

*ÉCOLE DOCTORALE EDMSII*

**IRMA**

**THÈSE** présentée par :

**Antoine SCHORGEN**

soutenue le : **21 septembre 2012**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : **Mathématiques Appliquées**

**Valeurs Extrêmes : covariables et cadre  
bivarié**

**THÈSE dirigée par :**

**Mme GUILLOU Armelle**

Professeur, université de Strasbourg

**RAPPORTEURS :**

**M. BEIRLANT Jan**

Professeur, KU Leuven Kulak

**Mme CHAVEZ-DEMOULIN Valérie**

Professeur, Université de Lausanne

---

**AUTRES MEMBRES DU JURY :**

**M. GARDES Laurent**

Professeur, université de Strasbourg

**M. NAVEAU Philippe**

Chargé de Recherche, LSCE CNRS

# Préface

Mon parcours universitaire s'achève maintenant, et avec lui l'ensemble de ma scolarité. A cette occasion je tenais à en faire un bref bilan.

Une fois le bac passé, le chemin naturel pour un très bon élève est celui de la classe préparatoire. Dans une grande prépa nantaise, je n'ai tenu qu'un mois. Sommet de l'horreur éducative atteint, mental cassé, je n'avais plus d'autre choix que de tout arrêter pour me préserver. Plus en état de continuer mes études, j'ai refusé de rejoindre tout de suite l'université, me promettant par la même occasion que je n'étudierai plus jamais dans un système élitiste.

Trois années à l'Université de Nantes suivies de cinq ans à celle de Strasbourg. Mon parcours n'ayant pas été un long fleuve tranquille, j'ai parfois souffert à l'université. À la fin de chaque année je voulais arrêter. Arrêter d'accord, mais pour faire quoi de mieux ?

J'ai eu la chance d'aimer fortement les mathématiques. Et d'aimer mes enseignants. Grâce à eux, j'ai pu trouver du plaisir, j'ai pu comprendre des choses et continuer jusqu'au bout. Ils ont rendu mon parcours valable. Mes enseignants du supérieur évidemment, mais aussi tous les enseignants depuis la première classe.

L'éducation publique, je l'ai vécue, du début à la fin. Je pourrai donc en parler. D'une certaine façon, j'en suis fier. Je ne sais pas si j'ai réussi à apprendre les sujets enseignés aussi profondément que je l'aurais souhaité...faute de temps, faute de moyens comme on dit. Mais je suis sûr que j'ai un peu appris la vie. C'est déjà ça.

Je termine cette préface par une phrase, dénichée dans un livre pioché au hasard dans la bibliothèque de maths à Nantes. Une phrase d'André Weil, alsacien d'origine qui a enseigné plusieurs années à l'Université de Strasbourg, qui disait (de mémoire) :

« Peu importe le système, les hommes [et les femmes], eux, importent ! »



# Remerciements

Je tiens tout d'abord à remercier vivement tous les membres du jury. Merci à Valérie Chavez-Demoulin et Jan Beirlant d'avoir accepté de rapporter mon travail de thèse, et encore merci à eux ainsi qu'à Laurent Gardes et Philippe Naveau d'avoir pu se libérer pour être présents aujourd'hui.

Je souhaite maintenant adresser un très très grand merci à Armelle Guillou, ma directrice de thèse. Pour commencer, merci de m'avoir proposé cette thèse à la fin du master. Peu de professeurs proposent une thèse aux nuls qui n'ont ni l'agrégation ni un label d'excellence type « normale sup » ou « centrale ». Merci pour le très grand soin avec lequel elle a encadré ma thèse pendant ces trois ans. Toujours disponible pour me recevoir malgré ses nombreuses occupations et responsabilités, toujours prête à répondre à mes questions et à discuter de différentes choses. Merci aussi pour sa grande maîtrise de la théorie des valeurs extrêmes. Et enfin, merci pour le soutien dans les moments difficiles. Il y en a eu quelques uns (beaucoup ?) pendant cette thèse, et elle a su, à sa façon, me remotiver à chaque fois pour que j'en termine aujourd'hui. Par ailleurs, j'admets (seulement maintenant !) être sûrement quelqu'un de difficile à encadrer ; j'ai donc dû lui causer du soucis plus d'une fois. J'ai pourtant essayé de faire du mieux que je pouvais. . . En tout cas, merci pour tout.

Je voudrais ensuite remercier les trois autres chercheurs, Philippe Naveau, Laurent Gardes et Yuri Goegebeur, avec qui j'ai eu la chance de travailler, et qui m'ont beaucoup appris, sur la recherche, sur mon sujet de thèse et sur d'autres choses.

Merci à Philippe, pour l'accueil toujours chaleureux au LSCE à Saclay, pour les discussions sur le madogramme, sur les valeurs extrêmes et d'autres sujets. Merci pour les données réelles. Merci aussi de m'avoir permis de rencontrer d'autres chercheurs, Martin Schlather notamment, et des jeunes chercheurs à Aussois, Pierre Ribereau, Julie Carreau, Mathieu Ribatet entre autres. . .

Merci à Laurent, pour la disponibilité qu'il a pu avoir pour faire avancer mon travail. Merci d'avoir toujours été prêt à m'aider ou à répondre aux questions que je pouvais avoir. Et merci de m'avoir fait découvrir le monde de la covariable :-)

Merci à Yuri, pour le travail effectué ensemble, pour ses deux visites à Strasbourg, qui m'ont permis d'échanger avec un chercheur étranger, de découvrir une approche parfois différente de celle qu'on a l'habitude de connaître. Cela a été très enrichissant !

Merci aussi à toute l'équipe de statistique, pour les séminaires, les repas et les discussions.

Merci aux doctorants de l'IRMA, avec qui j'ai pu discuter et passer un peu de temps pendant ces trois ans.

Merci à mon collègue de bureau, il m'a fait comprendre ce qu'était la motivation extrême !

Je remercie également toutes les personnes qui m'ont permis de mieux travailler, ou de travailler dans une meilleure atmosphère.

Merci à Nicolas Poulin pour sa bonne humeur ! Et merci pour les discussions variées et les BDs ;-)

Merci aux personnes du 4e étage pour la bonne ambiance. Merci Nicolas (x2), Laurent, Armelle, Michaël, Eugen, Jade, Anais, Christophe . . .

Merci à l'équipe d'info, Alain et Alexis, de nous fournir de très bonnes conditions de travail numérique (réseau, ordinateurs . . .). Merci aussi pour toutes les discussions (j'aime bien venir vous voir !) et pour toutes les solutions que vous avez trouvées pour régler les problèmes que j'avais (je vous ai bien embêtés hein ?). Merci aussi pour l'exposé sur les serveurs et le « nohup », cela m'a été d'une très grande utilité pour mener à bien mes simulations les plus lourdes.

Merci aux personnes qui s'occupent de la bibliothèque : l'endroit est très agréable et cela a toujours été pour moi un grand plaisir d'y venir.

Merci au personnel administratif et technique, qui m'ont facilité la vie à la fac !

Finalement, un grand merci à toutes celles (et ceux), que j'ai pu croiser ou côtoyer à l'université, qui ont toujours le sourire. C'est important dans ce monde de brutes.

Concernant la partie enseignement - parce que j'ai eu la chance d'avoir, pendant trois ans, une mission d'enseignement - j'aimerais remercier quelques personnes.

Merci à mon voisin de bureau, Michaël Gutnic, qui m'a permis de m'aguerrir dans l'enseignement, de découvrir le travail pédagogique en équipe, etc.

Merci encore à Armelle pour les enseignements bien cadrés de ProbaStat en L2 Mathéco. Merci à tous les enseignants avec qui j'ai pu collaborer (Nicolas Juillet, Laurent, Jade . . .).

Et évidemment, merci à tous les étudiants que j'ai eu pendant ces trois ans, j'ai passé de bons moments avec eux (j'espère que eux aussi !) qui m'ont permis de souffler et de faire des pauses avec la recherche. C'était nécessaire.

Ensuite, je tiens à remercier très chaleureusement toutes les personnes qui s'impliquent dans la vie de l'université, en plus de leur tâche quotidienne, et que j'ai pu croiser dans ce contexte. Surtout celles qui le font, non pas uniquement en vue d'une promotion, mais aussi parce qu'elles estiment que lorsqu'on fait partie d'une structure (telle que l'université), qui nous donne une certaine liberté et un certain confort de vie, cela peut sembler légitime et honnête de lui rendre un peu, notamment en participant à son bon fonctionnement (et pour un éventuel meilleur fonctionnement !). J'ai eu énormément de mal à supporter l'égoïsme ambiant qui règne à l'université. Chacun pour sa pomme et tout ira bien.

Merci à Fabien qui organise et gère les séminaires doctorants depuis trois ans. Merci à Camille, Anais et les autres doctorants qui ont représenté les doctorants au conseil scientifique du labo. Merci aux autres représentants des doctorants à l'École Doctorale, Lionel, Yves et Noëlle.

Merci à tous les doctorants qui font fonctionner l'Addal, l'association des doctorants et docteurs d'Alsace, Kenneth, Cécilia, Elise, Olivier, Tiphaine, Daniel, Benjamin, Natalia, Lionel et les autres.

Merci à Michaël Gutnic pour sa grande implication dans différents organes de l'université, et pour toutes les discussions que j'ai pu avoir avec lui à ce sujet.

Merci à Véronique Bertrand pour les quelques échanges que nous avons eus.

Pour en terminer avec la vie universitaire, je remercie les institutions qui m'ont permis d'étudier dans de très bonnes conditions : l'Université de Strasbourg, l'UFR de math-info et l'IRMA.

Merci également au ministère de la recherche pour le financement de ma thèse pendant trois ans.

Voilà, j'espère ne pas avoir oublié grand monde, sinon il faudra me le dire !

Je ne peux pas ne pas penser maintenant à tous ceux qui comptent énormément pour moi, une fois que je quitte l'université. Je vais essayer de faire bref pour cette partie dont la longueur sera inversement proportionnelle à son importance.

Merci à mes parents, pour tout ce qu'ils ont fait pour mon frère et moi depuis que nous sommes petits. Si j'en suis là, c'est grâce à eux.

Merci à mon père de m'avoir transmis le goût des mathématiques, de m'avoir formé et d'avoir toujours répondu à toutes mes questions à leur sujet (jusqu'à un certain point ;-)

Merci à ma mère de m'avoir transmis le goût de l'effort et du travail, et de m'avoir toujours offert un soutien sans faille.

Merci à toute ma famille, mon frère, mes grands-parents, mes oncles et tantes, mes cousins et ma cousine.

Merci à tous mes ami-e-s. Celles et ceux que je connais depuis longtemps maintenant (le

temps du lycée s'éloigne chaque année mais pas nous;-). Celles et ceux que je connais depuis plus récemment. Merci à ceux qui ont pu venir aujourd'hui, ça me fait très très plaisir !

Merci à ceux qui ont partagé mon quotidien à Strasbourg pendant ces cinq ans, qui m'ont aidé dans les moments difficiles et avec qui j'ai passé beaucoup de bon temps.

Merci à Maria évidemment pour tous les très bons moments vécus ensemble et pour le reste. C'était important pour moi, je n'oublierai pas.

Merci à ceux avec qui j'ai passé du temps quand ils étaient là. Et qui s'éclipsent dès qu'ils peuvent. . . c'est pas normal, faut rester ici !

Et merci à tous les gens que j'ai rencontrés grâce au tennis : la liste commence à être très longue maintenant. Énormément de gens biens et intéressants. Merci à eux de m'avoir permis de respirer autre chose que des études.

Je pense à Léo aussi.

Voilà, c'est terminé. Je suis content.

