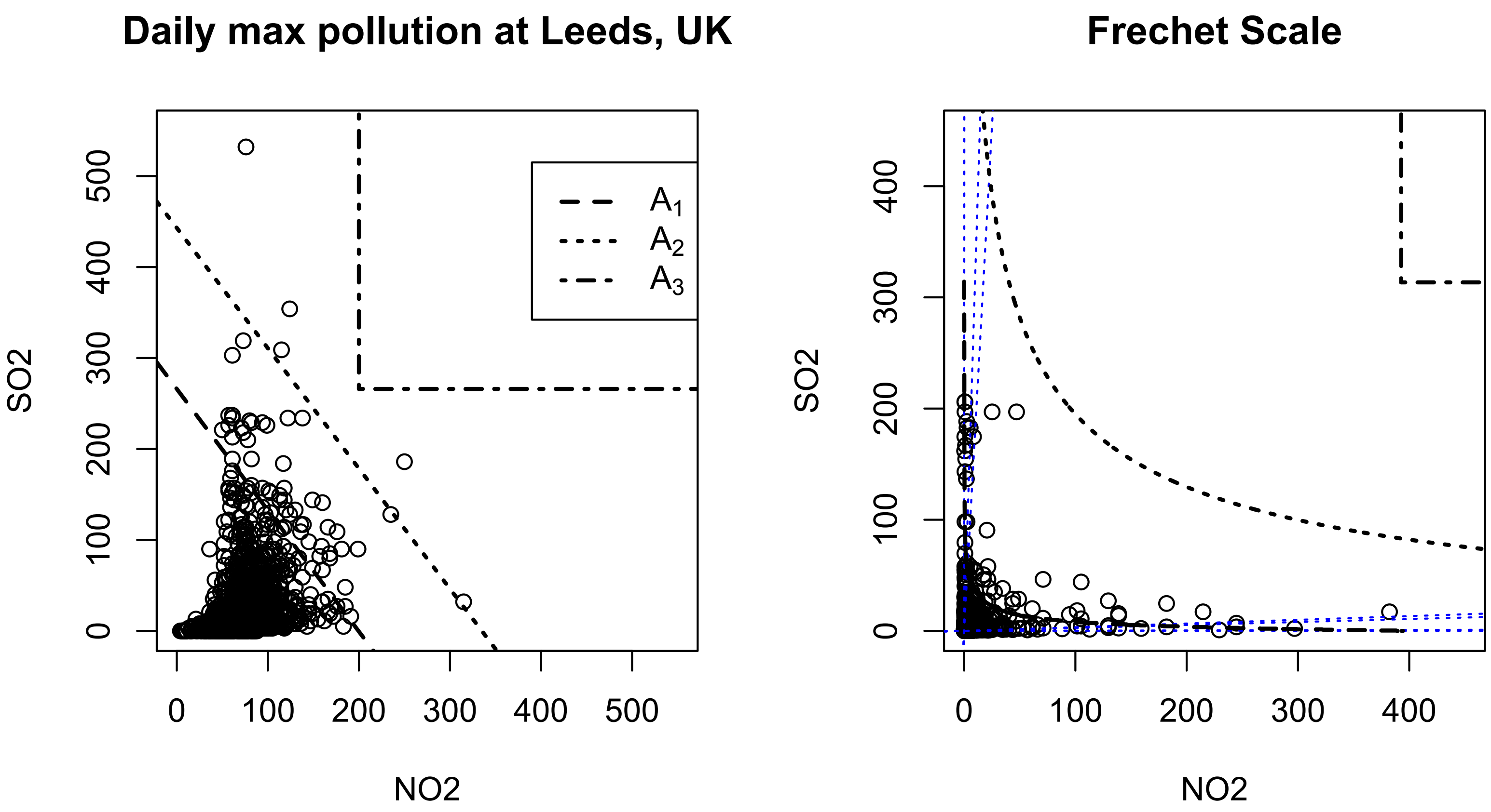
Hidden Regular Variation in Joint Tail Modeling with Likelihood Inference via the MCEM Algorithm

Dan Cooley
Grant Weller
Department of Statistics
Colorado State University

Funding:
NSF-DMS-0905315
Weather and Climate Impacts Assessment Program (NCAR)
2011-2012 SAMSI UQ program

Motivating Example: Daily Air Pollution, Leeds UK



Data exhibit asymptotic independence (Heffernan and Tawn, 2004).

Outline

- Hidden Regular Variation
- Sum Characterization of HRV
- Estimation via MCEM
- Application: air pollution data

When Multivariate Regular Variation Fails

Multivariate Regular Variation:

$$t\mathbb{P}\left[\frac{R}{b(t)} > r, \mathbf{W} \in B\right] \xrightarrow{v} r^{-\alpha} H(B).$$

In some cases, the angular measure H degenerates on some regions of $\mathcal{N}$, masking sub-asymptotic dependence features.

Example: asymptotic independence in $d = 2$:

$$\lim_{z \to z_+} \mathbb{P}(Z_1 > z | Z_2 > z) = 0.$$

- H consists of point masses at $\{0\}$ and $\{1\}$ (using $\|\cdot\|_1$)
- e.g. bivariate Gaussian with correlation $\rho < 1$

Normalization by $b(t)$ kills off sub-asymptotic dependence structure.

