Wasserstein–Aitchison GAN for angular measures of multivariate extremes

Stéphane Lhaut[1] Holger Rootzén[2] Johan Segers[1,3]

[1]UCLouvain, LIDAM, Institute of Statistics, Biostatistics and Actuarial Sciences
[2]Chalmers University of Technology, Department of Mathematical Sciences
[3]KU Leuven, Department of Mathematics

Workshop on Generative AI Modelling for Extreme Events
Edinburgh, June 13–14 2025

Outline

Angular measure modeling

Multivariate Generalized Pareto modeling

Numerical results

- Simulated data: logistic model
- Real data: daily returns of industry portfolios

Conclusions and extensions

Angular measure modeling

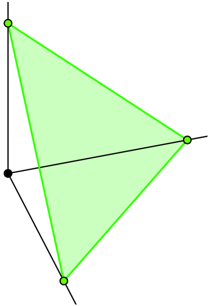
Multivariate Generalized Pareto modeling

Numerical results

- Simulated data: logistic model
- Real data: daily returns of industry portfolios

Conclusions and extensions

Multivariate tail modeling and angular measure



$$\boldsymbol{X} \in \mathbb{R}^d \xrightarrow{\text{Pareto(1) margins}} \boldsymbol{V} = \left(\frac{1}{1-F_j(X_j)}\right)_{j=1}^{d}$$

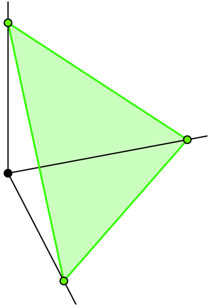
Regular variation

$$\lim_{t\to\infty} t \, \mathrm{P}\left(t^{-1}\boldsymbol{V} \in B\right) = \nu(B)$$

$R = |\boldsymbol{V}|_1 \qquad \boldsymbol{W} = R^{-1}\boldsymbol{V}$

$$\lim_{t\to\infty} \mathrm{P}\left(\boldsymbol{W} \in A \mid R \geq t\right) = \Phi(A)$$

Multivariate tail modeling and angular measure



$\boldsymbol{X} \in \mathbb{R}^d \xrightarrow{\text{Pareto(1) margins}} \boldsymbol{V} = \left(\frac{1}{1-F_j(X_j)}\right)_{j=1}^{d}$

Regular variation

$$\lim_{t\to\infty} t \, \mathrm{P}\left(t^{-1}\boldsymbol{V} \in B\right) = \nu(B)$$

$R = |\boldsymbol{V}|_1 \qquad \boldsymbol{W} = R^{-1}\boldsymbol{V}$

$$\lim_{t\to\infty} \mathrm{P}\left(\boldsymbol{W} \in A \mid R \geq t\right) = \Phi(A)$$

Main goal: “model” Φ via a generative method

Empirical data from Φ

- $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ i.i.d. distributed as $\boldsymbol{X}$

Empirical data from Φ

- $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ i.i.d. distributed as $\boldsymbol{X}$
- Empirical Pareto(1) standardization:

$$\widehat{\boldsymbol{V}}_i = \left(\frac{1}{1 - \widehat{F}_j(X_{ij})} \right)_{j=1}^{d}, \qquad \widehat{F}_j(t) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbb{I}\{X_{ij} \leq t\}$$

Empirical data from Φ

- $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ i.i.d. distributed as $\boldsymbol{X}$
- Empirical Pareto(1) standardization:

$$\widehat{\boldsymbol{V}}_i = \left(\frac{1}{1 - \widehat{F}_j(X_{ij})} \right)_{j=1}^{d}, \qquad \widehat{F}_j(t) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbb{I}\{X_{ij} \le t\}$$

- Extracting "large angles": fix some large $r > 0$ (here $r = n/k$ with $k \in \{1, \dots, n\}$)

$$\text{If} \quad R_i = |\widehat{\boldsymbol{V}}_i|_1 \ge r, \qquad \boldsymbol{W}_i = R_i^{-1} \widehat{\boldsymbol{V}}_i$$

$\boldsymbol{W}_1, \dots, \boldsymbol{W}_K$, with $K = \sum_{i=1}^{n} \mathbb{I}\{R_i \ge r\}$, considered as Φ-distributed

Data transformation to linear space

Two facts will hamper the learning process:

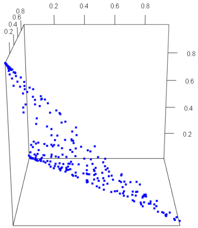
- $\boldsymbol{W}_i$ take values in the simplex $\Delta_{d-1} = \{\boldsymbol{x} \in [0,1]^d : |\boldsymbol{x}|_1 = 1\}$
- $K \overset{\mathcal{L}}{\approx} \mathrm{Bin}(n, k/n)$

Data transformation to linear space

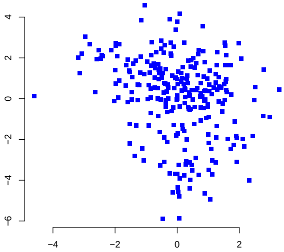
Two facts will hamper the learning process:

- $\boldsymbol{W}_i$ take values in the simplex $\Delta_{d-1} = \{\boldsymbol{x} \in [0,1]^d : |\boldsymbol{x}|_1 = 1\}$
- $K \overset{\mathcal{L}}{\approx} \mathrm{Bin}(n, k/n)$

Transform $\boldsymbol{W}_i$ to coordinates $\boldsymbol{W}_i^*$ in an orthonormal basis of Δ_{d-1}



$\longrightarrow$



Aitchison simplex: the CLR space

Assumption: concentration on the open simplex

$$\Phi(\Delta^{\circ}_{d-1}) = 1 \text{ where } \Delta^{\circ}_{d-1} = \{\boldsymbol{x} \in (0,1)^d : |\boldsymbol{x}|_1 = 1\}$$

Aitchison simplex: the CLR space

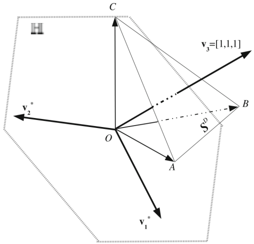
Assumption: concentration on the open simplex

$$\Phi(\Delta^{\circ}_{d-1}) = 1 \text{ where } \Delta^{\circ}_{d-1} = \{\boldsymbol{x} \in (0,1)^d : |\boldsymbol{x}|_1 = 1\}$$

- J. Aitchison (1926–2016) inner product space structure on Δ°_{d-1} via the isometry

$$\operatorname{clr} : \boldsymbol{w} \in \Delta^{\circ}_{d-1} \mapsto \operatorname{clr}(\boldsymbol{w}) = \left(\log \frac{w_j}{(\prod_{i=1}^d w_i)^{1/d}}\right)_{j=1}^d \in \mathbb{H} = \{\boldsymbol{x} \in \mathbb{R}^d : \langle \boldsymbol{x}, \mathbf{1}_d \rangle = 0\}$$

with inverse being the (restriction to $\mathbb{H}$) of the softmax function



Aitchison simplex: ON basis

- **Orthonormal basis** of $\Delta^{\mathrm{o}}_{d-1}$ given by $\{\boldsymbol{e}_j^* = \mathrm{softmax}(\boldsymbol{e}_j) : j = 1, \ldots, d-1\}$ with

$$\boldsymbol{e}_j = \sqrt{\frac{j}{j+1}} \left(\underbrace{j^{-1}, \ldots, j^{-1}}_{j \text{ times}}, -1, 0 \ldots, 0 \right) \in \mathbb{H}$$

Aitchison simplex: ON basis

- Orthonormal basis of Δ°_{d-1} given by $\{\boldsymbol{e}_j^* = \mathrm{softmax}(\boldsymbol{e}_j) : j = 1, \ldots, d-1\}$ with

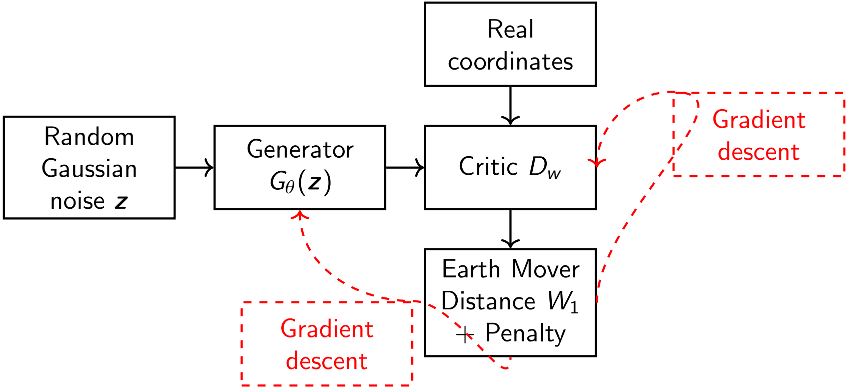
$$\boldsymbol{e}_j = \sqrt{\frac{j}{j+1}} \left(\underbrace{j^{-1}, \ldots, j^{-1}}_{j \text{ times}}, -1, 0 \ldots, 0 \right) \in \mathbb{H}$$

