand, display little tail interdependence with the Old Economy as manifested by their relatively large negative impact on the system MI.

[Figure 4]

5.4. *Stock and factor interdependence*

In this subsection, we focus on the interdependence of the 250 constituents of the S&P250 index and relate it to the Fama-French-Carhart (FFC) factors. The FFC factors for the period 02 January 1990 and 29 December 2017 are obtained from Kenneth French's website. Summary statistics are reported in Table 1.

[Table 1]

For all four factors, daily log returns are zero, negatively skewed and leptokurtic. As observed above and shown in the right panel of Figure 5, the S&P250 constituents' raw returns are highly interdependent and asymmetric, while AR(1)-GARCH(1,1)-standardised returns remain strongly dependent and display dependence asymmetry, albeit lower than for raw returns.

Considering the residuals of a regression of the S&P250 constituents' returns on the returns of the FFC factors, it appears that the latter account for a high degree

of this dependence in the central part of the distribution but they are unable to account for the strong system dependence of the S&P250 returns in the tails of the distribution. Moreover, although not shown, most of the dependence in the residuals is accounted for by the first factor (the market return) while the other factors account for very little of the interdependence.

The inability of the FFC factors to account for the interdependence of the S&P250 returns in the tails is a direct manifestation of the tail interdependence of the factors themselves (see also Christoffersen and Langlois, 2013). The system CTI depicted in the left panel of Figure 5 reveals that the FFC factors are interdependent for $\alpha < 0.25$ and $\alpha > 0.8$. Standardising the regression residuals with a GARCH(1,1) decreases further the dependence but this still remains statistically significant.

[Figure 5]

6. CTI and portfolio allocation

CTI can be useful in a number of applications, including hedging and risk management, multidimensional option pricing and system risk measurement. An interesting application is also portfolio allocation.⁷ Since CTI is a measure of dependence with a particular focus on the tails in high dimensions, it can be directly applied to tail risk diversification. Intuitively, for the CTI to add value in this context, the joint returns of the underlying assets have to be non-normal. Under such a scenario, as is indeed the case in reality, the portfolio return will depend not only on higher moments, which have been shown elsewhere to have a considerable impact on the optimal portfolio weights (Ghysels et al., 2016; Harvey et al., 2010; Guidolin and Timmerman, 2008), but also on higher co-moments (see, for example, Martellini and

⁷We would like to thank the editor Eric Ghysels for suggesting this analysis.

Zieman, 2009).

Intuition suggests that CTI, as a measure of dependence in the tails, is related to higher co-moments and, in particular, co-skewness which captures the asymmetry of the joint distribution of returns of the underlying assets. If an asset has a high coefficient of co-skewness with some reference asset such as the market index, then high returns on the asset are associated with very high and very low returns on the market. This leads to more probability mass in extreme negative and in extreme positive tails. Moreover, a joint distribution that displays such characteristics would also feature asymmetric tail dependence. Indeed, it is this intuition that Bücher et al. (2017) exploit to construct an asymmetric dependence copula function from co-skewness which they then use to model data with complex relationships from hydrology and finance. Further, Cerrato et al. (2017) investigate the relationship between higher-order co-moments and the dependence structure of equity portfolios in the U.S. and U.K. and find that co-skewness is positively associated with tail dependence. Moreover, the asymmetric tail dependence parameters that they employ⁸ are the classic coefficients of tail dependence proposed by Sibuya (1960) and Joe (1989) to which our CTI converges, as discussed in Section 3.2 above. Therefore, it is no surprise we also find a positive relationship between tail dependence asymmetry and co-skewness as shown in Figure 6. This figure highlights that the stronger the asymmetry of the joint distribution of the S&P250 index constituents with the market and therefore, the larger the (negative value of) their co-skewness coefficient, then the larger the asymmetry of the dependence on the market in the lower and upper tails.

[Figure 6]

Having confirmed that the relationship between tail dependence asymmetry and

⁸See equations A.12 and A13 in their Appendix A3.

co-skewness found in the literature is strong in the S&P250 index constituents that we employ in this paper, we examine whether these measures help improve portfolio performance. Since we include all 250 index constituents in our portfolio, it is too demanding to model their conditional joint density from which to obtain the optimal weights. Therefore, we follow the parametric framework of Ghysels et al. (2016). To obtain the optimal portfolio weights, they maximize the power utility function by changing the weights which are a function of some benchmark portfolio weights and other asset-specific characteristics such as the skewness of the assets. We adopt the Ghysels et al. (2016) approach but instead of the conditional skewness, we examine the impact of the conditional co-skewness of the assets with the market on portfolio weights. Further, we also examine whether the conditional tail dependence of the assets with the market can improve the portfolio performance. Specifically, we construct two dynamic portfolios spanning the $n = 250$ constituents of the S&P250 index where the weights depend on conditional co-skewness and conditional CTI difference respectively, as follows

$$w_{i,t-1}^{CSK} = \frac{1}{n} + \frac{\lambda}{n} \widetilde{CSK}_{t-1}(r_{it}, r_{mt}), \quad w_{i,t-1}^{CTI} = \frac{1}{n} + \frac{\lambda}{n} \widetilde{\Delta}_{t-1}(r_{it}, r_{mt}), \quad i = 1, \dots, n, \quad (11)$$

where λ is the parameter to estimate, $\widetilde{CSK}_{t-1}(r_{it}, r_{mt})$ is the estimate at time $t - 1$ of the conditional co-skewness of asset i return r_{it} with the market return r_{mt} at time t and $\widetilde{\Delta}_{t-1}(r_{it}, r_{mt})$ is the estimate at time $t - 1$ of the difference between the conditional CTI in the upper tail ($\alpha = 0.9$) and in the lower tail ($\alpha = 0.1$) for the same r_{it} and r_{mt} returns. As in Ghysels et al. (2016), in expression (11), the “tilde” indicates that the variables are standardized to have mean zero and unit variance at every point in time. This standardization ensures that the relative contribution of a predictor becomes clear as the magnitudes of the coefficients are directly comparable.

Moreover, demeaning the predictors ensures that

$$\sum_{i=1}^n w_{i,t-1}^{CSK} = \sum_{i=1}^n w_{i,t-1}^{CTI} = \sum_{i=1}^n \frac{1}{n} = 1.$$

These weights then have an intuitive interpretation. For example, the estimated $w_{i,t-1}^{CTI}$ can be interpreted as the “tail-dependence-managed” allocation in a long-short portfolio that tilts the optimal investment on asset i toward or away from the equally-weighted allocation, depending on that asset’s $\tilde{\Delta}_{t-1}(r_{it}, r_{mt})$ relative to the cross-sectional mean. For instance, if the investor prefers assets with a larger positive than negative tail-dependence on the market (i.e., λ is positive), then assets with higher $\tilde{\Delta}_{t-1}(r_{it}, r_{mt})$ will have a relatively higher portfolio weight. A similar argument applies to $w_{i,t-1}^{CSK}$ and $\widetilde{CSK}_{t-1}(r_{it}, r_{mt})$.

