Improving estimation for asymptotically independent bivariate extremes via global estimators for the angular dependence function

C. J. R. Murphy-Barltrop[1,2], J. L. Wadsworth[3] and E. F. Eastoe[3]

[1] Technische Universität Dresden, Institut Für Mathematische Stochastik, Helmholtzstraße 10, 01069 Dresden, Germany

[2] Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Dresden/Leipzig, Germany

[3]Department of Mathematics and Statistics, Lancaster University LA1 4YF, United Kingdom

STEW September 2023

- Finished my PhD at Lancaster University in June 2023, supervised by Jennifer Wadsworth and Emma Eastoe.
- I am now a Research Associate at TU Dresden.
- **NB:** The work I am discussing today was completed during my PhD.

Overview

Background

- Multivariate extreme value model $=$ framework for evaluating extremal dependence of a random vector (X, Y).

Background

- Multivariate extreme value model = framework for evaluating extremal dependence of a random vector (X, Y).
- Classical approaches based on the framework of multivariate regular variation (MRV).

Background

- Suppose (X, Y) has standard Pareto margins and define $R := X + Y$, $V := X/(X + Y)$.

Background

- Suppose (X, Y) has standard Pareto margins and define $R := X + Y$, $V := X/(X + Y)$.
- (X, Y) is MRV if, for any measurable $B \subset [0, 1]$,

$$\lim_{r \to \infty} \Pr(V \in B, R > sr \mid R > r) = H(B)s^{-1}, s \geq 1.$$

Background

- Suppose (X, Y) has standard Pareto margins and define $R := X + Y$, $V := X/(X + Y)$.
- (X, Y) is MRV if, for any measurable $B \subset [0, 1]$,

$$\lim_{r\to\infty} \Pr(V \in B, R > sr \mid R > r) = H(B)s^{-1}, s \geq 1.$$

- H is the spectral measure and summarises the extremal dependence (Resnick, 1987).
- Interpretation: as R gets big, V and R become independent.

Background

- Fundamental classification of extremal dependence given by the coefficient.

$$\chi = \lim_{u \to 1} \Pr(F_Y(Y) > u \mid F_X(X) > u) \in [0, 1].$$

Background

- Fundamental classification of extremal dependence given by the coefficient.

$$\chi = \lim_{u \to 1} \Pr(F_Y(Y) > u \mid F_X(X) > u) \in [0, 1].$$

- $\chi = 0 \Rightarrow$ asymptotic independence.
- $\chi > 0 \Rightarrow$ asymptotic dependence.

Background

- Classical approaches based on MRV (most approaches) cannot distinguish between asymptotic independence and complete independence.

$$H(\{0\}) = H(\{1\}) = 0.5.$$

Background

- Classical approaches based on MRV (most approaches) cannot distinguish between asymptotic independence and complete independence.

$$H(\{0\}) = H(\{1\}) = 0.5.$$

- Consequently, MRV cannot accurately extrapolate for asymptotically independent vectors (Ledford and Tawn, 1996, 1997; Heffernan and Tawn, 2004).

Background

- More recent techniques have been developed that can capture both dependence regimes.

Background

- More recent techniques have been developed that can capture both dependence regimes.
- Throughout the remainder of this talk, let (X, Y) be a random vector on standard exponential margins.

Background

- First such approach given by Ledford and Tawn (1996, 1997).

Background

- First such approach given by Ledford and Tawn (1996, 1997).

  $$\Pr(X > u, Y > u) = \Pr(\min(X, Y) > u) \to L(e^u) \exp(-u/\eta),$$

  as $u \to \infty$, with L slowly varying and $\eta \in (0, 1]$.
- Asymptotic dependence: $\eta = 1$.

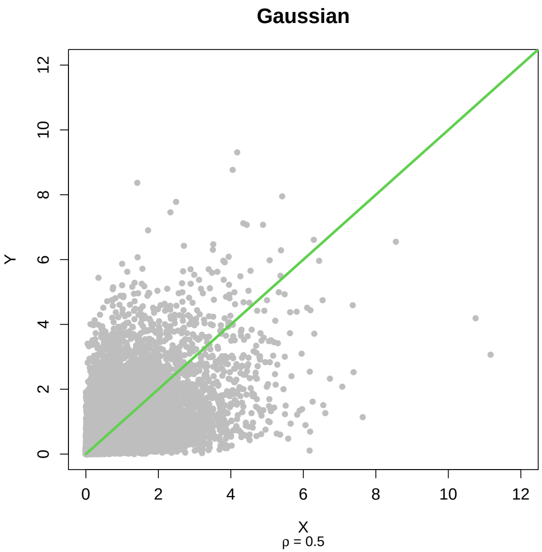
Background

- First such approach given by Ledford and Tawn (1996, 1997).

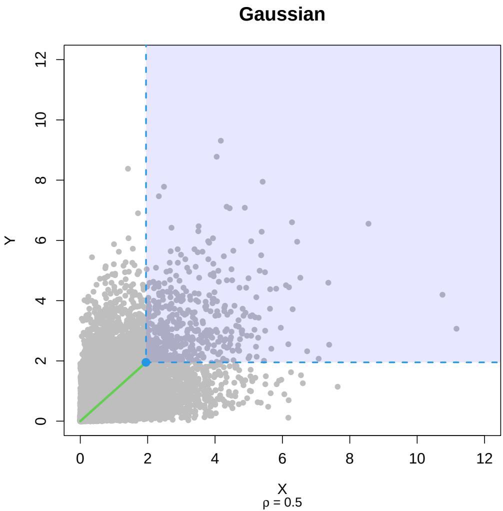
  $$\Pr(X > u, Y > u) = \Pr(\min(X, Y) > u) \to L(e^u)\exp(-u/\eta),$$

  as $u \to \infty$, with L slowly varying and $\eta \in (0, 1]$.
- Asymptotic dependence: $\eta = 1$.
- Interpretation: joint survivor function decays exponentially along the $y = x$ line.

Background



Background



As we drag this box out, the probability of being in the blue region decays exponentially.

Background

- Ledford and Tawn (1996, 1997) model was extended in Wadsworth and Tawn (2013).

