Modelling, Estimation and Visualization of Multivariate Dependence for High-frequency Data

Erik Brodin and Claudia Klüppelberg

Abstract Dependence modelling and estimation is a key issue in the assessment of financial risk. It is common knowledge meanwhile that the multivariate normal model with linear correlation as its natural dependence measure is by no means an ideal model. We suggest a large class of models and a dependence function, which allows us to capture the complete extreme dependence structure of a portfolio. We also present a simple nonparametric estimation procedure of this function. To show our new method at work we apply it to a financial data set of high-frequency stock data and estimate the extreme dependence in the data. Among the results in the investigation we show that the extreme dependence is the same for different time scales. This is consistent with the result on high-frequency FX data reported in Hauksson et al. (2001). Hence, the different asset classes seem to share the same time scaling for extreme dependence. This time scaling property of high-frequency data is also explained from a theoretical point of view.

Key words: Risk management; extreme risk assessment; high-frequency data; multivariate extreme value statistics; multivariate models; tail dependence function

1 Multivariate Risk Assessment for Extreme Risk

Estimation of dependence within a portfolio based on high-frequency data faces various problems:

- data are not normal: they are skewed and heavy-tailed

Erik Brodin
Department of Mathematical Sciences, Chalmers University of Technology, SE-412 96 Göteborg, Sweden, URL: www.chalmers.se/math/EN/, e-mail: ebrodin@math.chalmers.se

Claudia Klüppelberg
Center for Mathematical Sciences, Technische Universität München, D-85747 Garching, Germany, URL: www-m4.ma.tum.de, e-mail: cklu@ma.tum.de

T. Kneib, G. Tutz (eds.), *Statistical Modelling and Regression Structures*,
DOI 10.1007/978-3-7908-2413-1_15,

- one-dimensional data are uncorrelated but not iid
- multivariate high-frequency data are not synchronised
- data are discrete-valued for a very high frequency
- most likely there is microstructure noise in the data
- there is seasonality in the data
- higher moments may not exist
- the multivariate distribution may not be elliptical
- dependence may not be symmetric

We are interested here in the influence of the multivariate dependence within the portfolio. We first recall that under the condition that the portfolio P/L follows a multivariate normal distribution and, if there is no serial dependence, the portfolio P/L standard deviation σ is calculated by the square root of its variance

$$\sigma^2 = \sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i \neq j} w_i w_j \sigma_i \sigma_j \rho_{ij}\,, \tag{1}$$

where the portfolio consists of n different instruments with nominal amount w_i invested into asset i. The standard deviation of asset i is given by σ_i and the pairwise correlation coefficients are ρ_{ij} $(i, j = 1, \ldots, n)$.

Definition 1. For two random variables X and Y their *linear correlation* is defined as

$$\rho_L(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X)\operatorname{var}(Y)}}\,,$$

where $\operatorname{cov}(X,Y) = E((X - EX)(Y - EY))$ is the covariance of X and Y, and $\operatorname{var}(X)$ and $\operatorname{var}(Y)$ are the variances of X and Y, respectively.

Correlation measures linear dependence: we have $|\rho_L(X,Y)| = 1$ if and only if $Y = aX + b$ with probability 1 for $a \in \mathbb{R} \setminus \{0\}$ and $b \in \mathbb{R}$. Furthermore, correlation is invariant under strictly increasing *linear* transformations; i.e. for $\alpha, \gamma \in \mathbb{R} \setminus \{0\}$ and $\beta, \delta \in \mathbb{R}$

$$\rho_L(\alpha X + \beta, \gamma Y + \delta) = \operatorname{sign}(\alpha\gamma)\, \rho_L(X,Y)\,.$$

Also for high-dimensional models correlation is easy to handle. For random (column) vectors $X, Y \in \mathbb{R}^n$ we denote by $\operatorname{cov}(X,Y) = E((X - EX)(Y - EY)^T)$ the covariance matrix of X and Y. Then for $m \times n$ matrices A, B and vectors $a, b \in \mathbb{R}^m$ we calculate

$$\operatorname{cov}(AX + a, BY + b) = A \operatorname{cov}(X,Y)\, B^T\,,$$

where B^T denotes the transpose of the matrix B. From this it follows for $w \in \mathbb{R}^n$ that

$$\operatorname{var}(w^T X) = w^T \operatorname{cov}(X,X) w\,,$$

which is exactly formula (1) above. The popularity of correlation is also based on the fact that it is very easy to calculate and estimate. It is a natural dependence measure for elliptical distributions such as the multivariate normal or t distributions, provided

second moments exist. Within the context of linear models correlation has also proved as a useful tool for dimension reduction (e.g. by factor analysis), an important issue in risk management; see Klüppelberg & Kuhn (2009) for a new approach to dimension reduction for financial data.

Multivariate portfolios, however, are often not elliptically distributed, and there may be a more complex dependence structure than linear dependence. Indeed, data may be uncorrelated, i.e. with correlation 0, but still may be highly dependent. In the context of risk management, when measuring extreme risk, modelling dependence by correlation may be grossly misleading; see e.g. Embrechts et al. (2002).

We turn to a measure for tail dependence, which relates large values of the components of a portfolio; see e.g. Joe (1997). In the bivariate context, consider random variables X and Y with marginal distribution functions G_X and G_Y and (generalized) inverses $G_X^{\leftarrow}$ and $G_Y^{\leftarrow}$. For any distribution function G its *generalized inverse or quantile function* is defined as

$$G^{\leftarrow}(t) = \inf\{x \in \mathbb{R} \mid G(x) \geq t\}, \quad 0 < t < 1.$$

If G is strictly increasing, then $G^{\leftarrow}$ coincides with the usual inverse of G.

Definition 2. The *upper tail dependence coefficient* of (X,Y) is defined by

$$\rho_U = \lim_{u \uparrow 1} P(Y > G_X^{\leftarrow}(u) \mid X > G_Y^{\leftarrow}(u)), \tag{2}$$

provided the limit exists. If $\rho_U \in (0,1]$, then X and Y are called *asymptotically upper tail dependent*, if $\rho_U = 0$, they are called *asymptotically upper tail independent*.

For some situations, this measure may be an appropriate extreme dependence measure; this is true, in particular, when the bivariate distribution is symmetric; see Example 1. However, ρ_U is not a very informative measure, since the extreme dependence around the line with angle $\pi/4$ does not reveal much about what happens elsewhere; see e.g. the asymmetric model in Example 2. As a remedy we suggest an extension of the upper tail dependence coefficient to a function of the angle, which measures extreme dependence in any direction in the first quadrant of $\mathbb{R}^2$. Its derivation is based on multivariate extreme value theory and we indicate this relationship in Section 2. We shall, however, refrain from a precise derivation and rather refer to Hsing et al. (2004) for details. We also want to emphasize that one-dimensional extreme value theory has been applied successfully to risk management problems; see Embrechts (2000). We remark further that one-dimensional extreme value theory has meanwhile reached a consolidated state; we refer to Embrechts et al. (1997) or Coles (2001) as standard references.

We will illustrate our results by a direct application to a real data set. The complete data set we investigated consists of high-frequency data for three different stocks: Intel, Cisco and General Motors (GM). We have full sample paths of the price data of the stocks between February and October 2002 from the Trades and Quotes TAQ database of the New York Stock Exchange (NYSE); i.e. our data consists of all trading dates [in seconds] and corresponding prices [in cents]. They are depicted in

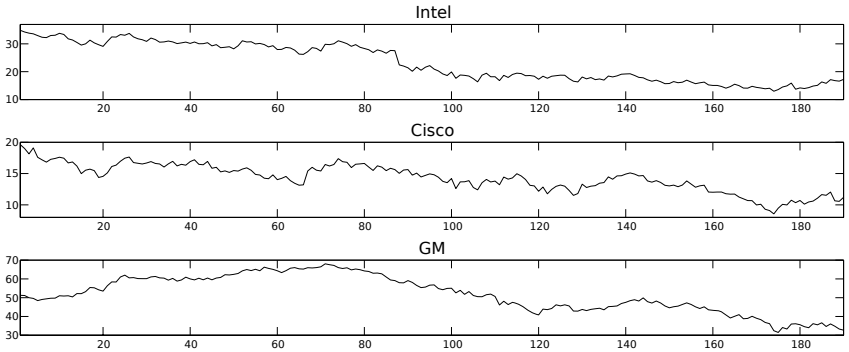

Fig. 1 Example of stock prices: Intel, Cisco and GM: Feb-Oct 2002.

Figure 1. This dataset has to be filtered in different steps due to false data, seasonality and serial dependence; see Section 4 for details.

After these filtering steps the residuals can be assumed iid and the dependence structure between the three stocks can be investigated. The first row of Figure 2 shows scatter plots of different combinations of the filtered stocks. We have estimated the means, variances and the correlation of the data in each scatter plot. The second row shows simulated normal data with the estimated parameters.

For extreme risk assessment one is particularly interested in the left lower corner and we have zoomed into this corner to get a more precise account of the dependence there; see Figure 3. None of the normal models seem to be able to capture the dependence structure in this area.

Our paper is organized as follows. After introducing the tail dependence function in Section 2 we shall present some examples including an asymmetric Pareto model and the bivariate normal model.

In Section 3 we introduce a simple nonparametric estimation procedure of the tail dependence function. We show its performance in various simulation examples and plots.

In Section 4 we investigate our high-frequency data in more detail and estimate their tail dependence function. We also show various plots to visualize our results. Finally, in Section 5 we conclude the paper with a summary of our findings.

2 Measuring Extreme Dependence

Although the upper tail dependence coefficient and its functional extension we are aiming at can be defined for random vectors of any dimension, we restrict ourselves in our presentation to the bivariate case. For a general treatment in any dimension we refer to Hsing et al. (2004).

Suppose $(X_i, Y_i)_{i=1,\dots,n}$ is a sequence of iid vectors and (X, Y) is a generic random vector with the same distribution function $G(x,y) = P(X \le x, Y \le y)$ for $(x,y) \in \mathbb{R}^2$

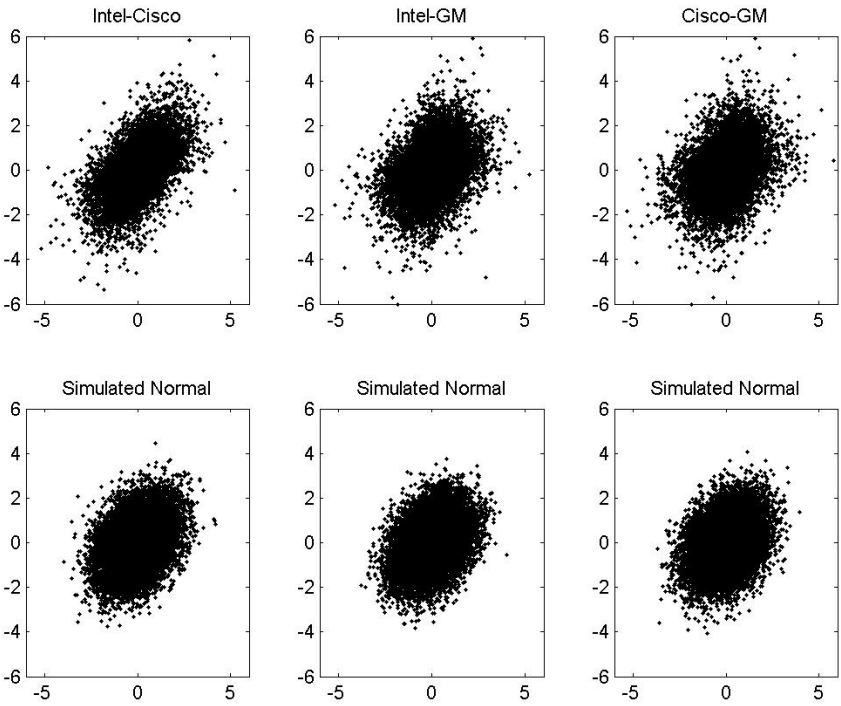

Fig. 2 Bivariate stock data versus simulated normal with same (estimated) means and variances.

with continuous marginals. For $n \in \mathbb{N}$ define the vector of componentwise maxima $M_n = (\max_{i=1,\ldots,n} X_i\,, \max_{i=1,\ldots,n} Y_i)$. As a first goal we want to describe the behavior of M_n for large n.

It is a standard approach in extreme value theory to first transform the marginals to some appropriate common distribution and then model the dependence structure separately. As copulas have become a fairly standard notion for modelling dependence we follow this approach and transform the marginal distributions G_X and G_Y to uniform (0,1). Then we have a bivariate uniform distribution, which is called a *copula* and is given for $0 < u, v < 1$ by

$$C_G(u,v) = P(G_X(X) \le u\,, G_Y(Y) \le v) = P(X \le G_X^{\leftarrow}(u)\,, Y \le G_Y^{\leftarrow}(v))\,.$$

For more details on copulas and dependence structures in general we refer to Joe (1997); for applications of copulas in risk management see Embrechts et al. (2001).The transformation of the marginals to uniforms is illustrated in Figure 4.