When $\lambda = 0$ the portfolios defined in (11) boil down to an equally weighted portfolio. The latter has shown good performance in practice. Indeed, DeMiguel et al. (2009) find that of the 14 more sophisticated models evaluated across seven empirical datasets, none is consistently better than the equally-weighted portfolio in terms of Sharpe ratio, certainty-equivalent return or turnover. Table 2 shows the results for portfolio p with returns

$$r_{pt} = \sum_{i=1}^n w_{i,t-1} \cdot r_{it}$$

where $w_{i,t-1} = w_{i,t-1}^{CSK}$ or $w_{i,t-1} = w_{i,t-1}^{CTI}$, that solves,

$$\max_{\lambda} \frac{1}{T} \sum_{t=1}^{T-1} \frac{(1 + r_{pt})^{1-\gamma}}{1 - \gamma} \tag{12}$$

subject to the relevant portfolio weights specification in expression (11). As in Ghysels et al. (2016), we set the relative risk aversion γ to five, but our main conclusions hold for a relatively wide range of risk aversion parameters. Conditional co-skewness

and conditional CTI are estimated by EWMA (8) for each asset i and date t . For comparison, we report also the statistics for the market (M), equally-weighted (EW) and value-weighted (VW) portfolios.

Table 2 reveals that conditional co-skewness and conditional CTI indeed improve portfolio performance as compared to the standard benchmarks. The conditional co-skewness portfolio has the highest annual expected return (18.44%) among the portfolios examined and also higher utility (-0.2418) and certainty equivalent (10.43%) than the benchmarks. The conditional CTI portfolio achieves comparably high utility (-0.2416) and certainty equivalent (10.82%) but at the cost of a lower annual return (16.21%). The latter finding is unsurprising given that the CTI portfolio, by construction, hedges against extreme tail events. In particular, it scales down the weight of an asset that often co-exceeds with the market in the negative tails but not in the positive. The hedging character of the conditional CTI portfolio is confirmed by its standard deviation (σ_p), which at 13.27% is much lower than the standard deviations of its competitors, in particular that of the conditional co-skewness portfolio (17.32%). Overall, the increase in the certainty equivalent is large with respect to the market benchmark: 9.22% for the conditional co-skewness portfolio (from 1.21% to 10.43%) and 9.61% (from 1.21% to 10.82%) for the conditional CTI portfolio. The increase is still significant with respect to the equally-weighted and the value-weighted portfolios. Regarding optimal portfolios, the second row in Table 2 indicates that they are heavily exposed to co-skewness and to CTI difference: the (relative) weight of \widetilde{CSK} and $\widetilde{\Delta}_{t-1}$ in (11) is 61.97% and 66.78%, respectively. Table 2 shows also that the conditional co-skewness (conditional CTI) portfolio features 9.12% (7.06%) annualized alpha with respect to the market portfolio ($\bar{r}_p - \bar{r}_m$), high correlation (ρ_{pm}) and a negative co-skewness (CSK_{pm}) with this benchmark. Interestingly, whereas the skewness (SK_p) of the \widetilde{CSK} portfolio is positive, it is negative for the $\widetilde{\Delta}_{t-1}$ portfolio.

[Table 2]

7. Conclusion

Tail interdependence captures the likelihood of the constituents of a system to get into distress or exuberance (or other, more complex extreme scenarios) simultaneously. However, in practice, its measurement is challenging due, among other reasons, to the small number of tail observations and the curse of dimensionality. We propose a versatile non-parametric framework to analyze and quantify tail dependence with several extensions. We apply it to data from the constituents of the S&P250 index and report some intriguing findings. For example, we find that while the Fama-French-Carhart factors do well to account for interdependence in the central part of the distribution, they do not in the tails of the distribution because they are themselves tail dependent. We also find that the interdependence of extreme equity returns is strongly asymmetric. In an application to portfolio allocation, we show that (conditional) CTI improves portfolio performance and hedges against extreme tail events. Interesting avenues for future research would be a rigorous examination of the dynamic properties of CTI and an examination of the dynamic relationship between the (system) CTI of main asset classes or financial institutions with macroeconomic indicators with MIDAS technique along the lines of Engle et al (2008).

References

Alexander, C. O. and Leigh, C. T. 1997, On the covariance matrices used in value at risk models, *Journal of Derivatives*, 4, 50-62.

Andrade, P., Ghysels, E. and Idier, J. 2013, Tails of Inflation Forecasts and Tales of Monetary Policy, UNC Kenan-Flagler Research Paper No. 2013-17 available at <https://ssrn.com/abstract=2169976>

Asimit, A.V. Gerrard, R. Hou, Y. and Peng, L. 2016, Tail dependence measure for examining financial extreme, co-movements, *Journal of Econometrics*, 194, 330-348.

Backus, D. Chernov, M. and Zin, S. 2014, Sources of entropy in representative agent models, *Journal of Finance*, 69, 51-99.

Bae, K.H., Karolyi, G.A. and Stulz, R. 2003, A new approach to measuring financial contagion, *Review of Financial Studies* 16, 717–763.

Beine, M. Cosma, A. and Vermeulean, R. 2015, The dark side of global integration: Increasing tail dependence, *Journal of Banking and Finance* 34, 184-192.

Boudoukh, J., Richardson, M. and Whitelaw, R. F. 1997, Investigation of a class of volatility estimators, *Journal of Derivatives*, 4, 63-71.

Boyson, N.M., Stahel, C.W. and Stulz, R. 2010, Hedge Fund Contagion and Liquidity Shocks, *Journal of Finance*, 65, 1789-1816.

Bücher, A. Jäschke, S. and Wied, D. 2015, Nonparametric tests for constant tail dependence with an application to energy and finance, *Journal of Econometrics*, 187, 154-168.

Bücher, A., Irresberger, F. and Weiss, G.N.F. 2017, Testing Asymmetry in Dependence with Copula-Coskewness, *North American Actuarial Journal*, 21, 267-280.

Cameron, A. and Windmeijer, F. 1997, An R-squared measure of goodness of fit for some common nonlinear regression models, *Journal of Econometrics*, 77, 329-342.