Background

- Ledford and Tawn (1996, 1997) model was extended in Wadsworth and Tawn (2013).
- Given any ray (angle) $w \in [0, 1]$,

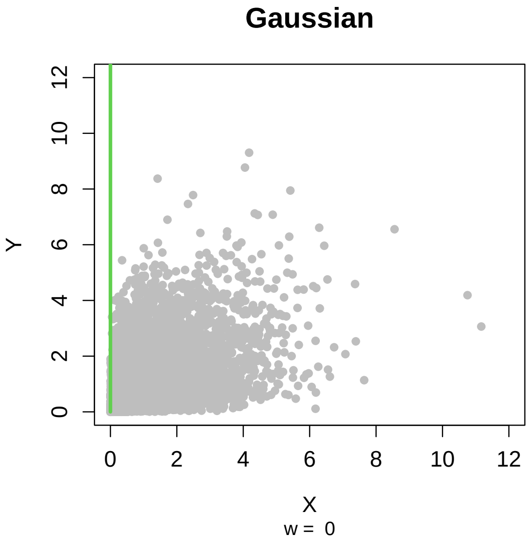
$$\Pr(X > wu, Y > (1-w)u) =$$
$$\Pr\left(\min\left\{\frac{X}{w}, \frac{Y}{1-w}\right\} > u\right) \to L(e^u \mid w)\exp(-\lambda(w)u),$$

  as $u \to \infty$, with L slowly varying.
- $\lambda(w), w \in [0, 1]$ is known as the **angular dependence function** (ADF).

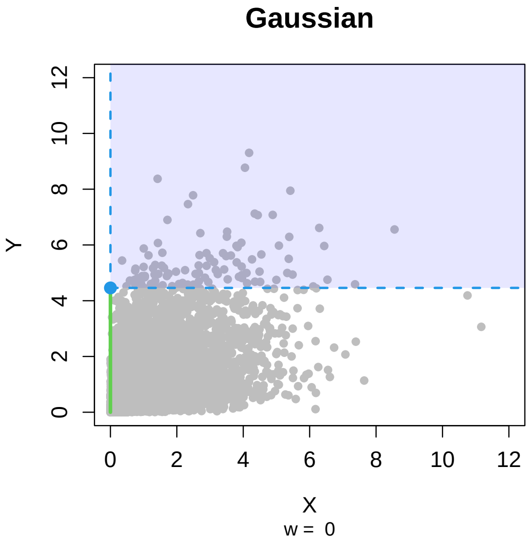
Background

- Interpretation: joint survivor function decays exponentially in all directions.
- $\lambda(w) \geq \max(w, 1-w)$.
- Asymptotic dependence: $\lambda(w) = \max(w, 1-w)$.
- $\eta = 1/(2\lambda(0.5))$.

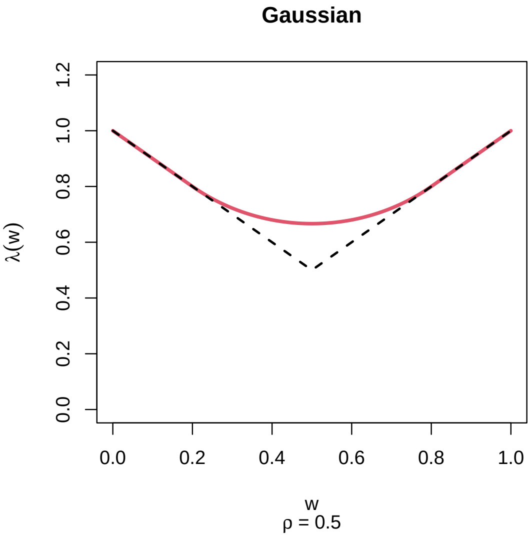
Background



Background



Background



Background

- Alternative representation for multivariate extremes given by Heffernan and Tawn (2004).

Background

- Alternative representation for multivariate extremes given by Heffernan and Tawn (2004).

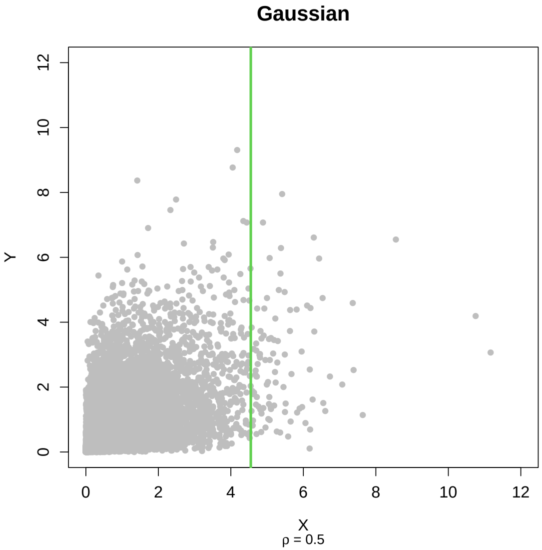
$$(Y \mid X = x) = \alpha_{y|x} x + x^{\beta_{y|x}} Z$$

for extreme x, where Z is a residual process.

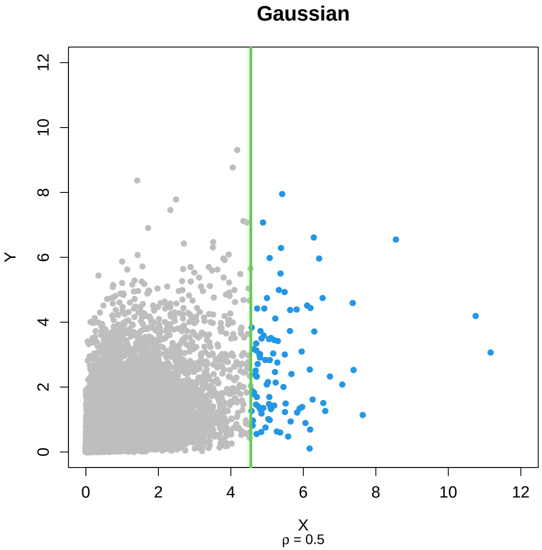
Background

- $\alpha_{y|x} \in [0, 1]$, $\beta_{y|x} \in [0, 1)$.
- Could alternatively condition on Y being extreme.
- Asymptotic dependence: $\alpha_{y|x} = 1$, $\beta_{y|x} = 0$.
- Interpretation: fancy regression.

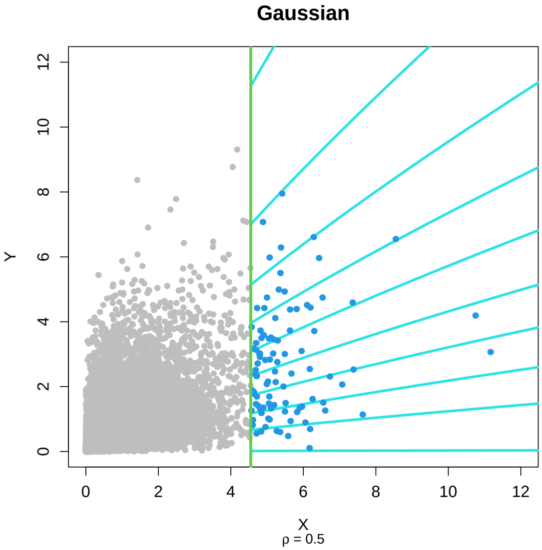
Background



Background



Background



$\alpha_{y|x}$ determines slope.