Under weak regularity conditions on the bivariate distribution function G we obtain

$$\lim_{n\to\infty} P\left(\max_{i=1,\ldots,n} G_X(X_i) \le 1 + \frac{1}{n}\ln u\,,\ \max_{i=1,\ldots,n} G_Y(Y_i) \le 1 + \frac{1}{n}\ln v\right)$$

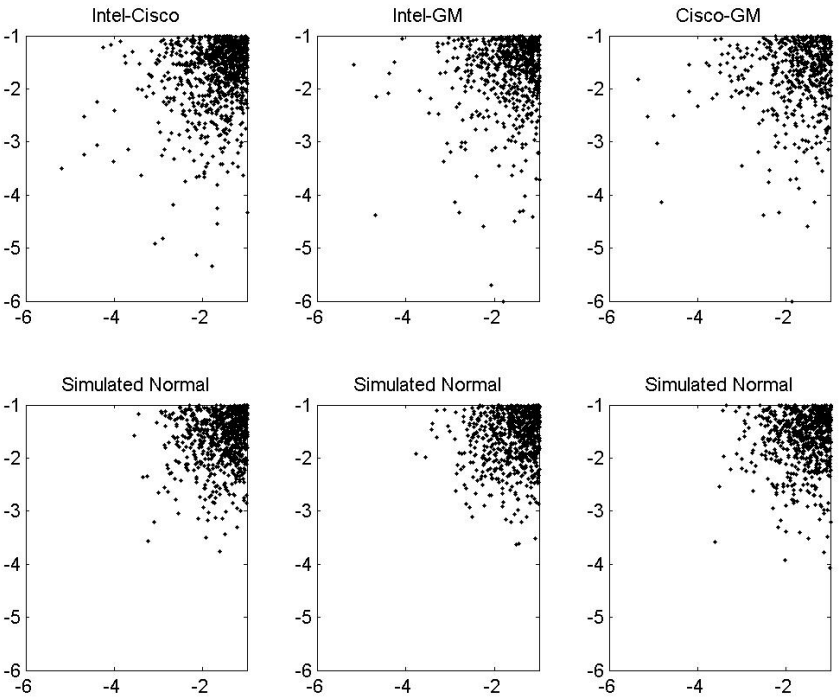

Fig. 3 Bivariate stock data versus simulated normal with same (estimated) means and variances.

$$= \exp(-\Lambda(-\ln u, -\ln v))) = C(u,v), \quad 0 \le u,v \le 1.$$

Such a copula is called *extreme copula* and satisfies for all $t > 0$

$$C^t(u,v) = C(u^t, v^t), \quad 0 < u,v < 1.$$

$C(u,v)$ has various integral representations. The *Pickands' representation* yields an extreme event intensity measure (we write $a \wedge b = \min(a,b)$ and $a \vee b = \max(a,b)$):

$$\begin{aligned} \Lambda(x,y) &= \lim_{n\to\infty} nP\left(G_X(X) > 1 - \frac{x}{n} \text{ or } G_Y(Y) > 1 - \frac{y}{n}\right) \\ &= \int_0^{\pi/2} \left(\frac{x}{1 \vee \cot\theta} \vee \frac{y}{1 \vee \tan\theta}\right) \Phi(d\theta), \quad x,y \ge 0. \end{aligned} \tag{1}$$

Φ is a finite measure on $(0, \pi/2)$ satisfying $\int_0^{\pi/2}(1 \wedge \tan\theta)\,\Phi(d\theta) = \int_0^{\pi/2}(1 \wedge \cot\theta)\,\Phi(d\theta) = 1$. The definition of Λ as a limit of $n \times$ *success probability* is a version of the classical limit theorem of Poisson. For large n the measure Λ can be interpreted as the mean number of data in a strip near the upper and right boundary of the uniform distribution; see Figure 4. We also recall some properties of $\tan\theta = \frac{1}{\cot\theta} = \frac{\sin\theta}{\cos\theta}$: $\tan 0 = 0$, $\tan\theta$ is increasing in $\theta \in (0, \pi/2)$ and $\lim_{\theta\to\pi/2} \tan\theta = \infty$. Then $\cot\theta$ is its reflection on the 45 degree line, correspond-

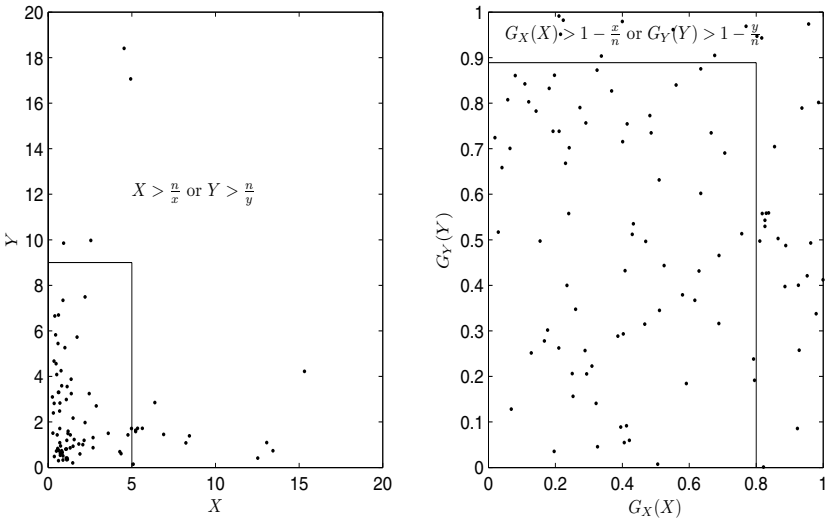

Fig. 4 Left plot: Simulated data for X and Y Fréchet distributed with distribution functions $G_X(x) = G_Y(x) = \exp(-1/x)$ for $x > 0$ with region of large data points indicated. The range of the data is in total $[0,410]$ for X and $[0,115]$ for Y; for reasons of presentation 14 extremely large points had to be left out.
Right plot: Illustration of the intensity measure Λ as defined in equation (1): Λ measures the probability in the strip near the upper and right boundary of the uniform distribution.

ing to $\theta = \pi/4$. Moreover, $\tan(\pi/4) = \cot(\pi/4) = 1$ and $\cot(\frac{\pi}{2} - \theta) = 1/\cot\theta$ for $\theta \in (0, \pi/2)$. Finally, arctan is the inverse function of tan.

The fact that $\Lambda(x,y) = x\Lambda(1, y/x)$ motivates the following definition.

Definition 3. For any random vector (X,Y) such that (1) holds we define the *dependence function* as

$$\psi(\theta) = \Lambda(1, \cot\theta), \quad 0 < \theta < \pi/2.$$

Note that $\psi(\cdot)$ is a function of the angle θ only and measures dependence in any direction of the positive quadrant of a bivariate distribution.

The following result shows that $\psi(\cdot)$ allows us to approximate for large x_1 and y_1 the probability for X or Y to become large. We write $a(x) \sim b(x)$ as $x \to x_0$ for $\lim_{x \to x_0} a(x)/b(x) = 1$. We also denote by $\overline{G}(\cdot) = 1 - G(\cdot)$ the tail of G.

Proposition 1. *Let (X,Y) be a random vector. If $x_1, y_1 \to \infty$ such that $P(X > x_1)/P(Y > y_1) \to \tan\theta$, then the following quotient converges for all $\theta \in (0, \pi/2)$,*

$$\frac{P(X > x_1 \text{ or } Y > y_1)}{P(X > x_1)}.$$

Furthermore, the limit is the dependence function $\psi(\theta)$.

Proof. From (4) we have for large x_1, y_1 and $x = n\overline{G}_X(x_1)$ and $y = n\overline{G}_Y(y_1)$ as $n \to \infty$ (note that x, x_1, y, y_1 depend on n):

$$P(X > x_1 \text{ or } Y > y_1) \sim \frac{1}{n}\Lambda(n\overline{G}_X(x_1), n\overline{G}_Y(y_1))$$
$$= \overline{G}_X(x_1)\Lambda\left(1, \frac{\overline{G}_Y(y_1)}{\overline{G}_X(x_1)}\right) = \overline{G}_X(x_1)\psi\left(\arctan\left(\frac{\overline{G}_X(x_1)}{\overline{G}_Y(y_1)}\right)\right). \tag{2}$$

We set

$$\theta = \arctan\left(\frac{\overline{G}_X(x_1)}{\overline{G}_Y(y_1)}\right)$$

and obtain the result.

The following corollary summarizes some obvious results; the symmetry property of part (d) is new and will prove useful for estimation purposes.

Corollary 1. *(a) For X and Y independent we calculate*

$$\frac{P(X > x_1 \text{ or } Y > y_1)}{P(X > x_1)} \sim \frac{P(X > x_1) + P(Y > y_1)}{P(X > x_1)} \to 1 + \cot\theta =: \psi_0(\theta)$$

for $x_1, y_1 \to \infty$ such that $P(Y > y_1)/P(X > x_1) \to \cot\theta$.
(b) For X and Y completely dependent, i.e. $X = g(Y)$ with probability 1 for some increasing function g, we obtain

$$\frac{P(X > x_1 \text{ or } Y > y_1)}{P(X > x_1)} = \frac{P(X > x_1) \vee P(X > y_1)}{P(X > x_1)} \to 1 \vee \cot\theta =: \psi_1(\theta)$$

for $x_1, y_1 \to \infty$ such that $P(Y > y_1)/P(X > x_1) \to \cot\theta$.
(c) $\psi_1(\theta) \le \psi(\theta) \le \psi_0(\theta)$ for $0 < \theta < \pi/2$.
(d) $\psi_{Y,X}(\theta) = \cot\theta\, \psi_{X,Y}(\pi/2 - \theta)$.

Proof. It only remains to proof part (d). By Example 4.3 in Hsing et al. (2004) we have together with the change of variables $x = t\tan\theta$,

$$1 + \cot\theta - \psi_{X,Y}(\theta) = \lim_{t\to\infty} P(X > G_X^{\leftarrow}(1 - 1/(t\tan\theta)) | Y > G_Y^{\leftarrow}(1 - 1/t))$$
$$= \lim_{t\to\infty} P(Y > G_Y^{\leftarrow}(1 - 1/t) | X > G_X^{\leftarrow}(1 - 1/(t\tan\theta))) \frac{P(X > G_X^{\leftarrow}(1 - 1/(t\tan\theta)))}{P(Y > G_Y^{\leftarrow}(1 - 1/t))}$$
$$= \cot\theta\,(1 + \tan\theta - \psi_{Y,X}(\pi/2 - \theta)) = \cot\theta + 1 - \cot\theta\, \psi_{Y,X}(\pi/2 - \theta).$$

We normalize $\psi(\cdot)$ to the interval $[0, 1]$ as follows.

Definition 4. The normalized function

$$\rho(\theta) = \frac{\psi_0(\theta) - \psi(\theta)}{\psi_0(\theta) - \psi_1(\theta)} = \frac{1 + \cot\theta - \psi(\theta)}{1 \wedge \cot\theta}, \quad 0 < \theta < \pi/2,$$

we call *tail dependence function*.

Note that ρ describes the tail dependence of (X,Y) in any direction of the bivariate distribution on the positive quadrant of $\mathbb{R}^2$.

By this definition we have $\rho(\theta) \in [0,1]$ for all $0 < \theta < \pi/2$, $\rho(\theta) \equiv 0$ in case of independence and $\rho(\theta) \equiv 1$ in case of complete dependence. Consequently, $\rho(\theta)$ being close to 0/1 corresponds to weak/strong extreme dependence.

Remark 1. (i) (Relation between tail dependence function and Pickands' dependence function.) We can write an extreme copula as

$$C(u,v) = \exp\left(\log(uv) A\left(\frac{\log(v)}{\log(uv)}\right)\right), \quad 0 < u,v < 1.$$

The function $A : [0,1] \to [\frac{1}{2}, 1]$ is called *Pickands' dependence function*. $A \equiv 1$ corresponds to independence and $A(t) = t \vee (1-t)$ to total dependence. Using $-\Lambda(-\log(u), -\log(v)) = \log(C(u,v))$ we have the following relation between ρ and A:

$$\rho(\theta) = \frac{(1 + \cot\theta)(1 - A\left(\frac{\cot\theta}{1+\cot\theta}\right))}{1 \wedge \cot\theta}, \quad 0 < \theta < \pi/2.$$

(ii) For elliptical copula models a new semi-parametric approach for extreme dependence modelling was suggested and investigated in Klüppelberg et al. (2007, 2008).

The function $\rho(\cdot)$ is invariant under monotone transformation of the marginal distributions. We show this by calculating it as a function of the copula.

