Contributions to extreme value theory : Trend detection for heteroscedastic extremes

Aline Mefleh

Université Libanaise

UBFC
UNIVERSITÉ
BOURGOGNE FRANCHE-COMTÉ

Thèse en cotutelle

pour obtenir le grade de Docteur délivré par

L'Université de Bourgogne Franche-Comté
(École doctorale Carnot Pasteur)
et
L'Université Libanaise
(École doctorale des Sciences et de Technologie)

Spécialité: Mathématiques Appliquées

Présentée et soutenue publiquement par

Aline MEFLEH

le 26 Juin 2018

Contributions à la théorie des valeurs extrêmes. Détection de tendance pour les extrêmes hétéroscédastiques.

Jury	Mme. Anne-Laure Fougères	Rapporteur
	M. Joseph Ngatchou-Wandji	Rapporteur
	Mme. Camélia Goga	Examinatrice (Présidente)
	M. Nasser Hoteit	Examinateur
	M. Clément Dombry	Directeur de thèse
	M. Hassan Zeineddine	Directeur de thèse
	M. Romain Biard	Co-directeur de thèse
	M. Zaher Khraibani	Co-directeur de thèse

I still remember the days I prayed for the things I have now...

Acknowledgement

First, I am deeply grateful to my advisors and co-advisors Clément Dombry, Hassan Zeineddine, Romain Biard and Zaher Khraibani for believing in me and giving me the chance to accomplish my purpose. Thank you Clément and Romain for welcoming me and supporting me at both the human and professional levels. Thank you for the time you gave me, for the knowledge you transferred to me and for making me evolve during the past three years.
Thank you Hassan and Zaher for your time, your comments and recommendations that help me improve my work.

Also, I would like to thank the Lebanese National Council for Scientific Research CNRS-L and the Lebanese University LU for giving me the opportunity to get my doctorate.

I express my gratitude to Anne-Laure Fougères, Joseph Ngatchou-Wandji, Camélia Goga and Nasser Hoteit for accepting to be members of the jury. In particular, I would like to thank the reviewers for all their comments.

Thank you for all the members of the LMB: the library, the academic, the administrative and the technical staff, the doctoral students and especially my colleagues in the office.
I would also like to thank the doctoral school of sciences and technology at the Lebanese University with its administration.

I owe my deepest gratitude to my family. Without their love, support and encouragement I couldn't have reached this moment.

Last but not least, I would like to thank God, the reason behind every success.

List of publications and communications

Submitted articles

[1] Permutation Bootstrap and the block maxima method. Submitted to *Communications in Statistics - Simulation and Computation*.

[2] Trend detection for heteroscedastic extremes. Submitted to *Scandinavian Journal of Statistics*.

Communications

May 11, 2018 Trend detection for heteroscedastic extremes,
LSMS 2018, Lebanese American University.

May 10, 2018 Trend detection for heteroscedastic extremes,
Doctoral forum at EDST, Lebanese University.

January 15, 2018 Trend detection for heteroscedastic extremes,
Institut Camille Jordan, Université Claude Bernard Lyon 1.

November 22, 2017 Trend detection for heteroscedastic extremes,
Forum des jeunes chercheurs, Nancy.

March 13, 2017 Trend detection for heteroscedastic extremes,
Laboratoire de Mathématiques de Besançon.

May 18, 2016 Permutation Bootstrap and the block maxima method (Poster),
Doctoral forum at EDST, Lebanese University.

Table of contents

List of Figures

List of Tables

Introduction générale

Contexte des valeurs extrêmes

La théorie des valeurs extrêmes (TVE) a été d'abord introduite durant les années 1920 par Fisher and Tippett (1928). Elle a été développée plus tard dans le livre *Statistics of Extremes* de Gumbel (1958) connu aussi pour sa citation: *Il est impossible que l'improbable n'arrive jamais.* Ce n'est qu'à partir de l'inondation catastrophique qui a frappé les Pays-Bas en 1953 et qui a causé la mort de plus de 1800 personnes que la TVE a connu son vrai développement. En effet, le gouvernement était confronté à la question: Quelle doit être la hauteur des digues construites de manière à avoir un compromis entre le coût de construction et la protection contre les inondations. La TVE était un atoût statistique puissant pour répondre à cette question (de Haan and Ferreira (2006)).
Comme son nom l'indique, la TVE s'intéresse donc aux événements extrêmes ou rares qui ont une faible probabilité d'occurrence mais qui une fois réalisés peuvent avoir des conséquences désastreuses. La théorie est largement utilisée dans plusieurs domaines d'application notamment en finance pour l'estimation de la Value at Risk (VaR) (Singh et al. (2013), Klüppelberg and Zhang (2015)), en hydrologie pour la protection contre les inondations (Buishand et al. (2008), Butler et al. (2007)), en météorologie pour la prédiction des vitesses de vent extrêmes (Walshaw (2000), Coles and Pericchi (2003)) etc. Un des problèmes rencontrés en TVE est que les évènements extrêmes sont rares et la théorie doit donc établir des prédictions sur des périodes beaucoup plus longues que celle des données disponibles. En d'autres termes, on pourra s'intéresser à l'estimation des niveaux de retour pour 100 ou 500 ans partant d'un jeu de données sur 50 ans par exemple. Cela nécessite donc une extrapolation des périodes observées à des périodes non observées. La TVE assure cette extrapolation en imposant des hypothèses de régularité de la queue de distribution des observations.

Les premiers travaux en théorie de valeurs extrêmes traitent le cas d'une seule variable aléatoire dont les observations sont supposées indépendamment et identiquement distribuées (i.i.d.). La première généralisation au cas non i.i.d. a été réalisée par Leadbetter et al. (1983) dans leur livre où ils considèrent le cas sta-

tionnaire ou le cas de dépendance à court terme. Des études supplémentaires sur la stationnarité ont été aussi faites par Hsing (1991). Le cas hétéroscédastique, i.e. le cas où les observations sont indépendantes mais non identiquement distribuées, a été lui aussi traité notamment par Davison and Smith (1990), Coles (2001) et récemment par de Haan et al. (2015) et Einmahl et al. (2016). Quant au cadre multivarié, il consiste à étudier des processus extrêmes simultanément afin de tester par exemple la dépendance entre plusieurs variables aléatoires. La TVE multivariées est plus compliquée à traiter et une première difficulté réside dans le fait de définir qu'est-ce qu'un évènement extrême multivarié. Parmi les travaux faits sur le cadre multivarié citons par exemple ceux de de Haan and Resnick (1977), Pickands (1975) et Buishand (1984).

Contributions personnelles

La thèse est composée de cinq chapitres. Le premier chapitre est une introduction générale à la TVE univariée. Les chapitres deux et trois constituent mes contributions personnelles qui ont été soumises dans des journaux scientifiques. Le chapitre quatre est un Data challenge pour l'estimation de quantiles extrêmes dont les résultats sont présentés durant la conférence EVA 2017 au Delft. Le chapitre cinq fait l'objet d'un travail d'application sur de vraies données au Liban. Nous présenterons brièvement dans ce qui suit les chapitres deux à cinq.

Chapitre 2: Bootstrap par permutation et méthode des maxima par blocs

Dans le second chapitre, nous proposons la méthode de Bootstrap par permutation pour améliorer l'estimation des paramètres de la loi des extrêmes généralisés (GEV) et par conséquent des quantiles extrêmes. Nous appliquons cette méthode aux blocs maxima (BM) dans un cadre particulier afin de réduire la variance des estimations. Si on dispose de n observations i.i.d. $X_1, \ldots, X_n$ divisées en k blocs de taille m ($n = k \times m$), nous allons extraire les BM

$$M_j = \max_{(j-1)m \leq i \leq jm} X_i, \;\; j = 1, \cdots, k.$$

Dans le cas où $n = k \times m + r$, les r dernières observations ne seront pas incluses dans les BM. Il est vrai qu'un changement dans l'ordre des observations $X_1, \ldots, X_n$ entraînera un changement dans l'échantillon des BM et par suite dans les paramètres estimés. Nous proposons ainsi une méthode de Bootstrap par permutation particulière basée sur les rangs des BM et les statistiques d'ordre qui évite de permuter les observations, les diviser en des blocs puis d'extraire les maxima permettant ainsi de réduire le temps des calculs. Les statistiques d'ordre des observations $X_1, \ldots, X_n$ sont notées par $X_{1:n} \leq \cdots \leq X_{n,n}$. D'une manière similaire, les statistiques d'ordre des BM $M_1, \ldots, M_k$ sont notées par $M_{1:k} \leq \cdots \leq M_{k,k}$. Nous définissons alors les

rangs des BM $R_1, \ldots, R_k$ par la relation suivante:

$$M_{k+1-j:k} = X_{R_j:n}, \quad j = 1, \ldots, k.$$

Notons ici qu'on a bien $n \geq R_1 > R_2 > \ldots > R_k \geq 1$ et que $X_{R_1:n}$ est le maximum de tous les blocs, en particulier $R_1 = n$ si $n = k \times m$. La distribution des rangs des BM est exprimée en fonction d'une distribution $\mathcal{G}$ définie comme suit.

Définition. *Soient $1 \leq n_1 < n_2$ des entiers naturels. On définit par $\mathcal{G}(n_1, n_2)$ la distribution d'une variable aléatoire N telle que*

$$\mathbb{P}(N = i) = \frac{\binom{n_2 - n_1}{i}}{\binom{n_2}{i}} - \frac{\binom{n_2 - n_1}{i+1}}{\binom{n_2}{i+1}}, \quad i = 0, \ldots, n_2 - n_1.$$

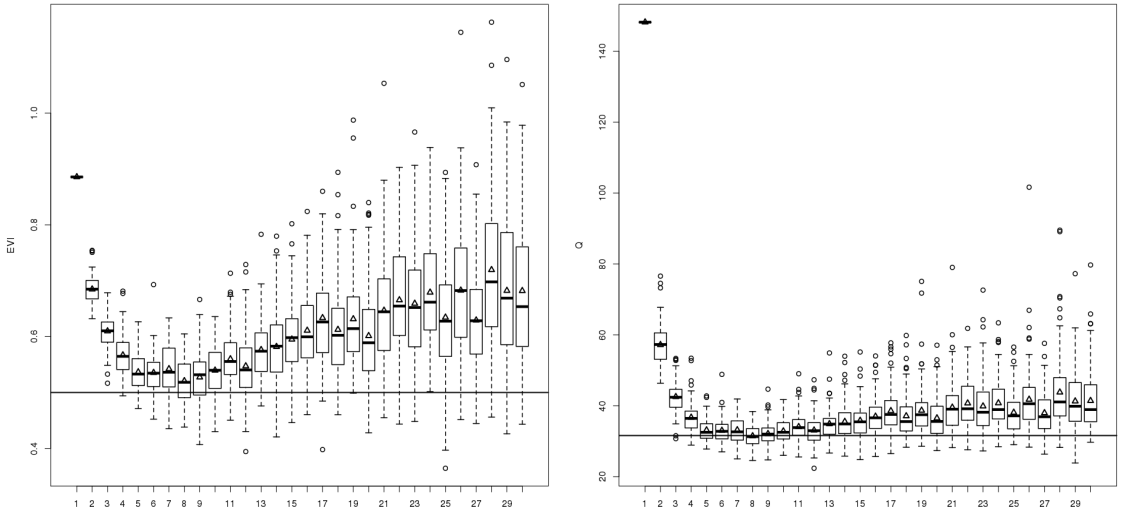
Dans le cas où $n_1 = n_2$, on définit $\mathcal{G}(n_1, n_2)$ par la masse de dirac en 0, i.e $\mathbb{P}(N = 0) = 1$.

Ainsi, la distribution des rangs des BM $(R_1, \ldots, R_k)$ sera donnée par la proposition suivante.

Proposition. *Soit $n = k \times m + r$, $r \in \{0, m - 1\}$. La distribution des rangs des BM $(R_1, .., R_k)$ d'un vecteur aléatoire échangeable $X = (X_1, .., X_n)$ est donnée par les relations suivantes:*

$$\begin{cases} \mathcal{L}(n - R_1) = \mathcal{G}(k \times m, n) \\ \mathcal{L}\left(R_{j-1} - 1 - R_j | R_1, .., R_{j-1}\right) = \mathcal{G}\left((k + 1 - j) \times m, R_{j-1} - 1\right), \quad j = 2, .., k. \end{cases}$$

La distribution des rangs des BM est alors introduite dans les simulations afin de tester la performance de la méthode. L'effet du Bootstrap par permutation est étudié dans un premier temps par une analyse numérique. On simule $n = 1000$ observations de loi Pareto de paramètre de forme $\gamma = \frac{1}{2}$. On fait varier la taille des blocs $(m = 1, \ldots, 30)$ et pour chaque valeur de m on estime les paramètres de la GEV ainsi que le quantile extrême en effectuant $B = 100$ rééchantillonnages de l'ordre des observations par notre méthode proposée. Nous illustrons la variabilité des estimations de γ et des quantiles par des boîtes à moustaches dans la figure suivante.



Boîtes à moustaches de $\hat{\gamma}$ (à gauche) et du quantile extrême estimé (à droite) en fonction de la taille des blocs.

Les boîtes à moustaches montrent la variabilité des estimations vis à vis de l'ordre des observations. En plus, nous remarquons la présence de valeurs aberrantes dans les estimations ce qui nous pousse à penser à utiliser la médiane Bootstrap (valeur estimée égale à la médiane des B estimations) en plus que la moyenne Bootstrap (valeur estimée égale à la moyenne des B estimations). En effet, la figure montre bien que la médiane Bootstrap (ligne gras dans la boite) est toujours plus proche de la vraie valeur (ligne horizontale) que la moyenne Bootstrap (triangle). Pour calculer l'erreur quadratique moyenne des estimations, on répète la simulation N' fois. Les résultats montrent que la méthode de Bootstrap par permutation proposée aboutit bien à une réduction de l'erreur quadratique moyenne, cette réduction étant plus remarquable avec la médiane-Bootstrap qu'avec la moyenne-Bootstrap. Nous discutons enfin le nombre de Bootstrap par permutation B afin de réaliser un compromis entre le temps de simulation et la réduction de l'erreur quadratique moyenne.

Chapitre 3: Détection de la tendance en extrême hétéroscédastique

Le troisième chapitre est une généralisation de la TVE classique au cas non i.i.d. particulièrement au cas où les observations sont indépendantes mais non identiquement distribuées. Les queues des distributions varient en fonction du temps et sont asymptotiquement liées par une fonction c dite *skedasis function* qui représente la fréquence des extrêmes. Ainsi le fait de tester la tendance de c nous informe sur la variation des évènements extrêmes. Par exemple, dans le cadre du réchauffement climatique il est toujours question de savoir si les évènements météorologiques extrêmes deviennent plus ou moins fréquents au cours du temps. Cette étude a pour but de développer un cadre théorique permettant de répondre à cette question en

testant la tendance de c au cours du temps.
Le modèle hétéroscédastique a été d'abord introduit par Einmahl et al. (2016) dans le papier *Statistics of heteroscedastic extremes* comme suit. Supposons qu'on dispose de n observations indépendantes $X_1^n, \ldots, X_n^n$ où X_i^n a une distribution continue $F_{n,i}$, $n \geq 1$, $1 \leq i \leq n$. On suppose en plus que toutes les distributions possèdent le même point terminal

$$x^\star = \sup\{x : F_{n,i}(x) < 1\} \in (-\infty, \infty]$$

et qu'il existe une fonction de répartition F ayant aussi $x^\star$ comme point terminal et une fonction positive c définie sur [0,1] qui caractérise la variation des extrêmes de manière à ce que les queues des distributions soient asymptotiquement proportionnelles, i.e.

$$\lim_{x \to x^\star} \frac{1 - F_{n,i}(x)}{1 - F(x)} = c(i/n), \quad \text{uniformément sur } n \geq 1,\ 1 \leq i \leq n.$$

La relation précédente décrit un modèle de queues proportionnelles déterminé par la fonction c où l'indice i/n est interprété comme étant le temps de l'observation X_i^n. Elle explicite le fait que c représente la fréquence des extrêmes. Par exemple, si c a une tendance croissante les fonctions de survie $1 - F_{n,i}$ seront aussi croissantes et cela veut dire que la fréquence des extrêmes augmente en fonction du temps. Notons que le cas $c = 1$ correspond à l'homoscédasticité et donc au cas i.i.d..
Dans le but de garder le modèle dans le cadre de la TVE, on considère que F est dans le domaine d'attraction d'une GEV de paramètre de forme γ (en particulier d'une Fréchet avec $\gamma > 0$ pour la simplicité du modèle). Par conséquent, toutes les distributions $F_{n,i}$ auront le même paramètre de forme γ. Einmahl et al. proposent une estimation non paramétrique de l'intégrale de la fonction skedasis

$$C(s) = \int_0^s c(t)\, dt, \ \ s \in [0,1],$$

donnée par

$$\hat{C}(s) = \frac{1}{k} \sum_{i=1}^{[ns]} \mathbf{1}_{\{X_i^n > X_{n-k:n}\}}.$$

Les valeurs $X_{1:n} \leq \cdots \leq X_{n:n}$ sont les statistiques d'ordre de $X_1^n, \ldots, X_n^n$ et $X_{n-k:n}$ est un seuil bien choisi en fonction de $k = k(n)$ tels que $\lim_{n\to\infty} k = \infty$ et $\lim_{n\to\infty} \frac{k}{n} = 0$. Des conditions supplémentaires sont imposées au modèle afin d'obtenir le théorème suivant sur l'estimateur non paramétrique $\hat{C}$ et l'estimateur de Hill $\hat{\gamma}_H$ défini par

$$\hat{\gamma}_H = \frac{1}{k} \sum_{i=1}^{k} \log X_{n-i+1:n} - \log X_{n-k:n}.$$

Théorème. *Sous les conditions proposées par Einmahl et al. (2016) nous avons:*

$$\sup_{0 \leq s \leq 1} |\sqrt{k} \left(\hat{C}(s) - C(s) \right) - B(C(s))| \underset{n\to\infty}{\longrightarrow} 0 \qquad p.s,$$

où B est un pont Brownien standard.
En plus, sous des hypothèses de régularité de F, l'estimateur de Hill $\hat{\gamma}_H$ satisfait

$$\sqrt{k}(\hat{\gamma}_H - \gamma) \underset{n\to\infty}{\xrightarrow{d}} \mathcal{N}(0, \gamma^2),$$

et est asymptotiquement indépendant de $\hat{C}$.

Le but de notre travail est de partir du modèle de Einmahl et al. (2016) et de considérer des modèles paramétriques de c afin de tester la présence d'une tendance. Nous introduisons d'abord la variable T_l^n qui aura un rôle primordial dans l'étude.

Définition. *Pour $1 \leq l \leq n$, on note par T_l^n le 'temps' de la $l^{\text{ème}}$ plus grande observation. En d'autre terme nous avons $X^n_{nT_l^n} = X_{n+1-l:n}$.*

Nous montrons dans un premier temps la convergence du modèle d'un point de vue heuristique. Pour cela nous présentons d'abord la proposition suivante qui a été demontrée par Resnick (1987) pour le cas i.i.d.. Nous l'introduisons et la prouvons pour le cas hétéroscédastique.

Proposition. *On suppose n observations positives $\{X_i^n,\ n \geq 1, 1 \leq i \leq n\}$ qui satisfont le modèle de queues proportionnelles avec F dans le domaine d'attraction d'une α-Fréchet, $\alpha > 0$. On considère la constante de normalisation $a_n = F^{\leftarrow}(1 - 1/n)$, $n \geq 1$, avec $F^{\leftarrow}$ la fonction quantile associée à F et on définit le processus ponctuel normalisé*

$$\Pi_n = \sum_{i=1}^{n} \varepsilon_{\left(\frac{i}{n}, \frac{X_i^n}{a_n}\right)}, \quad n \geq 1,$$

où ε représente la mesure de dirac. Alors, quand $n \to \infty$, Π_n converge en distribution dans $\mathcal{M}_p$ vers un processus de poisson Π d'intensité de mesure $c(s)\mathrm{d}s\alpha x^{-\alpha-1}\mathrm{d}x$ sur $[0,1] \times (0, \infty]$ où $\mathcal{M}_p$ représente l'espace de mesures ponctuelles $\pi = \sum_{i \geq 1} \varepsilon_{(t_i, x_i)}$.

Partant de la proposition précédente, on prouve le corollaire suivant qui montre la convergence du modèle pour un seuil fixé.

Corollaire. *Pour un k fixé $k \geq 1$, nous avons la convergence en distribution des temps et des valeurs des excès au dessus de $X_{n-k:n}$:*

$$\left(T_l^n, \frac{X_{n+1-l:n}}{X_{n-k:n}}\right)_{1 \leq l \leq k} \xrightarrow{d} (\widetilde{T}_l, \widetilde{X}_{k+1-l:k})_{1 \leq l \leq k}, \quad \textit{quand } n \to \infty,$$

où $\widetilde{X}_{1:k} \leq \cdots \leq \widetilde{X}_{k:k}$ sont les statistiques d'ordre d'observations i.i.d. $\widetilde{X}_1, \ldots, \widetilde{X}_k$ de loi α-Pareto, i.e. $\mathbb{P}(\widetilde{X}_l > x) = x^{-\alpha}$, $x > 1$, et, indépendamment, $\widetilde{T}_1, \ldots, \widetilde{T}_k$ sont des variables aléatoires indépendantes de densité c.

Nous étudions ensuite les modèles paramétriques $c = c_\theta$, $\theta \in \Theta$, de la forme log-linéaire, linéaire et log-linéaire discret ainsi que les résultats de consistance et de normalité asymptotique du paramètre θ représentant la tendance. On se limite dans ce chapitre introductif au modèle log-linéaire défini par

$$c\,(t) = c_\theta\,(t) = e^{\theta t - h(\theta)} \quad \text{où} \quad \theta \in \mathbb{R},$$

avec

$$h(\theta) = \log\left(\int_0^1 e^{\theta t} dt\right) = \log\left(\frac{e^\theta - 1}{\theta}\right).$$

Nous obtenons les résultats suivants.

Proposition. *Considérons le modèle log-linéaire* $\{c_\theta(t), \theta \in \mathbb{R}\}$ *et* T *une variable aléatoire de densité* c_θ*. On a,*

$$\mathbb{E}_\theta(T) = \frac{\partial h}{\partial \theta}(\theta) = \begin{cases} \dfrac{\theta e^\theta - e^\theta + 1}{\theta e^\theta - \theta} & si\ \theta \neq 0, \\ \dfrac{1}{2} & si\ \theta = 0, \end{cases}$$

et

$$\mathrm{Var}_\theta(T) = \frac{\partial^2 h}{\partial \theta^2}(\theta) = \begin{cases} \dfrac{(e^\theta - 1)^2 - \theta^2 e^\theta}{(\theta e^\theta - \theta)^2} & si\ \theta \neq 0, \\ \dfrac{1}{12} & si\ \theta = 0. \end{cases}$$

En plus, l'information de Fisher en θ *est*

$$I(\theta) = \mathrm{Var}_\theta(T).$$

A noter que $\dfrac{\partial h}{\partial \theta} : (-\infty, \infty) \longrightarrow (-1, 1)$ *est un difféomorphisme.*

Par conséquent, l'estimateur par maximum de vraisemblance de θ sera

$$\hat{\theta}_k = \left(\frac{\partial h}{\partial \theta}\right)^{-1} (\bar{T_k}) \text{ avec } \bar{T_k} = \frac{1}{k}\sum_{i=1}^{k} T_i^n.$$

Les résultats de consistance et de normalité asymptotique de $\hat{\theta}_k$ sont donnés dans le théorème suivant.

Théorème. *Supposons le modèle log-linéaire* c_{θ_0} *de* c*. Alors sous les conditions de Einmahl et al. (2016), l'estimateur de maximum vraisemblance* $\hat{\theta}_k$ *de* θ_0 *satisfait*

$$\hat{\theta}_k \xrightarrow{d} \theta_0 \quad quand\ n \to \infty$$

et

$$\sqrt{k}\left(\hat{\theta}_k - \theta_0\right) \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0, I(\theta_0)^{-1}\right).$$

Puisque le but est de tester la tendance de c, nous proposons alors le test paramétrique

$$H_0 : \theta_0 \leq 0 \qquad \text{contre} \qquad H_1 : \theta_0 > 0,$$

défini plus formellement par:

$$\varphi_k = \begin{cases} 0 & \text{si } \hat{\theta}_k \leq z_{1-\alpha}\frac{\sigma_0}{\sqrt{k}}, \\ 1 & \text{sinon}, \end{cases}$$

où ${\sigma_0}^2 = \frac{1}{I(\theta=0)} = \frac{1}{\mathrm{Var}_{\theta=0}(T)} = 12$. Le test paramétrique φ est consistant sous l'hypothèse de monotonicité de c.

Proposition. *Supposons la fonction skedasis c comme monotone et le test*

$$H_0 : c \text{ est décroissante} \quad \text{contre} \quad H_1 : c \text{ est croissante non constante.}$$

Le test φ_k satisfait

$$\lim_{n\to\infty} \mathbb{P}(\varphi_k = 0) \geq 1 - \alpha \quad \text{si } c \text{ est décroissante},$$
$$\lim_{n\to\infty} \mathbb{P}(\varphi_k = 1) = 1 \quad \text{si } c \text{ est croissante non constante},$$

avec α le seuil de signification.

Nous passons ensuite aux simulations afin de discuter deux idées principales: le choix du seuil et la puissance des tests pour la détection de la tendance. Nous proposons en premier temps la méthode de Lepski présentée par Boucheron and Thomas (2015) pour le choix du seuil et nous comparons dans un deuxième temps la puissance du test paramétrique T_1: $H_0 : \theta = 0$ contre $H_1 : \theta \neq 0$ (présenté dans le modèle log-linéaire) à la puissance du test non paramétrique de Kolmogorov-Smirnov T_2. Dans le but de réaliser cette comparaison, nous simulons un échantillon de $n = 1000$ observations qui vérifient le modèle de queues proportionnelles telle que la fonction c est monotone mais n'est pas issue du modèle paramétrique log-linéaire définie par

$$c(s) = 1 - \delta \cos(\pi s), \quad s \in [0, 1],$$

avec $\delta = (0, 0.2, 0.4, 0.6)$. Le cas $\delta = 0$ correspond à l'homoscédasticité. Nous faisons ensuite varier la valeur de k ($k = 100, 200, 300$). Pour chaque valeur de k et δ nous testons l'hypothèse d'homoscédasticité pour les deux tests T_1 et T_2 de niveaux 5%. Nous répétons la simulation $N = 1000$ fois et nous comptons le nombre de fois qu'on accepte H_0. Les résultats sont donnés dans le tableau suivant:

		$k = 100$	$k = 200$	$k = 300$
$\delta = 0$	T_1	0.963	0.975	0.982
	T_2	0.972	0.988	0.988
$\delta = 0.2$	T_1	0.697	0.479	0.285
	T_2	0.755	0.601	0.428
$\delta = 0.4$	T_1	0.159	0.005	0.000
	T_2	0.238	0.017	0.000
$\delta = 0.6$	T_1	0.004	0.000	0.000
	T_2	0.014	0.000	0.000

Probabilité estimée d'accepter H_0 pour le test paramétrique T_1 et non paramétrique T_2 pour détecter la tendance.

Les résultats montrent que c'est bien le test paramétrique T_1 qui rejette le plus l'hypothèse nulle d'homoscédasticité dans le cas où elle est fausse. Ainsi, le test paramétrique est plus puissant que le test non paramétrique d'où l'importance de considérer des modèles paramétriques de la tendance.
Finalement, nous appliquons la méthodologie aux données de températures minimales et maximales dans la région de Fort Collins Colorado durant le $20^{ème}$ siècle pour détecter la présence d'une tendance dans les extrêmes sur cette période.

Chapitre 4: Modèles GEV avec covariables pour les maxima annuels et sélection de modèles pour la prédiction des quantiles extrêmes

Dans ce chapitre nous présentons nos résultats obtenus au Data challenge proposé par le comité d'organisation de la conférence internationale EVA 2017. On dispose d'un jeu de données de précipitations journalières maximales sur 24 ans dans 40 stations. On réalise une prédiction spatio-temporelle des quantiles correspondants à un niveau de retour de 20 ans pour les précipitations mensuelles dans chaque station. Nous utilisons des modèles de GEV en introduisant les covariables mois, latitude et longitude des stations dans les paramètres μ et σ de la loi GEV (γ = cte pour la simplicité). Nous proposons trois grands modèles pour décrire les interactions entre les covariables:

- Le modèle additif Add_h où

$$\mu = month + f_1(latitude) + f_2(longitude),$$
$$\sigma = month + g_1(latitude) + g_2(longitude),$$

  avec f_1, f_2, g_1 et g_2 des splines cubiques naturelles de degré de liberté $df = h$.

- Le modèle mixte Mix_h où:

$$\mu = month + f_1(latitude) \times f_2(longitude),$$
$$\sigma = month + g_1(latitude) \times g_2(longitude).$$

- Le modèle multiplicatif $Mult_h$ où:

$$\mu = month \times f_1(latitude) \times f_2(longitude),$$
$$\sigma = month \times g_1(latitude) \times g_2(longitude).$$

Le meilleur modèle est choisi en terme d'Akaike information criterion (AIC) et par la méthode de validation croisée. Par exemple, les modèles proposés dans un premier temps avec leur critère AIC sont les suivants:

Model	K	AIC
Add_4	41	3917.797
Add_6	49	2286.818
Mix_2	41	2723.934
Mix_4	**73**	**2189.872**
Mix_6	121	2594.673
$Mult_4$	601	4290.686

AIC des 6 modèles

Le modèle choisi est celui qui a la plus petite valeur AIC et donc le modèle Mix_4. Dans un deuxième temps, nous proposons d'autres modèles et appliquons la méthode de validation croisée pour le choix du meilleur modèle. Les modèles proposés avec leur nombre de paramètre K ainsi que leur score (fonction score proposée par le challenge) sont présentés ci-dessous:

Model	K	Score
Add_1	**29**	**1018.448**
Add_2	33	1018.415
Add_3	37	1017.906
Add_4	41	1016.575
Add_5	45	1073.358
Add_6	49	1053.673
Mix_1	31	1018.573
Mix_2	41	1100.853
Mix_3	55	1052.124
Mix_4	73	1052.265
Mix_5	95	1067.313
Mix_6	121	1089.812
$Mult_1$	97	1014.848
$Mult_2$	217	1025.043
$Mult_3$	385	1028.709

Score obtenu par validation croisée ($k = 4$)

Le modèle choisi est celui qui assure un compromis entre le nombre de paramètres et la plus petite valeur du score. Nous choisissons donc le modèle Add_1.
Pour chacun des deux modèles choisis, nous estimons les quantiles extrêmes. En utilisant les scores de prédiction calculés grâce à l'échantillon test, nous déduisons qu'il est plus favorable de choisir le meilleur modèle par validation croisée que par AIC puisque la première méthode réduit le score du benchmark plus que la deuxième et donne donc de meilleures estimations des quantiles extrêmes.

Chapitre 5: Protection de la plateforme pétrolière contre les vents et les vagues extrêmes au Liban

Le dernier chapitre est une application de la TVE unvariée et bivariée sur la vitesse du vent et la hauteur des vagues hivernales dans la région de Beddawi au nord du Liban en vue de protéger la plateforme pétrolière qui y sera installée de ces risques environnementaux. En effet la plateforme pétrolière doit être conçue pour résister aux risques environnementaux notamment la vitesse du vent et la hauteur des vagues. Ses composantes sont construites pour faire face à des niveaux de retour associés à des périodes de retour de $T = 50$, 100 voire 500 ans. Notons dans cette partie le vecteur de valeurs extrêmes par $X^\star$ dans le but de simplifier les notations. On entend par niveau de retour de T ans noté $q_{X^\star,T}$ dans le cas de maxima annuels la valeur qui sera dépassée en moyenne une fois tous les T ans avec une probabilité de $1/T$. Ce niveau de retour correspond donc au quantile d'ordre $p = 1 - 1/T$ de la loi GEV des maxima et est donné par:

$$q_{X^\star,T} = \begin{cases} \mu + \frac{\sigma}{\gamma}[(-\log p)^{-\gamma} - 1] & \text{si } \gamma \neq 0 \\ \mu - \sigma \log(-\log p) & \text{si} \gamma = 0 \end{cases}.$$

On commence d'abord par compléter les données manquantes des hauteurs des vagues par l'algorithme de forêt aléatoire disponible dans le package **missForest**

du logiciel R. Nous choisissons cet algorithme puisqu'il utilise la méthode de forêt aléatoire appropriée aux cas où les données peuvent présenter des interactions non linéaires et quand nous ne connaissons pas leur distribution à priori. En plus, cet algorithme a été comparé à d'autres méthodes utilisées pour compléter les données manquantes et les a surpassées. Une fois les données complétées, on applique d'abord la théorie univariée sur la vitesse du vent et la hauteur des vagues séparément en utilisant la méthode des BM pour estimer les paramètres de la GEV et les niveaux de retour des maxima associés à des périodes de retour de 50, 100 et 500 années. Dans notre cas, nous appliquons la méthode des BM aux données journalières et nous choisissons des blocs de taille $m = 15$ jours. En effet, on étudie les données hivernales disponibles pour les années 2000 à 2015. Le nombre total des observations est donc $n = 1353$ et le choix de la taille des blocs doit être un compromis biais-variance. Avec le choix de $m = 15$ nous obtenons 90 maxima avec $\ell = 6$ maxima par an. Ainsi, le niveau de retour de T ans correspondra à un quantile d'ordre $p = 1 - \frac{1}{6\times T}$ et présentera la valeur qui sera dépassée en moyenne une fois tous les $6 \times T$ blocs. Les résultats des niveaux de retour estimés pour la vitesse du vent extrême ($X^\star$) et la hauteur des vagues extrême ($Y^\star$) ainsi que leur intervalle de confiance sont donnés dans les tableaux suivants:

T (ans)	$\hat{q}_{X^\star,T}$ (m/s)	$\hat{q}_{Y^\star,T}$ (cm)
50	43.570	284.839
100	47.530	298.510
500	57.197	327.238

Niveaux de retour pour $X^\star$ et $Y^\star$.

T (ans)	Intervalle de confiance de $q_{X^\star,T}$	Intervalle de confiance de $q_{Y^\star,T}$
50	[32.014,55.126]	[237.160,332.517]
100	[32.831,62.228]	[240.854,356.166]
500	[33.271,81.123]	[244.034,410.442]

Intervalle de confiance des niveaux de retour

Le premier tableau montre par exemple que la vitesse du vent extrême qu'on s'attend à être dépassée une fois toutes les 50 années ou 50×6 blocs est 43.570 m/s. En plus, la valeur 298.510 cm est la hauteur des vagues qui sera dépassée en moyenne une fois toutes les 100 années ou 600 blocs avec une probabilité de 1/600. Nous passons ensuite à l'application de la théorie bivariée en utilisant aussi la méthode des BM qui utilise les vecteurs des maxima par composante. Soit $(X_1, Y_1), (X_2, Y_2), ..$ une séquence de vecteurs indépendants de loi $F(x, y)$ avec F_1 et F_2 les distributions marginales respectives de X et Y. On définit alors

$$X^\star = \max_{i=1,..,m} X_i \qquad \text{et} \qquad Y^\star = \max_{i=1,..,m} Y_i.$$

Ainsi,

$$M_m = (X^\star, Y^\star)$$

est le vecteur de maxima par composante obtenu en combinant $X^\star$ and $Y^\star$ et dont les composantes ne sont pas nécessairement une vraie réalisation. On suppose qu'il existe des constantes de normalisation $a_m = (a_{1,m}, a_{2,m}) \in \mathbb{R}_+^{\star 2}$ et $b_m = (b_{1,m}, b_{2,m}) \in \mathbb{R}^2$ ainsi qu'une distribution non dégénérée G tel que

$$\mathbb{P}\left(\frac{M_m - b_m}{a_m} \leq z\right) = F^m(a_m z + b_m) \underset{m\to\infty}{\longrightarrow} G(z),$$

avec $z = (x, y)$, $\left(\frac{M_m - b_m}{a_m}\right) = \left(\frac{X^\star - b_{1,m}}{a_{1,m}}, \frac{Y^\star - b_{2,m}}{a_{2,m}}\right)$ et $a_m z = (a_{1,m}x, a_{2,m}y)$. F est alors dite dans le domaine d'attraction de G. Il est vrai que G ne présente pas une forme bien paramétrique mais en transformant les marginales F_1 et F_2 en Fréchet standard nous pouvons simplifier la représentation de G. Nous nous ramenons enfin à des modèles paramétriques de G notamment le modèle logistique où la distribution G est définie comme suit:

$$G(x, y) = \exp\{-(x^{-1/\alpha} + y^{-1/\alpha})^\alpha\},\ x > 0,\ y > 0,\ \alpha \in [0, 1],$$

avec α le paramètre de dépendance entre $X^\star$ et $Y^\star$:
- $\alpha \to 1$, alors $G(x, y) \to \exp\left(-(\frac{1}{x} + \frac{1}{y})\right)$ ce qui correspond à une indépendence.
- $\alpha \to 0$, alors $G(x, y) \to \exp\left(-\max(\frac{1}{x}, \frac{1}{y})\right)$ ce qui correspond à une dépendance totale.

A noter que G peut aussi être aussi exprimée en fonction de copules qui sont des fonctions $C(.,.) : [0, 1] \times [0, 1] \to [0, 1]$, qui vérifient les conditions suivantes:

- $C(u, 0) = C(0, v) = 0$, $C(u, 1) = u$ et $C(1, v) = v$, $\forall u, v \in [0, 1]$,
- $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$, $\forall\, u_1,\ u_2,\ v_1,\ v_2 \in [0, 1]$ tels que $u_1 \leq u_2$ et $v_1 \leq v_2$.

Nous estimons donc la dépendance entre les vents et les vagues extrêmes et les probabilités jointes de dépassement des niveaux de retour univariés:

$$\mathbb{P}(X^\star > \hat{q}_{X^\star,T}, Y^\star > \hat{q}_{Y^\star,T'}) = 1 - G_1(\hat{q}_{X^\star,T}) - G_2(\hat{q}_{Y^\star,T'}) + G(\hat{q}_{X^\star,T}, \hat{q}_{Y^\star,T'})$$

où G_1 et G_2 les distributions marginales des maxima univariés. Les résultats de l'estimation des probabilités précédentes sont présentés ci-dessous:

Niveaux de retour	$\hat{q}_{Y^\star,50} = 284.839$	$\hat{q}_{Y^\star,100} = 298.510$	$\hat{q}_{Y^\star,500} = 327.238$
$\hat{q}_{X^\star,50} = 43.570$	0.00106	0.00071	0.00019
$\hat{q}_{X^\star,100} = 47.530$	0.00071	0.00053	0.00018
$\hat{q}_{X^\star,500} = 57.197$	0.00019	0.00018	0.00011

$$\mathbb{P}(X^\star > \hat{q}_{X^\star,T}, Y^\star > \hat{q}_{Y^\star,T'})$$

Nous associons ces probabilités jointes de dépassement à des périodes de retour jointes par la relation suivante: $\mathbb{P}(X^\star > \hat{q}_{X^\star,T}, Y^\star > \hat{q}_{Y^\star,T'}) = \frac{1}{\ell T'_{X^\star Y^\star}}$ où $T'_{X^\star Y^\star}$ est la période de retour jointe associée à l'évènement ($X^\star > \hat{q}_{X^\star,T}$ et $Y^\star > \hat{q}_{Y^\star,T'}$) et ℓ le nombre de blocs par an. Les résultats sont donnés dans le tableau suivant:

Niveaux de retour	$\hat{q}_{Y^\star,50} = 284.839$	$\hat{q}_{Y^\star,100} = 298.510$	$\hat{q}_{Y^\star,500} = 327.238$
$\hat{q}_{X^\star,50} = 43.570$	157.273	236.309	868.057
$\hat{q}_{X^\star,100} = 47.530$	236.309	315.498	947.920
$\hat{q}_{X^\star,500} = 57.197$	868.057	947.920	1581.323

Période de retour jointe $T'_{X^\star Y^\star}$

Ce tableau montre que la période de retour jointe correspondante à l'évènement que la vitesse du vent extrême $X^\star$ dépasse le niveau de retour $\hat{q}_{X^\star,T}$ et la hauteur des vagues extrême dépasse le niveau de retour $\hat{q}_{Y^\star,T'}$ est supérieure au maximum entre les périodes de retour marginales T et T'. Par exemple la période de retour jointe de 157.273 ans correspond à la probabilité que la vitesse du vent et la hauteur de vague extrême dépassent simultanément les niveaux de retour de période 50 ans.

1

Fundamental of extreme value theory

Abstract

We present in this chapter a short introduction to univariate extreme value theory. We start by introducing the asymptotic distribution of the maxima. We then present the two main methods used to extract an extreme sample given a series of i.i.d. observations: the block maxima and the peak over threshold methods. The block maxima are approximated by a generalized extreme value distribution and the threshold excesses by a generalized Pareto distribution. Once the parameters distribution are estimated by either the maximum likelihood or the probability weighted moments, we can obtain an estimation of the return levels.

Contents

1.1 Block maxima model

1.1.1 Generalized Extreme Value distribution GEV

Suppose we have a sample of m i.i.d. random variables $X_1, \ldots, X_m$ with continuous distribution function F. In EVT we are interested in the behavior of $M_m = \max(X_1, \ldots, X_m)$, i.e. in the probability $\mathbb{P}(M_m \leq x)$. In fact, we have

$$\begin{aligned}\mathbb{P}(M_m \leq x) &= \mathbb{P}(X_1 \leq x, X_2 \leq x, \ldots, X_m \leq x) \\ &= \mathbb{P}(X_1 \leq x) \times \mathbb{P}(X_2 \leq x) \times \ldots \times \mathbb{P}(X_m \leq x) \\ &= F(x)^m.\end{aligned}$$

Unfortunately, since the function F is unknown it is the same for F^m and hence for the distribution of M_m. One can think of estimating F but even this can't solve the problem because a little discrepancy for F results in a significant discrepancy for F^m (Coles (2001)). The idea is to directly find the asymptotic behavior of F^m. Since $F^m(x) \underset{m\to\infty}{\longrightarrow} 0$ for every $x < x^\star = \sup\{x : F_{m,i}(x) < 1\}$, an alternative is to normalize the maximum M_m in order to get a non degenerate asymptotic distribution for the normalized maxima. This is solved by Gnedenko-Fisher-Tippett theorem, a fundamental in EVT (Gnedenko (1943); Fisher and Tippett (1928)).