Campbell, H.R., Liechty, J.C., Liechty, M.W. and Muller, P. 2010, Portfolio selection with higher moments, *Quantitative Finance* 10, 469–485.

Cerrato, M., Crosby, J., Kim, M. and Zhao, Y. 2017, Relation between higher order comoments and dependence structure of equity portfolio, *Journal of Empirical Finance* 40, 101–120.

Chicharro, D., Ledberg, A., 2012, Framework to study dynamic dependencies in networks of interacting processes, *Physical Review E*, 86(4).

Christoffersen, P. and Langlois, H. 2013, The Joint Dynamics of Equity Market

Factors, *Journal of Financial and Quantitative Analysis* 48, 1371–1404.

Colangelo, A., Scarsini, M. and Shaked, M., 2005, Some notions of multivariate positive dependence, *Insurance: Mathematics and Economics*, 37, 13 - 26.

Cover, T.M., Thomas, J.A., 2006, *Elements of Information Theory*, 2nd ed., Wiley & Sons.

DeMiguel, V., Garlappi, L. and Uppal, R. 2009, Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *Review of Financial Studies* 22 1915-1953.

Dungey, M. and Renault, E. 2018, Identifying contagion, *Journal of Applied Econometrics*, 33, 227-250.

Dungey, M., Jacobs, J.P.A.M., and Lestano, 2015, The internationalisation of financial crises: Banking and currency crises 1883–2008, *North American Journal of Economics and Finance* 32, 29–47.

Engle, R., Ghysels, R. and Sohn, B. 2008, On the Economic Sources of Stock Market Volatility, NYU Working Paper No. FIN-08-043.

Ghysels, E. Plazzi, A. and Valkanov, R. 2016, Why Invest in Emerging Markets? The Role of Conditional Return Asymmetry, *Journal of Finance*, 71, 4145-4192.

Guermat, C. and Harris, R. D.F. 2002, Robust conditional variance estimation and value-at-risk, *Journal of Risk*, 4, 25-42.

Guidolin, M. and Timmermann, A. 2008, International asset allocation under regime switching, skew, and kurtosis preferences, *Review of Financial Studies* 21, 889–935.

Heffernan, J.E. 2001, A Directory of Coefficients of Tail Dependence, *Extremes*, 3, 279-290.

Jansen, D.W., de Vries, C.G. 1991, On the Frequency of Large Stock Returns: Putting Booms and Busts into Perspective, *Review of Economics and Statistics*, 73, 18-24.

Joe, H., 1989, Relative Entropy Measures of Multivariate Dependence, *Journal of the American Statistical Association*, 84, 157-164.

Joe, H. 1997, *Multivariate Models and Dependence Concepts*, Chapman and Hall, London.

Li, H., 2009, Orthant tail dependence of multivariate extreme value distributions, *Journal of Multivariate Analysis* 100, 243—256.

Longin, F. and Solnik, B. 2001, Extreme Correlation of International Equity Markets, *Journal of Finance*, 562, 649-676.

Lucas, A. Schwaab, B. and Zhang, X. 2016, Modeling Financial Sector Joint Tail Risk in the Euro Area, *Journal of Applied Econometrics*, 32, 171-191.

Martellini, L. and Ziemann, V. 2010, Improved Estimates of Higher-Order Comoments and Implications for Portfolio Selection, *The Review of Financial Studies* 23, 467–1502.

McCloud, N. and Hong, Y. 2011, Testing the Structure of Conditional Correlations in Multivariate GARCH Models: A Generalized Cross-Spectrum Approach, *International Economic Review*, 52, 991–1037.

McCullagh, P., 1986, The Conditional Distribution of Goodness-of-Fit Statistics for Discrete Data, *Journal of American Statistical Association*, 81, 104-107.

Mehta, C.R. and Hilton, J.F. 1993, Exact power of conditional and unconditional tests: Going beyond the 2 times 2 contingency tables. *The American Statistician*, 47, 91-98.

Nelsen, R.B. 2007, *An Introduction to Copulas*, Springer Series in Statistics.

Oh, D. H. and Patton, A.J. 2017, Modelling Dependence in High Dimensions with Factor Copulas, *Journal of Business and Economic Statistics*, 35, 139-154.

Opitz, S. Seidel, H. and Szimayer, A. 2017, Measuring crisis risk using conditional copulas: An empirical analysis of the 2008 shipping crisis, *Journal of Applied Econometrics*, 33, 271-289.

Patton, A.J. 2006, Modelling Asymmetric Exchange Rate Dependence, *International Economic Review*, 47, 527-556.

Patton, A.J., 2009, Are Market Neutral Hedge Funds Really Market Neutral?, *Review of Financial Studies* 227, 2495-2530.

Quine, M.P. and Robinson, J. 1985, Efficiencies of chi-square and likelihood ratio goodness-of-fit tests. *Annals of Statistics* 13, 727–742.

Schneidman, E., Still, S., Berry, M. and Bialek, W. 2003, Network Information and Connected Correlations, *Physical Review Letters* 91(23).

Segoviano, M. A. and Goodhart, C.A.E. 2009, Banking stability measures. Number 627. *International Monetary Fund*.

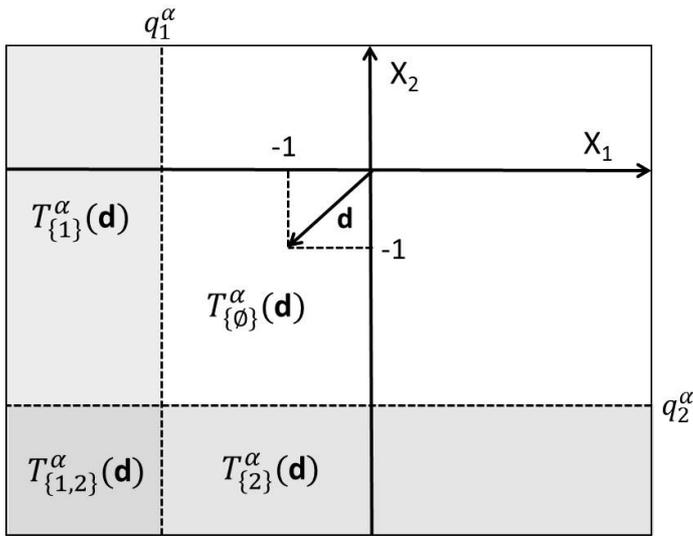
Sibuya, M., 1960, Bivariate extreme statistics, *Annals of the Institute of Statistical Mathematics*, 11, 195–210.