Hidden Regular Variation

(Resnick, 2002)

A regular varying random vector $\mathbf{Z}$ exhibits hidden regular variation on a subcone $\mathfrak{C}_0 \subset \mathfrak{C}$ if $\nu(\mathfrak{C}_0) = 0$ and there exists $\{b_0(t)\}$, $b_0(t) \to \infty$ with $b_0(t)/b(t) \to 0$ s.t.

$$t\mathbb{P}\left[\frac{\mathbf{Z}}{b_0(t)} \in \cdot\right] \xrightarrow{v} \nu_0(\cdot)$$

as $t \to \infty$ in $M_+(\mathfrak{C}_0)$.

- Scaling: $\nu_0(tA) = t^{-\alpha_0}\nu_0(A)$ for measurable $A \in \mathfrak{C}_0$, $\alpha_0 \geq \alpha$
- ν_0 is Radon but *not necessarily finite.*

Equivalently,

$$t\mathbb{P}\left[\frac{R}{b_0(t)} > r, \mathbf{W} \in B\right] \xrightarrow{v} r^{-\alpha_0} H_0(B)$$

for B a Borel set of $\mathcal{N}_0 = \mathfrak{C}_0 \cap \mathcal{N}$ (e.g. $\mathcal{N}_0 = (0,1)$).

H_0 is called the *hidden angular measure.*

Example: bivariate Gaussian

Consider $\mathbf{Z}$ with Fréchet margins and Gaussian dependence, $\rho \in [0, 1)$. Recall ν places mass only on the axes of $\mathfrak{C}$.

Define $\eta = (1 + \rho)/2$, the *coefficient of tail dependence* (Ledford and Tawn, 1997).

- $\mathbf{Z}$ exhibits hidden regular variation of order $\alpha_0 = 1/\eta$
- The density of the hidden measure ν_0 can be written

$$\nu_0(dr \times dw) = \frac{1}{\eta} r^{-(1+1/\eta)} dr \times \underbrace{\frac{1}{4\eta}\{w(1-w)\}^{-1/2\eta-1} dw}_{H_0(dw)}$$

H_0 is infinite on $(0, 1)$.

Tail Equivalence

(Maulik and Resnick, 2004)

Two random vectors $\mathbf{X}$ and $\mathbf{Y}$ are *tail equivalent* on the cone $\mathfrak{C}^*$ if

$$t\mathbb{P}\left[\frac{\mathbf{X}}{b^*(t)} \in \cdot\right] \xrightarrow{v} \nu(\cdot) \qquad \text{and} \qquad t\mathbb{P}\left[\frac{\mathbf{Y}}{b^*(t)} \in \cdot\right] \xrightarrow{v} c\nu(\cdot)$$

as $t \to \infty$ in $M_+(\mathfrak{C}^*)$ for $c > 0$.

'Extremes of $\mathbf{X}$ and $\mathbf{Y}$ samples taken in $\mathfrak{C}^*$ will have the same asymptotic properties.'

Mixture Characterization of HRV

(Maulik and Resnick, 2004)

Suppose $\mathbf{Z}$ is regular varying on $\mathfrak{C}$ with hidden regular variation on $\mathfrak{C}_0$:

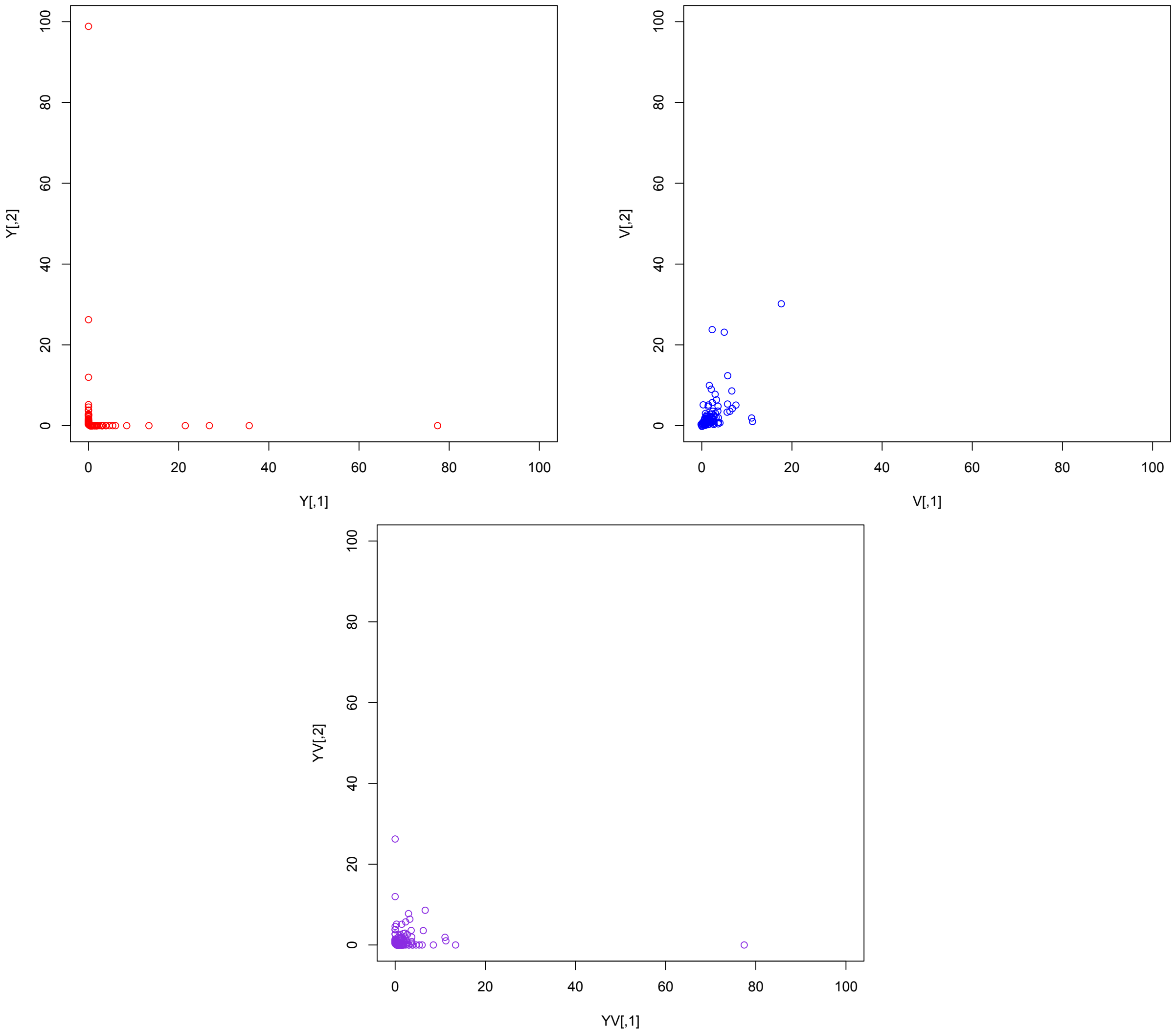
$$t\mathbb{P}\left[\frac{\mathbf{Z}}{b(t)} \in \cdot\right] \xrightarrow{v} \nu(\cdot) \qquad \text{in } M_+(\mathfrak{C}) \qquad \text{and}$$

