Hsing, Klüppelberg, Kuhn:

Dependence Estimation and Visualization in Multivariate Extremes with Applications to Financial Data

Sonderforschungsbereich 386, Paper 374 (2004)

Online unter: http://epub.ub.uni-muenchen.de/

Projektpartner

Dependence Estimation and Visualization in Multivariate Extremes with Applications to Financial Data

Tailen Hsing * Claudia Klüppelberg † Gabriel Kuhn†

December 3, 2003

Abstract

We investigate extreme dependence in a multivariate setting with special emphasis on financial applications. We introduce a new dependence function which allows us to capture the complete extreme dependence structure and present a nonparametric estimation procedure. The new dependence function is compared with existing measures including the spectral measure and other devices measuring extreme dependence. We also apply our method to a financial data set of zero coupon swap rates and estimate the extreme dependence in the data.

AMS 2000 Subject Classifications: primary: 62G32, 62H12;
secondary: 62E20.

Keywords: Extreme dependence function, nonparametric estimation, financial data analysis.

*Department of Statistics, Texas A&M University, College Station TX 77843 USA, email: thsing@stat.tamu.edu

†Center of Mathematical Sciences, Munich University of Technology, D-85747 Garching, Germany, email: {cklu , gabriel}@ma.tum.de, http://www.ma.tum.de/stat/

1 Extreme dependence structure

One of the general goals of statistical extreme value theory is to understand the behavior of the *extreme* observations in a set of data generated by a random process and how that information can be used to draw inference about the corresponding aspect of the true distribution. Extreme observations here may be very large or very small observations, or more generally, observations in some *rare* set. Some considerable progress has been made in past decades on the statistical inference of extremes. See Coles (2001), Embrechts, Klüppelberg and Mikosch (1997), and Smith (2003). In this paper, we focus on the very large observations in a data set when the observations are multivariate. Specifically, let m be a positive integer and consider an iid sequence of random vectors $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,m})$, $i \in \mathbb{N}$. We are interested in the statistical inference of the joint distribution of the componentwise maxima

$$M_{n,j} = \bigvee_{i=1}^{n} X_{i,j}\,, \qquad 1 \le j \le m\,,$$

for large n. This topic is of relevance in many problems of practical interest; examples can be found in Tawn (1988) (sea levels data), Coles and Tawn (1991) (tidal wave data), Schlather and Tawn (2003) (rainfall data), de Haan and de Ronde (1998) (sea-level and wind-speed data), Hauksson, Dacorgna, Domenig, Müller and Samorodnitsky (2001) (currency exchange rate data), to name a few.

Among the most important problems in multivariate statistical extremes are the description and inference of dependence between the components of $\boldsymbol{M}_n := (M_{n,1}, \ldots, M_{n,m})$ when n is large. For example, in designing an investment portfolio it is crucial to understand the relative behavior of the various assets in the portfolio in the event of large losses so that the risks can be balanced, or in the event of possible floods, it is important to understand of how extreme rainfall leads to dangerously high river levels so that losses of lives can be prevented.

It is well known that the dependence structure of a random vector can be fully captured by the *copula* or *dependence function*. A *copula* C is a multivariate cumulative distribution function (cdf) with standard uniform marginals. The copula C_G of an arbitrary random vector $(X_1, \ldots, X_m)$ with a joint cdf G and marginal cdf's G_j is given by

$$C_G(u_1, \ldots, u_m) = P\left(X_1 \le G_1^{\leftarrow}(u_1), \ldots, X_m \le G_m^{\leftarrow}(u_m)\right), \quad (u_1, \ldots, u_m) \in [0,1]^m \tag{1.1}$$

where $G_j^{\leftarrow}$ denotes the left-continuous inverse of G_j. See Joe (1997) for details. We focus on the copula of $\boldsymbol{M}_n$ for large n. Assume that there exist linear normalizing functions $f_{n,1}, \ldots, f_{n,m}$, such that

$$\lim_{n\to\infty} P\left(M_{n,j} \le f_{n,j}(x_j), 1 \le j \le m\right) = F(x_1, \ldots, x_m), \tag{1.2}$$

where F is a nondegenerate m-variate cdf. Any possible limit cdf F in (1.2) is called an *multivariate extreme value cdf* (mevdf). It can be seen that a cdf F is an mevdf if and only if the marginals F_j, $1 \le j \le m$, are one-dimensional extreme value cdf's (cf. Embrechts et al. 1997) and the copula C satisfies (Joe 1997, Section 6.2)

$$C^t(u_1, \ldots, u_m) = C(u_1^t, \ldots, u_m^t), \quad (u_1, \ldots, u_m) \in [0,1]^m, \ t > 0. \tag{1.3}$$

Any copula C satisfying (1.3) is called an *extreme copula*.

Since applying monotone transformations to the marginals do not change the copula, (1.2) implies that the copula of $\boldsymbol{M}_n$, for large n, can be approximated by that of F and hence approximately satisfies (1.3). By the same token, it is clear that the particular normalizations $f_{n,1}, \ldots, f_{n,m}$ in (1.2) do not play a role in (1.3). Consequently, (1.3) is a very general property for the limiting copula of $\boldsymbol{M}_n$.

It is also known that any extreme copula can be written in the form of the *Pickands representation* (Resnick 1997, Section 5.4):

$$C(u_1, \ldots, u_m) = \exp\left\{ \int_{S_m} \left(\bigwedge_{j=1}^{m} w_j (\ln u_j) \right) \mu(d\boldsymbol{w}) \right\}, \ (u_1, \ldots, u_m) \in [0,1]^m \tag{1.4}$$

where μ is a finite measure on $S_m = \{\boldsymbol{y} \ge 0 : \sum_{i=1}^m y_i = 1\}$ satisfying

$$\int_{S_m} w_j \mu(d\boldsymbol{w}) = 1, \quad j = 1, \ldots, m.$$

Further, by changing the variable $\boldsymbol{w}$ in the integral in (1.4), the extreme copula can be described in infinitely many different but equivalent forms; for instance, Einmahl, de Haan and Piterbarg (2001) adopts the following representation for the case $m = 2$:

$$C(u_1, u_2) = \exp\left\{ \int_{[0,\pi/2]} \left(\frac{\ln u_1}{1 \vee \cot\theta} \wedge \frac{\ln u_2}{1 \vee \tan\theta} \right) \Phi(d\theta) \right\}, \ u_1, u_2 \in [0,1], \tag{1.5}$$

where Φ is a finite measure, called *spectral 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 focal point of this paper is the inference of the copula of F in (1.2), namely the limiting copula of $\boldsymbol{M}_n$, based on a random sample. In view of (1.4), this is equivalent to the inference of the measure μ in the Pickands representation. We will discuss a purely nonparametric approach of estimating the extreme copula. In conjunction, we will introduce a method to visualize extreme tail dependence, a topic which has not received much attention. We believe that simple and effective visualization tools are crucial in this context in order to bridge theory and application. The literature of multivariate extremes has focused almost exclusively on the bivariate case $m = 2$.

See Section 2 for a brief review of the literature of this case. The case $m \geq 3$ in contrast has received little attention. Our approach of estimating dependence can be implemented for any general m. Needless to say the curse of dimensionality is even stronger here than in most other contexts so that the general procedure will not achieve the intended purpose unless enough data are available. We will illustrate our procedures by theoretical computations as well as simulations. We will also apply the results on the analysis of a portfolio of zero coupon swap rates.

Throughout the paper we write $a(u) \sim b(u)$ as $u \to \infty$, if $a(u)/b(u) \to 1$ as $u \to \infty$; we write $a(u) \approx b(u)$ for crude approximations.

2 Measuring bivariate extreme sets

As mentioned, the statistical estimation of F in (1.2) is of substantial interest in applications. There are three main approaches. Coles & Tawn (1991) and Tawn (1988) assume a parametric form for F and approach the estimation problem by maximum likelihood. While the parametric approach is efficient when the model is correct, the conclusion can be grossly misleading if the model is incorrect. The second approach estimates the measure μ in (1.4) based on the empirical measure for the transformed data where the transformation involves parameter estimation on the marginals. Such a procedure is semiparametric in nature and examples of it can be found in Embrechts, de Haan & Huang (2000), Einmahl, de Haan & Sinha (1997), de Haan & Resnick (1977), and de Haan & de Ronde (1998). A completely nonparametric approach for estimating μ was introduced in Einmahl et al. (2001). We next review this approach in detail.

Consider the bivariate case where $m = 2$. Suppose that the $\boldsymbol{X}_i = (X_{i,1}, X_{i,2})$ are iid random vectors with continuous marginal cdf's G_1, G_2. Assume that there exist continuous and nondecreasing normalizing functions $f_{n,1}, f_{n,2}$ such that

$$\lim_{n\to\infty} P\left(\bigvee_{i=1}^{n} X_{i,1} \leq f_{n,1}(x_1),\ \bigvee_{i=1}^{n} X_{i,2} \leq f_{n,2}(x_2)\right) = F(x_1, x_2)\,, \quad x_1, x_2 \in \mathbb{R}\,, \quad (2.1)$$

where F has continuous margins. As explained in Section 1, the copula C_F of F is an extreme copula and it is independent of the normalizations $(f_{n,1}, f_{n,2})$. Hence we consider instead the normalized limit of $(\bigvee_{i=1}^{n} G_1(X_{i,1}), \bigvee_{i=1}^{n} G_2(X_{i,2}))$. Since $G_j(X_{i,j}) \sim$ uniform$[0,1]$, $j = 1, 2$, we have for $j = 1, 2$,

$$\lim_{n\to\infty} P\left(\bigvee_{i=1}^{n} G_j(X_{i,j}) \leq \frac{1}{n}\ln u + 1\right) = u\,, \quad u \in [0,1]\,.$$

Consequently, the following computations yield the copula of F in (2.1):

$$P\left(\bigvee_{i=1}^{n} G_1(X_{i,1}) \le \frac{1}{n}\ln u_1 + 1, \bigvee_{i=1}^{n} G_2(X_{i,2}) \le \frac{1}{n}\ln u_2 + 1\right)$$

$$\to \quad C_F(u_1,u_2) \;=\; \exp\left\{\int_0^{\pi/2}\left(\frac{\ln u_1}{1\vee\cot\theta}\wedge\frac{\ln u_2}{1\vee\tan\theta}\right)\Phi(d\theta)\right\},$$

where the representation (1.5) is adopted in order to be consistent with the presentation of Einmahl et al. (2001). It follows that

$$\begin{aligned} & nP\left(G_1(X_{1,1}) > 1-\frac{x_1}{n} \text{ or } G_2(X_{1,2}) > 1-\frac{x_2}{n}\right) \\ = \;& nP\left(n(\overline{G}_1(X_{1,1}),\overline{G}_2(X_{1,2})) \in ([x_1,\infty]\times[x_2,\infty])^C\right) \\ \to \;& \int_0^{\pi/2}\left(\frac{x_1}{1\vee\cot\theta}\vee\frac{x_2}{1\vee\tan\theta}\right)\Phi(d\theta). \end{aligned} \tag{2.2}$$