Tarashev, N., Tsatsaronis, K., Borio, C. 2016, Risk Attribution Using the Shapley Value: Methodology and Policy Applications, *Review of Finance* 20, 1189-1213.

Wennekers, T. and Ay, N., 2003, Spatial and temporal stochastic interaction in neuronal assemblies. *Theory in Biosciences*, 122, 5–18.

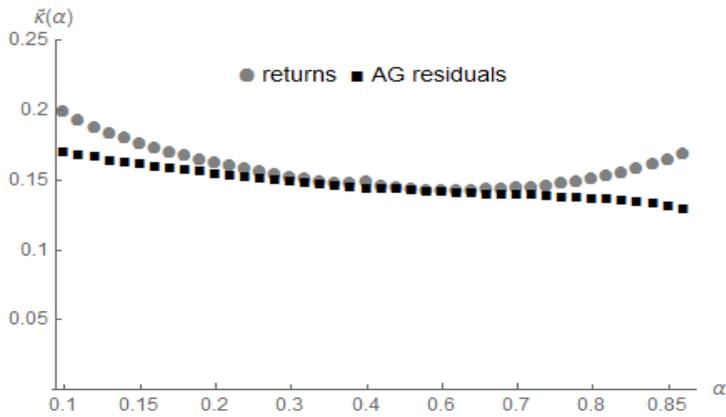
White, H. Kim, T-H. Manganelli, S. 2015, VAR for VaR: Measuring tail dependence using multivariate regression quantiles, *Journal of Econometrics* 187, 169-188.

Figure 1: The partition of the two-dimensional outcome space into joint tails



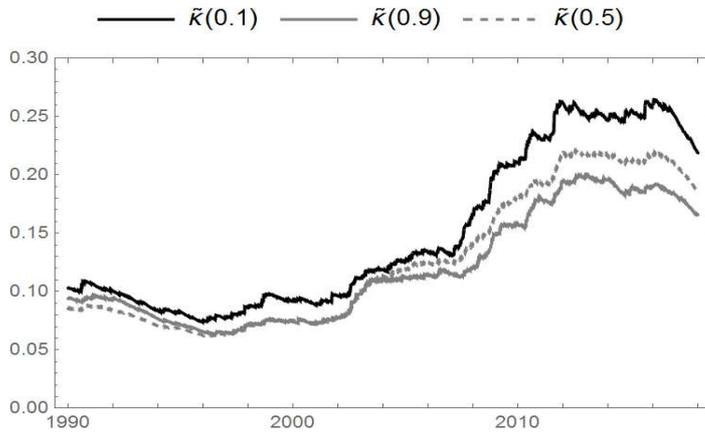
Notes: Partition of a two-dimensional outcome space into joint tails T^α for $d = -(1, 1)$, where $q_i^\alpha = F_i^{-1}(\alpha)$.

Figure 2: System CTI



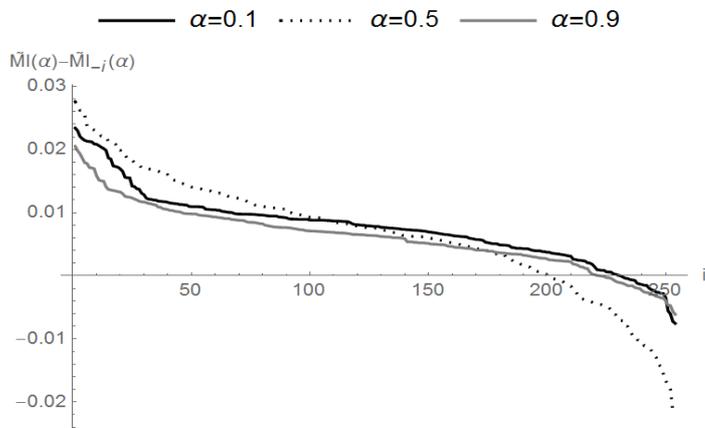
Notes: System CTI for the raw returns and the AR(1)-GARCH(1,1)-standardised returns of the constituents of the S&P250 index.

Figure 3: Evolution of the system CTI for S&P250 returns over time



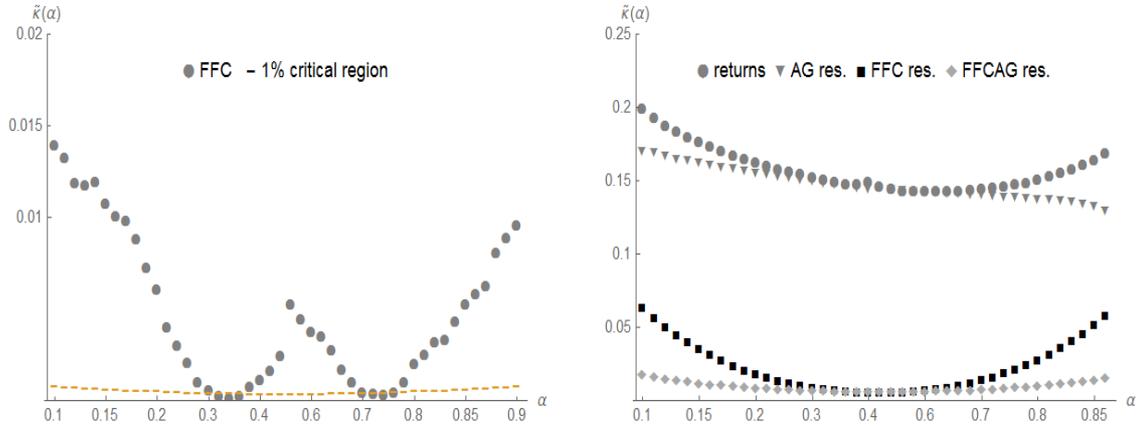
Notes: The evolution of the system CTI $\tilde{\kappa}(\alpha)$ for $\alpha = 0.1$, $\alpha = 0.5$ and $\alpha = 0.9$ computed by (8) with the estimated value of the parameter $\delta \approx 0.9975$ from the returns of S&P250 index constituent stocks.