Theorem 1.1 (Gnedenko-Fisher-Tippett)**.** *Consider a sequence of i.i.d. random variables (r.v.) $X = X_1, X_2, \ldots$ with distribution function $F(x) = \mathbb{P}(X \leq x)$ and let $M_m = \max(X_1, .., X_m)$. If there exist normalizing constants $a_m > 0$ and b_m such that*

$$\lim_{m\to\infty} \mathbb{P}\left(\frac{M_m - b_m}{a_m} \leq x\right) = \lim_{m\to\infty} F^m(a_m x + b_m) := G(x)$$

with G a non-degenerate distribution function, then G must be of one of the following types:

- *Gumbel:* $G(x) = \exp(-\exp(-x)) \quad$ *for all* $x \in \mathbb{R}$;
- *Fréchet:* $G(x) = \begin{cases} \exp(-x^{-\alpha}) & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$, $\alpha > 0$;
- *Weibull:* $G(x) = \begin{cases} \exp(-(-x)^{\alpha}) & \text{if } x \leq 0 \\ 1 & \text{if } \quad x > 0 \end{cases}$, $\alpha > 0$.

The variable X is said to be in the domain of attraction of G. The strength of this theorem is that regardless of the distribution F, the normalized maxima will have necessary one of the three previous distributions. Jenkinson (1969) combined them in a single parametric form called the generalized extreme value (GEV) distribution of the form:

$$G_{\mu,\sigma,\gamma}(x) := \begin{cases} \exp\{-[1 + \gamma\left(\frac{x-\mu}{\sigma}\right)]^{-1/\gamma}\}, & 1 + \gamma\frac{x-\mu}{\sigma} > 0, \gamma \neq 0, \\ \exp\left(-\exp(-\frac{x-\mu}{\sigma})\right), & x \in \mathbb{R}, \gamma = 0, \end{cases}$$

where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ the scale parameter and $\gamma \in \mathbb{R}$ the extreme value index. Based on the sign of the extreme value index γ, we could identify to which type of the three families does the GEV belong.

- If $\gamma > 0$, the GEV is a Fréchet which is the domain of attraction of the heavy-tailed distributions such as the Pareto distribution.
- If $\gamma = 0$, the GEV is a Gumbel which is the domain of attraction of the light-tailed distributions such as the Normal distribution.
- If $\gamma < 0$, the GEV is a Weibull which is the domain of attraction of the short tailed distributions with finite upper bounds such as the Uniform distribution.

Now that the asymptotic behavior of the normalized maxima $\frac{M_m - b_m}{a_m}$ is known, it remains to identify the normalizing constants a_m and b_m. In fact, we can get around the problem of identifying these constants in the following way: Theorem 1.1 states that as $m \to \infty$, $\mathbb{P}\left(\frac{M_m - b_m}{a_m} \leq x\right) \simeq G(x)$. Hence,

$$\mathbb{P}(M_m \leq x) \simeq G_{\mu,\sigma,\gamma}\left(\frac{x - b_m}{a_m}\right) = G_{a_m\mu + b_m, a_m\sigma,\gamma}(x) = G^\star(x)$$

where $G^\star$ is a GEV distribution with parameters $a_m\mu + b_m$, $a_m\sigma$ and γ. Consequently, the maxima M_m follow also approximately a GEV distribution whose parameters should be estimated.

1.1.2 The Block maxima method

In practice, the method used to extract the maxima sample from a series of i.i.d. observations is the block maxima (BM) method. It consists in dividing the observations into blocks and to extract the maximum within each block. Suppose we have a sequence of n i.i.d. observations $X_1, \ldots, X_n$ with distribution F divided in k blocks of size m. The set of the k maxima can be denoted by $X_1^\star, \ldots, X_k^\star$ for the sake of simplicity and will follow approximately a GEV distribution.
The BM method has its pros and cons: in fact, the method can be useful to ensure the assumption of i.i.d. maxima. For example in the case of seasonality in the observations, by choosing a block size of one year, the yearly maxima can be considered as identically distributed. On the other hand, the BM method leads to a loss of information: by choosing only the maximum value in a block, it may ignore other existing extreme values in the same block. These ignored values can even be more extreme than the maxima of other blocks.

1.1.3 Parameters estimation

The GEV parameters can be estimated by either the maximum likelihood (ML) method or the probability weighted moments (PWM) method.

Maximum likelihood estimation

Suppose the i.i.d. r.v. $X_1^\star, \ldots, X_k^\star$ following a GEV distribution and whose realizations are $x_1^\star, \ldots, x_k^\star$. The log-likelihood function is given by (see Beirlant et al. (2004))

$$\ell(\mu, \sigma, \gamma) = -k \log \sigma - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^{k} \log\left(1 + \gamma \frac{x_i^\star - \mu}{\sigma}\right) - \sum_{i=1}^{k} \left(1 + \gamma \frac{x_i^\star - \mu}{\sigma}\right)^{-1/\gamma},$$

if $\gamma \neq 0$ and $1 + \gamma \frac{x_i^\star - \mu}{\sigma} > 0$. In the case $\gamma = 0$,

$$\ell(\mu, \sigma, 0) = -k \log \sigma - \sum_{i=1}^{k} \exp\left(-\frac{x_i^\star - \mu}{\sigma}\right) - \sum_{i=1}^{k} \frac{x_i^\star - \mu}{\sigma}.$$

The ML estimator $(\hat{\mu}, \hat{\sigma}, \hat{\gamma})$ is then the solution of the following score equations:

$$\frac{\partial \ell(\mu, \sigma, \gamma)}{\partial \mu} = 0, \quad \frac{\partial \ell(\mu, \sigma, \gamma)}{\partial \sigma} = 0 \quad \text{and} \quad \frac{\partial \ell(\mu, \sigma, \gamma)}{\partial \gamma} = 0.$$

More computational details on the ML estimation of the GEV parameters date back to Prescott and Walden (1980, 1983).

Probability weighted moments estimation

The PWM method was first introduced by Greenwood et al. (1979). In a general framework, the PWM of a r.v. X with distribution function F are defined by

$$M_{p,r,s} := \mathbb{E}\left[X^p \{F(X)\}^r \{1 - F(X)\}^s\right]$$

with p, r and $s \in \mathbb{R}$ and can be used to estimate the parameter of a distribution. For the GEV distribution, the PWM with $p = 1$, $r = 0, 1, 2$ and $s = 0$ are used: they are well defined for $\gamma < 1$ and given by

$$\beta_r = \mathbb{E}\left[X \{F(X)\}^r\right] = \frac{1}{r+1} \left\{\mu - \frac{\sigma}{\gamma}\left[1 - (1+r)^\gamma \Gamma(1-\gamma)\right]\right\}, \quad r = 0, 1, 2,$$

(see Beirlant et al. (2004)). These three PWM allow to retrieve the three GEV parameters thanks to the relations

$$\begin{aligned}
\beta_0 &= \mu - \frac{\sigma}{\gamma}\left[1 - \Gamma(1-\gamma)\right], \\
2\beta_1 - \beta_0 &= \frac{\sigma}{\gamma}\Gamma(1-\gamma)\left(2^\gamma - 1\right), \\
\frac{3\beta_2 - \beta_0}{2\beta_1 - \beta_0} &= \frac{3^\gamma - 1}{2^\gamma - 1}.
\end{aligned}$$

In practice, the PWM are estimated by

$$\hat{\beta}_r = \frac{1}{k} \sum_{j=1}^{k} \frac{(j-1)(j-2)..(j-r)}{(k-1)(k-2)..(k-r)} X_{j:k},$$

where $X_{1:k} \leq X_{2:k} \leq ...X_{k:k}$ are the order statistics of $X_1, ..., X_k$ (see Caeiro and Gomes (2011)). Plugging-in these estimates for β_r, $r = 0, 1, 2$, in the three equations above, we obtain the so-called PWM estimators $\hat{\gamma}$, $\hat{\mu}$ and $\hat{\sigma}$.

1.1.4 Quantile estimation

Extreme quantile estimation is central in EVT. Based on BM method, the quantile $q_{X^\star,p}$ of order p of the vector of maxima $X^\star = X_1^\star, \ldots, X_k^\star$ is given by

$$q_{X^\star,p} = G^{\leftarrow}_{\mu,\sigma,\gamma}(p), \quad p \in (0,1).$$

The relation $G = F^m$ implies that the quantile of order p of X corresponds to the quantile of order p^m of $X^\star$. Hence an estimator of $q_{X,p}$ is given by the following formula where μ, σ and γ have to be replaced by their estimated values (see Beirlant et al. (2004)):

$$q_{X,p} = q_{X^\star,p^m} = \begin{cases} \mu + \frac{\sigma}{\gamma}[(-m\log p)^{-\gamma} - 1] & \text{if } \gamma \neq 0 \\ \mu - \sigma\log(-m\log p) & \text{if } \gamma = 0 \end{cases}.$$

Recall that m is the block size in the BM method.

1.2 Threshold models

We have already presented the first method to extract an extreme sample from a series of observations which is the BM method. As we have said, its main disadvantage is that it may lead to a loss of information. We present here an alternative method which takes into consideration all the values that exceed a high threshold known as the peak over threshold (POT) method.

1.2.1 Generalized Pareto Distribution GPD

Suppose again that we have a sequence $X = X_1, X_2, \ldots, X_n$ of i.i.d. r.v. with distribution F. For a high threshold u, we are interested in the asymptotic behavior of the conditional probability

$$F_u(y) := \mathbb{P}(X - u \leq y | X > u) = \frac{F(u+y) - F(u)}{1 - F(u)} \tag{1.1}$$

which is given in the following theorem.

Theorem 1.2 (Balkema and de Haan (1974))**.** *Let $X = X_1, X_2, \ldots$ be a sequence of i.i.d. r.v. with distribution F. For large u, the conditional excess distribution F_u defined in (1.1) satisfies*

$$F_u(y) \simeq H_{\beta,\gamma}(y)$$

where $H_{\beta,\gamma}$ is the generalized Pareto distribution (GPD) defined by

$$H_{\beta,\gamma}(y) := \begin{cases} 1 - \left(1 + \gamma\frac{y}{\beta}\right)^{-1/\gamma} & \text{if } \gamma \neq 0 \\ 1 - \exp\left(-\frac{y}{\beta}\right) & \text{if } \gamma = 0 \end{cases}.$$

The GPD distribution can be obtained from the GEV distribution in the following way (Coles (2001)). Recall the maximum $M_m = \max(X_1, \ldots, X_m)$. We have

$$\mathbb{P}(M_m \leq x) = F^m(x) \simeq G(x)$$

with $G(x) = \exp\{-[1+\gamma\left(\frac{x-\mu}{\sigma}\right)]^{-1/\gamma}\}$, for $\mu, \sigma > 0$ and $\gamma \in \mathbb{R}$.
Consequently,

$$m \log F(x) \simeq -[1+\gamma\left(\frac{x-\mu}{\sigma}\right)]^{-1/\gamma}.$$

It follows from the Taylor approximation for large x, $\log(F(x)) \simeq -(1-F(x))$, that

$$1 - F(x) \simeq \frac{\left[1+\gamma(\frac{x-\mu}{\sigma})\right]^{-1/\gamma}}{m}.$$

Consequently for $y > 0$ and $1 + \gamma\frac{y}{\beta} > 0$,

$$\begin{aligned}
\mathbb{P}(X - u \leq y | X > u) &= \frac{F(u+y) - F(u)}{1 - F(u)} \\
&\simeq \frac{\left[1+\gamma(\frac{u-\mu}{\sigma})\right]^{-1/\gamma} - \left[1+\gamma(\frac{u+y-\mu}{\sigma})\right]^{-1/\gamma}}{\left[1+\gamma(\frac{u-\mu}{\sigma})\right]^{-1/\gamma}} \\
&= 1 - \left[\frac{1+\gamma(\frac{u+y-\mu}{\sigma})}{1+\gamma(\frac{u-\mu}{\sigma})}\right]^{-1/\gamma} \\
&= 1 - \left(1+\gamma\frac{y}{\beta}\right)^{-1/\gamma},
\end{aligned}$$

with $\beta = \sigma + \gamma(u - \mu)$. The result for $\gamma = 0$ is directly obtained by taking the limit $\gamma \to 0$. Hence, we can conclude that if the maxima are approximated by a GEV distribution then the threshold excesses are approximated by a GPD distribution. Moreover, the two distributions share the same parameter γ called the extreme value index.

1.2.2 Parameters estimation

Similarly to the GEV distribution, the GPD parameters can be estimated by the ML or the PWM method.

Maximum likelihood estimation

Consider n i.i.d. r.v. $X_1, \ldots, X_n$ with distribution F. The threshold excesses $Y_i = X_i - u$, $i = 1, \ldots, k$ follow approximately a GPD distribution of the form:

$$H_{u,\beta,\gamma}(y) = 1 - \left(1+\gamma\frac{y}{\beta}\right)^{-1/\gamma}, \quad \gamma \neq 0.$$

The density is then

$$h(y) = \frac{1}{\beta}\left(1+\gamma\frac{y}{\beta}\right)^{-\frac{1}{\gamma}-1},$$

and the log-likelihood function equation given that $1 + \gamma\frac{y_i}{\beta} > 0$ is

$$\ell(\beta, \gamma) = -k\log\beta - \left(\frac{1}{\gamma}+1\right)\sum_{i=1}^{k}\log\left(1+\frac{\gamma}{\beta}y_i\right),$$

with y_i, $i = 1, \ldots, k$ the realizations of Y_i. If $1 + \gamma \frac{y}{\beta} \leq 0$, $\ell(\beta, \gamma) = -\infty$ (Coles (2001)). The log-likelihood function in the case $\gamma = 0$ is equal to

$$\ell(\beta) = -k \log(\beta) - \beta^{-1} \sum_{i=1}^{k} y_i.$$

The estimated parameter $\hat{\beta}$ and $\hat{\gamma}$ are then derived from the score equations:

$$\frac{\partial \ell(\beta, \gamma)}{\partial \beta} = 0 \quad \text{and} \quad \frac{\partial \ell(\beta, \gamma)}{\partial \gamma} = 0.$$

Further details on this method can be found in Smith (1987), Coles (2001) and Embrechts et al. (1997).

Probability weighted moments estimation

The starting point for the estimation of the GPD parameters with the PWM method is the quantity

$$w_r = \mathbb{E}[Y(\bar{H}_{\beta,\gamma}(Y))^r] = \frac{\beta}{(r+1)(r+1-\gamma)}, \quad r = 0, 1,$$

with Y a random variable following the GPD distribution $H_{\beta,\gamma}$ and $\bar{H}_{\beta,\gamma} = 1 - H_{\beta,\gamma}$ the survival function. The parameters β and γ are obtained via the relations:

$$\beta = \frac{2w_0 w_1}{w_0 - 2w_1} \quad \text{and} \quad \gamma = 2 - \frac{w_0}{w_0 - 2w_1}.$$

By replacing w_0 and w_1 by the empirical moment estimators we obtain the estimations $\hat{\beta}$ and $\hat{\gamma}$ (Embrechts et al. (1997)).

1.2.3 Return level estimation

We now move to the aim of the EVT which is the return level estimation. More precisely, we want to find the t-observation return level x_t that is exceeded on average once every t observations. Hence, x_t is the solution of the equation

$$\mathbb{P}(X > x_t) = \frac{1}{t}.$$

We have for $\gamma \neq 0$ and $x > u$,

$$\begin{aligned} \mathbb{P}(X > x | X > u) &= 1 - \mathbb{P}(X \leq x | X > u) \\ &= \left(1 + \gamma \frac{x-u}{\beta}\right)^{-1/\gamma}. \end{aligned}$$

Thus,

$$\mathbb{P}(X > x) = \xi_u \left(1 + \gamma \frac{x-u}{\beta}\right)^{-1/\gamma},$$

where $\xi_u = \mathbb{P}(X > u)$. Finally, the t-observation return level x_t is given by

$$x_t = u + \frac{\beta}{\gamma}\left((t\xi_u)^\gamma - 1\right). \tag{1.2}$$

The estimated return level $\hat{x}_t$ is then obtained by replacing the parameters β and γ by their estimations with $\hat{\xi}_u = \dfrac{N_u}{n}$ where N_u is the number of exceedances above u and n the number of observations (Coles (2001)). If $\gamma = 0$,

$$x_t = u + \beta \log\left(t\xi_u\right).$$

In application, we are generally interested in the T-year return level. In this case, if we have n_y observations per year the parameter t in (1.2) is equal to $T \times n_y$.

1.3 Point process approach

We expose in this section the behavior of extremes using a point process approach. This approach unifies the BM and threshold excesses models used to characterize the behavior of extremes and deals more naturally with non stationarity in threshold excesses due to the derived likelihood function (Coles (2001)).

1.3.1 Some basics of point process

By definition, a point process on a set $\mathcal{A}$ is a stochastic or random process for the occurrence and position of point events. When $\mathcal{A}$ represents a period of time, the point process is used to model the occurrence of a given event in this period. Let $N(A)$ denotes the set of non negative integer-valued random variables representing the number of points in $A \subset \mathcal{A}$. $N(A)$ characterizes statistically a point process whose intensity measure Λ is given by the expected number of points in any subset $A \subset \mathcal{A}$:

$$\Lambda(A) = \mathbb{E}(N(A)).$$

Assuming the existence of the derivative function of $\Lambda(A)$ for $A = [a_1, x_1] \times \ldots \times [a_k, x_k] \subset \mathbb{R}^k$ we obtain the intensity or density function of the process

$$\lambda(x) = \frac{\partial^k \Lambda(A)}{\partial x_1 \ldots \partial x_k}.$$

Homogeneous point process

A one-dimensional homogeneous point process on $A \subset \mathbb{R}$ with parameter $\lambda > 0$ is a process that satisfies

- $N(A) \sim \text{Poisson}(\lambda(t_2 - t_1))$, for all $A = [t_1, t_2] \subset \mathcal{A}$,

- $N(A)$ and $N(B)$ are independent random variables where A and B are non-overlapping subsets of $\mathcal{A}$.

Hence, the number of points in the interval $[t_1, t_2]$ follows a Poisson distribution whose mean is proportional to the interval length $(t_2 - t_1)$ and the number of points in separate intervals are mutually independent (Coles (2001)). Consequently, the set of points occurring at rate λ in a given period of time $[t_1, t_2]$ can be statistically modeled by a Poisson process with parameter λ, intensity measure $\Lambda([t_1, t_2]) = \lambda(t_2 - t_1)$ and density function $\lambda(t) = \lambda$.

Non-homogeneous Poisson process

A non-homogeneous Poisson process is a generalized version of the homogeneous type where the rate $\lambda(t)$ varies with time. Hence, the only difference with the homogeneous type is in the first characterization. More precisely, the non-homogeneous Poisson process satisfies

- $N(A) \sim \text{Poisson}(\Lambda(A))$, for all $A = [t_1, t_2] \subset \mathcal{A}$, with

$$\Lambda(A) = \int_{t_1}^{t_2} \lambda(t)dt.$$

- $N(A)$ and $N(B)$ are independent random variables where A and B are non-overlapping subsets of $\mathcal{A}$.

Its intensity measure and density function are $\Lambda(.)$ and $\lambda(.)$ respectively.
More generally, a k-dimensional non-homogeneous Poisson process with intensity measure $\lambda(.)$ on $\mathcal{A} \subset \mathbb{R}^k$ satisfies the property of independent non-overlapping random variables $N(A)$ and $N(B)$ and for all $A \subset \mathcal{A}$,

$$N(A) \sim \text{Poisson}(\Lambda(A)),$$

with

$$\Lambda(A) = \int_A \lambda(x)dx.$$

Definition 1.1. A sequence of point processes $N_1, N_2, \ldots$ on $\mathcal{A}$ converges in distribution to N, i.e

$$N_m \xrightarrow{d} N,$$

if the joint distribution of $(N_m(A_1), \ldots, N_m(A_\ell))$ converges in distribution to $(N(A_1), \ldots, N(A_\ell))$ for all ℓ and all bounded sets $A_1, \ldots, A_\ell$ such that $P\{N(\partial A_j) = 0\} = 1$, $j = 1, \ldots, \ell$ with ∂A the boundary of A (Coles (2001)).

1.3.2 Poisson process and extremes

Suppose again a sequence of i.i.d. r.v. $X_1, \ldots, X_m$ with distribution F and denote by $M_m = \max(X_1, \ldots, X_m)$. We have already seen that if there exist normalizing constants $a_m > 0$ and b_m such that

$$\lim_{m\to\infty} \mathbb{P}\left(\frac{M_m - b_m}{a_m} \leq x\right) = G(x),$$

then

$$G(x) = G_{\mu,\sigma,\gamma}(x) := \begin{cases} \exp\{-[1+\gamma\left(\frac{x-\mu}{\sigma}\right)]^{-1/\gamma}\}, & 1+\gamma\frac{x-\mu}{\sigma} > 0, \gamma \neq 0, \\ \exp\left(-\exp(-\frac{x-\mu}{\sigma})\right), & x \in \mathbb{R}, \gamma = 0. \end{cases}$$

Let

$$N_m = \left\{ \left(\frac{i}{m+1}\right), \left(\frac{X_i - b_m}{a_m}\right) : i = 1, \ldots, m \right\}$$

be the sequence of point processes defined on $\mathbb{R}^2$. The component $\left(\frac{i}{m+1}\right)$ can be interpreted as the time of occurrence defined on $(0,1)$ and the second component $\left(\frac{X_i - b_m}{a_m}\right)$ represents the asymptotic stabilized behavior of extremes (Coles (2001)). The asymptotic convergence of the sequence N_m is given in the following fundamental theorem (Coles (2001)).

Theorem 1.3. *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables such that $M_m = \max(X_1, \ldots, X_m)$. Suppose there exist sequences of constants $\{a_m > 0\}$ and $\{b_m\}$ such that*

$$\lim_{m\to\infty} \mathbb{P}\left(\frac{M_m - b_m}{a_m} \leq x\right) = G(x),$$

with

$$G(x) = \exp\{-[1+\gamma\left(\frac{x-\mu}{\sigma}\right)]^{-1/\gamma}\}$$

whose lower and upper endpoints are denoted by x_- and x_+ respectively i.e $G(x_-) = 0$ and $G(x_+) = 1$. Let

$$N_m = \left\{ \left(\frac{i}{m+1}\right), \left(\frac{X_i - b_m}{a_m}\right) : i = 1, \ldots, m \right\}$$

be the sequence of point processes defined on $\mathbb{R}^2$, then for $u > x_-$, N_m converges in distribution on $(0,1) \times [u, \infty)$ to a Poisson process with intensity measure

$$\Lambda(A) = (t_2 - t_1)\left[1 + \gamma\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\gamma}$$

on $A = [t_1, t_2] \times [x, x_+)$ with $[t_1, t_2] \subset (0,1)$.

Link with block maxima and threshold excesses models

Using Theorem 1.3, we can obtain the result stating that the BM follow approximately a GEV distribution as it follows. Suppose the sequence $X_1, X_2, \ldots$ of i.i.d. r.v. such that $M_m = \max(X_1, \ldots, X_m)$ and recall the sequence of point process

$$N_m = \left\{ \left(\frac{i}{m+1}\right), \left(\frac{X_i - b_m}{a_m}\right) : i = 1, \ldots, m \right\}.$$

If we consider the region $A = (0,1) \times (x, \infty)$ then the event $\{\frac{(M_m - b_m)}{a_m} \leq x\}$ corresponds to no points in the interval (x, ∞) and consequently in A. Hence, this event is equivalent to the event $\{N(A) = 0\}$ and consequently

$$\mathbb{P}\left(\frac{M_m - b_m}{a_m} \leq x\right) = \mathbb{P}\left(N_m(A) = 0\right).$$

It follows from Theorem 1.3 that

$$\begin{aligned}
\mathbb{P}\left(\frac{M_m - b_m}{a_m} \leq x\right) &\to \mathbb{P}\left(N(A) = 0\right) \\
&= \exp\{-\Lambda(A)\} \\
&= \exp\{-\left[1 + \gamma\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\gamma}\},
\end{aligned}$$

which means that the normalized maxima follow approximately a GEV distribution. As for the threshold model, the result stated in Theorem 1.2 can also be derived from Theorem 1.3. We first proceed to the factorization of $\Lambda(A_z)$ with $A_z = [t_1, t_2] \times (z, \infty)$ such that $[t_1, t_2] \subset (0,1)$:

$$\Lambda(A_z) = \Lambda_1([t_1, t_2]) \times \Lambda_2(z, \infty),$$

where $\Lambda_1([t_1, t_2]) = (t_2 - t_1)$ and $\Lambda_2(z, \infty) = \left[1 + \gamma\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\gamma}$. It follows that

$$\begin{aligned}
\mathbb{P}\left(\frac{X_i - b_m}{a_m} > z | \frac{X_i - b_m}{a_m} > u\right) &= \frac{\Lambda_2([z, \infty))}{\Lambda_2([u, \infty))} \\
&= \frac{\left[1 + \gamma\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\gamma}}{\left[1 + \gamma\left(\frac{u-\mu}{\sigma}\right)\right]^{-1/\gamma}} \\
&= \left[1 + \frac{\gamma(z-u)/\sigma}{1 + \gamma(u-\mu)/\sigma}\right]^{-1/\gamma} \\
&= \left[1 + \gamma\left(\frac{z-u}{\beta}\right)\right]^{-1/\gamma},
\end{aligned}$$

with $\beta = \sigma + \gamma(u - \mu)$. By setting $y = z - u$ we have

$$\mathbb{P}\left(\frac{X_i - b_m}{a_m} - u \leq y | \frac{X_i - b_m}{a_m} > u\right) = 1 - \left[1 + \gamma\left(\frac{y}{\beta}\right)\right]^{-1/\gamma}$$

which is the result obtained in Theorem 1.2 once the constants a_m and b_m are absorbed.

2

Permutation Bootstrap and the block maxima method

The chapter consists of an article submitted for publication to the journal *Communications in Statistics - Simulation and Computation.*

Abstract

We present a permutation bootstrap method for reducing the variance of estimation in the so-called block maxima (BM) method in extreme value theory. In the case of independent and identically distributed observations, it is sensible to use the permutation bootstrap to reduce the variance of the parameter and quantile estimators. The method is analyzed and we propose an implementation of the permutation bootstrap based on a particular sampling from the data based on the BM-ranks whose distribution is derived and easy to simulate. The performance of the method is discussed in a numerical study on simulated and then real data.

Contents

2.1 Introduction

Extreme value theory is widely used in many fields such as finance for the computation of value at risk (Singh et al. (2013), Klüppelberg and Zhang (2015)), hydrology for the protection against flood (Buishand et al. (2008), Butler et al. (2007)), meteorology for assessing extreme wind risks (Walshaw (2000), Coles and Pericchi (2003)). A fundamental problem is the estimation of extreme quantiles, beyond the observed data, so that empirical quantiles cannot be used. The estimation relies on an extrapolation of the tail of the distribution based on a regular variation assumption. In the so-called block maxima (BM) method, it is assumed that the BM are well approximated by a generalized extreme value (GEV) distribution. Estimating the GEV parameters allows for tail extrapolation and extreme quantile estimation. The confidence intervals associated to large quantiles are typically large and enhancing their accuracy is an important issue.

The problem of estimator accuracy is a common issue in statistics. Trying to deal with this problem, Efron (1979) introduced a technique called Bootstrap which consists in resampling the original data in order to assess the variability of the estimation procedures (Efron and Tibshirani (1986)). Bootstrap aggregating or bagging is commonly used to reduce the variance of an estimator (Kitagawa and Konishi (2010)). In extreme value theory though, bootstrap has to be applied carefully because only the extreme observations are important, bootstrapping heavy tailed phenomena is difficult and the bootstrap sample size must be chosen carefully (see Resnick (2007)).

We propose here a version of the bootstrap, called permutation bootstrap, that fits well with the BM method from univariate extreme value theory. The ultimate goal is to reduce the variance or mean square error (MSE) of estimators of the GEV parameters or of extreme quantiles. In Section 2.1, we introduce the framework of extreme value theory: we expose the BM and peaks over threshold (POT) methods as well as the maximum likelihood (ML) method for parameters and extreme quantile estimation. In Section 2.2, we discuss the permutation bootstrap in a general framework and then show that in the framework of the BM method, it is naturally associated with the BM-ranks that will be later introduced in Definition 2.3, analyzed and used to implement the permutation bootstrap. Finally in Section 2.3, we discuss in a numerical study the performance of the method and compare the MSE for various implementations of the permutation bootstrap. The effect of the permutation bootstrap sample size is also discussed in order to reduce computation time.

2.1.1 Basic notions on univariate extreme value theory

The BM and POT are fundamental methods in univariate extreme value theory. They are used to extract an extreme sample from a series of observations. The BM method consists in dividing the series of observations into blocks and keeping only the largest observation within each block. The POT method uses a high threshold and selects the observations that are above the threshold.

The BM method relies on the assumption that the maxima of independent and

identically distributed (i.i.d.) observations is well approximated by a GEV distribution. The choice of the GEV distribution is justified by the Gnedenko-Fisher-Tippett theorem (Gnedenko (1943); Fisher and Tippett (1928)): if $X_1, X_2, \ldots$ is a sequence of i.i.d. random variables (r.v.) with distribution function $F(x) = P(X \leq x)$, then the only possible limit in distribution for the rescaled maximum $X_{n,n} = \max(X_1, .., X_n)$ is a GEV distribution. More precisely, if

$$\lim_{n\to\infty} P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) = \lim_{n\to\infty} F^n(a_n x + b_n) := G(x)$$

exists and is non-degenerate (with $a_n > 0$ and b_n an existing normalizing constants), then G must be of one of the following types (see Embrechts et al. (1997)):

- Gumbel: $G(x) = \exp(-\exp(-x)) \quad \text{for all } x \in \mathbb{R}$;
- Fréchet: $G(x) = \begin{cases} \exp(-x^{-\alpha}) & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}, \ \alpha > 0$;
- Weibull: $G(x) = \begin{cases} \exp(-(-x)^{\alpha}) & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases}, \ \alpha > 0.$

From a statistical point of view, the parametrization due to Jenkinson (1969) of the three families into a single three parameters family is fundamental. The GEV distribution is defined as follows:

$$G_{\mu,\sigma,\gamma}(x) := \begin{cases} \exp\{-[1 + \gamma\left(\frac{x-\mu}{\sigma}\right)]^{-1/\gamma}\}, & 1 + \gamma\frac{x-\mu}{\sigma} > 0, \gamma \neq 0, \\ \exp\left(-\exp(-\frac{x-\mu}{\sigma})\right), & x \in \mathbb{R}, \gamma = 0, \end{cases}$$

where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ the scale parameter and $\gamma \in \mathbb{R}$ the extreme value index.

The POT method relies on the assumption that the exceedances over high threshold u are well approximated by a generalized Pareto distribution (GPD). The distribution function of the conditional exceedances $X - u | X > u$ is

$$F_u(y) := P(X - u \leq y | X > u) = \frac{F(u + y) - F(u)}{1 - F(u)}.$$

Balkema and de Haan (1974) and Pickands (1975) proved that if the rescaled exceedances converge in distribution to a non-degenerate limit, then the limit is a GPD defined by

$$F_{u,\sigma,\gamma}(y) := \begin{cases} 1 - \left(1 + \gamma\frac{y}{\sigma}\right)^{-1/\gamma} & \text{if } \gamma \neq 0 \\ 1 - \exp\left(-\frac{y}{\sigma}\right) & \text{if } \gamma = 0 \end{cases},$$

where $\sigma > 0$ is the scale parameter and $\gamma \in \mathbb{R}$ is the extreme value index.

It must be noticed here that the POT method is invariant under permutation of the observations. If π is a permutation of $\{1, \cdots, n\}$, the exceedances above threshold u in the original sample $X_1, \ldots, X_n$ and in the permuted sample $X_{\pi(1)}, \ldots, X_{\pi(n)}$

are the same. On the other hand, the BM method is not invariant under permutation of the observations. For instance, if the first two largest observations are in different blocks, they are both kept, while only the largest is kept if they are in the same block. That is why the permutation bootstrap can be useful in the BM method.

2.1.2 Estimation of the GEV parameters

Two main methods are used for estimating the parameters of a GEV distribution: ML and probability weighted moments (PWM) introduced in Greenwood et al. (1979). We apply in this paper only the ML method which is presented as follows. Assuming i.i.d. observations $Y_1, \ldots, Y_m$ from a GEV distribution, the log-likelihood function is given by (see Beirlant et al. (2004))

$$\ell(\mu, \sigma, \gamma) = -m \log \sigma - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^{m} \log\left(1 + \gamma \frac{Y_i - \mu}{\sigma}\right) - \sum_{i=1}^{m} \left(1 + \gamma \frac{Y_i - \mu}{\sigma}\right)^{-1/\gamma},$$

if $\gamma \neq 0$ and $1 + \gamma \frac{Y_i - \mu}{\sigma} > 0$. In the case $\gamma = 0$,

$$\ell(\mu, \sigma, 0) = -m \log \sigma - \sum_{i=1}^{m} \exp\left(-\frac{Y_i - \mu}{\sigma}\right) - \sum_{i=1}^{m} \frac{Y_i - \mu}{\sigma}.$$

The ML estimator $(\hat{\mu}, \hat{\sigma}, \hat{\gamma})$ solves the score equations:

$$\frac{\partial \ell(\mu, \sigma, \gamma)}{\partial \mu} = 0, \quad \frac{\partial \ell(\mu, \sigma, \gamma)}{\partial \sigma} = 0 \quad \text{and} \quad \frac{\partial \ell(\mu, \sigma, \gamma)}{\partial \gamma} = 0.$$

More computational details on the ML estimation of the GEV parameters date back to Prescott and Walden (1980, 1983). Many studies have been made on the existence and consistency of the estimators used in the ML method for the BM. Smith (1985) proved the asymptotic normality of the ML estimations for $\gamma > -\frac{1}{2}$ and consistency for $\gamma > -1$. Moreover, Zhou (2009) proved the existence of the ML estimator for $\gamma > -1$ and Dombry and Ferreira (2017) provided asymptotic theory for the ML estimators based on the BM method. One important advantage of the ML method is that it can be extended to more complicated models by doing small changes in the basic method (see chapter 6, Embrechts et al. (1997)).

2.1.3 Quantile estimation

Extreme quantile estimation is central in extreme value theory. Based on the BM method, extreme quantiles can be estimated as follows. Given a sample $X_1, ..., X_n$ of i.i.d. observations with distribution function F, one divides the sample into k blocks of size m, $n = k \times m$, and compute BM $M_1, \ldots, M_k$. The distribution of the M_i is approximatively GEV and its parameters μ, σ, γ are estimated by ML or PWM methods. The quantile $q_{M,p}$ of order p of M is given by

$$q_{M,p} = G^{\leftarrow}_{\mu,\sigma,\gamma}(p), \quad p \in (0, 1).$$

The relation $G = F^m$ implies that the quantile of order p of X corresponds to the quantile of order p^m of M. Hence an estimator of $q_{X,p}$ is given by the following formula where μ, σ and γ have to be replaced by their estimated values (see Beirlant et al. (2004)):

$$q_{X,p} = q_{M,p^m} = \begin{cases} \mu + \frac{\sigma}{\gamma}[(-m \log p)^{-\gamma} - 1] & \text{if } \gamma \neq 0 \\ \mu - \sigma \log(-m \log p) & \text{if } \gamma = 0 \end{cases} . \tag{2.1}$$

Recall that m is the block size in the BM method.

2.2 Permutation Bootstrap

2.2.1 Preliminaries

The permutation bootstrap is a general statistical methodology to reduce the variance of estimators based on i.i.d. observations, or more generally on exchangeable observations. We denote by S_n the set of permutations $\pi = (\pi_1, .., \pi_n)$ of $\{1, .., n\}$.

Definition 2.1. A random vector $X = (X_1, \ldots, X_n)$ is called *exchangeable* if, for all permutations $\pi \in S_n$, $X_\pi = (X_{\pi_1}, .., X_{\pi_n})$ has the same distribution as X.

Obviously a random vector $X = (X_1, \ldots, X_n)$ with i.i.d. components is exchangeable. The idea of the permutation bootstrap is to exploit the invariance of the distribution of X under permutation of its components to reduce the variance of statistics based on X.

Definition 2.2. Let $X = (X_1, .., X_n)$ be a random vector and $\hat{\theta} = \hat{\theta}(X)$ be a statistic based on X. The permutation bootstrapped statistic $\hat{\theta}^{PB}$ is defined by

$$\hat{\theta}^{PB} = \frac{1}{n!} \sum_{\pi \in S_n} \hat{\theta}(X_\pi), \tag{2.2}$$

where $X_\pi = (X_{\pi_1}, .., X_{\pi_n})$.

The next proposition shows the effect of the permutation bootstrap on the mean and variance of the estimator $\hat{\theta}$.

Proposition 2.1. *Assume that the random vector $(X_1, .., X_n)$ is exchangeable and that the statistic $\hat{\theta} = \hat{\theta}(X)$ has finite variance. Then*

$$\mathbb{E}(\hat{\theta}^{PB}) = \mathbb{E}(\hat{\theta}) \tag{2.3}$$

and

$$\mathrm{Var}(\hat{\theta}^{PB}) \leq \mathrm{Var}(\hat{\theta}). \tag{2.4}$$

Furthermore, equality holds in (2.4) if and only if $\hat{\theta}$ is permutation invariant, i.e

$$\hat{\theta}(X_\pi) = \hat{\theta}(X) \quad \text{a.s. for all } \pi \in S_n.$$

Proof. Equation (2.3) is a simple consequence of the linearity of expectation and exchangeability of X:

$$\begin{aligned}\mathbb{E}(\hat{\theta}^{PB}) &= \frac{1}{n!}\sum_{\pi\in S_n}\mathbb{E}(\hat{\theta}(X_\pi)),\\ &= \frac{n!}{n!}\mathbb{E}(\hat{\theta}(X)) = \mathbb{E}(\hat{\theta}).\end{aligned}$$

It follows that the variance of the permutation bootstrapped statistics is equal to

$$\mathrm{Var}(\hat{\theta}^{PB}) = \mathbb{E}\left[\left(\frac{1}{n!}\sum_{\pi\in S_n}\Big(\hat{\theta}(X_\pi) - \mathbb{E}(\hat{\theta})\Big)\right)^2\right].$$

Cauchy-Schwartz inequality implies

$$\left(\frac{1}{n!}\sum_{\pi\in S_n}\Big(\hat{\theta}(X_\pi) - \mathbb{E}(\hat{\theta})\Big)\right)^2 \le \frac{1}{n!}\sum_{\pi\in S_n}(\hat{\theta}(X_\pi) - \mathbb{E}(\hat{\theta}))^2$$

with equality if and only if $\hat{\theta}(X_\pi)$ does not depend on $\pi \in S_n$. Taking the expectation and using the exchangeability of X, we obtain

$$\begin{aligned}\mathrm{Var}(\hat{\theta}^{PB}) &\le \frac{1}{n!}\sum_{\pi\in S_n}\mathbb{E}\left[\Big(\hat{\theta}(X_\pi) - \mathbb{E}(\hat{\theta})\Big)^2\right]\\ &= \frac{n!}{n!}\mathbb{E}\left[\Big(\hat{\theta}(X) - \mathbb{E}(\hat{\theta})\Big)^2\right] = \mathrm{Var}(\hat{\theta})\end{aligned}$$

with equality if and only if $\hat{\theta}(X_\pi) = \hat{\theta}(X)$ a.s. for all $\pi \in S_n$. □

Remark 2.1. The permutation bootstrap reduces the variance and thus the estimation error while maintaining the mean constant so that it has no effect on the bias.

Remark 2.2. In practice, it is not possible to compute $\hat{\theta}^{PB}$ exactly because the cardinality $n!$ of S_n explodes. Instead, one uses the Monte-Carlo estimation $\hat{\theta}^B$ based on B independent random permutations $\pi^{(1)}, .., \pi^{(B)}$ with uniform distribution on S_n and defined by

$$\hat{\theta}^B = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}(X_{\pi^{(b)}}).$$

The choice of B will be discussed later in the numerical part in section 2.3 .

2.2.2 Permutation bootstrap and block maxima: a rank method

For applying the BM method, we consider a sample $X_1, \ldots, X_n$ of size n divided into k blocks of size m each and consider the BM

$$M_j = \max_{(j-1)m<i\le jm} X_i \quad j = 1, .., k.$$

We suppose that $n = k \times m + r$, with $r \in \{0, \ldots, m-1\}$ and note that when $r \geq 1$, the last r observations are not included in the BM. This can be done since the observations ignored at a given permutation can be included in the BM at a different permutation whence we will not have a loss of information. Clearly, the order of the observations has a strong impact on the BM which are not invariant under permutation of the observations $X_1, \ldots, X_n$.

A statistic $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ built on BM, i.e. of the form $\tilde{\theta} = \tilde{\theta}(M_1, \ldots, M_k)$ is generally not permutation invariant and the permutation bootstrap can be used in order to reduce its variance. Since repeatedly permuting the random vector X, forming blocks and computing BM is expensive from a numerical point of view, we propose a method based on ranks and order statistics that directly selects a sample of the order statistics with suitable rank distribution. This results in the following definition of the BM ranks that can be used for a more efficient implementation of the permutation bootstrap that avoids permutation and BM computation. The order statistics associated to $X_1, \ldots, X_n$ are denoted by $X_{1:n} \leq \cdots \leq X_{n:n}$. Similarly, the order statistics for the BM $M_1, \ldots, M_k$ are $M_{1:k} \leq \cdots \leq M_{k:k}$.

Definition 2.3. The BM-rank, $(R_1, \ldots, R_k)$, is defined by the relations

$$M_{k+1-j:k} = X_{R_j:n}, \quad j = 1, \ldots, k.$$

Note that $n \geq R_1 > \ldots > R_k \geq 1$ and that $X_{R_1:n}$ is the maximum within the k blocks, i.e. $\max_{1 \leq i \leq km} X_i$. In particular, if $n = k \times m$, $\max_{1 \leq i \leq n} X_i = X_{n:n}$, so that $R_1 = n$.