Since $P(\cdot)$ is monotone, the discrete index $n\to\infty$ in (2.2) can be replaced by a continuous index $t\to\infty$ and the limit remains the same. On $[0,\infty]^2\backslash\{(\infty,\infty)\}$ define the measures Λ_t and Λ on the Borel σ-algebra of $[0,\infty]^2\backslash\{(\infty,\infty)\}$ by

$$\Lambda_t(\boldsymbol{A}) = tP\left(t(\overline{G}_1(X_{1,1}),\overline{G}_2(X_{1,2}))\in\boldsymbol{A}\right),$$

and

$$\Lambda\left(([x_1,\infty]\times[x_2,\infty])^C\right) = \int_0^{\pi/2}\left(\frac{x_1}{1\vee\cot\theta}\vee\frac{x_2}{1\vee\tan\theta}\right)\Phi(d\theta)\,,\; x_1,x_2\in[0,\infty). \tag{2.3}$$

Note that the latter relation indeed defines a measure since the sets $([x_1,\infty]\times[x_2,\infty])^C$, $0<x_1,x_2<\infty$, form a π-class which generates the Borel σ-algebra of $[0,\infty]^2\backslash\{(\infty,\infty)\}$. It follows from the continuous-index version of (2.2) that for all Borel sets $\boldsymbol{A}\subset[0,\infty]^2\backslash\{(\infty,\infty)\}$ with $\Lambda(\partial\boldsymbol{A})=0$, we have (cf. Resnick 1987)

$$\lim_{t\to\infty}\Lambda_t(\boldsymbol{A}) = \lim_{t\to\infty} tP\left(t(\overline{G}_1(X_{1,1}),\overline{G}_2(X_{1,2}))\in\boldsymbol{A}\right) = \Lambda(\boldsymbol{A}). \tag{2.4}$$

Given an iid sample $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$, where $\boldsymbol{X}_i=(X_{i,1},X_{i,2})$, and a Borel set $\boldsymbol{A}$ in $[0,\infty]^2\backslash\{(\infty,\infty)\}$, an intuitive estimator of $\Lambda_t(\boldsymbol{A})$ is

$$\widetilde{\Lambda}_{t,n}(\boldsymbol{A}) = tP_n\left(t\left(\overline{G}_1(X_{1,1}),\overline{G}_2(X_{1,2})\right)\in\boldsymbol{A}\right) := \frac{t}{n}\sum_{i=1}^{n} I\left(t\left(\overline{G}_1(X_{i,1}),\overline{G}_2(X_{i,2})\right)\in\boldsymbol{A}\right).$$

Furthermore, each $\overline{G}_j(X_{i,j})$ is uniformly distributed on $[0,1]$ and hence can be estimated by $R_{i,j}/n$ where $R_{i,j}$ is the rank of $-X_{i,j}$ among $-X_{1,j},\ldots,-X_{n,j}$. Writing $\varepsilon=t/n$, the estimator $\widetilde{\Lambda}_{t,n}$ is approximated by

$$\widehat{\Lambda}_{\varepsilon,n}(\boldsymbol{A}) = \varepsilon\sum_{i=1}^{n} I\left(\varepsilon\left(R_{i,1},R_{i,2}\right)\in\boldsymbol{A}\right). \tag{2.5}$$

This simple and natural estimator works very well both in theory and in practice. The fact that it does not require estimating the marginal tail distributions eliminates an important source of error in the estimation of tail dependence. Generally speaking, the variance and bias of the estimator increases and decreases with ε, respectively, and ε should satisfy $\varepsilon \to 0$ and $n\varepsilon \to \infty$ in order for consistent estimation to be achieved. A result in Einmahl et al. (2001) shows that the estimator can achieve a quick rate of convergence in estimating $\Lambda(\boldsymbol{A})$ for $\boldsymbol{A}$ of a certain form when ε is chosen properly. See Einmahl et al. (2001), Huang (1992) and Qi (1997) for additional details on the theoretical aspects of this estimation approach.

However, in practice when the procedure is implemented we have to select a suitable ε from the data. This is always a difficult issue. In the examples in the next section, we show how to do this by a practical approach.

3 Inference of dependence through measure determining classes

We continue our discussions from Section 2 and use the notation developed there. To fully estimate the measure Λ, it suffices to estimate $\Lambda(\boldsymbol{A})$ for sets $\boldsymbol{A}$ in a measure-determining class of Λ. There are obviously infinitely many such classes. The key criteria for selecting such a class are that the measures $\Lambda(\boldsymbol{A})$ are easy to interpret, directly useful for describing tail probabilities, and can be estimated efficiently. Below we mention two examples of such classes for the case $m = 2$.

Definition 3.1 *For $\theta \in [0, \pi/2]$,*

$$\begin{aligned} \boldsymbol{C}_\theta &:= \left\{(x_1, x_2) \in [0,\infty]^2 : x_1 \wedge x_2 \le 1,\, x_2 \le x_1 \tan\theta\right\} \quad \textit{and} \\ \boldsymbol{D}_\theta &:= \left\{(x_1, x_2) \in [0,\infty]^2 : x_1 \wedge x_2 \tan\theta \le 1\right\}. \end{aligned}$$

Both sets $\boldsymbol{C}_\theta$ and $\boldsymbol{D}_\theta$ have clear geometric interpretations. For $\theta_1 < \theta_2$ in $[0, \pi/2]$, $\boldsymbol{C}_{\theta_2} \backslash \boldsymbol{C}_{\theta_1}$ contains those points in $[0,\infty]^2$ for which at least one of the components is no bigger than 1 and are trapped in the cone between angles θ_1 and θ_2; $\boldsymbol{D}_\theta$ defines the union of two sets

$$\{(x_1, x_2) : 0 \le x_1 \le 1, 0 \le x_2 \le \infty\} \cup \{(x_1, x_2) : 0 \le x_1 \le \infty, 0 \le x_2 \le \cot\theta\}$$

where the factor $\cot\theta$ allows us to control the boundary of the second set. Define

$$\Lambda(x_1, x_2) = \Lambda\left(([x_1, \infty] \times [x_2, \infty])^C\right), \quad x_1, x_2 \in [0,\infty]^2. \tag{3.1}$$

Immediately by (2.3),

$$\Lambda(x_1, x_2) = x_1 \Lambda(1, x_2/x_1). \tag{3.2}$$

The following holds.

Proposition 3.2 *For each* $\theta \in [0, \pi/2]$,

(1) $\Lambda(\boldsymbol{C}_\theta) = \Phi[0, \theta]$, *where* Φ *is the spectral measure in (2.3), and*

(2) $\Lambda(\boldsymbol{D}_\theta) = \Lambda(1, \cot\theta)$. □

In view of Proposition 3.2(2) together with (3.2), $\{\boldsymbol{D}_\theta : 0 \le \theta \le \pi/2\}$ is measure-determining for Λ. The corresponding result of (1), which is proved in the Appendix, shows that $\{\boldsymbol{C}_\theta : 0 \le \theta \le \pi/2\}$ is also measure-determining for Λ. We note that Proposition 3.2(1) was obtained in Einmahl et al. (2001) from an entirely different perspective.

Definition 3.3 *For all* $\theta \in [0, \pi/2]$ *we define*

$$\Phi(\theta) = \Lambda(\boldsymbol{C}_\theta) = \Phi[0, \theta], \quad \textit{and} \quad \psi(\theta) = \Lambda(\boldsymbol{D}_\theta) = \Lambda(1, \cot\theta).$$

□

By Proposition 3.2 and (2.4),

$$\Phi(\theta) = \lim_{t\to\infty} \Lambda_t(\boldsymbol{C}_\theta) \quad \text{and} \quad \psi(\theta) = \lim_{t\to\infty} \Lambda_t(\boldsymbol{D}_\theta),$$

provided that $\Lambda(\partial \boldsymbol{C}_\theta) = \Lambda(\partial \boldsymbol{D}_\theta) = 0$, and therefore $\Phi(\theta)$ and $\psi(\theta)$ can be estimated statistically by the nonparametric procedures $\widehat{\Lambda}_{\varepsilon,n}(\boldsymbol{C}_\theta), \widehat{\Lambda}_{\varepsilon,n}(\boldsymbol{D}_\theta)$, respectively, if an iid sample is available. From this perspective, we discuss below the relevance of $\Phi(\theta)$ and $\psi(\theta)$.

Estimating $\Phi(\theta)$ is a central theme in Einmahl et al. (2001). Let $\overline{G}_j(x) = P(X_j > x) = 1/x\,, x > 1$. Observe that for $0 < \theta_1 < \theta_2 < \pi/2$,

$$\begin{aligned} & P(X_1 \vee X_2 > n, \tan\theta_1 < X_1/X_2 \le \tan\theta_2) \\ = \;& P(n\overline{G}_1(X_1) \wedge n\overline{G}_2(X_2) < 1, \tan\theta_1 < \overline{G}_2(X_2)/\overline{G}_1(X_1) \le \tan\theta_2) \\ = \;& n^{-1}(\Lambda_n(\boldsymbol{C}_{\theta_2}) - \Lambda_n(\boldsymbol{C}_{\theta_1})) \\ \sim \;& n^{-1}(\Phi(\theta_2) - \Phi(\theta_1)), \end{aligned}$$

provided $\Lambda(\partial \boldsymbol{C}_{\theta_i}) = 0, i = 1, 2$. However, if the $\overline{G}_i$ are highly non-linear, the quantity $\Phi(\theta_2) - \Phi(\theta_1)$ may be difficult to interpret. It is also somewhat cumbersome to use an estimated $\Phi(\theta)$ to estimate the distribution of the coordinate-wise maxima

$$P\left(\bigvee_{i=1}^{n} G_j(X_{i,j}) \le \frac{1}{n}u_j + 1,\ j = 1, 2\right);$$

one could conceivably proceed with this using the integral representation of the copula, but in doing $\Phi(\theta)$ has to be estimated for every θ followed by a numerical integration. The function $\psi(\theta)$ compliments $\Phi(\theta)$ in that respect, as explained below.