Figure 4: Contributions (sorted in the decreasing order) to the system MI of the constituents of the S&P250 index



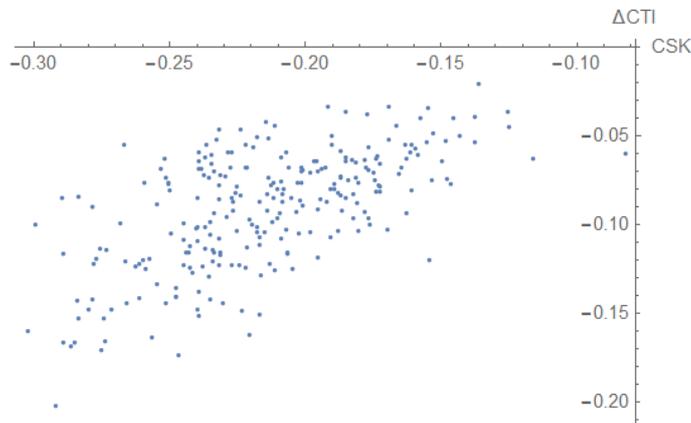
Notes: Contributions (sorted in the decreasing order) to the system MI of the constituents of the S&P250 index at the extremity levels $\alpha = 0.1$, $\alpha = 0.5$ and $\alpha = 0.9$.

Figure 5: Interdependence of Fama-French-Carhart factors and S&P250 index constituent stocks



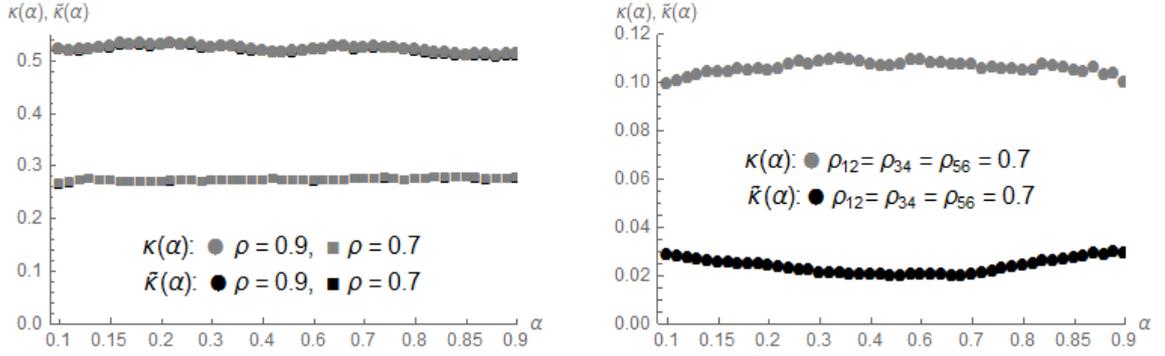
Notes: The left panel shows the system CTI for the Fama-French-Carhart (FFC) factors. The dashed line marks the 1% critical values for the test statistic (10) in the test of independence. The right panel shows the system CTI for the S&P250 index constituent returns as well as for the residuals of a regression of these constituents on the FFC factors.

Figure 6: Coskewness vs. CTI difference



Notes: Coskewness with the market vs. CTI difference with the market at $\alpha = 0.1$ (CTI in the upper JTs minus CTI in the lower JTs) for S&P250 returns between 2 January 1990 and 29 December 2017.

Figure 7: CTI vs system CTI



Notes: Total ($\kappa(\alpha)$) and system ($\tilde{\kappa}(\alpha)$) CTI computed from a sample of 10,000 observations from a standardized $n = 6$ dimensional multinormal X with $\text{corr}(X_i, X_k) = \rho$ for all $i, k = 1, \dots, 6$, $i \neq k$ (left panel) and $\text{corr}(X_1, X_2) = \text{corr}(X_3, X_4) = \text{corr}(X_5, X_6) = \rho$ and all other correlations equal to zero (right panel). The results for $\alpha \in [0.1, 0.5]$ correspond to the lower JTs $T^\alpha(-\mathbf{1})$ and for $\alpha \in (0.5, 0.9]$ to the upper JTs $T^{1-\alpha}(\mathbf{1})$.

Table 1: Summary Statistics for the Fama-French-Carhart Factor Returns

	RPm	SMB	HML	MOM
Mean	0.000	0.000	0.000	0.000
SD	0.012	0.006	0.006	0.009
Skewness	-0.105	-0.268	0.108	-0.956
Kurtosis	10.99	7.163	9.337	14.690

Notes: The table reports the mean, standard deviation, skewness, kurtosis for the Fama-French-Carhart factors Market Risk Premium (RPm), Small minus Big (SMB), High minus Low (HML) and Momentum (MOM). The data spans the period from 2 January 1990 through 29 December 2017 (7057 observations obtained from Kenneth French's website).

Table 2: Portfolios of S&P250 constituents

	M	EW	VW	\widetilde{CSK}	$\widetilde{\Delta}$
$\lambda/(1 + \lambda)$	–	–	–	0.6197	0.6678
\bar{r}_p	0.0860	0.1746	0.1663	0.1844	0.1621
σ_p	0.1587	0.1597	0.1542	0.1732	0.1327
SK_p	–0.8093	–0.5518	–0.3402	1.1846	–0.7145
$\rho_{p,m}$	1.0000	0.9195	0.9392	0.8852	0.7400
$CSK_{p,m}$	0.0000	–0.7004	–0.6470	–0.2009	–0.6379
$u(r_p, 5)$	–0.2490	–0.2427	–0.2427	–0.2418	–0.2416
$CE(u(r_p, 5))$	0.0121	0.0925	0.0925	0.1043	0.1082
$\bar{r}_p - \bar{r}_m$	0	0.0821	0.0744	0.0912	0.0706

Notes: Statistics for portfolio p computed from monthly S&P250 returns ($T = 300$)

between 02 January 1990 and 29 December 2017, where p is one of the following portfolios: M – market, EW – equally weighted, VW – value weighted, \widetilde{CSK} and $\widetilde{\Delta}$ defined in (11). $\lambda/(1 + \lambda)$ – relative weight of conditional co-skewness or CTI difference, where λ is obtained from (12) subject to (11), \bar{r}_p – annualized mean return, σ_p – annualized standard deviation, SK_p – skewness, $\rho_{p,m}$ – correlation with market, $CSK_{p,m}$ – co-skewness with market, $u(r_p, 5)$ – power utility of portfolio p for $\gamma = 5$, $CE(\cdot)$ – annualized certainty equivalent. Conditional co-skewness and conditional CTI difference for the last two portfolios are estimated by EWMA (8) with $\delta = 0.9975$ for each asset i and date t .