- $\boldsymbol{w} \in \Delta^{\circ}_{d-1}$ is decomposed as

$$\boldsymbol{w} = \bigoplus_{j=1}^{d-1} \langle \boldsymbol{w}, \boldsymbol{e}_j^* \rangle_A \, \boldsymbol{e}_j^* = \bigoplus_{j=1}^{d-1} \langle \mathrm{clr}(\boldsymbol{w}), \boldsymbol{e}_j \rangle \, \boldsymbol{e}_j^*$$

Wasserstein–Aitchison Generative Adversarial Networks

WGAN algorithm on the coordinates $\boldsymbol{w}_i^* = (\langle \mathrm{clr}(\boldsymbol{w}_i), \boldsymbol{e}_j \rangle)_{j=1}^{d-1}$



- the GAME:

$$\min_\theta \max_w \left\{ \int_{\mathcal{X}} D_w(x)\,\mathrm{dP}(x) - \int_{\mathcal{Z}} D_w\left(G_\theta(z)\right)\,\mathrm{dP}_Z(z) \right\},$$

Penalty term

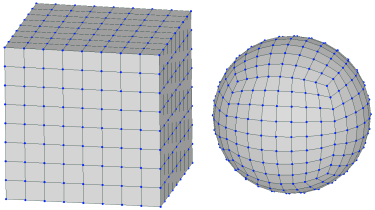
- $\lambda, \rho > 0$

$$\underbrace{\lambda \frac{1}{m} \sum_{i=1}^{m} (|\nabla_{\boldsymbol{w}^*} D_w(\widehat{\boldsymbol{w}}_i^*)|_2 - 1)^2}_{\text{Lipschitz constraint}} + \underbrace{\rho \left| \frac{1}{m} \sum_{i=1}^{m} \mathrm{softmax}(VG_\theta(\boldsymbol{z}_i)) - \frac{\mathbf{1}_d}{d} \right|}_{\text{marginal constraints}}$$

Change of norm

- $\boldsymbol{\Theta}_1, \dots, \boldsymbol{\Theta}_n$ sample from Φ on Δ_{d-1}
- Estimator for an arbitrary norm $\| \cdot \|$:

$$\sum_{i=1}^{n} \Lambda_i \delta_{\frac{\boldsymbol{\Theta}_i}{\|\boldsymbol{\Theta}_i\|}}, \qquad \Lambda_i = \frac{\|\boldsymbol{\Theta}_i\|}{\sum_{i=1}^{n} \|\boldsymbol{\Theta}_i\|}$$



Angular measure modeling

Multivariate Generalized Pareto modeling

Numerical results

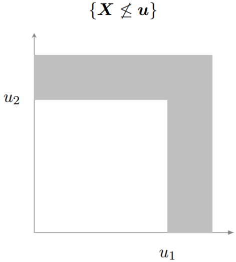
- Simulated data: logistic model
- Real data: daily returns of industry portfolios

Conclusions and extensions

Going beyond the dependence structure

$\boldsymbol{u}_* =$ upper endpoint of $\boldsymbol{X}$

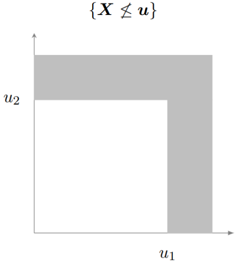
$$\boxed{\boldsymbol{X} \mid \boldsymbol{X} \not\leq \boldsymbol{u}, \qquad \text{as } \boldsymbol{u} \to \boldsymbol{u}_*}$$



Going beyond the dependence structure

$\boldsymbol{u}_* =$ upper endpoint of $\boldsymbol{X}$

$$\boldsymbol{X} \mid \boldsymbol{X} \not\leq \boldsymbol{u}, \qquad \text{as } \boldsymbol{u} \to \boldsymbol{u}_*$$