# Table des matières

Publications et Conférences	11
<b>1 Introduction</b>	<b>13</b>
1.1 Cas univarié . . . . .	13
1.2 Cas multidimensionnel . . . . .	21
1.3 Présentation des travaux de recherche . . . . .	25
<b>2 Estimating the conditional tail index by integrating a kernel conditional quantile estimator</b>	<b>29</b>
2.1 Introduction . . . . .	29
2.2 A class of conditional tail index estimators . . . . .	30
2.3 Main result . . . . .	32
2.4 A small simulation study . . . . .	36
2.4.1 Optimization of $(m_x, k_x)$ - AMSE minimization . . . . .	37
2.4.2 Optimization of $h_{2,n}$ - Subsample division . . . . .	38
2.4.3 Graphs and computational issues . . . . .	38
2.5 Proofs . . . . .	41
2.5.1 Preliminary results . . . . .	41
2.5.2 Proofs of the main results . . . . .	44
<b>3 Nonparametric regression estimation of conditional tails - the random</b>	



---

<b>covariate case</b>	<b>49</b>
3.1 Introduction . . . . .	50
3.2 Construction of the estimator and asymptotic properties . . . . .	52
3.3 Bias-corrected estimation of $\gamma(x)$ . . . . .	58
3.4 Simulation results . . . . .	62
3.5 Proofs . . . . .	69
3.5.1 Proof of Lemma 3.1 . . . . .	69
3.5.2 Proof of Lemma 3.2 . . . . .	70
3.5.3 Proof of Theorem 3.1 . . . . .	75
3.5.4 Proof of Theorem 3.2 . . . . .	76
3.5.5 Proof of Theorem 3.3 . . . . .	77
3.5.6 Proof of Theorem 3.4 . . . . .	78
3.5.7 Proof of Theorem 3.5 . . . . .	80
3.5.8 Proof of Theorem 3.6 . . . . .	83
<b>4 Madogram and asymptotic independence among maxima</b>	<b>85</b>
4.1 Introduction . . . . .	85
4.2 The $F$ -madogram in the asymptotic independence case . . . . .	90
4.3 Inference . . . . .	91
4.3.1 A method-of-moment approach . . . . .	91
4.3.2 The maximum likelihood approach . . . . .	94
4.4 Examples . . . . .	94
4.4.1 A small simulation . . . . .	94
4.4.2 Orgeval Rainfall data . . . . .	96
4.5 Proof . . . . .	97
<b>Conclusions et Perspectives</b>	<b>101</b>





## Publications et Conférences

### Articles publiés :

[1] Gardes, L., Guillou, A. & Schorgen, A. (2012). Estimating the conditional tail index by integrating a kernel conditional quantile estimator, *Journal of Statistical Planning and Inference*, **142**, 1586–1598.

### Articles soumis :

[2] Goegebeur, Y., Guillou, A. & Schorgen, A. (2012). Nonparametric regression estimation of conditional tails - the random covariate case.

[3] Guillou, A., Naveau, P. & Schorgen, A. (2012). Madogram and asymptotic independence among maxima.

### Conférences :

[1] Madogram and asymptotic independence, Workshop Young Researchers, Aussois, septembre 2010.

[2] Estimation de l'indice des valeurs extrêmes et programmation, Premières Rencontres R, Bordeaux, juillet 2012.

### Séminaires :

[1] Introduction à la Statistique, séminaire doctorants, Strasbourg, mars 2011.

[2] Indépendance asymptotique en théorie des valeurs extrêmes, séminaire équipe statistique, Strasbourg, mai 2011.

[3] Marketing et tests statistiques, séminaire doctorants, Strasbourg, novembre 2011.



# Chapitre 1

## Introduction

L'analyse des valeurs extrêmes est un sujet très répandu ces dernières décennies dans la littérature car elle permet d'apporter des réponses à de nombreux problèmes pratiques (e.g. en assurance, en environnement). Une telle analyse requiert l'estimation d'un indice de queue qui donne une indication essentielle sur la forme de la queue de distribution. De nombreux estimateurs de ce paramètre existent mais par contre peu dans le cas où la variable d'intérêt  $Y$  est observée simultanément avec une covariable, et ceci malgré le vaste champ d'applications qui en résultent (e.g. en hydrologie). Par ailleurs, l'extension de cette théorie des valeurs extrêmes dans le cadre multivarié pose de nombreux problèmes encore ouverts. Nous allons donc aborder dans cette thèse ces deux sujets, mais avant cela nous allons rappeler quelques éléments essentiels de la théorie existante, sans pour autant se vouloir exhaustif, mais juste pour donner les éléments utiles à la compréhension de nos contributions.

### 1.1 Cas univarié

L'analyse des valeurs extrêmes repose principalement sur les distributions limites des extrêmes et leurs domaines d'attraction. Ces distributions apparaissent comme les seules

distributions limites possibles du maximum d'un échantillon de variables aléatoires indépendantes et identiquement distribuées (iid)  $X_1, \dots, X_n$  de loi  $F$ , ayant un point terminal noté  $\tau_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$ . Plus spécifiquement, on dit que  $F$  est dans le domaine d'attraction  $D_\gamma$  si et seulement si il existe deux suites normalisantes  $(a_n)$ ,  $a_n \in \mathbb{R}_+^*$  et  $(b_n)$ ,  $b_n \in \mathbb{R}$  telles que  $\forall x \in \mathbb{R}$  :

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( a_n^{-1} \left\{ \max_{1 \leq i \leq n} X_i - b_n \right\} \leq x \right) = H_\gamma(x), \quad (1.1)$$

où

$$H_\gamma(x) = \begin{cases} \exp\left(- (1 + \gamma x)^{-\frac{1}{\gamma}}\right) & \text{pour tout } x \text{ tel que } 1 + \gamma x > 0, \text{ si } \gamma \neq 0, \\ \exp\left(- \exp(-x)\right) & \text{pour tout } x \in \mathbb{R}, \text{ si } \gamma = 0, \end{cases}$$

s'appelle la distribution des valeurs extrêmes généralisées (distribution  $\text{GEV}_\gamma$ ).

De ce théorème limite, il découle immédiatement que le comportement de la queue de distribution d'une fonction est complètement caractérisé par un unique paramètre, noté  $\gamma$ , et appelé indice des valeurs extrêmes. Le signe de ce paramètre est un indicateur essentiel sur la forme de la queue de distribution. Il faut donc distinguer les trois cas possibles :

- $\gamma > 0$  correspondant au domaine d'attraction de Fréchet. On a alors  $\tau_F = \infty$  et  $F$  est de type Pareto, c'est-à-dire  $1 - F(x) = x^{-1/\gamma} \ell_F(x)$  où  $\ell_F$  est une fonction à variations lentes en l'infini, i.e. une fonction vérifiant  $\lim_{x \rightarrow \infty} \frac{\ell_F(tx)}{\ell_F(x)} = 1$ , pour tout  $t > 0$ . On peut citer comme exemple classique de lois dans ce domaine, dites lois à queue lourde : lois Pareto strict, Burr, Fréchet, ... ;
- $\gamma = 0$  correspondant au domaine d'attraction de Gumbel. Le point terminal  $\tau_F$  peut alors être fini ou non et les queues des lois appartenant à ce domaine décroissent de manière exponentielle. Des exemples standards sont la Gamma, Logistic, Lognormale, Weibull, ... ;

•  $\gamma < 0$  correspondant au domaine d'attraction de Weibull. On a alors  $\tau_F < \infty$  et  $1 - F(x) = (\tau_F - x)^{-1/\gamma} \ell_F((\tau_F - x)^{-1})$  où  $\ell_F$  est également une fonction à variations lentes en l'infini. On peut citer par exemple la Reverse Burr ou l'uniforme.

Bien que très utilisée en pratique, cette approche des extrêmes est très controversée dans la littérature dans la mesure où elle ne tient compte que d'une seule observation, la plus grande. Pour résoudre ce problème une approche concurrentielle a été proposée par Balkema et de Haan (1974) ainsi que Pickands (1975). Il s'agit de regarder la loi des excès au-delà d'un seuil fixe  $u$ . Plus précisément, partant d'un échantillon  $X_1, \dots, X_n$ , on se fixe un seuil  $u$  grand et on ne considère que les  $N_u$  observations dépassant ce seuil. On note  $Y_i$ ,  $i = 1, \dots, N_u$ , les excès au-delà du seuil  $u$ , définis comme l'écart entre l'observation et  $u$ . La fonction de répartition des excès au-delà de  $u$  est alors donnée par

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

Le théorème de Pickands nous donne alors la convergence uniforme suivante :

**Théorème de Pickands (1975).** *F appartient au domaine d'attraction  $D_\gamma$  si et seulement si il existe une fonction  $\sigma(\cdot)$  positive et un réel  $\gamma$  tels que la loi des excès  $F_u$  peut être uniformément approchée par une distribution de Pareto généralisée (GPD) notée  $G_{\gamma, \sigma}$ , i.e.*

$$\lim_{u \uparrow \tau_F} \sup_{x \in (0, \tau_F - u)} |F_u(x) - G_{\gamma, \sigma(u)}(x)| = 0,$$

$$\text{où } G_{\gamma, \sigma}(x) = \begin{cases} 1 - (1 + \frac{\gamma x}{\sigma})^{-\frac{1}{\gamma}} & \text{si } \gamma \neq 0, x \geq 0 \text{ et } x < -\frac{\sigma}{\gamma} \text{ si } \gamma < 0 \\ 1 - \exp(-\frac{x}{\sigma}) & \text{si } \gamma = 0, x \geq 0. \end{cases}$$

Cette approche est communément appelée dans la littérature l'approche « pics au-delà d'un seuil » (approche POT, « Peaks-Over-Threshold »).



Suivant la méthode utilisée, par maximum ou par excès, on doit estimer l'indice des valeurs extrêmes  $\gamma$ , ou un couple de paramètres  $(\gamma, \sigma)$ . Pour cela, toutes les méthodes classiques d'estimation sont à notre disposition. Nous allons succinctement énumérer quelques unes de ces approches ci-dessous.

## Estimation des paramètres

### a. Estimateur de Hill dans le cas $\gamma > 0$

Intéressons-nous aux excès relatifs  $E_{j,u}$ , au-delà du seuil  $u$ , i.e.  $E_{j,u} := \frac{X_j}{u}$  avec  $X_j > u$ . On peut alors facilement vérifier que  $\mathbb{P}(E_{j,u} > x | E_{j,u} > 1) \rightarrow x^{-1/\gamma}$ , quand  $u \rightarrow \infty$ ,  $x > 1$ . De cette convergence, en formant la vraisemblance basée sur cette distribution limite, on vérifie facilement que l'estimateur de Hill (1975) défini comme suit :

$$\hat{\gamma}_{k,n}^{(H)} = \frac{1}{k} \sum_{j=1}^k \ln X_{n-j+1,n} - \ln X_{n-k,n}$$

avec  $X_{1,n}, \dots, X_{n,n}$  les statistiques d'ordre associées à l'échantillon  $X_1, \dots, X_n$ , n'est rien d'autre que l'estimateur du maximum de vraisemblance dans le cas où le seuil  $u = X_{n-k,n}$ . Cet estimateur est sans conteste le plus célèbre dans la littérature puisque non seulement il est l'estimateur du maximum de vraisemblance mais il peut aussi être interprété graphiquement, via un Pareto quantile plot, correspondant au graphe de  $(\ln \frac{n+1}{j}, \ln X_{n-j+1,n})$ . Dans le cas d'une distribution dans le domaine d'attraction de Fréchet, ce graphe est approximativement linéaire, avec une pente  $\gamma$ , pour les petites valeurs de  $j$ , c'est-à-dire les points extrêmes. Tout ceci rend cet estimateur de Hill très intéressant d'un point de vue pratique.

Par ailleurs, d'un point de vue théorique, il possède toutes les propriétés asymptotiques classiques que l'on peut demander à un estimateur : il converge en probabilité (Mason, 1982), presque sûrement (Deheuvels *et al.*, 1988) et il est asymptotiquement normal (Beirlant et Teugels, 1989). Concernant cette normalité, elle requiert des hypothèses supplémentaires, et notamment de spécifier la vitesse de convergence de  $\ell(tx)/\ell(x)$  vers

1 quand  $x$  tend vers l'infini. Cette hypothèse est régie notamment par un paramètre de second ordre,  $\rho$ , nécessairement négatif et difficile à estimer en pratique. Plus ce paramètre  $\rho$  sera proche de 0, plus l'estimateur de Hill exhibera un biais important.

### b. Estimateur à noyau dans le cas $\gamma > 0$

Csörgő *et al.* (1985) ont proposé les estimateurs à noyau  $K_{k,n}$ , comme extension de l'estimateur de Hill définis de la façon suivante :

$$\widehat{\gamma}_{k,n}^{(K)} = \frac{\sum_{j=1}^k \frac{j}{k} K\left(\frac{j}{k}\right) \left(\ln X_{n-j+1,n} - \ln X_{n-j,n}\right)}{\frac{1}{k} \sum_{j=1}^k K\left(\frac{j}{k}\right)}$$

où  $K$  représente un noyau d'intégrale égale à 1.

Cet estimateur peut à nouveau s'interpréter graphiquement à partir d'un Pareto quantile plot et ses propriétés asymptotiques sont du même type que celles de l'estimateur de Hill.