Background

- Heffernan and Tawn (2004) model widely used in practice.

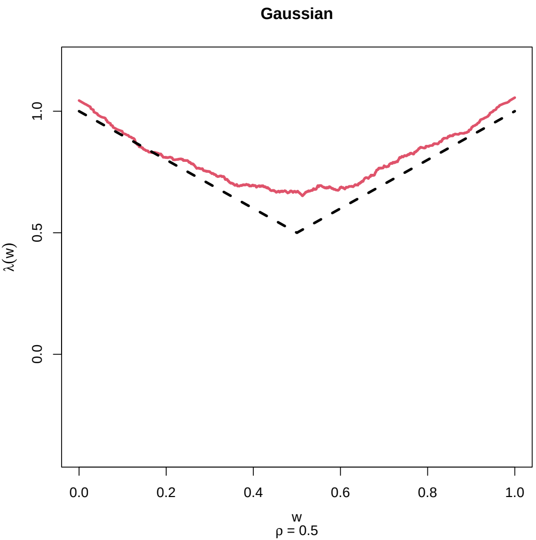
Background

- Heffernan and Tawn (2004) model widely used in practice.
- Few applications of Wadsworth and Tawn (2013) model exist, even though this model outperforms Heffernan and Tawn (2004) in certain scenarios (Murphy-Barltrop et al., 2023b).

Background

- **The problem:** until recently, the ADF has been estimated in a pointwise manner using the Hill estimator.
- Consider each $w \in [0, 1]$ in turn, obtain an estimate $\hat{\lambda}(w)$, stitch estimates together.

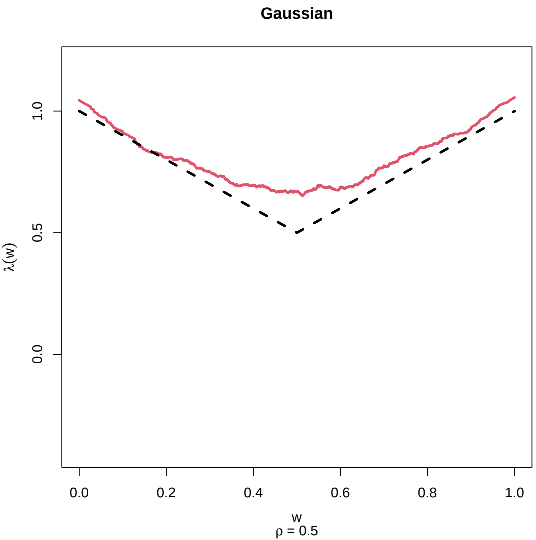
Background



Background

- Does not necessarily satisfy endpoint conditions $\lambda(0) = \lambda(1) = 1$ or lower bound $\lambda(w) \geq \max(w, 1 - w), w \in [0, 1]$.

Background



Background

Our goal:

Provide smooth ADF estimators that satisfy theoretical constraints.

Existing ADF estimators

- Denote non-smooth Hill estimator $\hat{\lambda}_H$
- Simpson and Tawn (2022) recently provided the first smooth estimator (others now available).
- This estimator exploits the results of Nolde and Wadsworth (2022).
- Ryan has thankfully described this theory to you already!

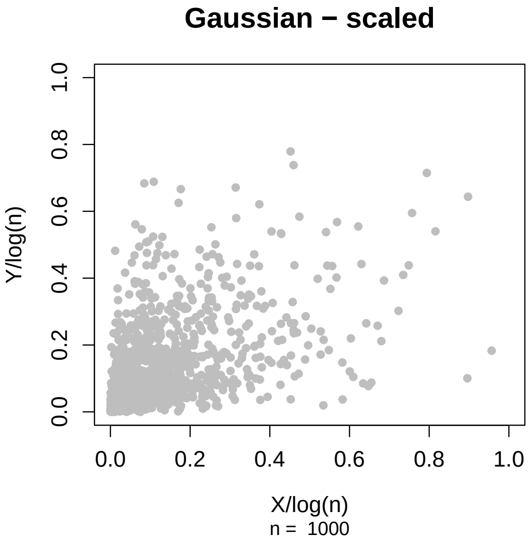
Existing ADF estimators

- Given n independent realisations from (X, Y), we consider the shape of

$$C_n^* := \{(X_i, Y_i)/\log n; \ i = 1, \ldots, n\},$$

as $n \to \infty$.

Existing ADF estimators



Existing ADF estimators

- As $n \to \infty$, we have that C_n^* converges onto the set

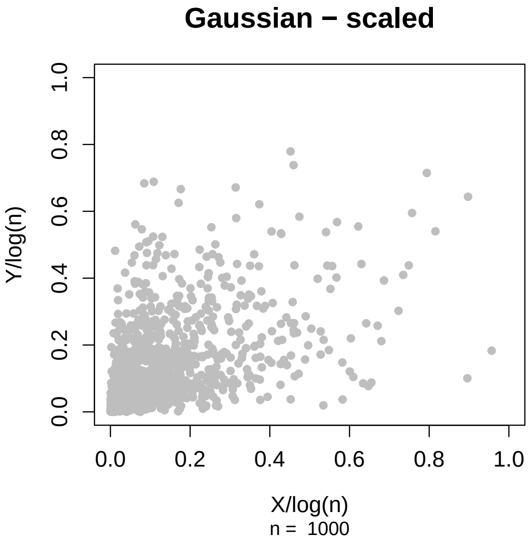
$$G^* = \{(x, y) : g(x, y) \leq 1\} \subseteq [0, 1]^2.$$

  where g is the so-called Gauge function.
- Interest lies in studying the boundary set given by

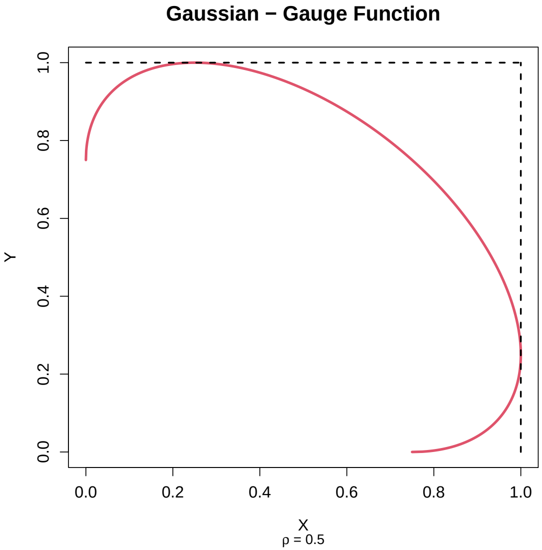
$$G = \{(x, y) : g(x, y) = 1\} \subset [0, 1]^2.$$

- Interpretation: scaled data points converge onto sets with nice shapes.

