A NOTE ON A NON-PARAMETRIC TAIL DEPENDENCE ESTIMATOR

Matthias Fischer & Marco Dörflinger
Department of Statistics and Econometrics
University of Erlangen-Nürnberg, Germany
Email: Matthias.Fischer@wiso.uni-erlangen.de

SUMMARY

We present a non-parametric tail dependence estimator which arises naturally from a specific regression model. Above that, this tail dependence estimator also results from a specific copula mixture.

Keywords and phrases: Upper tail dependence; nonparametric estimation; copula

1 Coefficients of Tail Dependence (TDC)

Let X and Y denote two random variables with joint distribution $F_{X,Y}(x,y)$ and continuous marginal distribution functions $F_X(x)$ and $F_Y(y)$. According to Sklar's (1960) fundamental theorem, there exists a unique decomposition

$$F_{X,Y}(x,y) = C(F_X(x), F_Y(y))$$

of the joint distribution into its marginal distribution functions and the so-called copula (function)

$$C(u,v) = P(U \leq u, V \leq v), \quad U \equiv F_X(X), \quad V \equiv F_Y(Y)$$

on $[0,1]^2$ which comprises the information about the underlying dependence structure (For details on copulas we refer to Joe, 1997). The concept of tail dependence provides, roughly speaking, a measure for extreme co-movements in the lower and upper tail of $F_{X,Y}(x,y)$, respectively. The upper tail dependence coefficient (TDC) is usually defined by

$$\lambda_U \equiv \lim_{u \to 1^-} P(Y > F_Y^{-1}(u) | X > F_X^{-1}(u)) = \lim_{u \to 1^-} \frac{1 - 2u + C(u,u)}{1-u} \in [0,1]. \tag{1.1}$$

noting that λ_U is solely depending on $C(u,v)$ and not on the marginal distributions. Analogously, the lower TDC is defined as

$$\lambda_L \equiv \lim_{u \to 0^+} P(Y \leq F_Y^{-1}(u) | X \leq F_X^{-1}(u)) = \lim_{u \to 0^+} \frac{C(u,u)}{u}. \tag{1.2}$$

Coles et al. (1999) provide asymptotically equivalent versions of (1.1) and (1.2),

$$\lambda_L = 2 - \lim_{u \to 0^+} \frac{\log(1 - 2u + C(u,u))}{\log(1-u)} \quad \text{and} \quad \lambda_U = 2 - \lim_{u \to 1^-} \frac{\log C(u,u)}{\log(u)}. \tag{1.3}$$

For reason of brevity, we focus on the upper TDC λ_U. Results on the lower TDC can be obtained in a similar manner.

2 Reviewing non-parametric TDC-estimators

For a given (bivariate) random sample of length n $(X_1, Y_1), \ldots, (X_n, Y_n)$ from (X, Y) let

$$X_{(1)} \equiv \min\{X_1, \ldots, X_n\} \leq \ldots \leq X_{(n)} \equiv \max\{X_1, \ldots, X_n\}$$

denote the corresponding order statistics. All of the relevant non-parametric TDC-estimator $\widehat{\lambda}_U$ of λ_U (See, e.g., Schmidt & Stadtmüller, 2006, Frahm, Junker & Schmidt, 2005 and Dobric & Schmid, 2005) rest upon the non-parametric copula estimator

$$C_n(i/n, j/n) = \frac{1}{n} \sum_{l=1}^{n} \mathbf{1}(X_l \leq X_{(i)}, Y_l \leq Y_{(j)}). \tag{2.1}$$

Plugging (2.1) into equation (1.1) and its asymptotically equivalent version (1.3), respectively, a first pair of estimators is obtained (via "simple replacement")

$$\widehat{\lambda}_U^{[1]} \equiv \frac{C_n\left((1-k/n, 1] \times (1-k/n, 1]\right)}{1-(1-k/n)} \quad \text{and} \quad \widehat{\lambda}_U^{[2]} = 2 - \frac{\log C_n(1-k/n, 1-k/n)}{\log(1-k/n)},$$

where $k \approx \sqrt{n}$ seems to be appropriate (cp. Dobric & Schmid, 2005, section 4). Secondly, Dobric & Schmid (2005) interpret equation (1.1) after suitable re-formulations as regression equation

$$C_n\left((1-i/n, 1] \times (1-i/n, 1]\right) = \lambda_U \cdot \frac{i}{n} + \varepsilon_i, \quad i = 1, \ldots, k \tag{2.2}$$

which motives $\widehat{\lambda}_U^{[3]}$ as OLS-estimator of equation (2.2). Thirdly, Dobric & Schmid (2005) propose to approximate the unknown copula $C(u, v)$ by the convex-combination $\widetilde{C}(u, v) \equiv \alpha \min\{u, v\} + (1-\alpha)uv$ of the maximum (co-monotonicity) copula $\min\{u, v\}$ and the independence copula $C_\perp(u, v) = uv$ (i.e. copula **B11** in Joe, 1997). Noting that λ_U of $\widetilde{C}$ is given by α, Dobric & Schmid (2005) introduce $\widehat{\lambda}_U^{[4]}$ which corresponds to that α which minimizes

$$F(\alpha) \equiv \sum_{i=1}^{k} \left(C_n\left(1-\frac{i}{n}, 1-\frac{i}{n}\right) - \widetilde{C}\left(1-\frac{i}{n}, 1-\frac{i}{n}\right)\right)^2.$$