Example 2.1. We illustrate the definition with a simple example. Suppose $n = 10$, $k = m = 3$ and the observations are

$$1.18 \quad 2.05 \quad 4.31 \quad 1.13 \quad 1.64 \quad 3.03 \quad 1.19 \quad 2.55 \quad 9.40 \quad 1.87.$$

We compute the ranks and the BM

$$\underbrace{2 \quad 6 \quad 9}_{M_1 = X_{9:10}} \quad \underbrace{1 \quad 4 \quad 8}_{M_2 = X_{8:10}} \quad \underbrace{3 \quad 7 \quad 10}_{M_3 = X_{10:10}} \quad 5.$$

It follows that the BM-ranks are $R_1 = 10$, $R_2 = 9$ and $R_3 = 8$. For the sorted sample

$$1.13 \quad 1.17 \quad 1.19 \quad 1.64 \quad 1.87 \quad 2.05 \quad 2.55 \quad 3.03 \quad 4.31 \quad 9.40,$$

we obtain similarly the ranks and BM

$$\underbrace{1 \quad 2 \quad 3}_{M_1 = X_{3:10}} \quad \underbrace{4 \quad 5 \quad 6}_{M_2 = X_{6:10}} \quad \underbrace{7 \quad 8 \quad 9}_{M_3 = X_{9:10}} \quad 10$$

and the BM-ranks are now $R_1 = 9$, $R_2 = 6$ and $R_3 = 3$.

Our goal is to compute the distribution of the BM-ranks and to explain how this relates to the permutation bootstrap and the BM method.

Definition 2.4. Let $1 \leq n_1 < n_2$ be integers and define $\mathcal{G}(n_1, n_2)$ as the distribution of a random variable N such that

$$\mathbb{P}(N = i) = \frac{\binom{n_2-n_1}{i}}{\binom{n_2}{i}} - \frac{\binom{n_2-n_1}{i+1}}{\binom{n_2}{i+1}}, \quad i = 0, \ldots, n_2 - n_1.$$

In the case $n_1 = n_2$, we define $\mathcal{G}(n_1, n_2)$ as the dirac mass at 0 so that $\mathbb{P}(N = 0) = 1$.

Remark 2.3. In the case $1 \leq n_1 < n_2$, the distribution $\mathcal{G}(n_1, n_2)$ appears naturally in the following urn problem. Consider an urn that contains n_1 white balls and $n_2 - n_1$ black balls from which we extract balls sequentially and without replacement. A black ball is considered as a failure and a white ball as a success. Define the random variable N as the number of failures until the first success. For $i = 0, \ldots, n_2 - n_1$, the event $\{N \geq i\}$ corresponds to the situation where the first i balls are black so that $\mathbb{P}(N \geq i) = \binom{n_2-n_1}{i} / \binom{n_2}{i}$. The formula in the definition follows from the relation $\mathbb{P}(N = i) = \mathbb{P}(N \geq i) - \mathbb{P}(N \geq i + 1)$.

In the following proposition, we express the distribution of the BM-ranks $(R_1, \ldots, R_k)$ in terms of the distribution $\mathcal{G}(n_1, n_2)$.

Proposition 2.2. *Let $n = k \times m + r$, $r \in \{0, m - 1\}$. For an exchangeable random vector $X = (X_1, .., X_n)$, the distribution of the BM ranks $(R_1, .., R_k)$ is given by the relations:*

$$\begin{cases} \mathcal{L}(n - R_1) = \mathcal{G}(k \times m, n) \\ \mathcal{L}\left(R_{j-1} - 1 - R_j | R_1, .., R_{j-1}\right) = \mathcal{G}\left((k + 1 - j) \times m, R_{j-1} - 1\right), \quad j = 2, .., k. \end{cases}$$

Proof. The relative ranks of the vector $X = (X_1, \ldots, X_n)$ define a random permutation $\pi = (\pi_1, \ldots, \pi_n) \in S_n$. Since X is exchangeable, the distribution of the ranks π is uniform on S_n. One can think that the random permutation $\pi = (\pi_1, \ldots, \pi_n)$ is obtained by first choosing the place of the integer n randomly among n places, then the place of the integer $n - 1$ randomly among $n - 1$ remaining places and so on. In the context of Example 2.1, think that the relative ranks are random and that one places the integers $10, 9, 8 \ldots$ in the ten places. The BM rank R_1 corresponds to the largest relative rank in the $k \times m$ first places among n. One can see that $n - R_1$ corresponds to the number of places chosen in the last r places (failures) until one of the first $k \times m$ places is chosen (success), whence the distribution of $n - R_1$ is $\mathcal{G}(k \times m, n)$ (see Remark 2.3 above). Then, for the second BM rank R_2, one proceeds with choosing a place for $R_1 - 1, R_1 - 2, \ldots$ until a new block is met, that is one of the $(k - 1) \times m$ places within new blocks is chosen among the $R_1 - 1$ remaining places. We deduce that, conditionally on R_1, the distribution of $R_1 - 1 - R_2$ is $\mathcal{G}((k - 1) \times m, R_1 - 1)$. This reasoning is easily generalized so as to prove that conditionally on $R_1, \ldots, R_{j-1}$, the distribution of $R_{j-1} - 1 - R_j$ is $\mathcal{G}((k + 1 - j) \times m, R_{j-1} - 1)$. □

The following proposition relates the BM rank and the permutation bootstrap in the framework of the BM method.

Proposition 2.3. *Consider a statistic* $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ *built on BM,*

$$\hat{\theta} = \tilde{\theta}(M_1, \ldots, M_k).$$

Assume that $\tilde{\theta}$ *is invariant by permutation, i.e*

$$\tilde{\theta}(M_1, \ldots, M_k) = \tilde{\theta}(M_{\pi_1}, \ldots, M_{\pi_k}), \ \ \pi \in S_k.$$

Then the permutation bootstrap estimator $\hat{\theta}^{PB}$ *is equal to*

$$\hat{\theta}^{PB} = \sum_{n \geq r_1 > \cdots > r_k \geq 1} \mathbb{P}(R_1 = r_1, \ldots, R_k = r_k)\, \tilde{\theta}(X_{r_1:n}, \ldots, X_{r_k:n}). \tag{2.5}$$

Remark 2.4. Note that the sum (2.5) involves $\binom{n}{k}$ terms which is of smaller order than $n!$, the number of terms in the sum (2.2). Similarly as in Remark 2.2 , we use in practice a Monte Carlo estimation of the permutation bootstrap based on BM-ranks. Using B independent copies of the BM-ranks denoted by $(R_1^b, \ldots, R_k^b)_{1 \leq b \leq B}$, we consider

$$\hat{\theta}^B = \frac{1}{B} \sum_{b=1}^{B} \tilde{\theta}\left(X_{R_1^b:n}, \ldots, X_{R_k^b:n}\right).$$

Proof. Recall that the permutation bootstrapped statistics is defined by

$$\hat{\theta}^{PB} = \frac{1}{n!} \sum_{\pi \in S_n} \hat{\theta}(X_\pi).$$

We can interpret $\hat{\theta}^{PB}$ as the expectation of $\hat{\theta}(X_\pi)$ where X is fixed and π is a random permutation with uniform distribution on S_n. Then the random variable X_π (with randomness on π) is exchangeable and Proposition 2.2 provides the distribution of the BM-rank of X_π. Furthermore, using the facts that $\hat{\theta}$ is built on BM and that $\tilde{\theta}$ is invariant by permutation, we have

$$\hat{\theta}(X_\pi) = \tilde{\theta}(M_1^\pi, \ldots, M_k^\pi) = \tilde{\theta}(M_{k:k}^\pi, \ldots, M_{1:k}^\pi) = \tilde{\theta}(X_{R_1^\pi:n}, \ldots, X_{R_k^\pi:n}).$$

We use the superscript π to recall that the randomness is now on π. Finally, noting $\mathbb{E}_\pi$ the expectation with respect to π, we get

$$\begin{aligned}
\hat{\theta}^{PB} &= \mathbb{E}_\pi[\hat{\theta}(X_\pi)] \\
&= \mathbb{E}_\pi[\tilde{\theta}(X_{R_1^\pi:n}, \ldots, X_{R_k^\pi:n})] \\
&= \sum_{n \geq r_1 > \cdots > r_k \geq 1} \mathbb{P}(R_1 = r_1, \ldots, R_k = r_k)\, \tilde{\theta}(X_{r_1:n}, \ldots, X_{r_k:n})
\end{aligned}$$

where the distribution of the BM-ranks $R_1, \ldots, R_k$ is given by Proposition 2.2 . □

2.2.3 Median permutation bootstrap

The effect of permuting observations in the BM method is illustrated in Figure 2.1. For a fixed sample of $n = 1000$ observations following a Pareto distribution with $\gamma = 1/2$, we compute different estimators of the extreme value index and of

an extreme quantile. The estimators are obtained by letting the block size vary ($m = 1, \ldots, 30$), and for each block size by resampling the order of observations using $B = 100$ random permutations. For different block sizes m, a boxplot represents the variability of the estimator with respect to the observation order. As usual for boxplot, the box represents the first and third quartiles, the bold line within the box represents the median and the mean of the estimations is presented by a triangle. The figure evidences a large variability of the estimators, both with respect to the block size and the observation order. Furthermore, we can note the presence of outliers, that lie far away from the box. This suggests that the median can be a better estimate than the mean, because it is more robust to outliers. Indeed, we can see that the permutation bootstrap median is closer to the target value (horizontal line) than the permutation bootstrap mean. In the next section, we will assess the performance of both the permutation bootstrap mean and median. Note however, that we have no guarantee that the median based permutation bootstrap reduces the variance of estimation, as it is the case for the mean-based permutation bootstrap (see Proposition 2.1).
On the other hand, the variance reduction is not the only advantage of the permutation bootstrap. Figure 2.1 suggests that some (unlikely) orderings of the observations yield outliers for the γ estimates. The permutation bootstrap provides a way to get rid of these outliers and results in more robust estimates.

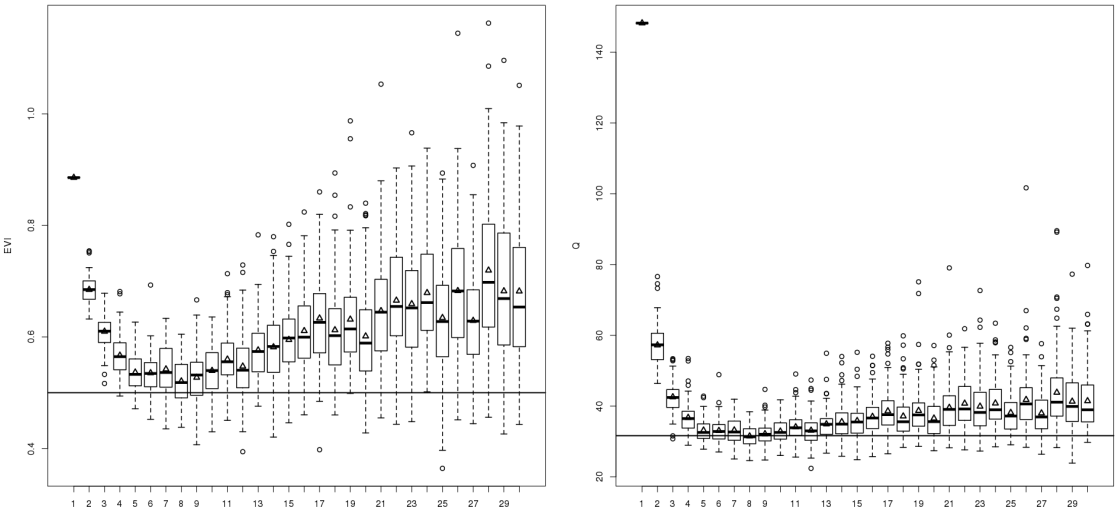


Figure 2.1 – Boxplot (with respect to the observation order) for $\hat{\gamma}$ (left) and $\hat{q}_{X,p}$ (right) in function of the block size.

2.2.4 Simulation algorithms

We provide here simple algorithms in order to obtain simulation for the distribution $\mathcal{G}(n_1, n_2)$ and the distribution of the BM-ranks $(R_1, \ldots, R_k)$. The first function called SIMULATIONG(n_1, n_2) is explained in the context of the urn problem (see Remark 2.3).

function SIMULATIONG(n_1, n_2)

initial state: $\begin{cases} n_b = n_2 - n_1 & \text{(number of black balls in the urn)} \\ n_f = 0 & \text{(number of failures)} \\ s = F & \text{(state is either failure F or success S)} \end{cases}$

  while $s = F$ **do**
    sample $U \sim \text{Unif}([0, 1])$
    if $U \leq \frac{n_b}{n_1+n_b}$ (probability of drawing a black ball) **then**
      $n_b = n_b - 1$
      $n_f = n_f + 1$
    else
      $s = S$
    end if
  end while
  return n_f
end function

To simulate a random vector of BM-ranks associated with n observations divided into blocks of size m each, we use the following function RANK(n, m). We use a R-like syntax for vectors.

function RANK(n, m)
  set (k, r) the quotient and the remainder in the division $n = m \times k + r$
  set $R[1] = m \times k$, $p = m - 1$ and $q = k \times m - m$
  for i=2 to k **do**
    $Y =$ SIMULATIONG$(q, p + q)$
    $R[i] = R[i - 1] - Y$, $q = k \times m - i \times m$, $p = R[i] - 1 - q$
  end for
  if $r = 0$ **then**
    return R
  else
    set I a random sample of size r from $\{1, \ldots, n\}$ (without replacement)
    $V = (1 : n)[-I]$ (V is the vector of numbers 1 to n without numbers in I)
    $R = V[R]$
  end if
end function

2.3 Numerical analysis

2.3.1 Experimental design

We perform in this section a simulation study. We generate a sample of size $n = 1000$ of r.v. $X_1, .., X_n$ with a Pareto distribution, that is

$$F(x) = 1 - \left(\frac{1}{x}\right)^{1/\gamma}, \quad x > 1,$$

with $\gamma = 0.2, 0.5, 0.8$. We then apply the BM method by considering k blocks of size m each and extract the maximum from each block based on the BM-ranks using the algorithms presented in Section 2.2.4. First, we estimate the GEV parameters and the quantile $q_{X,p}$ of order $p = 1 - \frac{1}{n}$ using the ML method since this method ensures consistency and asymptotic normality of the estimator when $\gamma > -1$ and $\gamma > -\frac{1}{2}$ respectively. In a second step, we apply the permutation bootstrap with $B = 100$ permutations (using both method based on the mean and the median) to estimate the parameters and the quantile. Note that the quantile can be estimated in two ways using the Bootstrap:

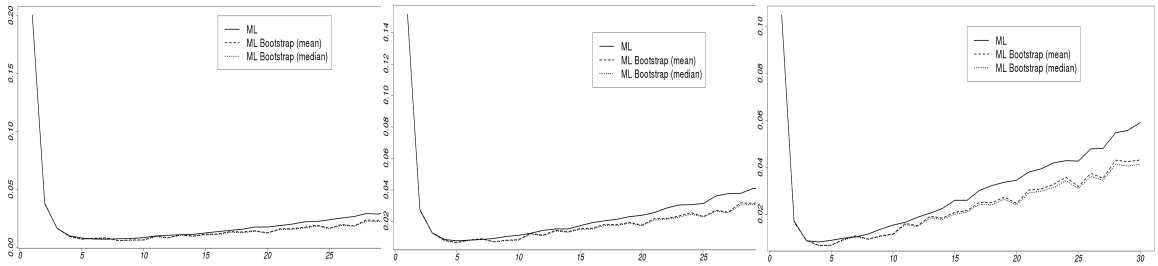
I. We estimate the parameters μ, σ and γ by the permutation bootstrap and then replace their values directly in (2.1);

II. We calculate the quantile (2.1) for each permutation and then apply the permutation bootstrap on the different estimations of the quantile.

We obtain estimates of γ and $q_{X,p}$ with or without bootstrap, the bootstrap can be performed with mean or median, and in the case of quantile estimation, methods I and II can be used. This leads us to 3 estimators of the extreme value index γ and 5 estimators of the extreme quantile $q_{X,p}$. Based on $N' = 1000$ repetitions of the experiment, we compute the MSE of the different estimators when the block size $m = 1, \cdots, 30$ varies.

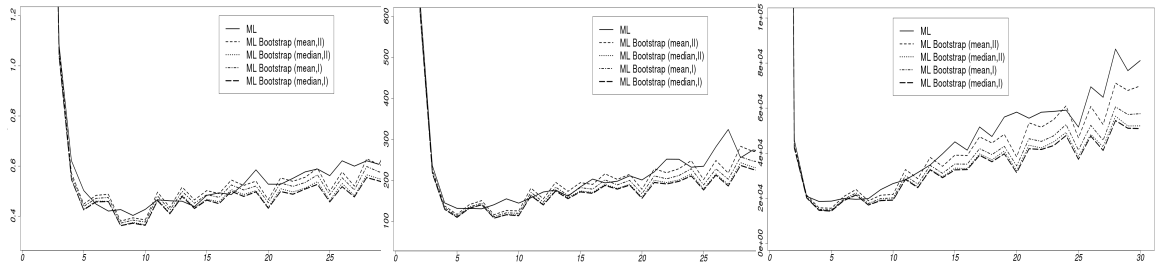
2.3.2 The performance of the permutation bootstrap

The MSE of $\hat{\gamma}$ and $\hat{q}_{X,p}$ as a function of the block size m are presented respectively in Figures 2.2 and 2.3 below.



(a) $\gamma = 0.2$ (b) $\gamma = 0.5$ (c) $\gamma = 0.8$

Figure 2.2 – Mean squarred error for $\hat{\gamma}$ as a function of the block size m.



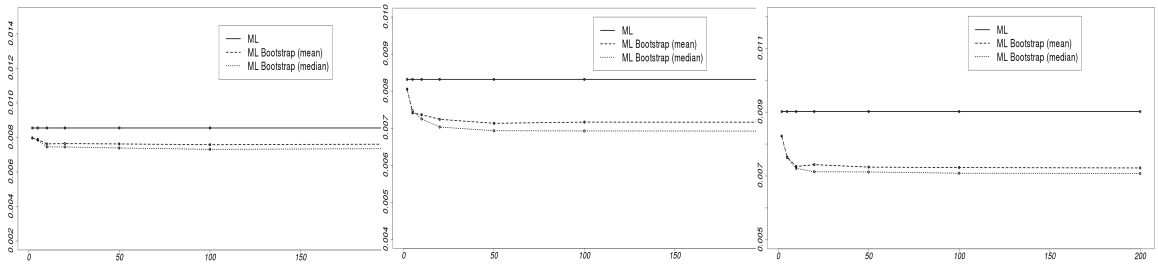
(a) $\gamma = 0.2$ (b) $\gamma = 0.5$ (c) $\gamma = 0.8$

Figure 2.3 – Mean squarred error for $\hat{q}_{X,p}$ as a function of the block size m.

In both figures the MSE is high when $m = 1$, because with this block size the bias is very high and the variance is low leading to a large MSE. The MSE decreases strongly when $m \simeq 5$ and then continues to fluctuate. For $\hat{\gamma}$, we can observe in Figure 2.2 a reduction of the MSE when using the permutation bootstrap. This reduction is modest, and more effective for larger block size and larger γ. The mean- and median-based bootstrap are essentially equivalent, the median bootstrap being slightly better. As for $\hat{q}_{X,p}$, we observe in Figure 2.3 that bootstrap reduces the MSE and the best combination is often (median,I), that is median-bootstrap performed on the GEV parameters before computing the quantile. The reduction of the MSE becomes more remarkable as γ increases for both estimators.

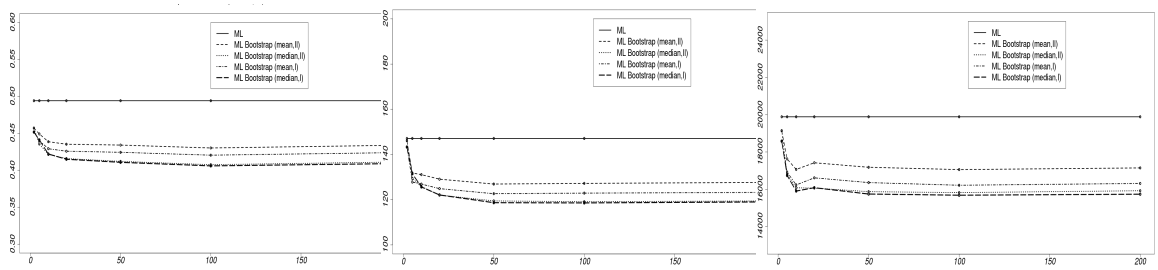
2.3.3 Discussion on the choice of B in the bootstrap

We consider the effect of the choice of B, the number of random permutation used in the bootstrap. We propose to fix the block size (here we choose $m = 5$ which shows almost the lowest MSE) and consider different values of $B = 2, 5, 10, 20, 50, 100, 200$. The results are shown in Figures 2.4 , 2.5 below. Here, the reduction of the MSE is clear. We also remark that the median-bootstrap is better than the mean-bootstrap with no preferences between method I and II for $\hat{q}_{X,p}$. Moreover, we can see that the MSE decreases from $B = 2$ to $B = 20$, after that the MSE is slightly reduced or almost constant.



(a) $\gamma = 0.2$ (b) $\gamma = 0.5$ (c) $\gamma = 0.8$

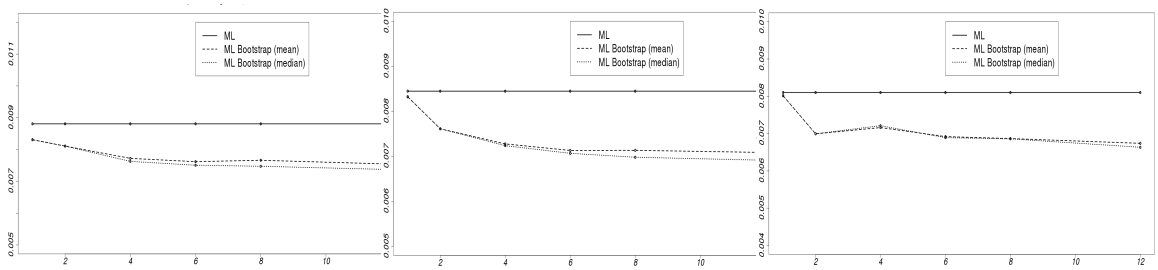
Figure 2.4 – Mean squarred error for $\hat{\gamma}$ as a function of B.



(a) $\gamma = 0.2$ (b) $\gamma = 0.5$ (c) $\gamma = 0.8$

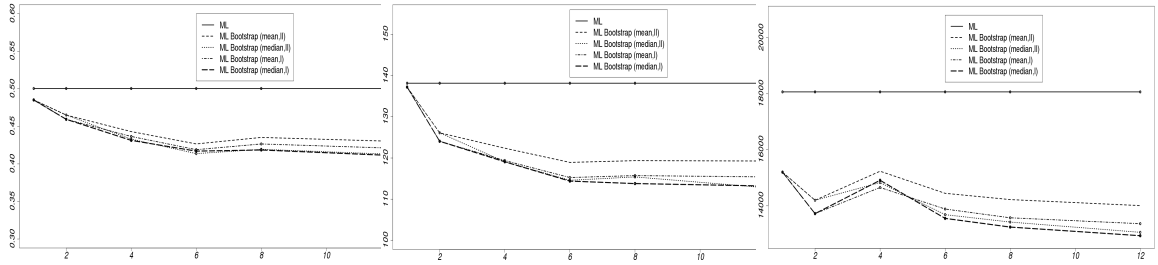
Figure 2.5 – Mean squarred error for $\hat{q}_{X,p}$ as a function of B.

This suggests that using small values of B can reduce the computational cost while still providing an effective MSE reduction. In Figure 2.6 and 2.7 below, smaller values of B are tested ($B = 1, 2, 4, 6, 8, 12$). We observe that a choice of $B = 6$ is acceptable since the MSE decreases insignificantly after that.



(a) $\gamma = 0.2$ (b) $\gamma = 0.5$ (c) $\gamma = 0.8$

Figure 2.6 – Mean squarred error for $\hat{\gamma}$ as a function of B.



(a) $\gamma = 0.2$ (b) $\gamma = 0.5$ (c) $\gamma = 0.8$

Figure 2.7 – Mean squarred error for $\hat{q}_{X,p}$ as a function of B.

2.4 Application

In order to apply the study on real data, we use the one of integer valued daily maximum temperature (degrees Fahrenheit) from Fort Collins, Colorado, USA on the period from 1900 to 1999 provided by the **extRemes** package of R. As in the simulations, we divide the data into blocks of size $m = 365$ (one year) and get our BM-ranks using the algorithms of Section 2.2.4. Then we estimate the extremal index γ

and the quantile $q_{X,p}$ of order $p = 1 - \frac{1}{m \times 100}$ (return level of 100 years) using the best combination found in the simulation study i.e median-bootstrap and method I for the quantile estimation with B taking values in $\{1, 2, 3, 4, 5, 10, 50, 100, 500, 1000\}$. The aim is to repeat the same procedure $N' = 100$ times in order to check how the choice of B affects the variance reduction. We present in Figure 2.8 the boxplot of $\hat{\gamma}$ and $\hat{q}_{X,p}$ in function of B in order to have graphical interpretations.

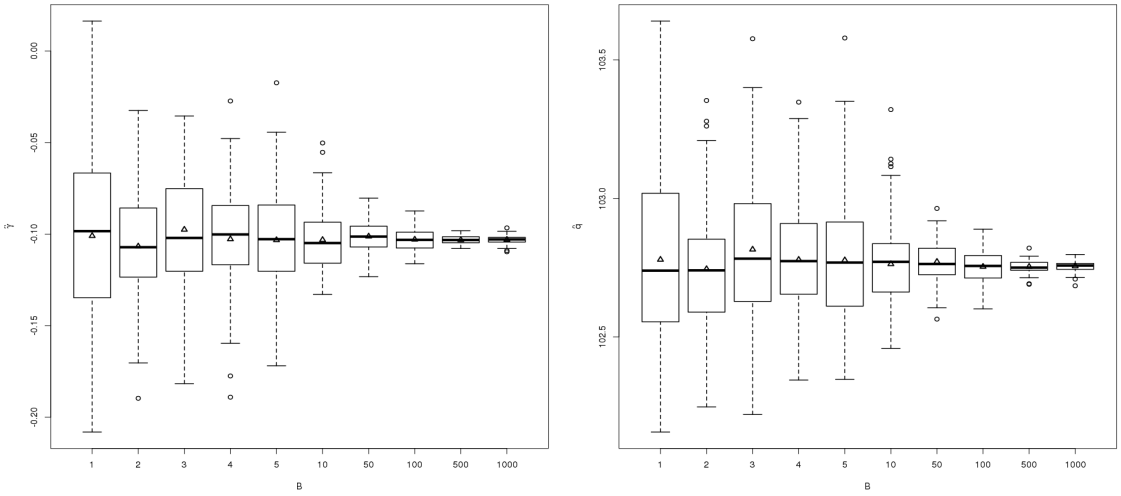


Figure 2.8 – Boxplot (with respect to the observation order) for $\hat{\gamma}$ (left) and $\hat{q}_{X,p}$ (right) in function of B.

One can first note here, that in the BM method, the order of the data influences significantly the value of the estimator. For some specific order of the data, the estimator can be far away from the true value. The use of the bootstrap procedure is a solution to face this problem. We remark also how the variance of the estimations is reduced with the increase of B for $\hat{\gamma}$ and $\hat{q}_{X,p}$ especially for large values of B: we can see through the boxplots of $B = 500$ and $B = 1000$ how the variance is very small especially with $B = 1000$ where the estimations become very close to each other. Here we can conclude that in application, one can get a very accurate estimation if he applies the Bootstrap method with large B. On the other hand, if it costs him long computation time, he can proceed to the use of smaller values of B which ensure a less accurate but nevertheless a very good estimation.

2.5 Conclusion

In extreme value theory, the use of the Bootstrap technique based on ranks and order statistics helps us save time in the algorithm and reduces the MSE of the estimations. The median Bootstrap method seems to be better than the mean Bootstrap. Increasing the number of resampled data B doesn't necessarily guarantee a more important reduction of the MSE. This is why detecting the optimal B allows us to make a trade-off between the fastest algorithm and the best estimation. Moreover, the permutation bootstrap provides a way to deal with unlucky ordering of observations and results in more robust estimates.

3

Trend detection for heteroscedastic extremes

The chapter consists of an extended version of an article submitted for publication to the journal *Scandinavian Journal of Statistics*.

Abstract

Extreme weather events are known to be more frequent due to global warming. From a statistical point of view, this raises the question of trend detection in the extremes of a series of observations. We build upon the heteroscedastic extremes framework by Einmahl et al. (2016) where the observations are assumed independent but not identically distributed and the variation in their tail distributions is modeled by the so-called skedasis function. While the original paper focuses on non parametric estimation of the skedasis function, we consider here parametric models and prove the consistency and asymptotic normality of the parameter estimators. A parametric test for trend detection in the case where the skedasis function is monotone is introduced. A short simulation study shows that the parametric test can be more powerful than the non parametric Kolmogorov-Smirnov type test, even for misspecified models. We also discuss the choice of threshold based on Lepski's method. The methodology is finally illustrated on a dataset of minimal/maximal daily temperatures in Fort Collins, Colorado, during the 20th century.

Contents

3.1 Heteroscedastic extremes

In the framework of global warming, there are some suggestions that extreme weather events (e.g. heat waves, flood, hurricanes) are becoming more frequent and with larger magnitude (Tank and Können (2003), Groisman et al. (2005), Zolina et al. (2009)). This paper addresses the question of detecting a temporal trend for such extreme events.

Classical extreme value theory is based on independent and identically distributed (i.i.d.) random variables. However, in many practical situations, the variables may deviate from this assumption by being either dependent or non identically distributed. A nice treatment of extreme value theory for dependent observations can be found in the monograph by Leadbetter et al. (1983). More recent contributions in the domain of dependent extremes include Hsing (1991) where tail index estimation is considered and Drees and Rootzén (2010) where empirical limit theory for clusters of extremes in stationary sequences is discussed.

Despite its importance in practice, theoretical results for independent but non identically distributed extremes are quite sparse. Davison and Smith (1990) describe models for exceedances over high thresholds where they impose a linear trend on the shape and scale parameters of the generalized Pareto distributions (GPD). Coles (2001) considers a log-linear trend for the scale parameter of the GPD. de Haan et al. (2015) consider tail trend detection in a discrete time model with groups of observations at several different times. Recently, Einmahl et al. (2016) have introduced a theoretical framework called "heteroscedastic extremes" where the distribution of the observations evolves in continuous time in such a way that a proportional tail condition is satisfied. They propose a non parametric estimation of the so-called skedasis function as well as for the extreme value index and discuss their asymptotic properties.

Since our work elaborates on this heteroscedastic extremes framework, we first recall more precisely the model and results from Einmahl et al. (2016). We consider a triangular array of independent observations $X_1^n, \ldots, X_n^n$, where X_i^n has a continuous distribution function $F_{n,i}$, $n \geq 1$, $1 \leq i \leq n$. We assume all the distributions share the same right endpoint

$$x^\star = \sup\{x : F_{n,i}(x) < 1\} \in (-\infty, \infty]$$

and satisfy the proportional tail condition

$$\lim_{x \to x^\star} \frac{1 - F_{n,i}(x)}{1 - F(x)} = c(i/n), \quad \text{uniformly in } n \geq 1,\ 1 \leq i \leq n, \tag{3.1}$$

where F is a baseline continuous distribution function with right endpoint $x^\star$ and the *skedasis function* $c : [0,1] \to (0, \infty)$ is a continuous function characterizing the evolution of extremes over time. The normalization condition

$$\int_0^1 c(s)\, ds = 1 \tag{3.2}$$

ensures uniqueness of the skedasis function c. The case $c \equiv 1$ is called *homoscedastic extremes*. To develop a theory for extremes, the baseline distribution F is supposed

to belong to the domain of attraction of a generalized extreme value (GEV) distribution with extreme value index $\gamma > 0$ (only positive extreme value index are considered in Einmahl et al. (2016), mainly for simplicity). Note that the proportional tail condition (3.1) implies that all the $F_{n,i}$ belong to the same domain of attraction and have the same extreme value index γ.

In this framework, Einmahl et al. (2016) propose to estimate the integrated skedasis function $C(s) = \int_0^s c(t)\,dt$, $s \in [0,1]$ by

$$\hat{C}(s) = \frac{1}{k} \sum_{i=1}^{[ns]} \mathbf{1}_{\{X_i^n > X_{n-k:n}\}}, \tag{3.3}$$

where $X_{1:n} \leq \cdots \leq X_{n:n}$ denote the order statistics of $X_1^n, \ldots, X_n^n$. The estimator $\hat{C}$ uses the first k largest order statistics. As usual for the purpose of asymptotic in extreme value theory , the sequence $k = k(n)$ is an intermediate sequence satisfying

$$\lim_{n\to\infty} k = \infty \quad \text{and} \quad \lim_{n\to\infty} \frac{k}{n} = 0. \tag{3.4}$$

The growth rate of k must also be balanced with the rate of convergence in (3.1) and the oscillations of the skedasis function c. Assume there exists a positive decreasing function A_1 such that $\lim_{t\to\infty} A_1(t) = 0$ and

$$\sup_{n\in\mathbb{N}} \max_{1\leq i\leq n} \left| \frac{1 - F_{n,i}(x)}{1 - F(x)} - c(i/n) \right| = \mathcal{O}\left(A_1\left(\frac{1}{1-F(x)} \right) \right) \quad \text{as } x \to \infty. \tag{3.5}$$

It is also required that

$$\sqrt{k} A_1(n/k) \underset{n\to\infty}{\longrightarrow} 0 \tag{3.6}$$

and

$$\lim_{n\to\infty} \sqrt{k} \sup_{|u-v|\leq 1/n} |c(u) - c(v)| = 0. \tag{3.7}$$

Finally, γ is estimated thanks to the Hill estimator defined by

$$\hat{\gamma}_H = \frac{1}{k} \sum_{i=1}^{k} \log X_{n-i+1:n} - \log X_{n-k:n}. \tag{3.8}$$

The asymtptotic theory for the estimators is given in the following theorem.

Theorem 3.1 (Einmahl et al. (2016))**.** *Under conditions (3.2),(3.4),(3.5),(3.6) and (3.7) we have*

$$\sup_{0\leq s\leq 1} \left| \sqrt{k} \left(\hat{C}(s) - C(s) \right) - B\left(C(s) \right) \right| \underset{n\to\infty}{\longrightarrow} 0 \qquad a.s, \tag{3.9}$$

where B is a standard Brownian bridge.
Furthermore, under some second order regular variation conditions for F, $\hat{\gamma}_H$ satisfies

$$\sqrt{k}(\hat{\gamma}_H - \gamma) \underset{n\to\infty}{\longrightarrow} \gamma N_0 \; a.s, \tag{3.10}$$

(where N_0 is a standard normal variable) and is asymptotically independent of $\hat{C}$.

Section 3.2 is devoted to a point process analysis of the heteroscedastic extremes framework. We prove in Proposition 3.1 that the empirical point measure converges to a Poisson point process. We deduce in Corollary 3.1 that the times of exceedances and the values of exceedances are asymptotically independent with distributions respectively equal to the skedasis density function and the Pareto distribution. This shades some new light on Theorem 3.1 and explains why the estimators of the integrated skedasis function and the Hill estimator are asymptotically independent. Section 3.4 is devoted to the study of the log-linear trend model where the ML estimator of the trend parameter is shown to be consistent and asymptotically normal in Theorem 3.2. A parametric test for trend detection is deduced and Proposition 3.4 states its consistency under a monotonicity assumption of the skedasis function. Section 3.5 and 3.6 are dedicated respectively to the linear trend model and the discrete log-linear trend model introduced in de Haan et al. (2015). In Section 3.7, we propose a simulation study emphasizing two main issues: threshold selection and test power. We introduce and illustrate Lepski's method for threshold selection and then compare the power of parametric and non parametric tests. We finally present in Section 3.8 an application of the methodology to a dataset of minimal/maximal daily temperatures in Fort Collins, Colorado, during the 20th century.

3.2 A point process approach

The connection between extremes and point process is well-known and is an important conceptual tool in extreme value theory, see e.g. the monograph by Resnick (1987). A standard result available for i.i.d. observations $X_1, X_2, \ldots$ is the following: if $\max_{1\leq i\leq n}(X_i - b_n)/a_n$ converges to the extreme value distribution G_γ with the normalizing constants $a_n > 0$ and b_n, then the empirical point measure $\sum_{i=1}^n \varepsilon_{(i/n,(X_i-b_n)/a_n)}$ converges in distribution to a Poisson random measure. See Theorem 7.1 in Coles (2001) or Proposition 3.21 in Resnick (1987) for a precise statement.

We provide a similar result in the framework of the heteroscedastic model (3.1), that is for independent but not identically distributed observations. For the sake of simplicity, we focus on non-negative heavy-tailed observations. We denote by $\mathcal{M}_p = \mathcal{M}_p([0,1]\times(0,\infty])$ the space of point measures $\pi = \sum_{i\geq 1}\varepsilon_{(t_i,x_i)}$ with finitely many points in compact sets. We equip $\mathcal{M}_p$ with the topology of vague convergence, that is the smallest topology that makes the applications $\pi \mapsto \int f \mathrm{d}\pi$ continuous for all compactly supported continuous function $f : [0,1]\times(0,\infty]$. A random point measure (or point process) is a measurable application from a probability space into $\mathcal{M}_p$ equipped with the Borel σ-algebra. Recall that $\mathcal{M}_p$ is a Polish space (separable completely metrizable topological space).

Proposition 3.1. *Assume* $\{X_i^n,\ n \geq 1, 1 \leq i \leq n\}$ *are non-negative and satisfy the proportional tail condition (3.1) with* F *in the domain of attraction of the Fréchet distribution with tail index* $\alpha > 0$. *Consider the normalizing constant* $a_n = F^{\leftarrow}(1-1/n)$, $n \geq 1$, *with* $F^{\leftarrow}$ *the quantile function associated with* F *and*

define the normalized sample point measure

$$\Pi_n = \sum_{i=1}^{n} \varepsilon_{\left(\frac{i}{n}, \frac{X_i^n}{a_n}\right)}, \quad n \geq 1,$$

with ε the dirac measure. Then, as $n \to \infty$, Π_n converges in distribution in $\mathcal{M}_p$ to a Poisson random measure Π with intensity measure $c(s)\mathrm{d}s\alpha x^{-\alpha-1}\mathrm{d}x$ on $[0,1]\times(0,\infty]$.

Proof. The proof is mostly an adaptation of the proof of Proposition 3.21 in Resnick (1987) and relies on the convergence of Laplace functional. We need to prove that

$$\begin{aligned} &\sum_{k=1}^{n} \int_{[0,\infty]} \left(1 - e^{-f\left(\frac{k}{n},x\right)}\right) P\left(\frac{X_k^n}{a_n} \in dx\right) \\ \to &\int\int_{(0,1]\times[0,\infty]} \left(1 - e^{-f(s,x)}\right) c(s) ds \alpha x^{-\alpha-1} dx, \quad \text{as } n \to \infty, \end{aligned}$$

for all $f : [0,1] \times (0,\infty]$ continuous with compact support. This is equivalent to the vague convergence $\pi_n \xrightarrow{v} \pi$, where

$$\pi_n(ds, dx) = \sum_{k=1}^{n} \varepsilon_{\frac{k}{n}}(ds)\, P\left(\frac{X_k^n}{a_n} \in dx\right)$$

and

$$\pi(ds, dx) = c(s) ds \alpha x^{-\alpha-1} dx$$

are the intensity measures of Π_n and Π respectively. For $s \in [0,1]$ and $x > 0$, we have

$$\pi_n([0,s] \times (x,\infty]) = \sum_{k=1}^{[ns]} \bar{F}_{n,k}(a_n x),$$

where $\bar{F}_{n,k}(a_n x) = 1 - F_{n,k}(a_n x)$. Equation (3.1) implies that for all $\varepsilon > 0$, there is $A > 0$ such that

$$\left| \frac{\bar{F}_{n,k}(y)}{\bar{F}(y)} - c(k/n) \right| \leq \varepsilon, \quad y > A$$

with $\bar{F}(y) = 1 - F(y)$, whence we deduce, for n large enough,

$$\begin{aligned} \sum_{k=1}^{[ns]} |\bar{F}_{n,k}(a_n x) - c(k/n)\bar{F}(a_n x)| &\leq \sum_{k=1}^{[ns]} \left| \frac{\bar{F}_{n,k}(a_n x)}{\bar{F}(a_n x)} - c(k/n) \right| \bar{F}(a_n x) \\ &\leq \sum_{k=1}^{[ns]} \varepsilon \bar{F}(a_n x) \\ &\leq \varepsilon n \bar{F}(a_n x). \end{aligned} \tag{3.11}$$

The assumption that F belongs to the domain of attraction of the Fréchet distribution implies that the sequence $n\bar{F}(a_n x)$ converges to $x^{-\alpha}$ and is hence bounded.

Since $\varepsilon > 0$ is arbitrary, the right hand side in (3.11) is arbitrary small so that

$$\begin{aligned}\pi_n([0,s]\times(x,\infty]) &= \sum_{k=1}^{[ns]} \bar{F}_{n,k}\left(a_n x\right) \\ &= \sum_{k=1}^{[ns]} c(k/n)\bar{F}\left(a_n x\right) + o(1) \\ &= n\bar{F}\left(a_n x\right) \times \frac{1}{n}\sum_{k=1}^{[ns]} c(k/n) + o(1). \end{aligned} \tag{3.12}$$

We have already seen $n\bar{F}\left(a_n x\right) \to x^{-\alpha}$, and the Riemann sum associated to the function $t \in [0,1] \longrightarrow c(t)\mathbf{1}_{t<s}$ satisfies

$$\frac{1}{n}\sum_{k=1}^{[ns]} c(k/n) = \int_0^s c(t)\mathrm{d}t + o(1). \tag{3.13}$$

Equations (3.12) and (3.13) together imply

$$\pi_n([0,s]\times(x,\infty]) \longrightarrow \int_0^s c(t)x^{-\alpha}\mathrm{d}t, \quad \text{as } n \to \infty.$$

On the other hand,

$$\begin{aligned}\pi\left([0,s]\times(x,\infty]\right) &= \int_{[0,s]\times(x,\infty]} c(t)\mathrm{d}t\, \alpha x^{-\alpha-1}\mathrm{d}x \\ &= \int_0^s c(t)x^{-\alpha}dt.\end{aligned}$$

This proves $\pi_n([0,s]\times(x,\infty]) \to \pi([0,s]\times(x,\infty])$ for all $s \in [0,1]$ and $x > 0$. The required vague convergence $\pi_n \xrightarrow{v} \pi$ follows. □

In the following, we explain how Proposition 3.1 sheds some light on Theorem 3.1, at least on a heuristic level. We need to introduce the random variable T_l^n that will play a central role in the present paper.