Suppose that $x_i = x_{i,n},\ i = 1, 2$, are such that

$$0 < \liminf_{n\to\infty} n\overline{G}_i(x_i) \le \limsup_{n\to\infty} n\overline{G}_i(x_i) < \infty,\ i = 1, 2.$$

Then it follows from (2.2) that for $n \to \infty$,

$$\begin{aligned} P(X_1 > x_1 \text{ or } X_2 > x_2) &\sim \frac{1}{n}\Lambda(n\overline{G}_1(x_1), n\overline{G}_2(x_2)) \\ = \overline{G}_1(x_1)\Lambda\left(1, \frac{\overline{G}_2(x_2)}{\overline{G}_1(x_1)}\right) &= \overline{G}_1(x_1)\psi\left(\arctan\left(\frac{\overline{G}_1(x_1)}{\overline{G}_2(x_2)}\right)\right). \end{aligned} \tag{3.3}$$

As a result,

$$\begin{aligned} P^n(X_1 \le x_1, X_2 \le x_2) &\approx \exp\left(-n\overline{G}_1(x_1)\psi\left(\arctan\left(\frac{\overline{G}_1(x_1)}{\overline{G}_2(x_2)}\right)\right)\right) \\ &\approx P^{n\xi}(X_1 \le x_1), \end{aligned} \tag{3.4}$$

where

$$\xi = \xi(x_1, x_2) = \psi\left(\arctan\left(\frac{\overline{G}_1(x_1)}{\overline{G}_2(x_2)}\right)\right).$$

If $G_1 = G_2$ then

$$\xi(x, x) = \psi(\pi/4), \tag{3.5}$$

which is what Schlather & Tawn (2000) refers to as *extremal coefficient*, a notion related to the extremal index (cf. Leadbetter, Lindgren and Rootzén (1983) or Embrechts et al. 1997) in univariate extreme value theory for time series.

4 Bivariate tail dependence function

In this section we continue to explore the properties of $\psi(\theta)$ defined in Definition 3.3 and how it can be useful for describing multivariate extremes. First, we have:

Proposition 4.1 **(1)** *ψ is convex.*

(2) $\psi_1(\theta) \le \psi(\theta) \le \psi_0(\theta), \quad \theta \in [0, \pi/2]$, *where $\psi_0(\theta) := 1 + \cot\theta$ corresponds to independence and $\psi_1(\theta) := 1 \vee \cot\theta$ to complete dependence.* □

The proof of this proposition can be found in the Appendix.

The function ψ becomes a much more effective tool for visualizing dependence if it is normalized, as follows.

Definition 4.2 *We define the bivariate* tail dependence function *as*

$$\rho(\theta) = \frac{\psi_0(\theta) - \psi(\theta)}{\psi_0(\theta) - \psi_1(\theta)} = \frac{1 + \cot\theta - \psi(\theta)}{1 \wedge \cot\theta}\,, \quad \theta \in (0, \pi/2)\,. \tag{4.1}$$

□

By Proposition 4.1(2) the function $\rho(\theta)$ takes values in $[0,1]$, with $\rho(\theta)$ being close to 0/1 corresponds to weak/strong dependence.

The quantity $\rho(\pi/4) = 2 - \psi(\pi/4)$ (cf. (3.5)) is referred to as the *(upper) tail dependence coefficient* in Joe (1997), which, as the name suggests, is meant to describe the degree of dependence in the upper tails of the marginals. Thus, the function ρ extends this notion from a single direction, $\pi/4$, to all directions in $(0, \pi/2)$. This is illustrated by the following example, which is similar to an example in Ledford & Tawn (1996).

Example 4.3 Let $X_1 \sim G_1$, $X_2 \sim G_2$ where G_1 and G_2 are continuous distributions. Note that $\left(1/\overline{G}_1(X_1), 1/\overline{G}_2(X_2)\right)$ has Pareto(1) margins and the same copula as (X_1, X_2). It follows from (2.4) and Definition 3.3 that for all $\theta \in (0, \pi/2)$, we have $\Lambda_t(1, \cot\theta) \to \Lambda(1, \cot\theta) = \psi(\theta)$ as $t \to \infty$, and hence

$$\begin{aligned}
&\lim_{t\to\infty} P\left(\frac{1}{\overline{G}_2(X_2)} > t\tan\theta \,\Big|\, \frac{1}{\overline{G}_1(X_1)} > t\right) \\
&= \lim_{t\to\infty} t\left(1 - P\left(\frac{1}{\overline{G}_1(X_1)} \le t\right) - P\left(\frac{1}{\overline{G}_2(X_2)} \le t\tan\theta\right)\right. \\
&\quad \left. + P\left(\frac{1}{\overline{G}_1(X_1)} \le t\,,\, \frac{1}{\overline{G}_2(X_2)} \le t\tan\theta\right)\right) \\
&= 1 + \cot\theta - \lim_{t\to\infty} t\left(1 - P\left(\frac{1}{\overline{G}_1(X_1)} \le t\,,\, \frac{1}{\overline{G}_2(X_2)} \le t\tan\theta\right)\right) \\
&= 1 + \cot\theta - \lim_{t\to\infty} tP\left(t\left(\overline{G}_1(X_1), \overline{G}_2(X_2)\right) \in \left([1,\infty] \times [\cot\theta, \infty]\right)^C\right) \\
&= 1 + \cot\theta - \psi(\theta) = (1 \wedge \cot\theta)\,\rho(\theta)\,.
\end{aligned}$$

Hence for all $\theta \in (0, \pi/2)$,

$$\lim_{t\to\infty} P\left(X_2 > G_2^{-1}\left(1 - \frac{1}{t\tan\theta}\right) \Big| X_1 > G_1^{-1}\left(1 - \frac{1}{t}\right)\right) \;=\; (1 \wedge \cot\theta)\,\rho(\theta)\,.$$

□

Our examples below show that ρ provides an effective tool to visualize dependence in the extreme tails of the bivariate distribution. In practice, when G is unknown, $\rho(\theta)$ can be estimated from a set of iid data $(X_i, Y_i), 1 \le i \le n$, by the nonparametric estimate

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

where

$$\begin{aligned}\widehat{\psi}_{\varepsilon,n}(\theta) &:= \widehat{\Lambda}_{\varepsilon,n}(\boldsymbol{D}_\theta) \\ &= \varepsilon \sum_{i=1}^{n} I\left(\varepsilon\left(R_{i,1}, R_{i,2}\right) \in \boldsymbol{D}_\theta\right) \\ &= \varepsilon \sum_{i=1}^{n} I(R_{i,1} \le \varepsilon^{-1} \text{ or } R_{i,2} \le \varepsilon^{-1} \cot\theta).\end{aligned}$$

As mentioned in Section 2, theoretically ε and $1/(n\varepsilon)$ should be both small in order for the estimator to perform well. In practice, we will plot $\widehat{\rho}_{\varepsilon,n}(\theta)$ for ε in some sensible range for which ε and $1/(n\varepsilon)$ are "small" and pick an ε_0 for which the estimates $\widehat{\psi}_{\varepsilon,n}(\pi/4)$ behave stably in the neighborhood of ε_0. While it is convenient to use the same ε for all θ, allowing ε to vary with θ in simple ways may improve the quality of the estimation. Indeed, when θ approaches $\pi/2$, increasingly fewer points of $\varepsilon\left(R_{i,1}, R_{i,2}\right)$ are captured by $\boldsymbol{D}_\theta$, which has the effect of inflating the variance of the estimate in that region. A practical way to overcome this is to choose a baseline $\varepsilon = \varepsilon_0$ at $\theta = \pi/4$ and allow ε to decrease slightly as θ approaches $\pi/2$. Another practical consideration is a simple smoothing. At least visually if not theoretically, the quality of the estimate of $\widehat{\rho}_{\varepsilon,n}(\theta)$ improves if some smoothing is incorporated. In that regard, one can perform a simple averaging over a box window or use something more sophisticated such as spline smoothing.

We also recommend plotting $(1/R_{i,1}, 1/R_{i,2})$, $1 \le i \le n$, alongside that of $\widehat{\rho}_{\varepsilon,n}(\theta)$ to fully appreciate the information in the latter. Recall that

$$\lim_{n\to\infty} P^n\left(1/(n\overline{G}_1(X_1)) \le 1, 1/(n\overline{G}_2(X_2)) \le \tan\theta\right) = \rho(\theta).$$

As such, $\widehat{\rho}_{\varepsilon,n}(\theta)$ describes the degree of dependence reflected by the pattern of points of $(1/R_{i,1}, 1/R_{i,2})$, $1 \le i \le n$, in the box $[0,1] \times [0, \tan\theta]$. The following simple example demonstrates these points.

Example 4.4 Let $p_1, p_2 \in (0,1)$ and consider the model

$$X_1 = p_1 Z_1 \vee (1-p_1) Z_2 \quad \text{and} \quad X_2 = p_2 Z_1 \vee (1-p_2) Z_3,$$

with Z_1, Z_2, Z_3 distributed as iid Pareto(1). Clearly, the dependence between X_1 and X_2 arises from the common component Z_1. Hence the dependence is stronger for larger values of p_1, p_2. It is easy to see that both X_1 and X_2 are asymptotically distributed as Pareto(1) in the tails. It is also easy to see that

$$P(X_1 > x \text{ or } X_2 > x\tan\theta) \sim \frac{1}{x}\left(1 + \cot\theta - p_1 \wedge p_2 \cot\theta\right).$$

Applying (3.3), we have

$$\psi(x) = 1 + \cot\theta - p_1 \wedge p_2 \cot\theta,$$

and

$$\rho(\theta) \;=\; \frac{p_1 \wedge p_2 \cot\theta}{1 \wedge \cot\theta}. \tag{4.2}$$

In Figure 1 we simulated this model for $n = 10\,000$ iid observations of (X_1, X_2). The three sets of plots on the three rows correspond to the cases: $p_1 = 0.7$, $p_2 = 0.3$, $p_1 = 0.5$, $p_2 = 0.5$ and $p_1 = 0.2$, $p_2 = 0.8$. On each row the left-most plot is the true functions $\rho(\theta)$ in (4.2) (dashed line) overlaid with the smoothed version of $\widehat{\rho}_{\varepsilon,n}(\theta)$ (solid line) based on one simulated sample of size 10 000, where ε is $1/200$ for $\theta \in [0, \pi/4]$ and thereafter, ε decreases linearly to $1/210$ when θ reaches $\pi/2$. We computed $\widehat{\rho}_{\varepsilon,n}(\theta)$ for $\theta \in \{\theta_i = i\pi/200, 1 \le i \le 100\}$ and produced the smoothed version $\widehat{\rho}^{(s)}_{\varepsilon,n}(\theta_i)$ by averaging $\widehat{\rho}_{\varepsilon,n}(\theta_j), |j - i| \le s = 5$, i.e.

$$\widehat{\rho}^{(s)}_{\varepsilon,n}(\theta_i) = \frac{1}{2s+1} \sum_{j=-s}^{s} \widehat{\rho}_{\varepsilon,n}(\theta_{i-j})\,.$$

The plots in the middle column illustrate the root of the mean squared error

$$\mathrm{MSE}(\theta_i) = \sum_{k=1}^{100} \left(\widehat{\rho}^{(s),k}_{\varepsilon,n}(\theta_i) - \rho(\theta_i)\right)^2$$

for the three cases based on 100 simulations with $n = 10\,000$ iid observations each and $\widehat{\rho}^{(s),k}_{\varepsilon,n}(\theta_i)$ represents the smoothed estimator of simulation k, $1 \le k \le 100$. The right-most plots contain the simulated points $(1/R_{i,1}, 1/R_{i,2})$, $1 \le i \le n$, of one single sample of size 10 000 but with points close to (1,1) truncated for easy viewing.