### c. Estimateurs par la méthode des moments dans le cas $\gamma \in \mathbb{R}$

Dans le cas d'une distribution GPD  $G_{\gamma,\sigma}$ , l'estimateur des moments est basé sur le fait que

$$\mathbb{E} \left[ \left( 1 + \frac{\gamma X}{\sigma} \right)^r \right] = \frac{1}{1 - r\gamma} \text{ si } 1 - r\gamma > 0.$$

Il en découle l'expression des paramètres en fonction des deux premiers moments  $\mu_1 := \mathbb{E}(X)$  et  $\mu_2 := \mathbb{E}(X^2)$

$$\begin{aligned} \gamma &= \frac{1}{2} \left( 1 - \frac{\mu_1^2}{\mu_2 - \mu_1^2} \right) \\ \sigma &= \frac{1}{2} \mu_1 \left( \frac{\mu_1^2}{\mu_2 - \mu_1^2} + 1 \right). \end{aligned}$$

En remplaçant dans les expressions précédentes  $\mu_1$  et  $\mu_2$  par leur version empirique, nous obtenons l'estimateur des moments du couple  $(\widehat{\gamma}_n^{MOM}, \widehat{\sigma}_n^{MOM})$  défini comme suit :

$$\widehat{\gamma}_n^{MOM} = \frac{1}{2} \left( 1 - \frac{\overline{X^2}}{\overline{X^2} - \overline{X}^2} \right), \quad \widehat{\sigma}_n^{MOM} = \frac{1}{2} \overline{X} \left( \frac{\overline{X^2}}{\overline{X^2} - \overline{X}^2} + 1 \right), \quad (1.2)$$

où  $\overline{X}$  et  $\overline{X^2}$  sont les estimateurs empiriques des moments d'ordre 1 et 2 de l'échantillon.

La normalité asymptotique du couple  $(\hat{\gamma}_n^{MOM}, \hat{\sigma}_n^{MOM})$  peut être établie sous la condition  $\gamma < 1/4$ . Pour affaiblir cette condition d'autres alternatives ont été proposées, telle que la méthode des moments pondérés (éventuellement généralisés).

#### d. Estimateurs par la méthode des moments pondérés

Hosking et Wallis en 1987 ont également proposé d'utiliser les deux moments pondérés  $\nu_0$  et  $\nu_1$  suivants :

$$\nu_s = \mathbb{E}[X(1 - G_{\gamma,\sigma}(X))^s] = \frac{\sigma}{(s+1)(s+1-\gamma)}, s \geq 0,$$

pour estimer  $\gamma$  et  $\sigma$ . Les paramètres  $(\gamma, \sigma)$  découlent alors de  $\nu_0$  et  $\nu_1$  par les relations

$$\gamma = 2 - \frac{\nu_0}{\nu_0 - 2\nu_1} \quad \text{et} \quad \sigma = \frac{2\nu_0\nu_1}{\nu_0 - 2\nu_1}.$$

L'estimation est possible en utilisant les versions empiriques de ces moments. Le domaine de validité nécessaire à la normalité asymptotique de ces estimateurs est encore assez restrictif :  $\gamma \in (-1, 1/2)$ ; néanmoins, cette méthode présente un certain nombre d'avantages tant au niveau de la facilité du calcul que des performances à distance finie (cf. Hosking et Wallis, 1987). Une généralisation de cette méthode permettant d'étendre le domaine de validité à  $\gamma \in (-1, 3/2)$  a été proposée par Diebolt *et al.* (2007).

#### e. Estimateurs par la méthode des moments pondérés généralisés

Les moments pondérés généralisés (GPWM) ont été récemment introduit par Diebolt *et al.* (2007) pour une distribution  $G$  donnée de la façon suivante :

$$\nu_\omega = \mathbb{E}(X\omega(G)) = \int_{-\infty}^{\infty} x \omega(G(x))dG(x),$$

où  $\omega$  est une fonction continue convenablement choisie. Par changement de variable, ce moment peut se réécrire comme suit :

$$\nu_\omega = \int_0^1 G^{-1}(u)\omega(u)du.$$

En notant  $W$  la primitive de  $\omega$ , nulle en 0, i.e.  $W(t) = \int_0^t \omega(u)du$ , Diebolt *et al.* (2007)

proposent d'estimer  $\nu_\omega$  par

$$\widehat{\nu}_{\omega,n} = \int_0^1 \mathbb{F}_n^{-1}(u) \omega(u) du \quad (1.3)$$

où  $\mathbb{F}_n$  est la fonction de répartition empirique basée sur l'échantillon  $(X_1, \dots, X_n)$ . Ils ont établi les propriétés asymptotiques de  $\widehat{\nu}_{\omega,n}$  pour des distributions  $\text{GEV}_{\gamma,\sigma,\mu}$  dont la fonction de répartition est définie par  $H_{\gamma,\sigma,\mu}(x) := H_\gamma\left(\frac{x-\mu}{\sigma}\right)$  sous les hypothèses suivantes sur la fonction  $\omega$  :

$$\begin{cases} \omega(t) = O((1-t)^b) & \text{pour } t \text{ proche de } 1, b \geq 0 \\ \omega(t) = O(t^{a'}) & \text{pour } t \text{ proche de } 0, a' > 0. \end{cases}$$

Un exemple de telles fonctions est la fonction  $\omega(t) = t^a(-\ln t)^b, a > a'$ . Dans ce cas, les GPWM pour une distribution  $\text{GEV}_{\gamma,\sigma,\mu}$  peuvent se réécrire comme suit

$$\nu_\omega = \frac{\sigma}{\gamma} \frac{1}{(a+1)^{b-\gamma+1}} \Gamma(b-\gamma+1) - \left(\frac{\sigma}{\gamma} - \mu\right) \frac{1}{(a+1)^{b+1}} \Gamma(b+1) \quad (1.4)$$

et la normalité asymptotique de (1.3) est établie dès lors que  $\gamma < b + \frac{1}{2}$ .

## f. Estimateurs par maximum de vraisemblance

Si on utilise l'approche POT, l'estimation du couple  $(\gamma, \sigma)$  devient alors naturelle, puisqu'elle découle de la maximisation de la fonction de log-vraisemblance. En effet, cette dernière dans le cas d'un échantillon  $\mathbf{X} = (X_1, \dots, X_n)$  issu d'une distribution de Pareto généralisée est donnée par :

$$\mathcal{L}(\mathbf{X}; \gamma, \sigma) = -n \ln \sigma - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^n \ln \left(1 + \frac{\gamma}{\sigma} X_i\right).$$

En dérivant cette expression par rapport aux deux paramètres d'intérêt, nous obtenons un système de deux équations à deux inconnues  $\gamma$  et  $\sigma$ . C'est en résolvant ces équations que nous obtenons les estimateurs du maximum de vraisemblance pour lesquels Smith (1987) a établi la normalité asymptotique sous la condition  $\gamma > -1/2$ .

## 1.2 Cas multidimensionnel

L'approche par maximum peut être généralisée dans le cas multidimensionnel en supposant que, partant de  $n$  vecteurs aléatoires,  $\mathbf{X}_i = \{\mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,d}\}, i = 1, \dots, n$ , iid de dimension  $d$  et de fonction de répartition  $F$  :

il existe des suites positives  $(a_{n,1})_n, \dots, (a_{n,d})_n$  et des suites  $(b_{n,1})_n, \dots, (b_{n,d})_n$  ainsi qu'une fonction de répartition  $G$  de marges non dégénérées telles que, lorsque  $n \rightarrow \infty$ , on ait :

$$\begin{aligned} \mathbb{P} \left( a_{n,1}^{-1} \left\{ \max_{i=1, \dots, n} X_{i,1} - b_{n,1} \right\} \leq x_1, \dots, a_{n,d}^{-1} \left\{ \max_{i=1, \dots, n} X_{i,d} - b_{n,d} \right\} \leq x_d \right) \\ = F^n(a_{n,1}x_1 + b_{n,1}, \dots, a_{n,d}x_d + b_{n,d}) \rightarrow G(x_1, \dots, x_d). \end{aligned} \quad (1.5)$$

La fonction de répartition  $G$  est alors appelée fonction de répartition multivariée de valeurs extrêmes et l'on dit que  $F$  est dans le domaine d'attraction (multivarié) de  $G$  (pour les maxima).

Cette convergence illustre le fait que en dimension  $d$  l'ensemble des lois des valeurs extrêmes perd le caractère paramétrique qu'il possédait en dimension 1.

Dans toute la suite de cette thèse, on se concentrera sur le cas bivarié ( $d = 2$ ) et donc on supposera disposer de  $n$  paires  $(X, Y)$  de loi jointe  $F$ . Sans perdre de généralité, il est classique de supposer que les lois marginales  $F_X$  et  $F_Y$  sont connues de loi Fréchet standard, i.e.  $F_X(t) = F_Y(t) = \exp(-1/t), t > 0$ . Il ne s'agit pas là d'une contrainte puisqu'on peut toujours s'y ramener moyennant une transformation des marges, en posant par exemple  $Z := \frac{1}{\ln F_X(X)}$ .

Dans ce cas, en notant  $M_{X,n} = \max_{1 \leq i \leq n} X_i$  et  $M_{Y,n} = \max_{1 \leq i \leq n} Y_i$ , (1.5) devient

$$\mathbb{P} \left( \frac{M_{X,n} - b_n}{a_n} \leq x, \frac{M_{Y,n} - b_n}{a_n} \leq y \right) = F^n(b_n + a_n x, b_n + a_n y) \longrightarrow G(x, y). \quad (1.6)$$

Puisque les marges sont Fréchet unitaires,  $G(\cdot, \cdot)$  peut se réécrire sous la forme suivante

$$G(x, y) = \exp[-V(x, y)] \quad \text{où} \quad V(x, y) = \int_0^1 \max \left( \frac{w}{x}, \frac{1-w}{y} \right) dH(w), \quad (1.7)$$

avec  $H(\cdot)$  une mesure positive finie sur  $[0, 1]$  telle que  $\int_0^1 \omega dH(\omega) = \int_0^1 (1-\omega) dH(\omega) = 1$ ,  $V$  est homogène d'ordre -1, i.e.  $V(tx, ty) = t^{-1}V(x, y)$ ,  $\forall t > 0$  et  $G$  est max-stable, i.e.  $G^t(tx, ty) = G(x, y)$ . La fonction  $V(\cdot, \cdot)$  est appelée fonction de dépendance extrême.

Tout le travail statistique de l'étude en bivarié réside alors dans l'estimation des structures de dépendance. Selon les auteurs et les points de vue utilisés, ces structures peuvent se décliner sous différentes formes, chaque fois liées les unes aux autres : la fonction de dépendance extrême  $V(\cdot)$ , la fonction de dépendance de Pickands  $A(\cdot)$  (Pickands, 1981) liée à  $V$  par  $V(x, y) = (1/x + 1/y)A(y/(x+y))$ , la fonction de dépendance  $l(\cdot)$  (Huang, 1992) liée à  $V$  par  $l(x, y) = V(1/x, 1/y)$ , ...

Pour estimer cette fonction de dépendance extrême  $V(\cdot)$ , différentes stratégies peuvent être utilisées. Par exemple, Capéràa *et al.* (1997) ont proposé une procédure d'estimation non-paramétrique pour des maxima bivariés, alors que des familles paramétriques ont également été étudiées dans le passé. Plus récemment, une nouvelle approche, appelée madogramme (Naveau *et al.*, 2009), a été introduite en ayant recours à des outils de géostatistique. Elle repose sur le variogramme au 1er ordre (ou madogramme) défini par

$$\nu := \frac{1}{2} \mathbb{E} |F(M_{X,n}) - F(M_{Y,n})|,$$

où  $F$  est la loi commune de  $M_{X,n}$  et  $M_{Y,n}$ .

Matheron (1987) en a obtenu quelques propriétés mais pas dans le cadre des extrêmes.

Pour cela il faut attendre Cooley *et al.* (2006) pour voir apparaître les propriétés classiques du madogramme dans le cas de champs spatiaux extrêmes. En particulier, ils ont établi le lien suivant

$$\theta = \frac{1 + 2\nu}{1 - 2\nu}$$

entre le madogramme et le coefficient extrémal  $\theta = V(1, 1)$  défini par

$$\mathbb{P}(\max(M_{X,n}, M_{Y,n}) \leq u) = \mathbb{P}^\theta(M_{X,n} \leq u)$$

(cf. e.g. Schlather (2002) et Schlather et Tawn (2003)). Le principal inconvénient de ce coefficient est qu'il se concentre uniquement sur les valeurs  $V(x, x)$ , mais ne fournit aucune information sur  $V(x, y)$  pour  $x \neq y$ . Pour cela, Naveau *et al.* (2009) étendent le madogramme en un  $\lambda$ -madogramme défini par

$$\nu(\lambda) = \frac{1}{2} \mathbb{E} |F^\lambda(M_{X,n}) - F^{1-\lambda}(M_{Y,n})|, \text{ pour } \lambda \in (0, 1)$$

et montrent en particulier que

$$V(\lambda, 1 - \lambda) = \frac{c(\lambda) + \nu(\lambda)}{1 - c(\lambda) - \nu(\lambda)} \quad \text{où} \quad c(\lambda) = \frac{3}{2(1 + \lambda)(1 - 2\lambda)}.$$

Cette égalité indique clairement que le  $\lambda$ -madogramme caractérise complètement la fonction de dépendance  $V(x, y)$ .