$$t\mathbb{P}\left[\frac{\mathbf{Z}}{b_0(t)} \in \cdot\right] \xrightarrow{v} \nu_0(\cdot) \qquad \text{in } M_+(\mathfrak{C}_0)$$

with $\nu(\mathfrak{C}_0) = 0$ and $b_0(t)/b(t) \to 0$ as $t \to \infty$.

- Let $\mathbf{Y}$ be $RV(\alpha)$ with support only on $\mathfrak{C} \setminus \mathfrak{C}_0$.
- Let $\mathbf{V} = R_0\theta_0$, $R_0 \sim F_{R_0}(t) = 1/b^{\rightarrow}(t)$ and $\theta_0 \sim H_0$, finite.
- Then $\mathbf{Z}$ is tail equivalent to a mixture of $\mathbf{Y}$ and $\mathbf{V}$ on both $\mathfrak{C}$ and $\mathfrak{C}_0$.

Works because $\mathbf{Y}$'s support doesn't mess with the HRV.



Y[,2]
Y[,1]
V[,2]
V[,1]
YV[,2]
YV[,1]
0
20
40
60
80
100

Construction of $\mathbf{Y} + \mathbf{V}$

Define $\mathbf{Y} = R\mathbf{W}$, with $\mathbb{P}(R > r) \sim 1/b^{\leftarrow}(r)$ and $\mathbf{W}$ drawn from limiting angular measure H. *Notice that* $\mathbf{Y}$ *has support only on* $\mathfrak{C} \setminus \mathfrak{C}_0$.

Let $\mathbf{V} \in [0, \infty)^d$ be regular varying on $\mathfrak{C}_0$ with limit measure ν_0:

$$t\mathbb{P}\left[\frac{\mathbf{V}}{b_0(t)} \in \cdot\right] \xrightarrow{v} \nu_0(\cdot) \quad \text{in } M_+(\mathfrak{C}_0).$$

Further assume that on $\mathfrak{C}$,

$$\mathbb{P}(\|\mathbf{V}\| > r) \sim cr^{-\alpha^*}$$

as $r \to \infty$, with $c > 0$ and

$$\alpha^* > \alpha \vee (\alpha_0 - \alpha).$$

Assume R, $\mathbf{W}$, $\mathbf{V}$ are independent.

Tail Equivalence Result

Then

$$t\mathbb{P}\left[\frac{\mathbf{Y}+\mathbf{V}}{b(t)} \in \cdot\right] \xrightarrow{v} \nu(\cdot) \text{ in } M_+(\mathfrak{C})$$

(Jessen and Mikosch, 2006).

Furthermore, tail equivalence (Maulik and Resnick, 2004) also holds on $\mathfrak{C}_0$:

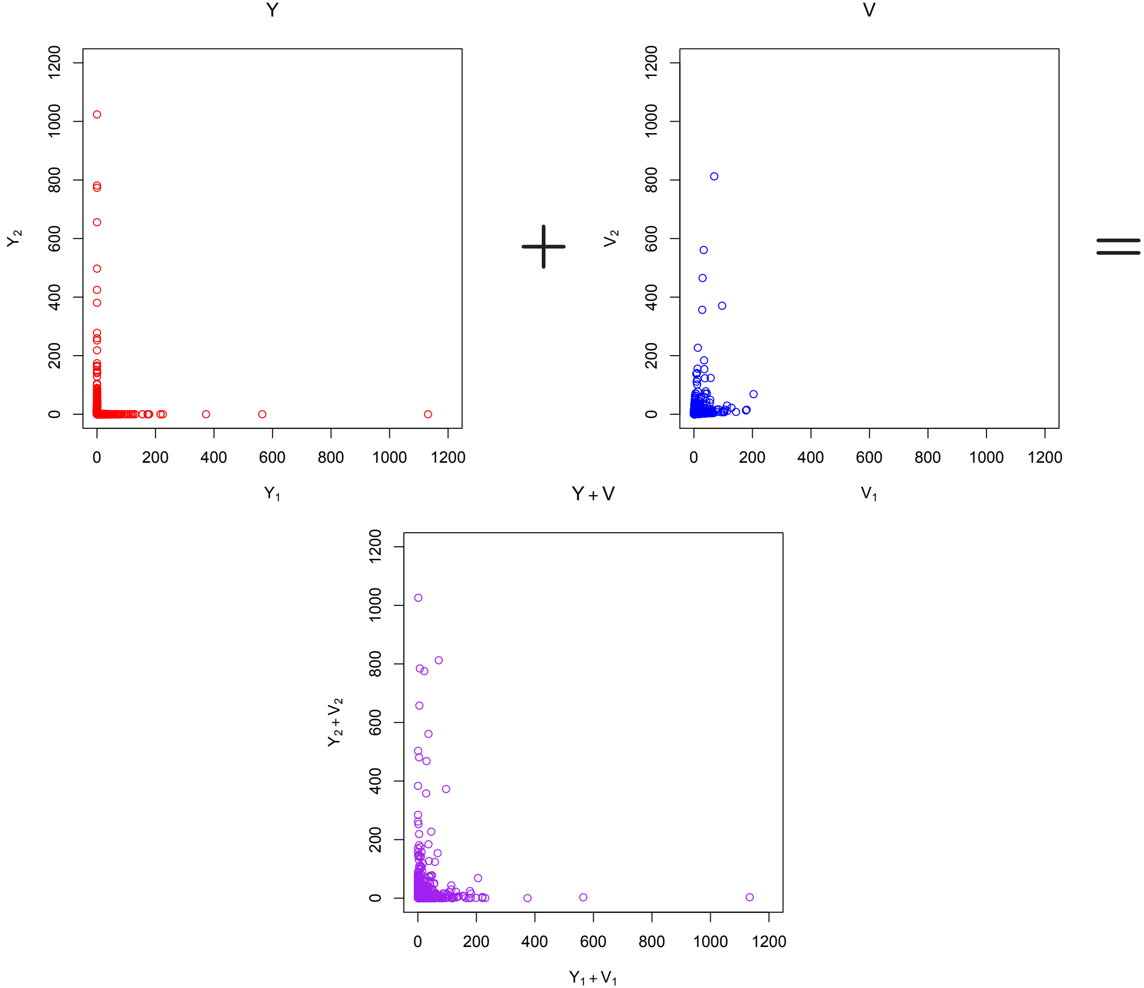
Theorem. *With* $\mathbf{Y}$ *and* $\mathbf{V}$ *as defined above,*

$$t\mathbb{P}\left[\frac{\mathbf{Y}+\mathbf{V}}{b_0(t)} \in \cdot\right] \xrightarrow{v} \nu_0(\cdot) \textit{ in } M_+(\mathfrak{C}_0).$$

View $\mathbf{Z}$ as a sum of 'first-order' $\mathbf{Y}$ and 'second-order' $\mathbf{V}$.

The sum $\mathbf{Y}+\mathbf{V}$ is *tail equivalent* to $\mathbf{Z}$ on *both* $\mathfrak{C}$ *and* $\mathfrak{C}_0$.

Simulation when ν_0 is finite.



No point falls exactly on an axis.

Infinite Measure Example: Bivariate Gaussian

$\mathbf{Z}$ has Fréchet margins and Gaussian dependence ($\rho < 1$). Recall: H_0 is *infinite* on $\mathfrak{N}_0 = (0, 1)$.

Poses difficulty near the axes of $\mathfrak{C}$.

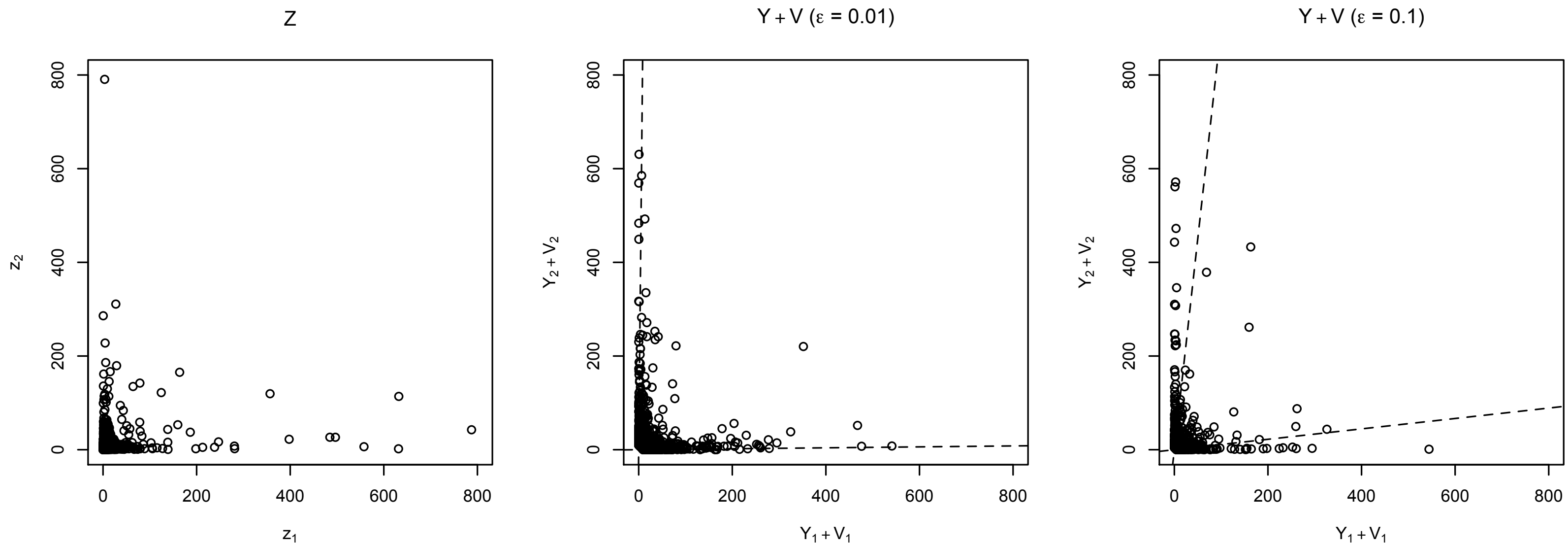
Proposed construction of $\mathbf{V}$:

- Restrict to $\mathfrak{C}_0^\epsilon = \mathfrak{C}_0 \cap \mathfrak{N}_0^\epsilon$, where $\mathfrak{N}_0^\epsilon = [\epsilon, 1-\epsilon]$ for $\epsilon \in (0, 1/2)$.
- Simulate W_0 from probability density $H_0(dw)/H_0(\mathfrak{N}_0^\epsilon)$
- Let R_0 follow a Pareto distribution with $\alpha = 1/\eta$
- $\mathbf{V} = [R_0 W_0, R_0(1 - W_0)]^T$

$\mathbf{Y} + \mathbf{V}$ is tail equivalent to $\mathbf{Z}$ on $\mathfrak{C}$ and $\mathfrak{C}_0^\epsilon$.

Sum representation of bivariate Gaussian

Example with $\rho = 0.5$ ($n = 2500$):



For any set completely contained in $\mathfrak{C}_0^\epsilon$ we achieve the correct limit measure ν_0.

Choice of ϵ involves a trade-off between:

- Size of the subcone on which tail equivalence holds
- Threshold at which $\mathbf{Y} + \mathbf{V}$ is a useful approximation
- Biases due to choice of ϵ calculated.

Inference via the EM Algorithm

Observe realizations from $\mathbf{Z}$, tail equivalent to $\mathbf{Y}+\mathbf{V}$. Assume parametric forms and perform ML inference via EM.

If we assume $\mathbf{Z}=\mathbf{Y}+\mathbf{V}$,

$$\begin{aligned}\log f(\mathbf{z};\boldsymbol{\theta}) &= \int \log f(\mathbf{z},\mathbf{y},\mathbf{v};\boldsymbol{\theta}) f(\mathbf{y},\mathbf{v}|\mathbf{z};\boldsymbol{\theta}^{(k)}) d\mathbf{y} d\mathbf{v} \\ &\quad - \int \log f(\mathbf{y},\mathbf{v}|\mathbf{z};\boldsymbol{\theta}) f(\mathbf{y},\mathbf{v}|\mathbf{z};\boldsymbol{\theta}^{(k)}) d\mathbf{y} d\mathbf{v} \\ &:= Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) - H(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}).\end{aligned}$$

Here: $\mathbf{Z}$ and $\mathbf{Y}+\mathbf{V}$ are only *tail equivalent*; $\boldsymbol{\theta}$ governs tail behavior of $\mathbf{Y}$ and $\mathbf{V}$. Requires a modification of the EM setup.

EM for Extremes

Consider distributions with densities $g_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta})$ and $g_{\mathbf{V}}(\mathbf{v};\boldsymbol{\theta})$ which are tail equivalent to the true distributions; i.e.,

$$g_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}) \cong f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}) \quad \text{for} \quad \|\mathbf{y}\| > r^*_{\mathbf{Y}}$$
$$g_{\mathbf{V}}(\mathbf{v};\boldsymbol{\theta}) \cong f_{\mathbf{V}}(\mathbf{v};\boldsymbol{\theta}) \quad \text{for} \quad \|\mathbf{v}\| > r^*_{\mathbf{V}},$$

Complete likelihood is based on limiting Poisson point processes for $\mathbf{Y}$ and $\mathbf{V}$.

- E step: expectation is taken with respect to $g(\mathbf{y},\mathbf{v}|\mathbf{z};\boldsymbol{\theta})$.
- M step: maximization is taken over only 'large' $\mathbf{y}$ and $\mathbf{v}$.

We show

$$H(\boldsymbol{\theta}^{(k)}|\boldsymbol{\theta}^{(k)}) - H(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) \geq 0$$

using Jensen's inequality.

MCEM

Natural framework for MCEM.

At the E step of the $(k+1)^{th}$ iteration, simulate from

$$g_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}^{(k)})g_{\mathbf{V}}(\mathbf{z}-\mathbf{y};\boldsymbol{\theta}^{(k)}) \propto g(\mathbf{y},\mathbf{v}|\mathbf{z};\boldsymbol{\theta}^{(k)})$$

for all $\mathbf{z}$ and use the simulated realizations to compute

$$\hat{Q}_m(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) = \frac{1}{m}\sum_{j=1}^{m} \ell(\boldsymbol{\theta};\mathbf{z},\mathbf{y}_j,\mathbf{v}_j).$$

employing Poisson point process likelihoods for *large* realizations of $\mathbf{Y}$ and $\mathbf{V}$.

Key idea: likelihood only depends on $\boldsymbol{\theta}$ for 'large' $\mathbf{y}$ and $\mathbf{v}$!

Uncertainty estimates obtained via Louis' method.