In the first row of plots, ρ is larger for small θ than for large θ; this is reflected by the right-most 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 right-most plot. The third row of plots is the converse of the first row of plots which is reflected by the pattern of extreme points above the diagonal. This is an example of a situation where Joe's 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. □

Example 4.5 Let $\boldsymbol{X} = (X_1, X_2)$ be a bivariate random vector with dependence structure given by a *Gumbel*-copula

$$C_{\boldsymbol{X}}(u, v) = \exp\left\{-\left[(-\ln u)^\delta + (-\ln v)^\delta\right]^{1/\delta}\right\}, \quad \delta \in [1, \infty)\,. \tag{4.3}$$

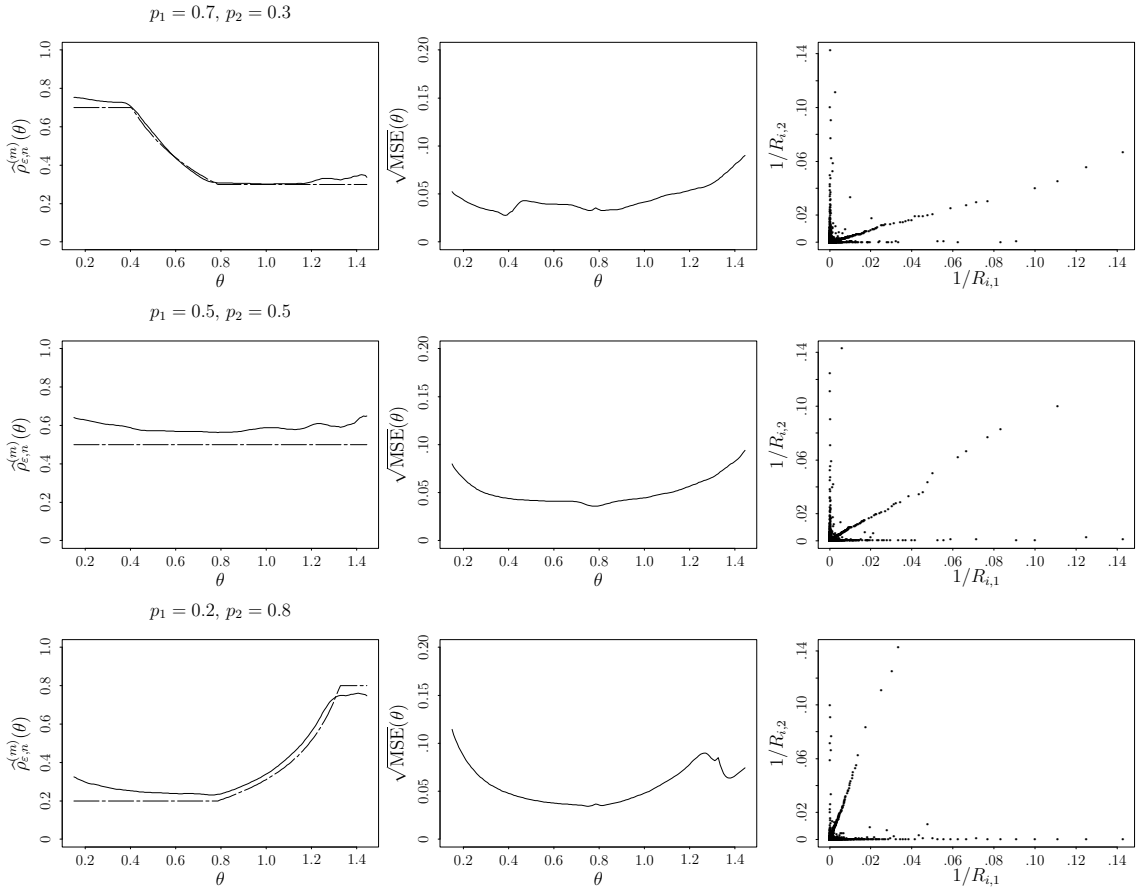

Figure 1: Left column: smoothed version of $\widehat{\rho}^{(5)}_{\varepsilon,n}(\theta)$ (solid line) overlaid with true function $\rho(\theta)$.
Middle column: $\sqrt{\mathrm{MSE}}(\theta)$.
Right column: plots of $(1/R_{i,1}, 1/R_{i,2})$, with points close to (1,1) truncated, $p_1 = 0.7$, $p_2 = 0.3$ (upper row), $p_1 = 0.5$, $p_2 = 0.5$ (middle row) and $p_1 = 0.2$, $p_2 = 0.8$ (lower row).

The dependence arises from δ. It is a symmetric model and by Example 4.3 it has (upper) tail dependence coefficient $\rho(\pi/4) = \lambda_U = \ln 2/\ln(2-\delta)$. Since $C_{\boldsymbol{X}}$ is an extreme copula, $\psi(\theta) = \left(1 + (\cot\theta)^\delta\right)^{1/\delta}$ and hence

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

We simulated this model for $n = 10\,000$, and in Figure 2 the plots are given in the same order as in Figure 1 based on Example 4.4. We have chosen $\rho(\pi/4) = 0.3$ (upper row), $\rho(\pi/4) = 0.7$ (middle row) and $\rho(\pi/4) = 0.9$ (lower row). The level of dependence is manifested by the data scattered around the diagonal. □

5 Multivariate extensions

One advantage of the functions of ψ and ρ in Definitions 3.3 and 4.2 is that they can be readily extended to higher dimensions by incorporating additional angles θ_j. Let

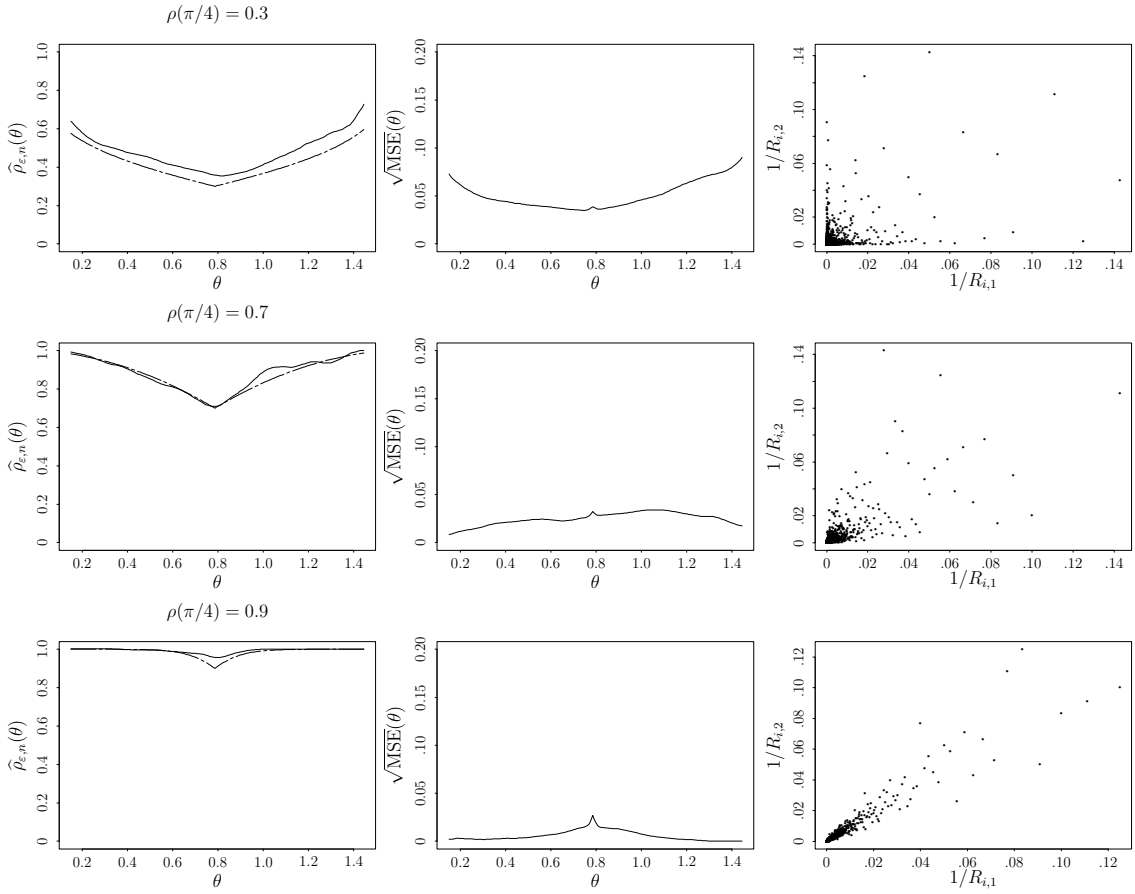

Figure 2: Left column: smoothed versions $\widehat{\rho}^{(5)}_{\varepsilon,n}(\theta)$ (solid line) overlaid with true function $\rho(\theta)$.
Middle column: $\sqrt{\mathrm{MSE}}(\theta)$.
Right column: plots of $1/R_{i,j}$, $j = 1, 2$, for $C_{\boldsymbol{X}}$ given in (4.3) and $\rho(\pi/4) = 0.3$ (upper row), $\rho(\pi/4) = 0.7$ (middle row), $\rho(\pi/4) = 0.9$ (lower row).