Proposition 2. *Let (X,Y) be a random vector with continuous marginal distribution functions G_X and G_Y. Then $G_X(X) \stackrel{d}{=} U$ and $G_Y(Y) \stackrel{d}{=} V$ for uniform random variables U and V with the same dependence structure as (X,Y). Denote by $C(u,v) = P(U \le u, V \le v)$ the corresponding copula. We also relate the arguments by $G_X(x_1) = u$ and $G_Y(y_1) = v$. Then, provided that the limits exist,*

$$\rho(\theta) = \lim_{\substack{u,v \to 1 \\ (1-u)/(1-v) \to \tan\theta}} \frac{1 - u - v + C(u,v)}{(1-u) \wedge (1-v)}, \quad 0 < \theta < \pi/2.$$

Proof.

$$\psi(\theta) = \lim_{\substack{x_1, y_1 \to \infty \\ \overline{G}_X(x_1)/\overline{G}_Y(y_1) \to \tan\theta}} \frac{1 - P(X \le x_1, Y \le y_1)}{P(X > x_1)} = \lim_{\substack{u,v \to 1 \\ (1-u)/(1-v) \to \tan\theta}} \frac{1 - C(u,v)}{1-u}.$$

Remark 2. Note also that the quantity $\rho(\pi/4)$ is nothing but the *(upper) tail dependence coefficient* ρ_U as defined in (2). Thus, the function ρ extends this notion from

a single direction, the 45 degree line corresponding to $\theta = \pi/4$, to all directions in $(0, \pi/2)$.

This extension is illustrated by the following examples.

Example 1. [Gumbel copula]
Let (X,Y) be a bivariate random vector with dependence structure given by a Gumbel copula for $\delta \in [1,\infty)$:

$$C(u,v) = \exp\left\{-\left[(-\ln u)^\delta + (-\ln v)^\delta\right]^{1/\delta}\right\}, \quad 0 < u,v < 1\,. \tag{3}$$

The dependence arises from δ. To calculate $\psi(\theta)$ we use the relationship of ψ to its copula. We use also the fact that for $u, v \to 1$ we have

$$\frac{-\ln v}{-\ln u} \sim \frac{1-v}{1-u} \to \cot\theta\,.$$

Then by continuity of u^x in x we obtain for $u, v \to 1$ such that $(1-v)/(1-u) \to \cot\theta$

$$1 - C(u,v) = 1 - \exp\left(\ln u\left[1 + \left(\frac{-\ln v}{-\ln u}\right)^\delta\right]^{1/\delta}\right) \sim 1 - u^{(1+(\cot\theta)^\delta)^{1/\delta}}\,.$$

Using the l'Hospital rule and the fact that $u \to 1$, we obtain

$$\frac{1-C(u,v)}{1-u} \to \left(1 + (\cot\theta)^\delta\right)^{1/\delta},$$

and hence

$$\rho(\theta) = \frac{1 + \cot\theta - \left(1 + (\cot\theta)^\delta\right)^{1/\delta}}{1 \wedge \cot\theta}, \quad 0 < \theta < \pi/2\,.$$

We also obtain the well-known upper tail dependence coefficient $\rho_U = \rho(\pi/4) = 2 - 2^{1/\delta}$.

Our next result concerns models, whose extreme dependence vanishes in the limit.

Proposition 3. *Let (X,Y) be a random vector with continuous marginal distribution functions G_X and G_Y. If $\rho(\theta_0) = 0$ for some $\theta_0 \in (0,\pi/2)$ then $\rho(\theta) = 0$ for all $\theta \in (0,\pi/2)$.*

Proof. From Corollary 1(d) we have

$$\rho_{Y,X}(\pi/2 - \theta) = \rho(\theta), \quad 0 < \pi/2 < 1\,. \tag{4}$$

Now note that $P(X > G_X^{\leftarrow}(1 - 1/(t\tan\theta)) | Y > G_Y^{\leftarrow}(1-1/t))$ is decreasing in θ, hence if $\rho(\theta_0) = 0$ then $\rho(\theta) = 0$ for $\theta > \theta_0$. Now, assume that $\rho(\pi/4) = 0$ so that

$\rho_{Y,X}(\pi/4) = 0$ by (4). This results in $\rho(\theta) = 0$ and $\rho_{Y,X}(\theta) = 0$ for $\theta > \pi/4$, i.e $\rho \equiv 0$ by (4) and monotonicity. Hence, we only have to show that $\rho(\theta_0) = 0$ for some $\theta_0 \in (0, \pi/2)$ implies $\rho(\pi/4) = 0$. This is trivial for $\theta_0 < \pi/4$ by monotonicity. For $\theta_0 > \pi/4$, (4) gives $\rho_{Y,X}(\pi/2 - \theta_0) = 0$ for $\pi/2 - \theta_0 < \pi/4$, so that $\rho_{Y,X}(\pi/4) = \rho(\pi/4) = 0$ and this finishes the proof.

We conclude with the multivariate normal distribution. It is well-known (see e.g. Embrechts et al. 2001, 2002) that for correlation $\rho < 1$ the upper tail dependence coefficient is $\rho_U = 0$. Consequently, Proposition 3 gives the following result.

Corollary 2. *For a bivariate normal distribution with correlation $\rho < 1$ we have $\rho \equiv 0$.*

The following example is a typical model to capture risk in the extremes.

Example 2. [Asymmetric Pareto model]
For $p_1, p_2 \in (0,1)$ set $\overline{p}_1 = 1 - p_1$ and $\overline{p}_2 = 1 - p_2$ and consider the model

$$X = p_1 Z_1 \vee \overline{p}_1 Z_2 \quad \text{and} \quad Y = p_2 Z_1 \vee \overline{p}_2 Z_3$$

with Z_1, Z_2, Z_3 iid Pareto(1) distributed; i.e., $P(Z_i > x) = x^{-1}$ for $x \geq 1$. Clearly, the dependence between X and Y arises from the common component Z_1. Hence the dependence is stronger for larger values of p_1, p_2. We calculate the function ρ, and observe first that by independence of the Z_i for $x \to \infty$,

$$\begin{aligned} P(X > x) &= 1 - P(p_1 Z_1 \vee \overline{p}_1 Z_2 \leq x) = 1 - P(p_1 Z_1 \leq x) P(\overline{p}_1 Z_2 \leq x) \\ &= 1 - \left(1 - \frac{p_1}{x}\right)\left(1 - \frac{\overline{p}_1}{x}\right) \sim \frac{1}{x}(p_1 + \overline{p}_1) = \frac{1}{x}. \end{aligned}$$

Consequently, we choose $y = x \tan\theta$, which satisfies the conditions of Proposition 1 and calculate similarly,

$$\begin{aligned} &P(X > x \text{ or } Y > x \tan\theta) = 1 - P(X \leq x, Y \leq x \tan\theta) \\ &= 1 - P\left(Z_1 \leq \frac{x}{p_1} \wedge \frac{x \tan\theta}{p_2}\right) P\left(Z_2 \leq \frac{x}{\overline{p}_1}\right) P\left(Z_3 \leq \frac{x \tan\theta}{\overline{p}_2}\right) \\ &\sim \frac{1}{x}(p_1 \vee p_2 \cot\theta + \overline{p}_1 + \overline{p}_2 \cot\theta), \end{aligned}$$

which implies $\psi(\theta) = 1 + \cot\theta - p_1 \wedge p_2 \cot\theta$ for $0 < \theta < \pi/2$ and

$$\rho(\theta) = \frac{p_1 \wedge p_2 \cot\theta}{1 \wedge \cot\theta}, \quad 0 < \theta < \pi/2.$$

An important class of distributions are those with Pareto-like tails. Proposition 4 ensures that, within this class, multivariate returns on different timescales have the same extremal (spatial) dependence, provided the observations are independent and have no time series structure. Hence, one can take advantage of the fact that a higher frequency results in a larger sample and is easier to estimate. We shall illustrate this in

Section 4.5. This version of the proof of Proposition 4 was kindly communicated to the first author by Patrik Albin. Also, one can find a similar proposition in Hauksson et al. (2001) in the setting of multivariate regular variation.

Proposition 4. *Let (X,Y) be a random vector with marginal tails $\overline{G}_X$ and $\overline{G}_Y$ that are regularly varying at infinity, with indices $\alpha < 0$ and $\beta < 0$, respectively. Denote by X^{*n} the sum of n iid copies of X and define Y^{*n} analogously. If the limit*

$$\lim_{t\to\infty} P\big(X > G_X^{\leftarrow}(1-\lambda/t)\,\big|\,Y > G_Y^{\leftarrow}(1-1/t)\big) = L(\lambda) \quad \textit{exists for } \lambda > 0, \tag{5}$$

then the following hold:
(a) $P\big(X^{*n} > x, Y^{*n} > y\big) \sim nP(X > x, Y > y)$ *as* $x,y \to \infty$*;*
(b) The marginal tails $\overline{G}_{X^{*n}}$ *and* $\overline{G}_{Y^{*n}}$ *of* X^{*n} *and* Y^{*n} *satisfy for all* $n \geq 2$

$$\lim_{t\to\infty} P\big(X^{*n} > G_{X^{*n}}^{\leftarrow}(1-\lambda/t)\,\big|\,Y^{*n} > G_{Y^{*n}}^{\leftarrow}(1-1/t)\big) = L(\lambda) \quad \textit{for } \lambda > 0\,.$$

Proof. (a) The one-dimensional version of this result goes back to Feller and has been extended to the larger class of subexponential random variables (see e.g. Embrechts et al. 1997, Appendix A3); i.e. we have

$$P\big(X^{*n} > t\big) \sim nP(X > t) \quad \text{and} \quad P\big(Y^{*n} > t\big) \sim nP(Y > t) \quad \text{as } t \to \infty. \tag{6}$$

We prove a bivariate version of this result. For $\varepsilon > 0$ sufficiently small, we have

$$\begin{aligned}
P\big(X^{*n} > x, Y^{*n} > y\big) &\leq \sum_{i=1}^{n}\sum_{j=1}^{n} P\big(X_i > (1-(n-1)\varepsilon)x, Y_j > (1-(n-1)\varepsilon)y\big) \\
&\quad + \sum_{1\leq i\neq k\leq n}\sum_{j=1}^{n} P\big(X_i > \varepsilon x, X_k > \varepsilon x, Y_j > (1-(n-1)\varepsilon)y\big) \\
&\quad + \sum_{i=1}^{n}\sum_{1\leq j\neq l\leq n} P\big(X_i > (1-(n-1)\varepsilon)x, Y_j > \varepsilon y, Y_l > \varepsilon y\big) \\
&\quad + \sum_{1\leq i\neq k\leq n}\sum_{1\leq j\neq l\leq n} P\big(X_i > \varepsilon x, X_k > \varepsilon x, Y_j > \varepsilon y, Y_l > \varepsilon y\big) \\
&\leq nP(X > (1-(n-1)\varepsilon)x, Y > (1-(n-1)\varepsilon)y) \\
&\quad + n^2 P(X > (1-(n-1)\varepsilon)x)\,P(Y > (1-(n-1)\varepsilon)y) \\
&\quad + 2n^2 P(X > \varepsilon x, Y > \varepsilon y)\,(P(X > \varepsilon x) + P(Y > \varepsilon y)) \\
&\quad + n^3 P(X > \varepsilon x)\,P(Y > \varepsilon y)\,(P(X > \varepsilon x) + P(Y > \varepsilon y)) \\
&\quad + n^2 P(X > (1-(n-1)\varepsilon)x)\,P(Y > (1-(n-1)\varepsilon)y) \\
&\quad + n^2 P(X > \varepsilon x, Y > \varepsilon y)^2 \\
&\quad + n^3 P(X > \varepsilon x, Y > \varepsilon y)\,P(X > \varepsilon x)\,P(Y > \varepsilon y) \\
&\quad + n^4 P(X > \varepsilon x)^2\,P(Y > \varepsilon y)^2 \\
&\sim nP(X > (1-(n-1)\varepsilon)x, Y > (1-(n-1)\varepsilon)y) \quad \text{as } x,y \to \infty
\end{aligned} \tag{7}$$

by (5) together with the regular variation properties. Now, using again an $\varepsilon > 0$ and properties of disjoints sets together with the Boolean inequality, we estimate

$$\begin{aligned}
&P\big(X^{*n} > x, Y^{*n} > y\big) \\
&\geq \sum_{i=1}^{n} P\big(X_i > (1+(n-1)\varepsilon)x, Y_i > (1+(n-1)\varepsilon)y \\
&\qquad\qquad \bigcap_{j\neq i}\{-\varepsilon x \leq X_j \leq x, -\varepsilon y \leq Y_j \leq y\}\big) \\
&\geq \sum_{i=1}^{n} P\big(X_i > (1+(n-1)\varepsilon)x, Y_i > (1+(n-1)\varepsilon)y\big) \\
&\quad - \sum_{i=1}^{n}\sum_{j\neq i} P\big(X_i > (1+(n-1)\varepsilon)x, Y_i > (1+(n-1)\varepsilon)y, X_j \notin [-\varepsilon x, x]\big) \\
&\quad - \sum_{i=1}^{n}\sum_{j\neq i} P\big(X_i > (1+(n-1)\varepsilon)x, Y_i > (1+(n-1)\varepsilon)y, Y_j \notin [-\varepsilon y, y]\big) \\
&\sim nP(X > (1+(n-1)\varepsilon)x, Y > (1+(n-1)\varepsilon)y) \quad \text{as } x,y \to \infty.
\end{aligned}$$

(b) Proposition 1.5.15 of Bingham et al. (1987) ensures that the generalized inverses satisfy as $t \to \infty$,

$$G_X^{\leftarrow}(1-1/t) \sim G_{X^{*n}}^{\leftarrow}(1-n/t) \quad \text{and} \quad G_Y^{\leftarrow}(1-1/t) \sim G_{Y^{*n}}^{\leftarrow}(1-n/t). \tag{8}$$

In particular, $G_X^{\leftarrow}(1-1/\cdot)$ and $G_{X^{*n}}^{\leftarrow}(1-1/\cdot)$ are regularly varying with index $1/\alpha$, while $G_Y^{\leftarrow}(1-1/\cdot)$ and $G_{Y^{*n}}^{\leftarrow}(1-1/\cdot)$ are regularly varying with index $1/\beta$.