Existing ADF estimators



Existing ADF estimators



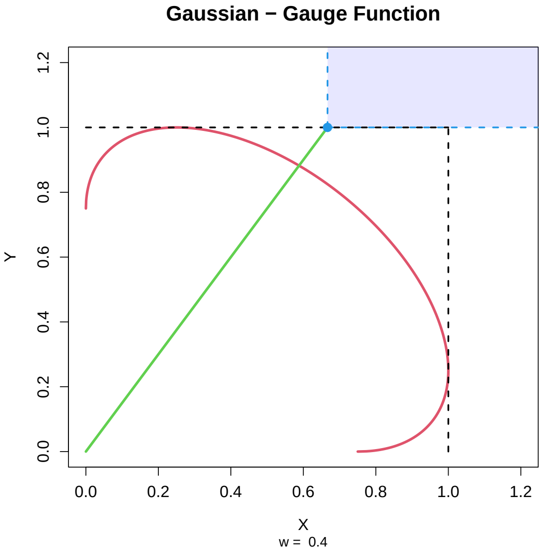
Existing ADF estimators

- The boundary G links the models of Heffernan and Tawn (2004) and Wadsworth and Tawn (2013).

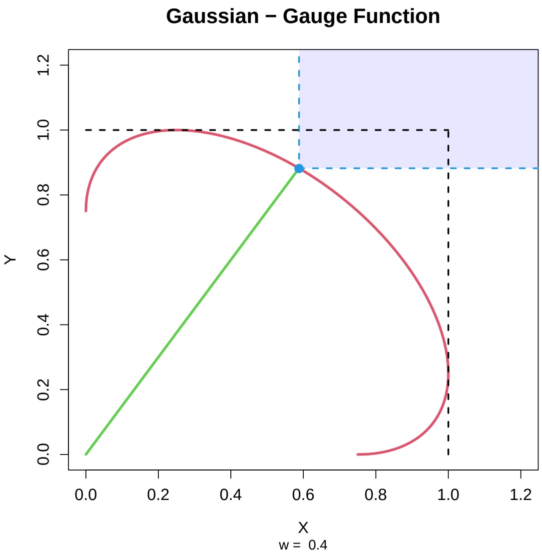
Existing ADF estimators

- We have that
$$\lambda(w) = \frac{\max(w, 1 - w)}{s_w},$$
where
$$s_w = \min \{ s \in [0, 1] : sS_w \cap G = \emptyset \}.$$

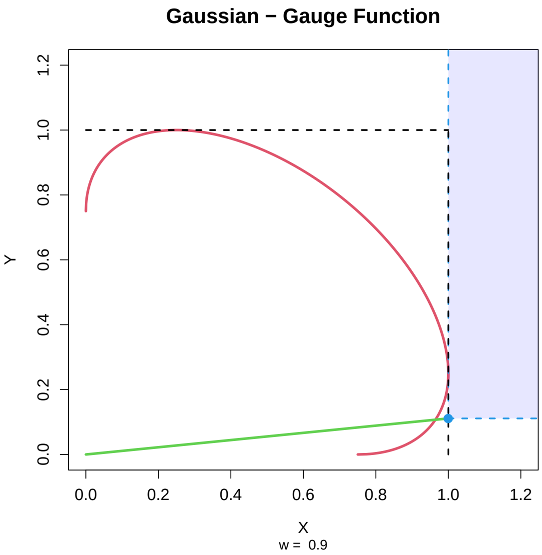
Existing ADF estimators



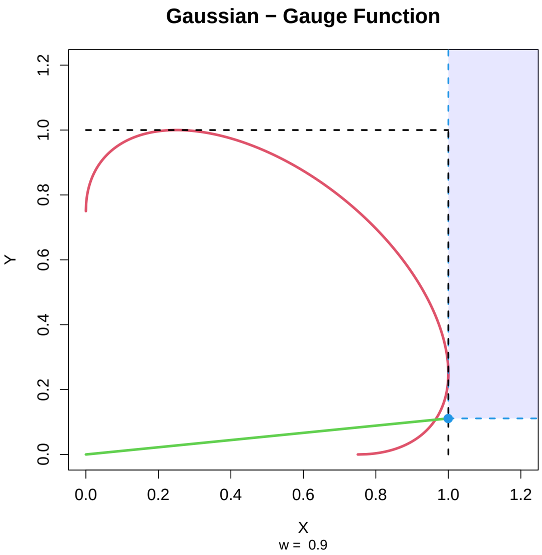
Existing ADF estimators



Existing ADF estimators



Existing ADF estimators

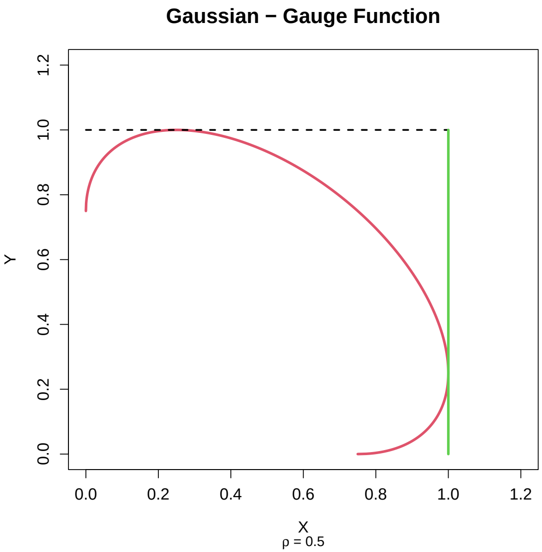


Existing ADF estimators

- Now consider $\alpha_{y|x}$ (i.e., the fancy regression 'slope').
- We have that

$$\alpha_{y|x} = \max\left\{\tilde{\alpha} \in [0,1] : g\left(1, \tilde{\alpha}\right) = 1\right\}.$$

Existing ADF estimators



Existing ADF estimators

Once we have G, we get the rest for free.

- Simpson and Tawn (2022) provided the first approach for estimating G.
- Let $\hat{\lambda}_{ST}$ denote the smooth ADF estimator obtained using this approach.

Novel estimators

- Recall our objective: to provide smooth estimates of the ADF.

Novel estimators

- Recall our objective: to provide smooth estimates of the ADF.
- Lots of approaches are available for smooth estimation of the Pickands' dependence function (e.g. Guillotte and Perron, 2016; Marcon et al., 2016).
- Many use Bernstein-Bézier polynomials.