$m \geq 2$ and $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,m})$ be iid with a distribution G, where the margins G_j are assumed to be continuous. Assume that (1.2) holds and the copula of F has the representation (1.4). Define the measures Λ_t and Λ on the Borel σ-algebra of $[0, \infty]^m \backslash \{\infty, \ldots, \infty\}$ by

$$\Lambda_t(\boldsymbol{A}) := tP\left(t\left(\overline{G}_1(X_{1,1}), \ldots, \overline{G}_m(X_{1,m}\right) \in \boldsymbol{A}\right),$$

and

$$\Lambda\left(([x_1, \infty] \times \cdots \times [x_m, \infty])^C\right) := \int_{S_m} \left(\bigvee_{j=1}^{m} w_j x_j\right) \mu(d\boldsymbol{w}).$$

As in the two-dimensional case, we have

$$\lim_{t \to \infty} \Lambda_t(\boldsymbol{A}) = \Lambda(\boldsymbol{A})$$

for any Borel set $\boldsymbol{A} \subset [0, \infty]^m \backslash \{\infty, \ldots, \infty\}$ with $\Lambda(\partial \boldsymbol{A}) = 0$. Now set

$$\Lambda(x_1, \ldots, x_m) := \Lambda\left(([x_1, \infty] \times \cdots \times [x_m, \infty])^C\right),$$

and, for $\theta_2, \ldots, \theta_m \in [0, \pi/2]$,

$$\boldsymbol{D}_{\theta_2,\ldots,\theta_m} := \{(x_1, \ldots, x_m) \in [0,\infty]^m : x_1 \wedge x_2 \tan\theta_2 \wedge \cdots \wedge x_m \tan\theta_m \leq 1\} .$$

The sets $\boldsymbol{D}_{\theta_2,\ldots,\theta_m}$ for $\theta_2, \ldots, \theta_m \in [0, \pi/2]$ are measure-determining for Λ. Define

$$\psi(\theta_2, \ldots, \theta_m) := \Lambda\left(\boldsymbol{D}_{\theta_2,\ldots,\theta_m}\right) = \Lambda(1, \cot\theta_2, \ldots, \cot\theta_m).$$

Hence by the same arguments as in Proposition 4.1, ψ is convex and

$$\psi_1(\theta_2, \ldots, \theta_m) \leq \psi(\theta_2, \ldots, \theta_m) \leq \psi_0(\theta_2, \ldots, \theta_m), \quad \theta_2, \ldots, \theta_m \in [0, \pi/2], \tag{5.1}$$

where

$$\begin{aligned} \psi_0(\theta_2, \ldots, \theta_m) &= 1 + \cot\theta_2 + \cdots + \cot\theta_m, \\ \psi_1(\theta_2, \ldots, \theta_m) &= 1 \vee \cot\theta_2 \vee \cdots \vee \cot\theta_m; \end{aligned}$$

ψ_0 and ψ_1 correspond to the independent and completely dependent cases, respectively.

Definition 5.1 *The* tail dependence function*, for $m \geq 2$, is defined as*

$$\rho(\theta_2, \ldots, \theta_m) = \frac{(1 + \cot\theta_2 + \cdots + \cot\theta_m) - \psi(\theta_2, \ldots, \theta_m)}{(1 + \cot\theta_2 + \cdots + \cot\theta_m) - (1 \vee \cot\theta_2 \vee \cdots \vee \cot\theta_m)}.$$

□

By (5.1), ρ is in $[0, 1]$ and ρ being close to 0 and 1 correspond to weak and strong dependence, respectively.

In practice, when G is unknown, $\Lambda(\boldsymbol{A})$ can be estimated for any Borel set $\boldsymbol{A}$ from a set of data $(X_{i,1}, \ldots, X_{i,m})$, $1 \leq i \leq n$, using the nonparametric estimator

$$\widehat{\Lambda}_{\varepsilon,n}(\boldsymbol{A}) = \varepsilon \sum_{i=1}^{n} I\left(\varepsilon\left(R_{i,1}, \ldots, R_{i,m}\right) \in \boldsymbol{A}\right). \tag{5.2}$$

The theoretical properties of the bivariate estimator as explained after (2.5) can also be verified in higher dimensions. Accordingly, the estimate $\rho(\theta_2, \ldots, \theta_m)$ is defined as

$$\widehat{\rho}_{\varepsilon,n}(\theta_2, \ldots, \theta_m) := \frac{\psi_0(\theta_2, \ldots, \theta_m) - \widehat{\Lambda}_{\varepsilon,n}\left(\boldsymbol{D}_{\theta_2,\ldots,\theta_m}\right)}{\psi_0(\theta_2, \ldots, \theta_m) - \psi_1(\theta_2, \ldots, \theta_m)},$$

where

$$\widehat{\Lambda}_{\varepsilon,n}\left(\boldsymbol{D}_{\theta_2,\ldots,\theta_m}\right) = \varepsilon \sum_{i=1}^{n} I\left(\varepsilon\left(R_{i,1}, \ldots, R_{i,m}\right) \in \boldsymbol{D}_{\theta_2,\ldots,\theta_m}\right).$$

All practical considerations made in the previous section continue to be applicable here. To visualize extreme dependence in the data, plot $\widehat{\rho}_{\varepsilon,n}(\theta_2, \ldots, \theta_m)$ for a discrete set of $(\theta_2, \ldots, \theta_m)$. When $m \geq 3$, plotting the estimated ρ requires considerable creativity. In the following example the tail dependence function can be calculated explicitly.

Example 5.2 Let $c_{ji} \in [0,1]$ for $1 \le j \le m, 1 \le i \le k$, such that $\sum_{i=1}^{k} c_{ji} = 1$ for all j. Consider

$$X_j = \bigvee_{i=1}^{k} c_{ji} Z_i \,, \quad j = 1, \ldots, m,$$

where $Z_1, \ldots, Z_k$ are iid Pareto(1). Generalizing (3.3), we obtain

$$P\left(X_1 > x \text{ or } X_2 > x \tan\theta_2 \text{ or } \cdots \text{ or } X_m > x \tan\theta_m\right) \sim \frac{1}{x}\psi(\theta_2, \ldots, \theta_m)\,, \quad x \to \infty\,.$$

On the other hand,

$$\begin{aligned} & P\left(X_1 > x \text{ or } X_2 > x \tan\theta_2 \text{ or } \cdots \text{ or } X_m > x \tan\theta_m\right) \\ &= 1 - P\left(X_1 \le x, X_2 \le x \tan\theta_2, \ldots, X_m \le x \tan\theta_m\right) \\ &= 1 - \prod_{j=1}^{k} P\left(Z_j \le x \left(c_{1j}^{-1} \wedge c_{2j}^{-1} \tan\theta_2 \wedge \cdots \wedge c_{mj}^{-1} \tan\theta_m\right)\right) \\ &\sim \frac{1}{x} \sum_{j=1}^{k} c_{1j} \vee c_{2j} \cot\theta_2 \vee \cdots \vee c_{mj} \cot\theta_m\,. \end{aligned}$$

Hence,

$$\psi(\theta_2, \ldots, \theta_m) = \sum_{i=1}^{k} \left(c_{1i} \vee c_{2i} \cot\theta_2 \vee \cdots \vee c_{mi} \cot\theta_m\right)\,,$$

and

$$\begin{aligned} & \rho(\theta_2, \ldots, \theta_m) \\ &= \frac{(1 + \cot\theta_2 + \cdots + \cot\theta_m) - \sum_{i=1}^{k} \left(c_{1i} \vee c_{2i} \cot\theta_2 \vee \cdots \vee c_{mi} \cot\theta_m\right)}{(1 + \cot\theta_2 + \cdots + \cot\theta_m) - (1 \vee \cot\theta_2 \vee \cdots \vee \cot\theta_m)}\,. \end{aligned}$$

Note that this example generalizes Example 4.4 which is the special case of $m = 2$, $k = 3$, $c_{11} = p_1$, $c_{12} = 1 - p_1$, $c_{13} = 0$, $c_{21} = p_2$, $c_{22} = 0$, $c_{23} = 1 - p_2$. □

Example 5.3 We estimate the dependence structure of the model given in Example 5.2 with $m = 3$ and $k = 5$. We choose the constants c_{ji}, $1 \le j \le 3$, $1 \le i \le 5$, as

$c_{11} = 0.2$	$c_{12} = 0.2$	$c_{13} = 0$	$c_{14} = 0.6$	$c_{15} = 0$
$c_{21} = 0.6$	$c_{22} = 0$	$c_{23} = 0.2$	$c_{24} = 0$	$c_{25} = 0.2$
$c_{31} = 0.2$	$c_{32} = 0.6$	$c_{33} = 0.2$	$c_{34} = 0$	$c_{35} = 0$

Figures 3 and 4 contain the simulation results of this model for $n = 10\,000$ iid observations of (X_1, X_2, X_3). We chose $\varepsilon = 1/200$ and computed the estimate $\widehat{\rho}_{\varepsilon,n}(\theta_2, \theta_3)$

for $\theta_2, \theta_3 \in \{\theta_i = i\pi/200,\, 1 \le i \le 100\}$ and smoothed $\widehat{\rho}_{\varepsilon,n}(\theta_i, \theta_j)$ by averaging $\widehat{\rho}_{\varepsilon,n}(\theta_k, \theta_l)$, $|k-i|, |l-j| \le s = 3$, i.e.

$$\widehat{\rho}^{(s)}_{\varepsilon,n}(\theta_i, \theta_j) = \frac{1}{(2s+1)^2} \sum_{k,l=-s}^{s} \widehat{\rho}_{\varepsilon,n}(\theta_{i-k}, \theta_{j-l})\,.$$

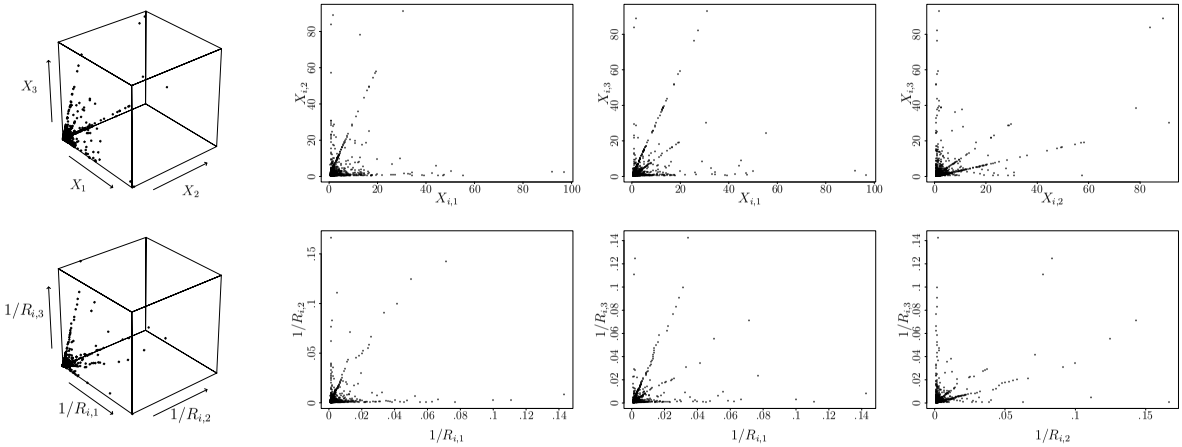

Figure 3: First row: data (3-d and 2-d projections) of the model given in Example 5.2.
Second row: ranks $1/R_{i,j}$, $1 \le j \le 3$, in the same order as in row one.