Definition 3.1. For $1 \le l \le n$, denote by T_l^n the 'time' at which we observe the l^{th} highest observation, that is T_l^n is defined by $X^n_{nT_l^n} = X_{n+1-l:n}$.

Note that the random variables $T_1^n, \dots T_k^n$ represent the times of the exceedances above threshold $X_{n-k:n}$ so that the estimator $\hat{C}(s)$ defined by (3.3) can be rewritten as

$$\hat{C}(s) = \frac{1}{k}\sum_{i=1}^{[ns]} \mathbf{1}_{\{X_i^n > X_{n-k:n}\}} = \frac{1}{k}\sum_{l=1}^{k} \mathbf{1}_{\{T_l^n \le s\}}. \tag{3.14}$$

Hence $\hat{C}$ is the empirical distribution function associated with the sample $T_1^n, \dots, T_k^n$ of exceedance times.

Based on Proposition 3.1 and the point process approach, the next corollary proves that the exceedance times distribution is approximately $c(s)\mathrm{d}s$, providing some intuition why $\hat{C}$ is a good estimator of the distribution function C. Furthermore, we also obtain the asymptotic independence of the exceedance times and values, explaining the asymptotic independence of $\hat{C}$ and $\hat{\gamma}$ in Theorem 3.1.

Corollary 3.1. *For all fixed $k \geq 1$, we have the convergence in distribution of the times and values of the exceedances above $X_{n-k:n}$:*

$$\left(T_l^n, \frac{X_{n+1-l:n}}{X_{n-k:n}}\right)_{1\leq l\leq k} \xrightarrow{d} (\widetilde{T}_l, \widetilde{X}_{k+1-l:k})_{1\leq l\leq k}, \quad \text{as } n \to \infty, \tag{3.15}$$

where $\widetilde{X}_{1:k} \leq \cdots \leq \widetilde{X}_{k:k}$ are the order statistics of independent and identically distributed random variables $\widetilde{X}_1, \ldots, \widetilde{X}_k$ with α-Pareto distribution, i.e. $\mathbb{P}(\widetilde{X}_l > x) = x^{-\alpha}$, $x > 1$, and, independently, $\widetilde{T}_1, \ldots, \widetilde{T}_k$ are independent random variables with density c.

Equation (3.9) in Theorem 3.1 states that the empirical process $\hat{C}$ associated via (3.14) with the sample $T_1^n, \ldots, T_k^n$ of exceedance times has the same asymptotic behavior as the empirical process of i.i.d. random variables with density c. Note that $k = k(n) \to \infty$ there whereas k remains fixed in Corollary 3.1.

Proof. Consider $\mathcal{M}_p$ the space of point measures on $E = [0,1] \times (0, \infty]$. Let $\pi = \sum_{1\leq i\leq N} \varepsilon_{(t_i,x_i)}$ be a point measure with N points, possibly $N = +\infty$, and assume that the $k+1$ largest values within $(x_i)_{1\leq i\leq N}$ are reached uniquely and hence can be reordered as $x_{i_1} > \cdots > x_{i_{k+1}}$. We consider the mapping F that associates to any such π the vector $F(\pi) \in E^{k+1}$ defined by

$$F(\pi) = \left(t_{i_l}, \frac{x_{i_l}}{x_{i_{k+1}}}\right)_{1\leq l\leq k+1}.$$

For any sequence $\pi_n \to \pi$ in $\mathcal{M}$, it is easy to prove that $F(\pi_n) \to F(\pi)$ because the points of π_n converge to those of π , see Proposition 3.13 in Resnick (1987) for a more precise statement. This proves the continuity of F at π.

The definition of F is meant so that the left hand side of Equation (3.15) is equal to $F(\Pi_n)$, with Π_n the sample point process defined in Proposition 3.1. By the continuous mapping theorem, the convergence of $\Pi_n \xrightarrow{d} \Pi$ stated in Proposition 3.1 together with the fact that F is continuous at Π almost surely imply $F(\Pi_n) \xrightarrow{d} F(\Pi)$.

It remains to prove that the distribution of $F(\Pi)$ is equal to the distribution of $(\widetilde{T}_l, \widetilde{X}_{k+1-l:k})_{1\leq l\leq k}$ as in the right hand side of (3.15). The Poisson point measure Π with intensity $c(s)\mathrm{d}s\alpha x^{-\alpha-1}\mathrm{d}x$ can be represented as

$$\Pi = \sum_{i\geq 1} \varepsilon_{(T_i, \Gamma_i^{-1/\alpha})}$$

where $\Gamma_1 < \Gamma_2 < \cdots$ are the points of a homogeneous point process on $(0, \infty)$ and, independently, $T_1, T_2, \ldots$ are i.i.d. with distribution $c(t)\mathrm{d}t$. Then,

$$F(\Pi) = \left(T_l, \left(\frac{\Gamma_l}{\Gamma_{k+1}}\right)^{-1/\alpha}\right)_{1\leq l\leq k}.$$

The result follows since for the homogeneous Point process on $(0, \infty)$, the random variables $\Gamma_1/\Gamma_{k+1}, \ldots, \Gamma_k/\Gamma_{k+1}$ have the same joint distribution as the order statistics $U_{1:k} \leq \cdots \leq U_{k:k}$ of i.i.d. random variables with uniform distribution on $[0,1]$ (see Theorem 1 in Seshadri et al. (1969)) and furthermore $U^{-1/\alpha}$ is α-Pareto distributed for U uniform on $[0,1]$. □

3.3 Useful propositions

We present in this section two standard results in the theory of empirical process that will be useful in the rest of the paper together with their proof. The first lemma links the empirical distribution of $T_1^n, \dots, T_l^n$ and the empirical process $\hat{C}$ defined by (3.14).

Lemma 3.1. *For all continuously differentiable function $g : [0,1] \to \mathbb{R}$, we have*

$$\frac{1}{k}\sum_{i=1}^{k} g(T_i^n) = g(1) - \int_0^1 \hat{C}(t)g'(t)dt.$$

Proof. This follows by integration by part. Indeed,

$$\begin{aligned}\frac{1}{k}\sum_{i=1}^{k} g(T_i^n) &= \frac{1}{k}\sum_{i=1}^{k}\left(g(1) - \int_{T_i^n}^1 g'(t)dt\right)\\ &= g(1) - \frac{1}{k}\sum_{i=1}^{k}\int_0^1 \mathbf{1}_{\{T_i^n \le t\}} g'(t)dt\\ &= g(1) - \int_0^1 \hat{C}(t)g'(t)dt.\end{aligned}$$

□

The second lemma provides a simple expression for the variance of some integral of the Brownian bridge.

Proposition 3.2. *Let X be a $[0,1]$- valued random variable with distribution function F and B be a Brownian-Bridge on $[0,1]$. Then*

$$\mathrm{Var}\left(\int_0^1 B(F(s))ds\right) = \mathrm{Var}(X).$$

Proof. The Brownian bridge and hence its integral have zero mean, so that

$$\begin{aligned}\mathrm{Var}\left(\int_0^1 B\left(F(s)\right)ds\right) &= \mathbb{E}\left[\left(\int_0^1 B\left(F(s)\right)ds\right)^2\right]\\ &= \int_{[0,1]^2}\mathbb{E}\left(B(F(s_1)B(F(s_2))\right)ds_1ds_2\\ &= \int_{[0,1]^2}\mathrm{cov}(B(F(s_1)), B(F(s_2)))ds_1ds_2\\ &= \int_{[0,1]^2}\Big(\min(F(s_1), F(s_2)) - F(s_1)F(s_2)\Big)ds_1ds_2\\ &= 2\int_{[0,1]^2}\mathbf{1}_{s_1<s_2}F(s_1)ds_1ds_2 - \left(\int_{[0,1]}F(s_1)ds_1\right)^2.\end{aligned}$$

We have

$$\begin{aligned}
2\int_{[0,1]^2} \mathbf{1}_{s_1<s_2} F(s_1) ds_1 ds_2 &= 2\int_{[0,1]^2} \mathbf{1}_{s_1<s_2} \int_0^{s_1} dF(x) ds_1 ds_2 \\
&= 2\int_{[0,1]^3} \mathbf{1}_{s_1<s_2} \mathbf{1}_{x<s_1} dF(x) ds_1 ds_2 \\
&= 2\int_{x=0}^1 \int_{s_1=x}^1 \left(\int_{s_2=s_1}^1 ds_2 \right) ds_1 dF(x) \\
&= \int_0^1 (1-x)^2 dF(x).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\int_{[0,1]} F(s_1) ds_1 &= \int_{[0,1]} \int_0^{s_1} dF(x) ds_1 \\
&= \int_{[0,1]^2} 1_{x<s_1} dF(x) ds_1 \\
&= \int_{x=0}^1 \left(\int_x^1 ds_1 \right) dF(x) \\
&= \int_0^1 (1-x) dF(x).
\end{aligned}$$

Finally,

$$\begin{aligned}
\mathrm{Var}\left(\int_0^1 B\left(F(s)\right) ds \right) &= \int_0^1 (1-x)^2 dF(x) - \left(\int_{[0,1]} (1-x) dF(x) \right)^2 \\
&= \int_0^1 x^2 dF(x) - \left(\int_0^1 x dF(x) \right)^2 \\
&= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\
&= \mathrm{Var}(X).
\end{aligned}$$

□

3.4 Trend detection in the log-linear case

3.4.1 The log-linear trend model

We consider in this section the case of a log-linear trend model, corresponding to a skedasis function defined on $[0,1]$ of the form

$$c\left(t\right) = c_\theta\left(t\right) = e^{\theta t - h(\theta)} \quad \text{where} \quad \theta \in \mathbb{R}, \tag{3.16}$$

with

$$h(\theta) = \log\left(\int_0^1 e^{\theta t} dt \right) = \log\left(\frac{e^\theta - 1}{\theta} \right).$$

The value $h(\theta)$ ensures that the skedasis function has integral 1 as in (3.2). An important feature of this model is that the density family (3.16) forms an exponential family (Barndorff-Nielsen (1978)) making inference relatively easy.

Proposition 3.3. *Consider the statistical model $\{c_\theta(t), \theta \in \mathbb{R}\}$ given by (3.16) and T a random variable with density c_θ. We have,*

$$\mathbb{E}_\theta(T) = \frac{\partial h}{\partial \theta}(\theta) = \begin{cases} \dfrac{\theta e^\theta - e^\theta + 1}{\theta e^\theta - \theta} & \text{if } \theta \neq 0, \\ \dfrac{1}{2} & \text{if } \theta = 0, \end{cases} \tag{3.17}$$

and

$$\mathrm{Var}_\theta(T) = \frac{\partial^2 h}{\partial \theta^2}(\theta) = \begin{cases} \dfrac{(e^\theta - 1)^2 - \theta^2 e^\theta}{(\theta e^\theta - \theta)^2} & \text{if } \theta \neq 0, \\ \dfrac{1}{12} & \text{if } \theta = 0. \end{cases}$$

Furthermore, the Fisher information matrix at θ is

$$I(\theta) = \mathrm{Var}_\theta(T) \tag{3.18}$$

and $\dfrac{\partial h}{\partial \theta} : (-\infty, \infty) \longrightarrow (-1, 1)$ is a diffeomorphism (i.e. a continuously differentiable bijection with continuously differentiable inverse function).

Proof. We compute with the definition (3.16),

$$\mathbb{E}_\theta(T) = \int_0^1 t c_\theta(t) dt = \int_0^1 t e^{\theta t - h(\theta)} dt.$$

On the other hand,

$$\frac{\partial h}{\partial \theta}(\theta) = \frac{\int_0^1 t e^{\theta t} dt}{\int_0^1 e^{\theta t} dt} = \int_0^1 t e^{\theta t - h(\theta)} ds = \mathbb{E}_\theta(T) = \frac{\theta e^\theta - e^\theta + 1}{\theta e^\theta - \theta}.$$

Similarly,

$$\begin{aligned} \frac{\partial^2 h(\theta)}{\partial \theta^2} &= \frac{\int_0^1 t^2 e^{\theta t} dt \int_0^1 e^{\theta t} dt - (\int_0^1 t e^{\theta t} dt)^2}{(\int_0^1 e^{\theta t} dt)^2} \\ &= \int_0^1 t^2 e^{\theta t - h(\theta)} - \left(\int_0^1 t e^{\theta t - h(\theta)} \right)^2 = \mathbb{E}_\theta(T^2) - (\mathbb{E}_\theta(T))^2 \\ &= \mathrm{Var}_\theta(T) = \frac{(e^\theta - 1)^2 - \theta^2 e^\theta}{(\theta e^\theta - \theta)^2}. \end{aligned}$$

As for $I(\theta)$, we have

$$\begin{aligned} I(\theta) &= \mathrm{Var}_\theta \left(\frac{\partial \log c_\theta(T)}{\partial \theta} \right) = \mathrm{Var}_\theta \left(T - \frac{\partial h}{\partial \theta}(\theta) \right) \\ &= \mathrm{Var}_\theta(T). \end{aligned}$$

It can be checked directly from Equation (3.17) that $\frac{\partial h}{\partial \theta}$ is a diffeomorphism or this also follows from the general theory of exponential families (see Theorem 9.13 in Barndorff-Nielsen (1978)). □

3.4.2 Maximum likelihood estimator

Assume the observations $X_1^n, \dots, X_n^n$ follow the assumptions of the heteroscedastic model (3.1) with skedasis function $c = c_\theta$ from the log-linear trend model (3.16). In view of Corollary 3.1, the times of exceedances $T_1^n, \dots, T_k^n$ are approximatively independent with distribution $c_\theta(t)\mathrm{d}t$ so that it seems sensible to use maximum likelihood (ML) estimation for the inference of the parameter θ. Note that the model is misspecified since $T_1^n, \dots, T_k^n$ are not i.i.d. with density c_θ even if Equation (3.9) suggests that there should be asymptotically no difference .

The ML method works indeed well as proved in Theorem 3.2 below. The log-likelihood function is

$$\begin{aligned} L\left(\theta; T_1^n, \dots, T_k^n\right) &= \frac{1}{k} \log\left(\prod_{i=1}^{k} c_\theta(T_i^n)\right) \\ &= \frac{1}{k} \sum_{i=1}^{k} \left(\theta T_i^n - h(\theta)\right) \\ &= \theta \bar{T}_k - h(\theta) \end{aligned}$$

with

$$\bar{T}_k = \frac{1}{k} \sum_{i=1}^{k} T_i^n.$$

The ML estimator of θ is defined as

$$\hat{\theta}_k = \underset{\theta \in \mathbb{R}}{\operatorname{argmax}}\, L\left(\theta; T_1^n, \dots, T_k^n\right).$$

It is unique and satisfies the score equation

$$\bar{T}_k - \frac{\partial h}{\partial \theta}(\hat{\theta}_k) = 0,$$

so that

$$\hat{\theta}_k = \left(\frac{\partial h}{\partial \theta}\right)^{-1} (\bar{T}_k) \tag{3.19}$$

where $\left(\frac{\partial h}{\partial \theta}\right)^{-1}$ is the inverse function of $\frac{\partial h}{\partial \theta}$ and is continuously differentiable according to Proposition 3.3.

Theorem 3.2. *In the log-linear trend model (3.16) and under assumptions (3.4),(3.5),(3.6) and (3.7), the ML estimator (3.19) satisfies*

$$\hat{\theta}_k \xrightarrow{d} \theta_0 \quad \text{as } n \to \infty \tag{3.20}$$

and

$$\sqrt{k}\left(\hat{\theta}_k - \theta_0\right) \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0, I(\theta_0)^{-1}\right), \tag{3.21}$$

with $I(\theta)$ the Fisher information given by (3.18).

Proof. We start by proving (3.20). By Lemma 3.1, we have

$$\bar{T}_k = 1 - \int_0^1 \hat{C}(s)ds.$$

On the other hand

$$\mathbb{E}_{\theta_0}(T) = 1 - \int_0^1 C_{\theta_0}(s)ds,$$

with C_{θ_0} the distribution function associated with the density c_{θ_0}. We deduce

$$\begin{aligned}\left|\bar{T}_k - \mathbb{E}_{\theta_0}(T)\right| &= \left|\int_0^1 C_{\theta_0}(s)ds - \int_0^1 \hat{C}(s)ds\right| \\ &\leq \int_0^1 \left|C_{\theta_0}(s) - \hat{C}(s)\right| ds \\ &= \mathcal{O}_p\left(\frac{1}{\sqrt{k}}\right)\end{aligned}$$

where the last estimate follows from Theorem 3.1. Hence, we deduce $\bar{T}_k \xrightarrow{d} \mathbb{E}_{\theta_0}(T)$. By Equation (3.17), $\mathbb{E}_{\theta_0}(T) = \frac{\partial h}{\partial \theta}(\theta_0)$ and consequently thanks to the continuity of $\left(\frac{\partial h}{\partial \theta}\right)^{-1}$,

$$\hat{\theta_k} = \left(\frac{\partial h}{\partial \theta}\right)^{-1}(\bar{T}_k) \xrightarrow{d} \left(\frac{\partial h}{\partial \theta}\right)^{-1}\left(\frac{\partial h}{\partial \theta}(\theta_0)\right) = \theta_0.$$

As for the proof of (3.21), let us first demonstrate the asymptotic normality for $\bar{T}_k$. From (3.9) we have

$$\hat{C}(s) = C_{\theta_0}(s) + \frac{1}{\sqrt{k}}B(C_{\theta_0}(s)) + \frac{1}{\sqrt{k}}\epsilon_k(s),$$

so that

$$\begin{aligned}\bar{T}_k &= \int_0^1 (1 - \hat{C}(s))ds \\ &= \int_0^1 (1 - C_{\theta_0}(s))ds - \frac{1}{\sqrt{k}}\int_0^1 B(C_{\theta_0}(s))ds - \frac{1}{\sqrt{k}}\int_0^1 \epsilon_k(s)ds \\ &= \mathbb{E}_{\theta_0}(T) - \frac{1}{\sqrt{k}}\int_0^1 B(C_{\theta_0}(s))ds - \frac{1}{\sqrt{k}}\int_0^1 \epsilon_k(s)ds.\end{aligned}$$

Equation (3.9) implies that the error term $\epsilon_k(s)$ converges to zero uniformly for all $s \in [0,1]$, hence

$$\left|\sqrt{k}\left(\bar{T}_k - E_{\theta_0}(T)\right) + \int_0^1 B(C_{\theta_0}(s))ds\right| = \int_0^1 \epsilon_k(s)ds \xrightarrow{a.s} 0,$$

and we conclude that

$$\bar{T}_k = \mathbb{E}_{\theta_0}(T) - \frac{1}{\sqrt{k}}\int_0^1 B(C_{\theta_0}(s))ds + o_P(1).$$

The integral of the Brownian bridge is a Gaussian random variable with mean 0 and variance $\mathrm{Var}_{\theta_0}(T)$ according to Proposition 3.2. We deduce

$$\sqrt{k}\left(\bar{T}_k - \mathbb{E}_{\theta_0}(T)\right) \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0, \mathrm{Var}_{\theta_0}(T)\right). \tag{3.22}$$

Applying the delta-method (Theorem 3.1 in van der Vaart (1998)) for $\hat{\theta}_k = \left(\frac{\partial h}{\partial \theta}\right)^{-1}\left(\bar{T}_k\right)$, we deduce from (3.22)

$$\sqrt{k}\left(\left(\frac{\partial h}{\partial \theta}\right)^{-1}(\bar{T}_k) - \left(\frac{\partial h}{\partial \theta}\right)^{-1}(\mathbb{E}_{\theta_0}(T))\right) \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0, \frac{1}{\left(\frac{\partial^2 h(\theta_0)}{\partial \theta^2}\right)^2}\mathrm{Var}_{\theta_0}(T)\right).$$

From Proposition 3.3,

$$\frac{\partial^2 h(\theta_0)}{\partial \theta^2} = \mathrm{Var}_{\theta_0}(T) = I(\theta_0)$$

so that this simplifies into

$$\sqrt{k}\left(\hat{\theta}_k - \theta_0\right) \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0, I(\theta_0)^{-1}\right).$$

□

3.4.3 Testing for a log-linear trend

For the purpose of detecting an increasing log-linear trend, in the framework of climate change for instance, one may be interested in testing

$$H_0 : \theta_0 \leq 0 \qquad \text{versus} \qquad H_1 : \theta_0 > 0.$$

Based on the asymptotic normality of $\hat{\theta}_k$ stated in (3.21), it is natural to accept H_0 if $\dfrac{\hat{\theta}_k}{\sigma_0/\sqrt{k}} \leq z_{1-\alpha}$ the quantile of order $1-\alpha$ of the standard normal distribution and ${\sigma_0}^2 = \frac{1}{I(\theta=0)} = \frac{1}{\mathrm{Var}_{\theta=0}(T)} = 12$.

Interestingly, this test is still valid in the non parametric framework where the skedasis function c is monotone. Define formally the test φ_k by

$$\varphi_k = \begin{cases} 0 & \text{if } \hat{\theta}_k \leq z_{1-\alpha}\frac{\sigma_0}{\sqrt{k}}, \\ \\ 1 & \text{otherwise.} \end{cases} \tag{3.23}$$

Proposition 3.4. *Assume the skedasis function c is monotone and consider the test*

$$H_0 : c \text{ is non increasing} \quad \text{versus} \quad H_1 : c \neq 1 \text{ is non decreasing.}$$

The test φ_k defined in (3.23) satisfies

$$\lim_{n\to\infty} \mathbb{P}(\varphi_k = 0) \geq 1-\alpha \quad \text{if } c \text{ is non-increasing,}$$
$$\lim_{n\to\infty} \mathbb{P}(\varphi_k = 1) = 1 \quad \text{if } c \text{ is non decreasing; } c \neq 1,$$

with α the significance level.

Remark 3.1. *It is straightforward to modify the test above to consider bilateral alternatives*

$$H_0 : \theta_0 = 0 \qquad versus \qquad H_1 : \theta_0 \neq 0$$

and, in the non parametric case,

$$H_0 : c = 1 \quad versus \quad H_1 : c \neq 1 \text{ monotone}$$

The proof of Proposition 3.4 relies on the following lemma.

Lemma 3.2. *If c is a monotone density function on $[0,1]$, then*

i) $c = 1$ *if and only if* $\int_0^1 sc(s)ds = \frac{1}{2}$,

ii) c *is non decreasing if and only if* $\int_0^1 sc(s)ds \geq \frac{1}{2}$,

iii) c *is non increasing if and only if* $\int_0^1 sc(s)ds \leq \frac{1}{2}$.

Proof. We start by proving the direct implication in *ii*). Suppose that c is a non decreasing density function. Then, the cumulative distribution function $C(s) = \int_0^s c(t)dt$ is convex so that the conditions $C(0) = 0$ and $C(1) = 1$ imply $C(s) \leq s$ for all $s \in [0,1]$. We deduce $\int_0^1 C(s)ds \leq \frac{1}{2}$ and, with an integration by part,

$$\int_0^1 sc(s)ds = 1 - \int_0^1 C(s)ds \geq \frac{1}{2}.$$

This proves the direct implication in *ii*). The direct implication in *iii*) is proved in a similar way. Considering the contrapositive of these direct implications, we obtain the indirect implication in *ii*) and *iii*) (note that since c is monotone, the negation of c non increasing is c non-decreasing and non constant). Finally, *i*) is deduced from *ii*) and *iii*) together since $c = 1$ is the only density function which is at the same time non increasing and non decreasing. □

Proof of proposition 3.4. From Lemma 3.2, c is non-decreasing non constant if and only if $\int_0^1 sc(s)ds > 1/2$ and in this case $\bar{T}_k \xrightarrow{d} \mathbb{E}(T) = \int_0^1 sc(s)ds > 1/2$, so that

$$\hat{\theta}_k = \left(\frac{\partial h}{\partial \theta}\right)^{-1} (\bar{T}_k) \xrightarrow{d} \left(\frac{\partial h}{\partial \theta}\right)^{-1} (\mathbb{E}(T)) > \left(\frac{\partial h}{\partial \theta}\right)^{-1} \left(\frac{1}{2}\right) = 0$$

We deduce that in this case, $\mathbb{P}(\phi_k = 1) \to 1$. In a similar way we prove that for c non-increasing non constant $\mathbb{P}(\phi_k = 0) \to 1$. The last case is $c \equiv 1$ that corresponds to the log-linear trend model with $\theta = 0$, whence Theorem 3.21 provides the asymptotic normality of $\hat{\theta}_k$ and the test ϕ_k has been built so that $\mathbb{P}(\phi_k = 0) \to 1 - \alpha$ in this case. □

3.5 Trend detection in the linear trend model

3.5.1 The linear trend model

Considering a linear trend model is natural since this type of models is simple and encountered in many situations. The linear skedasis function is defined as

$$c_\theta(t) = \theta(2t-1)+1 \qquad \text{with} -1<\theta<1, \tag{3.24}$$

so the corresponding distribution function is

$$C_\theta(t) = \theta t^2 + (1-\theta)t. \tag{3.25}$$

The condition $-1 < \theta < 1$ follows from the fact that c_θ must remain positive. We obtain similar results as for the log-linear model and consider consistency and asymptotic normality of the moment and ML estimators. Nonetheless, the analysis of the ML estimator is more involved since there is no simple closed formula.

Proposition 3.5. *Consider the statistical model $\{c_\theta(t), \theta \in (-1,1)\}$ defined in (3.24) and T a random variable with density c_θ. We have,*

$$\mathbb{E}_\theta(T) = \frac{\theta+3}{6}, \tag{3.26}$$
$$\mathrm{Var}_\theta(T) = \frac{1}{12} - \frac{\theta^2}{36}$$

and

$$I(\theta) = \begin{cases} \dfrac{(\tanh)^{-1}(\theta)-\theta}{\theta^3} & \text{if } \theta \neq 0, \\ \dfrac{1}{3} & \text{if } \theta = 0. \end{cases} \tag{3.27}$$

Proof. The expectation and variance are computed readily:

$$\mathbb{E}_\theta(T) = \int_0^1 t c_\theta(t)dt = \int_0^1 \left(2\theta t^2 + (1-\theta)t\right)dt = \frac{\theta+3}{6}$$

and

$$\begin{aligned}\mathrm{Var}_\theta(T) &= \int_0^1 t^2 c_\theta(t)dt - (\mathbb{E}_\theta(T))^2 = \int_0^1 t^2[\theta(2t-1)+1]dt - \left(\frac{\theta+3}{6}\right)^2 \\ &= \frac{1}{12} - \frac{\theta^2}{36}.\end{aligned}$$

The information matrix is given by

$$I(\theta) = \mathrm{Var}_\theta\left(\frac{\partial \log(c_\theta(T))}{\partial\theta}\right) = \mathrm{Var}_\theta\left(\frac{2T-1}{1+\theta(2T-1)}\right) = \int_0^1 \frac{(2t-1)^2}{(1+\theta(2t-1))^2}\left(1+\theta(2t-1)\right)dt.$$

Thanks to the change of variable $u = 1+\theta(2t-1)$, we obtain

$$I(\theta) = \frac{(\tanh)^{-1}(\theta)-\theta}{\theta^3}, \quad \text{if } \theta \neq 0.$$

If $\theta = 0$ then,

$$I(\theta) = \int_0^1 (2t-1)^2 dt = \frac{1}{3}.$$

□

3.5.2 Moment estimator of the linear trend parameter

The simplest estimation method is the method of moments which consists in equating the empirical mean of a sample to the theoretical mean. From Equation (3.26), $\theta_0 = 6\mathbb{E}_{\theta_0}(T) + 3$, whence the moment estimator $\hat{\theta}_k$ of θ_0 can be expressed explicitly as

$$\hat{\theta}_k = 6\bar{T}_k - 3. \tag{3.28}$$

Proposition 3.6. *Under assumptions (3.4),(3.5),(3.6) and (3.7) in the linear trend model (3.24), the moment estimator (3.28) satisfies*

$$\hat{\theta}_k \xrightarrow{d} \theta_0 \ \ as\, n \to \infty \tag{3.29}$$

and

$$\sqrt{k}(\hat{\theta}_k - \theta_0) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, 3 - {\theta_0}^2).$$

Proof. The proof of (3.29) is straightforward. Since $\bar{T}_k \xrightarrow{d} \mathbb{E}_{\theta_0}(T)$ then

$$\hat{\theta}_k = 6\bar{T}_k - 3 \xrightarrow{d} 6\mathbb{E}_{\theta_0}(T) - 3 = \theta_0.$$

From (3.26) and (3.28) we have that $\mathbb{E}_{\theta_0}(T) = \dfrac{\theta_0 + 3}{6}$ and $\hat{\theta}_k = 6\bar{T}_k - 3$. Thanks to (3.22) we obtain,

$$\sqrt{k}(\hat{\theta}_k - \theta_0) = 6\sqrt{k}(\bar{T}_k - E_{\theta_0}(T)) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, 36\,\mathrm{Var}_{\theta_0}(T)) = \mathcal{N}(0, 3 - {\theta_0}^2).$$

□

3.5.3 Maximum likelihood estimator of the linear trend parameter

For the same reasons presented in the log-linear trend model for the use of the ML estimation, it also makes sense to estimate the parameter θ using this method in the linear trend model. The log-likelihood function of θ with observations $T_1^n, \ldots, T_k^n$ having c_θ as density function is given by:

$$L(\theta; T_1^n, \ldots, T_k^n) = \frac{1}{k}\sum_{i=1}^{k} \log\left(\theta(2T_i^n - 1) + 1\right).$$

This defines a strictly concave function on the domain

$$\left(\frac{1}{2\min(T_i^n) - 1}, \frac{1}{2\max T_i^n - 1}\right).$$

Assuming $\min(T_i^n) < \frac{1}{2} < \max(T_i^n)$, the domain contains $(-1, 1)$ and the derivative $\frac{\partial L}{\partial \theta}$ is a decreasing diffeomorphism from the domain onto $(-\infty, \infty)$. Hence, there exists a unique ML estimator $\hat{\theta}_k$ such that

$$\frac{\partial}{\partial \theta} L\left(\hat{\theta}_k; T_1^n, \ldots, T_k^n\right) = 0.$$

Denoting by $l_\theta(t) = \frac{\partial}{\partial\theta}\log(c_\theta(t))$ the score function, this writes

$$\frac{1}{k}\sum_{i=1}^{k} l_{\hat{\theta}_k}(T_i^n) = 0. \tag{3.30}$$

Theorem 3.3. *In the linear trend model (3.24), under conditions (3.4),(3.5),(3.6) and (3.7) the ML estimator $\hat{\theta}_k$ satisfying 3.30 is consistent*

$$\hat{\theta}_k \xrightarrow{\mathbb{P}} \theta_0 \ \ as\, n \to \infty$$

and asymptotically normal

$$\sqrt{k}\left(\hat{\theta}_k - \theta_0\right) \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0, \frac{1}{I(\theta_0)}\right),$$

where $I(\theta_0)$ is the Fisher information defined in (3.27) .

Proof. We start by proving the consistency. For this purpose we consider the function

$$\theta \longrightarrow \frac{\partial}{\partial\theta} L(\theta; T_1^n, \dots, T_k^n).$$

For $\varepsilon > 0$ such that $(\theta_0 - \varepsilon, \theta_0 + \varepsilon) \subset (-1, 1)$, an integration by part (Lemma 3.1) yields

$$\frac{\partial}{\partial\theta} L(\theta_0 \pm \varepsilon; T_1^n, \dots, T_k^n) = \frac{1}{k}\sum_{i=1}^{k} l_{\theta_0\pm\varepsilon}(T_i^n) = l_{\theta_0\pm\varepsilon}(1) - \int_0^1 \hat{C}(t) l'_{\theta_0\pm\varepsilon}(t)dt.$$

Then, Equation (3.9) together with another integration by parts yield

$$\frac{\partial}{\partial\theta} L(\theta_0 \pm \varepsilon; T_1^n, \dots, T_k^n) \xrightarrow{d} l_{\theta_0\pm\varepsilon}(1) - \int_0^1 C_{\theta_0}(t) l'_{\theta_0\pm\varepsilon}(t)dt = \mathbb{E}_{\theta_0}[l_{\theta_0\pm\varepsilon}(t)]. \tag{3.31}$$

The function $\theta \longrightarrow \mathbb{E}_{\theta_0}\left[l_\theta(T)\right]$ is strictly decreasing and vanishes at $\theta = \theta_0$ so that

$$\mathbb{E}_{\theta_0}[l_{\theta_0-\varepsilon}(t)] > 0 \quad \text{and} \quad \mathbb{E}_{\theta_0}[l_{\theta_0+\varepsilon}(t)] < 0.$$

Using the convergence (3.31), we deduce that for $\delta > 0$, there exists k_0 such that for all $k > k_0$, $\mathbb{P}\left(\frac{\partial}{\partial\theta} L(\theta_0 - \varepsilon, T_1^n, \dots, T_k^n) > 0\right) \geq 1-\delta$ and $\mathbb{P}\left(\frac{\partial}{\partial\theta} L(\theta_0 + \varepsilon, T_1^n, \dots, T_k^n) < 0\right) \geq 1 - \delta$. Then, for $k > k_0$, the event

$$\left\{\frac{\partial}{\partial\theta} L(\theta_0 - \varepsilon, T_1^n, \dots, T_k^n) > 0 > \frac{\partial}{\partial\theta} L(\theta_0 + \epsilon; T_1^n, \dots, T_k^n)\right\}$$

has probability at least $1 - 2\delta$, and $\hat{\theta}_k \in (\theta_0 - \varepsilon, \theta_0 + \varepsilon)$ on this event (recall that $\frac{\partial L}{\partial\theta}$ is a decreasing diffeomorphism). We deduce

$$\mathbb{P}(\hat{\theta}_k \in (\theta_0 - \varepsilon, \theta_0 + \varepsilon)) \geq 1 - 2\delta,$$

and the consistency $\hat{\theta}_k \xrightarrow{\mathbb{P}} \theta_0$.

We next consider the asymptotic normality of $\hat{\theta}_k$. The starting point is the relations

$$\int_0^1 \frac{\partial l_{\theta_0}}{\partial t}(t) C_{\theta_0}(t) dt = \frac{1}{1+\theta_0} \tag{3.32}$$

$$\int_0^1 \frac{\partial l_{\hat{\theta}_k}}{\partial t}(t) \hat{C}(t) dt = \frac{1}{1+\hat{\theta}_k}. \tag{3.33}$$

Equation (3.32) is obtained via the score equation

$$\int_0^1 l_{\theta_0}(t) dC_{\theta_0}(t) = \int_0^1 \frac{\frac{\partial c_{\theta_0}}{\partial \theta_0}(t)}{c_{\theta_0}(t)} c_{\theta_0}(t) dt = \int_0^1 \frac{\partial c_{\theta_0}}{\partial \theta}(t) dt = \frac{\partial}{\partial \theta} \int_0^1 c_\theta(t) dt \Big|_{\theta=\theta_0} = 0$$

thanks to an integration by part

$$\int_0^1 \frac{\partial l_{\theta_0}}{\partial t}(t) C_{\theta_0}(t) dt = [l_{\theta_0}(t) C_{\theta_0}(t)]_0^1 - \int_0^1 l_{\theta_0}(t) c_{\theta_0}(t) dt = l_{\theta_0}(1).$$

Equation (3.33) follows from (3.30)

$$\frac{1}{k} \sum_{i=1}^k l_{\hat{\theta}_k}(T_i^n) = \int_0^1 l_{\hat{\theta}_k}(s) d\hat{C}(s) ds = 0$$

and an integration by parts.

Multiplying Equation (3.33) by $1+\hat{\theta}_k$ and Equation (3.32) by $1+\theta_0$ and taking the difference of the two, we obtain

$$\int_0^1 \left((1+\hat{\theta}_k) \frac{\partial l_{\hat{\theta}_k}}{\partial t}(t) \hat{C}(t) dt - (1+\theta_0) \frac{\partial l_{\theta_0}}{\partial t}(t) C_{\theta_0}(t) \right) dt = 0.$$

Expliciting the derivative $\frac{\partial l_\theta}{\partial t}$, we obtain

$$\int_0^1 \left(\frac{1+\hat{\theta}_k}{(1+\hat{\theta}_k(2t-1))^2} \hat{C}(t) - \frac{1+\theta_0}{(1+\theta_0(2t-1))^2} C_{\theta_0}(t) \right) dt = 0. \tag{3.34}$$

Using the consistency $\hat{\theta}_k \to \theta_0$, it is natural to introduce the Taylor expansion

$$\begin{aligned} &\frac{1+\hat{\theta}_k}{(1+\hat{\theta}_k(2t-1))^2} \\ =&\frac{1+\theta_0}{(1+\theta_0(2t-1))^2} + \frac{1-(\theta_0+2)(2t-1)}{(1+\theta_0(2t-1))^3}(\hat{\theta}_k - \theta_0) + (\hat{\theta}_k - \theta_0)\varepsilon_t(\hat{\theta}_k - \theta_0) \end{aligned}$$

where $\varepsilon_t(h) \to 0$ as $h \to 0$. On the other hand, Equation (3.9) provides the asymptotic expansion

$$\hat{C}(t) = C_{\theta_0}(t) + \frac{1}{\sqrt{k}} B(C_{\theta_0}(t)) + \frac{\epsilon_k(t)}{\sqrt{k}}$$

with $\epsilon_k(t) \to 0$ as $k \to \infty$ uniformly in t. Plugging these two asymptotic expansions into Equation (3.34), the main term cancels out and it remains, after simplification,

$$(A_1 + R_1)\sqrt{k}(\hat{\theta}_k - \theta_0) + A_2 + R_2 + R_3 = 0 \tag{3.35}$$

with

$$A_1 = \int_0^1 \frac{1-(\theta_0+2)(2t-1)}{(1+\theta_0(2t-1))^3} C_{\theta_0}(t)dt,$$
$$A_2 = \int_0^1 \frac{(1+\theta_0)}{(1+\theta_0(2t-1))^2} B(C_{\theta_0}(t))dt,$$

and

$$R_1 = \int_0^1 \varepsilon_t(\hat{\theta}_k - \theta_0)\Big[C_{\theta_0}(t) + \frac{B(C_{\theta_0}(t))}{\sqrt{k}} + \frac{\epsilon_k(t)}{\sqrt{k}}\Big]dt,$$
$$R_2 = (\hat{\theta}_k - \theta_0)\int_0^1 \frac{1-(\theta_0+2)(2t-1)}{(1+\theta_0(2t-1))^3}\Big[B(C_{\theta_0}(t)) + \epsilon_k(t)\Big]dt,$$
$$R_3 = \int_0^1 \frac{1+\theta_0}{(1+\theta_0(2t-1))^2}\epsilon_k(t)dt.$$

We will prove below that the remainder terms R_i, $i = 1, 2, 3$, are asymptotically negligible, that is $R_i = o_P(1)$ as $k \to \infty$. We first proceed with the analysis of the main terms A_1, A_2 in (3.35). A direct computation shows that

$$A_1 = \begin{cases} \dfrac{(1+\theta_0)((\tanh)^{-1}(\theta_0)-\theta_0)}{2{\theta_0}^3} & \text{if } \theta_0 \neq 0 \\ \\ \dfrac{1}{6} & \text{if } \theta_0 = 0 \end{cases} = \frac{(1+\theta_0)}{2} I(\theta_0).$$

On the other hand, since $B(C_{\theta_0}(s))$ is a standard Brownian bridge , the stochastic term A_2 is a centered Gaussian variable with variance

$$\begin{aligned}
\mathrm{Var}(A_2) &= \mathbb{E}\left[\left(\int_{[0,1]} \frac{(1+\theta_0)}{(1+\theta_0(2t-1))^2} B(C_{\theta_0}(t))dt\right)^2\right] \\
&= \mathbb{E}\left[\int\int_{[0,1]^2} \frac{(1+\theta_0)}{(1+\theta_0(2t_1-1))^2} B(C_{\theta_0}(t_1)) \frac{(1+\theta_0)}{(1+\theta_0(2t_2-1))^2} B(C_{\theta_0}(t_2))dt_1dt_2\right] \\
&= \int\int_{[0,1]^2} \frac{(1+\theta_0)^2}{(1+\theta_0(2t_1-1))^2(1+\theta_0(2t_2-1))^2} \mathrm{cov}(B(C_{\theta_0}(t_1)), B(C_{\theta_0}(t_2)))dt_1dt_2 \\
&= \int\int_{[0,1]^2} \frac{(1+\theta_0)^2}{(1+\theta_0(2t_1-1))^2(1+\theta_0(2t_2-1))^2} C_{\theta_0}(t_1)(1-C_{\theta_0}(t_2))\mathbb{1}_{t_1<t_2}dt_1dt_2 \\
&\quad + \int\int_{[0,1]^2} \frac{(1+\theta_0)^2}{(1+\theta_0(2t_1-1))^2(1+\theta_0(2t_2-1))^2} C_{\theta_0}(t_2)(1-C_{\theta_0}(t_1))\mathbb{1}_{t_1>t_2}dt_1dt_2 \\
&= 2\int\int_{[0,1]^2} \frac{(1+\theta_0)^2}{(1+\theta_0(2t_1-1))^2(1+\theta_0(2t_2-1))^2} C_{\theta_0}(t_1)(1-C_{\theta_0}(t_2))\mathbb{1}_{t_1<t_2}dt_1dt_2 \\
&= \begin{cases} \dfrac{(1+\theta_0)^2}{4\theta_0^3}((\tanh)^{-1}(\theta_0)-\theta_0) & \text{if } \theta_0 \neq 0 \\ \\ \dfrac{1}{12} & \text{if } \theta_0 = 0 \end{cases} \\
&= \frac{(1+\theta_0)^2}{4} I(\theta_0).
\end{aligned}$$

Using this, Equation (3.35) writes

$$(A_1 + o_P(1))\sqrt{k}\Big(\hat{\theta}_k - \theta_0\Big) + A_2 + o_P(1) = 0$$

so that

$$\sqrt{k}\Big(\hat{\theta}_k - \theta_0\Big) = -\frac{A_2 + o_P(1)}{A_1 + o_P(1)} = -\frac{A_2}{A_1} + o_P(1)$$

yielding the announced asymptotic normality

$$\sqrt{k}(\hat{\theta}_k - \theta_0) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{I(\theta_0)}\right).$$

We finally prove that the remainder terms R_i, $i = 1, 2, 3$, are negligible. For the first term R_1, observe that $\hat{\theta}_k - \theta_0 = o_P(1)$ so that $\epsilon_t(\hat{\theta}_k - \theta_0) = o_P(1)$ uniformly in $t \in [0, 1]$. On the other hand,

$$C_{\theta_0}(t) + \frac{B(C_{\theta_0}(t))}{\sqrt{k}} + \frac{\epsilon_k(t)}{\sqrt{k}} = O_P(1)$$

uniformly in $t \in [0, 1]$. We deduce $R_1 = o_P(1)$. Similarly, the term R_2 appears as the product of $\hat{\theta}_k - \theta_0 = o_P(1)$ and an integral term which is $O_P(1)$, whence $R_2 = o_P(1)$. For the remainder term R_3, we observe that $\varepsilon_k(t) = o_P(1)$ uniformly in $t \in [0, 1]$ and we deduce $R_3 = o_P(1)$. □

3.5.4 Comparison of the two estimators

We compare the asymptotic variance of the moment and ML estimators. Recall the asymptotic normality results for the moment estimator

$$\sqrt{k}\Big(\hat{\theta}_k - \theta_0\Big) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, 3 - {\theta_0}^2)$$

and for the ML estimator

$$\sqrt{k}\Big(\hat{\theta}_k - \theta_0\Big) \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0, \frac{{\theta_0}^3}{(\tanh)^{-1}(\theta_0) - \theta_0}\right).$$

ML theory suggests that the ML estimator is asymptotically efficient and so the asymptotic variance of the ML estimator should be always lower. Denoting the asymptotic variances by

$$V_M(\theta_0) = 3 - {\theta_0}^2 \quad \text{and} \quad V_{ML}(\theta_0) = \frac{{\theta_0}^3}{(\tanh)^{-1}(\theta_0) - \theta_0},$$

we plot in Figure 3.1 the graphs of V_M and V_{ML} as functions of $\theta_0 \in (-1, 1)$. Indeed, we can check that the moment estimator has always higher asymptotic variance than the ML estimator. The variance of the two estimators are equal when $\theta_0 = 0$ and the relative efficiency of the MLE over the moment estimators increases as $|\theta_0|$ increases.