Novel estimators

- We adapt Bernstein-Bézier polynomials for ADF estimation.

Novel estimators

- We adapt Bernstein-Bézier polynomials for ADF estimation.
- Consider this family

$$\mathcal{B}_k^* = \left\{ (1-w)^k + \sum_{i=1}^{k-1} \beta_i \binom{k}{i} w^i (1-w)^{k-i} + w^k =: f(w) \,\middle|\, w \in [0,1], \beta \in [0,\infty)^{k-1} \text{ such that } f(w) \geq \max(w, 1-w) \right\}. \tag{1}$$

- Satisfies all theoretical conditions for the ADF.

Novel estimators

- We assume $\lambda \in \mathcal{B}_k^*$, so $\lambda(w) = \lambda(w; \beta)$.

Novel estimators

- We assume $\lambda \in \mathcal{B}_k^*$, so $\lambda(w) = \lambda(w; \boldsymbol{\beta})$.
- What remains is to estimate $\boldsymbol{\beta} \in [0, \infty)^{k-1}$, which we achieve through a composite likelihood approach.

Novel estimators

- We multiply this function over rays (components) $w \in [0,1]$ to give one overall likelihood function.

$$\mathcal{L}_C(\beta) = \prod_{w \in \mathcal{W}} \prod_{t_w^* \in \mathbf{t}_w^*} \lambda(w;\ \beta) e^{-\lambda(w;\ \beta) t_w^*}.$$

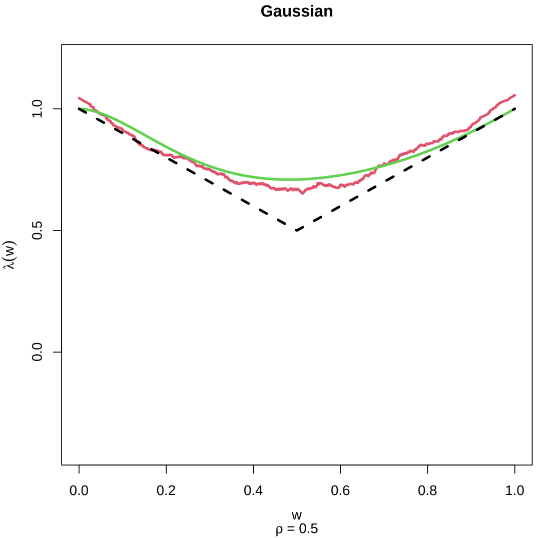
Novel estimators

- We multiply this function over rays (components) $w \in [0, 1]$ to give one overall likelihood function.

$$\mathcal{L}_C(\boldsymbol{\beta}) = \prod_{w \in \mathcal{W}} \prod_{t_w^* \in \mathbf{t}_w^*} \lambda(w;\ \boldsymbol{\beta}) e^{-\lambda(w;\ \boldsymbol{\beta}) t_w^*}.$$

- Let $\hat{\boldsymbol{\beta}}_{CL}$ denote the maximum likelihood estimator of $\boldsymbol{\beta}$.
- Corresponding ADF estimator given by $\hat{\lambda}_{CL}(\cdot) = \hat{\lambda}(\cdot; \boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_{CL})$.

Novel estimators



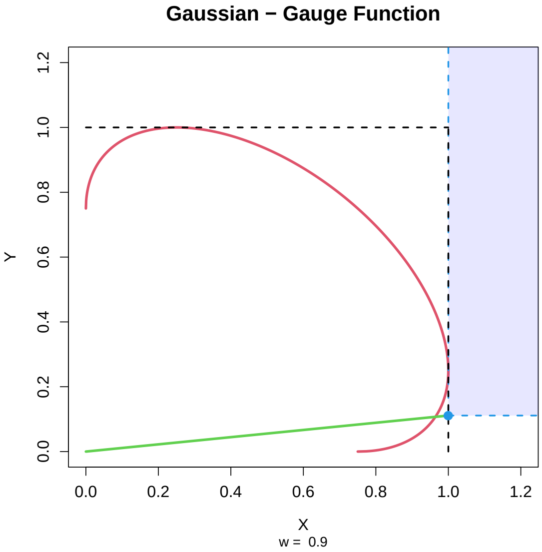
Novel estimators

- We also exploited a corollary from Nolde and Wadsworth (2022) to improve ADF estimation.

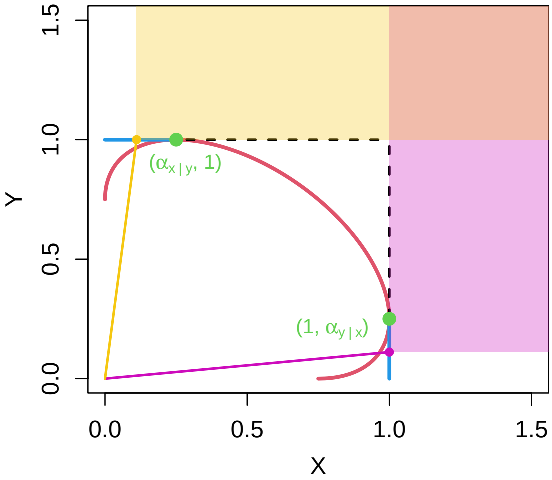
Novel estimators

- We also exploited a corollary from Nolde and Wadsworth (2022) to improve ADF estimation.
- In particular, we found that for all $w \in [0, \alpha_{x|y}/(1+\alpha_{x|y})] \bigcup [1/(1+\alpha_{y|x}), 1]$, $\lambda(w) = \max(w, 1-w)$.
- Interpretation: if we know $\alpha_{x|y}$, $\alpha_{y|x}$, we know where $\lambda(w, 1-w) = \max(w, 1-w)$ (and vice versa).

Novel estimators



Novel estimators



Novel estimators

Procedure:

1. Use 'standard' techniques to obtain $\hat{\alpha}_{y|x}$ and $\hat{\alpha}_{x|y}$.

Novel estimators

Procedure:

1. Use 'standard' techniques to obtain $\hat{\alpha}_{y|x}$ and $\hat{\alpha}_{x|y}$.
2. Set $\lambda(w) = \max(w, 1-w)$ for all $w \in [0, \hat{\alpha}_{x|y}/(1+\hat{\alpha}_{x|y})] \bigcup [1/(1+\hat{\alpha}_{y|x}), 1]$.

