Discussion and vote of thanks: *Graphical Models for Extremes*

Jennifer L. Wadsworth

I congratulate the authors on an excellent paper, which is both rich in new ideas and elegant mathematical detail.

The setting of the paper is multivariate regular variation (MRV), a broadly applicable regularity assumption on the extremal dependence structure of a random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$. Assuming standard Pareto margins, MRV implies convergence of normalized "threshold exceedances" to a multivariate Pareto distribution, with support $\mathcal{L}$, as in equation (6). It is further assumed that

(a) the d-dimensional joint density $\lambda(\boldsymbol{y})/\Lambda(\boldsymbol{1}) > 0$, and

(b) the full support is on the interior of $\mathcal{L}$, i.e., no mass lies on regions of the form $\{\boldsymbol{x} \in \mathcal{L} : \min(x_1, \ldots, x_d) = 0\}$.

A loosely-described consequence of these settings is that all variables will tend to take their largest values simultaneously, with no possibility that some groups of variables will tend to be large whilst others are small.

Assumptions (a) and (b) are reasonably common in the literature, and in this case facilitate the vast progress achieved on the notions of conditional independence and graphical structure for extremes. In particular, Proposition 1 and Theorem 1 expose very neatly how ideas from the world of graphical modelling pass through to extreme-value theory via the density of the exponent measure, $\lambda(\boldsymbol{y})$. For decomposable graphs, and more particularly block graphs, this leads to new ideas for high-dimensional model construction, and inference on coherent high-dimensional models via lower-dimensional subgroups. Several interesting results are obtained for the Hüsler–Reiss model, and its parameterization shown to reveal extremal conditional independence properties in a very natural way.

The new properties are explored via classical notions of conditional independence for the multivariate Pareto random vector $\boldsymbol{Y}$ with support restricted to $\mathcal{L}^k$, i.e., $\boldsymbol{Y}^k = \boldsymbol{Y} | Y_k > 1$. This presents an intriguing connection with the last RSS discussion paper on extremes, namely the so-called conditional model introduced by Heffernan and Tawn (2004) and Heffernan and Resnick (2007). In the setting of the paper with $\boldsymbol{X}$ standard Pareto,

$$\boldsymbol{X}/u \mid \|\boldsymbol{X}\|_\infty > u \xrightarrow{d} \boldsymbol{Y}, \qquad \boldsymbol{X}/u \mid X_k > u = \boldsymbol{X}/u \mid \{\|\boldsymbol{X}\|_\infty > u, X_k > u\} \xrightarrow{d} \boldsymbol{Y}^k, \qquad u \to \infty,$$

and

$$\boldsymbol{X}/X_k \mid X_k > u = \boldsymbol{X}/X_k \mid \{\|\boldsymbol{X}\|_\infty > u, X_k > u\} \xrightarrow{d} \boldsymbol{Y}^k/Y^k_k = \boldsymbol{U}^k, \qquad u \to \infty, \tag{1}$$

with $\boldsymbol{U}^k$ the extremal function relative to coordinate k. Working in exponential-tailed margins, Heffernan and Tawn (2004) made the assumption[1] that there exist $\boldsymbol{a}^k : \mathbb{R} \to \mathbb{R}^d$, $\boldsymbol{b}^k : \mathbb{R} \to (0,\infty)^d$ with $a^k_k(\log X_k) = \log X_k$, $b^k_k(\log X_k) = 1$ such that

$$\frac{\log \boldsymbol{X} - \boldsymbol{a}^k(\log X_k)}{\boldsymbol{b}^k(\log X_k)} \;\Big|\; \log X_k > u \xrightarrow{d} \boldsymbol{Z}^k, \tag{2}$$

where $\boldsymbol{Z}^k_{\setminus k}$ is non degenerate with no mass at $+\infty$, and $\log X_k - u | \log X_k > u \xrightarrow{d} E \sim \mathrm{Exp}(1)$ is independent of $\boldsymbol{Z}^k_{\setminus k}$. Convergence (1) is recovered from (2) with $\boldsymbol{a}^k(\log X_k) = \log X_k \boldsymbol{1}$, $\boldsymbol{b}^k(\log X_k) = \boldsymbol{1}$, and $\boldsymbol{Z}^k = \log \boldsymbol{U}^k$.

Under assumptions (a) and (b), the exponent measure density $\lambda(\boldsymbol{y})$ provides the "glue" linking the extremal functions together: $\lambda(\boldsymbol{y})$ yields the distribution of each $\boldsymbol{U}^k$ via equation (41). As a key example, the extremal functions for the Hüsler–Reiss model are log-Gaussian, and the paper shows how the conditional independence patterns in these log-Gaussians yield the overall graphical structure of $\boldsymbol{Y}$.