Assumption: Generalized Pareto limit of marginal tails

For every $j \in \{1, \ldots, d\}$, there exists $\xi_j \in \mathbb{R}$ and a function $\sigma_j(\cdot) : (-\infty, u_{*j}) \to (0, \infty)$

$$\forall y > 0, \qquad \lim_{u \nearrow u_{*j}} \mathrm{P}\left(\frac{X_j - u}{\sigma_j(u)} > y \,\Big|\, X_j > u \right) = (1 + \xi_j y)_+^{-1/\xi_j},$$

where $a_+ \stackrel{\text{def}}{=} \max(a, 0)$ for $a \in \mathbb{R}$; if $\xi_j = 0$, the limit is to be understood as $\exp(-y)$.

Regular variation + GP assumptions $\implies$ MGPD convergence [Lhaut et al., 2025, Proposition 1]

Proposition (Convergence to a Multivariate Generalized Pareto distribution)

Under the previous assumptions, along the curve

$$\boldsymbol{u} : t \in (1, \infty) \mapsto \boldsymbol{u}(t) = \left(F_j^{-1}(1 - 1/t)\right)_{j=1}^{d}$$

we have

$$\frac{\boldsymbol{X} - \boldsymbol{u}(t)}{\boldsymbol{\sigma}(\boldsymbol{u}(t))} \mid \boldsymbol{X} \not\leq \boldsymbol{u}(t) \xrightarrow{\mathcal{L}} \frac{\boldsymbol{Y}^{\boldsymbol{\xi}} - 1}{\boldsymbol{\xi}}, \qquad t \to \infty,$$

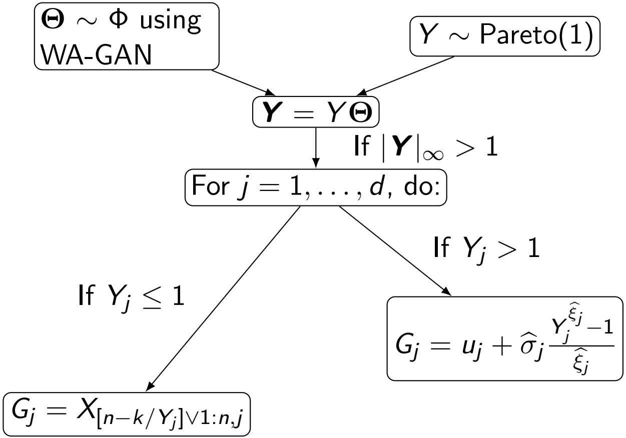
where $\boldsymbol{Y}$ is Multivariate Pareto distributed, i.e.,

$$\boldsymbol{Y} \stackrel{\mathcal{L}}{=} (Y\boldsymbol{\Theta} \mid Y\boldsymbol{\Theta} \not\leq 1),$$

with Y unit-Pareto distributed and $\boldsymbol{\Theta} \sim \Phi$ independent.

Sampling from $\boldsymbol{X} \mid \boldsymbol{X} \not\leq \boldsymbol{u}$:

Input: data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, integer $k \in [1, n]$, ML estimates $(\widehat{\boldsymbol{\xi}}, \widehat{\boldsymbol{\sigma}})$ of marginal GP parameters



Output: $\boldsymbol{G} = (G_j)_{j=1}^d$ sample from $\boldsymbol{X} \mid \boldsymbol{X} \not\leq \boldsymbol{u}$

Angular measure modeling

Multivariate Generalized Pareto modeling

Numerical results

- Simulated data: logistic model
- Real data: daily returns of industry portfolios

Conclusions and extensions

Performance measures

Dependence score:

$$E_{\overline{\theta}}(k) = \frac{1}{\binom{d}{k}} \sum_{\substack{J \subseteq \{1,\ldots,d\} \\ |J|=k}} \left| 1 - \frac{\theta_J^{\mathcal{G}}}{\theta_J^{\mathcal{T}}} \right|, \qquad k \in \{2, 3, \ldots\}$$

based on the extremal coefficients

$$\theta_J = d \cdot \int_{\Delta_{d-1}} \bigvee_{j \in J} w_j \, \mathrm{d}\Phi(\boldsymbol{w}) \in [1, |J|]$$