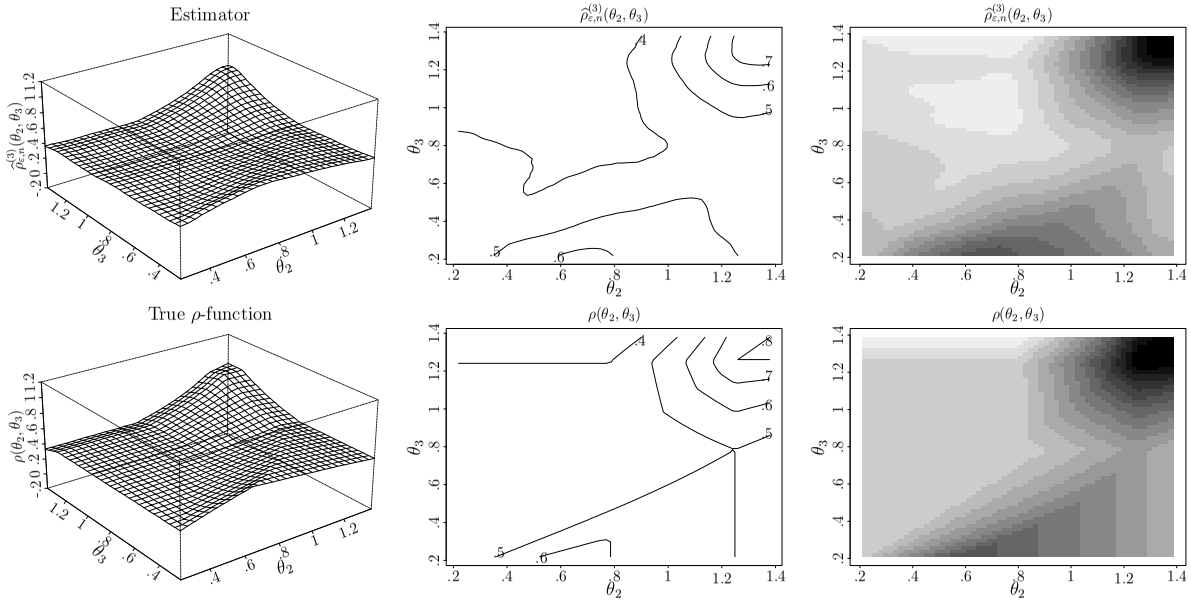

Figure 4: First row: smoothed estimate $\widehat{\rho}^{(3)}_{\varepsilon,n}(\theta_2, \theta_3)$ of the simulated data (see Figure 3), with perspective plot (left-most), contour plot (middle) and image plot (right-most).
Second row: true tail dependence function $\rho(\theta_2, \theta_3)$ for this model.

In the first row of Figure 3 the data are plotted, where in the left-most plot we show the 3-dimensional data, the three plots on the right hand side show the

projections of the data, $(X_{i,1}, X_{i,2})$, $(X_{i,1}, X_{i,3})$ and $(X_{i,2}, X_{i,3})$. The second row is given in the same order as the first row, showing the reciprocal ranks $1/R_{i,j}$, $1 \leq j \leq 3$. The first row of Figure 4 shows the estimate $\widehat{\rho}^{(3)}_{\varepsilon,n}(\theta_2, \theta_3)$, where the left plot is a perspective plot, the middle one is a contour plot and the right one is a grey-scale image plot. To see how the estimator performs the second row presents the true tail-dependence function $\rho(\theta_2, \theta_3)$ for this model.

Remark 5.4 Let $\rho_{1,2,3}$ be the tail dependence function of three rvs X_1, X_2, X_3 and $\rho_{1,j}$ be the tail dependence function of X_1, X_j, $j = 2, 3$, hence by definition $\rho_{1,2}(\theta_2) = \rho_{1,2,3}(\theta_2, \pi/2)$ and $\rho_{1,3}(\theta_3) = \rho_{1,2,3}(\pi/2, \theta_3)$ holds $\forall \theta_2, \theta_3 \in (0, \pi/2)$. Therefore $\rho_{1,2}$ can be estimated by the cross section of the estimated trivariate tail dependence function at a large and fixed angle θ_2, and similarly for $\rho_{1,3}$. To identify $\rho_{2,3}$ recall that $\Lambda_{1,2,3}(0, a, b) = \Lambda_{2,3}(a, b)$, hence

$$
\begin{aligned}
&\lim_{\varepsilon \to 0} \rho_{1,2,3}\left(\arctan \varepsilon, \arctan\left(\varepsilon \tan \theta\right)\right) \\
&= \lim_{\varepsilon \to 0} \frac{1 + 1/\varepsilon + \cot\theta/\varepsilon - \psi_{1,2,3}\left(\arctan \varepsilon, \arctan\left(\varepsilon \tan \theta\right)\right)}{1 + 1/\varepsilon + \cot\theta/\varepsilon - 1 \vee 1/\varepsilon \vee \cot\theta/\varepsilon} \\
&= \lim_{\varepsilon \to 0} \frac{\varepsilon + 1 + \cot\theta - \varepsilon \Lambda_{1,2,3}\left(1, 1/\varepsilon, \cot\theta/\varepsilon\right)}{\varepsilon + 1 + \cot\theta - \varepsilon \vee 1 \vee \cot\theta} \\
&= \lim_{\varepsilon \to 0} \frac{\varepsilon + 1 + \cot\theta - \Lambda_{1,2,3}\left(\varepsilon, 1, \cot\theta\right)}{\varepsilon + 1 + \cot\theta - \varepsilon \vee 1 \vee \cot\theta} \\
&= \frac{1 + \cot\theta - \Lambda_{2,3}\left(1, \cot\theta\right)}{1 + \cot\theta - 1 \vee \cot\theta} = \rho_{2,3}(\theta)\,.
\end{aligned}
$$

□

6 The swap rate data

The data consist of returns (daily differences) of *Annually Compounded Zero Coupon Swap Rates* with different maturities (between 7 days and 30 years) and different currencies (EUR, USD and GBP). Each of the time series consists of 257 daily returns during the year 2001. In an exploratory data analysis we investigated first each single time series. Plots of the autocorrelation functions of the single time series, their moduli and squares exhibited no significant temporal dependence structure; hence we assume the data being iid. Moreover, the histograms and a tail analysis showed that the marginals are well modelled (at least in the tails) by a two-sided exponential distribution. Concerning multivariate (spatial) dependence, for swap rates in the same currency we observed a high dependence for similar maturities, and a low dependence between very different maturities. Between different currencies we observed only very little dependence except for similarly long maturities, where

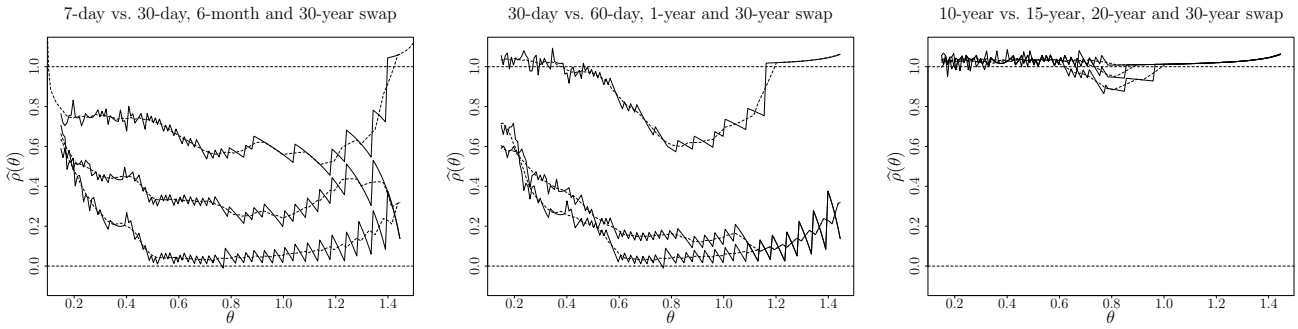

Figure 5: Estimates $\widehat{\rho}$ for some swap rates with smoothed versions (dashed lines).
Left plot: $\widehat{\rho}$ for 7-day vs 30-day, 7-day vs 6-month, and 7-days vs 30-year.
Middle plot: $\widehat{\rho}$ for 30-day vs 60-day, 30-day vs 1-year and 30-day vs 30-year.
Right plot: $\widehat{\rho}$ for 10-year vs 15-year, 10-year vs 20-year and 10-year vs 30-year.

we detected some moderate dependence. For plots and details on these effects we refer to Kuhn (2002).

To see the estimator $\widehat{\rho}$ at work we show plots of $\widehat{\rho}_{\varepsilon,n}(\theta)$, $\theta \in (0, \pi/2)$, as defined in (4.1) for the *swap rate data* described above for EUR. We use the nonparametric estimator given in (2.5). We stay away from the boundaries $\theta = 0$ and $\theta = \pi/2$ since $\widehat{\psi}_{\varepsilon,n}(\theta)$ tends to ∞ as $\theta \to 0$, and for θ near $\pi/2$ there is a lack of data.