The following tables classifies the above-mentioned estimators $\widehat{\lambda}_U^{[i]}$, $i = 1, \ldots, 4$. By the end

	Underlying method		
Underlying formula ↓	Simple replacement	Regression approach	Approximation
(1.1)	$\widehat{\lambda}_U^{(1)}$ √	$\widehat{\lambda}_U^{(3)}$ √	$\widehat{\lambda}_U^{(4)}$ √
(1.3)	$\widehat{\lambda}_U^{(2)}$ √	$\widehat{\lambda}_U^{(5)}$	$\widehat{\lambda}_U^{(6)}$

Table 1: A classification scheme of non-parametric TDC estimators.

of this work we present an TDC-estimator $\lambda_U^{(6)}$ which coincides with the TDC-estimator $\lambda_U^{(5)}$ which arises from a regression equation derived from TDC-formula (1.3).

3 Derivation of a new non-parametric TDC-estimator

Instead of considering the arithmetic mean of the independence copula $C_\perp$ and the comonotonicity copula C_U, we now focus on the geometric mean of $C_\perp$ and C_U (i.e. copula family **B12** in Joe, 1997), that is

$$C^*(u,v) = (\min\{u,v\})^\delta \cdot (uv)^{1-\delta}, \quad \delta \in [0,1]. \tag{3.1}$$

We first proof that δ corresponds to the lower TDC of $\overline{C}$: Using equation (1.3),

$$\lambda_U = 2 - \lim_{u\to 1^-} \frac{\log C^*(u,u)}{\log(u)} = 2 - \lim_{u\to 1^-} \frac{\log(u^{2-\delta})}{\log(u)} = \delta.$$

In accordance to Dobric & Schmid (2005), approximating the (unknown) copula $C(u,v)$ by $C^*(u,v)$, an estimator for λ_U is given by

$$\widehat{\lambda}_U^{(6)} = \text{argmin}_{\lambda\in[0,1]} \sum_{i=1}^{k} \left(C_n\left(1-\frac{i}{n}, 1-\frac{i}{n}\right) - \left(1-\frac{i}{n}\right)^{2-\lambda} \right)^2. \tag{3.2}$$

We next show that $\widehat{\lambda}_U^{(6)}$ equals $\widehat{\lambda}_U^{(5)}$, a new estimator which results from an LS-estimaton of the equation

$$\log C_n\left(1-\frac{i}{n}, 1-\frac{i}{n}\right) = (2-\lambda_U)\cdot \log\left(\frac{i}{n}\right) + \varepsilon_i, \quad i=1,\dots,k \tag{3.3}$$

which itself follows from the Coles et al. (1999) formula (1.3) in combination with (2.1).

Lemma 3.1. *The tail dependence estimator* $\widehat{\lambda}_U^{(6)}$ *and* $\widehat{\lambda}_U^{(5)}$ *are asymptotically equivalent.*

Proof: We first observe that the LS estimator can be represented as

$$\widehat{\lambda}_U^{(5)} = \text{argmin}_{\lambda\in[0,1]} \sum_{i=1}^{k} \left(\log C_n\left(1-\frac{i}{n}, 1-\frac{i}{n}\right) - (2-\lambda)\cdot\log\left(1-\frac{i}{n}\right) \right)^2.$$

Now, using the relationship $\log(y^d) \approx 1 - y^d$ for $y \approx 1$ and $d \in [0,1]$,

$$\begin{aligned}
\widehat{\lambda}_U^{(5)} &= \text{argmin}_{\lambda\in[0,1]} \sum_{i=1}^{k} \left(\log C_n\left(1-\frac{i}{n}, 1-\frac{i}{n}\right) - \log\left(1-\frac{i}{n}\right)^{2-\lambda} \right)^2. \\
&\approx \text{argmin}_{\lambda\in[0,1]} \sum_{i=1}^{k} \left(1 - C_n\left(1-\frac{i}{n}, 1-\frac{i}{n}\right) - 1 + \left(1-\frac{i}{n}\right)^{2-\lambda} \right)^2 \\
&= \text{argmin}_{\lambda\in[0,1]} \sum_{i=1}^{k} \left(C_n\left(1-\frac{i}{n}, 1-\frac{i}{n}\right) - \left(1-\frac{i}{n}\right)^{2-\lambda} \right)^2 = \widehat{\lambda}^{(6)}. \quad \square
\end{aligned}$$

References

[1] Coles, S.; Heffernan, J.; Tawn, J. Dependence measures for extreme value analyses. *Extremes*, **1999**, 2, 339-365.

[2] Dobric, J.; Schmid, F. Nonparametric Estimation of the Lower Tail Dependence l in Bivariate Copulas. *Journal of Applied Statistics*, **2005**, 32(4), 387-407 .

[3] Frahm, G.; Junker, M.; Schmidt, R. Estimating the tail dependence coefficient. *Insurance: Mathematics and Economics*, **2005**, 37, 80-100.

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

[5] Schmidt, R.; Stadtmüller, U. Non-parametric estimation of tail dependence. *Scandinavian Journal of Statistics*, **2006**, 33, 307-335.

[6] Sklar, A. Fonctions de répartitions á n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, **1959**, 8(1), 229-231.