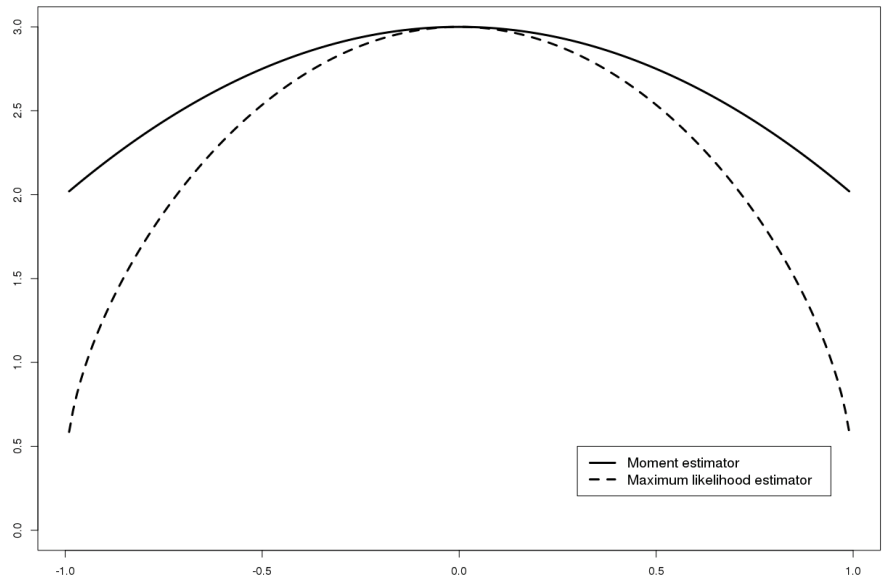


Figure 3.1 – Asymptotic variance of the moment and ML estimators.

3.6 Trend detection in the discrete log-linear trend model

The aim of this section is to use the present methodology to study the discrete log-linear model introduced by de Haan et al. (2015). This model corresponds to the situation when observations arise at different times $t_1 < \ldots < t_K$, with np_i observations at time $t_i \in [0,1]$, $i = 1, \ldots, K$. We assume $p_i \in (0,1)$ and $\sum_{i=1}^K p_i = 1$ so that n is the total number of observations. The distribution of observations is assumed to evolve over time according to a discrete log-linear skedasis function. As we will see, we can exploit in this case the exponential family structure of the limiting model and obtain a simple and more accurate estimator of the trend parameter.

3.6.1 The discrete log-linear trend model

The discrete log-linear model with parameter $\theta \in \mathbb{R}$, times $t_1 < \cdots < t_K$ and proportions $p_1, \ldots, p_K \geq 0$ (such that $\sum_{i=1}^K p_i = 1$) corresponds to the piecewise constant skedasis function

$$c_\theta(t) = \begin{cases} e^{\theta t_1 - h(\theta)} & \text{if} \quad 0 \leq t \leq p_1, \\ e^{\theta t_2 - h(\theta)} & \text{if} \quad p_1 < t \leq p_1 + p_2, \\ \vdots & \\ e^{\theta t_K - h(\theta)} & \text{if} \quad p_1 + \ldots + p_{K-1} < t \leq 1 \end{cases} \tag{3.36}$$

with

$$h(\theta) = \log \sum_{i=1}^K p_i e^{\theta t_i}.$$

The definition of h ensures that

$$\int_0^1 c_\theta(t)dt = \sum_{i=1}^{K} p_i e^{\theta t_i - h(\theta)} = 1.$$

We introduce the statistic $S(t)$

$$S(t) = \begin{cases} t_1 & \text{if} \quad 0 \leq t \leq p_1, \\ t_2 & \text{if} \quad p_1 < t \leq p_1 + p_2, \\ \quad \vdots & \\ t_K & \text{if} \quad p_1 + \ldots + p_{K-1} < t \leq 1. \end{cases}$$

Note that $S(t)$ is the sufficient statistic of the underlying exponential model, that is the skedasis function can be rewritten as

$$c_\theta(t) = e^{\theta S(t) - h(\theta)}.$$

Proposition 3.7. *Consider the statistical model $\{c_\theta(t), \theta \in \mathbb{R}\}$ defined in (3.36) and T a random variable with density c_θ. We have*

$$\mathbb{E}_\theta(S(T)) = \frac{\partial h}{\partial \theta}(\theta) = \frac{\sum\limits_{i=1}^{K} p_i t_i e^{\theta t_i}}{\sum\limits_{i=1}^{K} p_i e^{\theta t_i}}, \tag{3.37}$$

$$\mathrm{Var}_\theta(S(T)) = \frac{\partial^2 h}{\partial \theta^2}(\theta) = \frac{\left(\sum\limits_{i=1}^{K} p_i e^{\theta t_i}\right)\left(\sum\limits_{i=1}^{K} p_i t_i^2 e^{\theta t_i}\right) - \left(\sum\limits_{i=1}^{K} p_i t_i e^{\theta t_i}\right)^2}{\left(\sum\limits_{i=1}^{K} p_i e^{\theta t_i}\right)^2}, \tag{3.38}$$

$$I(\theta) = \mathrm{Var}_\theta(S(T)) \tag{3.39}$$

and the function $\dfrac{\partial h}{\partial \theta} : (-\infty, \infty) \longrightarrow (-1, 1)$ is a diffeomorphism.

Proof. We start first by proving (3.37).

$$\begin{aligned} \mathbb{E}_\theta(S(T)) &= \int_0^1 s(t) c_\theta(t) dt \\ &= \int_0^{p_1} t_1 \frac{e^{\theta t_1}}{a} dt + \ldots + \int_{p_1+\ldots+p_K}^{p_1+\ldots+p_K} t_K \frac{e^{\theta t_K}}{a} dt \\ &= \frac{\sum\limits_{i=1}^{K} p_i t_i e^{\theta t_i}}{\sum\limits_{i=1}^{K} p_i e^{\theta t_i}} = \frac{\partial h}{\partial \theta}(\theta). \end{aligned}$$

As for the proof of (3.38) we have,

$$\begin{aligned}\mathrm{Var}_\theta(S(T)) &= \int_0^1 S(t)^2 c_\theta(t)dt - \mathbb{E}_\theta(S(T))^2 \\ &= \int_0^{p_1} {t_1}^2 \frac{e^{\theta t_1}}{a}dt + \ldots + \int_{p_1+\ldots+p_{K-1}}^{p_1+\ldots+p_K} {t_K}^2 \frac{e^{\theta t_K}}{a}dt - \mathbb{E}_\theta(S(T))^2 \\ &= \frac{\sum_{i=1}^K p_i t_i^2 e^{\theta t_i}}{\sum_{i=1}^K p_i e^{\theta t_i}} - \left(\frac{\sum_{i=1}^K p_i t_i e^{\theta t_i}}{\sum_{i=1}^K p_i e^{\theta t_i}}\right)^2 = \frac{\partial^2 h}{\partial \theta^2}(\theta).\end{aligned}$$

Equation (3.39) is finally proved as

$$I(\theta) = \mathrm{Var}_\theta\left(\frac{\partial \log c_\theta(T)}{\partial \theta}\right) = \mathrm{Var}_\theta\left(S(T) - \frac{\partial h}{\partial \theta}\right) = \mathrm{Var}_\theta(S(T)).$$

□

3.6.2 Maximum likelihood estimation

We estimate in this section the parameter of interest θ using the ML method. The method gives an estimator with good asymptotic properties. The log-likelihood function of θ with observations $T_1^n, \ldots, T_k^n$ is defined as

$$\begin{aligned}L(\theta; T_1^n, \ldots, T_k^n) &= \frac{1}{k}\log\Big(\prod_{i=1}^k c_\theta(T_i^n)\Big) \\ &= \theta \bar{S}_k - h(\theta)\end{aligned}$$

with $\bar{S}_k$ the sufficient statistic defined by

$$\bar{S}_k = \frac{1}{k}\sum_{i=1}^k S(T_i^n).$$

The ML estimator $\hat{\theta}_k$ satisfies the score equation

$$\frac{\partial L}{\partial \theta}(\hat{\theta}_k; T_1^n, \ldots, T_k^n) = 0$$

which is equivalent to

$$\frac{\partial h}{\partial \theta}(\hat{\theta}_k) = \bar{S}_k.$$

We deduce

$$\hat{\theta}_k = \left(\frac{\partial h}{\partial \theta}\right)^{-1}\left(\bar{S}_k\right). \tag{3.40}$$

Theorem 3.4. *In the discrete log-linear trend model (3.36) and under conditions (3.4),(3.5),(3.6) and (3.7), the ML estimator (3.40) satisfies*

$$\sqrt{k}\left(\hat{\theta}_k - \theta_0\right) \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0, \frac{1}{I(\theta_0)}\right),$$

where $I(\theta)$ is the Fisher information defined in (3.39).

Proof. The proof relies on the asymptotic normality of the sufficient statistic $\bar{S}_k$ together with the delta-method. More precisely, we will prove below that

$$\sqrt{k}\left(\bar{S}_k - \mathbb{E}_{\theta_0}(S(T)\right) \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0, \mathrm{Var}_{\theta_0}(S(T))\right). \tag{3.41}$$

Then the delta-method entails

$$\sqrt{k}\left(\left(\frac{\partial h}{\partial \theta_0}\right)^{-1}(\bar{S}_k) - \left(\frac{\partial h}{\partial \theta_0}\right)^{-1}(\mathbb{E}_\theta(S(T))\right) \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0, \frac{1}{\left(\frac{\partial^2 h}{\partial {\theta_0}^2}(\theta_0)\right)^2}\mathrm{Var}_{\theta_0}(S(T))\right).$$

Thanks to Equations (3.37), (3.38), (3.39) and (3.40), this is equivalent to the announced result

$$\sqrt{k}\left(\hat{\theta}_k - \theta_0\right) \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0, \frac{1}{I(\theta_0)}\right).$$

We now focus on the proof of (3.41). Since $S(t)$ is piecewise constant, the sufficient statistic $\bar{S}_k$ can be written as

$$\begin{aligned}
\bar{S}_k &= \frac{1}{k}\sum_{i=1}^{k} S(T_i^n) = \int_0^1 S(t)d\hat{C}(t) \\
&= t_1\Big(\hat{C}(p_1) - \hat{C}(p_0)\Big) + \cdots + t_K\Big(\hat{C}(p_1 + \ldots + p_K) - \hat{C}(p_1 + \ldots + p_{K-1})\Big) \\
&= \sum_{i=1}^{K} t_i\Big(\hat{C}(q_i) - \hat{C}(q_{i-1})\Big)
\end{aligned}$$

with $q_0 = 0$, $q_i = p_1 + \cdots + p_i$, $i = 1, \ldots, K$.

Introducing the asymptotic expansion of $\hat{C}$, we get

$$\bar{S}_k = \sum_{i=1}^{K} t_i\Big(\Big(C(q_i) - C(q_{i-1})\Big) + \frac{1}{\sqrt{k}}\Big(B(C(q_i)) - B(C(q_{i-1}))\Big) + o_P(1/\sqrt{k})\Big),$$

whence we obtain

$$\begin{aligned}
\bar{S}_k &= \sum_{i=1}^{K} t_i\Big(C(q_i) - C(q_{i-1})\Big) + \frac{1}{\sqrt{k}} t_i\Big(B(C(q_i)) - B(C(q_{i-1}))\Big) + o_P(1/\sqrt{k}) \\
&= \mathbb{E}_{\theta_0}(S(T)) + \frac{1}{\sqrt{k}}\sum_{i=1}^{K} t_i\Big(B(C(q_i)) - B(C(q_{i-1}))\Big) + o_P(1/\sqrt{k}).
\end{aligned}$$

It follows

$$\sqrt{k}\left(\bar{S}_k - \mathbb{E}_{\theta_0}(S(T)\right) = \sum_{i=1}^{K} t_i\Big(B(C(q_i)) - B(C(q_{i-1}))\Big) + o_P(1).$$

In order to prove (3.41), it remains to show that

$$\sum_{i=1}^{K} t_i\Big(B(C(q_i)) - B(C(q_{i-1}))\Big) \overset{d}{=} \mathcal{N}\left(0, \mathrm{Var}_{\theta_0}(S(T))\right).$$

Since the left hand side is a centered Gaussian variable, it is enough to identify the variance. Using the identities

$$\sum_{i=1}^{K} t_i\Big(B(C(q_i)) - B(C(q_{i-1}))\Big) = \sum_{i=1}^{K-1}(t_i - t_{i+1})B(C(q_i))$$

and

$$S(T) = \sum_{i=1}^{K-1}(t_i - t_{i+1})\,\mathbf{1}_{T\le q_i} + t_K,$$

we have to prove

$$\mathrm{Var}_{\theta_0}\left[\sum_{i=0}^{K-1}(t_i - t_{i+1})\,B(C(q_i))\right] = \mathrm{Var}_{\theta_0}\left[\sum_{i=1}^{K-1}(t_i - t_{i+1})\,\mathbf{1}_{T\le q_i}\right].$$

This follows from the bilinearity of the covariance and from the equality

$$\begin{aligned}\mathrm{Cov}_{\theta_0}\left[B(C(q_i)), B(C(q_j))\right] &= C_{\theta_0}(q_i)\wedge C_{\theta_0}(q_j) - C_{\theta_0}(q_i)C_{\theta_0}(q_j)\\ &= \mathrm{Cov}_{\theta_0}\left[\mathbf{1}_{T\le q_i}, \mathbf{1}_{T\le q_j}\right]\end{aligned}$$

for all $1 \le i, j \le K-1$. □

3.6.3 Variance comparison between the MLE and the estimator by de Haan et al.

The discrete log-linear trend model was introduced by de Haan et al. (2015). The authors proposed three different estimators of the trend parameter θ_0 and it was found that their second estimator $\hat{\theta}_k^{(2)}$ was superior. Here we propose a comparison of the asymptotic variances of $\hat{\theta}_k^{(2)}$ and of our maximum likelihood estimator $\hat{\theta}_k^{ML}$. It is to be noted that the two estimators differ in the selection of the top order statistics from the whole sample. In de Haan et al. (2015), the largest k' observations at each time $t_1, \ldots, t_K$ are selected so that the effective sample size is $k = k' \times K$. Here we use the k largest observations of the whole sample, without stratification at the different times.

In the case $p_1 = \cdots = p_K$, it is shown in de Haan et al. (2015) that $\hat{\theta}_k^{(2)}$ is asymptotically normal with limit variance

$$\begin{aligned}&\frac{K}{\left(\sum_{i=1}^{K} {t_i}^2\right)^2}\left[\frac{\gamma^2+1}{K}\left(\sum_{i=1}^{K} t_i\frac{1-e^{\theta_0\gamma t_i}-\theta_0\gamma t_i}{\gamma^2}\right)^2 + \sum_{i=1}^{K} {t_i}^2\right.\\ &\qquad\left. + \left(\sum_{i=1}^{K} t_i e^{-\theta_0\gamma t_i}\right)^2 + (\gamma^2+2)\left(\sum_{i=1}^{K} t_i\frac{1-e^{-\theta_0\gamma t_i}}{\gamma}\right)^2\right]\end{aligned} \tag{3.42}$$

which we compare with the asymptotic variance of $\hat{\theta}_k^{ML}$ given by

$$\frac{\left(\sum_{i=1}^{K} p_i e^{\theta_0 t_i}\right)^2}{\left(\sum_{i=1}^{K} p_i e^{\theta_0 t_i}\right)\left(\sum_{i=1}^{K} p_i t_i^2 e^{\theta_0 t_i}\right) - \left(\sum_{i=1}^{K} p_i t_i e^{\theta_0 t_i}\right)^2}. \tag{3.43}$$

In Figure 3.2 , we plot on the same graph the quantities (3.42) and (3.43). Following de Haan et al. (2015), the selected parameters are $K = 17$, $\gamma = 1/2$, $t_i = i/K$ and $p_i = 1/K$, $i = 1, \ldots, K$. We observe that the asymptotic variance of $\hat{\theta}_k^{ML}$ is symmetric and always lower than the one of $\hat{\theta}_k^{(2)}$.

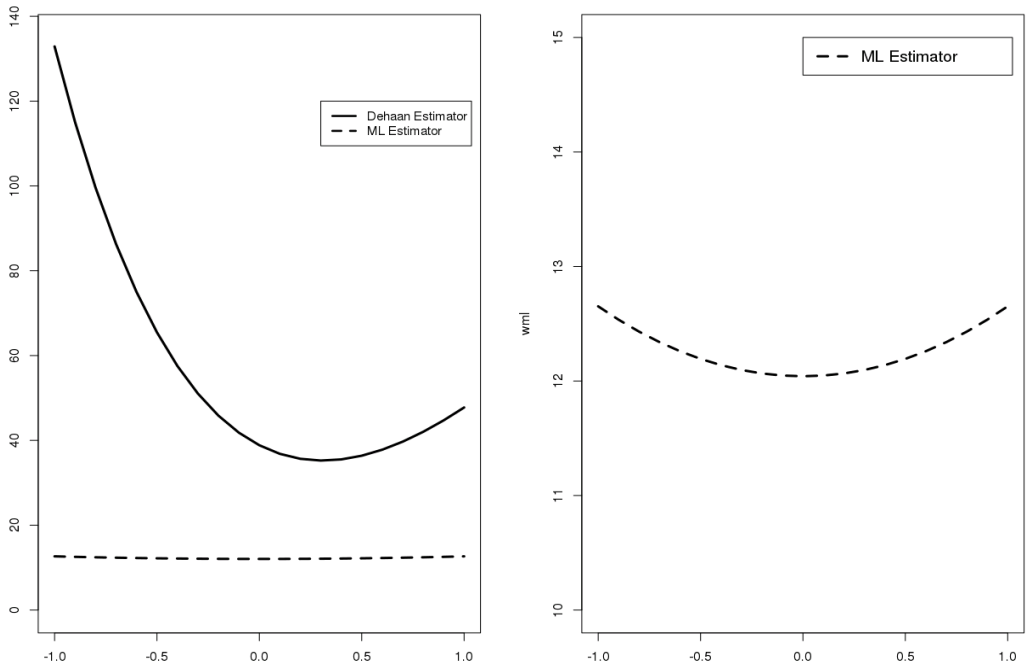


Figure 3.2 – Asymptotic variance comparison between $\hat{\theta}_k^{(2)}$ and $\hat{\theta}_k^{ML}$ (left) and asymptotic variance of $\hat{\theta}_k^{ML}$ (right) in the discrete log-linear model with $K = 17$, $\gamma = \frac{1}{2}$, $t_k = \frac{k}{K}$, $p_k = \frac{1}{K}$, $k = 1, \ldots, K$.

3.7 Simulation study

We propose simulation studies to evaluate the performances of the estimator of the trend parameter and extreme value index in the log-linear and linear trend models (3.16) and (3.24) respectively. Two different questions are investigated in two different simulation studies:

i) The issue of the threshold selection is an important and recurring issue in extreme value theory. The estimators of the trend parameter and the extreme value index are based on the top k order statistics of the sample and the asymptotic properties are proved for $k = k(n)$ an intermediate sequence satisfying (3.4) and (3.6). However in practice, the choice of k has an important impact of a finite sample estimator: selecting a larger number k of top order statistics leads to smaller variance since the estimator uses more observations, but this also increases the bias since the extreme value asymptotic holds only in the tail. We propose here to base the choice of k on Lepski's method that has been justified theoretically for the Hill estimator by Boucheron and Thomas (2015).

ii) We investigate the finite sample properties of the trend detection test proposed in Section 3.4.3 and compare its power with the power of nonparametric tests of Kolmogorov-Smirnov type proposed by Einmahl et al. (2016). We show that even if we deviate from the log-linear model, the parametric test advocated in Section 3.4.3 can exhibit a better power.

We now describe the data generating process used in our simulation study: let U_i, $i \geq 1$, be a sequence of i.i.d. random variables with uniform distribution on $[0,1]$ and define

$$X_i^n = \left(\frac{c(\frac{i}{n})}{U_i} \right)^{\frac{1}{\alpha}} + (1-p)^{-1/\alpha} \varepsilon_i, \tag{3.44}$$

where $\alpha > 0$ is the tail index, c is the skedasis function, $p \in (0,1)$ and ε_i is an i.i.d. noise. The first term in (3.44) has a Pareto distribution with tail index $\alpha > 0$ and scale parameter $c(i/n)^{1/\alpha}$. Provided the noise distribution ε has a tail lighter than α, that is $\lim_{x\to\infty} x^\alpha \mathbb{P}(\varepsilon > x) = 0$, then the extremes of X_i^n are given by the first Pareto term and the assumption (3.1) of the heteroscedastic model is satisfied with reference distribution F equal to the standard Pareto distribution (see Lemma 3.3 below). The term $(1-p)^{-1/\alpha}\varepsilon_i$ appears as a nuisance term that does not affect the tail asymptotics but that will come into play in finite sample study. The parameter p controls the strength of the nuisance: the larger the value of p, the larger the perturbation. The order of magnitude of the nuisance term $(1-p)^{-1/\alpha}\varepsilon_i$ is the p-th quantile of F and for this reason, we expect that a sensible choice for k is $k = (1-p)n$. Heuristically, only the lower p-th fraction of the sample is affected by the nuisance term. The results of the simulation study are quite robust with respect to the distribution of ε and for simplicity, we use here the uniform distribution.

Lemma 3.3 (Embrechts et al. (1979), Proposition 1)**.** *If X and Y are independent and such that*

$$\mathbb{P}(X > x) = x^{-\alpha}\ell(x) \quad \textit{and} \quad \lim_{x\to\infty} \frac{\mathbb{P}(Y > x)}{\mathbb{P}(X > x)} = 0$$

for some $\alpha > 0$ and slowly varying function ℓ, then

$$\mathbb{P}(X + Y > x) \underset{x\to\infty}{\sim} \mathbb{P}(X > x).$$

3.7.1 Lepski's method for the choice of k

Recently, Boucheron and Thomas (2015) advocated the use of Lepski's method for threshold selection in extreme value theory and provided theoretical guarantees for the Hill estimator of the extreme value index. The method works as follows: for each value of k, the Hill estimator $\hat{\gamma}_k$ is computed as well as the confidence interval

$$\hat{\gamma}_k^{\pm} = \hat{\gamma}_H \pm \sqrt{2.1 \log\log k} \frac{\hat{\sigma}_\gamma}{\sqrt{k}} \tag{3.45}$$

where $\hat{\sigma}_\gamma = \hat{\gamma}_k$ is an estimator of the asymptotic variance (recall that $\sqrt{k}(\hat{\gamma}_k - \gamma) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \gamma^2)$) and $\sqrt{2.1\log\log k}$ is a non-standard penalty term justified theoretically

by Lepski's method. The confidence intervals (3.45) shrink when k increases and Lepski's method recommends to select k as the largest value such that $\hat{\gamma}_k$ belongs to all the previous confidence intervals $(\hat{\gamma}_l^-, \hat{\gamma}_l^+)$, $l = 1, \ldots, k$. More formally, we define

$$k^* = \max \left\{ k \leq n : \hat{\gamma}_k \in [\max_{l \leq k} \gamma_l^-, \min_{l \leq k} \gamma_l^+] \right\}.$$

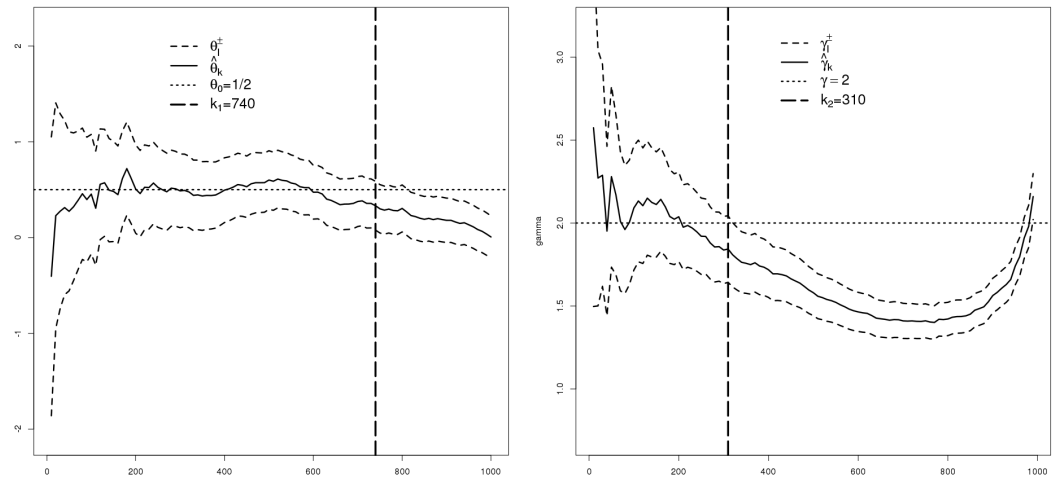
In our case, we estimate simultaneously the trend parameter θ of the skedasis function in the log-linear model (3.16) or the linear model (3.24) and the extreme value index, so we define furthermore the confidence intervals for θ by

$$\hat{\theta}_k \pm \sqrt{2.1 \log(\log(k))} \frac{\hat{\sigma}_\theta}{\sqrt{k}}, \tag{3.46}$$

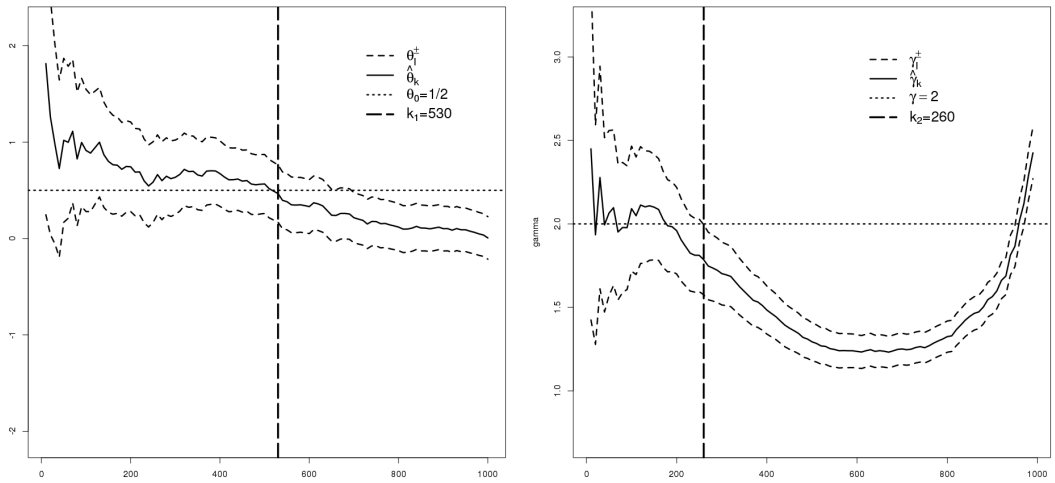
where $\hat{\sigma}_\theta^2$ is a consistent estimator of the asymptotic variance of $\hat{\theta}_k$ obtained by plugging $\hat{\theta}_k$ in the formula for the asymptotic variance $\sigma^2(\theta_0)$ of $\hat{\theta}_k$. We propose to select the threshold $k^* = \min(k_1^*, k_2^*)$ with

$$k_1^* = \max \left\{ k \leq n : \hat{\theta}_k \in [\max_{l \leq k} \theta_l^-, \min_{l \leq k} \theta_l^+] \right\},$$
$$k_2^* = \max \left\{ k \leq n : \hat{\gamma}_k \in [\max_{l \leq k} \gamma_l^-, \min_{l \leq k} \gamma_l^+] \right\}.$$

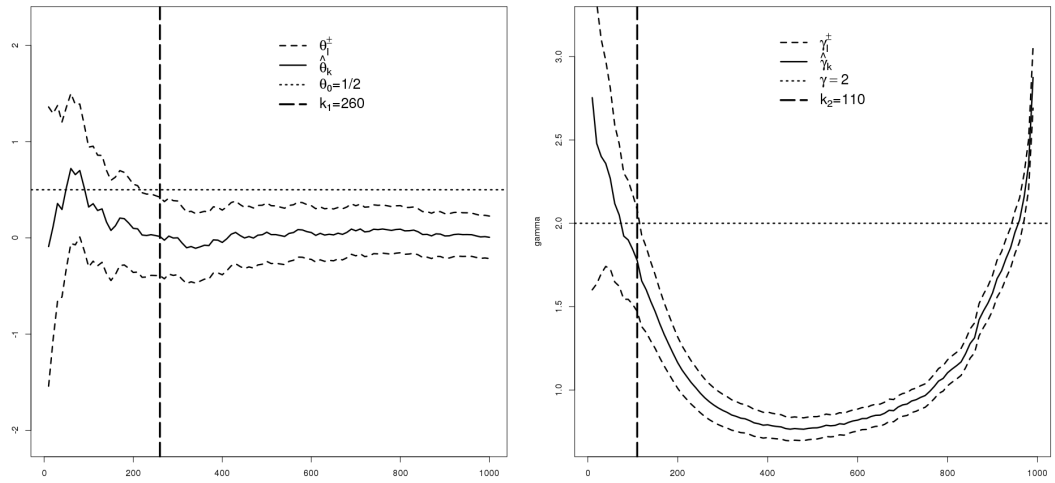
Our simulation study is run with $n = 1000$ observations of the data generating process (3.44) with $\alpha = 1/2$ and skedasis function $c = c_{\theta_0}$ either from the log-linear model (3.16) or from the linear model (3.24) with $\theta_0 = 1/2$ in both cases. The value of the nuisance parameter p is set to $p = 0.6$ (small noise), $p = 0.75$ (intermediate noise) or $p = 0.9$ (strong noise). Lepski's method is used to select an adequate threshold k and the results are plotted in Figure 3.3 (log-linear model), Figure 3.4 (linear model with moment estimator) and Figure 3.5 (linear model with ML estimator). Each pair of plots represents as a function of k the estimated value of θ (left) or γ (right) together with the corresponding confidence interval from Lepski's method. The vertical line represents the value of k selected by Lepski's method while the horizontal line corresponds to the theoretical value of the parameter. The main common feature for these different plots is that for higher values of p, that is for higher levels of tail perturbation, the selected threshold gets lower. This is in agreement with the intuition that a nicer tail behavior should correspond to a larger number of selected extreme observations.



(a) p=0.6

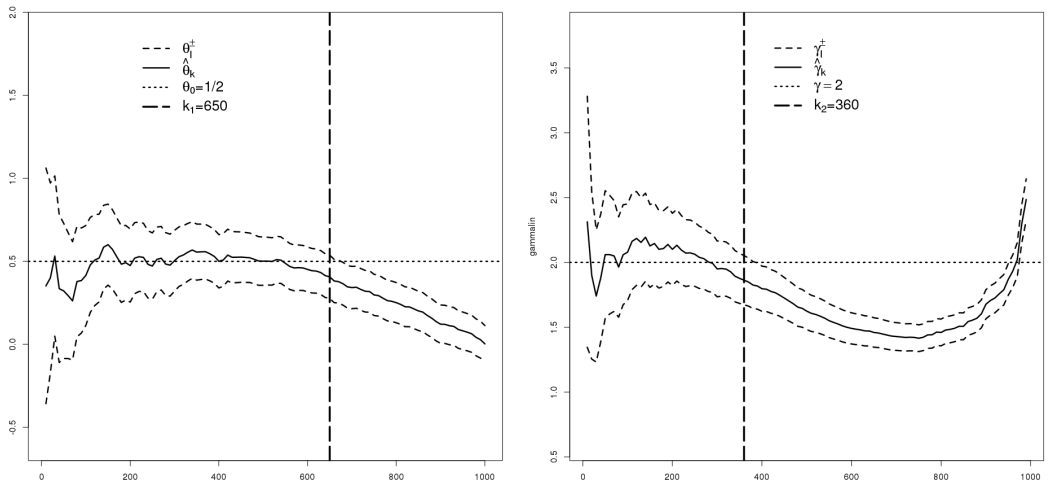


(b) p=0.75

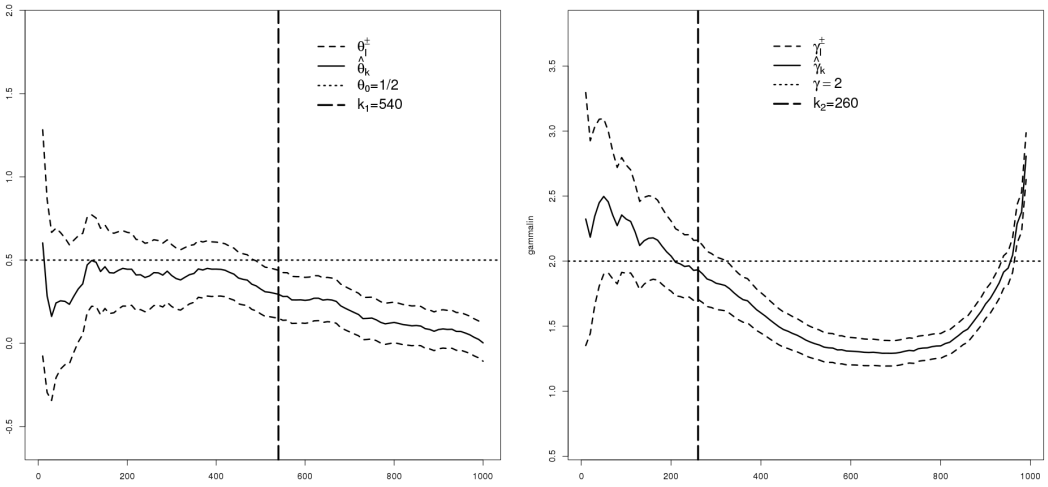


(c) p=0.9

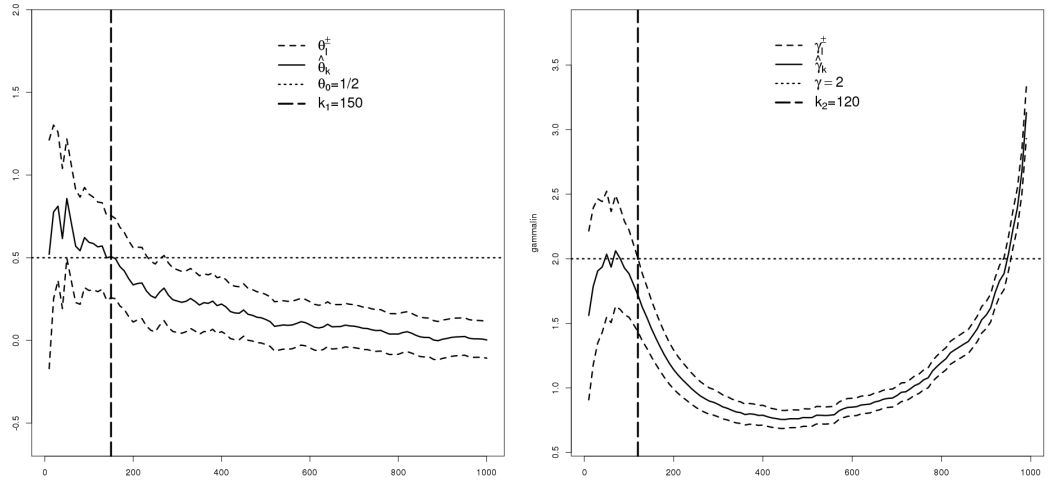
Figure 3.3 – Lepski's method for the selection of k, log-linear model.



(a) p=0.6

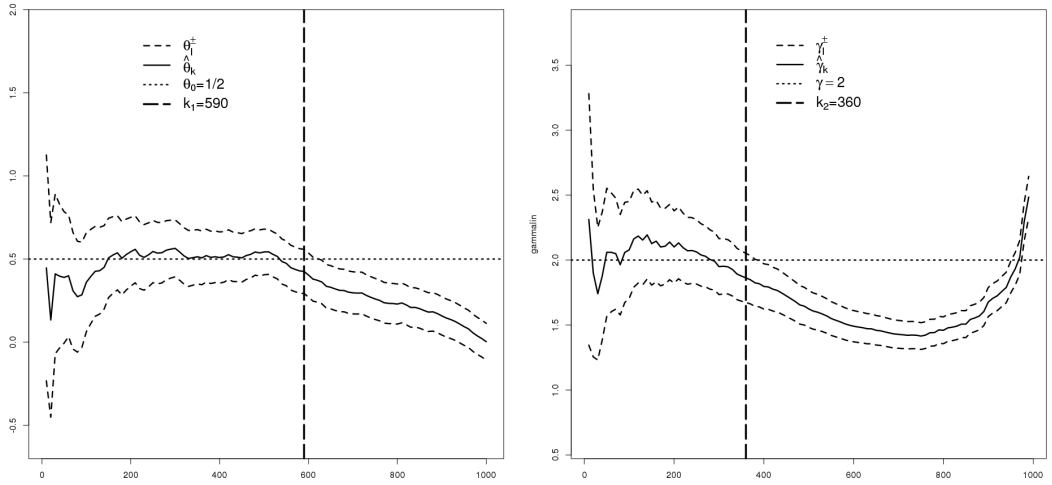


(b) p=0.75

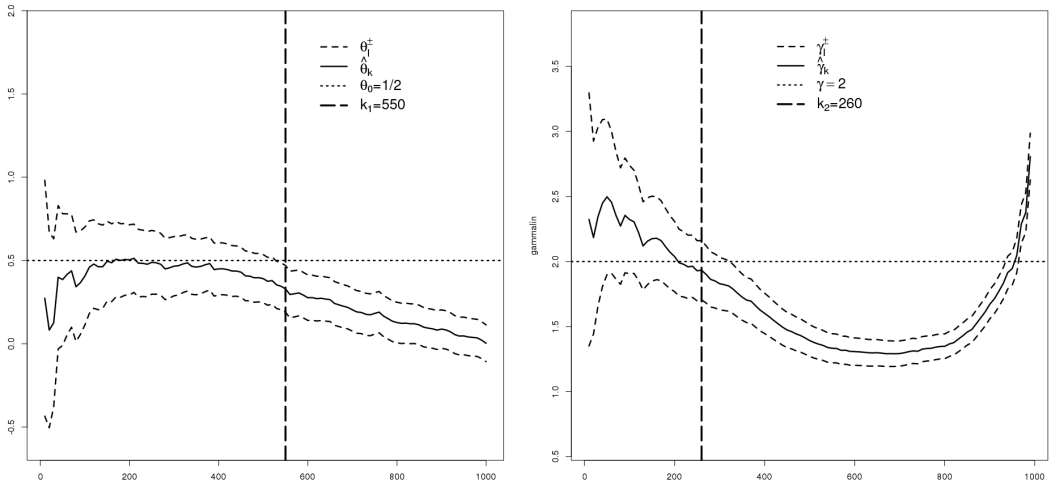


(c) p=0.9

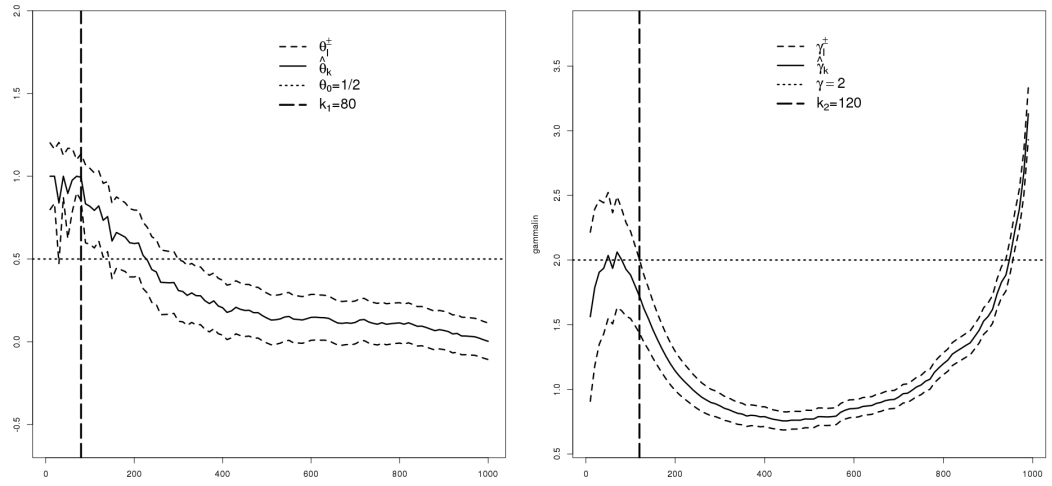
Figure 3.4 – Lepski's method for the selection of k, linear model with moment estimator.