Table 3: System CTI with 95% confidence intervals (in parentheses)

multinormal	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.9$
$\rho = 0.0$	0.000 (.000, .001)	0.000 (.000, .011)	0.000 (.000, .011)
$\rho = 0.5$	0.112 (.096, .130)	0.129 (.119, .140)	0.119 (.102, .139)
$\rho = 0.9$	0.501 (.460, .547)	0.497 (.479, .515)	0.496 (.454, .541)
t-distribution			
$\rho = 0.0$	0.028 (.021, .036)	0.000 (.000, .001)	0.026 (.020, .034)
$\rho = 0.5$	0.189 (.166, .215)	0.139 (.128, .150)	0.174 (.153, .198)
$\rho = 0.9$	0.546 (.502, .594)	0.506 (.488, .523)	0.546 (.501, .594)
Clayton copula			
$\rho = 0.0$	0.000 (.000, .001)	0.000 (.000, .011)	0.000 (.000, .001)
$\rho = 0.5$	0.346 (.313, .382)	0.148 (.137, .160)	0.026 (.019, .033)
$\rho = 0.9$	0.812 (.757, .871)	0.656 (.638, .673)	0.278 (.250, .310)

Notes: System CTI with 95% confidence intervals in parentheses. Each sample con-

tains 10,000 vectors of 5-dimensional realizations generated from the multinormal, the multivariate t-distribution with 3 degree of freedom and a Clayton copula with normal marginal densities. In all models, the pairwise correlation is set (approximately) to $\rho \in \{0.0, 0.5, 0.9\}$ for each pair of variables. In the case of the Clayton copula, the correlations 0, 0.5 and 0.9 correspond to dependence coefficients of 0.001, 1.4 and 9, respectively.

Table 4: System CTI with 95% bootstrap CIs (in parentheses).

multinormal	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.9$
$\rho = 0.0$	0.000 (.000, 0.000)	0.000 (.000, .000)	0.000 (.000, .000)
$\rho = 0.5$	0.258 (.250, 0.266)	0.256 (.250, .262)	0.257 (.248, .262)
$\rho = 0.9$	0.657 (.646, 0.667)	0.644 (.637, .652)	0.657 (.646, .667)
t-distribution			
$\rho = 0.0$	0.103 (.099, 0.107)	0.000 (.000, .000)	0.105 (.101, .108)
$\rho = 0.5$	0.360 (.350, 0.370)	0.249 (.244, .254)	0.360 (.351, .370)
$\rho = 0.9$	0.709 (.698, 0.719)	0.650 (.643, .657)	0.708 (.696, .717)

Notes: System CTI with 95% confidence intervals in parentheses. Each sample contains 10,000 vectors of 100-dimensional realizations generated from the multinormal and the multivariate t-distribution with 3 degree of freedom. In all models, the pairwise correlation is set (approximately) to $\rho \in \{0.0, 0.5, 0.9\}$ for each pair of variables.

8. Appendix: Technical proofs

Proof of Proposition 1: First, we note that the r.h.s. of (5) implies that $\kappa(\mathbf{p}^\alpha) \geq 0$ as $MI(\mathbf{p}^\alpha || \boldsymbol{\pi}^\alpha) \geq 0$ (Cover and Thomas, 2006) and $(1-n) \ln(\alpha^\alpha(1-\alpha)^{1-\alpha}) > 0$ for $n \geq 2$ and $\alpha \in (0, 1)$.

In order to show that $\kappa(\mathbf{p}^\alpha) \leq 1$, we write the MI (4) as the difference between the sum of n identical marginal entropies and the joint entropy (e.g. Schneidman et al., 2003),

$$MI(\mathbf{p}^\alpha || \boldsymbol{\pi}^\alpha) = n \cdot E((\alpha, 1 - \alpha)) - E(\mathbf{p}^\alpha),$$

where the entropy of \mathbf{p}^α is defined as $E(\mathbf{p}^\alpha) = -\sum_{C \subseteq \mathcal{N}} p_C^\alpha \ln p_C^\alpha$ and each marginal distribution of exceedances is, by construction, equal to $(\alpha, 1 - \alpha)$. On the other hand, joint entropy cannot be smaller than marginal entropy (Cover and Thomas, 2006),

$$E(\mathbf{p}^\alpha) \geq E((\alpha, 1 - \alpha)),$$

which implies,

$$\begin{aligned} MI(\mathbf{p}^\alpha || \boldsymbol{\pi}^\alpha) &= n \cdot E((\alpha, 1 - \alpha)) - E(\mathbf{p}^\alpha) \leq \\ (n - 1)E((\alpha, 1 - \alpha)) &= -(n - 1) \ln(\alpha^\alpha (1 - \alpha)^{1-\alpha}). \end{aligned}$$

The last expression is the normalizing factor in (5). \blacksquare

Proof of Proposition 2: First, we express the system CTI for $n = 2$ as,

$$\kappa(\tilde{\mathbf{p}}^\alpha(f, \mathbf{d})) = \frac{1}{H(\alpha)} \sum_{k=0}^2 \tilde{p}_k^\alpha \ln \frac{\tilde{p}_k^\alpha}{\tilde{\pi}_k^\alpha} = \frac{1}{H(\alpha)} \left\{ \tilde{p}_0^\alpha \ln \frac{\tilde{p}_0^\alpha}{\tilde{\pi}_0^\alpha} + \alpha \lambda_1^\alpha \ln \frac{\alpha \lambda_1^\alpha}{\tilde{\pi}_1^\alpha} + \alpha \lambda_2^\alpha \ln \frac{\alpha \lambda_2^\alpha}{\tilde{\pi}_2^\alpha} \right\}, \quad (13)$$

where,

$$\begin{aligned} \lambda_k^\alpha &= \lambda_k^\alpha(\mathbf{d}) \equiv \tilde{p}_k^\alpha(f, \mathbf{d})/\alpha, \quad \mathbf{d} \in \{-1, 1\}^2, \\ \tilde{p}_0^\alpha &= 1 - \alpha \lambda_1^\alpha - \alpha \lambda_2^\alpha, \quad H(\alpha) \equiv -\ln(\alpha^\alpha (1 - \alpha)^{1-\alpha}). \end{aligned}$$

The following limits, obtained e.g. by the L'Hôpital's rule, will be used repeatedly in the proof,

$$\lim_{\alpha \rightarrow 0} \frac{\alpha}{H(\alpha)} = 0, \quad \lim_{\alpha \rightarrow 0} \frac{\alpha \ln(\alpha)}{-H(\alpha)} = 1.$$

We note that λ_k^α is a continuous function of α with a finite (right-hand) derivative due to our assumptions on the joint pdf and the marginal cdfs. Moreover, $\lambda_1^\alpha \leq 2$ and $\lambda_2^\alpha \leq 1$ (for example, $\lambda_1^\alpha = 2$ when $\alpha = 1/2$ and there is exactly one exceedance in each period, i.e., $\tilde{p}_1^\alpha = 2\alpha$). Hence, there is a unique limit $\lambda_k \equiv \lim_{\alpha \rightarrow 0} \lambda_k^\alpha \in [0, 2]$.