By (5)-(7), we have (with ε not the same as before)

$$\begin{aligned}
&\limsup_{t\to\infty} P\big(X^{*n} > G_{X^{*n}}^{\leftarrow}(1-\lambda/t) \,\big|\, Y^{*n} > G_{Y^{*n}}^{\leftarrow}(1-1/t)\big) \\
&\leq \limsup_{t\to\infty} \frac{nP\big(X > (1-\varepsilon)^{\beta/\alpha} G_X^{\leftarrow}(1-\lambda/(nt)), Y > (1-\varepsilon)G_Y^{\leftarrow}(1-1/(nt))\big)}{nP\big(Y > (1+\varepsilon)G_Y^{\leftarrow}(1-1/(nt))\big)} \\
&\leq \limsup_{t\to\infty} \frac{P\big(X > G_X^{\leftarrow}(1-(1-2\varepsilon)^{\beta}\lambda/(nt)), Y > G_Y^{\leftarrow}(1-(1-2\varepsilon)^{\beta}/(nt))\big)}{((1+\varepsilon)/(1-3\varepsilon))^{\beta} P\big(Y > (1-3\varepsilon)G_Y^{\leftarrow}(1-1/(nt))\big]} \\
&\leq \Big(\frac{1-3\varepsilon}{1+\varepsilon}\Big)^{\beta} \limsup_{t\to\infty} P\Big(X > G_X^{\leftarrow}\Big(1-\frac{(1-2\varepsilon)^{\beta}\lambda}{nt}\Big) \,\Big|\, Y > G_Y^{\leftarrow}\Big(1-\frac{(1-2\varepsilon)^{\beta}}{(nt)}\Big)\Big) \\
&= \Big(\frac{1-3\varepsilon}{1+\varepsilon}\Big)^{\beta} L(\lambda) \\
&\to L(\lambda) \quad \text{as } \varepsilon \downarrow 0.
\end{aligned}$$

Analogously follows from the reverse inequality in (a)

$$\liminf_{t\to\infty} P\big(X^{*n} > G_{X^{*n}}^{\leftarrow}(1-\lambda/t) \,\big|\, Y^{*n} > G_{Y^{*n}}^{\leftarrow}(1-1/t)\big) \geq L(\lambda).$$

Remark 3. In Example 4.3 in Hsing et al. (2004) we have, for X and Y random variables with continuous distributions G_X and G_Y,

$$\lim_{t\to\infty} P(X > G_X^{\leftarrow}(1-1/(t\tan\theta))|Y > G_Y^{\leftarrow}(1-1/t)) = (1\wedge\cot\theta)\rho(\theta)\,, 0<\theta<\pi/2\,.$$

Hence, for X and Y random variables with Pareto-like tails, setting $\lambda = 1/\tan\theta$ and $L(\cot\theta) = (1\wedge\cot\theta)\rho(\theta)$ we conclude $L(\lambda) = (1\wedge\lambda)\rho(\arctan(1/\lambda))$ for $\lambda > 0$.

Corollary 3. *Denote by $\psi(\theta)$ the dependence function of (X,Y). Let X^{*n} and Y^{*n} be the sum of n iid copies of X and Y, respectively, and denote by $\psi^{*n}(\cdot)$ the dependence function of (X^{*n},Y^{*n}) for $n\geq 2$. Then $\psi^{*n}(\theta) = \psi(\theta)$ for all $0<\theta<\pi/2$. The same holds for the tail dependence function $\rho(\theta)$.*

3 Extreme Dependence Estimation

To assess extreme dependence in data we estimate the tail dependence function $\rho(\cdot)$ on the positive quadrant. We use a nonparametric estimator as suggested in Hsing et al. (2004) based on the empirical distribution function, which yields a simple nonparametric estimator of $\psi(\cdot)$ and hence of $\rho(\cdot)$. Recall that the empirical distribution function given by

$$\widehat{G}_X(x) = \widehat{P}_n(X\leq x) = \frac{1}{n}\sum_{j=1}^{n} I(X_i\leq x)\,,\quad x\in\mathbb{R}\,,$$

is the standard estimator for the distribution function G_X of iid data ($I(\mathbf{A})$ denotes the indicator function of the set $\mathbf{A}$). The empirical distribution function can be rewritten in terms of the ranks of the sample variables X_i for $i = 1,\ldots,n$ and we write

$$\widehat{G}_X(X_i) = \widehat{P}_n(X\leq X_i) = \frac{1}{n}\mathrm{rank}(X_i)\,.$$

We still have to explain one important issue of our estimation procedure. Recall from (1), denoting by $\overline{G}_X(\cdot) = 1-G_X(\cdot)$ and $\overline{G}_Y(\cdot) = 1-G_Y(\cdot)$ for continuous G_X and G_Y, that

$$\begin{aligned}\Lambda_n(x,y) &:= nP\left(G_X(X) > 1-\frac{x}{n} \text{ or } G_Y(Y) > 1-\frac{y}{n}\right)\\ &= nP\left(n\overline{G}_X(X)\leq x \text{ or } n\overline{G}_Y(Y)\leq y\right)\\ &= nP(n(\overline{G}_X(X),\overline{G}_Y(Y))\in\mathbf{A})\\ &\to \Lambda(x,y)\quad n\to\infty\,.\end{aligned} \tag{9}$$

By a continuity argument we can replace $n\in\mathbb{N}$ by $t\in(0,\infty)$ and also replace in a first step the probability measure P by its empirical counterpart $\widehat{P}_n$. Then we obtain

$$\widehat{\Lambda}_{t,n}(x,y) = t\widehat{P}_n\left(t\left(\overline{G}_X(X),\overline{G}_Y(Y)\right)\in \mathbf{A}\right) = \frac{t}{n}\sum_{i=1}^{n} I(t\left(\overline{G}_X(X),\overline{G}_Y(Y)\right)\in \mathbf{A}).$$

Now estimate the two distribution tails by their empirical counterparts:

$$\widehat{\overline{G}}_X(X_i) := \frac{1}{n}R_i^X := \frac{1}{n}\text{rank}(-X_i) \quad \text{and} \quad \widehat{\overline{G}}_Y(Y_i) := \frac{1}{n}R_i^Y := \frac{1}{n}\text{rank}(-Y_i).$$

Then setting $\varepsilon = t/n$ we obtain

$$\widehat{\Lambda}_{\varepsilon,n}(\mathbf{A}) = \varepsilon\sum_{i=1}^{n} I(\varepsilon\left(R_i^X, R_i^Y\right)\in \mathbf{A}).$$

This yields in combination with Definition 4 an estimator for the function ρ:

$$\widehat{\rho}_{\varepsilon,n}(\theta) = \frac{1+\cot\theta - \widehat{\Lambda}_{\varepsilon,n}(1,\cot\theta)}{1\wedge\cot\theta}, \quad 0\le\theta\le\frac{\pi}{2}, \tag{10}$$

where $\widehat{\Lambda}_{\varepsilon,n}(1,\cot\theta)$ can be rewritten as

$$\varepsilon\sum_{i=1}^{n} I(R_i^X \le \varepsilon^{-1} \text{ or } R_i^Y \le \varepsilon^{-1}\cot\theta), \quad 0\le\theta\le\frac{\pi}{2}. \tag{11}$$

Choosing ε is not an easy task and when θ approaches $\pi/2$ increasingly fewer points are used in the estimation. In Hsing et al. (2004) this problem was solved by letting ε decrease slightly as θ approaches $\pi/2$. A much better solution is provided by the symmetry proved in Corollary 1(d) in combination with (4): the extreme dependence of (X,Y) for $\theta\in[\pi/4,\pi/2]$ is the same as the extreme dependence of (Y,X) for $\theta\in[0,\pi/4]$. Consequently, we estimate $\rho_{\varepsilon,n}(\theta)$ by estimating $\rho_{X,Y}(\theta)$ by

$$\widehat{\rho}_{\varepsilon,n}(\theta) := \begin{cases} \widehat{\rho}_{\varepsilon,n}^{XY}(\theta), & 0<\theta<\pi/4, \\ \widehat{\rho}_{\varepsilon,n}^{YX}(\pi/2-\theta), & \pi/4\le\theta<\pi/2. \end{cases} \tag{12}$$

In the following remark we summarize some important properties of $\widehat{\rho}_{\varepsilon,n}$.

Remark 4. (i) Estimator (12) has good convergence properties: for appropriately small ε and $n\to\infty$ it converges in probability and almost surely; see Hsing et al. (2004) and references therein.
(ii) To assess asymptotic dependence involves passing to a limit function, which for a finite sample is simply impossible. Consequently, for X and Y independent, even for very small ε it is highly possible that the estimated tail dependence function will be positive. This can be made precise by calculating

$$\begin{aligned}
&\varepsilon \sum_{i=1}^{n} I(R_i^X \le \varepsilon^{-1} \text{ or } R_i^Y \le \varepsilon^{-1}\cot\theta) \\
&= \varepsilon \sum_{i=1}^{n} (I(R_i^X \le \varepsilon^{-1}) + I(R_i^Y \le \varepsilon^{-1}\cot\theta)) - I(R_i^X \le \varepsilon^{-1} \text{ and } R_i^Y \le \varepsilon^{-1}\cot\theta)) \\
&= 1 + \cot\theta - \varepsilon \sum_{i=1}^{n} I(R_i^X \le \varepsilon^{-1} \text{ and } R_i^Y \le \varepsilon^{-1}\cot\theta)\,.
\end{aligned}$$

Now, independent samples for X and Y yield for fixed n, ε and θ

$$\sum_{i=1}^{n} I(R_i^X \le \varepsilon^{-1} \text{ and } R_i^Y \le \varepsilon^{-1}\cot\theta) \sim \text{Bin}\left(\frac{\cot\theta}{\varepsilon^2 n^2}, n\right).$$

Hence,

$$E\left(\varepsilon \sum_{i=1}^{n} I(R_i^X \le \varepsilon^{-1} \text{ or } R_i^Y \le \varepsilon^{-1}\cot\theta)\right) = 1 + \cot\theta - \frac{\cot\theta}{\varepsilon n}, \quad 0 < \theta < \frac{\pi}{2},$$

giving

$$E\left(\widehat{\rho}_{\varepsilon,n}(\theta)\right) = \begin{cases} \dfrac{\cot\theta}{\varepsilon n}, & 0 < \theta < \dfrac{\pi}{4}, \\ \dfrac{\cot(\pi/2-\theta)}{\varepsilon n}, & \dfrac{\pi}{4} \le \theta < \dfrac{\pi}{2}. \end{cases} \tag{13}$$

In much the same fashion we get

$$\text{Var}(\widehat{\rho}_{\varepsilon,n}(\theta)) = \begin{cases} \dfrac{\cot\theta}{n} - \dfrac{\cot^2\theta}{\varepsilon^2 n^3}, & 0 < \theta < \dfrac{\pi}{4}, \\ \dfrac{\cot(\pi/2-\theta)}{n} - \dfrac{\cot^2(\pi/2-\theta)}{\varepsilon^2 n^3}, & \dfrac{\pi}{4} \le \theta < \dfrac{\pi}{2}. \end{cases}$$

(iii) Inspecting equation 11 one can see that choosing an ε is equivalent to the estimation of $\rho(\theta)$ based on the $1/\varepsilon$ largest values of X. Hence, it is natural to see $1/\varepsilon$ as a threshold of the data and we will therefore use this term.
(iv) The estimator $\widehat{\rho}_{\varepsilon,n}$ has the advantage that it is only based on the ranks of the data. Consequently, it can be smoothed in the usual way. For instance, by averaging it over a window of size $2m+1$ for $m \in \mathbb{N}$, we call this smoothed estimator $\widehat{\rho}^{(m)}_{\varepsilon,n}(\cdot)$.