Performance measures

Dependence score:

$$E_{\overline{\theta}}(k) = \frac{1}{\binom{d}{k}} \sum_{\substack{J \subseteq \{1,\dots,d\} \\ |J|=k}} \left| 1 - \frac{\theta_J^{\mathcal{G}}}{\theta_J^{\mathcal{T}}} \right|, \qquad k \in \{2, 3, \dots\}$$

based on the extremal coefficients

$$\theta_J = d \cdot \int_{\Delta_{d-1}} \bigvee_{j \in J} w_j \, \mathrm{d}\Phi(\boldsymbol{w}) \in [1, |J|]$$

Extremes score: generalized Fréchet Inception Distance (FID) score

$$W_2^2(\boldsymbol{X}_{\boldsymbol{u}}^{\mathcal{G}}, \boldsymbol{X}_{\boldsymbol{u}}^{\mathcal{T}}) = \inf_{\pi \in \Pi\left(\frac{\mathbf{1}_{n_\mathcal{G}}}{n_\mathcal{G}}, \frac{\mathbf{1}_{n_\mathcal{T}}}{n_\mathcal{T}}\right)} \sum_{i=1}^{n_\mathcal{G}} \sum_{j=1}^{n_\mathcal{T}} \|(\boldsymbol{X}_{\boldsymbol{u}}^{\mathcal{G}})_i - (\boldsymbol{X}_{\boldsymbol{u}}^{\mathcal{T}})_j\|_2^2 \, \pi_{ij}$$

Methodology and computations

- Architectures: fully connected MLP with leaky ReLU activation function
- Hyperparameter search: random search among 2000 possible models (batch size, dimension of the latent space, penalty parameters, number of layers, ...)
- Comparison with existing methods: we compare our performance to (own implementation of) Heavy–Tailed GAN [Girard et al., 2024] and Generalized Pareto GAN [Li et al., 2024]
- Software: training on PyTorch, validation and testing on R
- Hardware: Lemaitre4 and NIC5 computer clusters available from the Consortium of Équipements de Calcul Intensif [1]

[1] https://www.ceci-hpc.be/clusters.html

- Angular measure modeling
- Multivariate Generalized Pareto modeling
- Numerical results
  - Simulated data: logistic model
  - Real data: daily returns of industry portfolios
- Conclusions and extensions

Results

- $\boldsymbol{X}$ has Gumbel copula such that Kendall's correlation equals $\tau \in \{1/4, 1/2, 3/4\}$ and Pareto margins with parameter $\alpha = 2$ in dimension $d \in \{10, 20, 50\}$

Results

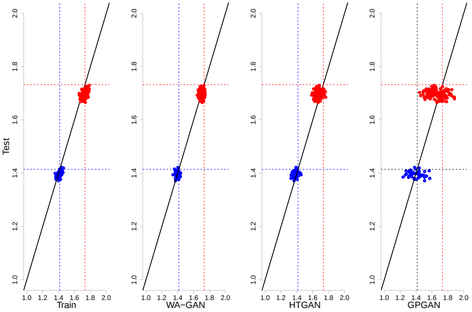
- $\boldsymbol{X}$ has Gumbel copula such that Kendall's correlation equals $\tau \in \{1/4, 1/2, 3/4\}$ and Pareto margins with parameter $\alpha = 2$ in dimension $d \in \{10, 20, 50\}$
- $n_{\text{train}} = 10\,000$, $n_{\text{val}} = 5\,000$, $n_{\text{test}} = 20\,000$ and $k = \sqrt{n_{\text{train}}}$ for each algorithm

Results

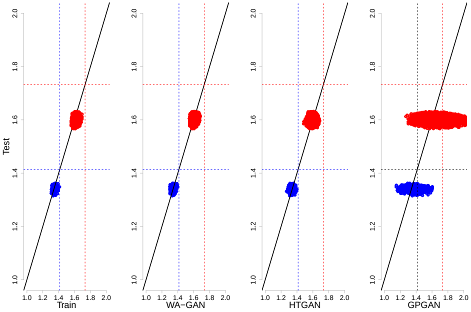
- $\boldsymbol{X}$ has Gumbel copula such that Kendall's correlation equals $\tau \in \{1/4, 1/2, 3/4\}$ and Pareto margins with parameter $\alpha = 2$ in dimension $d \in \{10, 20, 50\}$
- $n_{\text{train}} = 10\,000$, $n_{\text{val}} = 5\,000$, $n_{\text{test}} = 20\,000$ and $k = \sqrt{n_{\text{train}}}$ for each algorithm