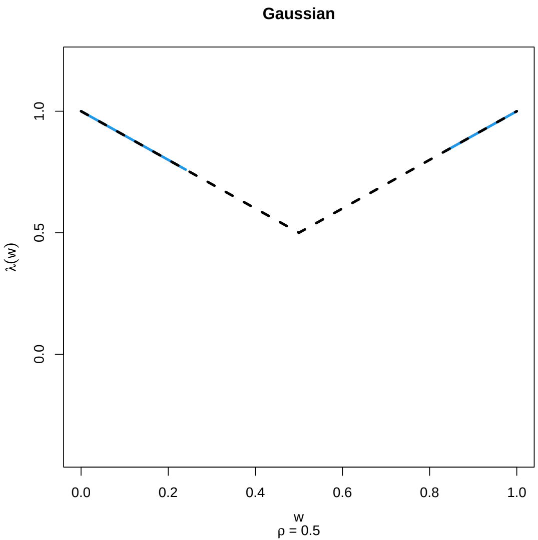
Novel estimators

Procedure:

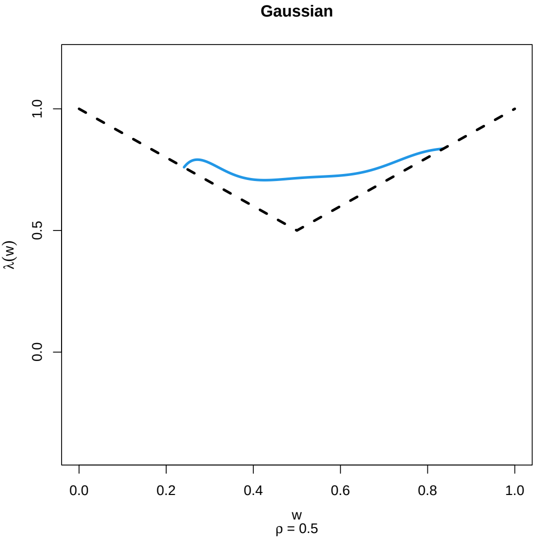
1. Use 'standard' techniques to obtain $\hat{\alpha}_{y|x}$ and $\hat{\alpha}_{x|y}$.
2. Set $\lambda(w) = \max(w, 1-w)$ for all $w \in [0, \hat{\alpha}_{x|y}/(1+\hat{\alpha}_{x|y})] \bigcup [1/(1+\hat{\alpha}_{y|x}), 1]$.
3. Estimate λ using composite likelihood for $w \in (\hat{\alpha}_{x|y}/(1+\hat{\alpha}_{x|y}), 1/(1+\hat{\alpha}_{y|x}))$.

This gives us a second composite likelihood estimator $\hat{\lambda}_{CL2}$.

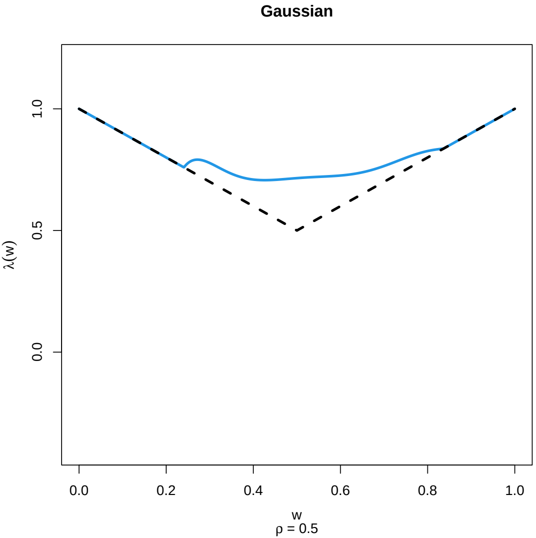
Novel estimators



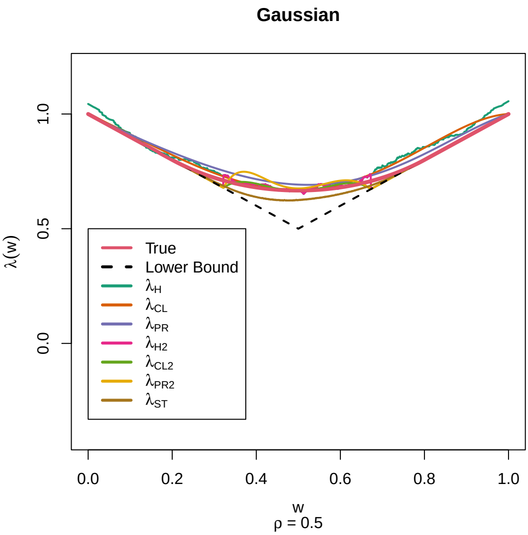
Novel estimators



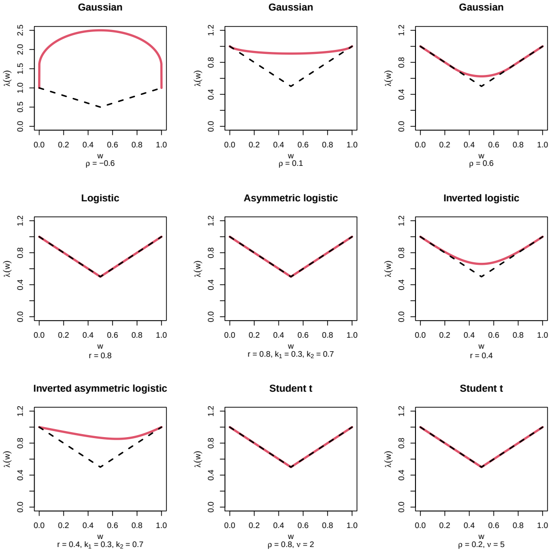
Novel estimators



Simulation study



Simulation study



Simulation study

Table: RMISE values (multiplied by 100) for each estimator and copula combination. Smallest RMISE values in each row are highlighted in bold, with values reported to 3 significant figures.

Copula	$\hat{\lambda}_H$	$\hat{\lambda}_{CL}$	$\hat{\lambda}_{PR}$	$\hat{\lambda}_{H2}$	$\hat{\lambda}_{CL2}$	$\hat{\lambda}_{PR2}$	$\hat{\lambda}_{ST}$
Copula 1	**61.1**	61.3	66.2	61.4	61.9	66.7	63.7
Copula 2	3.55	3.33	3.64	3.51	3.33	3.63	**2.95**
Copula 3	3.78	3.48	3.84	3.27	3.22	3.57	**1.09**
Copula 4	4.9	4.79	6.92	4.28	4.25	6.17	**2.77**
Copula 5	14.1	14.1	17.1	14.1	14.1	17	**12.1**
Copula 6	2.51	1.97	2.15	2	**1.74**	1.9	2.12
Copula 7	2.93	**2.64**	2.88	2.87	2.66	2.89	3.96
Copula 8	2.49	2.72	2.95	0.66	**0.6**	0.789	1.87
Copula 9	12.1	12	14.9	12	12	14.9	**11.1**