(a) p=0.6



(b) p=0.75



(c) p=0.9

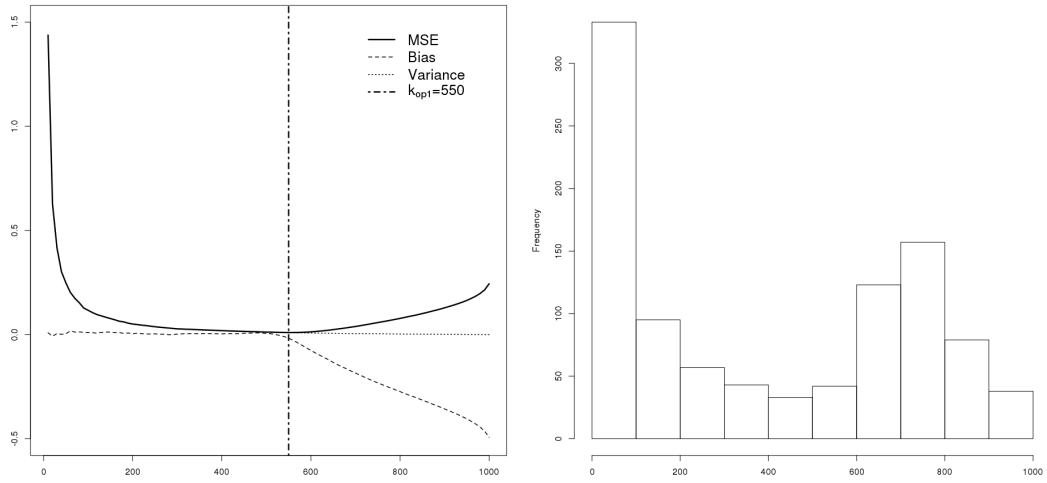
Figure 3.5 – Lepski's method for the selection of k, linear model with ML estimator.

In order to study further the performance of Lepski's method, we conduct a simulation study that compares the distribution of the selected threshold k and the estimator mean square error (MSE). We expect that the selected threshold distribution concentrates in a neighborhood where the MSE is closed to its minimum.

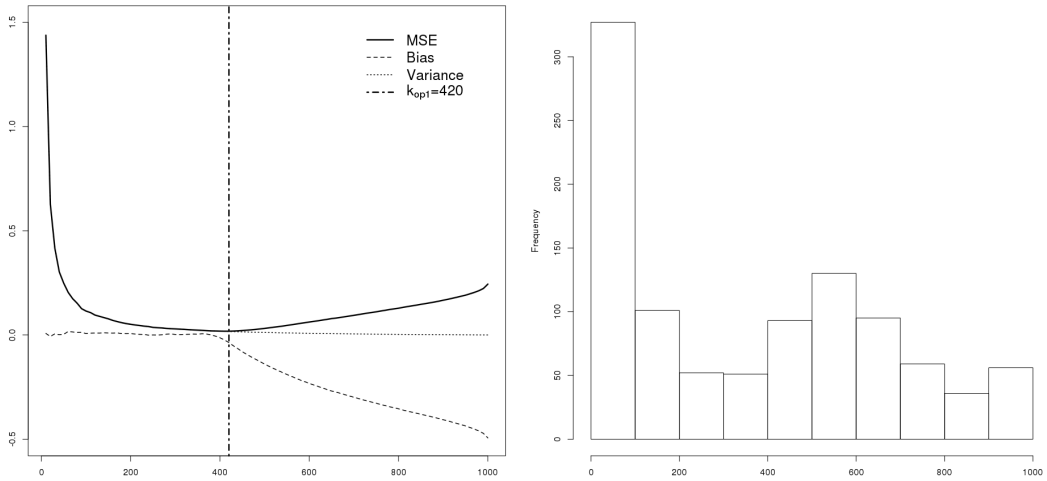
The simulation study consists in $N = 1000$ repetitions of the procedure leading to Figures 3.3, 3.4 and 3.5. Based on these N simulations, we assess the estimator bias, variance and mean square error as well as the distribution of the selected threshold k. The plots of the MSE for $\hat{\theta}$ as well as the histogram of the chosen values of k using the Lepski's method for the log-linear, linear (ML) and linear (moments) models are presented in Figures 3.6, 3.7 and 3.8 respectively. The plots of the MSE for $\hat{\gamma}$ as well as the histogram of k using the Lepski's method for the log-linear and linear models are presented in Figures 3.9 and 3.10 respectively.

The plots for the MSE are quite standard: the variance dominates for small k and the bias for large k; the MSE reaches a minimum for a suitable bias/variance trade off. Interestingly, the value of k that minimizes the MSE for the trend parameter θ depends quite strongly on the parameter p measuring the noise in the tail (vertical line in the left plots from Figures 3.6, 3.7, 3.8), while it is much more robust for the Hill estimator (vertical line in the left plots from Figures 3.9, 3.10).

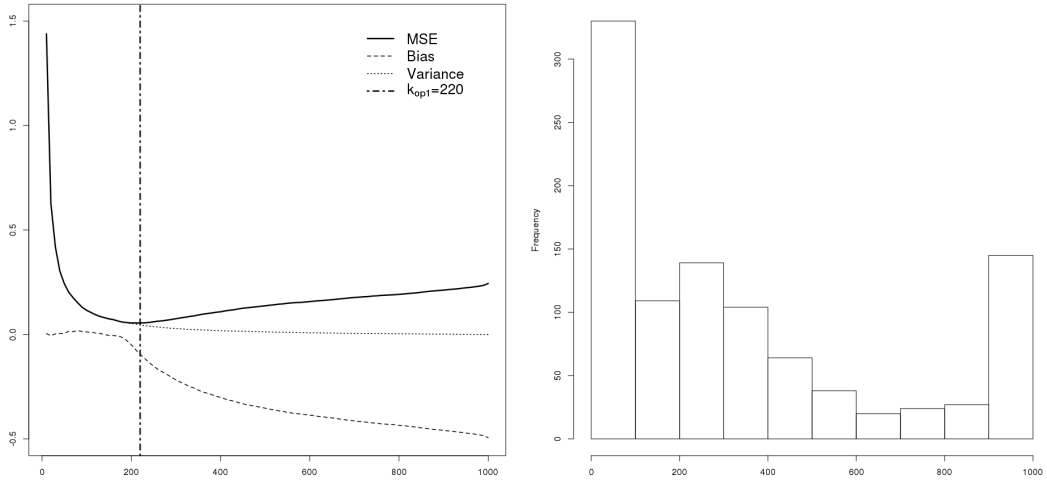
The histograms corresponding to the distribution of the selected threshold k show a strong bi-modality. A first mode corresponds to small values of k, indicating that quite often Lepski's method will return a too low threshold. Interestingly, the second mode appears to be quite close to the MSE minimizer, with a tendency to slightly overestimating it. In any case, the distribution of k is quite spread out and not really concentrated near the MSE minimizer.



(a) p=0.6

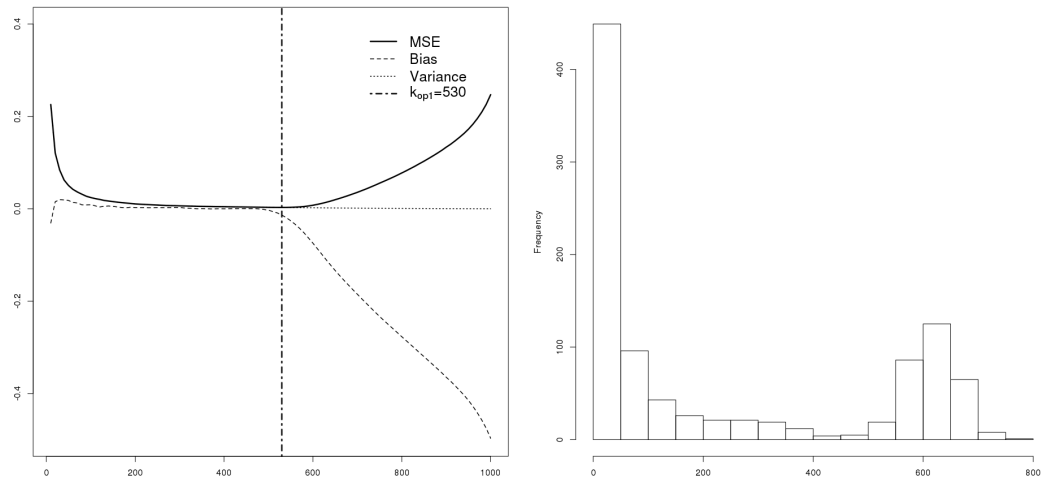


(b) p=0.75

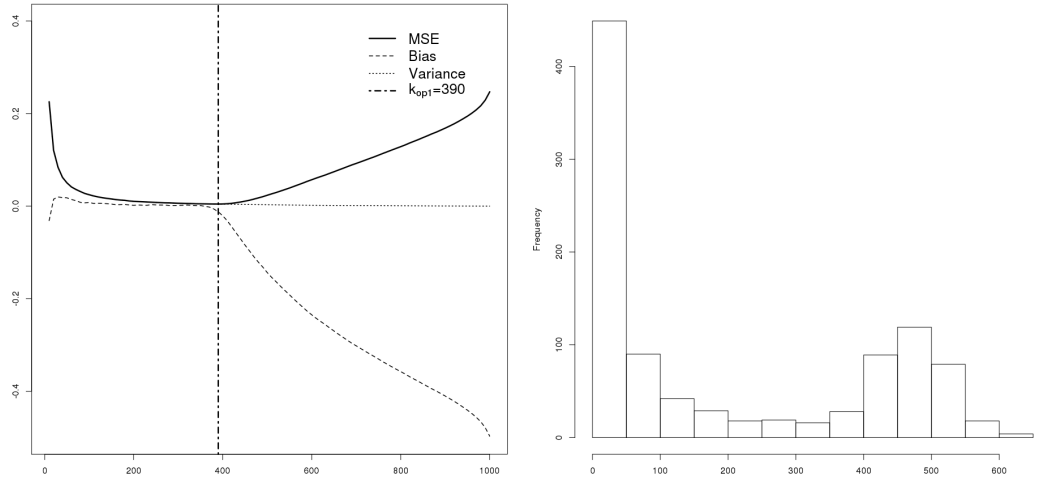


(c) p=0.9

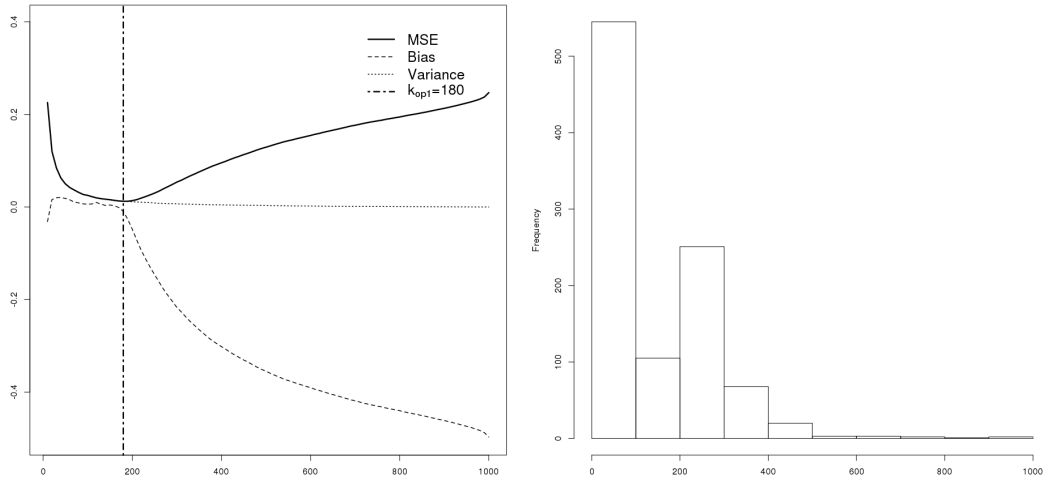
Figure 3.6 – MSE and histogram of k for $\hat{\theta}$ in the log-linear model.



(a) p=0.6

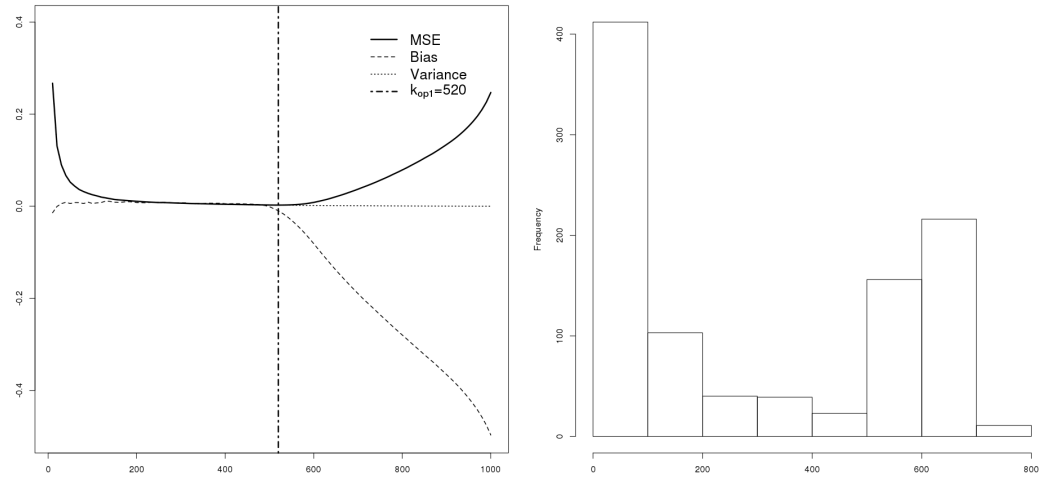


(b) p=0.75

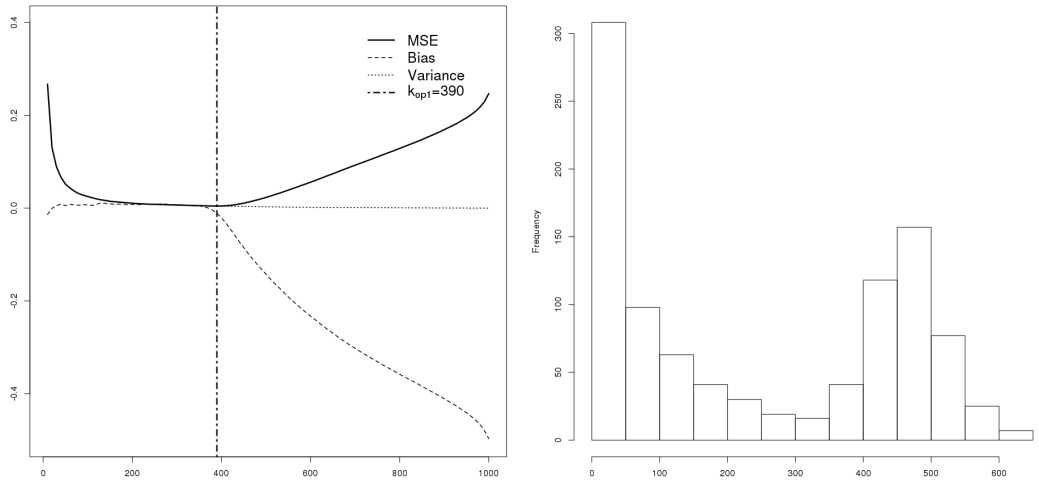


(c) p=0.9

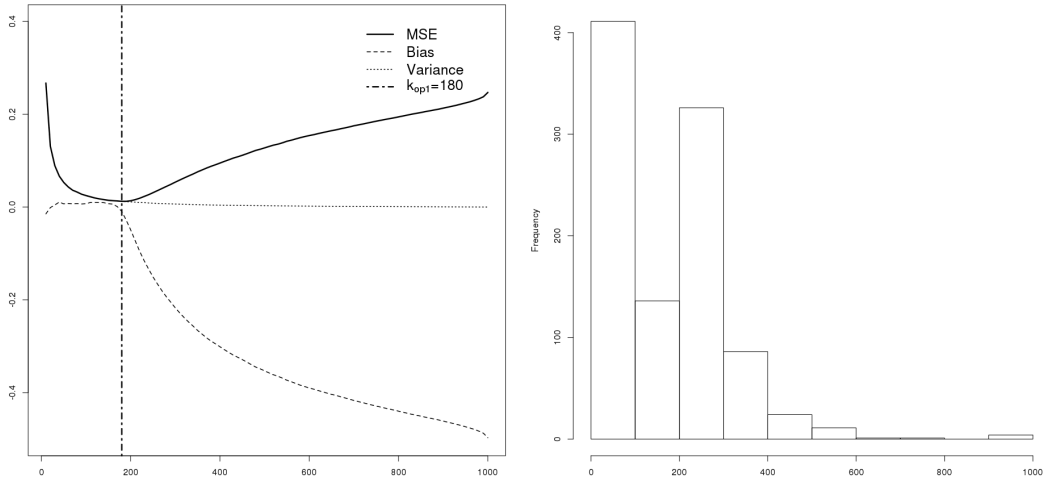
Figure 3.7 – MSE and histogram of k for $\hat{\theta}$ in the linear model (ML).



(a) p=0.6

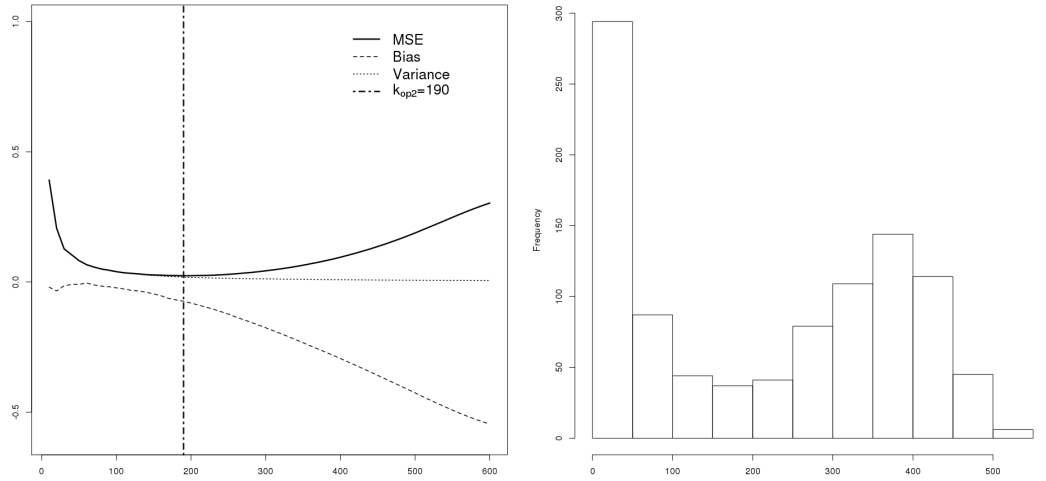


(b) p=0.75

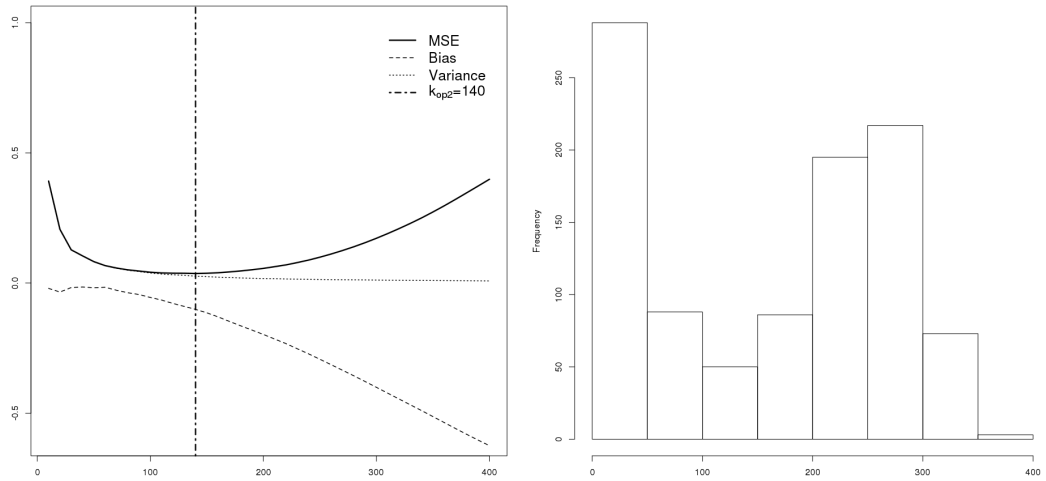


(c) p=0.9

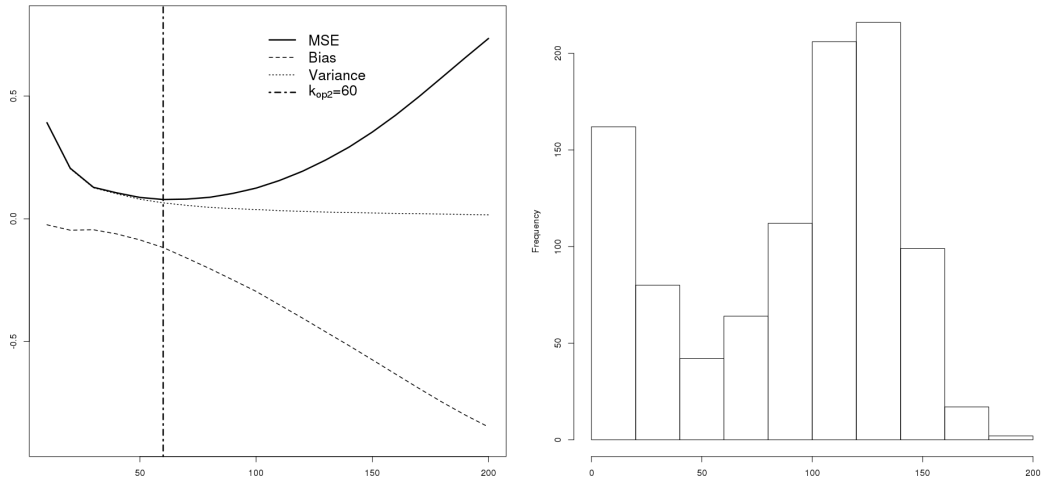
Figure 3.8 – MSE and histogram of k for $\hat{\theta}$ in the linear model (moment).



(a) p=0.6

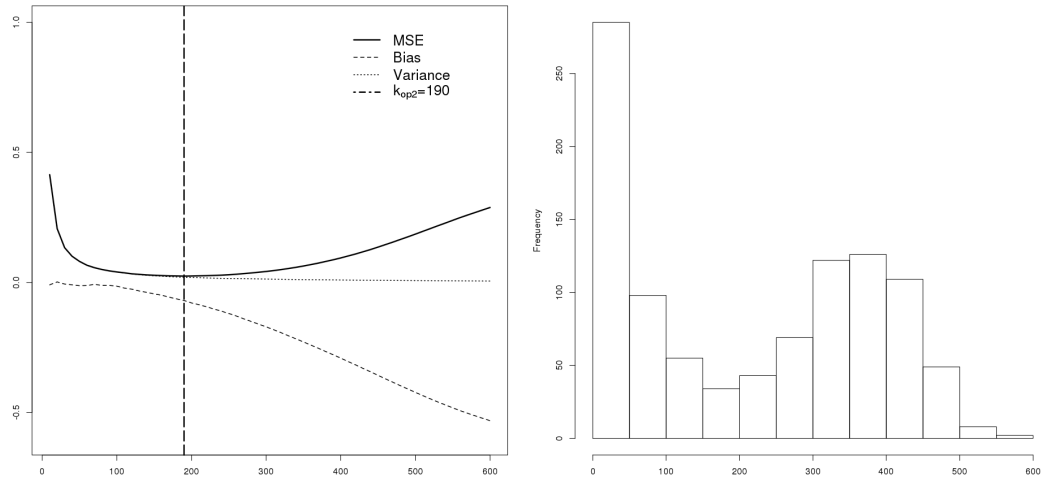


(b) p=0.75

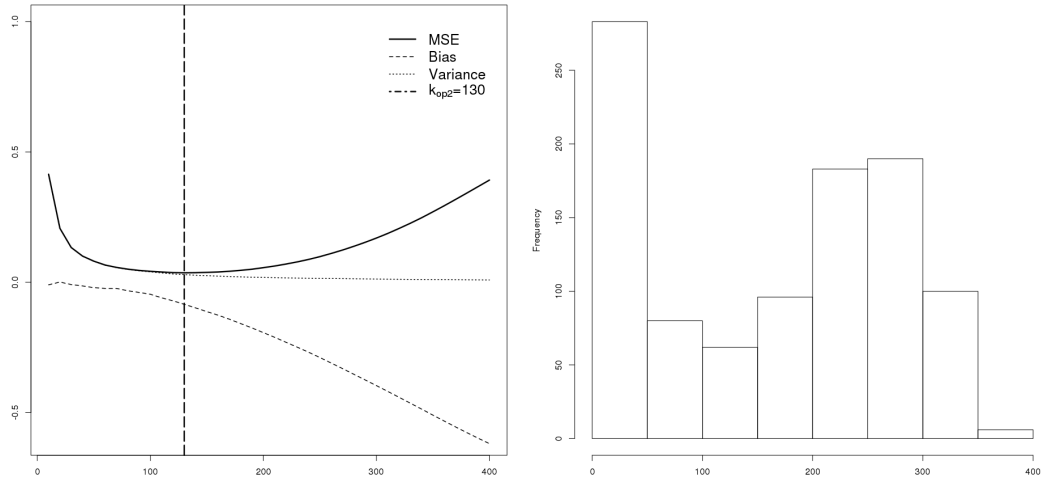


(c) p=0.9

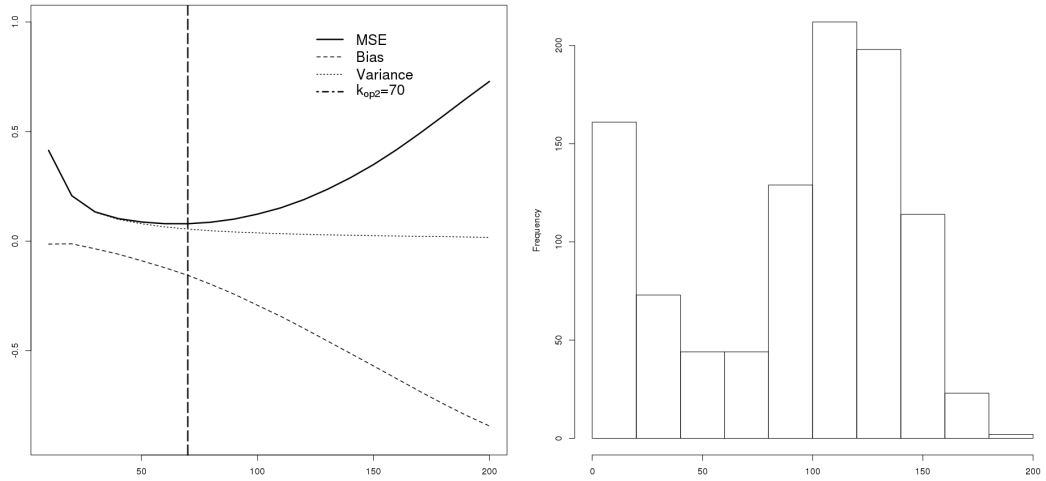
Figure 3.9 – MSE and histogram of k for $\hat{\gamma}$ in the log-linear model.



(a) p=0.6



(b) p=0.75



(c) p=0.9

Figure 3.10 – MSE and histogram of k for $\hat{\gamma}$ in the linear model.

3.7.2 Power of tests

A natural question that can be asked is what is the benefit of considering parametric models in heteroscedastic extremes while a non parametric study has been proposed by Einmahl et al. (2016). One reason is that parametric tests are generally more powerful than non parametric tests when the model is well specified. In the framework of trend detection in extremes, we have seen in Proposition 3.4 that the parametric test is justified even for misspecified skedasis function, under the mere monotonicity assumption.

We first perform a short simulation study to assess the power of the parametric test T_1: $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$ in the log-linear model with $\theta_0 = (0, 0.1, 0.3, 0.5)$. For different sample sizes $n = (500, 1000, 5000, 10000)$, we consider n observations from the data generating process with $\alpha = \frac{1}{2}$ and $p = 0.75$. The probability of acceptance of the parametric test from Proposition 3.4 is assessed thanks to $N = 1000$ independent repetitions of the procedure. The number of selected order statistics k is 10%, 5% or 1% of the sample size. The results are summarized in Table 3.1. For $\theta_0 = 0$, the probability of acceptance fluctuates at about 95% . As expected, the acceptance probability decreases as θ_0 increases, that is in the presence of a stronger trend. Note that the power of the test depends mostly on k, the number of selected extreme observations, more than on n, the total sample size.

		$\theta_0 = 0$	$\theta_0 = 0.1$	$\theta_0 = 0.3$	$\theta_0 = 0.5$
$n = 500$	$k = 50$	0.971	0.957	0.914	0.823
	$k = 25$	0.948	0.950	0.906	0.874
	$k = 5$	0.898	0.906	0.905	0.894
$n = 1000$	$k = 100$	0.961	0.944	0.873	0.679
	$k = 50$	0.947	0.952	0.912	0.824
	$k = 10$	0.941	0.940	0.924	0.911
$n = 5000$	$k = 500$	0.968	0.923	0.521	0.087
	$k = 250$	0.972	0.920	0.747	0.351
	$k = 50$	0.946	0.940	0.899	0.822
$n = 10000$	$k = 1000$	0.975	0.850	0.216	0.002
	$k = 500$	0.952	0.899	0.500	0.097
	$k = 100$	0.936	0.929	0.866	0.699

Table 3.1 – Estimated acceptance probabilities for the parametric test for trend detection in the case of log-linear skedasis function.

In a second simulation study, we compare the parametric test T_1 with the non parametric Kolmogorov-Smirnov test T_2 proposed by Einmahl et al. (2016) whose statistical significance level is 5%. For the purpose of a fair comparison, we do not use the log-linear trend model, but rather the skedasis function

$$c(s) = 1 - \delta \cos(\pi s), \quad s \in [0, 1],$$

where $\delta = (0, 0.2, 0.4, 0.6)$. Note that this skedasis function is non increasing with $\delta = 0$ corresponding to the homoscedastic case. We use a sample of size $n = 1000$

from the data generating process (3.44) with $\alpha = 0.5$ and $p = 0.75$. The results are summarized in Table 3.2.

		$k = 100$	$k = 200$	$k = 300$
$\delta = 0$	T_1	0.963	0.975	0.982
	T_2	0.972	0.988	0.988
$\delta = 0.2$	T_1	0.697	0.479	0.285
	T_2	0.755	0.601	0.428
$\delta = 0.4$	T_1	0.159	0.005	0.000
	T_2	0.238	0.017	0.000
$\delta = 0.6$	T_1	0.004	0.000	0.000
	T_2	0.014	0.000	0.000

Table 3.2 – Estimated acceptance probabilities for the parametric test T_1 and non parametric test T_2 for trend detection. The sample is of size $n = 1000$ with skedasis function $c(s) = 1 - \delta \cos(\pi s)$.

We remark that when $\delta = 0$ the probability of accepting H_0 is larger than 95%, so that the tests are slightly conservative. In the presence of a trend in extremes ($\delta \neq 0$), we see that the parametric test T_1 has always a lower acceptance probability than the Kolmogorov-Smirnov test T_2, indicating a higher power for trend detection in extremes. Clearly, the power is better for larger values of k, i.e. a larger effective sample, and higher values of δ, i.e. a stronger trend. We can conclude that the probability that the parametric tests correctly reject the null hypothesis H_0 when the alternative hypothesis H_1 is true is higher than the one for non parametric tests. This is why we say that the parametric test T_1 is more powerful than the non parametric test T_2 and it is important to use it in practice.

3.8 Application

We want now to apply our results on maximum and minimum daily temperatures (measured in degrees Fahrenheit) in Fort Collins, Colorado, USA from 1900 to 1999. The data is freely available in the R package **extRemes** which provides in addition the daily accumulated precipitation, snow amount and cover amount. The total sample size of our data, the maximum and minimum daily temperatures, is $n = 36524$. The aim is to test the presence of a trend in extremes for these two variables that could be associated with a global warming effect during the 20th century. Due to possible non positive extreme value index, we do not use here the Hill estimator but rather the Dekkers-Einmahl-deHaan estimator (de Haan and Ferreira (2006)) defined as the moment estimator

$$\hat{\gamma}_M := M_n^{(1)} + 1 - \frac{1}{2}\left(1 - \frac{\left(M_n^{(1)}\right)^2}{M_n^{(2)}}\right)^{-1}, \tag{3.47}$$

where

$$M_n^{(j)} := \frac{1}{k}\sum_{i=0}^{k-1}(\log X_{n-i:n} - \log X_{n-k:n})^j.$$

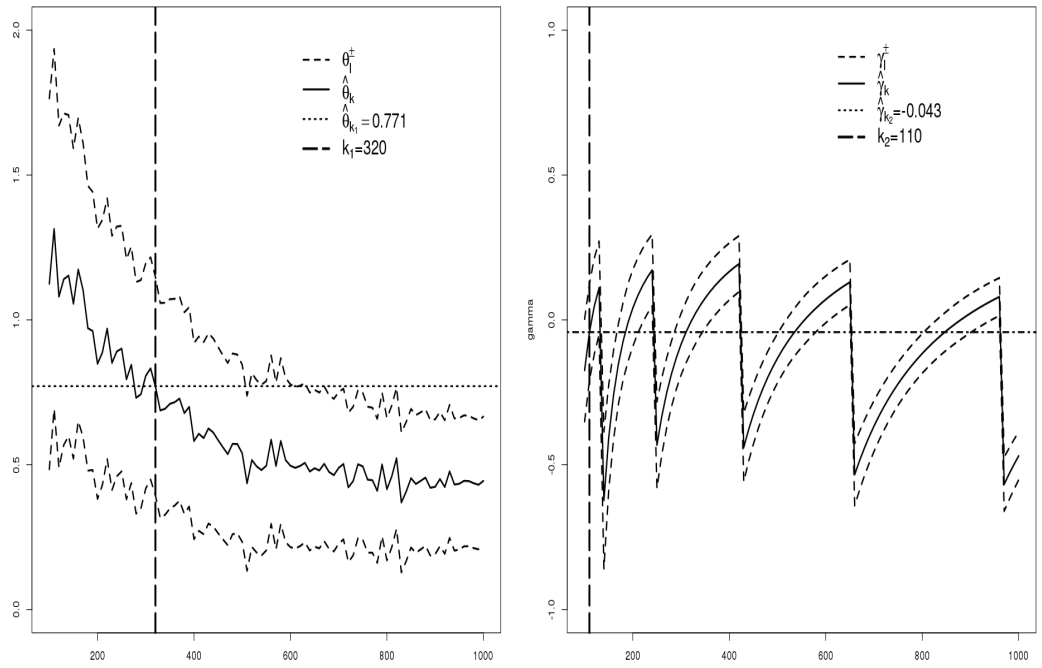
Note that asymptotic normality is satisfied

$$\sqrt{k}(\hat{\gamma}_M - \gamma) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, var_\gamma)$$

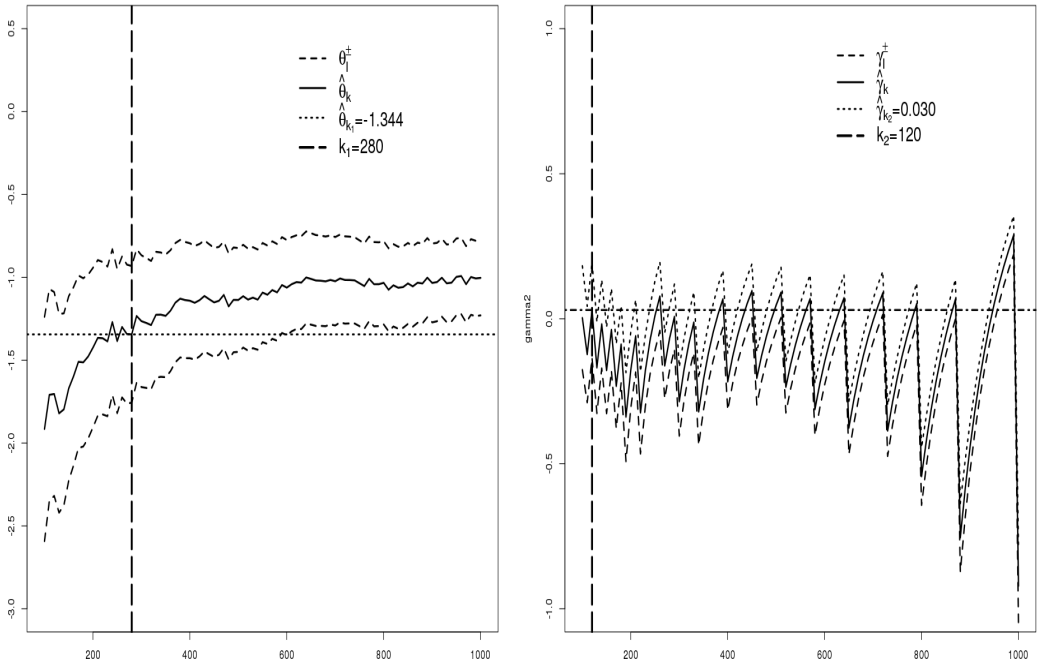
with

$$var_\gamma = \begin{cases} \gamma^2 + 1 & \text{if } \gamma \geq 0, \\ \dfrac{(1-\gamma^2)(1-2\gamma)(1-\gamma+6\gamma^2)}{(1-3\gamma)(1-4\gamma)} & \text{if } \gamma < 0. \end{cases}$$

The estimates of the trend parameter and extreme value index as functions of k are provided in Figure 3.11 as well as the selected threshold via Lepski's method. Note that the particular periodic shape for the estimator of γ is due to presence of many ties in the sample. Applying Lepski's method on the confidence intervals, we choose the optimal size k and consequently the corresponding estimators $\hat{\theta}_k$ and $\hat{\gamma}_k$. For the maximal temperature, we obtain $k = 110$, $\hat{\gamma}_k^{\max} = -0.043$ and $\hat{\theta}_k^{\max} = 1.314$. In order to deal with extremes of minimum daily temperature, we use the negative minimum daily temperature and obtain $k = 120$, $\hat{\gamma}_k^{\min} = 0.030$ and $\hat{\theta}_k^{\min} = -1.704$. The test proposed in Proposition 3.4 is strongly significant for rejecting the null hypothesis $H_0 : \theta_0 \leq 0$: indeed, the test statistic $\sqrt{12k}\hat{\theta}_k$ is compared to the quantile of the normal distribution 1.645 and is equal to 47.73 and 64.65 in the case of maximum and minimum daily temperatures respectively. In both cases, the p-value of the test is close to 0. The trend parameter $\hat{\theta}^{\max} > 0$ suggests that extreme (high values) of daily maximum temperatures are getting more frequent, while the trend parameter $\hat{\theta}^{\min} < 0$ indicates that extremes (low values) of daily minimum temperatures are getting less frequent. Hence both trend parameters corroborate global warming in the regime of extremes over the 20th century .



(a) Maximum daily temperatures



(b) Minimum daily temperatures

Figure 3.11 – Lepski's method for estimation of θ (left) and γ (right) related to the maximum and minimum daily temperatures

3.9 Conclusion

In the direction of heteroscedastic extremes, the study of the monotone skedasis function c allows us to conclude on the frequency of extreme events. In log-linear, linear and discrete log-linear trend models of c, the trend parameter θ is consistent and asymptotically normal. In the case where c is monotone but not parametric, it is possible to test the trend using a parametric test on θ and this will give more accurate results since parametric tests are more powerful than non parametric tests for the trend detection.

4

GEV modeling of precipitation of annual maxima and model selection for high quantile prediction

The chapter consists of a work devoted to the EVA 2017 Data challenge.

Abstract

Based on the daily precipitations over 24 years in 40 stations in the Netherlands, we predict the 20-year return level of precipitations for each month and station. We use the annual maximum method from extreme value theory and model the annual maximum precipitations by generalized extreme value distributions with scale and location parameters depending on the longitude, latitude and month. We propose different models and consider model selection based on the Akaike information criterion and k-fold cross validation. The latter method turns out to be much more efficient for the purpose of prediction and results in a fairly simple model with relatively good predictive power.

Contents

4.1 Introduction

Climate extremes is a long standing topic of interest, especially in the last decades. Extreme value theory (EVT) focuses on the modeling and statistical assessment of extreme events. One of the main issue in EVT is the estimation of an extreme quantile, or equivalently, a return level associated to large return period. The return level of T years corresponds to the value that will be exceeded on average once every T years.

This paper consists in a report of the participation to the data challenge held in the conference EVA 2017. The data consists in daily precipitation in the Netherlands over 40 stations over the period from 01/01/72 (with only 16 observations available in this year) to 12/31/16: the training sample period is from 01/01/72 to 12/31/95 and the test sample period is from 01/01/96 to 12/31/16. The aim is to estimate from the training sample a quantile of level corresponding to extreme monthly precipitation over the next 20 years, station by station. The general methodology that we used consists in taking annual maxima for each month and year (one observation per station, month and year) and to fit several generalized extreme value (GEV) models with covariates. In these types of models, the GEV parameters depend on latitude, longitude and month while the observations are considered as independent, that is we ignore completely the spatial dependence among the stations. For the data challenge participation, model selection had been performed via the simple Akaike information criterion (AIC). Discussions with the other participating teams suggested that k-fold cross validation should perform better as it incorporates the criterion on which the different methods are ranked. Accordingly, we have since then performed model selection based on k-fold cross validation and it turns out the selected model provides better predictions on the sample test.

The estimation is performed with the **extRemes** package in R (Gilleland and Katz (2016)). This package was designed to analyse extreme events and their changes over time. It focuses on geophysics with applications for weather and climate (Gilleland and Katz (2011)). It incorporates GEV, generalized Pareto (GP) and point process (PP) models. The functions used are **fevd** to estimate the model (with covariates) and compute AIC and **return.level** to estimate the quantiles.

4.2 The models

We compare several models for the dependence of the GEV parameters as functions of the covariates. For this purpose we use natural splines (James et al. (2013)), which offer a simple and flexible way for response surface modeling. Splines are piecewise polynomial functions, as presented in the following definition.

