BIVARIATE THRESHOLD MODELS AND THE WEIGHTED BOOTSTRAP

Pál Rakonczai, László Varga, András Zempléni Eötvös Loránd University Budapest

BGPD MODELS

Bivariate (multivariate) threshold models can be defined in two different ways.
If we claim exceedance in all of the coordinates (BGPD I), we usually get simpler models, with nice properties (marginals are univariate GPD etc.), but we may use less data.
If we use all data that exceed the threshold $\mathbf{u}$ in at least one coordinate, we get the BGPD II model ([4]).
This approach can be formulated as follows. Let $\mathbf{Y} = (Y_1, Y_2)$ denote a random vector, $\mathbf{u} = (u_1, u_2)$ be a suitably high threshold vector and $\mathbf{X} = \mathbf{Y} - \mathbf{u} = (Y_1 - u_1, Y_2 - u_2)$ be the vector of exceedances. Then the bivariate generalized Pareto distribution (BGPD) for the exceedances $\mathbf{X}$ can be defined by a bivariate extreme value distribution G with non-degenerate margins as

$$H(\mathbf{x}) = \frac{1}{\log G(0,0)} \log \frac{G(x_1, x_2)}{G(x_1 \wedge 0, x_2 \wedge 0)},$$

where $0 < G(0,0) < 1$. This definition provides a model for observations that are extreme in at least one component.
Remarks

1. Note that if $H_1(x) = H(x, \infty)$ is a margin, then H_1 is not a one dimensional GPD, only the conditional distribution of $X_1 | X_1 > 0$ is GPD.
2. All margins are dependent on all parameters, as the constant $1/\log G(0,0)$ factor remains in the formula even when the other component of X converges to infinity, so the parameters cannot be interpreted individually.
3. Some models put weight to the boundaries, so they will not remain absolutely continuous.
4. We can use more data, which hopefully helps in model fitting.

Simulation is possible by approximating the density over a suitable grid and using acceptance sampling. Parametric models can be introduced: let G be a bivariate d.f. with marginals G_i. Then

$$-\log G(\mathbf{x}) = l\{-\log G_1(x_1), -\log G_2(x_2)\}, \qquad \mathbf{x} \in \mathbb{R}^2$$

The dependence function $A(t) = l(1-t, t), \quad t \in [0,1]$ satisfies the following three properties: $A(t)$ is convex,

$$\max\{(1-t), t\} \le A(t) \le t$$

and $A(0) = A(1) = 1$.
The two most popular models:

- Logistic model: $l(v_1, v_2) = (v_1^{1/\alpha} + v_2^{1/\alpha})^\alpha, \quad v_j \ge 0$, with parameter $0 < \alpha \le 1$. Independence: $\alpha = 1$, complete dependence: $\alpha \downarrow 0$. Symmetric, absolutely continuous
- Negative logistic model: $l(v_1, v_2) = v_1 + v_2 - \{(\psi_1 v_1)^{1/\alpha} + (\psi_2 v_2)^{1/\alpha}\}^\alpha$, where $-\infty \le \alpha \le 0$ ï£¡s $0 \le \psi_1, \psi_2 \le 1$.

The BGPD type II model is not absolutely continuous in the asymmetric case.

NEW ASYMMETRIC MODEL

Let A be an arbitrary dependence function, $\Psi(t) : [0,1] \to [0,1]$ be strictly monotonic, $\Psi(0) = 0$ and $\Psi(1) = 1$.
$A_\Psi(t) = A(\Psi(t))$ defines a dependence function iff

1. $A_\Psi(0) = A_\Psi(1) = 0$,
2. $(1-t) \vee t \le A_\Psi(t) \le 1$, if $0 \le t \le 1$,
3. $A'_\Psi(0) = -1$, $A'_\Psi(1) = 1$, $A''_\Psi \ge 0$.