In the second column of Figures 5 and 6 we estimated $\rho(\theta)$ for the Gumbel copula (cf. Example 1) and the asymmetric Pareto model (cf. Example 2). The estimated tail dependence function is indeed (except for $\theta \in \{0, \pi/2\}$, where $E(\widehat{\rho}_{\varepsilon,n}(\cdot))$ has singularities) far away from $E(\widehat{\rho}_{\varepsilon,n}(\cdot))$. For our sample size and the chosen ε it is smaller than 0.075 for the interval depicted. Given that the variance is of the order n^{-1} the estimated extreme dependence in our data is significant.

Example 3. [Gumbel copula: continuation of Example 1]
In Figure 5 we simulated the model (with student-t marginals with 8 degrees of

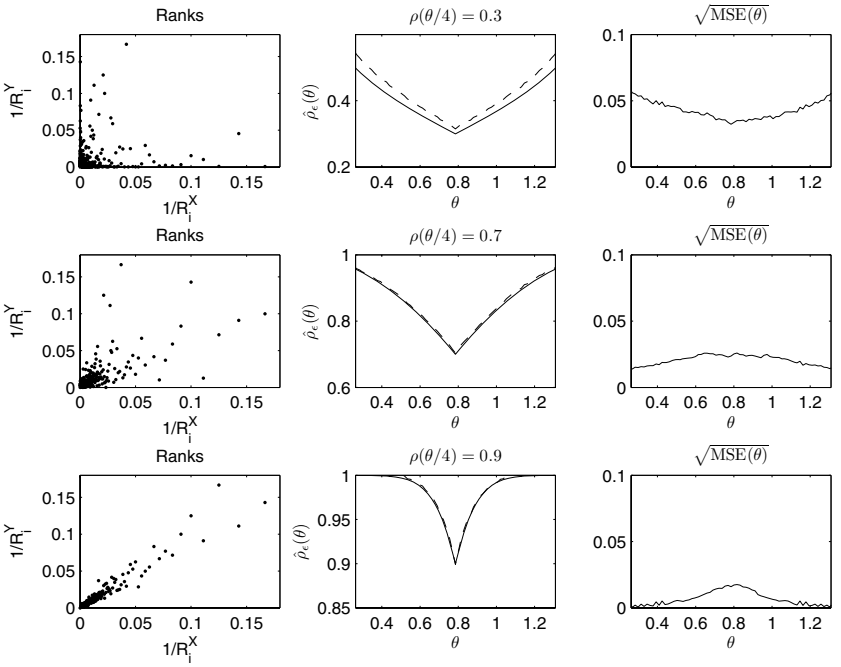

Fig. 5 Simulated Gumbel copula model for $\rho(\theta/4) = 0.3$ (upper row), $\rho(\theta/4) = 0.7$ (middle row), $\rho(\theta/4) = 0.9$ (lower row).
Left column: Plots of ranks $(1/R_i^X, 1/R_i^Y)$, with points close to $(1,1)$ truncated.
Middle column: Plots of $\widehat{\rho}(\theta)$ (dashed) overlaid with true function $\rho(\theta)$ (solid).
Right column: Estimation error in terms of $\sqrt{\mathrm{MSE}(\theta)}$.

freedom) for $n = 10\,000$ iid observations of (X,Y) 100 times. We estimate the tail dependence function $\rho(\cdot)$ for this model with $\varepsilon = 1/200$. We stay away from the boundaries $\theta = 0$ and $\theta = \pi/2$, since in the numerator of (10) we have the difference of two quantities which both tend to ∞ as $\theta \to 0$. The three sets of plots on the three rows correspond to the cases: $\rho(\pi/4) = 0.3$ (upper row), $\rho(\pi/4) = 0.7$ (middle row) and $\rho(\pi/4) = 0.9$ (lower row). On each row the left plots contain ranks $\left(1/R_i^X, 1/R_i^Y\right)$, $1 \le i \le n$, of a simulated sample of size 10 000. Points on the axes correspond to independent extreme points; all points in the open quadrant exhibit some extreme dependence structure. Completely dependent points are to be found on the 45-degree line. The level of dependence is manifested by the data scattered around this diagonal. The true functions $\rho(\theta)$ in (5) (solid) are overlaid with the estimated mean of $\widehat{\rho}_{\varepsilon,n}(\theta)$ (dashed) based on the simulated sample. The right plot depicts the squareroot of the estimated mean squared error. Note that $\rho(\pi/4)$ is the upper tail dependence coefficient, which is an appropriate and simple measure of extreme dependence for this symmetric model.

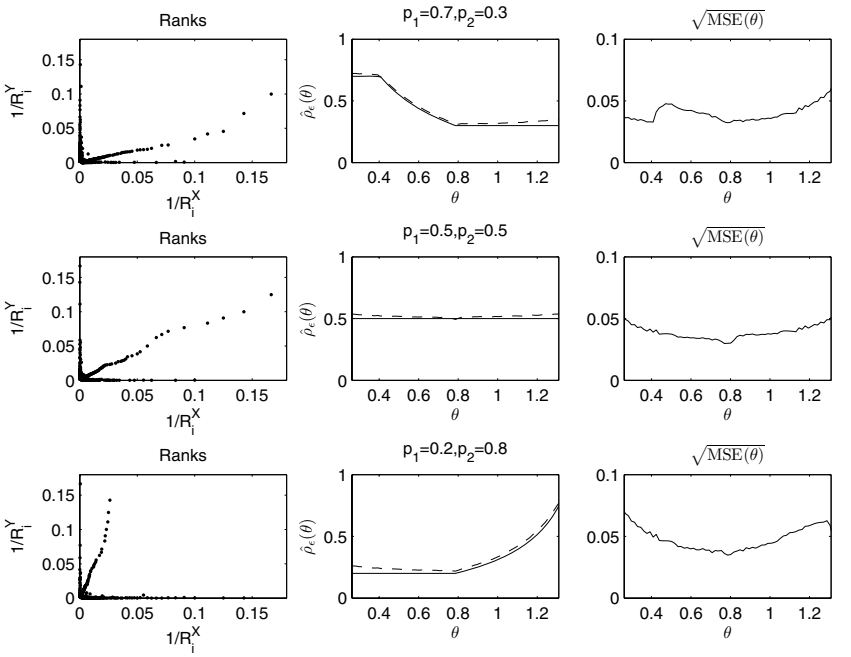

Fig. 6 Simulated asymmetric Pareto model with $\rho(\theta/4) = 0.3$ (upper row), $\rho(\theta/4) = 0.7$ (middle row), $\rho(\theta/4) = 0.9$ (lower row).
Left column: Plots of ranks $(1/R_i^X, 1/R_i^Y)$, with points close to $(1,1)$ truncated.
Middle column: Plots of $\widehat{\rho}(\theta)$ (dashed) overlaid with true function $\rho(\theta)$ (solid).
Right column: Estimation error in terms of $\sqrt{\mathrm{MSE}(\theta)}$.

Example 4. [Asymmetric Pareto model: continuation of Example 2]
In Figure 6 we simulated this model for $n = 10\,000$ iid observations of (X,Y) with $\varepsilon = 1/200$ 100 times. The three sets of plots on the three rows correspond to the cases: $(p_1\,, p_2) = (0.7\,, 0.3)$, $(p_1\,, p_2) = (0.5\,, 0.5)$ and $(p_1\,, p_2) = (0.2\,, 0.8)$. On each row the left plots contain ranks $\left(1/R_i^X, 1/R_i^Y\right)$, $1 \le i \le n$ of a simulated sample of size 10 000. The true functions $\rho(\theta)$ in (5) (solid) are overlaid with the estimated mean of $\widehat{\rho}_{\varepsilon,n}(\theta)$ (dashed) based on the simulated sample. The right plot depicts the squareroot of the estimated mean squared error.

In the first row of plots, ρ is larger for small θ than for large θ; this is reflected by the left plot in which the violation of independence can be seen to be more severe below the diagonal. In the second row of plots, ρ is constant; which is reflected by having a portion of extreme points lined up on the diagonal in the left plot. The third row of plots is the converse situation to the first row, which is reflected by the pattern of extreme points above the diagonal. This is an example of a situation where the tail dependence coefficient does not convey a good picture of extreme dependence, in that $\rho(\pi/4)$ is not sufficient to describe the full dependence structure of this model.

4 High-frequency Financial Data

We have tick-by-tick data of the *Trades and Quotes* database, in terms of trading times [in seconds] and prices [in 1 cent units] of three stocks traded between February and October 2002 on NYSE and Nasdaq. The stocks are General Motors (GM) from NYSE, and Intel and Cisco both from Nasdaq. One major difference between the two stock markets is that on NYSE trading is made on the floor while Nasdaq has electronic trading. We shall analyze the extreme dependence between the three stocks using the tail dependence function ρ. A study with focus on bivariate dependence structures on FX spot data has been performed by Breymann et al. (2003) and Dias & Embrechts (2003). Also, FX spot data was studied within the concept of multivariate regular variation in Hauksson et al. (2001). For cleaning and deseasonalizing our data we mainly follow the methods applied in these papers; see also there for further references. In these papers *parametric bivariate copulas* were fitted to FX spot data in both non-extreme and extreme regions. Our study considers the extremal dependence for stock data, which is estimated *nonparametrically*. One main difference between stock data and FX spot data is that FX spot data is traded 24 hours per day. In contrast, NYSE for instance, has regular opening hours between 9.30 and 16.00 on working days. This introduces additional complexity into our data analysis, and we have to deal with this problem.

When dealing with extremes it is of importance to use as much data as possible, since extremes are consequences of rare events. However, we can not simply use the full samples of all stocks as each single time series is not stationary and, even worse, for high-frequency data the different time series are not synchronized. As a remedy for the non-synchronous data we take subsamples of logreturns on specific timescales. If one chooses a relatively high frequency, one is confronted with the problem that tick prices are discrete, and also microstructure noise effects can enter. We chose 5 minutes logreturns as the lowest frequency, thus avoiding microstructure noise effects.

There are a number of issues which appear when dealing with high-frequency data and we will describe them in turns.

4.1 Cleaning the Data

A full sample path of stock data contains a huge amount of information. At Nasdaq there is almost a trade every second. However, some ticks are false, mostly due to fake quotes and decimal errors.

To be able to continue the analysis one has to clean the data. This is done by filtering the data and removing values that differ too much from their neighboring values in the sense of logarithmic differences. Also, sometimes false values may come in clusters, which one also has to deal with. The selection of thresholds for removing a bad tick was done by visually inspecting the time series before and after the cleaning. When a false tick was observed it was replaced by a value based on

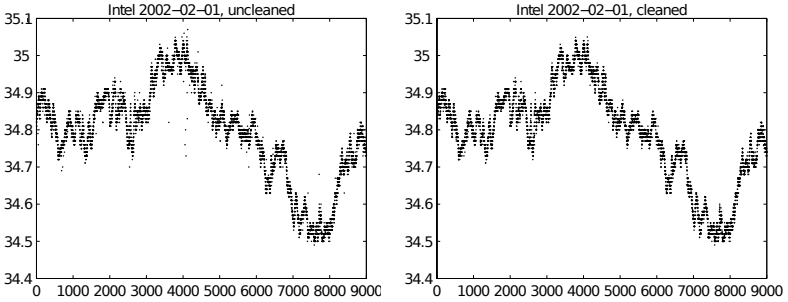

Fig. 7 Intel ticks during 9:35 to 11:45 on February 1, 2002. Left: Raw data. Right: Data cleaned from false ticks as described in Section 4.1.

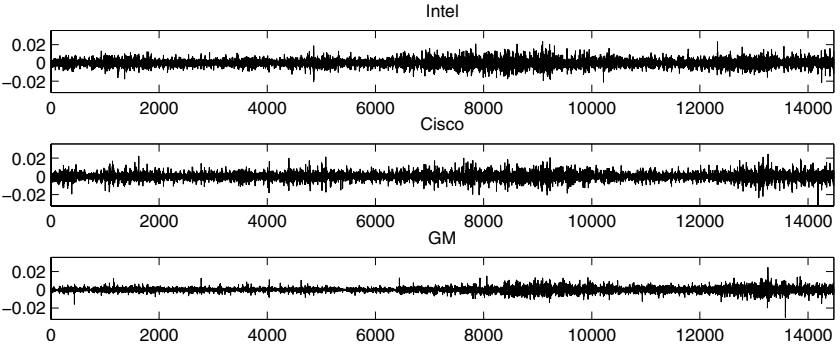

Fig. 8 Synchronized 5 minutes logreturns for Intel, Cisco and GM between 9:35 and 16:00 during February 1 to October 31, 2002

linear interpolation with its neighbors. In this way less than one percent of the data was removed.