Toute cette théorie des valeurs extrêmes bivariée apporte des réponses pertinentes à moult questions lorsque les observations présentent une dépendance asymptotique, i.e. dans la situation où

$$\lambda = \lim_{v \rightarrow 1} \mathbb{P}(F_Y(Y) > v | F_X(X) > v) > 0$$

(si cette limite existe). Dans le cas où  $\lambda = 0$ , on parle alors d'indépendance asymptotique et dans ce contexte des insuffisances de cette théorie surgissent incitant ainsi à proposer une alternative de modélisation. Plus précisément, nous dirons que  $M_{X,n}$  et  $M_{Y,n}$  sont



*asymptotiquement indépendant* si  $G(x, y) = G(x, \infty)G(\infty, y)$  pour tout  $x, y$ . En d'autres termes, l'indépendance asymptotique exprime le fait que la probabilité que deux composantes soient simultanément grandes est négligeable par rapport à la probabilité qu'une seule le soit. Ce cas est très fréquent en pratique comme dans le cas d'une distribution bivariée normale avec coefficient de corrélation  $|\rho| < 1$  où les marges sont asymptotiquement indépendantes. Malheureusement si on veut estimer la probabilité que à la fois  $X$  et  $Y$  excèdent un seuil élevé, la convergence (1.6) est sans intérêt. En effet, en passant au logarithme on peut montrer que

$$n\mathbb{P}\left(\frac{X - b_n}{a_n} > x, \frac{Y - b_n}{a_n} > y\right) = \ln G(x, y) - \ln G(x, \infty) - \ln G(\infty, y)$$

qui est exactement égal à 0 dans le cas de l'indépendance asymptotique. Pour résoudre ce problème, Ledford et Tawn (1996, 1997) ont introduit un sous-modèle en supposant que la fonction  $t \mapsto q(t) := \mathbb{P}(1 - F_X(X) < t, 1 - F_Y(Y) < t)$  est à variations régulières en 0 d'indice  $1/\eta$ , i.e.  $q(t) = t^{1/\eta}\ell(t)$  où  $\ell(\cdot)$  est à variations lentes en 0. Le paramètre  $\eta$  est appelé le coefficient de dépendance de queue. Ledford et Tawn (1996, 1997) motivent leur modèle par de nombreux exemples de loi bivariée satisfaisant cette condition ; nous faisons référence à Heffernan (2000) pour une liste de telles distributions.

Le cas  $\eta = 1$  et  $\lim_{t \rightarrow 0} \ell(t) = c$ , pour  $0 < c \leq 1$  correspond au cas de dépendance asymptotique de degré  $c$ , tandis que  $\eta \in (0, 1)$  ou  $\eta = 1$  avec  $\lim_{t \rightarrow 0} \ell(t) = 0$ , implique l'indépendance asymptotique de degré  $2\eta - 1$ . Ledford et Tawn (1996) identifient trois types d'indépendance asymptotique :

- si  $\eta \in (0, 1/2)$ , il est moins fréquent d'observer à la fois  $X$  et  $Y$  au delà d'un même seuil élevé que dans le cas  $X$  et  $Y$  indépendants ;
- si  $\eta = 1/2$  : les extrêmes de  $X$  et  $Y$  sont proches de l'indépendance et exactement indépendants si  $\ell(\cdot) = 1$  ;
- si  $\eta \in (1/2, 1)$  ou  $\eta = 1$  et  $\ell(t) \rightarrow 0$ , il est plus fréquent d'observer que  $(X, Y)$  dépasse un même seuil élevé que dans le cas d'indépendance.

Par conséquent le degré de la dépendance asymptotique dépend à la fois de  $\eta$  et de  $\ell$ . Des grandes valeurs de  $\eta$  indiquent une forte association entre les extrêmes des deux composantes  $X$  et  $Y$ .

D'autres mesures de dépendance ont été proposées dans la littérature, mais comme elles ne seront pas utilisées dans la suite de cette thèse, nous ne les introduirons pas ici. Nous renvoyons néanmoins le lecteur à Joe (1993) et Coles *et al.* (2000) notamment.

### 1.3 Présentation des travaux de recherche

L'estimation de l'indice de queue dans un contexte de régression avec covariables n'a été que très peu abordée dans la littérature des valeurs extrêmes. On peut distinguer selon la nature de la variable explicative deux types de modèles :

- le modèle dit à *design fixe* pour lequel les données sont des couples  $\{(Y_i, x_i), i = 1, \dots, n\}$  où les observations  $Y_i$  sont des variables aléatoires réelles indépendantes et les  $x_i$  sont des points d'observations non aléatoires ;
- le modèle dit à *design aléatoire* pour lequel les données sont des couples  $\{(Y_i, X_i), i = 1, \dots, n\}$  de variables iid.

Dans ce contexte, l'indice de queue dépend de la covariable et sera appelé dans la suite l'indice de queue conditionnel. Ce sujet a déjà été abordé par Smith (1989) ou Davison et Smith (1990) qui ont proposé des modèles paramétriques, tandis que des approches semi-paramétriques ont été introduites dans Hall et Tajvidi (2000) ou Beirlant et Goegebeur (2003). Des approches non paramétriques sont également possibles par exemple basées sur des splines (Chavez-Demoulin et Davison, 2005) ou des polynômes locaux (Davison et Ramesh, 2000). D'autres méthodes ont également été récemment étudiées dans la littérature, comme dans Beirlant et Goegebeur (2004), Gardes et Girard

(2008, 2010), Goegebeur et de Wet (2012). En dépit de ces contributions dans le cas de covariable fixe, l'estimation d'indice de queue conditionnel dans le cas de covariable aléatoire n'a reçu que peu d'attention. Wang et Tsai (2009) ont considéré l'estimation par maximum de vraisemblance dans le cas particulier du modèle de Hall (Hall, 1982) où l'indice de queue est relié à la covariable par une fonction de lien logarithmique. Plus récemment, Daouia *et al.* (2011) ont étudié l'estimation de quantiles extrêmes dans le cas d'un modèle de type Pareto conditionnel avec covariable aléatoire et ils ont utilisé un nombre fixe de ces estimations pour construire un estimateur de type Pickands (1975).

Dans le chapitre 2, nous proposons de nouveaux estimateurs d'indice de queue conditionnel basés sur une classe de fonctions satisfaisant des conditions faibles dans le cas d'un modèle à *design fixe*. Nous supposons juste que la distribution conditionnelle de  $Y$  sachant  $x$  est à queue lourde mais aucune hypothèse paramétrique n'est faite sur la covariable  $x$ . Plus spécifiquement en notant  $F(\cdot; x)$  la fonction de répartition conditionnelle de la variable réponse  $Y$  se donnant  $x$ , on a le modèle suivant :

$$\bar{F}(y; x) = 1 - F(y; x) = y^{-\frac{1}{\gamma(x)}} \ell(y; x), y > 0$$

où  $\ell$  est une fonction à variations lentes en l'infini, telle que

$$\lim_{y \rightarrow \infty} \frac{\ell(\lambda y; x)}{\ell(y; x)} = 1, \forall \lambda > 0.$$

La fonction  $\gamma(x)$  décrit la lourdeur de la queue de la distribution réponse conditionnelle et est une fonction qui doit être adéquatement estimée à partir des données disponibles. Dans ce but, nous proposons un estimateur à noyau basé uniquement sur les  $Y_i$  pour lesquels les covariables  $x_i$  appartiennent à un voisinage de  $x$  (une boule dont le rayon tend vers 0). Dans ce contexte, il est en effet naturel de se servir de l'information apportée par la covariable pour estimer les paramètres d'intérêt, i.e. de travailler localement, dans un voisinage du point étudié. L'utilisation des noyaux est alors une façon habituelle

d'y parvenir. Nous étudions dans ce chapitre 2 les propriétés asymptotiques de notre estimateur à noyau de  $\gamma(x)$  et son comportement sur simulations.

Dans le chapitre 3, un problème similaire est abordé, mais cette fois-ci dans le cadre d'un *design aléatoire*. Nous proposons des familles d'estimateurs non paramétriques de l'indice de queue conditionnel d'une distribution de type Pareto en présence de covariables aléatoires. Ces familles sont construites à partir de statistiques de la forme

$$T_n^{(s,t)}(x, K) := \frac{1}{n} \sum_{i=1}^n K_h^s(x - X_i) (\ln Y_i - \ln \omega_n)_+^t \mathbf{1}\{Y_i > \omega_n\},$$

où  $s \geq 1$ ,  $t \geq 0$ ,  $K_h(x) := K(x/h)/h^p$  avec  $K$  une densité sur  $\mathbb{R}^p$ ,  $h = h_n$  une suite non aléatoire tendant vers 0 et  $\omega_n$  un seuil, également non aléatoire, mais lui tendant vers l'infini quand  $n \rightarrow \infty$ .

Les propriétés de ces familles d'estimateurs ont été obtenues sous des hypothèses sur la distribution réponse conditionnelle, la fonction poids et la fonction densité des covariables. Nous introduisons aussi des versions à biais corrigé de ces estimateurs d'indice de queue conditionnel et proposons dans ce contexte un estimateur consistant du paramètre de second ordre  $\rho(x)$ . Les performances à distance finie de nos estimateurs sont illustrées sur des simulations.

Le chapitre 4 de cette thèse concerne la théorie des valeurs extrêmes dans un cadre multivarié et plus spécifiquement les mesures de dépendance entre extrêmes. Dans ce contexte, on s'intéresse souvent à l'étude du comportement de processus au-delà de certains niveaux, niveaux qui peuvent causer des dommages structurels irréversibles, comme par exemple les tremblements de terre sur les bâtiments. Dans le but de faire de l'inférence, il est nécessaire d'extrapoler au-delà des données et la théorie des valeurs extrêmes multivariées entre alors en jeu en fournissant le contexte mathématique approprié. En particulier, une description adéquate des dépendances extrémales entre

les variables d'intérêt est nécessaire. À cette fin, nous introduisons des estimateurs de paramètres décrivant la dépendance entre variables à des niveaux extrêmes et nous étudions leurs propriétés asymptotiques. Des simulations et l'étude d'un jeu de données réelles illustrent le bon comportement, à distance finie, de ces estimateurs.

Cette thèse offre de nombreuses perspectives, tant sur le plan pratique que théorique dont une liste non exhaustive sera présentée en conclusion.

# Chapitre 2

## Estimating the conditional tail index by integrating a kernel conditional quantile estimator

### Abstract

This chapter deals with the estimation of the tail index of a heavy-tailed distribution in presence of covariates. A class of estimators is proposed in this context and its asymptotic normality established under mild regularity conditions. These estimators are functions of a kernel conditional quantile estimator depending on some tuning parameters. The finite sample properties of our estimators are illustrated on a small simulation study.

### 2.1 Introduction

Tail index estimation has been extensively studied in the literature and several estimators proposed. The most popular semi-parametric estimator was proposed by Hill (1975) in the context of heavy-tailed distributions. Diebolt *et al.* (2008a) considered the case of

Weibull-tail distributions and the general case was studied by Dekkers *et al.* (1989). The aim of this chapter is to extend this framework to the case where some covariate  $x$  is recorded simultaneously with the variable of interest  $Y$ . In this case, the tail index depends on the covariate  $x$  and is referred in the following to the conditional tail index.

Such a context has been already considered by Smith (1989) or Davison and Smith (1990) who proposed parametric models, while semi-parametric approaches have been used in Hall and Tajvidi (2000) or Beirlant and Goegebeur (2003). Also, fully non-parametric methods can be used for instance based on splines (see Chavez-Demoulin and Davison, 2005) or local polynomials (see Davison and Ramesh, 2000). These latter estimators have been then extended by Beirlant and Goegebeur (2003) in case of multidimensional covariates. More recently, Gardes and Girard (2010) have addressed the problem of the estimation of a conditional extreme quantile via a nearest neighbor approach.