In Figure 5 the tail dependence function is estimated for various combinations of swap rates of different maturities with $\widehat{\rho}_{\varepsilon,n}(\theta_i)$ (zigzag-line) and the smoothed version $\widehat{\rho}^{(m)}_{\varepsilon,n}(\theta_i)$ (dashed line) for $\varepsilon = 0.06$, $m = 5$ and $\theta_i = \frac{i}{200}\frac{\pi}{2}$, $1 \le i \le 200$. The left plot shows strong dependence between the 7-day and 30-day rates, moderate dependence between the 7-day and 6-month rates, but very weak dependence between the 7-day and 30-year rates. The middle plot shows moderate dependence between the 30-day and 60-day rates for θ close to $\pi/4$ and exceptionally high dependence for θ small or large, but weak dependence between the 30-day and 1-year and 30-day and 30-year rates. The right plot shows strong dependence between the 10-year,15-year, 20-year and 30-year rates. □

Example 6.1 Figure 6 shows a comparison of the tail dependence function with the spectral measure Φ as defined in (3.2). We recall that in case of independence

$$\int_0^{\frac{\pi}{2}} \left(\frac{x}{1 \vee \cot\gamma} \vee \frac{y}{1 \vee \tan\gamma} \right) \Phi(d\gamma) = x + y,$$

and hence

$$\Phi(\theta) = \Phi([0,\theta]) = \begin{cases} 1, & \theta < \pi/2\,, \\ 2, & \theta = \pi/2\,, \end{cases}$$

and in case of complete dependence

$$\int_0^{\frac{\pi}{2}} \left(\frac{x}{1 \vee \cot\gamma} \vee \frac{y}{1 \vee \tan\gamma} \right) \Phi(d\gamma) = x \vee y,$$

and hence

$$\Phi(\theta) = \Phi([0,\theta]) = \begin{cases} 0, & \theta < \pi/4\,, \\ 1, & \pi/4 \le \theta \le \pi/2\,. \end{cases}$$

These results allow us to interpret the plots. We consider the 20-year vs. 30-year, 7-day vs. 30-day, and 7-day vs. 30-year swap rates. In the first row (high dependence) the estimated spectral measure Φ equals 0 for $\theta < 0.4$ and then quickly jumps to 1. In the third row (low dependence) the estimated Φ jumps quickly to 1 and remains there until close to $\pi/2$ where it jumps to 2. The middle row (moderate dependence) is a mixture of high and low dependence case. □

Example 6.2 Figures 7 and 8 show two trivariate examples. The first example is generated by the low dependent swap rates with 7 day maturity and currencies USD, EUR and GBP; $X_{i,1}$ corresponds to USD, $X_{i,2}$ to EUR and $X_{i,3}$ to GBP. In the first row we plotted the ranks $1/R_{i,j}$, $1 \le j \le 3$, where $R_{i,j} = \mathrm{rank}(-X_{i,j})$. In the left-most plot we show the 3-dimensional data, the three plots on the right hand side show the two-dimensional projections $(1/R_{i,1}, 1/R_{i,2})$, $(1/R_{i,1}, 1/R_{i,3})$ and $(1/R_{i,2}, 1/R_{i,3})$. The second row shows the smoothed estimator $\widehat{\rho}^{(s)}_{\varepsilon,n}(\theta_2,\theta_3)$ for $n = 257$, $\varepsilon = 0.06$ and $s = 3$; the left plot is a perspective plot, the middle one is a contour plot and the right one is a grey-scale image plot.

These 7-day swap rates show low and symmetric tail dependence which is reflected by many points lying near to the axes and the rest is scattered roughly uniformly with respect to the angles θ_2, θ_3 (first row of figure 7). The estimator $\widehat{\rho}^{(s)}_{\varepsilon,n}(\theta_2,\theta_3)$ (second row) is therefore between 0.15 and 0.35 showing no significant difference between small and large angles.

Figure 8 shows the same as figure 7 for the high dependent EUR swap rates with maturities 5, 6 and 7 years. These swap rates with long and similar maturities show

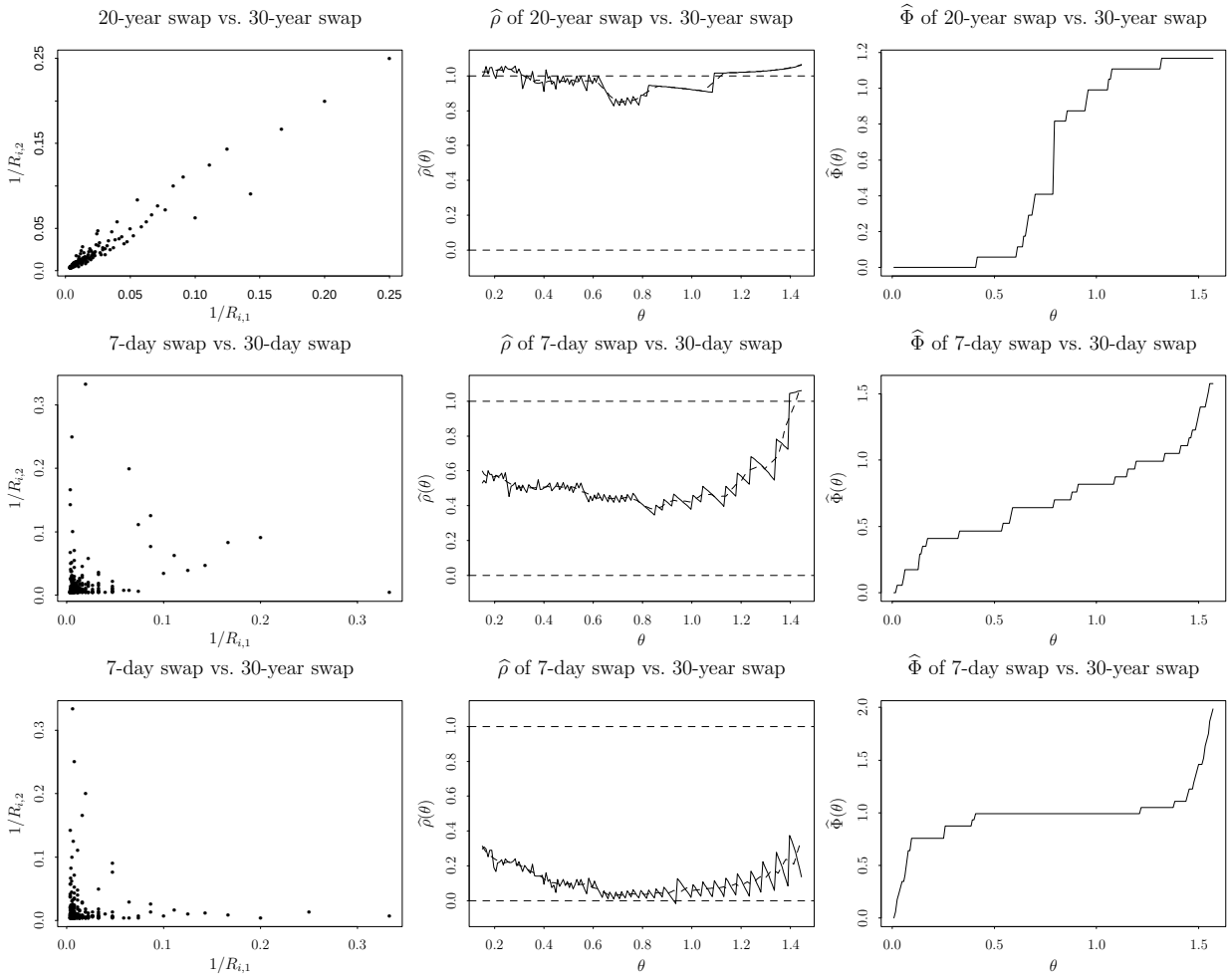

Figure 6: Estimators of ρ and Φ for some swap rates: 20-year vs. 30-year (first row), 7-day vs. 30-day (second row), and 7-day vs. 30-year swap rates (third row).
Left plots: transformed ranks $1/R_{i,j}, j = 1, 2$.
Middle plots: estimated tail dependence function $\widehat{\rho}$.
Right plots: estimated spectral measure $\widehat{\Phi}$ of the data.
In the first row we see high, in the second middle and in the third row low dependence.

high and symmetric tail dependence which is reflected by all points lying near the diagonal (first row of figure 7). The estimator $\widehat{\rho}^{(s)}_{\varepsilon,n}(\theta_2, \theta_3)$ (second row) is therefore almost everywhere close to 1, only for angles θ_2, θ_3 near $\pi/4$ the estimator becomes smaller which is illustrated by the points that are away from the diagonal. □

Appendix

Proof of Proposition 3.2 (1): The case $\theta = \pi/2$ is obvious. Let $y_i = i/n$ if

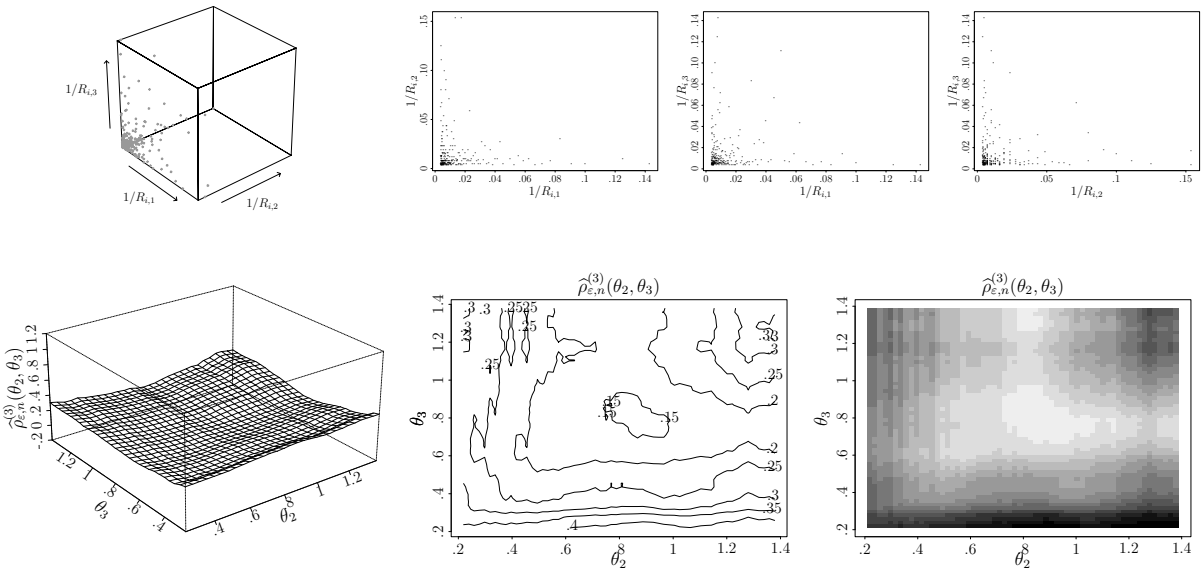

Figure 7: First row: data (3-d and 2-d projections) of the ranks $1/R_{i,j}$, $1 \le j \le 3$, of the low dependent 7-day swap rates rates in USD, EUR and GBP.
Second row: smoothed estimator $\widehat{\rho}^{(3)}_{\varepsilon,n}(\theta_2,\theta_3)$, perspective plot (left-most), contour plot (middle) and grey scale image plot (right-most)