Definition 4.1. Let $\xi_1 < \xi_2 < \cdots < \xi_N$ be N knots. A spline with degree m is a function of class $C^{(m-1)}$ whose restriction on each of the intervals $(-\infty, \xi_1)$, (ξ_1, ξ_2), $\ldots$, $(\xi_N, +\infty)$ is a polynomial function with degree at most m.
A natural spline is a spline that is linear on the boundary intervals $(-\infty, \xi_1)$ and $(\xi_N, +\infty)$.

For the sake of simplicity, we assume the shape parameter γ of the GEV distribution constant (not depending on covariates); its estimation is indeed quite complex. The location parameter μ and the scale parameter σ are expressed in function of the covariates: month, latitude and longitude. We define three types of models for the covariates of the GEV parameters. The first one is the additive model where the two parameters μ and σ are expressed as follows:

$$\mu = month + f_1(latitude) + f_2(longitude),$$

$$\sigma = month + g_1(latitude) + g_2(longitude),$$

with f_1, f_2, g_1 and g_2 natural cubic spline functions with df degrees of freedom. In what follows we denote by Add_h the additive model where the natural splines have $df = h$.

The second model is the mixed model in which there is not interaction between the time and space covariates, that is between $month$ and $(latitude, longitude)$. However, we consider possible interactions between $latitude$ and $longitude$ and set:

$$\mu = month + f_1(latitude) \times f_2(longitude),$$

$$\sigma = month + g_1(latitude) \times g_2(longitude).$$

Let Mix_h denote the mixed model with $df = h$ for the natural splines.

The third model is the multiplicative model where space/time interaction appears in the model and μ and σ are defined as:

$$\mu = month \times f_1(latitude) \times f_2(longitude),$$

$$\sigma = month \times g_1(latitude) \times g_2(longitude).$$

We denote by $Mult_h$ the multiplicative model in which natural splines have $df = h$.

4.3 Akaike information criterion

In this section, we present the AIC model selection which was used for our participation to the data challenge in EVA 2017. After a brief presentation of the method, we propose six models and compare them based on the AIC. We then present the results of the return level prediction associated with the selected model.

4.3.1 Model selection

The AIC was first introduced by Akaike (1973). It measures the quality of a given model by making a trade-off between its complexity and its goodness of fit. It is defined as

$$AIC = 2K - 2\log(L),$$

where K is the number of estimated parameters and $\log(L)$ is the maximized log-likelihood function of the model. The AIC as well as the number K for each of the six proposed models are given in Table 4.1.

Model	K	AIC
Add_4	41	3917.797
Add_6	49	2286.818
Mix_2	41	2723.934
$\mathbf{Mix_4}$	**73**	**2189.872**
Mix_6	121	2594.673
$Mult_4$	601	4290.686

Table 4.1 – *AIC for the six models*

We have also tried to introduce the year's effect in the models but we haven't obtained any improvements in their performance by adding this factor. The best model among the six is the one with the lowest AIC value which is model Mix_4. Consequently, this model has been selected for the quantile estimation.

4.3.2 Results

We now predict the return levels with period 20 years using the fitted model Mix_4. Recall that this model is of the form

$$\mu = month + f_1(latitude) \times f_2(longitude)$$

$$\sigma = month + g_1(latitude) \times g_2(longitude)$$

with f_1, f_2, g_1 and g_2 natural cubic spline functions with $df = 4$. The prediction is done using the function **return.level**. The same strategy of estimation is used for the observed sites (C_1 group) and the non observed sites ($C_2 \setminus C_1$ group) whose estimated return levels are presented in Table 4.2 and Table 4.3 respectively. The columns represent the months and the rows represent the stations.

"stations"	"X1"	"X2"	"X3"	"X4"	"X5"	"X6"	"X7"	"X8"	"X9"	"X10"	"X11"	"X12"
"2"	0.85	0.93	0.84	0.92	0.85	1.18	1.24	1.1	1.34	1.22	1.07	1.23
"4"	1.48	1.56	1.47	1.55	1.48	1.8	1.87	1.72	1.97	1.85	1.7	1.86
"5"	1.53	1.6	1.52	1.59	1.52	1.85	1.91	1.77	2.02	1.89	1.75	1.91
"6"	1.45	1.52	1.44	1.52	1.44	1.77	1.84	1.69	1.94	1.82	1.67	1.83
"11"	1.48	1.55	1.47	1.55	1.47	1.8	1.87	1.72	1.97	1.85	1.7	1.86
"12"	1.49	1.57	1.48	1.56	1.49	1.81	1.88	1.73	1.98	1.86	1.71	1.87
"13"	1.04	1.11	1.03	1.11	1.03	1.36	1.43	1.28	1.53	1.41	1.26	1.42
"15"	1.26	1.33	1.25	1.33	1.25	1.58	1.64	1.5	1.75	1.63	1.48	1.64
"16"	1.38	1.45	1.36	1.44	1.37	1.7	1.76	1.62	1.86	1.74	1.59	1.75
"18"	1.31	1.38	1.3	1.37	1.3	1.63	1.69	1.55	1.8	1.68	1.53	1.69
"19"	1.51	1.58	1.5	1.58	1.5	1.83	1.9	1.75	2	1.88	1.73	1.89
"20"	1.5	1.57	1.49	1.56	1.49	1.82	1.88	1.74	1.99	1.86	1.72	1.88
"21"	1.53	1.6	1.52	1.6	1.52	1.85	1.92	1.77	2.02	1.9	1.75	1.91
"22"	1.62	1.69	1.6	1.68	1.61	1.94	2	1.86	2.1	1.98	1.83	1.99
"23"	1.42	1.49	1.41	1.49	1.41	1.74	1.8	1.66	1.91	1.79	1.64	1.8
"24"	1.38	1.46	1.37	1.45	1.38	1.7	1.77	1.62	1.87	1.75	1.6	1.76
"25"	0.94	1.02	0.93	1.01	0.94	1.26	1.33	1.18	1.43	1.31	1.16	1.32
"26"	1.48	1.55	1.47	1.55	1.47	1.8	1.87	1.72	1.97	1.85	1.7	1.86
"28"	1.28	1.35	1.26	1.34	1.27	1.6	1.66	1.52	1.77	1.64	1.5	1.66
"29"	0.64	0.71	0.63	0.71	0.63	0.96	1.03	0.88	1.13	1.01	0.86	1.02
"30"	1.11	1.18	1.1	1.18	1.1	1.43	1.49	1.35	1.6	1.48	1.33	1.49
"32"	0.01	0.08	0	0.08	0	0.33	0.4	0.25	0.5	0.38	0.23	0.39
"33"	1.42	1.49	1.4	1.48	1.41	1.74	1.8	1.66	1.9	1.78	1.63	1.79
"34"	1.56	1.63	1.54	1.62	1.55	1.88	1.94	1.8	2.04	1.92	1.77	1.93
"35"	1.45	1.53	1.44	1.52	1.45	1.77	1.84	1.69	1.94	1.82	1.67	1.83
"36"	1.27	1.34	1.26	1.34	1.26	1.59	1.66	1.51	1.76	1.64	1.49	1.65
"38"	1.36	1.43	1.35	1.42	1.35	1.68	1.74	1.6	1.85	1.73	1.58	1.74
"39"	1.31	1.38	1.3	1.38	1.3	1.63	1.7	1.55	1.8	1.68	1.53	1.69
"40"	0.97	1.04	0.95	1.03	0.96	1.29	1.35	1.21	1.45	1.33	1.19	1.34

Table 4.2 – Return levels for group C_1 (AIC method)

"stations"	"X1"	"X2"	"X3"	"X4"	"X5"	"X6"	"X7"	"X8"	"X9"	"X10"	"X11"	"X12"
"7"	1.78	1.85	1.77	1.84	1.77	2.1	2.16	2.02	2.27	2.14	2	2.16
"8"	1.56	1.63	1.55	1.62	1.55	1.88	1.94	1.8	2.05	1.93	1.78	1.94
"9"	0.97	1.04	0.96	1.04	0.96	1.29	1.36	1.21	1.46	1.34	1.19	1.35
"10"	1.97	2.04	1.96	2.03	1.96	2.29	2.35	2.21	2.46	2.33	2.19	2.35
"37"	1.17	1.24	1.16	1.24	1.16	1.49	1.56	1.41	1.66	1.54	1.39	1.55

Table 4.3 – Return levels for group $C_2 \setminus C_1$ (AIC method)

The performance of the quantile predictions $\hat{q}$ are evaluated using the following score function:

$$S_i(\hat{q}) = \sum_{s \in C_i} \sum_{j=1}^{12} S_{s,j}(\hat{q}_{s,j}),\ i = 1, 2, \tag{4.1}$$

where $\hat{q}_{s,j}$ is the estimated return level for station s and month j with $s = 1, \ldots, 40$, $j = 1, \ldots, 12$.
C_1 is the set of stations

$$\begin{aligned} C_1 =& \{2, 4, 5, 6, 11, 12, 13, 15, 16, 18, 19, 20, 21, 22, 23, 24, 25, 26, 28, 29, 30, 32, 33, 34, 35, \\ & 36, 38, 39, 40\} \end{aligned}$$

and C_2 the set of stations $C_2 = C_1 \cup \{7, 8, 9, 10, 37\}$. The function $S_{s,j}(\hat{q}_{s,j})$ is defined as

$$S_{s,j}(\hat{q}_{s,j}) = \sum_{\text{day t of the test period in month j}} \ell(P_{s,t}, \hat{q}_{s,j}),$$

where ℓ is the quantile loss function

$$\ell(x, y) = \alpha(x - y)1_{x>y} + (1 - \alpha)(y - x)1_{y \geq x},$$

at $\alpha = 0.998$ with $x, y \in \mathbb{R}$ and $P_{s,t}$ is the precipitation in station s and day t of period j. The final score is compared to the score of a benchmark predicting the quantiles by the maxima of the training sample for the stations of C_1 and by the average of the monthly maxima of the stations in the training sample for $\{7, 8, 9, 10, 37\}$. The performance of the predicting method is assessed by a percentage representing how much the proposed method has reduced the score with respect to the benchmark methods. Our predicted quantiles in Tables 4.2 and 4.3 based on model Mix_4 has reduced the score with respect to the benchmarks by 43.626% for C_1 and by 41.861% for C_2.

4.4 k-fold cross validation method

Based on discussion with the other teams participating to the data challenge, we have considered model selection based on the k-fold cross validation method. It is indeed sensible to incorporate the score that is used to compare the different methods into the model selection procedure.

4.4.1 Model selection

The cross validation is a method used to evaluate the predictive performance of a given statistical model. It consists in removing a part of the data called the test set and to use the remaining part as a training set. The model is fitted on the training set and the prediction error for the learned model is then assessed on the test set (James et al. (2013)). There are several approaches for cross validation: the validation set approach, the leave-one-out cross validation and the k-fold cross validation. The main difference between the three methods is the way of splitting the observations into the test and training sets. We consider here k-fold cross validation: after

dividing the data set into k folds, we calculate the error based on the observations of the held-out fold (James et al. (2013)). We repeat the same procedure k times and at each time a fold plays the role of a test set and $k-1$ folds play the role of a training set. Finally the global error is the sum of the k computed errors.

Here some care is needed because of the different time scales: year for the annual maxima method, and day for the computation of the score equation (4.1). The data set consists in 24 years, with year 1972 having only 16 observations that we disregard here. The remaining 23 years from 1973 to 1995 are divided into 4 groups, i.e $k = 4$: the first 3 groups have 6 years data and the last group has 5 years data. The group sizes have the same magnitude and hence can be considered as approximately equal. At each time 3 groups will be considered as the training set on which we compute the annual maxima for each month and station from the daily precipitations and apply the block maxima method to estimate the GEV parameters and the 20-year return levels. The remaining group with daily precipitation data will be the test set and the error is computed in terms of the score (4.1).

We present in Table 4.4 the scores obtained for each of the 15 proposed models and the number of estimated parameters K. We can see that the cross validation scores of model Add_1, Add_2, Add_3, Add_4, Mix_1 and $Mult_1$ are very close to each other (all between 1014 and 1019). Because of its simplicity and interpretability, we have selected model Add_1 which corresponds to a linear response surface.

Model	K	Score
Add_1	**29**	**1018.448**
Add_2	33	1018.415
Add_3	37	1017.906
Add_4	41	1016.575
Add_5	45	1073.358
Add_6	49	1053.673
Mix_1	31	1018.573
Mix_2	41	1100.853
Mix_3	55	1052.124
Mix_4	73	1052.265
Mix_5	95	1067.313
Mix_6	121	1089.812
$Mult_1$	97	1014.848
$Mult_2$	217	1025.043
$Mult_3$	385	1028.709

Table 4.4 – *Score with* $4-$*fold cross validation*

4.4.2 Results

Model Add_1 has the very simple form

$$\hat{\mu}(month, station) = \hat{\mu}_{month} + \hat{\mu}_{lat} \times latitude + \hat{\mu}_{lon} \times longitude$$

$$\hat{\sigma}(month, station) = \hat{\sigma}_{month} + \hat{\sigma}_{lat} \times latitude + \hat{\sigma}_{lon} \times longitude.$$

month	$\hat{\mu}_{month}$	$\hat{\sigma}_{month}$
January	0.208	0.264
February	0.092	0.246
March	0.158	0.257
April	0.096	0.221
May	0.172	0.256
June	0.288	0.329
July	0.275	0.353
August	0.220	0.327
September	0.316	0.357
October	0.256	0.350
November	0.250	0.318
December	0.261	0.309

Table 4.5 – *$\hat{\mu}$ and $\hat{\sigma}$ for the months*

Latitude	$\hat{\mu}_{lat} = -0.066$	$\hat{\sigma}_{lat} = 0.018$
Longitude	$\hat{\mu}_{lon} = 0.476$	$\hat{\sigma}_{lon} = -0.047$

Table 4.6 – *$\hat{\mu}$ and $\hat{\sigma}$ for latitude and longitude*

The estimated values of the parameters are given in Tables 4.5 and 4.6. The shape parameter γ is obtained as $\hat{\gamma} = 0.202 > 0$ which means that the daily precipitations are heavy-tailed.

The estimated return levels by model Add_1 for the observed sites (C_1 group) and non observed sites ($C_2 \setminus C_1$ group) are given in Table 4.7 and Table 4.8 respectively. With these results, the reduction of the score of the benchmarks is 58.997% for C_1 and 56.740% for C_2. This is significantly better than for the Mix_4 model selected via the AIC criterion. Furthermore, the Add_1 model can be simply interpreted as the surface responses are linear.

4.5 Discussion

The results of the prediction show that the basic annual maxima method from univariate EVT can perform quite well compared to more advanced methods. The k-fold cross validation method is better than the AIC for model selection because it reduces more the score of the benchmarks and results in more accurate predictions. However, the study remains quite rough so far since many models could have been tested. The assumption γ is a constant could have been tested more carefully. It would be nice to have maps for the parameters and return level. Furthermore, the data suffers from missing values and presence of zeros which were totally disregarded: there are 4709/144413=3.3% missing daily data and 266/4940=5.4% zeros in annual maxima. Finally, it would make sense to incorporate spatial dependence in the model. The resulting spatial model for annual maxima is the max-stable process, but the theory and estimation is much more complicated and it is not clear whether this is useful only for univariate quantile prediction.

"stations"	"X1"	"X2"	"X3"	"X4"	"X5"	"X6"	"X7"	"X8"	"X9"	"X10"	"X11"	"X12"
"2"	1.38	1.19	1.3	1.09	1.31	1.72	1.81	1.65	1.86	1.78	1.64	1.61
"4"	1.4	1.21	1.32	1.11	1.33	1.74	1.82	1.66	1.88	1.8	1.66	1.63
"5"	1.4	1.21	1.32	1.11	1.33	1.74	1.82	1.66	1.88	1.79	1.66	1.63
"6"	1.41	1.22	1.33	1.12	1.34	1.75	1.84	1.68	1.9	1.81	1.67	1.64
"11"	1.42	1.23	1.34	1.13	1.35	1.76	1.84	1.68	1.9	1.81	1.68	1.65
"12"	1.42	1.23	1.34	1.13	1.35	1.76	1.85	1.69	1.91	1.82	1.68	1.65
"13"	1.43	1.24	1.35	1.14	1.36	1.77	1.85	1.69	1.91	1.83	1.69	1.66
"15"	1.43	1.25	1.36	1.15	1.37	1.78	1.86	1.7	1.92	1.83	1.7	1.67
"16"	1.45	1.26	1.37	1.16	1.38	1.79	1.87	1.71	1.93	1.85	1.71	1.68
"18"	1.45	1.26	1.37	1.17	1.39	1.8	1.88	1.72	1.94	1.85	1.72	1.69
"19"	1.45	1.26	1.37	1.16	1.38	1.79	1.88	1.72	1.94	1.85	1.71	1.68
"20"	1.47	1.28	1.39	1.18	1.4	1.81	1.9	1.74	1.96	1.87	1.73	1.7
"21"	1.47	1.28	1.39	1.18	1.4	1.81	1.9	1.74	1.96	1.87	1.73	1.71
"22"	1.48	1.3	1.41	1.2	1.42	1.83	1.91	1.75	1.97	1.88	1.75	1.72
"23"	1.49	1.3	1.41	1.2	1.42	1.83	1.92	1.76	1.97	1.89	1.75	1.72
"24"	1.49	1.3	1.41	1.2	1.42	1.83	1.92	1.76	1.98	1.89	1.75	1.72
"25"	1.52	1.33	1.44	1.23	1.45	1.86	1.94	1.78	2	1.91	1.78	1.75
"26"	1.5	1.31	1.42	1.21	1.43	1.84	1.93	1.77	1.99	1.9	1.76	1.74
"28"	1.34	1.15	1.26	1.05	1.27	1.68	1.76	1.6	1.82	1.73	1.6	1.57
"29"	1.35	1.16	1.27	1.06	1.28	1.69	1.77	1.61	1.83	1.75	1.61	1.58
"30"	1.35	1.16	1.27	1.06	1.28	1.69	1.78	1.62	1.84	1.75	1.61	1.58
"32"	1.36	1.17	1.28	1.07	1.29	1.7	1.79	1.63	1.85	1.76	1.62	1.6
"33"	1.4	1.21	1.32	1.11	1.34	1.74	1.83	1.67	1.89	1.8	1.67	1.64
"34"	1.41	1.22	1.33	1.13	1.35	1.76	1.84	1.68	1.9	1.81	1.68	1.65
"35"	1.42	1.24	1.35	1.14	1.36	1.77	1.85	1.69	1.91	1.82	1.69	1.66
"36"	1.44	1.25	1.36	1.15	1.37	1.78	1.87	1.71	1.93	1.84	1.7	1.68
"38"	1.44	1.25	1.36	1.15	1.37	1.78	1.87	1.71	1.93	1.84	1.7	1.67
"39"	1.44	1.25	1.36	1.16	1.38	1.79	1.87	1.71	1.93	1.84	1.71	1.68
"40"	1.47	1.28	1.39	1.18	1.4	1.81	1.89	1.73	1.95	1.86	1.73	1.7

Table 4.7 – Return levels for group C_1 (k-fold cross validation method)

"stations"	"X1"	"X2"	"X3"	"X4"	"X5"	"X6"	"X7"	"X8"	"X9"	"X10"	"X11"	"X12"
"7"	1.39	1.2	1.31	1.1	1.32	1.73	1.81	1.65	1.87	1.78	1.65	1.62
"8"	1.41	1.22	1.33	1.12	1.34	1.75	1.83	1.67	1.89	1.8	1.67	1.64
"9"	1.43	1.24	1.35	1.14	1.36	1.77	1.85	1.69	1.91	1.82	1.69	1.66
"10"	1.39	1.2	1.31	1.1	1.32	1.73	1.81	1.65	1.87	1.78	1.65	1.62
"37"	1.44	1.25	1.36	1.16	1.38	1.79	1.87	1.71	1.93	1.84	1.71	1.68

Table 4.8 – Return levels for group $C_2 \setminus C_1$ (k-fold cross validation method)

5

Oil rig protection against extreme wind and wave in Lebanon

Abstract

The oil rigs that will be installed in the Lebanese offshore are floating rigs. This type can be affected by environmental risks such as extreme wind speed and wave height. We study these risks in Beddawi region in the north of Lebanon during the winter season. We estimate the return levels associated to return periods of 50, 100 and 500 years for each risk separately using the univariate extreme value theory. Then, by using the multivariate extreme value theory we estimate the dependence between extreme wind speed and wave height as well as joint exceedance probabilities and joint return levels to take into consideration the risk of these two environmental factors simultaneously.

Contents

5.1 Introduction

Petroleum exploration in Lebanon started in the 1930s in the Lebanese onshore but was then stopped due to security circumstances. The offshore exploration has then started in the 1990s and in 2013, 2D and 3D seismic surveys suggest the presence of Gas in the Lebanese offshore. Due to the high depth of the offshore region where the drilling operation will take place, the most probably rig that will be installed is the semi-submersible rig which is a floating rig affected by environmental risks such as wind speed and wave height [1]. The position of Lebanon on the Mediterranean sea makes him susceptible to severe storms accompanied with extreme wind speed and wave height. For instance, the mistral storm of February 2005 hit the Mediterranean sea with wind speed exceeding 30 m/s and wave height reaching 14 meters (m) (Bertotti and Cavaleri (2008)). Furthermore, in the cyclone of January 1995 in the Mediterranean sea, the highest recorded wind speed was 135 km/h and in October 1996, Mediterranean tropical cyclones caused wind speed up to 108 km/h (Pytharoulis et al. (2000)). Hence, extreme wind speed and wave height are not far from being recorded in the Mediterranean sea making essential the problem of predicting them especially when talking about the protection of oil rigs in the offshore. The history has shown how dangerous and catastrophic can extreme wind speed and wave height be to an oil rig. For example, in October 2007, a severe storm with 130 km/h wind speed and 8 m wave height moved the Usumacinta jack-up rig and caused its collision with the Kab-101 platform during a drilling operation. The collision led to a rupture of the platform's production tree and leakage of the hydrocarbon which damaged the platform and caused the death of 21 people. Moreover in November 1979, the Bohai 2 jack-up rig installed in the Golf of Bohai in China encountered a storm with strong winds and high waves that broke a ventilator pump free. Its falling down created a hole in the deck of the rig which was flooded. The rig sank and 72 personnel died on board with only 2 survivals [2]. Note also the sinking of the semi-submersible Ocean Ranger rig in the Canadian waters on 15 February 1982 due to extremely high waves leading to 84 reported deaths. There exist many other oil rig disasters that were caused mainly by storm factors and resulted in dramatic consequences at the environmental, economical and human levels. Here arises the importance of predicting the extreme environmental factors such as wind speed and wave height to ensure a level of safety for the oil rig. One essential statistical tool for these predictions is the univariate extreme value theory (EVT) which predicts these risks separately by estimating their return levels. However, in practice the risk arises as the result of jointly extreme events that happen together and may be dependent. In fact, in offshore environment, oil rigs must be designed to face extreme climate conditions that occur simultaneously. Thus, the study of the multivariate EVT becomes unavoidable to find the joint distribution and the extreme dependence between several risks. Many studies using bivariate or multivariate extreme value distribution have been applied to environmental risks in order to calculate their joint and conditional probabilities (Morton and Bowers (1996), Vanem (2016), Jonathan et al. (2010), Jonathan et al. (2013), Cai et al. (2013)).

1. www.lpa.gov.lb
2. www.oilrigdisasters.co.uk

This chapter is an application of the univariate and bivariate EVT using the block maxima (BM) which ensures the assumption of independence between the extracted BM by considering them separated by a sufficient number of days. The study is done on wind speed and wave height data registered in the winter season because extreme values are mostly recorded in this season. Moreover, since Lebanon's offer for drilling operation includes blocks mainly in the north and the south of Lebanon, we have tried to find data relative to these regions. The data we could get is the one for Beddawi region located in the north of Lebanon. To sum up, the results of the application are relative to extreme wind speed and wave height during the winter season in the Beddawi northern region. The work will be divided as follows: In Section 5.2, we give a theoretical description of the univariate and bivariate EVT. We focus on the BM method used in our work by presenting its main results in both cases. In Section 5.3, we first move to the application of the univariate case and estimate the return levels of extreme wind speed and wave height separately. Secondly we apply the bivariate EVT to study the dependence between the two risks and estimate joint exceedances probabilities associated to joint return periods as well as joint return levels.

5.2 Classical extreme value theory

In this section we briefly recall the basic notions of univariate EVT and provide some fundamentals of the bivariate theory.

5.2.1 Univariate extreme value theory

The univariate EVT deals with extreme observations that are extracted from a given sample of observations using two methods: The BM and the peak over threshold (POT) methods. The first one divides the observations into blocks and extract the maximum value within each block whereas the second one fixes a high threshold and considers the values that exceed it. In our work we only use the BM method so that the BM are separated by a number of days making possible the assumption of independent maxima. The rescaled maxima follow approximately a generalized extreme value (GEV) distribution thanks to the following theorem.

Theorem 5.1 (Gnedenko-Fisher-Tippett theorem, see Embrechts et al. (1997))**.** *Statistically, suppose that $X_1, X_2, \ldots$ is a sequence of independent and identically distributed (i.i.d.) random variables whose distribution function is $F(x) = P(X \leq x)$ such that $X_{m,m} = \max(X_1, .., X_m)$, then: if there exist normalizing constants $a_m > 0$ and b_m such that*

$$\lim_{m\to\infty} \mathbb{P}\left(\frac{M_m - b_m}{a_m} \leq x\right) = \lim_{m\to\infty} F^m(a_m x + b_m) := G(x)$$

with G a non-degenerate distribution function, then G must be of one of the following types:

- *Gumbel:* $G(x) = \exp(-\exp(-x)) \quad$ *for all* $x \in \mathbb{R}$;

- *Fréchet:* $G(x) = \begin{cases} \exp(-x^{-\alpha}) & if\, x > 0 \\ 0 & if\, x \leq 0 \end{cases}$;

- *Weibull:* $G(x) = \begin{cases} \exp(-(-x)^{\alpha}) & if\, x \leq 0 \\ 1 & if \quad x > 0 \end{cases}$.

Jenkinson (1969) combined the three families in a single parametric one depending on three parameters so that the GEV distribution is defined as follows:

$$G_{\mu,\sigma,\gamma}(x) := \begin{cases} \exp\{-[1+\gamma\left(\frac{x-\mu}{\sigma}\right)]^{-1/\gamma}\}, & 1+\gamma\frac{x-\mu}{\sigma} > 0, \gamma \neq 0, \\ \exp\left(-\exp(-\frac{x-\mu}{\sigma})\right), & x \in \mathbb{R}, \gamma = 0, \end{cases}$$

where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ the scale parameter and $\gamma \in \mathbb{R}$ the extremal index.

The main purpose of the univariate EVT is the estimation of extreme quantiles. If we suppose a sample of n observations $X_1, \ldots, X_n$ and divide it into k blocks of size m each such that $n = k \times m$ and denote by $X_1^\star, \ldots, X_k^\star$ the BM, then as we have seen, the $X_i^\star$ will have approximately a GEV distribution whose parameters are estimated by either maximum likelihood or probability weighted moments. The quantile of order p of the vector of maxima $X^\star$ and the vector of original observations X are given in the following lemma.

Lemma 5.1. *The quantile $q_{X^\star,p}$ of order p of $X^\star$ is given by*

$$\begin{aligned} q_{X^\star,p} &= G^{\leftarrow}_{\mu,\sigma,\gamma}(p), \quad p \in (0,1) \\ &= \begin{cases} \mu + \frac{\sigma}{\gamma}[(-\log p)^{-\gamma} - 1] & if\, \gamma \neq 0 \\ \mu - \sigma \log(-\log p) & if\, \gamma = 0 \end{cases} . \end{aligned} \tag{5.1}$$

Because we have $G = F^m$ then the quantile of order p of X will correspond to the quantile of order p^m of $X^\star$. Consequently, an estimator of $q_{X,p}$ is given by the following formula where μ, σ and γ are replaced by their estimated values (see Beirlant et al. (2004)):

$$q_{X,p} = q_{X^\star,p^m} = \begin{cases} \mu + \frac{\sigma}{\gamma}[(-m\log p)^{-\gamma} - 1] & if\, \gamma \neq 0 \\ \mu - \sigma \log(-m \log p) & if\, \gamma = 0 \end{cases} . \tag{5.2}$$

5.2.2 Bivariate extreme value theory

Most of the theoretical works in multivariate EVT are done on i.i.d. observations and very few treat the case of dependent observations. Similarly to the univariate case, the bivariate EVT uses the BM and the POT methods. For the same reason as in the univariate case, we restrict the work to the BM method that uses componentwise maxima.

Componentwise Maxima

Let $(X_1, Y_1), (X_2, Y_2), ..$ be a sequence of independent vectors having $F(x, y)$ as a distribution function, with F_1 and F_2 the distribution function of X and Y respectively. We define

$$X^\star = \max_{i=1,..,m} X_i \qquad \text{and} \qquad Y^\star = \max_{i=1,..,m} Y_i.$$

Hence,

$$M_m = (X^\star, Y^\star)$$

is the vector of componentwise maxima obtained by combining the two vectors $X^\star$ and $Y^\star$. It is to be noted here that M_m is not necessarily a realization. It is assumed that there exist normalizing constants $a_m = (a_{1,m}, a_{2,m}) \in \mathbb{R}_+^{\star 2}$ and $b_m = (b_{1,m}, b_{2,m}) \in \mathbb{R}^2$ and a distribution with non-degenerate margins such that

$$\mathbb{P}\left(\frac{M_m - b_m}{a_m} \leq z\right) = F^m(a_m z + b_m) \underset{m\to\infty}{\longrightarrow} G(z), \tag{5.3}$$

with $z = (x, y)$, $\left(\frac{M_m - b_m}{a_m}\right) = \left(\frac{X^\star - b_{1,m}}{a_{1,m}}, \frac{Y^\star - b_{2,m}}{a_{2,m}}\right)$ and $a_m z = (a_{1,m}x, a_{2,m}y)$. Once (5.3) is verified, F is said to be in the domain of attraction of the bivariate extreme value distribution (BEVD) G (Capéraà and Fougères (2000)). In the bivariate case, we work on the marginal distributions (Coles (2001)). Generally, we can transform the margins to a unit Fréchet to simplify the representation of the BEVD under some parametric models or we can express it in function of extreme value copula as we will see in the next parts.

Unit Fréchet margins

Assume that X_i and Y_i have unit Fréchet margins F_1 and F_2 respectively which are GEV distributions with $\mu = \sigma = \gamma = 1$, i.e $F_1(x) = \exp(-\frac{1}{x})$, $x > 0$ and $F_2(y) = \exp(-\frac{1}{y})$, $y > 0$, and let

$$M_m^\star = (M_{X,m}^\star, M_{Y,m}^\star) = \left(\max_{i=1,..,m}\{X_i\}/m, \max_{i=1,..,m}\{Y_i\}/m\right). \tag{5.4}$$

The distribution of $M_m^\star$ is given by the following theorem (Coles (2001)).

Theorem 5.2. *If (X_i, Y_i) are independent vectors having standard Fréchet marginal distributions and $M_m^\star = (M_{X,m}^\star, M_{Y,m}^\star)$ the vector defined in (5.4), then if*

$$\mathbb{P}(M_{X,m}^\star \leq x, M_{Y,m}^\star \leq y) \xrightarrow{d} G(x, y) \text{ as } m \to \infty,$$

where G is a non-degenerate distribution function, then G can be written as

$$G(x, y) = \exp\{-V(x, y)\},\ x > 0,\ y > 0$$

with

$$V(x, y) = 2\int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right) dH(w) \tag{5.5}$$

called the exponent measure and H a distribution function defined on [0,1] called the spectral measure which satisfies the following mean constraint

$$\int_0^1 w dH(w) = 1/2. \tag{5.6}$$

If H is differentiable having a density h, we can replace $dH(w)$ by $h(w)dw$ in (5.2). As we have seen, $G(x, y)$ doesn't have a parametric representation. However, by choosing a parametric family of H on $[0, 1]$ having a mean of $1/2$ for every value of the parameter we obtain the corresponding family of G. Many models are hence used, the standard one being the logistic model. If we fix the density function of H as

$$h(w) = \frac{1}{2}(\alpha^{-1} - 1)\{w(1-w)\}^{-1-1/\alpha}\{w^{-1/\alpha} + (1-w)^{-1/\alpha}\}^{\alpha-2}$$

which satisfies the mean constraint (5.6) for $0 < w < 1$, then

$$G(x, y) = \exp\{-(x^{-1/\alpha} + y^{-1/\alpha})^{\alpha}\},\ x > 0,\ y > 0,\ \alpha \in [0, 1]. \tag{5.7}$$

The parameter α refers to the degree of dependence (Coles (2001)):
- $\alpha \to 1$, then $G(x, y) \to \exp\left(-(\frac{1}{x} + \frac{1}{y})\right)$ corresponding to an independence.
- $\alpha \to 0$, then $G(x, y) \to \exp\left(-\max(\frac{1}{x}, \frac{1}{y})\right)$ corresponding to a perfect dependence.
Many other models can also be used such as the asymmetric logistic, Husler Reiss and negative logistic. There exists another representation of the BEVD having unit Fréchet margins that uses the Pickands dependence function $A(.)$ (Pickands (1975)):

$$G(x, y) = \exp\left(-\left(\frac{1}{x} + \frac{1}{y}\right) A\left(\frac{x}{x+y}\right)\right),$$

where $A : [0, 1] \to [1/2, 1]$ is a convex function such that $A(0) = A(1) = 1$ and

$$\max(w, 1-w) \leq A(w) \leq 1,\ \forall\, w \in [0, 1].$$

The case $A(w) = 1$ corresponds to independence and $A(w) = \max(w, 1-w)$ to a perfect dependence. The following relation links A to the spectral measure H:

$$A(u) = 2\int_0^1 \max(u(1-w), (1-u)w) dH(w).$$

Extreme value copula

To represent the joint distribution of two variables independently of the margins, we use a bidimensional copula defined as

$$C(.,.) : [0, 1] \times [0, 1] \to [0, 1],$$

such that

- $C(u, 0) = C(0, v) = 0$, $C(u, 1) = u$ and $C(1, v) = v$, $\forall u, v \in [0, 1]$,
- $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$, $\forall\, u_1,\ u_2,\ v_1,\ v_2 \in [0, 1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$ (Nelsen (2006)).

Theorem 5.3 (Sklar, 1959 (Nelsen (2006))). *Let F be a bivariate distribution with margins F_1 and F_2. Then there exists a copula C such that*

$$F(x, y) = C(F_1(x), F_2(y)),$$

with x, $y \in [-\infty, +\infty]$.

Remark 5.1. *If the margins F_1 and F_2 are continuous, then C is unique.*

Most popular copula families include Gaussian copula, Student copula, Archimedean copulas (Clayton and Gumbel copula) etc.
Suppose that the i.i.d. pairs $(X_1, Y_1) \dots (X_m, Y_m)$ have a common copula C and denote by $C_{(m)}$ the copula of the componentwise maxima $X^\star$ and $Y^\star$. Hence we have for u, $v \in [0, 1]$ that $C_{(m)}(u, v) = C^m(u^{1/m}, v^{1/m})$. We give in what follows the definition of an extreme value copula (Nelsen (2006)).

Definition 5.1. A copula $C_\star$ is an extreme value copula if there exists a copula C such that

$$C_\star(u, v) = \lim_{m \to \infty} C^m(u^{1/m}, v^{1/m}),$$

with u, $v \in [0, 1]$.

Definition 5.2. A bidimensional copula C is max-stable if it satisfies the relationship

$$C(u, v) = C^m(u^{1/m}, v^{1/m}),$$

for every integer $m \geq 1$ and $(u, v) \in [0, 1] \times [0, 1]$.

Remark 5.2. *A copula is an extreme copula if and only if it is max-stable (Jaworski et al. (2010)).*

If C is an extreme value copula, it can be expressed in function of A by

$$C(u, v) = \exp\left[\log(uv) A\left(\frac{\log(v)}{\log(uv)}\right)\right]. \tag{5.8}$$

Corollary 5.1. *If G is a bivariate extreme value distribution with univariate margins G_1 and G_2, then*

$$G(x, y) = C_G(G_1(x), G_2(y)),$$

with C_G an extreme copula.

Measures of extremal dependence

Many measures of extremal dependence have been presented and discussed in theory such as Kendall's tau τ and Spearman's rho ρ which are given under different forms. Let (X_1, Y_1) and (X_2, Y_2) be independent random vectors having F as a bivariate distribution function. Kendall's tau is defined as the difference between the probability of concordance or dependence and the probability of discordance or independence (Beirlant et al. (2004)):

$$\tau = P\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - P\{(X_1 - X_2)(Y_1 - Y_2) < 0\}.$$

As for Spearman's rho, it is defined as

$$\rho = 3(P\{(X_1 - X_2)(Y_1 - Y_3) > 0\} - P\{(X_1 - X_2)(Y_1 - Y_3) < 0\}),$$

where $(X_1, Y_1), (X_2, Y_2), (X_3, Y_3)$ are independent copies.
Kendall's tau and Spearman's rho are developed in details in Kendall and Stuart (1979), Kruskal (1958), Lehmann (1975) and Capéraà and Genest (1993). If the margins F_1 and F_2 of F are continuous and C is the copula function of F then Kendall's tau can be expressed in function of C by

$$\tau = 4\mathbb{E}[C(U, V)] - 1 = 4 \int \int_{[0,1]^2} C(u, v) dC(u, v) - 1,$$

with $(U, V) = (F_1(X), F_2(Y))$ has C as a distribution function. The Spearman's rho is given by

$$\rho = 12\mathbb{E}(UV) - 3 = 12 \int \int_{[0,1]^2} C(u, v) du dv - 3.$$

Kendall's tau and Spearman's rho can also be expressed in function of the Pickands dependence function A and its derivative A' (Beirlant et al. (2004)):

$$\tau = \int_0^1 \frac{t(1-t)}{A(t)} dA'(t), \qquad \rho = 12 \int_0^1 \frac{1}{\{1 + A(t)\}^2} dt - 3.$$

Another measure of dependence is the upper tail dependence defined by

$$\psi = \lim_{u \to 1} \psi(u),$$

where

$$\psi(u) = \mathbb{P}(U > u | V > u) = \frac{\bar{C}(u, u)}{1 - u} = 2 - \frac{1 - C(u, u)}{1 - u} \quad u \in [0, 1],$$

and $\bar{C}(u, v) = \mathbb{P}(U > u, V > v)$ is the survival copula of C (Joe (1993)).
The upper tail dependence ψ takes values in $[0, 1]$. The case $\psi = 0$ corresponds to asymptotic independence whereas the asymptotic dependence is obtained if and only if $\psi > 0$.

5.3 Application

We apply in this section the theory on wind speed and wave height data given by the meteorological department of Beirut Rafic Hariri international airport using the **R** program. We tend to predict extreme quantiles of the two variables separately and jointly in order to protect oil rigs against these meteorological risks.

5.3.1 Wind speed and wave height in Lebanon

Lebanon is located on the east of the Mediterranean sea having almost 220 km of waterfront. The north of Lebanon is characterized by a cool and wet climate whereas the south is characterized by a dry and warm climate (Hassan (2011)). Some studies

have shown that on average the period from February to April is the windiest period of the year whereas the period from October to November is the least windy one (Hassan (2011)). The wind movement over the sea generates waves whose height is affected by the wind speed. During storm periods in Lebanon, the wind speed can exceed 27 m/s and the wave height can exceed 8 m (Kabbara (2005)).

5.3.2 The data

We consider now the two meteorological variables of interest the wind speed registered in Beddawi station in Lebanon located at an altitude of 5 m with latitude of $(34°, 27')$ and longitude of $(35°, 53')$ and the wave height registered by a maritime station facing Beddawi station located at an altitude of 0 m with a latitude of $(34°, 28')$ and longitude of $(35°, 53')$. Since the extreme values of the wind speed and wave height are registered in winter, we restrict the data to this season. Note that the study is done on a particular region, hence the results are proper to it and may be different for other regions or for a different time period data. The figure below shows Lebanon's map with white line border where Beddawi region in the north is shown by the blue item.

Figure 5.1 – Location of Beddawi station in Lebanon.

The wind speed data is constructed from the maximum wind speed per day (in m/s) and the wave height data is constructed from the mean of the hourly maximums of wave height registered per day (in cm). The winter data period is from 21/12 to 20/3 starting from year 2000 to 2015 (first date: 21/12/2000, last date: 20/3/2015). The total length of observations is $n = 1353$ observations. We denote by X the wind speed and by Y the wave height. The following table shows a brief summary of the available data.

Data	Min	1^{st} Quartile	Median	Mean	3^{rd} Quartile	Max	NA
X = Wind Speed (m/s)	2.00	5.00	8.00	8.99	11.00	43.00	0
Y= Wave height (cm)	13.40	48.30	69.10	78.14	98.25	298.80	606

Table 5.1 – Summary of available data.

It is to be noted here that the data has two problems: First, the Beddawi station is not a maritime one but facing the maritime station from which the wave height data is collected, thus the wind speed values are not registered in the sea but should be close to the values registered in the offshore. Second the wave height data suffers from a high percentage of missing data: we have 606 missing observations which is a percentage of 44.79%. To complete missing data we proceed to the use of the **missForest** package of R which provides an algorithm for missing values imputation based on random forest (RF) method which presents many advantages. In fact, the RF deals with high dimension and mixed-type data (continuous and categorical) and is a robust and accurate method. Furthermore, it is a non parametric method which means that it can handle data with complex interactions and non linear relation structures. Consequently, the RF doesn't need to make assumptions about the distribution of the data and this leads to a better performance of the method especially when we don't have a prior knowledge of the data. Moreover, the RF algorithm is able to estimate an out of bag (OOB) error rate without the need of a test set. In what follows a brief explanation of the OOB error: The RF is a special case of the bagging method where trees are repeatedly fit to bootstrapped subsets of observations (James et al. (2013)). It is shown that each bagged tree uses around two-thirds of the observations. Hence, there are one-third observations that are not used to fit the given bagged tree. These one-third observations are known as OOB observations. For each of the total n observations, we can predict the response using the OOB observations leading to $B/3$ predictions with B the number of bootstrapped sets. The final prediction is the mean of all predictions. The OOB error is the mean squared OOB error of the n predictions made (James et al. (2013)). Finally, the **missForest** algorithm is shown to be competitive or outperforms other missing values imputation methods (Stekhoven and Bühlmann (2012)). For more details on this method and package see Stekhoven and Bühlmann (2012). To increase the accuracy of the estimated missing values, we apply the **missForest** algorithm 50 times and the final estimation for each day is the mean of the 50 estimations obtained for this day at each iteration. Similarly, the final OOB error is the mean of the 50 OOB errors. The table below shows the summary of the new completed data using the **missForest** algorithm.