The thresholds for logarithmic difference were set to 0.1% for Intel and Cisco and to 0.2% for GM, respectively. The reason for different thresholds is that Intel and Cisco are traded at a much higher frequency. A result of the cleaning procedure can be seen in Figure 7. We repeated our analysis after altering the thresholds slightly. However, this sensitivity analysis did basically not change the results.

When dealing with information from a stock exchange one is faced with the problem that they do not trade for 24 hours resulting in a gap of information, when the stock market is closed over night. However, Nasdaq and NYSE have off-hour trading, but prices behave differently than prices during the regular opening times as the trading rules differ. To obtain synchronised data we only considered the stocks between 9:35 to 16:00 from Monday to Friday using the previous tick method, which results in 77 five minutes logreturns per day. Also, there were a couple of holidays where no data were available. Finally we had 14 476 synchronized observations (5 minutes logreturns) for each stock, which we plot in Figure 8

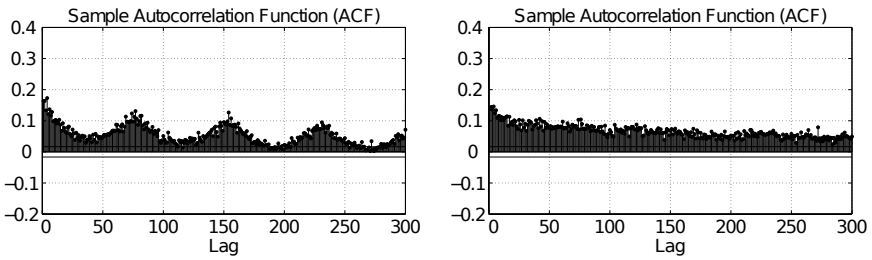

Fig. 9 Autocorrelation function of squared 5 minutes logreturns for Intel.
Left: Original data: Visible is the cycle of 77 lags indicating daily seasonality.
Right: Deseasonalized data as in (15) based on daily seasonality.

4.2 Deseasonalizing the Data

When investigating the 5 minutes logreturns closer one can detect seasonality in the data. In Figure 9, we depict the autocorrelation of the squared logreturns for Intel. Here one can see the daily seasonality. A comparison to the FX data in Breymann et al. (2003) shows that FX data have a much clearer weekly seasonality.

To be able to remove the seasonality, there are two main approaches. The first one is to time-change the logreturns to a business clock instead of the physical clock. The second is to use volatility weighting. We chose the second one as it is not clear how to choose a business clock for multivariate time series.

Volatility weighting divides a period (we first take a week) into several smaller subperiods and then estimates the seasonality effect in each subperiod in terms of volatility. Then each subperiod is deseasonalized separately by devolatization. We chose 5 minutes intervals as subperiods. This means that our observed returns, $\tilde{x}_t$, is a realization of the process

$$\tilde{x}_t = \mu + v_t x_t.$$

where x_t are the deseasonalized returns, μ is a constant drift and v_t is the seasonality coefficient (volatility weights), estimated by

$$\widehat{v}_\tau = \sqrt{\frac{1}{N^\tau}\sum_{i=1}^{N^\tau}(\tilde{x}_{t_i+\tau})^2}. \tag{14}$$

Here N^τ is the number of weeks, during which we have observed our stocks in the given subperiod $\tau \in \{0,5,10,15,\ldots\}$ (in minutes), and t_i denotes the start of week i which always is on Monday at 9:35. Also, τ has to be corrected for nights and weekends. We estimate μ with the sample mean $\widehat{\mu}$ of the logreturns. Hence, the deseasonalized 5 minutes logreturns are

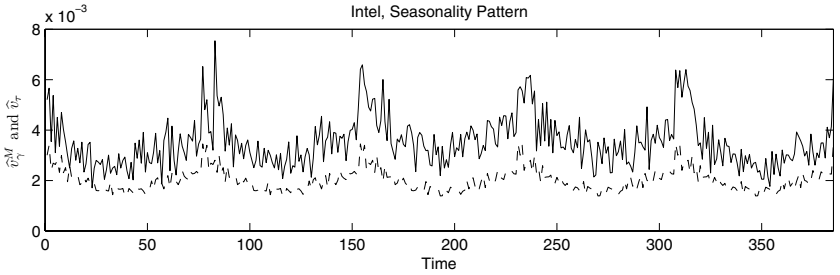

Fig. 10 Volatility weights $\widehat{v}_\tau$ and $\widehat{v}_\gamma^M$ for 5 minutes Intel logreturns: weekly estimated by (14), $\widehat{v}_\tau$, (solid) and daily robust estimated by (16), $\widehat{v}_\gamma^M$, (dashed).

$$x_t = \frac{\tilde{x}_t - \widehat{\mu}}{\widehat{v}_t}. \tag{15}$$

However, as we only have about 40 weeks the estimated volatility weights are quite noisy, see Figure 10. This is due to the fact that single large values can dominate $\widehat{v}_\tau^2$: the mean taken over 40 weeks is not sufficiently smooth.

To overcome this problem we first assume a daily seasonality instead of the weekly. This can be motivated by the fact that the different days do not seem to differ to a higher degree; see Figure 10. However, single large values still dominate the volatility weights, which is unsatisfactory.

Consequently, we use a robust estimator based on the median and absolute values:

$$\widehat{v}_\gamma^M = \text{median}_{i=1,\ldots,N^\gamma} |\tilde{x}_{t_i+\gamma}|. \tag{16}$$

Here N^γ is the number of days, during which we have observed our stocks in the given subperiod $\gamma \in \{0,5,\ldots,385\}$ (in minutes). We can now observe the stylistic pattern of the autocorrelation of squared logreturns in Figure 9 for our deseasonalized time series using the robustly estimated volatility weights.

The depicted volatility weights can be seen in Figure 10. One can clearly see that trading is more intense at the beginning and at the end of a day. We also observe that the robustly estimated volatility weights are much more stable. The deseasonalization removes seasonality in the squared logreturns, which are right skewed, hence the difference in magnitude for the two estimation methods.

When comparing the two different deseasonalization methods the robust one leaves more larger absolute values in the data, which occur in low trading time. The non-robust version decreases them as large values contribute much more to the volatility weights. Hence, the non-robust version of the deseasonalization makes the time series smoother than the robust method does.

Table 1 AIC-based values for (r,m,p,q,ν) and t and normally distributed residuals with corresponding likelihood; the last column presents our model.

Stock	t	normal	our model
Intel	55531 (2,3,5,2,8.2)	55869 (2,3,4,1,-)	55534 (0,1,5,2,8.2)
Cisco	55805 (0,1,3,3,6.5)	56596 (0,1,4,0,-)	55807 (0,1,1,1,6.5)
GM	57343 (2,2,5,1,5.5)	58467 (1,2,3,1,-)	57355 (0,3,1,1,5.5)

4.3 Filtering the Data

Because of the dependence, which we have observed in the autocorrelation for the squared logreturns, we will assume a stochastic volatility model for each stock. We model the mean by an ARMA process and use the standard GARCH(p,q) model for the martingale part. The model selection is based on the AIC criterion, the results are summarized in Table 1.

We model the logreturns for different equidistant frequencies by

$$x_t = \mu_t + \sigma_t z_t$$

with $\mu_t = c + \sum_{i=1}^{r} \phi_i x_{t-i} + \sum_{i=1}^{m} \theta_i \varepsilon_{t-i}$ and $\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \sigma_{t-i}^2 + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i}^2$, where $\varepsilon_t = \sigma_t z_t$. We model the z_t by a standardnormal or a student-t distribution with ν degrees of freedom. The overall fit of the model was assessed by a residual analysis. We applied the Ljung-Box test for serial correlation, where we tested the residuals and the squared residuals, and the Kolmogorov-Smirnov test for goodness-of-fit of the normal and student-t distribution.

As we only have logreturns for 9:35-16:00 Monday to Friday we will make an error if we fit the time series model to our data without taking the missing values into account. We have used three different approaches to circumvent this problem:

(1) We (wrongly) fit the ARMA-GARCH model directly to the deseasonalized data, ignoring the missing observations during the nights completely.
(2) We estimated the logreturns during the nights by 5 minutes logreturns using the (wrong) square root scaling (the correct but complicated scaling constants have been calculated by Drost & Nijman 1993). Then we deseasonalize and fit the time series.
(3) We fit different MA(1)-GARCH(1,1) models for each day. In this case we used the estimated volatility of the previous day as the initial value.

One comment to the second approach is that the deseasonalized nightly logreturns should have the same distribution as the deseasonalized daily logreturns. We have tested this assumption via QQ-plots with bootstrapped confidence interval. Using ordinary bootstrap we can conclude that the deseaonalized nightly logreturns do not have the same distribution as the deseasonalized daily logreturns. However, as we have dependence in our time series one should use a bootstrap method which takes this into consideration. Using block bootstrap we can not reject the hypothesis that

Table 2 Estimated α by the Hill estimator for the loss region.

Stock	Intel	Cisco	GM
$\widehat{\alpha}$	5.6	4.36	4.3

the deseaonalized nightly logreturns have the same distribution as the deseasonalized daily logreturns.

If we compare the methods (1)-(3) we conclude that the first and second behave very similar with respect to the parameter estimation. For the third method this estimation was difficult. Even if we only use a three parameter model the estimation is not stable. Based on the Ljung-Box test for serial correlation, both residuals and squared residuals, the two first methods out-perform the third. Also, for the final result in Section 4.4 the outcome is similar. We concluded that the error of using a false approach (among these three) is minimal and concentrated on the first method for simplicity.

In Table 1 we have selected the model by AIC criteria, also giving the likelihood of the selected model. The selected optimal order of the ARMA model $m, r \in (0, \ldots, 5)$ and the order of the GARCH model $p, q \in (0, \ldots, 6)$ for normal and also the degree of freedom ν for t-distributed innovations are given in the first two columns.

As we want to keep the number of parameters as low as possible, we performed a sensitivity analysis based on the likelihood of the model. In this way we found the model given in column 3 of Table 1, which we will use in the sequel. Our analysis also confirmed the common knowledge that residuals are heavy-tailed; i.e. the t-distribution outperforms by far the normal distribution.

Concerning the Ljung-Box test, we could not reject independence of the residuals or the squared residuals for all time series. In Table 3 we show the p-values for a selection of lags for squared logreturns. We have also looked at the p-values up to 50 lags. However, all time series failed the Kolmogorov-Smirnov test, actually for all models presented in Table 1.

Diagnostic tools from extreme value theory (see e.g. Embrechts et al. 1997, Section 6.1) show, however, clearly that all three filtered time series are heavy-tailed. Consequently, we model the far out distribution tail of all residuals as regularly varying and estimate the tail index α by the Hill estimator (see e.g. Embrechts et al. (1997), Section 6.4). We summarize the result in Table 2.

Due to the devolatizaton a 10 minutes logreturn is obtained as a linear combination of two 5 minutes logreturns and so logreturns should have the same tail-parameter for different frequencies. However, for higher timescales the tail-parameter increases slightly, even if one compares the filtered 5 minutes logreturns with 45 minutes, but still remains heavy-tailed. This is well known and reported, for instance, in Müller et al. (1998).

We have also investigated the cross-correlation between the stocks. In Table 4 we display the first four lags. The other lags were smaller in absolute magnitude.

From Table 4 one can see that GM tends to follow Intel and Cisco more than vice versa. A formal test for uncorrelation of two time series tests this hypothesis for each specific lag based on asymptotic normality of the cross-correlation function (see e.g.

Table 3 p-values from the Ljung-Box test of filtered squared residuals. We have tested 1, 5, and 10 lags of the 5 minutes logreturns.

Stock	1	5	10
Intel	0.39	0.06	0.07
Cisco	0.74	0.86	0.77
GM	0.92	0.96	0.84

Table 4 Cross-correlation of the first four lags for the filtered 5 minutes logreturns.

Stocks	-4	-3	-2	-1	0	1	2	3	4
Intel-Cisco	-0.01	0.01	0.01	0.05	0.56	0.03	0.02	0.01	-0.00
Intel-GM	0.02	0.02	0.05	0.04	0.35	0.02	0.01	-0.00	-0.02
Cisco-GM	0.03	0.01	0.04	0.04	0.33	0.03	0.01	-0.00	-0.02

Brockwell & Davis 1987, Theorem 11.2.2) The uncorrelation hypothesis is rejected if the corresponding estimate has absolute value larger than 0.017.

For the 15 minutes data, there is some cross-correlation between GM and Intel and GM and Cisco for the first lag, but none significant between Intel and Cisco. For the 30 minutes data, there is no significant cross-correlation at all.