It follows then that the term,

$$\frac{\tilde{p}_0^\alpha \ln(\tilde{p}_0^\alpha / \tilde{\pi}_0^\alpha)}{H(\alpha)} = \left(1 - \alpha \sum_{k=1}^2 \lambda_k^\alpha\right) \left\{ \frac{\ln(1 - \alpha \lambda_1^\alpha - \alpha \lambda_2^\alpha)}{H(\alpha)} - \frac{\ln(1 - \alpha)^2}{H(\alpha)} \right\},$$

vanishes in the limit $\alpha \rightarrow 0$ because,

$$\begin{aligned}\lim_{\alpha \rightarrow 0} \frac{\ln(1 - \alpha\lambda_1^\alpha - \alpha\lambda_2^\alpha)}{H(\alpha)} &= \lim_{\alpha \rightarrow 0} \frac{\alpha(\lambda_1^\alpha + \lambda_2^\alpha)}{-H(\alpha)} = 0, \\ \lim_{\alpha \rightarrow 0} \frac{2\ln(1 - \alpha)}{-H(\alpha)} &= \lim_{\alpha \rightarrow 0} \frac{2\alpha}{H(\alpha)} = 0,\end{aligned}$$

where we used the Taylor expansion of \ln about 1. On the other hand, for each $k = 1, 2$,

$$\begin{aligned}\lim_{\alpha \rightarrow 0} \frac{\alpha \cdot \lambda_k^\alpha \ln(\alpha \cdot \lambda_k^\alpha / \tilde{\pi}_k^\alpha)}{H(\alpha)} &= \\ \lim_{\alpha \rightarrow 0} \frac{\lambda_k^\alpha \cdot \alpha \ln(\lambda_k^\alpha)}{H(\alpha)} - \lim_{\alpha \rightarrow 0} \lambda_k^\alpha \frac{\alpha \ln(\binom{2}{k} \alpha^{k-1} (1 - \alpha)^{2-k})}{H(\alpha)} &= \\ -\lambda_k \lim_{\alpha \rightarrow 0} \frac{\alpha \ln(\binom{2}{k} \alpha^{k-1} (1 - \alpha)^{2-k})}{H(\alpha)} &= \lambda_k(k - 1),\end{aligned}$$

for $2 \geq k \geq 1$. The last two equalities follow because,

$$\begin{aligned}\lim_{\alpha \rightarrow 0} \frac{\alpha \ln(\alpha^{k-1})}{-H(\alpha)} &= (k - 1) \lim_{\alpha \rightarrow 0} \frac{\alpha \ln(\alpha)}{-H(\alpha)} = k - 1, \\ \lim_{\alpha \rightarrow 0} \frac{\lambda_k^\alpha \cdot \alpha \ln(\lambda_k^\alpha)}{-H(\alpha)} &= \lim_{\alpha \rightarrow 0} \frac{\alpha}{-H(\alpha)} \lim_{\alpha \rightarrow 0} (\lambda_k^\alpha \cdot \ln(\lambda_k^\alpha)) = 0,\end{aligned}$$

as the last limit is bounded given that λ_k^α is bounded. We obtain, therefore,

$$\lim_{\alpha \rightarrow 0} \kappa(\tilde{\mathbf{p}}^\alpha(f, \mathbf{d})) = \lambda_2(\mathbf{d}),$$

with the special cases $\lambda_2((-1, -1)) = \lambda_L$ and $\lambda_2((1, 1)) = \lambda_U$. \blacksquare

Proof of Proposition 3: In order to prove (7), we first show the decomposition of the MI (4),

$$MI(\mathbf{p}^\alpha || \boldsymbol{\pi}^\alpha) = MI(\tilde{\mathbf{p}}^\alpha || \tilde{\boldsymbol{\pi}}^\alpha) + \sum_{k=0}^n \tilde{p}_k^\alpha MI(\mathbf{p}^{\alpha,k} || \boldsymbol{\pi}^{\alpha,k}). \quad (14)$$

The proof consists in expanding the definition of MI and rearranging terms.

$$\begin{aligned}
MI(\mathbf{p}^\alpha || \boldsymbol{\pi}^\alpha) - MI(\tilde{\mathbf{p}}^\alpha || \tilde{\boldsymbol{\pi}}^\alpha) &= \sum_{C \subseteq \mathcal{N}} p_C^\alpha \ln \frac{p_C^\alpha}{\pi_C^\alpha} - \sum_{k=0}^n \tilde{p}_k^\alpha \ln \frac{\tilde{p}_k^\alpha}{\tilde{\pi}_k^\alpha} = \quad (15) \\
\sum_{k=0}^n \sum_{C \subseteq \mathcal{N}: |C|=k} p_C^\alpha \ln \frac{p_C^\alpha}{\pi_C^\alpha} - \sum_{k=0}^n \tilde{p}_k^\alpha \ln \frac{\tilde{p}_k^\alpha}{\tilde{\pi}_k^\alpha} &= \sum_{k=0}^n \tilde{p}_k^\alpha \left(\sum_{C \subseteq \mathcal{N}: |C|=k} \left(\frac{p_C^\alpha}{\tilde{p}_k^\alpha} \ln \frac{p_C^\alpha}{\pi_C^\alpha} \right) - \ln \frac{\tilde{p}_k^\alpha}{\tilde{\pi}_k^\alpha} \right) \\
&= \sum_{k=0}^n \tilde{p}_k^\alpha \left(\sum_{C \subseteq \mathcal{N}: |C|=k} \left(\frac{p_C^\alpha}{\tilde{p}_k^\alpha} \ln \frac{p_C^\alpha}{\pi_C^\alpha} - \frac{p_C^\alpha}{\tilde{p}_k^\alpha} \ln \frac{\tilde{p}_k^\alpha}{\tilde{\pi}_k^\alpha} \right) \right),
\end{aligned}$$

where the last equality follows from the fact that $\sum_{C \subseteq \mathcal{N}: |C|=k} p_C^\alpha / \tilde{p}_k^\alpha = 1$. We can write the last expression as,

$$\sum_{k=0}^n \tilde{p}_k^\alpha \sum_{C \subseteq \mathcal{N}: |C|=k} \left(\frac{p_C^\alpha}{\tilde{p}_k^\alpha} \ln \frac{p_C^\alpha / \tilde{p}_k^\alpha}{\pi_C^\alpha / \tilde{\pi}_k^\alpha} \right) = \sum_{k=0}^n \tilde{p}_k^\alpha MI(\mathbf{p}^{\alpha, k} || \tilde{\boldsymbol{\pi}}^{\alpha, k}), \quad (16)$$

which completes the proof of the decomposition (14).