The conditional model viewpoint seems to provide a possibility for alternative methodology to that outlined in the paper, as well as a first suggestion of how one might approach extending these ideas to cases where (a) and/or (b)

[1] The formulation with random normalization came later with Heffernan and Resnick (2007), but is equivalent to the Heffernan and Tawn (2004) case under the existence of densities.

may not hold. The latter becomes more likely as d grows, yielding a natural tension between the beautiful theory of this work and the messy reality of data.

Firstly, assuming a Hüsler–Reiss model, $\log \boldsymbol{X}_{\setminus k} - \log X_k | \log X_k > u \overset{d}{\approx} \boldsymbol{Z}^k_{\setminus k} \sim N(-\text{diag}(\Sigma^{(k)})/2), \Sigma^{(k)})$ for a high threshold u (Engelke et al., 2015). Applying a graphical lasso technique to $\Theta^{(k)} = (\Sigma^{(k)})^{-1}$ may give a sparse precision matrix, which in principle leads to Γ and the complete Hüsler–Reiss parameterization. A practical hurdle is the likelihood of inferring different graphs from different k, and that connections to the kth node are encoded in row/column sums of $\Theta^{(k)}$, which are not shrunk towards zero in a standard implementation. However, this could provide a starting point to explore beyond trees and block graphs, should these appear inadequate.

The potential of the conditional formulation is particularly apparent in the case where mass of the multivariate Pareto distribution lies on regions of the form $\{\boldsymbol{x} \in \mathcal{L} : \min(x_1, \ldots, x_d) = 0\}$. In this case, the normalizations in (2) allow differing strengths of extremal dependence between the components of $\boldsymbol{X}$ and X_k, such that components of $\boldsymbol{Z}^k_{\setminus k}$ may not have mass at $-\infty$ where those of $\log \boldsymbol{U}^k_{\setminus k}$ do. As such, the representation provides more detail about the extremal dependence, increasing its utility for statistical modelling. To exploit assumption (2) in practice, one poses parametric forms for $\boldsymbol{a}^k, \boldsymbol{b}^k$ and the distribution of $\boldsymbol{Z}^k$. Suppose that we take $\boldsymbol{a}^k(\log X_k) = \boldsymbol{\alpha}^k \log X_k$, $\boldsymbol{\alpha}^k_{\setminus k} \in [0,1]^{d-1}, \alpha^k_k = 1$, $\boldsymbol{b}^k(\log X_k) = \mathbf{1}$ and, similarly to the Hüsler–Reiss model, assume $\boldsymbol{Z}^k_{\setminus k} \sim N(\boldsymbol{\mu}^k, \Sigma^{(k)})$. Then, above a high threshold u,

$$\log \boldsymbol{X} | \log X_k > u \overset{d}{\approx} \boldsymbol{\alpha}^k(E + u) + \boldsymbol{Z}^k, \qquad E \sim \text{Exp}(1), \qquad E \perp\!\!\!\perp \boldsymbol{Z}^k; \tag{3}$$

the Hüsler–Reiss model is a special case with $\boldsymbol{\alpha}^k = \mathbf{1}$ and $\boldsymbol{\mu}^k = -\text{diag}(\Sigma^{(k)})/2$ for all k. For illustration, model (3) was fitted to the Danube river data both with $\boldsymbol{\alpha}^k = \mathbf{1}$ fixed and estimated. The threshold u was taken as the 0.85 marginal quantile; higher thresholds produced some errors in sparse precision matrix estimation. In each case the components of $\boldsymbol{Z}^k_{\setminus k}$ were estimated individually and a graphical lasso applied to $\Theta^{(k)}$ using `EBICglasso` in the R library `qgraph` (Epskamp et al., 2012). To give an impression of results across all k, Figure 1 displays connections selected at least half of the time. Notably, although most estimates $\widehat{\alpha}^k_j < 1$, the set of connections is fairly similar. As a diagnostic, Figure 2 displays $\chi_C(q) = \Pr(F_i(X_i) > q \ \forall i \in C)/(1-q)$ for three sets with $|C| = 2$, and $C = \{1, \ldots, 31\}$. For a multivariate Pareto distribution $\chi_C(q) \equiv \chi_C$ for q sufficiently large (Rootzén et al., 2018), and the bivariate estimates from the fitted model in the paper are displayed. For model (3) with $k \in C$ and $\min_{j \in C}\{\alpha^k_j\} < 1$, $\chi_C(q) \searrow 0$ as $q \to 1$. This is often realistic for environmental datasets, though the Danube data display a high degree of extremal dependence.