We propose here new conditional tail index estimators based on a class of functions satisfying some mild conditions. We only assume that the conditional distribution of  $Y$  given  $x$  is heavy-tailed whereas no parametric assumption on the covariates  $x$  is made. The conditional distribution of  $Y$  given  $x$  is then of polynomial-type, with a rate of convergence driven by the conditional tail index.

The remainder of this chapter is organized as follows. Our class of estimators is given in Section 2.2. Their asymptotic normality is established in Section 2.3. In Section 2.4, the finite sample properties of our estimators are illustrated. All the proofs are postponed to Section 2.5.

## 2.2 A class of conditional tail index estimators

Let  $E$  be a metric space associated to a distance  $d$ . For  $x \in E$ , denote by  $F(\cdot, x)$  (resp.  $q(\cdot, x)$ ) the conditional distribution function (resp. the conditional quantile function) of

$Y$  given  $x$ . We assume that for all  $x \in E$ ,

$$\bar{F}(y, x) = y^{-1/\gamma(x)} L(y, x), \quad (2.1)$$

or, equivalently, for  $\alpha \in (0, 1)$ ,

$$q(\alpha, x) = \bar{F}^{\leftarrow}(\alpha, x) = \alpha^{-\gamma(x)} \ell(\alpha^{-1}, x),$$

where  $\bar{F}^{\leftarrow}(\alpha, x) = \inf\{t, \bar{F}(t, x) \leq \alpha\}$  denotes the generalized inverse of the conditional survival function and, for  $x$  fixed,  $L(\cdot, x)$  and  $\ell(\cdot, x)$  are slowly varying functions, that is for all  $\lambda > 0$ ,

$$\lim_{y \rightarrow \infty} \frac{L(\lambda y, x)}{L(y, x)} = \lim_{y \rightarrow \infty} \frac{\ell(\lambda y, x)}{\ell(y, x)} = 1.$$

Here,  $\gamma(\cdot)$  is a unknown positive function of the covariate referred in the following to the conditional tail index. For a given  $x \in E$ , our aim is to propose an estimator of  $\gamma(x)$ . Suppose that we have at our disposal pairs  $(Y_1, x_1), \dots, (Y_n, x_n)$  of independent observations from model (2.1) where the design points  $x_1, \dots, x_n$  are assumed to be non random. We propose to estimate the conditional tail index as a function of the kernel conditional quantile estimator  $\hat{q}_n(\alpha, x) := \hat{F}_n^{\leftarrow}(\alpha, x)$ , with

$$\hat{F}_n(y, x) = \sum_{i=1}^n \mathbf{1} \left( \frac{d(x, x_i)}{h_{1,n}} \in (0, 1) \right) K \left( \frac{y - Y_i}{h_{2,n}} \right) / \sum_{i=1}^n \mathbf{1} \left( \frac{d(x, x_i)}{h_{1,n}} \in (0, 1) \right), \quad (2.2)$$

where  $h_{1,n}$  and  $h_{2,n}$  are non random positive sequences. This kernel estimator was defined for instance in Ferraty and Vieu (2006) and the function  $K(\cdot)$  is of the following form

$$K(v) = \int_v^{\infty} g(s) ds,$$

where  $g(\cdot)$  is a bounded probability function with support included in  $[-1, 1]$ . Our class of conditional tail index estimators is given by

$$\hat{\gamma}(u_{n,x}, x) = \int_0^{u_{n,x}} \Psi(\alpha, u_{n,x}, x) \ln \hat{q}_n(\alpha, x) d\alpha, \quad (2.3)$$



where for all  $u \in (0, 1)$ ,  $\Psi(\cdot, u, x)$  is a non-null continuous function in  $L_1(0, 1)$  such that  $\int_0^u \Psi(\alpha, u, x) d\alpha = 0$  and  $u_{n,x}$  is a positive sequence. The sequence  $u_{n,x}$  is introduced in order to select only the largest observations to estimate  $\gamma(x)$ . Its choice is thus very important in practice. Note that the estimator  $\widehat{\gamma}(u_{n,x}, x)$  depends only on the  $Y_i$ 's for which the corresponding  $x_i$ 's belong to the ball  $B(x, h_{1,n}) = \{t \in E; d(t, x) \leq h_{1,n}\}$ . The smoothness of the function  $\widehat{F}_n(\cdot, x)$  is also controlled by the bandwidth  $h_{2,n}$ . Finally, note that  $\widehat{F}_n(y, x) \sim \overline{F}_n(y, x)$  as  $h_{2,n} \rightarrow 0$ , where

$$\overline{F}_n(y, x) = \frac{\sum_{i=1}^n \mathbf{1}\left(\frac{d(x, x_i)}{h_{1,n}} \in (0, 1)\right) \mathbf{1}(Y_i > y)}{\sum_{i=1}^n \mathbf{1}\left(\frac{d(x, x_i)}{h_{1,n}} \in (0, 1)\right)},$$

is the empirical estimator of the conditional survival function. The same result is established for the associated conditional quantile in Lemma 2.1.

## 2.3 Main result

In this section, we give the useful notations and assumptions in order to establish the asymptotic normality of our estimators.

**(A.1)** The slowly varying function  $\ell(\cdot, x)$  is normalized.

Assumption **(A.1)** is equivalent to suppose that for all  $y > 1$ ,

$$\ell(y, x) = c(x) \exp \int_1^y \frac{\Delta(v, x)}{v} dv,$$

where  $c(x) > 0$  and  $\Delta(v, x) \rightarrow 0$  as  $v \rightarrow \infty$ .

**(A.2)** The function  $|\Delta(\cdot, x)|$  is ultimately decreasing.

The largest oscillation of the log-quantile function with respect to its second variable is defined for all  $a \in (0, 1/2)$  as :

$$\omega_n(a) = \sup \left\{ \left| \ln \frac{q(\alpha, x)}{q(\alpha, x')} \right|, \alpha \in (a, 1 - a), (x, x') \in B(x, h_{1,n}) \times B(x, h_{1,n}) \right\}.$$

**(A.3)** The function  $|\Delta(\cdot, x)|$  is regularly varying with index  $\rho(x) < 0$ , that is for all  $\lambda > 0$ ,

$|\Delta(\lambda y, x)|/|\Delta(y, x)| \rightarrow \lambda^{\rho(x)}$  as  $y \rightarrow \infty$ . Moreover  $\Delta(\cdot, x)$  is assumed to have a constant sign at infinity.

Conditions **(A.1)** and **(A.3)** imply that for all  $v > 0$ ,

$$\ln \frac{\ell(vy, x)}{\ell(y, x)} = \Delta(y, x) \frac{v^{\rho(x)} - 1}{\rho(x)} (1 + o(1)) \text{ as } y \rightarrow \infty,$$

which is the so-called second order condition classically used to establish the asymptotic normality of tail index estimators. Note that the second order parameter  $\rho(x)$  controls the rate of convergence of  $\ell(\lambda y, x)/\ell(y, x)$  to 1. In particular if  $\rho(x)$  is close to 0, this convergence is slow and thus the estimation of the conditional tail index is difficult. Let  $m_x = m_{n,x}$  be the number of covariates in the ball  $B(x, h_{1,n})$ . The covariates belonging to the ball  $B(x, h_{1,n})$  are denoted by  $\{x_i^*, i = 1, \dots, m_x\}$  and the associated observations by  $\{Y_i^*, i = 1, \dots, m_x\}$ . The order statistics are denoted by  $Y_{1,m_x}^* \leq \dots \leq Y_{m_x,m_x}^*$  and the corresponding covariates by  $\{x_{(1)}^*, \dots, x_{(m_x)}^*\}$ . We introduce

$$\beta_i = \frac{i}{m_x} \quad \text{for all } i = 0, \dots, m_x.$$

Let  $k_x = k_{n,x}$  be a sequence of integers such that  $1 < k_x < m_x$  and assume the following two conditions.

**(B.1)** Let  $u \in (0, 1)$ . There exists a positive function  $\Phi(\cdot, x)$  such that for all  $\beta \in (0, u]$  and for some  $\xi > 0$  :

$$\frac{u}{\beta} \int_0^\beta \Psi(\alpha, u, x) d\alpha = \Phi(\beta/u, x)$$

**(B.2)** There exists a positive function  $h(\cdot, x)$  such that for  $i = 1, \dots, k_x$

$$\text{with } \int_0^1 \Phi(\alpha, x) d\alpha = 1 \text{ and } \int_0^1 \Phi^{2+\xi}(\alpha, x) d\alpha < \infty.$$

$$\left| i\Phi\left(\frac{i}{k_x}, x\right) - (i-1)\Phi\left(\frac{i-1}{k_x}, x\right) \right| \leq h\left(\frac{i}{k_x}, x\right),$$

with

$$\int_0^1 h(\alpha, x) \max(\ln(1/\alpha), 1) d\alpha < \infty.$$

We consider the case where the threshold  $u_{n,x}$  in (2.3) is taken equal to  $\beta_{k_x} = \frac{k_x}{m_x}$ . Thus, the choice of the sequence  $u_{n,x} \in (0, 1)$  is replaced by the choice of a sequence of integers  $k_x \in \{2, \dots, m_x - 1\}$ . We are now in position to state our main result, that is the asymptotic normality of  $\widehat{\gamma}(k_x/m_x, x)$ .

**Theorem 2.1.** *Suppose (A.1), (A.2), (A.3), (B.1) and (B.2) hold. If  $k_x \rightarrow \infty$ ,  $m_x/k_x \rightarrow \infty$ ,*

$$k_x^{1/2} \Delta \left( \frac{m_x}{k_x}, x \right) \rightarrow \lambda(x) \in \mathbb{R}, \quad \frac{k_x^{1/2} h_{2,n}}{q(k_x/m_x, x)} \rightarrow 0 \text{ and } k_x^2 \omega_n(m_x^{-(1+\delta)}) \rightarrow 0,$$

for some  $\delta > 0$  then

$$k_x^{1/2} \left( \widehat{\gamma}(k_x/m_x, x) - \gamma(x) - \Delta \left( \frac{m_x}{k_x}, x \right) \mathcal{AB}(\Phi, \rho(x), x) \right)$$

converges to a  $\mathcal{N}(0, \gamma^2(x) \mathcal{AV}(\Phi, x))$  random variable with :

$$\mathcal{AB}(\Phi, \rho(x), x) = \int_0^1 \Phi(\alpha, x) \alpha^{-\rho(x)} d\alpha \text{ and } \mathcal{AV}(\Phi, x) = \int_0^1 \Phi^2(\alpha, x) d\alpha.$$

Note that in the more particular case  $h_{2,n} = 0$  (i.e. when  $\widehat{F}_n = \overline{F}_n$ ), estimator  $\widehat{\gamma}(k_x/m_x, x)$  is equal to

$$\sum_{i=1}^{k_x} \left( \frac{1}{i} \int_0^{i/m_x} \Psi(\alpha, k_x/m_x, x) d\alpha \right) Z_{i,m_x}(x), \quad (2.4)$$

where

$$Z_{i,m_x}(x) = i \left\{ \ln(Y_{m_x-i+1, m_x}^*) - \ln(Y_{m_x-i, m_x}^*) \right\}, \quad i = 1, \dots, m_x - 1,$$

are the rescaled log-spacings. Estimator (2.4) is similar to the estimator proposed in Gardes and Girard (2008).

Now as an example of functions satisfying the conditions of Theorem 2.1, one can mention for all  $\alpha < u$  :

$$\Psi_\theta(\alpha, u, x) := \begin{cases} \frac{(\theta + 1)^2}{\theta u^{\theta+1}} \left( \frac{u^\theta}{\theta + 1} - \alpha^\theta \right) & \text{if } \theta > 0 \\ \frac{1}{u} \left( \ln \left( \frac{u}{\alpha} \right) - 1 \right) & \text{if } \theta = 0 \\ \frac{1}{u} & \text{if } \theta = \infty. \end{cases}$$

Note that  $\Psi_0(\cdot, \cdot, \cdot)$  and  $\Psi_\infty(\cdot, \cdot, \cdot)$  correspond to the limiting cases of  $\Psi_\theta(\cdot, \cdot, \cdot)$ .

Denote by  $\tilde{\gamma}_\theta(x)$  the estimator obtained by using the function  $\Psi_\theta(\cdot, u, x)$ . Its asymptotic normality is a direct consequence of Theorem 2.1.