Example w/ Infinite Hidden Measure

Simulate $n = 10000$ realizations from a bivariate Gaussian distribution with correlation ρ, transform marginals to unit Fréchet.

Tail equivalent on $\mathfrak{C}$ and $\mathfrak{C}_0^\epsilon$ to $\mathbf{Y} + \mathbf{V}$, where $\mathbf{V}$ has angular measure

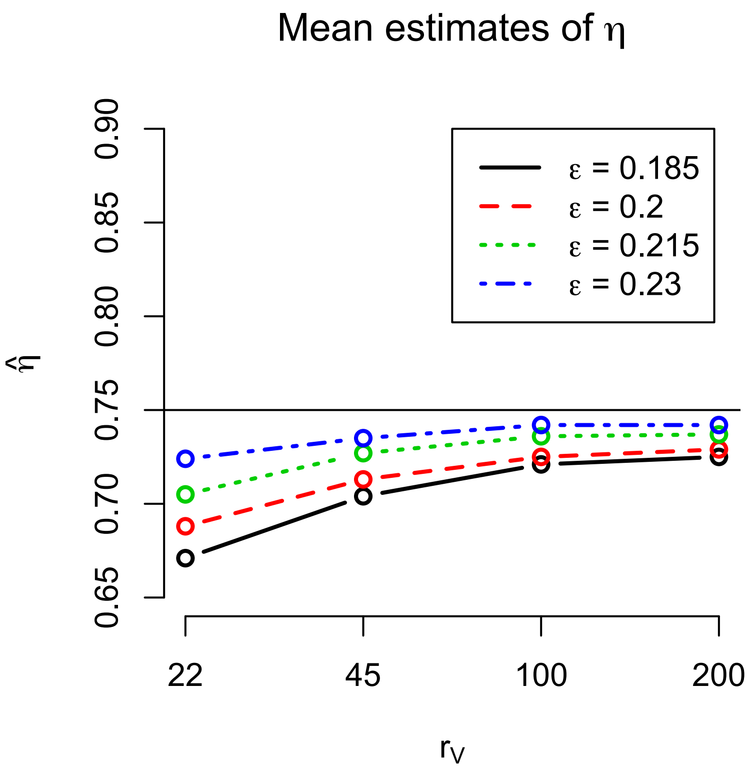
$$H_0(dw) = \frac{1}{4\eta}\{w(1-w)\}^{-1/2\eta-1}dw.$$

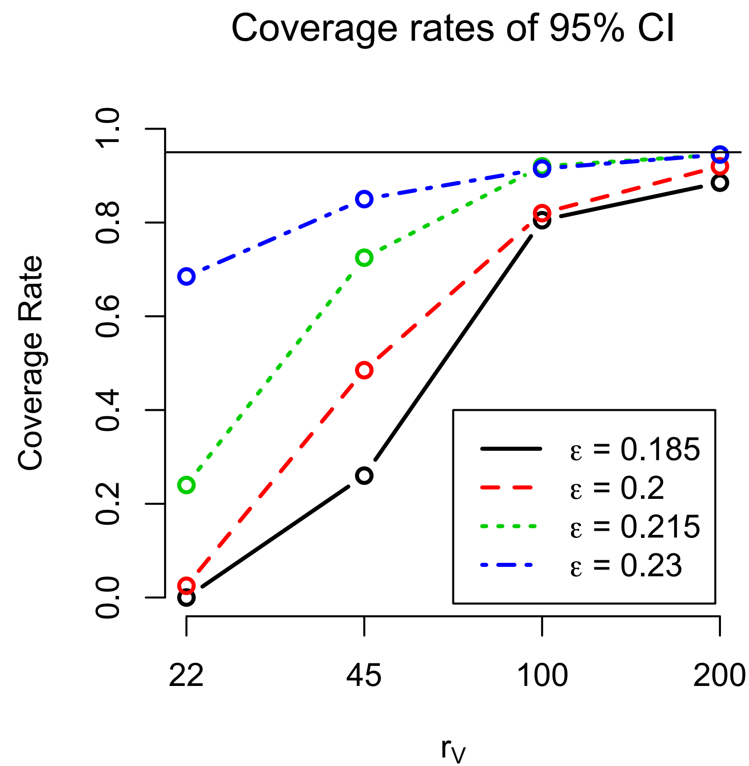
Aim: estimate $\eta = (1+\rho)/2$ from the ϵ-restricted model.

- Must select both ϵ and $r^*_{\mathbf{V}}$
- Trade-off in finite sample estimation problems