Certain conditional independences could be established for model (3), but an interpretation along the lines of Definition 1 in the paper is desirable. This seems to require a device like Proposition 1, where the fact that $\lambda(\boldsymbol{y})$ does not depend on k is crucial, and it remains to be seen whether useful and coherent notions of graphical structure can be established in this case. The ideas presented in the paper nonetheless form great inspiration for consideration of structured estimation and interpretation in the case of weaker extremal dependence.

I am very pleased to propose the vote of thanks for this thought-provoking work.

References

Engelke, S., Malinowski, A., Kabluchko, Z., and Schlather, M. (2015). Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):239–265.

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., and Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, 48(4):1–18.

Heffernan, J. E. and Resnick, S. I. (2007). Limit laws for random vectors with an extreme component. *The Annals of Applied Probability*, 17(2):537–571.

Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):497–546.

Rootzén, H., Segers, J., and Wadsworth, J. L. (2018). Multivariate generalized Pareto distributions: Parametrizations, representations, and properties. *Journal of Multivariate Analysis*, 165:117–131.

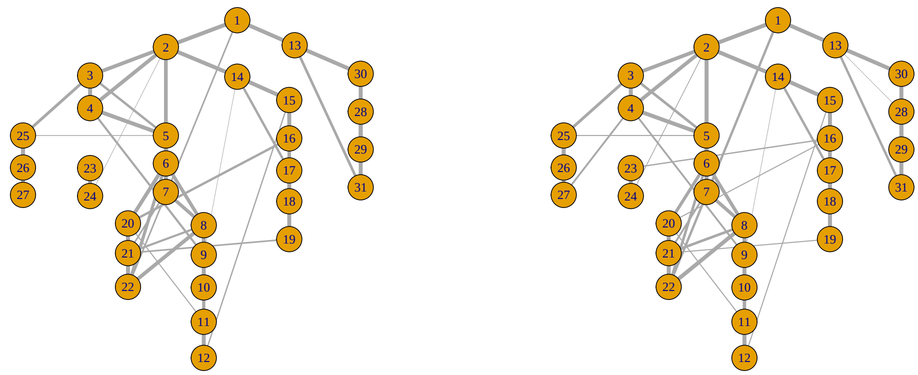

Figure 1: Left: connections estimated under model (3) with $\boldsymbol{\alpha}^k = \mathbf{1}$; right: connections estimated under model (3) with $\boldsymbol{\alpha}^k$ estimated. The line thickness is proportional to the number of times connections were included in the graph, with only those selected at least half of the time displayed. Some connections may not be visible due to the graph layout.

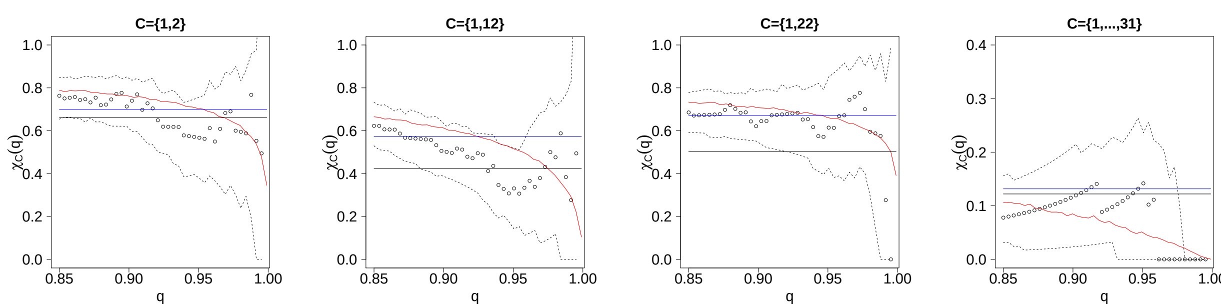

Figure 2: Estimates of $\chi_C(q)$ for $C = \{1,2\}, \{1,12\}, \{1,22\}$ and $\{1,\ldots,31\}$. Black dots: empirical estimates; red line: estimate from model (3) with $\boldsymbol{\alpha}^k$ estimated; blue line: estimate from model (3) with $\boldsymbol{\alpha}^k = \mathbf{1}$ fixed; black line: estimate from the fitted model in the paper. The red and blue lines were produced conditioning on $k = 1$. Dashed lines are approximate 95% confidence intervals for empirical estimates obtained using the nonparametric bootstrap.