		Dependence score			Extremes score		
τ	Model	$d=10$	$d=20$	$d=50$	$d=10$	$d=20$	$d=50$
$\tau = \frac{1}{4}$	WA-GAN	**0.0103**	**0.0074**	**0.0054**	67.311	**62.905**	**42.825**
	HTGAN	0.0146	0.0185	0.0199	1479.9	3106.7	1170.7
	GPGAN	0.0262	0.0321	0.0326	**66.663**	65.429	52.133
$\tau = \frac{1}{2}$	WA-GAN	**0.0176**	0.0207	**0.0097**	**66.734**	**52.832**	**32.468**
	HTGAN	0.0247	**0.0179**	0.0126	2725.3	1447.9	2367.4
	GPGAN	0.0511	0.0671	0.0672	67.841	55.308	46.996
$\tau = \frac{3}{4}$	WA-GAN	**0.0208**	**0.0266**	**0.0158**	**59.548**	**44.602**	**31.808**
	HTGAN	0.0284	0.0338	0.055	891.62	612.44	440.45
	GPGAN	0.0497	0.0726	0.0818	59.919	53.758	45.541

Illustration: extremal coefficients

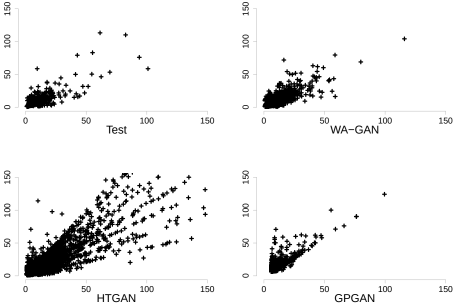


$d = 10, \;\; \theta = 2$

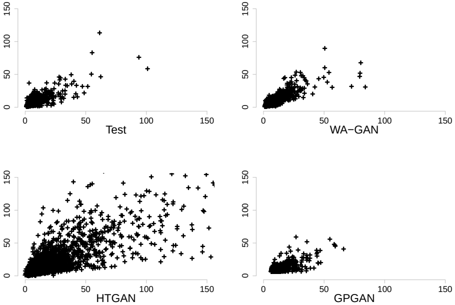


$d = 50, \;\; \theta = 2$

Illustration: generated extremes



$d = 10, \; \theta = 2$



$d = 50, \; \theta = 2$

Data description

- Value-averaged daily returns of $d = 30$ industry portfolios compiled and posted as part of the Kenneth French Data Library [2]

[2] `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

Data description

- Value-averaged daily returns of $d = 30$ industry portfolios compiled and posted as part of the Kenneth French Data Library [2]
- Between 1950 and 2015 ($n = 16\,694$). We take $n_{\text{train}} = 7\,000$, $n_{\text{val}} = 3\,000$, $n_{\text{test}} = 6\,694$ and $k = \sqrt{n_{\text{train}}}$

[2] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

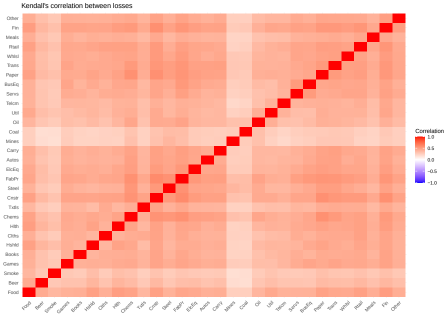
Data description

- Value-averaged daily returns of $d = 30$ industry portfolios compiled and posted as part of the Kenneth French Data Library [2]
- Between 1950 and 2015 ($n = 16\,694$). We take $n_{\text{train}} = 7\,000$, $n_{\text{val}} = 3\,000$, $n_{\text{test}} = 6\,694$ and $k = \sqrt{n_{\text{train}}}$
- Extreme losses $\rightarrow$ data multiplied by -1