**Corollary 2.1.** *Suppose (A.1), (A.2), (A.3) hold. If  $k_x \rightarrow \infty$ ,  $m_x/k_x \rightarrow \infty$ ,*

$$k_x^{1/2} \Delta \left( \frac{m_x}{k_x}, x \right) \rightarrow \lambda(x) \in \mathbb{R}, \quad \frac{k_x^{1/2} h_{2,n}}{q(k_x/m_x, x)} \rightarrow 0 \text{ and } k_x^2 \omega_n(m_x^{-(1+\delta)}) \rightarrow 0,$$

for some  $\delta > 0$  then

$$k_x^{1/2} \left( \tilde{\gamma}_\theta(x) - \gamma(x) - \Delta \left( \frac{m_x}{k_x}, x \right) \frac{\theta + 1}{(1 - \rho(x))(1 + \theta - \rho(x))} \right)$$

converges to a  $\mathcal{N}(0, 2\gamma^2(x)(\theta + 1)/(1 + 2\theta))$  random variable.

Note that when  $\theta = 0$ ,  $\mathcal{AB}(\Phi_0, \rho(x), x) = 1/(1 - \rho(x))^2$  and  $\mathcal{AV}(\Phi_0, x) = 2$  which are the same values than the ones obtained for the conditional Zipf estimator defined in Gardes and Girard (2008). This estimator is an extension in the conditional framework of the Zipf estimator introduced by Kratz and Resnick (1996) and Schultze and Steinebach (1996). When  $\theta = \infty$ ,  $\mathcal{AB}(\Phi_\infty, \rho(x), x) = 1/(1 - \rho(x))$  and  $\mathcal{AV}(\Phi_\infty, x) = 1$ , which correspond to the asymptotic bias and variance of the conditional Hill estimator proposed again in Gardes and Girard (2008). To choose the value of  $\theta$ , one can use for instance the criterion introduced in Gardes and Girard (2010). More precisely, we introduce a modified version

of the Asymptotic Mean Squared Error (AMSE) as follows :

$$\begin{aligned} AMSE^* &:= \Delta^2 \left( \frac{m_x}{k_x}, x \right) \int_{-\infty}^0 \mathcal{AB}^2(\Phi_\theta, \rho, x) d\rho + \frac{\gamma^2(x) \mathcal{AV}(\Phi_\theta, x)}{k_x} \\ &= \frac{\pi(\theta)}{k_x} \left\{ \frac{k_x \Delta^2 \left( \frac{m_x}{k_x}, x \right)}{\mathcal{AV}(\Phi_\theta, x)} + \frac{\gamma^2(x)}{\int_{-\infty}^0 \mathcal{AB}^2(\Phi_\theta, \rho, x) d\rho} \right\} \leq C(x) \frac{\pi(\theta)}{k_x} \end{aligned}$$

where  $C(x)$  is a constant independent of  $\theta$  and

$$\begin{aligned} \pi(\theta) &= \int_{-\infty}^0 \left( \frac{\theta + 1}{(1 - \rho)(1 + \theta - \rho)} \right)^2 d\rho \times \frac{2(1 + \theta)}{1 + 2\theta} \\ &= 2 \left( \frac{\theta + 1}{\theta} \right)^3 \frac{1}{1 + 2\theta} (\theta - 2 \ln(1 + \theta) - 1/(1 + \theta) + 1). \end{aligned}$$

Therefore, a good choice of  $\theta$  could be to take  $\theta_\pi$  minimizing the upper bound  $\pi(\theta)$ .

Numerical optimization leads to  $\theta_\pi \approx 0.6833$ .

## 2.4 A small simulation study

In this section, we illustrate the efficiency of our estimator  $\tilde{\gamma}_\theta(\cdot)$  proposed in Section 2.3 on a small simulation study. Keeping the notations of Sections 2.2 and 2.3 in mind, we recall the expression of our estimator

$$\tilde{\gamma}_\theta(x) = \int_0^{k_x/m_x} \Psi_\theta(\alpha, k_x/m_x, x) \ln \hat{q}_n(\alpha, x) d\alpha. \quad (2.5)$$

The integrated kernel  $K(\cdot)$  is chosen as the integrated biweight kernel :

$$K(x) = \int_x^\infty \frac{15}{16} (1 - s^2)^2 \mathbf{1}(s \in [-1, 1]) ds.$$

Concerning the parameter  $\theta$ , it will be fixed at the value  $\theta_\pi = 0.6833$  as explained in Section 2.3.

Moreover, we assume that the conditional survival function of  $Y$  given the covariate  $x$  is

a Burr-type distribution, that is

$$\bar{F}(y, x) = (1 + y^{-\rho/\gamma(x)})^{1/\rho}, \quad (2.6)$$

where  $\rho < 0$  is the second order parameter defined in condition **(A.3)**. In this particular setting, the slowly varying function  $\ell(\cdot, x)$  and  $\Delta(\cdot, x)$  are given by

$$\begin{aligned} \ell(z, x) &= (1 - z^\rho)^{-\gamma(x)/\rho}, \\ \Delta(z, x) &= \gamma(x) \frac{z^\rho}{1 - z^\rho}, \end{aligned}$$

with  $z > 1$ .

Concerning the conditional tail index, as in Daouia *et al.* (2011), we choose

$$\gamma(x) = \frac{1}{2} \left( \frac{1}{10} + \sin(\pi x) \right) \left( \frac{11}{10} - \frac{1}{2} \exp \left( -64 \left( x - \frac{1}{2} \right)^2 \right) \right). \quad (2.7)$$

In order to proceed to our simulations, the next step is to select  $k_x$  and the two bandwidths  $h_{1,n}$  and  $h_{2,n}$ , i.e to find their optimal values according to some criterion.

### 2.4.1 Optimization of $(m_x, k_x)$ - AMSE minimization

Our estimator only depends on the number of points in the ball  $B(x, h_{1,n})$ . Thus to find the optimal numbers  $(m_x, k_x)$ , we minimize :

$$AMSE(m_x, k_x) = \frac{2(\theta_\pi + 1)}{2\theta_\pi + 1} \frac{\gamma^2(x)}{k_x} + \left[ \frac{\theta_\pi + 1}{(1 - \rho)(1 + \theta_\pi - \rho)} \Delta \left( \frac{m_x}{k_x}, x \right) \right]^2 \quad (2.8)$$

with respect to these parameters.

Now, concerning the two unknown quantities  $\gamma(x)$  and  $\Delta(m_x/k_x, x)$ , they will be estimated using an exponential regression model (see (2.17) in the proof of Theorem 2.1), taking  $\rho(x) = \rho := -1$  as a canonical value. Indeed we tried also to estimate  $\rho$  as in Goegebeur *et*

*al.* (2010) but since it does not improve the estimation of  $\gamma(x)$ , we keep this usual choice for  $\rho$  (see for instance Feuerverger and Hall (1999) or Beirlant *et al.* (1999)). This leads to a least squares estimator for  $\gamma(x)$  and  $\Delta(\frac{m_x}{k_x}, x)$ , denoted respectively by  $\widehat{\gamma}_{LS}(x)$  and  $\widehat{\Delta}_{LS}(\frac{m_x}{k_x}, x)$ . Therefore, the AMSE can be estimated by

$$\widehat{AMSE}(m_x, k_x) = \frac{2(\theta_\pi + 1)}{2\theta_\pi + 1} \frac{\widehat{\gamma}_{LS}^2(x)}{k_x} + \left[ \frac{\theta_\pi + 1}{2(\theta_\pi + 2)} \widehat{\Delta}_{LS} \left( \frac{m_x}{k_x}, x \right) \right]^2 \quad (2.9)$$

from which we deduce our criterion of selection for  $(m_x, k_x)$  :

$$(\widehat{m}_x, \widehat{k}_x) = \underset{m_x, k_x}{\operatorname{argmin}} \widehat{AMSE}(m_x, k_x).$$

Note that the AMSE does not depend on the bandwidth parameter  $h_{2,n}$ . Therefore we need another criterion to find its optimal value.

### 2.4.2 Optimization of $h_{2,n}$ - Subsample division

The task is now to find an optimal value for the bandwidth  $h_{2,n}$  of the integrated kernel  $K(\cdot)$ . To this end, in the spirit of Ferraty and Vieu (2006), we propose to split our sample into two random subsamples  $I_1$  and  $I_2$  of the same sample size. Then, we fix the covariate on  $x_0$  (for instance taken in  $I_1$ ) and find the nearest neighbor  $x_0^*$  of  $x_0$  in  $I_2$ . Finally, we calculate the estimation  $\check{\gamma}_{\theta_\pi}^{(1)}(x_0)$  of  $\gamma(x_0)$  from  $I_1$  and  $\check{\gamma}_{\theta_\pi}^{(2)}(x_0^*)$  from  $I_2$ , using the optimal value for  $(m_x, k_x)$  as described in Section 2.4.1. The resulting optimization criterion for  $h_{2,n}$  can then be proposed :

$$\widehat{h}_{2,n} = \underset{h_{2,n}}{\operatorname{argmin}} |\check{\gamma}_{\theta_\pi}^{(1)}(x_0) - \check{\gamma}_{\theta_\pi}^{(2)}(x_0^*)|.$$

### 2.4.3 Graphs and computational issues

We simulate 100 replications of  $n = 5\,000$  pairs  $(Y_i, x_i), i = 1, \dots, n$  where  $x_i = i/n$  and the distribution of  $Y$  given  $x$  is the Burr-type distribution (2.6). The conditional quan-

tile estimator  $\hat{q}_n(\alpha, x)$  is obtained by dichotomous inversion of the conditional empirical survival function defined in (2.2).

For a fixed value of the covariate  $x$ , we perform the optimizations of  $(m_x, k_x)$  and  $h_{2,n}$  using grids :

- the number  $m_x$  of neighbors of  $x$  belongs to a grid from 2% to 10% of the sample size, i.e from  $m_x = 100$  to 500 in our case ; to avoid too small or too large values for  $k_x$ , this parameter has been taken in the interval  $(0.15 \times m_x; 0.85 \times m_x)$  ;
- $h_{2,n}$  goes from 0.05 to 1 by step of 0.05.

Finally, two values of  $\rho$  were used :  $\rho = -1$  and  $-0.8$ . Note that as usual in the absence of covariates, the closer  $\rho$  is to 0, the more difficult is the estimation of  $\gamma(\cdot)$ .

Figure 2.1 below illustrates the procedure presented in Section 2.4.1 for  $x = 0.4$ . In panel (a), we plot the estimated and theoretical AMSE for one realization (based on (2.8) and (2.9)). The boxplot of  $\hat{k}_x$  for the 100 replications is given in panel (b). As observed,  $\hat{k}_{0.4}$  is close to its optimal value  $k_{0.4}^{opt}$  (see Figure 2.1 a) with reasonable variability for such a parameter (see Figure 2.1 b).



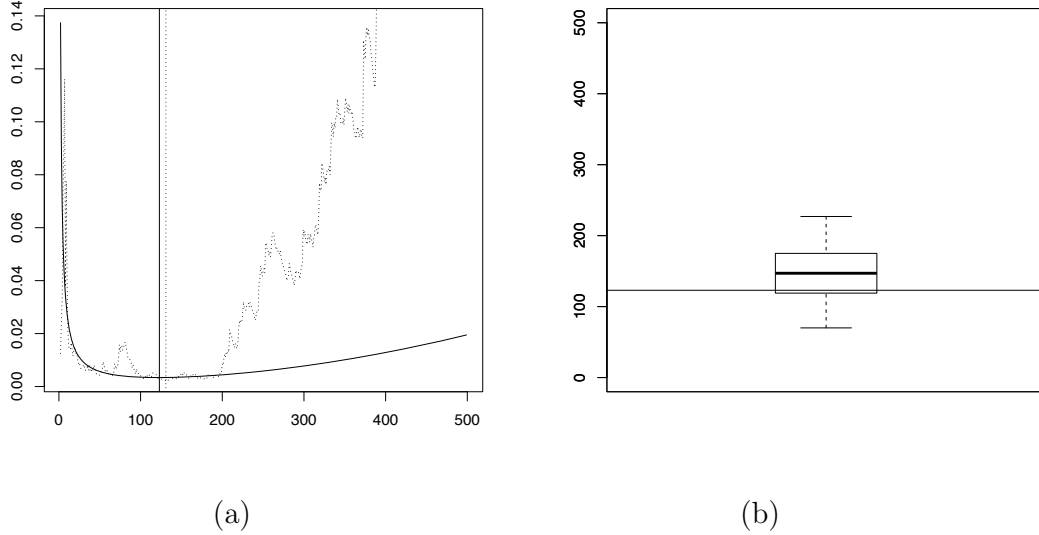


FIGURE 2.1 – (a) Theoretical (—) and estimated ( $\cdot\cdot\cdot$ ) AMSE for one realization of the Burr-type distribution (2.6) with  $\rho = -1$  and  $\gamma(0.4)$  defined in (2.7). The vertical lines correspond to the optimal values  $k_{0.4}^{opt}$  and  $\hat{k}_{0.4}$ ; (b) Boxplot of  $\hat{k}_{0.4}$  for the 100 replications. The horizontal line is the optimal value  $k_{0.4}^{opt}$ .