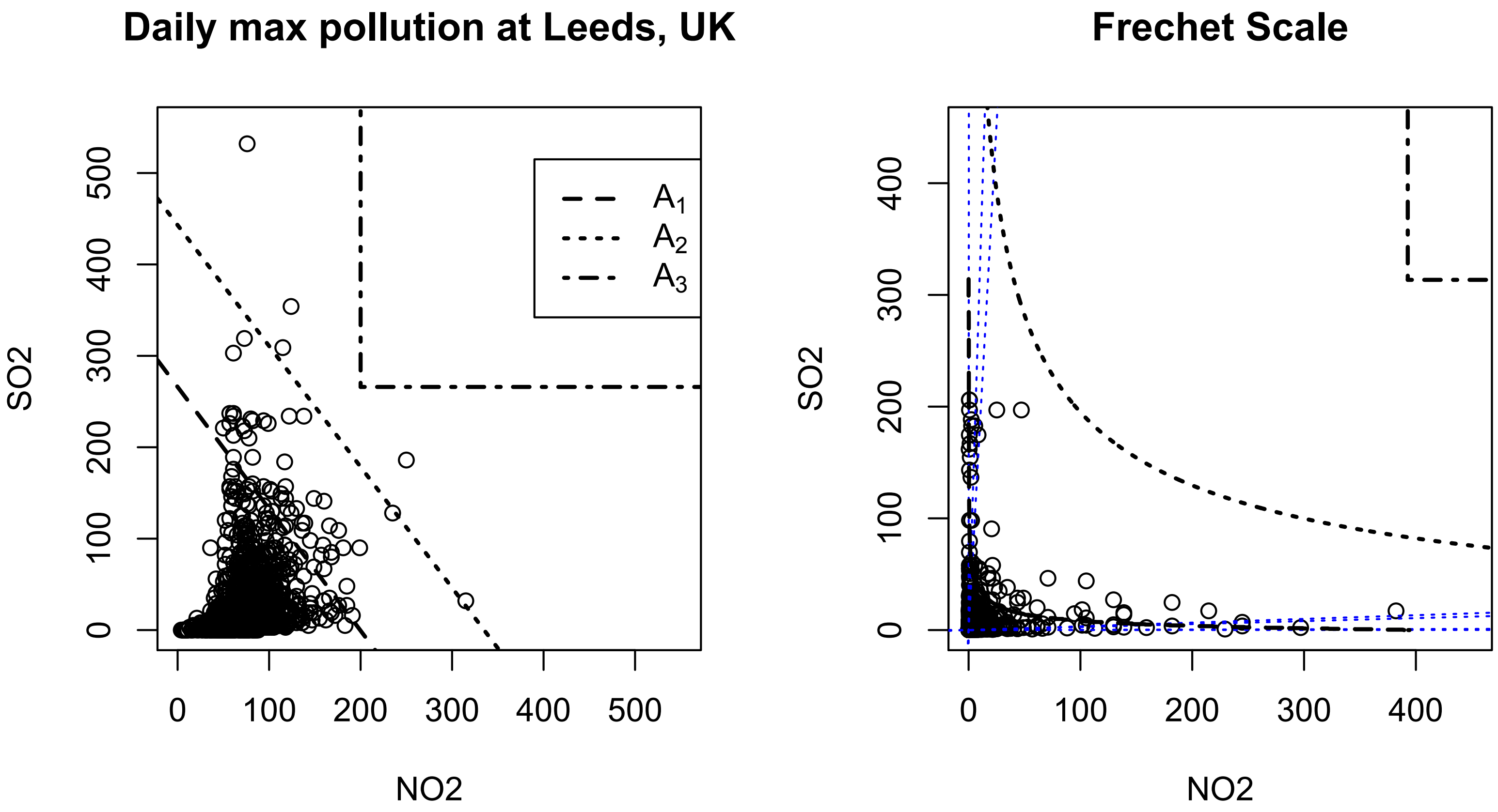
Infinite Hidden Measure Results

Shown for $\eta = 0.75$ ($\rho = 0.5$)





Air Pollution Data



- Strong evidence for asymptotic independence
- Aim: estimate risk set probabilities

Competing Approaches

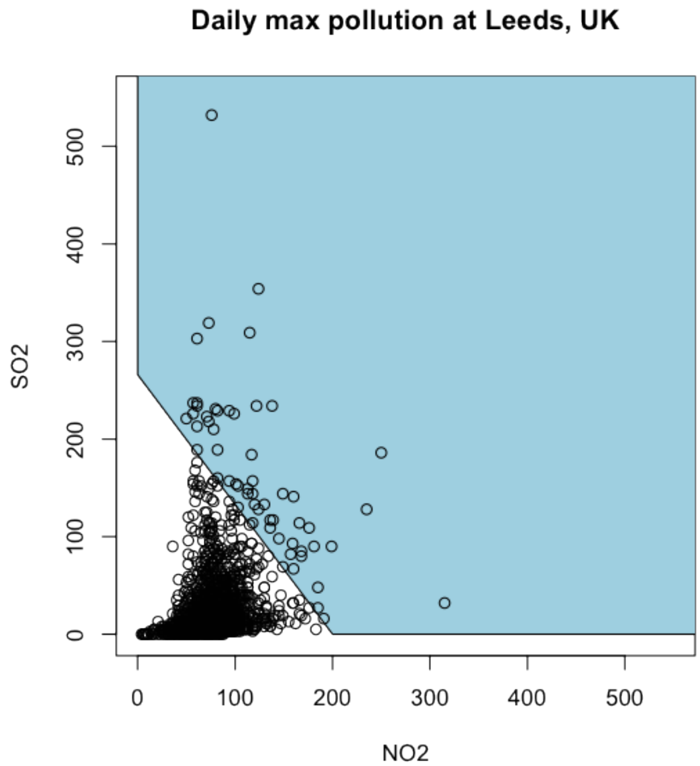
Examine three modeling approaches:

1. Assume asymptotic dependence; i.e. that $\nu(\cdot)$ places mass on the entire cone $\mathfrak{C}$. Fit a bivariate logistic angular dependence model to largest 10% of observations (in terms of L_1 norm). Estimate $\widehat{\beta} = 0.713$.