$\theta \in (0,\pi/4]$ and $x_i = i/n$ if $\theta \in (\pi/4,\pi/2)$, $1 \le i \le n$, then

$$\Lambda(C_\theta) = \begin{cases} \lim\limits_{n\to\infty} \sum\limits_{i=1}^{n} \left[\Lambda(y_i \cot\theta, y_i) - \Lambda(y_i \cot\theta, y_{i-1})\right], & \theta \in (0, \frac{\pi}{4}] \\ \lim\limits_{n\to\infty} \sum\limits_{i=1}^{n} \left[\Lambda(x_i, x_i \tan\theta) - \Lambda(x_i, x_{i-1} \tan\theta)\right] + \Lambda(1,1) - \Lambda(1, \tan\theta), & \theta \in (\frac{\pi}{4}, \frac{\pi}{2}). \end{cases}$$

Consider first $\theta \in (0,\pi/4]$. Note that for $x_1, x_2 \in [0,\infty]$,

$$\Lambda(x_1,x_2) = x_1 \int\limits_{\left(\arctan\frac{x_2}{x_1},\frac{\pi}{2}\right]} \frac{1}{1 \vee \cot\gamma}\Phi(d\gamma) + x_2 \int\limits_{\left[0,\arctan\frac{x_2}{x_1}\right]} \frac{1}{1 \vee \tan\gamma}\Phi(d\gamma). \quad \text{(A.1)}$$

Thus, letting $\theta_i := \arctan\left(\frac{i-1}{i}\tan\theta\right)$,

$$\begin{aligned} &\Lambda\left(y_i\cot\theta, y_i\right) - \Lambda\left(y_i\cot\theta, y_{i-1}\right) \\ &= \frac{1}{n}\left(i \int\limits_{(\theta_i,\theta]} \left[(1 \wedge \cot\gamma) - (\cot\theta)(1\wedge\tan\gamma)\right]\Phi(d\gamma) + \int\limits_{[0,\theta_i]} \left(1 \wedge \cot\theta\right)\Phi(d\gamma)\right) \\ &= \frac{1}{n}\left(i \int\limits_{(\theta_i,\theta]} \left[1 - (\cot\theta)(\tan\gamma)\right]\Phi(d\gamma) + \Phi[0,\theta_i]\right). \end{aligned}$$

Observe that $\sup_{\gamma\in(\theta_i,\theta]} i[1-(\cot\theta)(\tan\gamma)] \le 1$. Since $\theta_i \to \theta$, we have

$$\limsup_{i\to\infty} i \int\limits_{(\theta_i,\theta]} \left[1 - (\cot\theta)(\tan\gamma)\right]\Phi(d\gamma) \le \Phi(\{\theta\}),$$

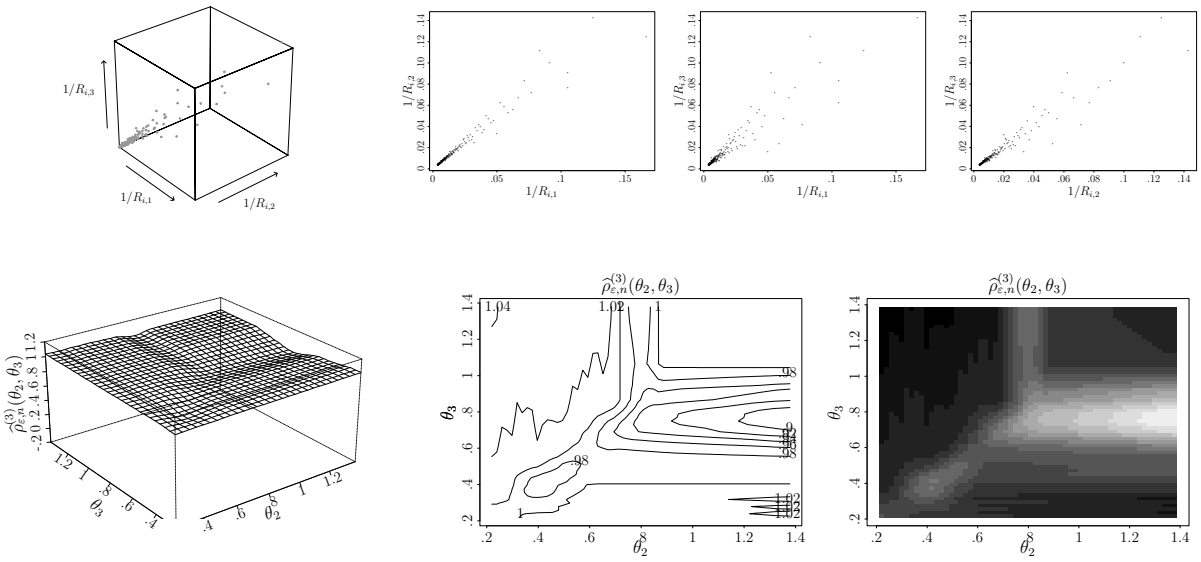

Figure 8: First row: data (3-d and 2-d projections) of the ranks $1/R_{i,j}$, $1 \le j \le 3$, of the high dependent 5-year, 6-year and 7-year EUR swap rates.
Second row: smoothed estimator $\widehat{\rho}^{(3)}_{\varepsilon,n}(\theta_2,\theta_3)$, perspective plot (left-most), contour plot (middle) and grey scale image plot (right-most)

whereas $\Phi[0,\theta_i] \to \Phi[0,\theta)$ as $i \to \infty$. Applying Cesaro's mean value theorem we conclude that $\Lambda(C_\theta) = \Phi[0,\theta]$ for all $\theta \in (0,\pi/4]$ with $\Phi(\{\theta\}) = 0$. The case $\theta \in (\pi/4,\pi/2)$ can be dealt with similarly and the two cases combine to give $\Lambda(C_\theta) = \Phi[0,\theta]$ for all $\theta \in (0,\pi/2)$ with $\Phi(\{\theta\}) = 0$. Note that both $\Lambda(C_\theta)$ and $\Phi[0,\theta]$ are nondecreasing and right-continuous functions in θ. Since they agree on a dense subset of points in $[0,\pi/2]$ they must agree on the entire interval of $[0,\pi/2]$. This concludes the proof. □

Proof of Proposition 4.1 (1): $\psi(\theta) = \int_0^{\pi/2} (1/(1 \vee \cot\gamma)) \vee (\cot\theta/(1 \vee \tan\gamma))\, \Phi(d\gamma)$ and since the integrand is convex with respect to θ, ψ is.
(2): Recall that $\int_0^{\pi/2} (1/(1 \vee \cot\gamma))\, \Phi(d\gamma) = \int_0^{\pi/2} (1/(1 \vee \tan\gamma))\, \Phi(d\gamma) = 1$. With (A.1) this gives $\Lambda(x_1,x_2) \le x_1 + x_2$ hence $\psi(\theta) = \Lambda(1,\cot\theta) \le 1 + \cot\theta$. The independence extreme copula being $C_{\mathrm{I}}(x,y) = xy$, it follows from (1.5) that $\psi(\theta) = \psi_0(\theta) = 1 + \cot\theta$ for the independence case.

The extreme copula that corresponds to complete dependence is $C_{\mathrm{U}}(x,y) = x \wedge y$, it follows from (1.5) that $\psi(\theta) = \psi_1(\theta) = 1 \vee \cot\theta$ for the complete dependence case. By *Fréchet-Hoeffding* bounds (see Nelsen 1999, p. 9) $C \le C_{\mathrm{U}}$ holds for all copulae C, hence $\psi_1 \le \psi$ for all ψ. □

References

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

[2] Coles, S. G. and Tawn, J. A. (1991) Modelling extreme multivariate events. *J. Royal Statist. Soc.* B **53**, 377-392.

[3] Einmahl, J., de Haan, L. and Piterbarg, V.I. (2001) Nonparametric estimation of the spectral measure of an extreme value distribution. *Ann. Statist.* **29**, 1401-1423.

[4] Einmahl, J., de Haan, L. and Sinha, A.K. (1997) Estimating the spectral measure of an extreme value distribution. *Stoch. Proc. Appl.* **70**, 143-171.

[5] Embrechts, P., Haan, L. de, Huang, X. (2000) Modelling multivariate extremes. In: Embrechts, P. (Ed.) *Extremes and Integrated Risk Management* RISK Books, pp. 59-67

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

[7] De Haan, L. and Resnick, S. (1977) Limit theory for multidimensional sample extremes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **40**, 317-337.

[8] De Haan, L. and De Ronde, J. (1998). Sea and wind: multivariate extremes at work. *Extremes* **1**, 7-45.

[9] Hauksson, H. A., Dacorogna, M., Domenig, T., Müller, U. and Samorodnitsky, G. (2001) Multivariate extremes, aggregation and risk estimation. *Quantitative Finance* **1**, 79-95.
Available at *http://papers.ssrn.com/sol3/papers.cfm?abstract_id=254392*

[10] Huang, Xin (1992) *Statistics of Bivariate Extremes.* Ph.D. Thesis, Erasmus University Rotterdam, Tinbergen Research Institute, no. **22**.

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

[12] Kuhn, G. (2002) *Multivariate Value at Risk Schätzung von Zins Swap Sätzen.* Diploma Thesis. Munich University of Technology.
Available at *www.ma.tum.de/stat/*.

[13] Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983) *Extremes and Related Properties of Random Sequences and Processes.* Springer, New York.

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

[15] Nelsen, R. (1999) *An Introduction to Copulas.* Lecture Note in Statistics, no. 139. Springer, New York.

[16] Qi, Y. (1997) Almost sure convergence of the stable tail empirical dependence function in multivariate extremes statistics. *Acta Math. Appl. Sinica (English ser.)* **13**, 167-175

[17] Resnick, S.I. (1987) *Extreme Values, Regular Variations and Point Processes.* Springer, New York.

[18] Schlather, M. and Tawn, J. (2000) Properties of extremal coefficients. Technical report ST 00-01, Department of Mathematics and Statistics, Lancaster University.

[19] Schlather, M. and Tawn, J. (2003) A dependence measure for multivariate and spatial extreme values: Properties and Inference. *Biometrika* **90**, 139-156.

[20] Smith, R. L. (2003) Statistics of extremes, with applications in environment, insurance and finance. Chapter 1 of, *Extreme Values in Finance, Telecommunications and the Environment*, pp. 1-78. Editors B. Finkenstadt and H. Rootzén, Chapman and Hall/CRC Press.

[21] Tawn, J. A. (1988) Bivariate extreme value theory: models and estimation. *Biometrika* **75**, 397-415.