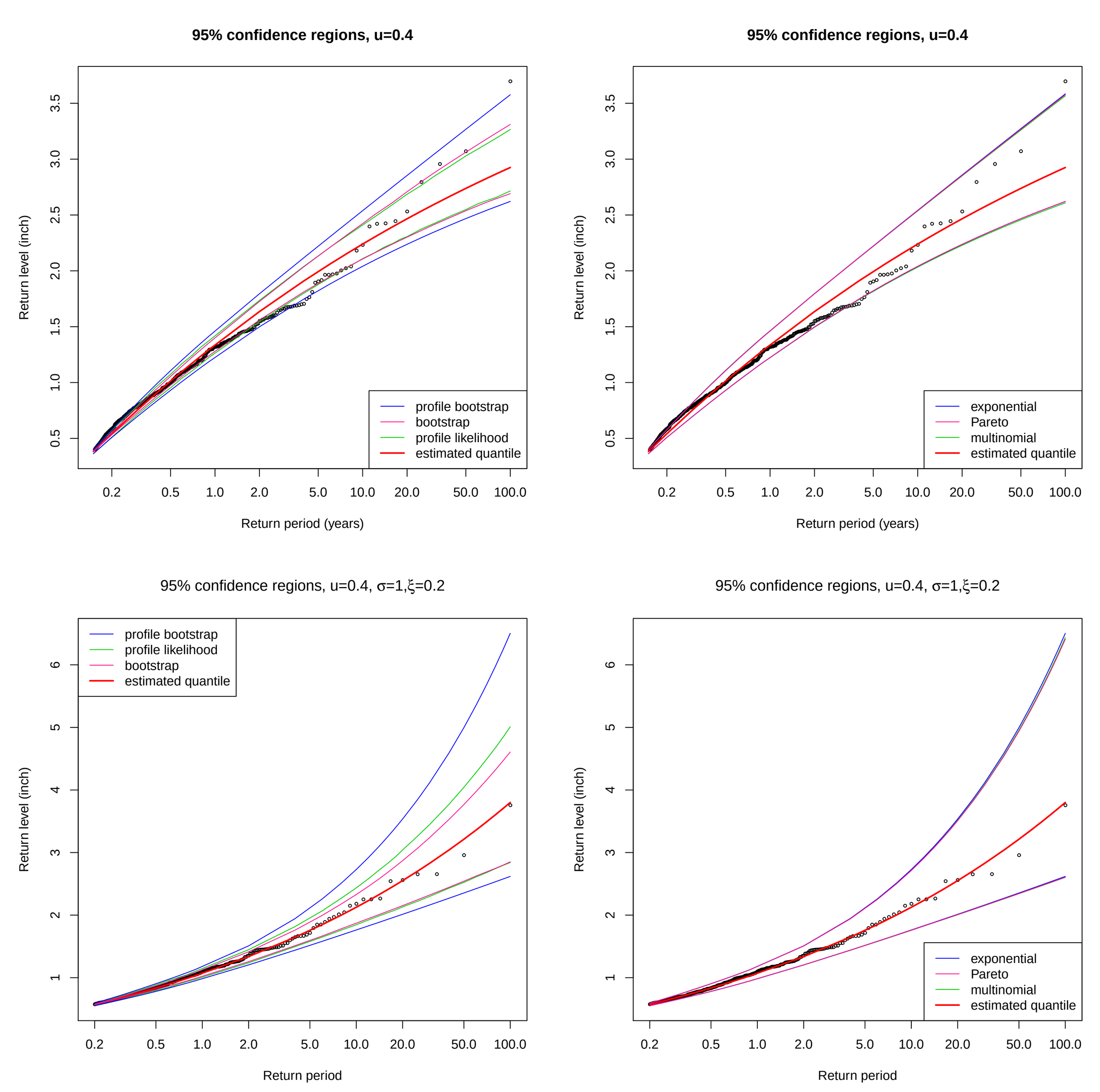
Let $\Psi(t) = t + f(t)$, e.g. $f_{\psi_1, \psi_2}(t) = \psi_1 [t(1-t)]^{\psi_2}$, where $\psi_1 \in \mathbb{R}$ and $\psi_2 \ge 2$ are the asymmetry parameters.

- The permitted parameter space can be determined, it is suitably rich.
- $\psi_1 = 0$ gives back the original model.
- The models are absolutely continuous.
- Although the transformations Ψ was defined now for the bivariate case only, it can be extended for the higher dimensional cases as well.

Other forms of f are also possible, like a two-parameter family of polynomials over$[0,1]$ with a root in $0 < p < 1$.

APPLICATIONS

We have used 100 years of daily precipitation data from two Hungarian cities: Budapest and Debrecen.



Estimated return level with bootstrap confidence bounds

The Ψ-logistic model follows the odd shape of the data extremely well.

ACKNOWLEDGEMENT

The European Union and the European Social Fund have provided financial support to the project under the grant agreement no. TÁMOP 4.2.1/B-09/KMR-2010-0003.

WEIGHTED BOOTSTRAP

This method is also called multiplier bootstrap and was investigated as early as in the 1990s. Newton and Raftery [1] used it in a Bayesian framework.
The bootstrap sample in this framework is denoted by $(\tau_{n,i} X_i)$. The so-called bootstrap weights $\tau_{n,i}$ $(1 \le i \le n,\ n \ge 1)$ are identically distributed and supposed to be independent from the original data. In case of dependent data, the weights are applied to the loglikelihood function only, thus keeping the supposed structure. We have used different distributions for τ: besides the traditional polynomial case, exponential and Pareto distributions were also considered (for its application to ARCH models, see [5]).
There are different approaches for bootstrapping in the extreme value models, parametric bootstrap being the most commonly used method. However, the new proposed method is much more flexible than the original one, and as in our bivariate investigations the model is not unique, the extremes are not especially heavy tailed and there is a relatively long data set available (Section **??**), so this nonparametric approach is more appealing.
The proposed bootstrap approach may be used to goodness-of-fit or for constructing confidence bounds, see Section **??**.
We combined the weighted bootstrap with the profile likelihood method for constructing confidence regions for the return levels of the precipitation data on hand - the elements of the log-likelihood function were multiplied by the appropriate bootstrap weights.

PLANS FOR THE FUTURE

- To clarify the theoretical properties of the weighted bootstrap estimators
- Comparison of the models,

REFERENCES

[1] M. Newton and A. Raftery, 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society*, **56**: p. 3-48.

[2] P. Rakonczai and N. Tajvidi, 2010. On prediction of bivariate extremes, *International Journal of Intelligent Technologies and Applied Statistics* **Vol.3**, No.2 p.115-139.

[3] P. Rakonczai and A. Zempléni, 2012. Bivariate generalized Pareto distribution in practice: models and estimation.*Environmetrics*, **23**, p. 219-227.

[4] H. Rootzén and N. Tajvidi, 2006. The multivariate generalized Pareto distribution. *Bernoulli* **12**: 917-930.

[5] L. Varga and A. Zempléni, 2013. Weighted bootstrap in GARCH models. Submitted.