Simulation study

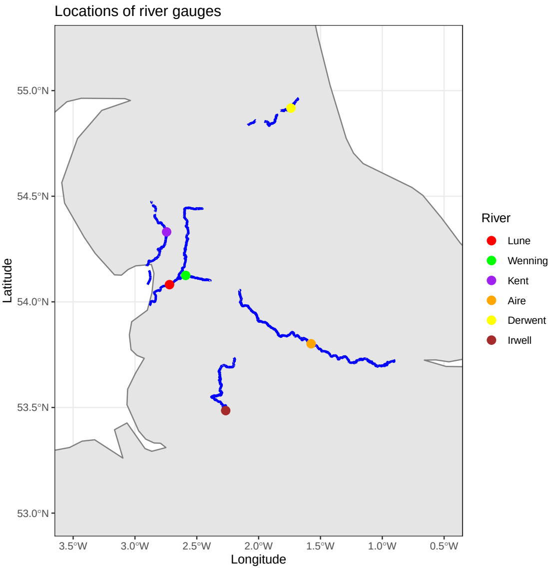
Smooth estimators $>$ pointwise estimators

Simulation study

Smooth estimators $>$ pointwise estimators

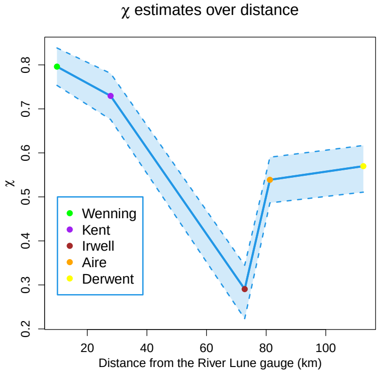
Important insight: estimators linked to Gauge functions/limit sets performed best.

Case study

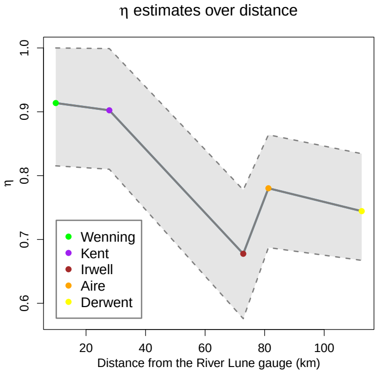


Locations of river gauges
55.0°N
54.5°N
54.0°N
53.5°N
53.0°N
Latitude
3.5°W
3.0°W
2.5°W
2.0°W
1.5°W
1.0°W
0.5°W
Longitude
River
Lune
Wenning
Kent
Aire
Derwent
Irwell

Case study

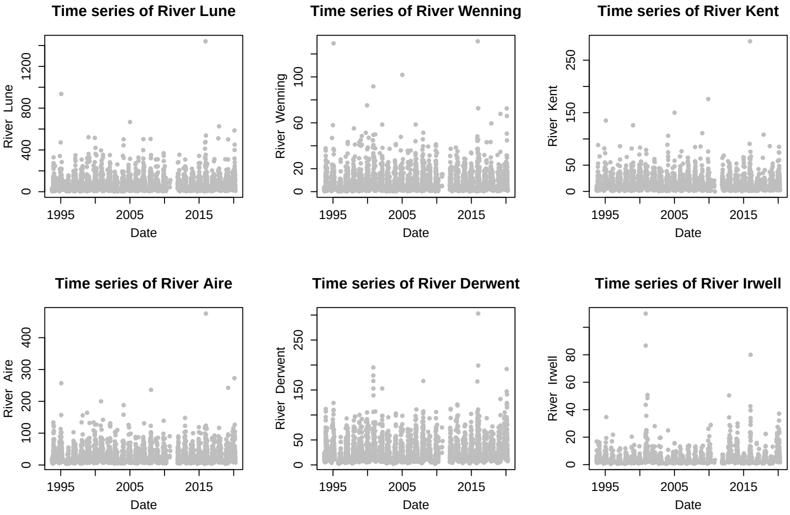


χ estimates over distance
χ
0.2
0.3
0.4
0.5
0.6
0.7
0.8
20
40
60
80
100
Distance from the River Lune gauge (km)
Wenning
Kent
Irwell
Aire
Derwent

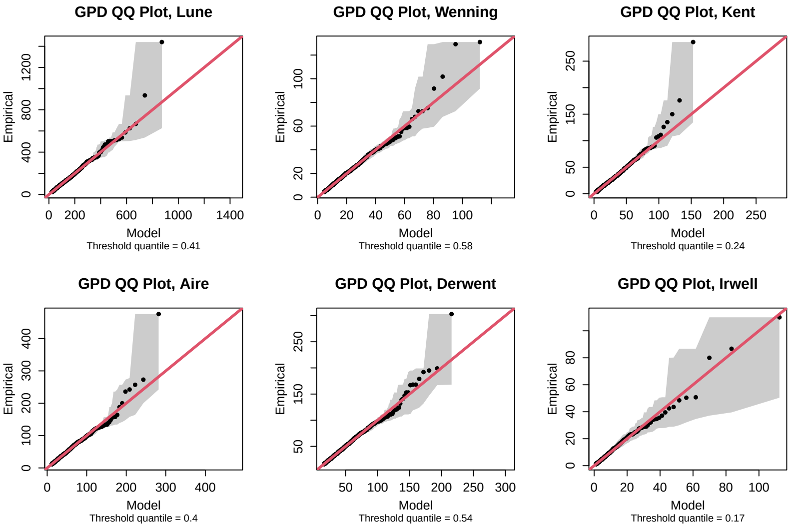


η estimates over distance
η
0.6
0.7
0.8
0.9
1.0
20
40
60
80
100
Distance from the River Lune gauge (km)
Wenning
Kent
Irwell
Aire
Derwent