Such tests have to be interpreted with caution for various reasons. First of all there is the usual problem that a test should be performed not only on each lag separately. Furthermore, the amount of high-frequency data is so large that a formal test rejects already for very small cross-correlation: for 5 min the rejection level is 0.017, for 30 min it is 0.042.

4.4 Analyzing the Extreme Dependence

Recall the estimator $\widehat{\rho}_{\varepsilon,n}$ from (12), where $\varepsilon = t/n$ represents the proportion of upper order statistics used for the estimation, which itself has to be estimated; cf. Remark 4(iii). The estimation of ε involves in extreme value statistics a variance-bias tradeoff; i.e. it is tricky and time-consuming, but important. We have used two approaches.

Firstly, by plotting the estimated tail dependence function for different choices of ε visual inspection clearly showed the influence of the variance/bias, when using different thresholds. For high threshold, i.e. small ε, the estimated tail dependence function was rather rough showing the high variation of the estimator. When decreasing the threshold the estimated tail dependence function became very smooth, which we interpreted in analogy to tail index estimation as a bias.

Secondly, we studied plots of $\widehat{\rho}_{\varepsilon}(\theta)$ as a function of ε for fixed θ. This was done for $\theta = \pi/4 \pm 0, \pi/12, \pi/6$. Here we looked for regions where $\rho(\theta)$ was stable. The case $\theta = \pi/4$ can be seen in Figure 11. We want to mention that for other choices of θ the stability plots were not equally convincing.

As a result of our diagnostics we fixed $\varepsilon = 1/650$, which represents about 4.5% of the data. In Figure 12 we can see the resulting estimated tail dependence function.

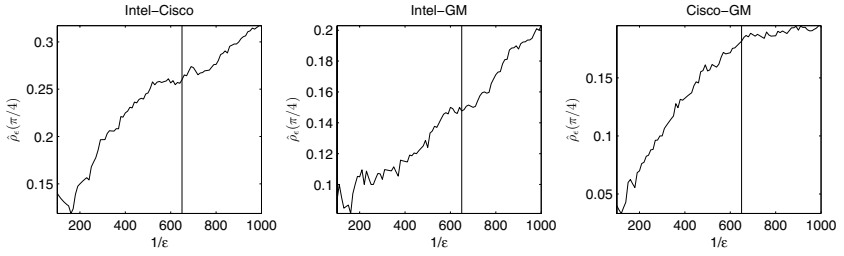

Fig. 11 For the 5 minutes logreturns: $\widehat{\rho}_\varepsilon(\pi/4)$ as a function of ε for $\varepsilon = 1/100, \ldots, 1/1000$.

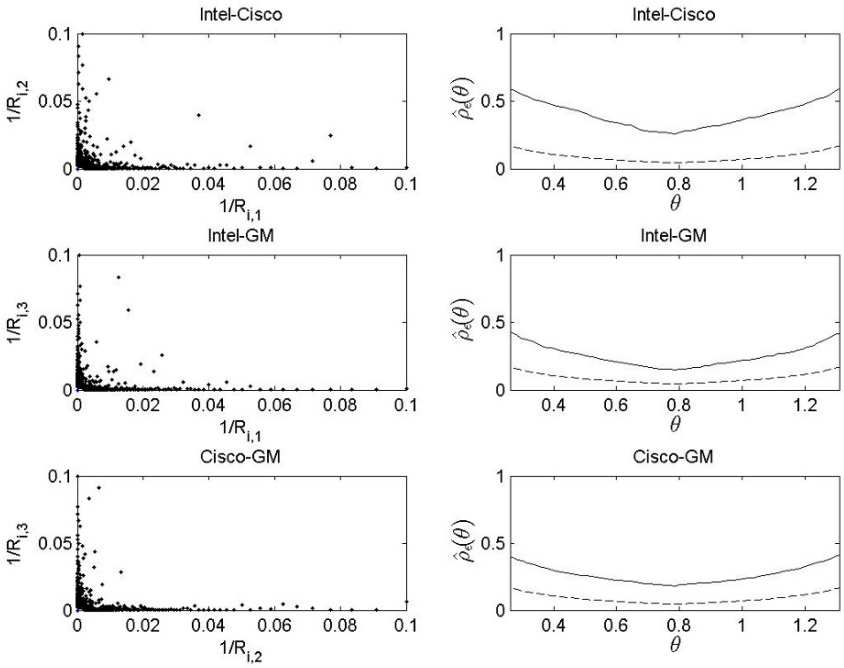

Fig. 12 Left plots: $1/R_{i,j}$, where $R_{i,j} = \text{rank}(-X_{i,j})$. Right plots: Estimators of $\rho(\theta)$ (solid line). For sake of reference we have also plotted the expected dependence $E(\widehat{\rho}_{\varepsilon,n}(\cdot))$ from (13) for independent samples (dashed line).

We conclude that for all bivariate combinations of our data tail dependence can be modelled symmetric and is significantly stronger than for the independent case. Not surprisingly, dependence is highest between Intel and Cisco, presumably due to branch dependence, besides being both traded at Nasdaq. The dependence of GM and Cisco is slightly higher than of GM and Intel. The symmetry in the dependence reflects that we have three major stocks and can also be viewed as underlying market dependence. We also notice that the estimated tail dependence function looks similar as the tail dependence function of a bivariate extreme value distribution with a Gumbel copula.

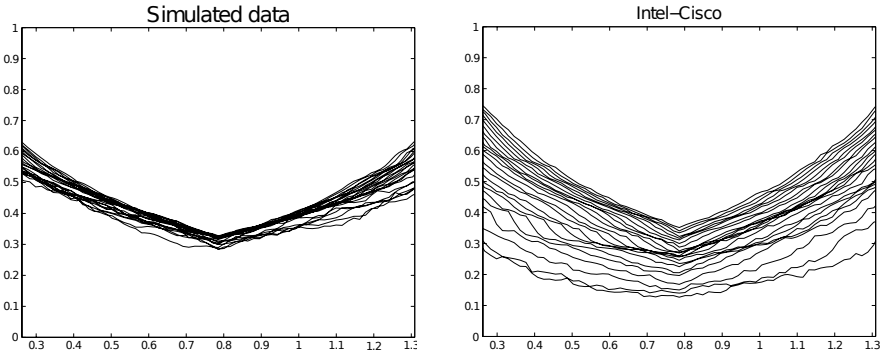

Fig. 13 Estimated tail dependence function $\widehat{\rho}_\varepsilon(\theta)$ for different ε (1/200 to 1/1200). Left: Simulated data from the t-Gumbel model with parameters estimated from the Intel-Cisco filtered 5 minutes logreturns. Right: Filtered 5 minutes logreturns, Intel-Cisco.

It is now tempting to fit a distribution with t marginals (a common model in econometrics, called the t-GARCH) with degree of freedom from Table 1 and a Gumbel copula (cf. Example 1,3) pairwise to our data or even to the three-dimensional sample. We know already from Section 4.3 that the t distribution is not a good model for the marginals. However, our concern is now for the extreme dependence structure, and it turns out that the Gumbel copula, although an extreme value copula, is not a valid model. The dependence structure in our data is far more complex. This can be illustrated by viewing the tail dependence function for different ε (1/200 to 1/1200) compared to data simulated from the above t-Gumbel model with the same sample size, presented in Figure 13. Recall from Example 1 that the Gumbel coupla gives tail dependence function

$$\rho(\theta) = \frac{1+\cot\theta-(1+(\cot\theta)^\delta)^{(1/\delta)}}{1\wedge\cot\theta}, \quad 0<\theta<\pi/2.$$

We estimated δ from the upper tail dependence coefficient $\widehat{\rho}_\varepsilon(\pi/4) = 2-2^{(1/\widehat{\delta})}$, the value of the estimated tail dependence function at $\pi/4$. We obtained $\widehat{\delta} = 1.25$. Now we generate a sample of the same size as the 5 minutes logreturns with a Gumble($\widehat{\delta}$) copula and t-distributed marginals. We compare the estimated tail dependence functions for different ε and present the results in Figure 13.

Notice that the simulated data behave much more stably with respect to changes of ε, while the real data reacts heavily on such changes. Using a parametric model such as the Gumble copula would only be an approximation based on one given threshold.

Table 5 Correlation change for different timescales.

Stock	Intel-Cisco	Intel-GM	Cisco-GM
5	0.56	0.35	0.33
15	0.62	0.41	0.38
30	0.65	0.43	0.39
45	0.66	0.46	0.41

4.5 Different Timescales

As a result of our statistical analysis of the marginal data, the one-dimensional logreturns exhibit Pareto-like tails. If the stocks came from a three-dimensional exponential Lévy process with appropriate dependence structure, then the extreme dependence would be the same for all time scales, i.e. 5 minutes logreturns of the three stocks would have the same dependence structure as daily logreturns. This applies in particular to extreme dependence and is in accordance with Proposition 4. Note that our data do not satisfy the independence condition of Proposition 4. Extreme value estimates, however, often extend properties from independent data to dependent data.

We shall at least perform a statistical test to our data, whether there is a change in the extreme dependence on different timescales by analyzing logreturns of 5, 15, 30 and 45 minutes frequencies.

To this end we performed the same filtering steps as for the 5 minutes logreturns again for the 15, 30 and 45 minutes logreturns obtained from the raw data. Then we fitted a MA(1)-GARCH(1,1) model with student-t distributed residuals to the deseasonalized data of the 15, 30 and 45 minutes logreturns. For the 5 minutes logreturns we keep the model from Section 4.4.

For the 15, 30 and 45 minutes logreturns we applied the Ljung-Box test for serial correlation, where we tested the residuals and the squared residuals, and the Kolmogorov-Smirnov test for goodness-of-fit of the student-t density. Observe that the degrees of freedom is not the same as in Table 1 for the different timescales. All the filtered time series passed the Ljung-Box test and the filtered 30 and 45 minutes logreturns passed the Kolmogorov-Smirnov test.

For the residuals we again estimate the dependence between the different stocks.

A comparisons of the linear dependence for different timescales is presented in Table 5. Here we can see that the correlation increases for higher frequencies; this effect is well-known and also called the Epps effect; see Zhang (2006).

Next we estimate the tail dependence function for the different frequencies. To compensate for the increasing lack of data for low frequencies, the ε is always chosen so that ε times the number of observations is the same for all frequencies. Hence we always consider the same quantile.

As the sample of the 45 minutes logreturns is only about 10 percent in size of the 5 minutes logreturns, they set the standard for the other frequencies. We increased the threshold $1/\varepsilon$ until the estimated tail dependence function behaved stably for the 45 minutes logreturns. We also studied a plot of $\widehat{\rho}_\varepsilon(\theta)$ for various values of θ,

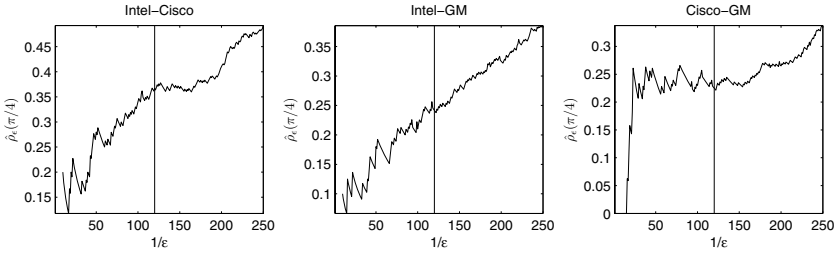

Fig. 14 For the 45 minutes filtered logreturns we depict $\widehat{\rho}_\varepsilon(\pi/4)$ as a function of ε for $\varepsilon = 1/10, \ldots, 1/250$.

when altering ε. For $\theta = \pi/4$ the result can be seen in Figure 14. Finally, we chose $\varepsilon = 1/120$, which represents about eight percent of the data. We want to remark that, in view of Figure 11, we presumably introduced a bias into our estimation.

Also, by using straight forward bootstrap techniques one can present bootstrap confidence intervals. In Figure 16 we depict the tail dependence function for Intel-Cisco on the timescales 5 minutes and 45 minutes.

From Figures 15 and 16 we can conclude that the tail dependence is approximately the same for different timescales. This also holds for different ε but there are some variations if we increase the threshold. If we lower the threshold, then the similarities between the different timescales become more pronounced. We recall that in Table 2 on p. 9 in Breymann et al. (2003) the tail dependence coefficient $\rho(\pi/4)$ is estimated via a parametric model for different timescales for DEM (Deutsche Mark) and JPY (Japanese Yen). Even for the unfiltered data in that paper the estimator for $\rho(\pi/4)$ looks stable. If we increase the timescale to, for instant, two hours the extreme dependence starts to deviate unless we lower the threshold and use as much as 15% of the data. This is consistent with the result on high-frequency FX data reported in Hauksson et al. (2001). Hence, the two different asset classes seem to share the same time scaling for extreme dependence. As pointed out earlier in this section, the time scaling is also explained from a theoretical point of view via Proposition 4.