[2] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

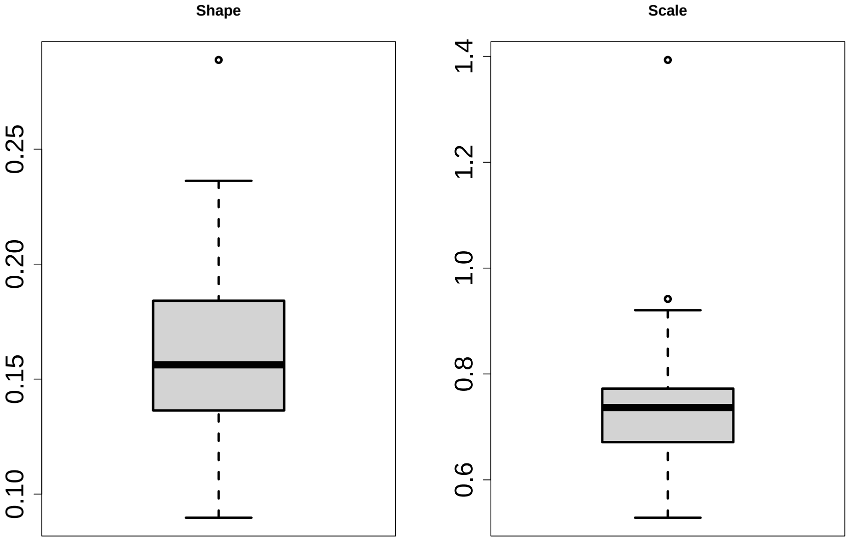
Data description

- Value-averaged daily returns of $d = 30$ industry portfolios compiled and posted as part of the Kenneth French Data Library [2]
- Between 1950 and 2015 ($n = 16\,694$). We take $n_{\text{train}} = 7\,000$, $n_{\text{val}} = 3\,000$, $n_{\text{test}} = 6\,694$ and $k = \sqrt{n_{\text{train}}}$
- Extreme losses $\rightarrow$ data multiplied by -1



[2] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Marginal parameters of the GPD

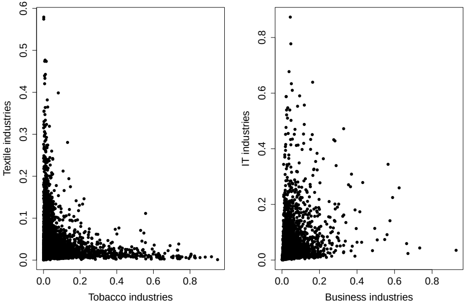


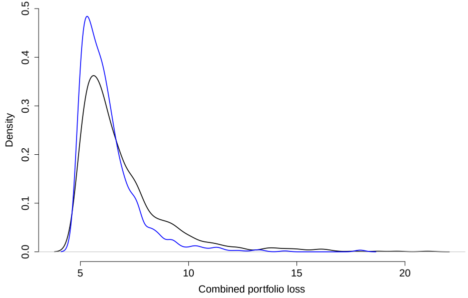
Results

method	dependence	extremes
WA-GAN	**0.018**	11.93
HTGAN	0.026	/
GPGAN	0.043	**8.46**

Results

method	dependence	extremes
WA-GAN	**0.018**	11.93
HTGAN	0.026	/
GPGAN	0.043	**8.46**





- WA-GAN method to sample from angular measures w.r.t. L_1-norm
- Simple post-processing allows for arbitrary norm
- Captures the dependence structure in moderately high dimensions
- May be used to sample from the associated MGPD if the margins are supposed to be in the DoA of univariate GPDs

- WA-GAN method to sample from angular measures w.r.t. L_1-norm
- Simple post-processing allows for arbitrary norm
- Captures the dependence structure in moderately high dimensions
- May be used to sample from the associated MGPD if the margins are supposed to be in the DoA of univariate GPDs
- No uncertainty quantification but can be done via “bootstrap”
- Many other architectures possible for the networks, some of which may be more adapted to particular data types

Thank you!

stephane.lhaut@uclouvain.be
Stéphane Lhaut

References

Girard, S., Gobet, E., and Pachebat, J. (2024).
Deep generative modeling of multivariate dependent extremes.
HAL preprint, hal-04700084v2, pages 1–35.

Lhaut, S., Rootzén, H., and Segers, J. (2025).
Wasserstein-Aitchison GAN for angular measures of multivariate extremes.

Li, J., Li, D., Li, P., and Samorodnitsky, G. (2024).
Generalized Pareto GAN: Generating extremes of distributions.
In *Proceedings of the 2024 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.