Case study

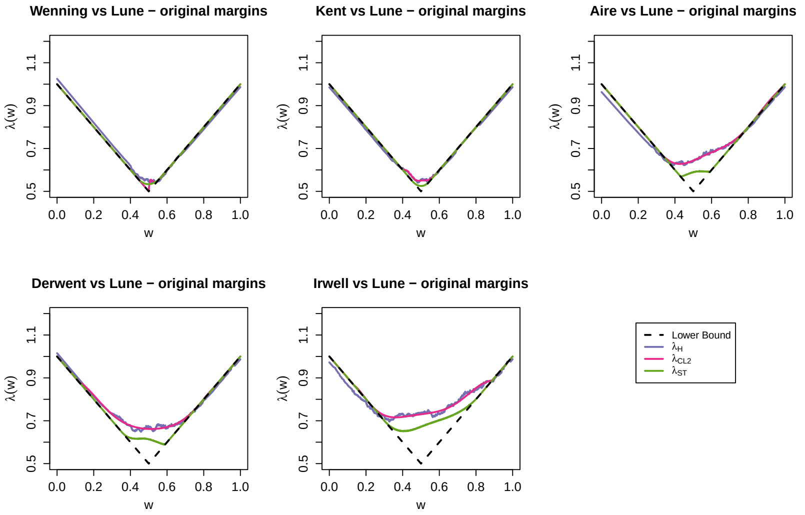

Time series of River Lune
River Lune
Date
0
400
800
1200
1995
2005
2015
Time series of River Wenning
River Wenning
Date
0
20
60
100
1995
2005
2015
Time series of River Kent
River Kent
Date
0
50
150
250
1995
2005
2015
Time series of River Aire
River Aire
Date
0
100
200
300
400
1995
2005
2015
Time series of River Derwent
River Derwent
Date
0
50
150
250
1995
2005
2015
Time series of River Irwell
River Irwell
Date
0
20
40
60
80
1995
2005
2015

Case study



See Conor Murphy's talk.

Case study


Wenning vs Lune − original margins
Kent vs Lune − original margins
Aire vs Lune − original margins
Derwent vs Lune − original margins
Irwell vs Lune − original margins
λ(w)
w
0.5
0.7
0.9
1.1
0.0
0.2
0.4
0.6
0.8
1.0
Lower Bound
λ_H
λ_{CL2}
λ_{ST}

Case Study

Return curves defined by the set

$$\mathrm{RC}(p) := \{(x, y) \in \mathbb{R}^2 \mid \Pr(X > x, Y > y) = p\},$$

where p is very small.

Case Study

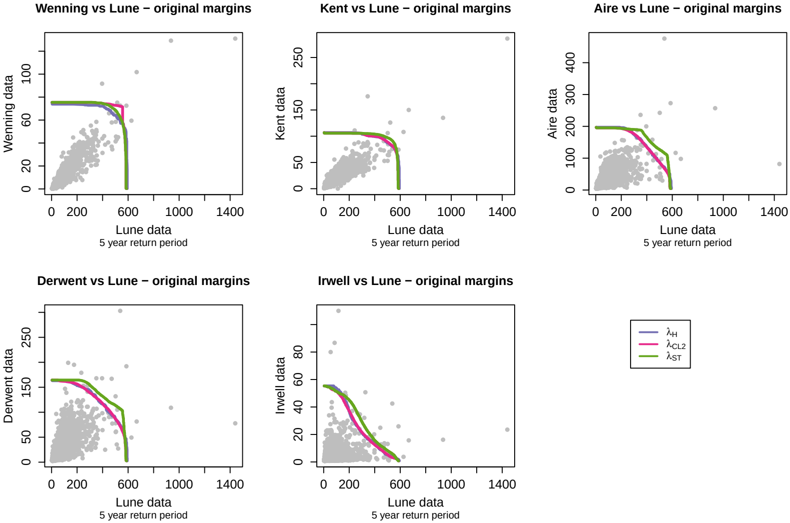
Return curves defined by the set

$$\mathrm{RC}(p) := \{(x, y) \in \mathbb{R}^2 \mid \Pr(X > x, Y > y) = p\},$$

where p is very small.

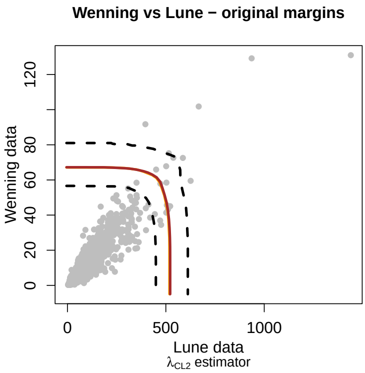
Used to quantify joint extreme risk in practice (Murphy-Barltrop et al., 2023b).

Case study

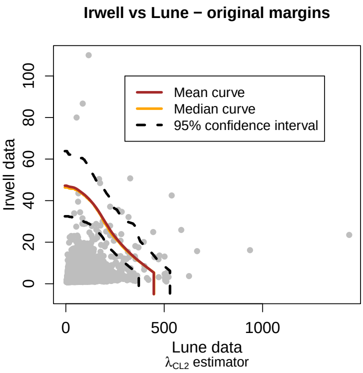


Return curve diagnostics indicate good accuracy.

Case study



Wenning vs Lune − original margins
Wenning data
0
20
40
60
80
120
0
500
1000
Lune data
λ_{CL2} estimator



Irwell vs Lune − original margins
Mean curve
Median curve
95% confidence interval
Irwell data
0
20
40
60
80
100
0
500
1000
Lune data
λ_{CL2} estimator

Discussion

In summary

- We have proposed a range of global estimators for the ADF and compared these to existing techniques.
- Global estimators consistently outperform pointwise techniques.
- ADF is a valuable tool for estimation of joint extremes under asymptotic independence.

Discussion

- More generally, this paper illustrates the **benefits** of applying the **limit set representation** for multivariate extremes in practice (both directly and indirectly).

Discussion

- Lack of theoretical results for estimators.
- Limited to bivariate setting.
- Selecting tuning parameters (not discussed here).

See Murphy-Barltrop et al. (2023a) for further details.

Thanks for listening!

Does anyone have any questions?

Email: *callum_john.murphy-barltrop@tu-dresden.de*

References I

Guillotte, S. and Perron, F. (2016). Polynomial pickands functions. *Bernoulli*, 22:213–241.

Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 66:497–546.

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

Ledford, A. W. and Tawn, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 59:475–499.

Marcon, G., Padoan, S. A., and Antoniano-Villalobos, I. (2016). Bayesian inference for the extremal dependence. *Electronic Journal of Statistics*, 10:3310–3337.

References II

Murphy-Barltrop, C. J. R., Wadsworth, J. L., and Eastoe, E. F. (2023a). Improving estimation for asymptotically independent bivariate extremes via global estimators for the angular dependence function. *arXiv*, 2303.13237.

Murphy-Barltrop, C. J. R., Wadsworth, J. L., and Eastoe, E. F. (2023b). New estimation methods for extremal bivariate return curves. *Environmetrics*.

Nolde, N. and Wadsworth, J. L. (2022). Linking representations for multivariate extremes via a limit set. *Advances in Applied Probability*, 54:688–717.

Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer New York.

Simpson, E. S. and Tawn, J. A. (2022). Estimating the limiting shape of bivariate scaled sample clouds for self-consistent inference of extremal dependence properties. *arXiv*, 2207.02626.

References III

Wadsworth, J. L. and Tawn, J. A. (2013). A new representation for multivariate tail probabilities. *Bernoulli*, 19:2689–2714.