Now, we come back to our original problem, that is the estimation of  $\gamma(\cdot)$ . Again we simulate 100 replications from the Burr-type distribution (2.6). The squared of the  $L^2$ -errors between  $\hat{\gamma}(\cdot)$  and  $\gamma(\cdot)$  are computed. The median, 10% and 90% quantiles of these errors are represented in Figure 2.2. From these plots, we observe that our estimator performs quite well with a deterioration in the estimation of  $\gamma(\cdot)$  if  $|\rho|$  decreases.

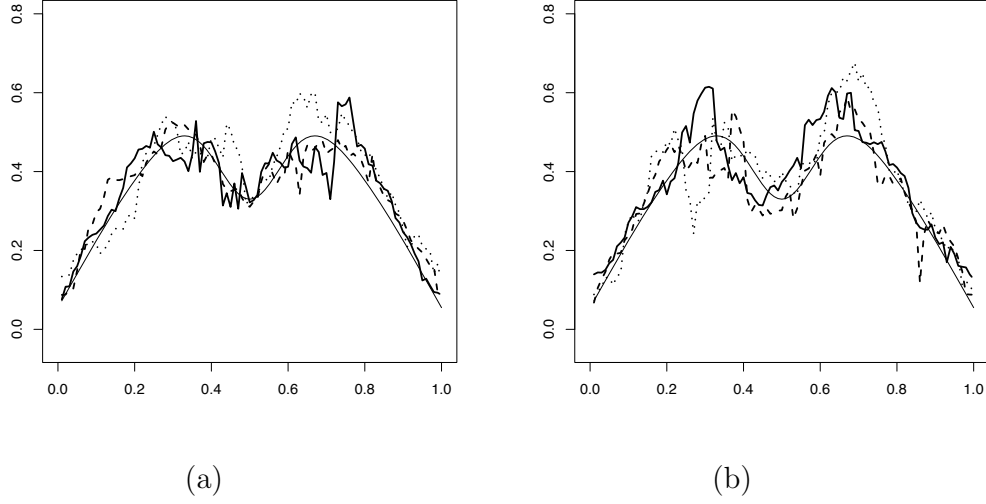


FIGURE 2.2 – Estimation of  $\gamma(\cdot)$  : Median (full line), 10% (dashed line) and 90% quantiles (dotted line) of the squared of the  $L^2$ -errors and true function (—) based on 100 replications from the Burr-type distribution (2.6) with  $\gamma(\cdot)$  defined in (2.7) and (a)  $\rho = -1$  or (b)  $\rho = -0.8$ .

## 2.5 Proofs

### 2.5.1 Preliminary results

The difference between the estimators  $\hat{q}_n(\cdot, x)$  and  $\bar{\mathbb{F}}_n^{\leftarrow}(\cdot, x)$  of the conditional quantile is controlled by the following lemma :

**Lemma 2.1.** *For all  $x \in E$ ,*

$$\sup_{\alpha \in [0,1]} |\hat{q}_n(\alpha, x) - \bar{\mathbb{F}}_n^{\leftarrow}(\alpha, x)| \leq h_{2,n}.$$

**Proof** – The empirical estimator of the survival function can be rewritten as :

$$\bar{\mathbb{F}}_n(y, x) = \begin{cases} \beta_0 = 0 & \text{if } y \geq Y_{m_x, m_x}^* \\ \beta_j = \frac{j}{m_x} & \text{if } y \in [Y_{m_x-j, m_x}^*, Y_{m_x-j+1, m_x}^*) \\ \beta_{m_x} = 1 & \text{if } y < Y_{1, m_x}^*. \end{cases}$$

It is then easy to check that :

$$\bar{\mathbb{F}}_n^{\leftarrow}(\alpha, x) = \sum_{i=0}^{m_x-1} Y_{m_x-i, m_x}^* \mathbf{1}(\alpha \in [\beta_i, \beta_{i+1})). \quad (2.10)$$

Let  $j \in \{0, \dots, m_x - 1\}$ . Since  $K(u) = 1$  for all  $u \leq -1$  and  $K(u) = 0$  for all  $u \geq 1$ , we have for all  $y \geq Y_{m_x-j, m_x}^* + h_{2,n}$  that

$$K\left(\frac{y - Y_{i, m_x}^*}{h_{2,n}}\right) = 0 \text{ for all } i \in \{1, \dots, m_x - j\}.$$

Furthermore, for all  $y \leq Y_{m_x-j, m_x}^* - h_{2,n}$ ,

$$K\left(\frac{y - Y_{i, m_x}^*}{h_{2,n}}\right) = 1 \text{ for all } i \in \{m_x - j, \dots, m_x\}.$$

Thus, for all  $y \geq Y_{m_x-j, m_x}^* + h_{2,n}$ ,  $\widehat{F}_n(y, x) \leq \beta_j$  and for all  $y \leq Y_{m_x-j, m_x}^* - h_{2,n}$ ,  $\widehat{F}_n(y, x) \geq \beta_{j+1}$ . Hence, for all  $\alpha \in [\beta_j, \beta_{j+1})$ ,

$$\widehat{F}_n^{\leftarrow}(\alpha, x) \in (Y_{m_x-j, m_x}^* - h_{2,n}, Y_{m_x-j, m_x}^* + h_{2,n}). \quad (2.11)$$

Equations (2.10) and (2.11) conclude the proof.

The following lemma establishes the convergence of Riemann sums. Its proof is straightforward and therefore omitted.

**Lemma 2.2.** *Let  $\varphi(\cdot)$  be an integrable function defined on  $(0, 1]$ . Then*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \varphi\left(\frac{i}{N}\right) = \int_0^1 \varphi(x) dx.$$

The next result provides an expansion of the conditional tail index estimator  $\widehat{\gamma}(k_x/m_x, x)$ .

**Lemma 2.3.** *Assume that conditions (A.1) and (A.2) hold. If  $k_x \rightarrow \infty$ ,  $m_x/k_x \rightarrow \infty$  and  $k_x^2 \omega_n(m_x^{-(1+\delta)}) \rightarrow 0$  for some  $\delta > 0$  then,*

$$\widehat{\gamma}(k_x/m_x, x) = \sum_{i=1}^{k_x} \left( \frac{1}{i} \int_0^{\beta_i} \Psi(\alpha, k_x/m_x, x) d\alpha \right) Z_{i, m_x}(x) + O_{\mathbb{P}}\left(\frac{h_{2,n}}{q(k_x/m_x, x)}\right).$$

**Proof** – From Lemma 2.1,

$$\sup_{\alpha \in [0, \beta_{k_x})} \left| \frac{\widehat{q}_n(\alpha, x)}{\overline{\mathbb{F}}_n^{\leftarrow}(\alpha, x)} - 1 \right| \leq \frac{h_{2,n}}{Y_{m_x - k_x, m_x}^*}, \quad (2.12)$$

since, for all  $\alpha \in [0, \beta_{k_x})$ ,  $\overline{\mathbb{F}}_n^{\leftarrow}(\alpha, x) \geq \overline{\mathbb{F}}_n^{\leftarrow}(\beta_{k_x}, x) = Y_{m_x - k_x, m_x}^*$ . From Gardes and Girard (2010, Lemma 1), there exists a sequence of events  $\mathcal{A}_n$  with  $\mathbb{P}(\mathcal{A}_n) \rightarrow 1$  as  $n \rightarrow \infty$  such that for some  $\delta > 0$ ,

$$\ln Y_{m_x - k_x, m_x}^* | \mathcal{A}_n \stackrel{d}{=} \ln q(U_{k_x+1, m_x}, x) + O_{\mathbb{P}}(\omega_n(m_x^{-(1+\delta)})) | \mathcal{A}_n, \quad (2.13)$$

where  $U_i$ ,  $i = 1, \dots, m_x$  are independent standard uniform variables and  $U_{1, m_x} \leq \dots \leq U_{m_x, m_x}$  are the associated order statistics. Hence, for all  $\varepsilon > 0$ ,

$$\mathbb{P} \left( \left| \ln Y_{m_x - k_x, m_x}^* - \ln q \left( \frac{k_x}{m_x}, x \right) \right| > \varepsilon \right) =: T_{1,n} + T_{2,n},$$

where

$$T_{1,n} := \mathbb{P} \left( \left| \ln Y_{m_x - k_x, m_x}^* - \ln q \left( \frac{k_x}{m_x}, x \right) \right| > \varepsilon \mid \mathcal{A}_n \right) \mathbb{P}(\mathcal{A}_n)$$

and

$$T_{2,n} := \mathbb{P} \left( \left| \ln Y_{m_x - k_x, m_x}^* - \ln q \left( \frac{k_x}{m_x}, x \right) \right| > \varepsilon \cap \mathcal{A}_n^C \right).$$

Clearly from (2.13),

$$T_{1,n} \stackrel{d}{=} \mathbb{P} \left( \left| \ln \frac{q(U_{k_x+1, m_x}, x)}{q(k_x/m_x, x)} + O_{\mathbb{P}}(\omega_n(m_x^{-(1+\delta)})) \right| > \varepsilon \mid \mathcal{A}_n \right) \mathbb{P}(\mathcal{A}_n).$$

Using the fact that for a positive sequence  $k_x$  such that  $k_x \rightarrow \infty$  and  $m_x/k_x \rightarrow \infty$ ,  $U_{k_x+1, m_x} \stackrel{\mathbb{P}}{\sim} k_x/m_x$  and since  $q(\cdot, x)$  is a regularly varying function we have  $T_{1,n} \rightarrow 0$  using the fact that  $\omega_n(m_x^{-(1+\delta)}) \rightarrow 0$ . Furthermore,  $T_{2,n} \leq \mathbb{P}(\mathcal{A}_n^C) \rightarrow 0$  since  $\mathbb{P}(\mathcal{A}_n) \rightarrow 1$ . Hence, for all  $\varepsilon > 0$ ,

$$\mathbb{P} \left( \left| \ln Y_{m_x - k_x, m_x}^* - \ln q \left( \frac{k_x}{m_x}, x \right) \right| > \varepsilon \right) \rightarrow 0,$$

and thus (2.12) implies that

$$\sup_{\alpha \in [0, \beta_{k_x}]} \left| \frac{\widehat{q}_n(\alpha, x)}{\overline{\mathbb{F}}_n^{\leftarrow}(\alpha, x)} - 1 \right| = O_{\mathbb{P}} \left( \frac{h_{2,n}}{q(k_x/m_x, x)} \right). \quad (2.14)$$

Now remark that

$$\widehat{\gamma}(k_x/m_x, x) = \int_0^{\beta_{k_x}} \Psi(\alpha, k_x/m_x, x) \ln \overline{\mathbb{F}}_n^{\leftarrow}(\alpha, x) d\alpha + \int_0^{\beta_{k_x}} \Psi(\alpha, k_x/m_x, x) \ln \frac{\widehat{q}_n(\alpha, x)}{\overline{\mathbb{F}}_n^{\leftarrow}(\alpha, x)} d\alpha.$$

From (2.14) and since  $\Psi(\cdot, u, x) \in L_1(0, 1)$ ,

$$\left| \int_0^{\beta_{k_x}} \Psi(\alpha, k_x/m_x, x) \ln \frac{\widehat{q}_n(\alpha, x)}{\overline{\mathbb{F}}_n^{\leftarrow}(\alpha, x)} d\alpha \right| = O_{\mathbb{P}} \left( \frac{h_{2,n}}{q(k_x/m_x, x)} \right). \quad (2.15)$$

Furthermore, from (2.10) and since  $\int_0^{\beta_{k_x}} \Psi(\alpha, k_x/m_x, x) d\alpha = 0$ , we have

$$\begin{aligned} \int_0^{\beta_{k_x}} \Psi(\alpha, k_x/m_x, x) \ln \overline{\mathbb{F}}_n^{\leftarrow}(\alpha, x) d\alpha &= \sum_{i=1}^{k_x} \left( \int_{\beta_{i-1}}^{\beta_i} \Psi(\alpha, k_x/m_x, x) d\alpha \right) \ln Y_{m_x-i+1, m_x}^* \\ &= \sum_{i=1}^{k_x} \left( \frac{1}{i} \int_0^{\beta_i} \Psi(\alpha, k_x/m_x, x) d\alpha \right) Z_{i, m_x}(x) \end{aligned}$$

which achieves the proof.