From the above analysis we conclude that we can estimate extreme dependence for lower frequencies by estimating it for high frequencies, where enough data are available.

Another possibility to achieve a more stable estimation procedure is to use subsampling, based on different samples of the same frequency, obtained by time shifts. We performed the estimation separately for each subsample and, at the end, averaged over all estimated tail dependence functions. The subsamples proved to be very stable in the basic statistics, the estimates for the ARMA and GARCH parameters, and also the properties of the residuals. However, for the estimated tail dependence functions we cannot report significant improvement, in particular, when compared to the tail dependence function estimated from higher frequencies.

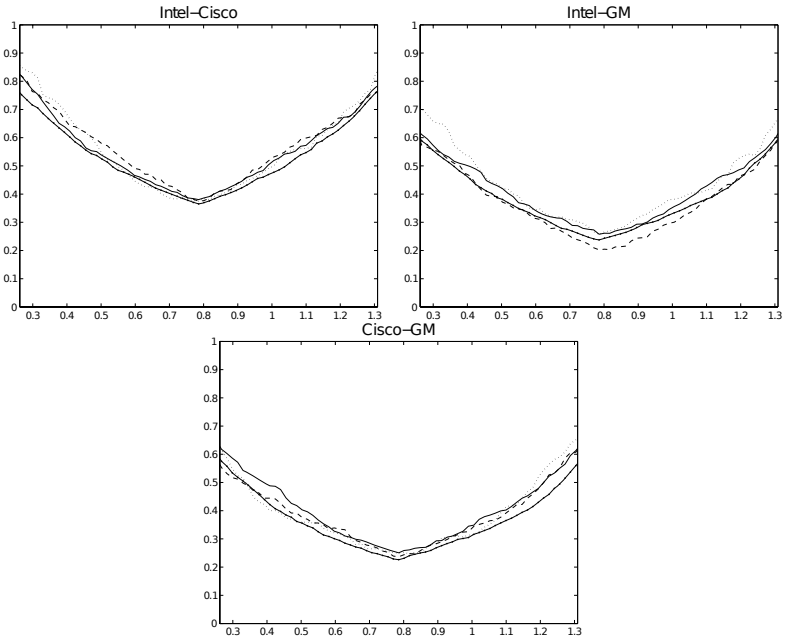

Fig. 15 Estimated tail dependence function $\widehat{\rho}_\varepsilon(\theta)$ of filtered logreturns for different frequencies. Five minutes (straight-dotted), 15 minutes (straight), 30 minutes (dashed) and 45 minutes (dotted).

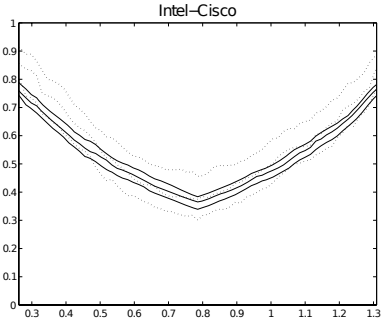

Fig. 16 Estimated tail dependence function $\widehat{\rho}_\varepsilon(\theta)$ with bootstrap confidence intervals (100 resamples) of filtered logreturns for timescale 5 and 45 minutes. Five minutes with corresponding confidence intervals (straight) and 45 minutes with corresponding confidence intervals (dotted).

4.6 Dependence Under Filtering

Recall that we have in principle prices which are multiples of one cent, but there are values our logreturn will never take. Moreover, we have an unnaturally large amount of zeroes and small values. However, concerning extreme dependence we can rest assured that this does not affect the tails.

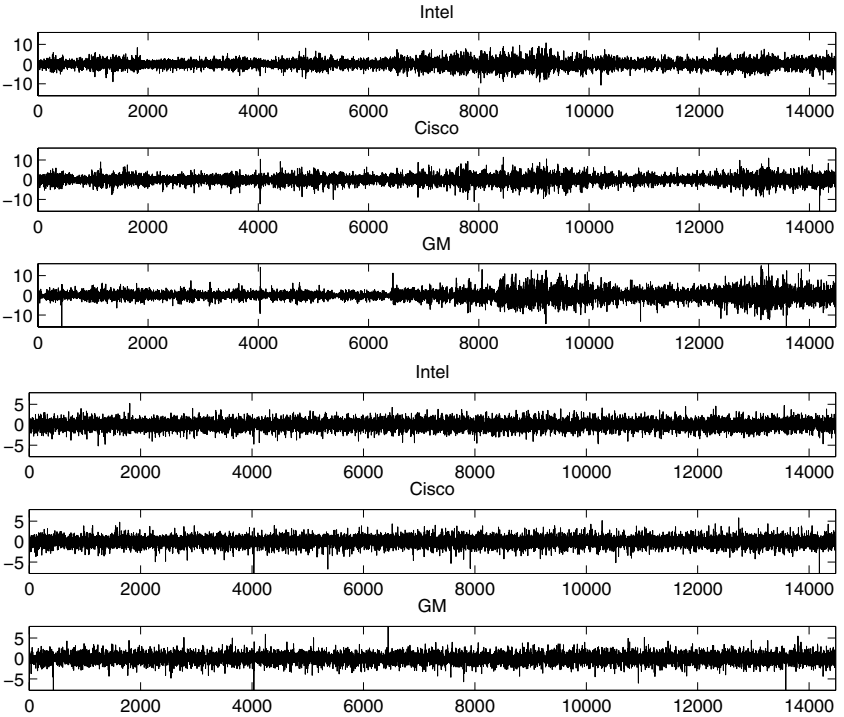

Fig. 17 Upper three plots: Deseasonalized 5 minutes logreturns modelled by daily seasonality and coefficients estimated by the robust method (16). Lower three plots: GARCH filtered logreturns based on our model in Table 1. Compare to the raw data in Figure 8.

Table 6 Estimated correlation for the 5 minutes logreturns for different steps in the data analysis.

Stocks	Original	Deseasonalized	Filtered
Intel-Cisco	0.57	0.56	0.56
Intel-GM	0.36	0.36	0.35
Cisco-GM	0.33	0.33	0.33

Now we shall investigate, how the dependence structure has changed during the filtering steps. In Table 6 we can see the correlation between the 5 minutes logreturns for the different steps of the filtering.

It is satisfactory to see that the different filtering steps have obviously not changed the correlation and hence not changed the linear dependence between the different stocks. This also holds for other timescales.

Now we turn to an account of extreme dependence before and after filtering. When examining the logreturns in Figure 8 one can clearly see the dominating volatile periods. The same holds for Figure 17. Taking the same ε for the raw data and the filtered returns yields for the raw data an over representation of the volatile periods. This implies that one would consider in fact only a much smaller time period for the extreme value analysis. So theoretically, there is no reason, why extreme dependence

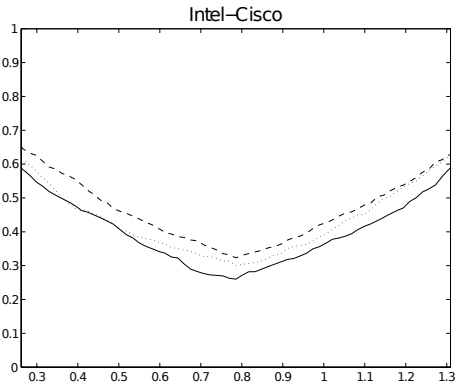

Fig. 18 Estimated tail dependence function $\widehat{\rho}_{\varepsilon}(\theta)$ for 5 minutes Intel logreturns. Dashed: Unfiltered data. Dotted: Deseasonalized data. Solid: GARCH filtered data.

before and after filtering should be similar. In Figure 18 we have plotted the estimated tail dependence function for Intel and Cisco after each filtering step for 5 minutes logreturns. We have used the same ε as in Section 4.4 and the same ε for the different filtering steps. One can see that there seems to be only a small difference in magnitude, not in shape. This also holds for different choices of ε and different timescales. Consequently, for our data the rather complicated filtering procedure seems to be obsolete for a realistic account of the extreme dependence.

5 Conclusion

We have introduced a new estimator for the tail dependence function, which is tailor made to assess the extreme dependence structure in data. As it measures dependence in every direction it is in principle also able to measure extreme dependence for data with asymmetric dependence structure. We show the performance of this function for high-frequency data for varying frequencies.

After giving some theoretical results, which are important in the high-frequency context, we clean the data carefully and perform some basic statistics. We then show the tail dependence function at work for our data and estimate extreme dependence for high-frequency stock data.

We have investigated the extremal dependence between Intel, Cisco and GM for different time scales. We can conclude for the filtered data:

- All three stocks have heavy tails. Within the 5,10,15,45 minutes frequencies we observed that a lower frequency gives lighter tails.
- We can work with the hypothesis that the square root scaled deseasonalized nightly logreturns have the same distribution as the deseasonalized daily logreturns.
- There is (weak) cross-correlation between the stocks for frequencies of up to 30 minutes, it disappears for lower frequencies.

- The extreme dependence is symmetric which means that the stocks influence each other to the same degree. This can be interpreted as market dependence.
- The IT stocks (Cisco and Intel) have stronger dependence indicating branch dependence.
- Extreme dependence is there, but moderate. We have the same extreme dependence for different timescales. This is consistent with the result on high-frequency FX data reported in Hauksson et al. (2001). Hence, the two different asset classes seems to share the same time scaling for extreme dependence. The time scaling is also explained from a theoretical point of view via Proposition 4.
- The filtering steps do not alter the extreme dependence to a high degree.
- Higher correlation does not necessarily lead to stronger extreme dependence.

Our analysis shows again that extreme value theory has to be applied with care. To obtain a realistic picture about the extreme dependence structure in real data it is not enough to describe it by one single number. Another obvious lesson to draw from our analysis is that it is important to use reference results such as simulations from exact models. Moreover, a message, which we can not repeat too often, one should be careful when selecting the threshold.

Acknowledgements E.B. takes pleasure to thank Patrik Albin for generously providing the version the proof of Proposition 4, Holger Rootzén for fruitful discussions, Catalin Starica for suggesting the median filter and the Stochastic Center, Chalmers for a travelling grant. He also thanks the Center for Mathematical Sciences of the Munich University of Technology for a very stimulating and friendly atmosphere during a much needed research stay.

References

Bingham, N.H., Goldie, C.M. & Teugels, J.L. (1987) *Regular Variation.* Cambridge University Press. Cambridge.

Breymann, W. Dias, A. & Embrechts, P. (2003). Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance* **3**: 1–14.

Brockwell, P.J. & Davis, R.A. (1991). *Time Series: Theory and Methods,* 2nd edition. Springer, New York.

Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, London.

Dias, A. & Embrechts, P. (2003). Dynamic copula models for multivariate high-frequency data in finance. *Preprint, ETH Zurich.*

Drost, F. C. & Nijman, T.E. (1993). Temporal aggregation of GARCH processes *Econometrica* **61**: 909–927.

Embrechts, P. (Ed.) (2000). *Extremes and Integrated Risk Management.* UBS Warburg and Risk Books.

Embrechts, P., Klüppelberg, C. & Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance.* Springer, Berlin.

Embrechts, P., Lindskog, F. & McNeil, A. (2001). Modelling dependence with copulas and applications to risk management. In: Rachev, S. (Ed.) *Handbook of Heavy Tailed Distributions in Finance.* Elsevier, Chapter 8, pp. 329–384.

Embrechts, P., McNeil, A. & Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In: Dempster, M. and Moffatt, H.K. (Eds.) *Risk Management: Value at Risk and Beyond.* Cambridge University Press, Cambridge.

Hauksson, H., Dacorogna, M., Domenig, T., Müller, U. & Samorodnitsky, G. (2001) Multivariate extremes, aggregation and risk estimation. *Quantitative Finance* **1**(1): 79–95.

Hsing, T., Klüppelberg, C. & Kuhn, G. (2004). Dependence estimation and visualization in multivariate extremes with applications to financial data. *Extremes* **7**: 99–121.

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

Klüppelberg, C. & Kuhn, G. (2009). Copula structure analysis. *J. Royal Stat. Soc., Series B* **71**(3), 737–753.

Klüppelberg, C., Kuhn, G. & Peng, L. (2008). Semi-parametric models for the multivariate tail dependence function - the asymptotically dependent case. *Scand. J. Stat.* **35**(4): 701–718.

Klüppelberg, C., Kuhn, G. & Peng, L. (2007). Estimating the tail dependence of an elliptical distribution. *Bernoulli* **13**(1): 229–251.

Müller U. A., A. Dacorogna M. M. & Pictet, O. V. (1998). Heavy tails in high-frequency financial data. In: R.J. Adler, R. E. Feldman and M. S. Taqqu (Eds.) *A Practical Guide to Heavy Tails: Statistical Techniques for Analysing Heavy Tailed Distributions*, Birkhäuser, Boston, MA, pp. 55–77.

Zhang, L. (2009). Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics.* To appear.