Data	Min	1^{st} Quartile	Median	Mean	3^{rd} Quartile	Max
X (m/s)	2.00	5.00	8.00	8.99	11.00	43.00
Y (cm)	13.40	51.76	71.00	78.52	94.55	298.80

Table 5.2 – Summary of completed data.

The difference between the two tables is surely in the wave height data which had missing values. A little difference is registered for the 1^{st} quartile, the median, the mean and the 3^{rd} quartile. The OOB error rate is 4.67%. From now on, the application is done using the completed data.

5.3.3 Univariate extreme value theory

We start the application of the univariate EVT by considering each data separately. We want to estimate the GEV's parameters and the return level corresponding to a return period T years. In the univariate case, the return period T is the average time between two successive extreme events. According to Equation (5.2), the return level of T years for the maxima $X^\star$ is the quantile $q_{X^\star,p}$ where $p = 1 - \frac{1}{\ell T}$. It is also equivalent to the quantile $q_{X,p^{1/m}}$ with $p^{1/m} \simeq 1 - \frac{1}{\ell m T}$ corresponding to the return level of T years for X with ℓ is the number of extreme events per year (or number of blocks per year) and m the block size (Fawcett and Walshaw (2012)). We denote in this section the return level of T years for a random variable $X^\star$ by $q_{X^\star,T}$ in order to make the notations more expressive.

BM method

The choice of block length is critical in EVT. It is a question of trade-off between bias and variance. A large block size leads to high variance and a small block size leads to bias. For this purpose, since we have a total number of observations $n = 1353$ we consider a block of 15 days, i.e with daily observation data we have $m = 15$ and $\ell = 6$ (we have 6 maximums per year since we considered just the winter season equivalent to 90 days per year). The maxima sample size is then equal to $6 \times 15 = 90$. The GEV parameters are estimated using the maximum likelihood method. The estimations $(\hat{\mu}, \hat{\sigma}, \hat{\gamma})$ follow approximately a multivariate normal distribution with mean (μ, σ, γ) and variance-covariance matrix equal to the inverse of the observed information matrix evaluated at the maximum likelihood estimate (Coles (2001)). Consequently the 95% confidence interval of say μ is $\hat{\mu} \pm 1.96 \times \sqrt{\text{Var}(\hat{\mu})}$ (simlarly for σ and γ) where the variances of the estimated parameters relie on the diagonals of the variance-covariance matrix. The estimated parameters and the confidence intervals are presented in Tables 5.3 and 5.4 respectively.

Data	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$
$X^\star$ (m/s)	15.245	4.369	0.044
$Y^\star$ (cm)	135.217	33.243	-0.086

Table 5.3 – *$(\hat{\mu}, \hat{\sigma}, \hat{\gamma})$ for $X^\star$ and $Y^\star$.*

Data	Confidence interval of μ	Confidence interval of σ	Confidence interval of γ
$X^\star$ (m/s)	[14.241,16.250]	[3.639,5.099]	[-0.092,0.181]
$Y^\star$ (cm)	[127.557,142.877]	[27.749,38.736]	[-0.222,0.049]

Table 5.4 – *Confidence interval of μ, σ and γ.*

We remark that the extreme wind speed's tail distribution is heavier ($\gamma > 0$) than the one of the extreme wave height ($\gamma < 0$). Also, the confidence intervals are quite large.
The estimated return level by maximum likelihood method $\hat{q}_{X^\star,T}$ (or $\hat{q}_{Y^\star,T}$) is obtained by replacing the maximum likelihood estimators $(\hat{\mu}, \hat{\sigma}, \hat{\gamma})$ in (5.1). The confidence interval is then $\hat{q}_{X^\star,T} \pm 1.96 \times \mathrm{Var}(\hat{q}_{X^\star,T})$ with

$$\mathrm{Var}(\hat{q}_{X^\star,T}) \sim \nabla^T_{q_{X^\star,T}} V \nabla_{q_{X^\star,T}}$$

where V denotes the variance-covariance matrix of $(\hat{\mu}, \hat{\sigma}, \hat{\gamma})$ and

$$\begin{aligned}\nabla^T_{q_{X^\star,T}} &= \left[\frac{\partial q_{X^\star,T}}{\partial \mu}, \frac{\partial q_{X^\star,T}}{\partial \sigma}, \frac{\partial q_{X^\star,T}}{\partial \gamma}\right] \\ &= \left[1, -\gamma^{-1}(1 - y_p^{-\gamma}), \sigma\gamma^{-2}(1 - y_p^{-\gamma}) - \sigma\gamma^{-1} y_p^{-\gamma} \log y_p\right]\end{aligned}$$

at $(\hat{\mu}, \hat{\sigma}, \hat{\gamma})$ and $y_p = -\log(1-p)$ (Coles (2001)). The return levels $q_{X^\star,T}$ and $q_{Y^\star,T}$ of the maxima wind speed and wave height respectively for a return period of 50, 100 and 500 years are the following.

T (years)	$\hat{q}_{X^\star,T}$	$\hat{q}_{Y^\star,T}$
50	43.570	284.839
100	47.530	298.510
500	57.197	327.238

Table 5.5 – *Return levels for $X^\star$ and $Y^\star$.*

The confidence interval of the return levels are given in Table 5.6.

T (years)	Confidence interval of $q_{X^\star,T}$	Confidence interval of $q_{Y^\star,T}$
50	[32.014,55.126]	[237.160,332.517]
100	[32.831,62.228]	[240.854,356.166]
500	[33.271,81.123]	[244.034,410.442]

Table 5.6 – *Confidence interval of the return levels for $X^\star$ and $Y^\star$.*

Table 5.5 shows for example that the maxima wind speed value in the winter season in Beddawi that will be exceeded on average once every 50 years or 300 blocks is 43.570 m/s. Similarly, the maxima wave height value in the winter season that will be exceeded on average once every 100 years or 600 blocks is 298.510 cm. The return levels $q_{X,T}$ and $q_{Y,T}$ of the original data for the winter season are approximately equal to those of Table 5.5 due to the approximation $G \simeq F^m$ and given in Table 5.7.

T (years)	$\hat{q}_{X,T}$	$\hat{q}_{Y,T}$
50	43.579	284.870
100	47.534	298.525
500	57.198	327.241

Table 5.7 – *Return levels for X and Y.*

5.3.4 Bivariate extreme value theory

In this part we will treat the bivariate EVT in order to measure the dependence between the two variables as well as the joint return levels. We restrict the application of the bivariate EVT to the componentwise maxima method using the logistic model and copulas because as in the univariate case we consider that maxima from different blocks are well separated and can then be considered as independent.

Componentwise Maxima

Similarly to the univariate case, we choose for each variable a block size of 15 days. The number of maxima is then equal to 90. We start by estimating the BEVD's parameters using the logistic model.

Logistic model

The parameters of the BEVD estimated by the logistic model are close to those obtained in the univariate case and given below.

Parameters	$X^\star$	$Y^\star$
$\hat{\mu}$	15.029	135.990
$\hat{\sigma}$	4.272	33.708
$\hat{\gamma}$	0.106	-0.049

Table 5.8 – *$(\hat{\mu}, \hat{\sigma}, \hat{\gamma})$ for $X^\star$ and $Y^\star$ in the logistic model.*

The estimated parameter of dependence α in the logistic model (5.7) is equal to 0.767 which means that there is a little but not complete dependence between the maxima of wind speed and wave height. Moreover, if we apply the χ^2 test under the null hypothesis $H_0 : \alpha = 1$ we obtain a p-value of $0.017 < 0.05$. Hence H_0 is rejected confirming also the dependence between the two variables.
Knowing the BEVD, we can now calculate the conditional probabilities since the two variables $X^\star = \max_{i=1,..,m} X_i$ and $Y^\star = \max_{i=1,..,m} Y_i$ are dependent. For instance, since the wind speed affects the wave height we would like to calculate the probability that $Y^\star$ exceeds its 95% quantile which is 209.425 cm knowing that $X^\star$ has exceeded itself its 95% quantile which is 29.55 m/s. Consequently,

$$\begin{aligned}&\mathbb{P}(Y^\star > y = 209.425 | X^\star > x = 29.55)\\ &= \frac{1 - \mathbb{P}(Y^\star \leq 209.425) - \mathbb{P}(X^\star \leq 29.55) + \mathbb{P}(X^\star \leq 29.55, Y^\star \leq 209.425)}{1 - \mathbb{P}(X^\star \leq 29.55)}\end{aligned} \tag{5.9}$$

such that

$$X^\star \sim G_1 = GEV(\hat{\mu_1} = 15.029, \hat{\sigma_1} = 4.272, \hat{\gamma_1} = 0.106)$$

and

$$Y^\star \sim G_2 = GEV(\hat{\mu_2} = 135.990, \hat{\sigma_2} = 33.708, \hat{\gamma_2} = -0.049).$$

To use the BEVD estimated by the logistic model (5.7), we consider the transformed sequences $M^\star_{1,m}$ and $M^\star_{2,m}$ having a unit Fréchet margin. Consequently, Equation (5.9) is equivalent to

$$\mathbb{P}(M^\star_{2,m} > y^\star | M^\star_{1,m} > x^\star) = \frac{1 - \mathbb{P}(M^\star_{2,m} \leq y^\star) - \mathbb{P}(M^\star_{1,m} \leq x^\star) + \mathbb{P}(M^\star_{1,m} \leq x^\star, M^\star_{2,m} \leq y^\star)}{1 - \mathbb{P}(M^\star_{1,m} \leq x^\star)}$$

where

$$x^\star = \left(1 + \hat{\gamma_1}\frac{29.55 - \hat{\mu_1}}{\hat{\sigma_1}}\right)^{1/\hat{\gamma_1}} = 18.224,$$

$$y^\star = \left(1 + \hat{\gamma_2}\frac{209.425 - \hat{\mu_2}}{\hat{\sigma_2}}\right)^{1/\hat{\gamma_2}} = 10.010$$

and $\mathbb{P}(M^\star_{1,m} \leq x^\star, M^\star_{2,m} \leq y^\star) = G(x^\star, y^\star)$ the distribution function defined in (5.7) with $\hat{\alpha} = 0.767$.

Finally,

$$\begin{aligned}\mathbb{P}(M^\star_{2,m} > y^\star = 10.010 | M^\star_{1,m} > x^\star = 18.224) &= \frac{1 - \exp(-\frac{1}{y^\star}) - \exp(-\frac{1}{x^\star}) + G(x^\star, y^\star)}{1 - \exp(-\frac{1}{x^\star})} \\ &= 0.442.\end{aligned}$$

Going back to the original conditional probability (5.9), the value 0.442 explains that given a wind speed that exceeds 29.55 m/s the wave height has a chance of 44.2% to exceed 209.425 cm. This is also because the dependence is not strong between extreme wind speed and wave height, hence the probability isn't close to 0 (independence) nor to 1 (complete dependence).

Kendall test

Kendall's rank correlation τ for two samples X and Y of size n with no ties is given by

$$\tau = \frac{n_c - n_d}{n(n-1)/2} = \frac{S}{n(n-1)/2},$$

with n_c the number of concordant pairs and n_d the number of discordant pairs (if $x_3 > y_3$ with X and Y ordered samples then the third pair is concordant and discordant otherwise) and $S = n_c - n_d$. The number $n(n-1)/2$ is the number of possible pairs from the samples X and Y. For large samples ($n > 10$), under the null hypothesis of no correlation, S has approximately a Normal distribution with mean zero and variance

$$\mathrm{Var}(S) = \frac{n(n-1)(2n+5)}{18}.$$

If there are ties, τ will be noted τ_b and is given by

$$\tau_b = \frac{S}{\sqrt{(n(n-1)/2 - n_1) \times (n(n-1)/2 - n_2)}},$$

where $n_1 = \sum_i t_i(t_i - 1)/2$ and $n_2 = \sum_j u_j(u_j - 1)/2$ with t_i and u_j the number of tied values in the i^{th} and j^{th} group of ties for the variable X and Y respectively (Bland (2010)). If there are no ties then $\tau_b = \tau$. In the case of ties, the variance has a complicated formula (Hipel and McLeod (1994)):

$$\mathrm{Var}(S) = \frac{(n^2 - n)(2n + 5)}{18} - T"_X - T"_Y + \frac{T'_X T'_Y}{9(n^2 - n)(n - 2)} + \frac{T_X T_Y}{2(n^2 - n)},$$

$$T'_X = \sum_i (t_i^2 - t_i)(t_i - 2),$$

$$T"_X = \sum_i (t_i^2 - t_i)(2t_i + 5),$$

$$T'_Y = \sum_j (u_i^2 - u_j)(u_j - 2),$$

$$T"_Y = \sum_j (u_j^2 - u_j)(2u_j + 5).$$

The variable

$$z = \frac{S + \delta}{\sqrt{Var(S)}}$$

has approximately a Normal distribution where

$$\delta = \begin{cases} -1 \text{ if } S > 0, \\ 1 \text{ if } S < 0. \end{cases}$$

The correlation between the wind speed and wave height maxima given by kendall τ is equal to 0.241. To confirm this dependence between these two variables we apply the kendall test under the null hypothesis

$$H_0 : \tau = 0 \text{ versus } H_1 : \tau \neq 0.$$

The test gives a p-value= 0.00103 $< 5\%$ so H_0 is rejected and H_1 is accepted. We further need to know whether there is a positive or negative correlation. In fact, we obtain p-value= 0.00051 for $H_1 : \tau > 0$ and a p-value of 0.99950 for $H_1 : \tau < 0$. Hence, the accepted alternative hypothesis is $\tau > 0$. Finally, we conclude that there is a positive dependence between the extreme wind speed and wave height which is not so strong. Note that the dependence between the wind speed and the wave height (not extreme values) by kendall tau is equal to 0.486. Hence, even the dependence between the wind speed and wave height is not complete, but still a little bit stronger than for extreme values.

Extreme value Copula

We start this part by recalling the general form of extreme value copula according to Pickands given by Equation (5.8):

$$C(u,v) = \exp\left[\log(uv)A\left(\frac{\log v}{\log uv}\right)\right].$$

Table 5.9 presents the extreme value copula distributions with their Pickands dependence function $A(t)$ and their dependence parameter (Jaworski et al. (2010), Ribatet and Sedki (2013), Tawn (1988)).

Copula	$A(t)$	$C(u,v)$	Dependence parameter
Gumbel	$A(t) = [t^\theta + (1-t)^\theta]^{1/\theta}$	$C_\theta(u,v) = \exp(-[(-\log u)^\theta + (-\log v)^\theta]^{1/\theta})$	$1 \leq \theta < \infty$. When $\theta \to \infty$, we have a perfect dependence and when $\theta = 1$ we have an independence.
Galambos	$A(t) = 1 - \left[t^{-\theta} + (1-t)^{-\theta}\right]^{-1/\theta}$	$C_\theta(u,v) = uv\exp\{[(-\log u)^{-\theta} + (-\log v)^{-\theta}]^{-1/\theta}\}$	$\theta \geq 0$. The case $\theta = 0$ corresponds to an independence and $\theta = \infty$ to a perfect dependence.
Husler Reiss	$A(t) = t\phi(\frac{1}{\theta} + \frac{\theta}{2}\log(\frac{t}{1-t})) + (1-t)\phi(\frac{1}{\theta} - \frac{\theta}{2}\log(\frac{t}{1-t}))$	$C_\theta(u,v) = \exp(\phi[\frac{1}{\theta} + \frac{\theta}{2}\log(\frac{\log v}{\log u})]\log u + \phi[\frac{1}{\theta} + \frac{\theta}{2}\log(\frac{\log u}{\log v})]\log v)$	$\theta \in [0,\infty)$: If $\theta = 0$ we have an independence and if $\theta \to \infty$ we have a perfect dependence.
Tawn	$A(t) = [\delta^\theta(1-t)^\theta + \rho^\theta t^\theta]^{\frac{1}{\theta}} + (\delta-\rho)t + 1 - \delta$ with $\delta, \rho \in [0,1]$	$C_\theta(u,v) = \exp\{\log u^{1-\delta} + \log v^{1-\rho} - [(-\delta\log u)^\theta + (-\rho\log v)^\theta]^{\frac{1}{\theta}}\}$	$1 \leq \theta < \infty$. When $\theta \to \infty$ (with $\delta = \rho = 1$), we have a perfect dependence and when $\theta = 1$ (or $\delta = 0$ or $\rho = 0$) we have an independence.
t-EV	$A(t) = tT_{\psi+1}(z_t) + (1-t)T_{\psi+1}(z_{1-t})$, $z_t = (1+\psi)^{1/2}[\{t/(1-t)\}^{1/\psi} - \rho](1-\rho^2)^{-1/2}$ where $\psi > 0$, $T_{\psi+1}$ the distribution function of the univariate student T-distribution having $\psi + 1$ as a degree of freedom	$C_{\psi,\rho}(u,v) = \int_{-\infty}^{t_\psi^{-1}(u)}\int_{-\infty}^{t_\psi^{-1}(v)} \frac{1}{\pi\psi\lvert P\rvert^{1/2}} \times \frac{\Gamma(\frac{\psi}{2}+1)}{\Gamma(\frac{\psi}{2})}\left(1 + \frac{x'P^{-1}x}{\psi}\right)^{-\psi/2+1}$	$\rho \in (-1,1)$: If $\rho = \pm 1$ we have a perfect dependence and if $\rho = 0$ we have an independence.

Table 5.9 – *Extreme value copulas*

We use from now on the parameters of the univariate GEV distributions G_1 and G_2 for the extreme wind speed and wave height respectively that were estimated in Table 5.3. Using the package **evd** in R we estimate the copula dependence parameter of the couple $(X^\star, Y^\star)$ as well as the upper tail dependence ψ and the conditional probability

$$\mathbb{P}(Y^\star > 209.425 | X^\star > 29.55) = \frac{1 - G_1(29.55) - G_2(209.425) + G(29.55, 209.425)}{1 - G_1(29.55)}$$

where

$$G(x,y) = C_G(G_1(x), G_2(y)).$$

The results are given in Table 5.10.

Copula	Estimated dependence parameter	$\hat{\psi}$	$\mathbb{P}(Y^\star > 209.425 \mid X^\star > 29.55)$
Gumbel	$\hat{\theta} = 1.318$	0.308	0.442
Galambos	$\hat{\theta} = 0.582$	0.304	0.439
Husler Reiss	$\hat{\theta} = 0.967$	0.301	0.437
Tawn	$\hat{\theta} = 1.582$	0.316	0.447
t-EV	$\hat{\rho} = 0.590$	0.308	0.442

Table 5.10 – *Estimated dependence parameter, ψ and conditional probability*

The estimated dependence parameters as well as the upper tail dependence $\psi > 0$ show a slight dependence between the extreme wind speed and wave height. Also, the extreme value copulas are all comparable since their corresponding $\hat{\psi}$ are all close to 0.3. Moreover, the conditional probabilities are all close to 0.4 meaning that once the extreme wind speed exceeds its 95% quantile the extreme wave height will exceed its 95% quantile with approximately a 40% probability.

Choice of copula

To choose the best copula among the five, we can use the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) defined as

$$AIC = 2k - 2\log L,$$

with $\log L$ the optimized value of the log-likelihood function and k the number of estimated parameters and

$$BIC = k\log n - 2\ln L,$$

where n is the number of observations. The best model is the one with the lowest AIC or BIC. Below are the results of the these two criteria for the Gumbel, Galambos, Husler Reiss, Tawn and t-EV copula.

Extreme Copula	Gumbel	Galambos	Husler Reiss	Tawn copula	t-EV Copula
AIC	-6.732	-5.648	-5.165	-8.103	-7.692
BIC	-4.232	-3.148	-2.665	-5.603	-2.692

Table 5.11 – *AIC and BIC for Gumbel, Galambos, Husler Reiss and t-EV copulas*

As we see, the Tawn copula has the lowest AIC and BIC and hence it will be chosen for the rest of the work.

Joint Probability, Conditional Probability and Return levels

In the bivariate case, the joint return period $T_{X^\star Y^\star}$ corresponding to the event that either x or y or both are exceeded in the case of annual maxima (i.e $X^\star > x$ or $Y^\star > y$ or ($X^\star > x$ and $Y^\star > y$)) is given by (Shiau et al. (2006))

$$T_{X^\star Y^\star} = \frac{1}{\mathbb{P}(X^\star > x \text{ or } Y^\star > y)} = \frac{1}{1 - G(x,y)} = \frac{1}{1 - C(G_1(x), G_2(y))}.$$

The joint return period corresponding to the event that both x and y are exceeded (i.e $X^\star > x$ and $Y^\star > y$) denoted $T'_{X^\star Y^\star}$ is given by (Shiau et al. (2006))

$$\begin{aligned} T'_{X^\star Y^\star} &= \frac{1}{\mathbb{P}(X^\star > x \text{ and } Y^\star > y)} \\ &= \frac{1}{1 - G_1(x) - G_2(y) + G(x,y)} \\ &= \frac{1}{1 - G_1(x) - G_2(y) + C(G_1(x), G_2(y))}. \end{aligned} \tag{5.10}$$

It is to be noted that the calculation made are based on the Tawn copula which is the most appropriate copula. We recall the return levels for 50, 100 and 500 years that were estimated in Section 5.3.3 and that will be used in this part:

T (years)	$\hat{q}_{X^\star,T}$	$\hat{q}_{Y^\star,T}$
50	43.570	284.839
100	47.530	298.510
500	57.197	327.238

Return levels for the maxima wind speed and wave height.

Table 5.12 presents the following exceedance probability

$$\mathbb{P}(X^\star > \hat{q}_{X^\star,T}, Y^\star > \hat{q}_{Y^\star,T'}) = 1 - G_1(\hat{q}_{X^\star,T}) - G_2(\hat{q}_{Y^\star,T'}) + G(\hat{q}_{X^\star,T}, \hat{q}_{Y^\star,T'})$$

for the different return levels above.

Return levels	$\hat{q}_{Y^\star,50} = 284.839$	$\hat{q}_{Y^\star,100} = 298.510$	$\hat{q}_{Y^\star,500} = 327.238$
$\hat{q}_{X^\star,50} = 43.570$	0.00106	0.00071	0.00019
$\hat{q}_{X^\star,100} = 47.530$	0.00071	0.00053	0.00018
$\hat{q}_{X^\star,500} = 57.197$	0.00019	0.00018	0.00011

Table 5.12 – $\mathbb{P}(X^\star > \hat{q}_{X^\star,T}, Y^\star > \hat{q}_{Y^\star,T'})$

Since we have $\ell = 6$ blocks per year, the probability $\mathbb{P}(X^\star > \hat{q}_{X^\star,T}, Y^\star > \hat{q}_{Y^\star,T'})$ is then equal to $\frac{1}{\ell T'_{X^\star Y^\star}}$. The corresponding return period $T'_{X^\star Y^\star}$ are the following:

Return levels	$\hat{q}_{Y^\star,50} = 284.839$	$\hat{q}_{Y^\star,100} = 298.510$	$\hat{q}_{Y^\star,500} = 327.238$
$\hat{q}_{X^\star,50} = 43.570$	157.273	236.309	868.057
$\hat{q}_{X^\star,100} = 47.530$	236.309	315.498	947.920
$\hat{q}_{X^\star,500} = 57.197$	868.057	947.920	1581.323

Table 5.13 – *Joint return period* $T'_{X^\star Y^\star}$

Table 5.13 shows that we have $T'_{X^\star Y^\star} \geq \max(T, T')$. In fact, this was also given and proved in Yue and Rasmussen (2002). It is explained as follows: when we are considering the risk of $X^\star$ and $Y^\star$ jointly, the probability of exceedance

$\mathbb{P}(X^\star > \hat{q}_{X^\star,T}, Y^\star > \hat{q}_{Y^\star,T'})$ is lower than the minimum of the marginal exceedance probabilities. The table also shows that the joint return periods $T'_{X^\star Y^\star}$ corresponding to the exceedance probabilities $\mathbb{P}(X^\star > \hat{q}_{X^\star,T}, Y^\star > \hat{q}_{Y^\star,T'})$ and $\mathbb{P}(X^\star > \hat{q}_{X^\star,T'}, Y^\star > \hat{q}_{Y^\star,T})$ are the same. The joint return periods $T_{X^\star Y^\star}$ corresponding to the event that either $\hat{q}_{X^\star,T}$ or $\hat{q}_{Y^\star,T'}$ or both are exceeded are presented in Table 5.14.

Return levels	$\hat{q}_{Y^\star,50} = 284.839$	$\hat{q}_{Y^\star,100} = 298.510$	$\hat{q}_{Y^\star,500} = 327.238$
$\hat{q}_{X^\star,50} = 43.570$	29.725	38.807	47.966
$\hat{q}_{X^\star,100} = 47.530$	38.807	59.416	91.365
$\hat{q}_{X^\star,500} = 57.197$	47.966	91.365	296.946

Table 5.14 – *Joint return period* $T_{X^\star Y^\star}$

Table 5.14 shows that $T_{X^\star Y^\star} \leq \min(T, T')$ as shown in Yue and Rasmussen (2002) when there is a dependence between $X^\star$ and $Y^\star$.
One can also be interested in the computation of the conditional probabilities $\mathbb{P}(Y^\star > \hat{q}_{Y^\star,T'} | X^\star > \hat{q}_{X^\star,T})$ whose results are given in Table 5.15.

Return levels	$\hat{q}_{Y^\star,50} = 284.839$	$\hat{q}_{Y^\star,100} = 298.510$	$\hat{q}_{Y^\star,500} = 327.238$
$\hat{q}_{X^\star,50} = 43.570$	0.318	0.212	0.058
$\hat{q}_{X^\star,100} = 47.530$	0.423	0.317	0.105
$\hat{q}_{X^\star,500} = 57.197$	0.576	0.527	0.316

Table 5.15 – $\mathbb{P}(Y^\star > \hat{q}_{Y^\star,T'} | X^\star > \hat{q}_{X^\star,T})$

Obviously we have $\mathbb{P}(Y^\star > \hat{q}_{Y^\star,T'} | X^\star > \hat{q}_{X^\star,T}) > \mathbb{P}(X^\star > \hat{q}_{X^\star,T}, Y^\star > \hat{q}_{Y^\star,T'})$. We remark that the probability increases with the increase of the return period of $X^\star$ for a fixed $\hat{q}_{Y^\star,T'}$ and decreases with the increase of the return period of $Y^\star$ for a fixed $\hat{q}_{X^\star,T}$. Also in the main diagonal of the table (i.e. when $T = T'$), the probabilities are very close to each other.
Furthermore, it is very interesting to find $(x, y) | \mathbb{P}(X^\star > x, Y^\star > y) = p = \frac{1}{\ell \times T'_{X^\star Y^\star}}$ for a given return period $T'_{X^\star Y^\star}$. Heuristically we have found these couples using an algorithm and presented them in the table below for a period $T'_{X^\star Y^\star}$ of 50,100 and 500 years. The probability of exceedance for each of the presented couples (x, y) is an approximation of p rounded to the nearest 3×10^{-6}, 3×10^{-7} and 9×10^{-8}. We present the couple having the closest exceedance probability to $\frac{1}{\ell \times T'_{X^\star Y^\star}}$ in Table 5.16.

$T'_{X^\star Y^\star}$	50	100	500
Joint return levels	(40,238)	(38,283)	(54,266)

Table 5.16 – *Return levels for 50,100 and 500 years return period*

From this table we conclude that for example the vector (x, y) corresponding to a joint return period of 50 years is (40,238) which means that with a probability of $\frac{1}{6 \times 50}$ extreme wind speed will exceed 40 m/s and the wave height will exceed 238 cm simultaneously once every 50 years or 600 blocks. Note that the values 40 and 238

are lower than the univarite 50-year return levels for wind speed and wave height respectively since the couple (43.570, 284.839) of univariate 50 years return period corresponds to a joint return period $T'_{X^\star Y^\star}$ of 157.273 years.

5.4 Conclusion & perspectives

Extreme wind speed and wave height in the winter season in the north of Lebanon are dependent but not perfectly. The study of their jointly risk is an important issue to protect to the oil rig that will be installed in the northern Lebanese offshore. The joint return period corresponding to the event that extreme wind speed and wave height exceed simultaneously their univariate return levels of periods T and T' respectively is higher than $\max(T, T')$. The joint return period corresponding to the event that either extreme wind speed or extreme wave height or both exceed simultaneously their univariate return levels of periods T and T' respectively is lower than $\min(T, T')$. Finally, the results were found after completing missing data by the missForest algorithm. Hence, it is important to check the existence or not of an imputation method that outperforms the random forest algorithm. Moreover, given the estimation of the univariate and joint return levels with return period of 50, 100 and 500 years, it remains to decide which of these combinations is the best for the design of the oil rig's components. The choice should be a trade-off between the economical cost and the desired level of safety.

Conclusion & Perspectives

The thesis has contributed to EVT at different levels.
The second chapter has shown that the permutation Bootstrap method based on BM ranks and the order statistics of the original observations enhances the accuracy of the GEV parameters. The BM ranks follow a distribution that is used in their simulation. The method saves computation time by avoiding the permutation of the original observations, forming blocks and computing BM at each time. It also leads to a reduction of the mean square error (MSE) of the estimators.
The third chapter allows for a trend detection of the heteroscedastic extremes. Based on the model of Einmahl et al. (2016), we propose three parametric models (log-linear, linear and discrete log-linear) for the skedasis function $c = c_\theta$ which represents the variation of extremes in order to test the trend of extremes through time. The estimators $\hat{\theta}$ in all the models are consistent and effective. The parametric tests $H_0 : \theta = 0$ (resp. $\theta \leq 0$) versus $H_1 : \theta \neq 0$ (resp. $\theta > 0$) are proposed in order to test the trend of c and consequently conclude on the frequency of extremes. The parametric test $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$ is compared to the non parametric Kolmogorov-Smirnov test. The results show that the parametric test is more powerful than the non parametric test which explains the usefulness of considering parametric models of c.
In the fourth chapter the univariate EVT that incorporates covariates in the GEV parameters is used and performs quite well for the prediction of the 20-year return level of precipitations for every month and station. The k-fold cross validation method is more appropriate than the Akaike information criterion for model selection and results in more robust estimates.
In the last chapter we apply the univariate and bivariate EVT on wind speed and wave height winter season data in a Lebanese northern region. We predict at a first time the return levels of 50, 100 and 500 years for each environmental factor in order to protect the oil rig that will be installed in the northern offshore from these risks. To increase the level of safety, we study the risks of these two factors simultaneously by using the bivariate EVT. We estimate joint exceedance probabilities above the marginal return levels and correspond them to joint return periods. Heuristically we finally estimate the joint return levels associated to a joint return period of 50,

100 and 500 years.
The different chapters give rise to many questions and perspectives that should be studied in the future. Some were already presented at the end of the chapters but will be recalled here.
In the second chapter, it would be interesting to check the MSE reduction or the behavior of the estimators based on simulated data that follow a different distribution than the Pareto.
In the framework of heteroscedastic extremes, many interesting parametric models of the skedasis function c can still be studied such as the change point model. In this type of models, the function c is defined as

$$c_{p,s}(t) = \begin{cases} \dfrac{p}{s} & \text{if} \quad 0 \leq t \leq s, \\ \dfrac{1-p}{1-s} & \text{if} \quad s < t \leq 1 \end{cases} .$$

Can the maximum likelihood and the moment methods be applied here for the estimation of the parameters s and p? Are the estimators consistent and asymptotically normal? The estimation seems more complex in this case because the statistical model is not regular since the likelihood function is not differential with respect to s. Also, are there any additional interesting parametric models that can be studied? Could the study be done on abstract parametric models $\{c_\theta, \theta \in \Theta\}$?
Moreover in Chapter 4, can we consider more complex models that improve the accuracy of the estimations? Should we take the spatial dependence between stations into account and incorporate it in the models?
Finally in Chapter 5, the wave height data was completed by the random forest algorithm. Is there any other method that outperforms this algorithm and gives more accurate estimations of the missing data? Also, how does the choice of the block size ($m = 15$) affect the parameters estimation and consequently the return levels estimation? Furthermore, it would be interesting to apply the results of Chapter 3 in order to test the presence of a trend for the wind speed and wave height data.

Bibliography

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. pages 267–281.

Balkema, A. A. and de Haan, L. (1974). Residual life time at great age. *Ann. Probability*, 2:792–804.

Barndorff-Nielsen, O. (1978). *Information and exponential families in statistical theory*. John Wiley & Sons, Ltd., Chichester. Wiley Series in Probability and Mathematical Statistics.

Beirlant, J., Goegebeur, Y., Teugels, J., and Segers, J. (2004). *Statistics of extremes*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester. Theory and applications, With contributions from Daniel De Waal and Chris Ferro.

Bertotti, L. and Cavaleri, L. (2008). Analysis of the voyager storm. 35:1–5.

Bland, M. (2010). An introduction to medical statistics (3rd ed.). 96:82.

Boucheron, S. and Thomas, M. (2015). Tail index estimation, concentration and adaptivity. *Electron. J. Stat.*, 9(2):2751–2792.

Buishand, T. (1984). Bivariate extreme-value data and the station-year method. *Journal of Hydrology*, 69(1):77–95.

Buishand, T. A., de Haan, L., and Zhou, C. (2008). On spatial extremes: with application to a rainfall problem. *Ann. Appl. Stat.*, 2(2):624–642.

Butler, A., Heffernan, J. E., Tawn, J. A., and Flather, R. A. (2007). Trend estimation in extremes of synthetic North Sea surges. *J. Roy. Statist. Soc. Ser. C*, 56(4):395–414.

Caeiro, F. and Gomes, M. I. (2011). Semi-parametric tail inference through probability-weighted moments. *J. Statist. Plann. Inference*, 141(2):937–950.

Cai, J. J., Fougères, A.-L., and Mercadier, C. (2013). Environmental data: multivariate extreme value theory in practice. *J. SFdS*, 154(2):178–199.

Capéraà, P. and Fougères, A.-L. (2000). Estimation of a bivariate extreme value distribution. *Extremes*, 3(4):311–329 (2001).

Capéraà, P. and Genest, C. (1993). Spearman's ρ is larger than Kendall's τ for positively dependent random variables. *J. Nonparametr. Statist.*, 2(2):183–194.

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer Series in Statistics. Springer-Verlag London, Ltd., London.

Coles, S. and Pericchi, L. (2003). Anticipating catastrophes through extreme value modelling. *J. Roy. Statist. Soc. Ser. C*, 52(4):405–416.

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *J. Roy. Statist. Soc. Ser. B*, 52(3):393–442. With discussion and a reply by the authors.

de Haan, L. and Ferreira, A. (2006). *Extreme value theory*. Springer Series in Operations Research and Financial Engineering. Springer, New York. An introduction.

de Haan, L., Klein Tank, A., and Neves, C. (2015). On tail trend detection: modeling relative risk. *Extremes*, 18(2):141–178.

de Haan, L. and Resnick, S. I. (1977). Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 40(4):317–337.

Dombry, C. and Ferreira, A. (2017). Maximum likelihood estimators based on the block maxima method. *ArXiv e-prints*.

Drees, H. and Rootzén, H. (2010). Limit theorems for empirical processes of cluster functionals. *Ann. Statist.*, 38(4):2145–2186.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.*, 7(1):1–26.

Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Sci.*, 1(1):54–77. With a comment by J. A. Hartigan and a rejoinder by the authors.

Einmahl, J. H. J., de Haan, L., and Zhou, C. (2016). Statistics of heteroscedastic extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 78(1):31–51.

Embrechts, P., Goldie, C. M., and Veraverbeke, N. (1979). Subexponentiality and infinite divisibility. *Z. Wahrsch. Verw. Gebiete*, 49(3):335–347.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling extremal events*, volume 33 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin. For insurance and finance.

Fawcett, L. and Walshaw, D. (2012). Estimating return levels from serially dependent extremes. *Environmetrics*, 23(3):272–283.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 180–190. Cambridge Univ Press.

Gilleland, E. and Katz, R. W. (2011). New software to analyze how extremes change over time. *Eos, Transactions American Geophysical Union*, 92(2):13–14.

Gilleland, E. and Katz, R. W. (2016). extRemes 2.0: An extreme value analysis package in R. *Journal of Statistical Software*, 72(8):1–39.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. of Math. (2)*, 44:423–453.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., and Wallis, J. R. (1979). Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form. *Water Resources Research*, 15(5):1049–1054.

Groisman, P. Y., Knight, R. W., Easterling, D. R., Karl, T. R., Hegerl, G. C., and Razuvaev, V. N. (2005). Trends in intense precipitation in the climate record. *Journal of Climate*, 18(9):1326–1350.

Gumbel, E. J. (1958). Statistics of extremes. *Columbia University Press, New York*.

Hassan, G. (2011). The national wind atlas of lebanon. *UNDP/CEDRO*.

Hipel, K. W. and McLeod, A. I. (1994). *Time series modelling of water resources and environmental systems*. Elsevier Amsterdam; New York.

Hsing, T. (1991). On tail index estimation using dependent data. *Ann. Statist.*, 19(3):1547–1569.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*, volume 103 of *Springer Texts in Statistics*. Springer, New York. With applications in R.

Jaworski, P., Durante, F., Härdle, W., and Rychlik, T., editors (2010). *Copula theory and its applications*, volume 198 of *Lecture Notes in Statistics—Proceedings*. Springer, Heidelberg.

Jenkinson, A. (1969). Statistics of extremes. *Estimation of Maximum Flood, Technical Note*, 98:183–227.

Joe, H. (1993). Parametric families of multivariate distributions with given margins. *J. Multivariate Anal.*, 46(2):262–282.

Jonathan, P., Ewans, K., and Randell, D. (2013). Joint modelling of extreme ocean environments incorporating covariate effects. *Coastal Engineering*, 79:22 – 31.

Jonathan, P., Flynn, J., and Ewans, K. (2010). Joint modelling of wave spectral parameters for extreme sea states. *Ocean Engineering*, 37(11):1070 – 1080.

Kabbara, N. (2005). Wind and wave data analysis for the lebanese coastal area-preliminary results. 6.

Kendall, M. and Stuart, A. (1979). Handbook of statistics. *Griffin & Company, London*.

Kitagawa, G. and Konishi, S. (2010). Bias and variance reduction techniques for bootstrap information criteria. *Ann. Inst. Statist. Math.*, 62(1):209–234.

Klüppelberg, C. and Zhang, J. (2015). Time-consistency of risk measures with GARCH volatilities and their estimation. *Stat. Risk Model.*, 32(2):103–124.

Kruskal, W. H. (1958). Ordinal measures of association. *J. Amer. Statist. Assoc.*, 53:814–861.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*. Springer Series in Statistics. Springer-Verlag, New York-Berlin.

Lehmann, E. L. (1975). *Nonparametrics: statistical methods based on ranks*. Holden-Day, Inc., San Francisco, Calif.; McGraw-Hill International Book Co., New York-Düsseldorf. With the special assistance of H. J. M. d'Abrera, Holden-Day Series in Probability and Statistics.

Morton, I. and Bowers, J. (1996). Extreme value analysis in a multivariate offshore environment. *Applied Ocean Research*, 18(6):303 – 317.

Nelsen, R. B. (2006). *An introduction to copulas*. Springer Series in Statistics. Springer, New York, second edition.

Pickands, III, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.*, 3:119–131.

Prescott, P. and Walden, A. (1983). Maximum likeiihood estimation of the parameters of the three-parameter generalized extreme-value distribution from censored samples. *Journal of Statistical Computation and Simulation*, 16(3-4):241–250.

Prescott, P. and Walden, A. T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika*, 67(3):723–724.

Pytharoulis, I., Craig, G. C., and Ballard, S. P. (2000). The hurricane-like mediterranean cyclone of january 1995. *Meteorological Applications*, 7(3):261–279.

Resnick, S. I. (1987). *Extreme values, regular variation, and point processes*, volume 4 of *Applied Probability. A Series of the Applied Probability Trust*. Springer-Verlag, New York.

Resnick, S. I. (2007). *Heavy-tail phenomena*. Springer Series in Operations Research and Financial Engineering. Springer, New York. Probabilistic and statistical modeling.

Ribatet, M. and Sedki, M. (2013). Extreme value copulas and max-stable processes. *J. SFdS*, 154(1):138–150.

Seshadri, V., Csorgo, M., and Stephens, M. A. (1969). Tests for the exponential distribution using kolmogorov-type statistics. *Journal of the Royal Statistical Society. Series B (Methodological)*, 31(3):499–509.

Shiau, J.-T., Wang, H.-Y., and Tsai, C.-T. (2006). Bivariate frequency analysis of floods using copulas. *JAWRA Journal of the American Water Resources Association*, 42(6):1549–1564.

Singh, A. K., Allen, D. E., and Robert, P. J. (2013). Extreme market risk and extreme value theory. *Math. Comput. Simulation*, 94:310–328.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90.

Smith, R. L. (1987). Estimating tails of probability distributions. *Ann. Statist.*, 15(3):1174–1207.

Stekhoven, D. and Bühlmann, P. (2012). Missforest?non-parametric missing value imputation for mixed-type data. 28:112–8.

Tank, A. M. G. K. and Können, G. P. (2003). Trends in indices of daily temperature and precipitation extremes in europe, 1946–99. *Journal of Climate*, 16(22):3665–3680.

Tawn, J. A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika*, 75(3):397–415.

van der Vaart, A. W. (1998). *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

Vanem, E. (2016). Joint statistical models for significant wave height and wave period in a changing climate. *Marine Structures*, 49:180 – 205.

Walshaw, D. (2000). Modelling extreme wind speeds in regions prone to hurricanes. *J. Roy. Statist. Soc. Ser. C*, 49(1):51–62.

Yue, S. and Rasmussen, P. (2002). Bivariate frequency analysis: discussion of some useful concepts in hydrological application. *Hydrological Processes*, 16(14):2881–2898.

Zhou, C. (2009). Existence and consistency of the maximum likelihood estimator for the extreme value index. *J. Multivariate Anal.*, 100(4):794–815.

Zolina, O., Simmer, C., Belyaev, K., Kapala, A., and Gulev, S. (2009). Improving estimates of heavy and extreme precipitation using daily records from european rain gauges. *Journal of Hydrometeorology*, 10(3):701–716.