Dividing both sides of Equation (14) by $-(n-1) \ln \alpha^\alpha (1-\alpha)^{1-\alpha} > 0$ for $0 < \alpha < 1$ yields the decomposition of the CTI,

$$\kappa(\mathbf{p}^\alpha) = \tilde{\kappa}(\mathbf{p}^\alpha) + \sum_{k=0}^n \tilde{p}_k^\alpha \kappa^k(\mathbf{p}^\alpha).$$

We note that $\kappa(\mathbf{p}^\alpha) \geq \tilde{\kappa}(\mathbf{p}^\alpha) \geq 0$ follows from the non-negativity of $\tilde{\kappa}(\mathbf{p}^\alpha)$ and $\kappa^k(\mathbf{p}^\alpha)$ as the KL divergence is always non-negative (Cover and Thomas, 2006). ■

Proof of Corollary 1: The proof of this corollary is obtained by observing that $MI(\mathbf{p}^{\alpha, k} || \boldsymbol{\pi}^{\alpha, k}) = 0$ (see formulae (6) and (7)). The latter equality follows from the fact that $\mathbf{p}^{\alpha, k} = \boldsymbol{\pi}^{\alpha, k}$ in the bivariate case for any pdf \mathbf{p}^α and $k = 0, 1, 2$. ■

9. Online Appendix

9.1. Monte Carlo experiments

9.1.1. Interdependence across tails

Generally, interdependence varies across the support of a distribution. When measured over the entire support, it may lead to misleading conclusions, especially for applications where a measure of dependence for a specific area (or direction) is required. To illustrate this, Table 1 shows the system CTIs with their estimated 95% bootstrap confidence intervals for samples generated from three widely used multidimensional parametric models: the multinormal, the multivariate t-distribution with 3 degree of freedom and a Clayton copula with normal marginal densities. Each sample contains 10,000 vectors of 5-dimensional realizations of the relevant model. In all models, the pairwise correlation is set (approximately) to $\rho \in \{0.0, 0.5, 0.9\}$ for each pair of variables.⁹ The CTIs for $\alpha = 0.1$ corresponds to the negative joint tails $\mathcal{T}^\alpha(-\mathbf{1})$, while for $\alpha = 0.5$ and $\alpha = 0.9$ they correspond to the central part of the distribution and positive joint tails $\mathcal{T}^{1-\alpha}(\mathbf{1})$ respectively. The CIs are estimated by bootstrapping. First, we draw a sample S of size 10,000 from a given parametric distribution and estimate the system CTI. The empirical distribution of this parameter is then generated by resampling from S and its empirical quantiles are computed.

Around the median ($\alpha = 0.5$), the dependence structure is similar for all three models (except for the copula with $\rho = 0.9$). In particular, for $\rho = 0.0$ and $\rho = 0.5$, the intersection of the three confidence intervals is non-empty, which suggests that the underlying interdependence structure is the same for all three models.¹⁰ While the multinormal appears to generate the same system CTI for all α , this is not the

⁹In the case of the Clayton copula, the correlations 0, 0.5 and 0.9 correspond to dependence coefficients of 0.001, 1.4 and 9, respectively.

¹⁰The symmetry of two interdependence structures and other hypotheses can be formally tested by our test procedures discussed in the paper.

case for the other models. In particular, the CTI is higher in the extreme tails than around the median for the t-distribution. Interestingly, this distribution has a significant tail interdependence even when its marginals are uncorrelated. This result is driven by the fat tails of this distribution which increase the likelihood of joint exceedances thus leading to a higher CTI than would otherwise prevail under complete independence. For the Clayton copula, on the other hand, we observe a clear asymmetry among the negative and the positive tails for $\rho = 0.5$ and $\rho = 0.9$. This is not surprising since the essence of the Clayton copula is asymmetric tail dependence.

[Table 3]

9.1.2. Total vs. system interdependence

The left panel of Figure 7 shows the total ($\kappa(\alpha)$) and the system ($\tilde{\kappa}(\alpha)$) CTI computed for a standardized $n = 6$ dimensional normal \mathbf{X} with $\text{corr}(X_i, X_k) = \rho$ for all $i, k = 1, \dots, 6, i \neq k$. Note that the tail interdependence from multinormal samples (with a fixed correlation for all pairs of variables) is constant across the entire range of α . Hence, the interdependence in this case neither increases nor decreases as the tails become more extreme. Moreover, the total and the system CTIs are identical for all α implying that this is all system interdependence and that the additional CTI is close to zero in this case. This is intuitive because all pairs have the same correlation and there are no pairs with higher or lower correlation which would generate subsystem tail dependence.

The right panel of Figure 7 shows the results when the correlation is the same for three pairs but zero for the remaining pairs ($\text{corr}(X_1, X_2) = \text{corr}(X_3, X_4) = \text{corr}(X_5, X_6) = 0.7$ and zero for all other pairs). In this case, while the patterns of the total and the system CTIs are similar for all α , $\kappa(\alpha)$ is about three times larger than $\tilde{\kappa}(\alpha)$ correctly identifying that interdependence originates primarily in

interactions within subsets of variables.

[Figure 7]

9.1.3. System CTI in higher dimensions

In the following experiment, we estimate the system CTI and its 95% bootstrap confidence intervals (CI) from samples generated from a high-dimensional multinormal and multivariate t-distribution. First, we estimate the system CTI from the sample S of size 10,000 drawn from a given parametric distribution of $n = 100$ random variables, where the pairwise correlation is set (approximately) to $\rho \in \{0.0, 0.5, 0.9\}$ for each pair of variables. The empirical distribution of the CTI is then generated by resampling from S and its quantiles are used to construct the CIs.

[Table 4]

Around the median ($\alpha = 0.5$), the dependence structure is similar for both models. In particular, for each n , the intersection of the confidence intervals is non-empty for both models and each correlation ρ , which suggests that the underlying interdependence structure could be the same for both models.¹¹ While the multinormal appears to generate the same system CTI for all α , this is not the case for the t-distribution. In this case, the CTI is higher in the extreme tails than around the median. Interestingly, this distribution has a significant tail interdependence even when its marginals are uncorrelated. This result is driven by the fat tails of this distribution which increase the likelihood of joint exceedances thus leading to a higher CTI that would otherwise prevail under complete independence. Overall, this exercise highlights the complex patterns of interdependence and hence, the importance of flexible techniques for its analysis.

¹¹The symmetry of two interdependence structures and other hypotheses can be formally tested by our test procedures outlined in Section 4.