2. Assume asymptotic independence and ignore any possible hidden regular variation.

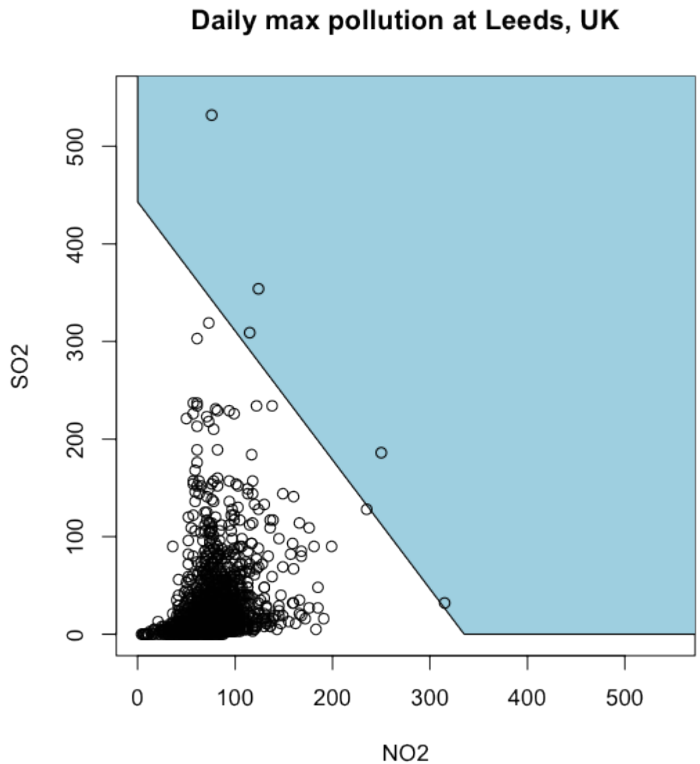
3. Assume asymptotic independence and hidden regular variation. Fit the ϵ-restricted infinite hidden measure model via MCEM. Select $r^*_{\mathbf{V}} = 7.5$ and $\epsilon = 0.3$. Estimate $\widehat{\eta} = 0.748$.

Results - risk set estimates



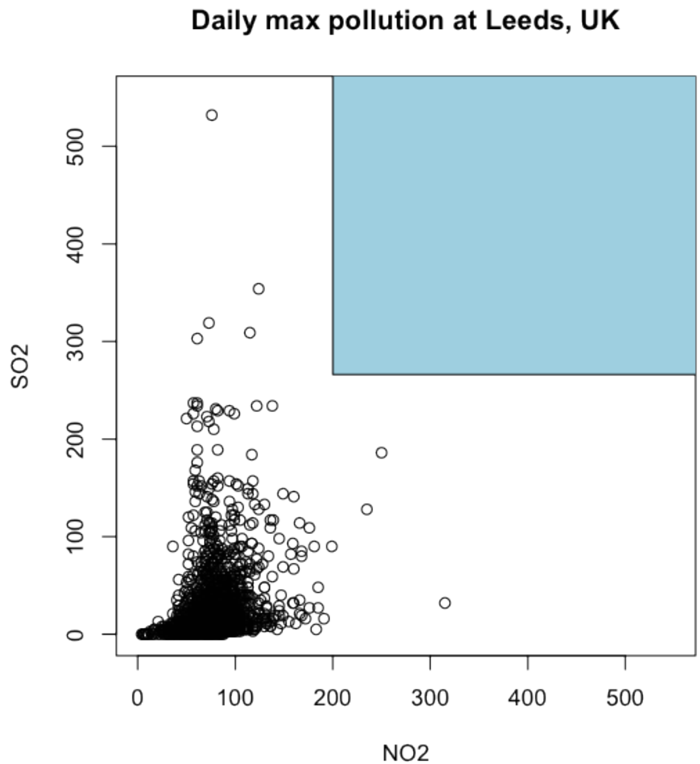
Model	$\widehat{\mathbb{P}}(\mathbf{Z} \in A_1)$	Expected #	p-val
1 (asy. dep.)	0.0297	59.04	0.480
2 (asy. indep.)	0.0120	23.86	8.17×10^{-5}
3 $(\mathbf{Y} + \mathbf{V})$	0.0261	51.89	0.210
Empirical	0.0292	58	–

Results - risk set estimates



Model	$\widehat{\mathbb{P}}(\mathbf{Z} \in A_2)$	Expected #	p-val
1 (asy. dep.)	0.0044	8.74	0.132
2 (asy. indep.)	0.0002	0.40	0.009
3 $(\mathbf{Y} + \mathbf{V})$	0.0018	3.58	0.274
Empirical	0.0025	5	–

Results - risk set estimates



Model	$\widehat{\mathbb{P}}(\mathbf{Z} \in A_3)$	Expected #	p-val
1 (asy. dep.)	0.0010	1.99	0.130
2 (asy. indep.)	0	0	1
3 $(\mathbf{Y} + \mathbf{V})$	0.0002	0.40	0.704
Empirical	0	0	–

Summary

This work introduces a sum representation for regular varying random vectors possessing hidden regular variation.

- Useful representation for finite samples
- Asymptotically justified by tail equivalence result
- Difficulty arises when H_0 is infinite - restrict to a compact cone to simulate $\mathbf{V}$
- Likelihood estimation via modified MCEM algorithm
- Captures tail dependence in the presence of asymptotic independence
- Improved estimation of tail risk set probabilites

References

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

Jessen, A. and Mikosch, T. (2006). *Regularly varying functions*. University of Copenhagen, laboratory of Actuarial Mathematics.

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

Maulik, K. and Resnick, S. (2004). Characterizations and examples of hidden regular variation. *Extremes*, 7(1):31–67.

Resnick, S. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, 5(4):303–336.

Weller, G. and Cooley, D. (2013). A sum decomposition for hidden regular variation in joint tail modeling with likelihood inference via the mcem algorithm. *Submitted*.