## 2.5.2 Proofs of the main results

**Proof of Theorem 2.1** – Let

$$\widetilde{\gamma}(k_x/m_x, x) = \sum_{i=1}^{k_x} \left( \frac{1}{i} \int_0^{\beta_i} \Psi(\alpha, k_x/m_x, x) d\alpha \right) Z_{i, m_x}(x).$$

First remark that by Lemma 2.3 and our assumptions we have

$$\begin{aligned} \sqrt{k_x} \left( \widehat{\gamma}(k_x/m_x, x) - \gamma(x) - \Delta \left( \frac{m_x}{k_x}, x \right) \mathcal{AB}(\Phi, \rho(x), x) \right) \\ &= \sqrt{k_x} \left( \widetilde{\gamma}(k_x/m_x, x) - \gamma(x) - \Delta \left( \frac{m_x}{k_x}, x \right) \mathcal{AB}(\Phi, \rho(x), x) \right) + O_{\mathbb{P}} \left( \frac{\sqrt{k_x} h_{2,n}}{q(k_x/m_x, x)} \right) \\ &= \sqrt{k_x} \left( \widetilde{\gamma}(k_x/m_x, x) - \gamma(x) - \Delta \left( \frac{m_x}{k_x}, x \right) \mathcal{AB}(\Phi, \rho(x), x) \right) + o_{\mathbb{P}}(1). \end{aligned}$$

Now, from Lemma 1 in Gardes and Girard (2010), there exists a sequence of events  $\mathcal{A}_n$  with  $\mathbb{P}(\mathcal{A}_n) \rightarrow 1$  as  $n \rightarrow \infty$  such that  $\{Z_{i,m_x}(x), i = 1, \dots, k_x | \mathcal{A}_n\}$  has the same distribution as

$$\left\{ i \ln \frac{q(U_{i,m_x}, x)}{q(U_{i+1,m_x}, x)} + i O_{\mathbb{P}}(\omega_n(m_x^{-(1+\delta)})), i = 1, \dots, k_x \middle| \mathcal{A}_n \right\}.$$

We thus have that  $\tilde{\gamma}(k_x/m_x, x) | \mathcal{A}_n$  has the same distribution as

$$\begin{aligned} \sum_{i=1}^{k_x} \left( \frac{1}{i} \int_0^{\beta_i} \Psi(\alpha, k_x/m_x, x) d\alpha \right) i \ln \frac{q(U_{i,m_x}, x)}{q(U_{i+1,m_x}, x)} \\ + O_{\mathbb{P}}(k_x \omega_n(m_x^{-(1+\delta)})) \sum_{i=1}^{k_x} \left( \frac{1}{i} \int_0^{\beta_i} \Psi(\alpha, k_x/m_x, x) d\alpha \right) \middle| \mathcal{A}_n. \end{aligned} \quad (2.16)$$

Let us introduce the following notation  $\Delta_x = \Delta(m_x/k_x, x)$ . From Gardes and Girard (2010, Lemma 4) and under our assumptions, to show that  $\sqrt{k_x}(\tilde{\gamma}(k_x/m_x, x) - \gamma(x) - \Delta_x \mathcal{AB}(\Phi, \rho(x), x))$  converges to a  $\mathcal{N}(0, \gamma^2(x) \mathcal{AV}(\Phi, x))$  random variable, it is sufficient to prove that  $\sqrt{k_x}(T_{1,n} - \gamma(x) - \Delta_x \mathcal{AB}(\Phi, \rho(x), x))$  converges to a  $\mathcal{N}(0, \gamma^2(x) \mathcal{AV}(\Phi, x))$  random variable, where

$$T_{1,n} = \sum_{i=1}^{k_x} \left( \frac{1}{i} \int_0^{\beta_i} \Psi(\alpha, k_x/m_x, x) d\alpha \right) i \ln \frac{q(U_{i,m_x}, x)}{q(U_{i+1,m_x}, x)}.$$

From Beirlant *et al.* (2002, Theorem 2.1),  $\{i(\ln q(U_{i,m_x}, x) - \ln q(U_{i+1,m_x}, x)), i = 1, \dots, k_x\}$  has the same distribution as

$$\left\{ \left( \gamma(x) + \Delta_x \left( \frac{i}{k_x + 1} \right)^{-\rho(x)} \right) F_i + \nu_{i,n}(x) + o_{\mathbb{P}}(\Delta_x), i = 1, \dots, k_x \right\}, \quad (2.17)$$

where  $F_1, \dots, F_{k_x}$  are independent standard exponential random variables and, uniformly in  $i = 1, \dots, k_x$ ,

$$\sum_{j=i}^{k_x} \frac{|\nu_{j,n}(x)|}{j} = o_{\mathbb{P}} \left( \Delta_x \max \left( \ln \frac{k_x + 1}{i}, 1 \right) \right). \quad (2.18)$$

Thus by assumptions and Lemma 2.2,  $\sqrt{k_x}(T_{1,n} - \gamma(x) - \Delta_x \mathcal{AB}(\Phi, \rho(x), x))$  has the same

distribution as

$$\sqrt{k_x} \left( \gamma(x)T_{1,1,n} + \Delta_x T_{1,2,n} + T_{1,3,n} + \sum_{i=1}^{k_x} \left( \frac{1}{i} \int_0^{\beta_i} \Psi(\alpha, k_x/m_x, x) d\alpha \right) o_{\mathbb{P}}(\Delta_x) \right),$$

with

$$\begin{aligned} T_{1,1,n} &= \sum_{i=1}^{k_x} \left( \frac{1}{i} \int_0^{\beta_i} \Psi(\alpha, k_x/m_x, x) d\alpha \right) (F_i - 1) \\ T_{1,2,n} &= \sum_{i=1}^{k_x} \left( \frac{1}{i} \int_0^{\beta_i} \Psi(\alpha, k_x/m_x, x) d\alpha \right) \left( \frac{i}{k_x + 1} \right)^{-\rho(x)} (F_i - 1) \\ T_{1,3,n} &= \sum_{i=1}^{k_x} \left( \frac{1}{i} \int_0^{\beta_i} \Psi(\alpha, k_x/m_x, x) d\alpha \right) \nu_{i,n}(x). \end{aligned}$$

Let us first consider the term  $T_{1,1,n}$ . From Lyapunov's Theorem, if for some  $\xi > 0$ ,

$$\left( \sum_{i=1}^{k_x} \left( \frac{1}{i} \int_0^{\beta_i} \Psi(\alpha, k_x/m_x, x) d\alpha \right)^2 \right)^{-(1+\xi/2)} \sum_{i=1}^{k_x} \left| \frac{1}{i} \int_0^{\beta_i} \Psi(\alpha, k_x/m_x, x) d\alpha \right|^{2+\xi} \rightarrow 0, \quad (2.19)$$

then,

$$\left( \sum_{i=1}^{k_x} \left( \frac{1}{i} \int_0^{\beta_i} \Psi(\alpha, k_x/m_x, x) d\alpha \right)^2 \right)^{-1/2} T_{1,1,n} \xrightarrow{d} \mathcal{N}(0, 1). \quad (2.20)$$

From condition **(B.1)**,

$$\begin{aligned} &\left( \sum_{i=1}^{k_x} \left( \frac{1}{i} \int_0^{\beta_i} \Psi(\alpha, k_x/m_x, x) d\alpha \right)^2 \right)^{-(1+\xi/2)} \sum_{i=1}^{k_x} \left| \frac{1}{i} \int_0^{\beta_i} \Psi(\alpha, k_x/m_x, x) d\alpha \right|^{2+\xi} \\ &= k_x^{-\xi/2} \left( \frac{1}{k_x} \sum_{i=1}^{k_x} \Phi^2 \left( \frac{i}{k_x}, x \right) \right)^{-(1+\xi/2)} \left( \frac{1}{k_x} \sum_{i=1}^{k_x} \Phi^{2+\xi} \left( \frac{i}{k_x}, x \right) \right). \end{aligned}$$

Using Lemma 2.2 and condition **(B.1)**, it is then easy to check that condition (2.19) is satisfied. Thus (2.20) holds or equivalently

$$\sqrt{k_x} T_{1,1,n} \xrightarrow{d} \mathcal{N}(0, \mathcal{AV}(\Phi, x)).$$

Now, let us focus on the term  $T_{1,2,n}$ . We have  $\mathbb{E}(T_{1,2,n}) = 0$  and

$$\text{Var}(T_{1,2,n}) = \frac{1}{k_x^2} \sum_{i=1}^{k_x} \left( \frac{k_x}{i} \int_0^{\beta_i} \Psi(\alpha, k_x/m_x, x) d\alpha \right)^2 \left( \frac{i}{k_x + 1} \right)^{-2\rho(x)} = O(k_x^{-1}),$$

from Lemma 2.2 and condition **(B.1)**. Thus,

$$T_{1,2,n} = o_{\mathbb{P}}(1). \quad (2.21)$$

Furthermore,

$$\begin{aligned} T_{1,3,n} &= \sum_{i=1}^{k_x} \left( \frac{1}{i} \int_0^{\beta_i} \Psi(\alpha, k_x/m_x, x) d\alpha \right) \nu_{i,n}(x) \\ &= \sum_{i=1}^{k_x} \left( \sum_{j=1}^i \int_{\beta_{j-1}}^{\beta_j} \Psi(\alpha, k_x/m_x, x) d\alpha \right) \frac{\nu_{i,n}(x)}{i} \\ &= \sum_{j=1}^{k_x} \int_{\beta_{j-1}}^{\beta_j} \Psi(\alpha, k_x/m_x, x) d\alpha \sum_{i=j}^{k_x} \frac{\nu_{i,n}(x)}{i} \\ &= \sum_{j=1}^{k_x} \left[ \int_0^{\beta_j} \Psi(\alpha, k_x/m_x, x) d\alpha - \int_0^{\beta_{j-1}} \Psi(\alpha, k_x/m_x, x) d\alpha \right] \sum_{i=j}^{k_x} \frac{\nu_{i,n}(x)}{i} \\ &= \frac{1}{k_x} \sum_{j=1}^{k_x} \left[ j\Phi\left(\frac{j}{k_x}, x\right) - (j-1)\Phi\left(\frac{j-1}{k_x}, x\right) \right] \sum_{i=j}^{k_x} \frac{\nu_{i,n}(x)}{i}. \end{aligned}$$

Thus, from **(B.2)**, (2.18) and using Lemma 2.2

$$\begin{aligned} |T_{1,3,n}| &= \Delta_x o_{\mathbb{P}} \left( \frac{1}{k_x} \sum_{j=1}^{k_x} h\left(\frac{j}{k_x}, x\right) \max\left(\ln \frac{k_x + 1}{j}, 1\right) \right) \\ &= o_{\mathbb{P}}(\Delta_x). \end{aligned} \quad (2.22)$$

The proof of Theorem 2.1 is then achieved.

**Proof of Corollary 2.1** – It is sufficient to prove that the function  $\Psi_{\theta}(\cdot, u, x)$  satisfies conditions **(B.1)** and **(B.2)**. Clearly, **(B.1)** is satisfied with

$$\Phi_{\theta}(\alpha, x) = \begin{cases} \frac{\theta + 1}{\theta} (1 - \alpha^{\theta}) & \text{if } \theta > 0 \\ -\ln(\alpha) & \text{if } \theta = 0 \\ 1 & \text{if } \theta = \infty, \end{cases}$$



for  $\alpha < 1$ . Note that  $\Phi_0(\cdot, \cdot)$  and  $\Phi_\infty(\cdot, \cdot)$  correspond to the limiting cases of  $\Phi_\theta(\cdot, \cdot)$ .

Concerning the condition **(B.2)**, the mean value theorem leads to :

$$\begin{aligned} \left| i\Phi_\theta\left(\frac{i}{k_x}, x\right) - (i-1)\Phi_\theta\left(\frac{i-1}{k_x}, x\right) \right| &\leq \sup_{\alpha \in I_i} \left| \frac{\theta+1}{\theta}(1-\alpha^\theta) - \alpha^\theta(1+\theta) \right| \\ &\leq \sup_{\alpha \in I_i} \frac{\theta+1}{\theta}(1+\alpha^\theta(\theta-1)), \end{aligned}$$

where  $I_i = [(i-1)/k_x, i/k_x]$ . Clearly, if  $\theta \geq 1$ , for all  $i = 1, \dots, k_x$ ,

$$\left| i\Phi_\theta\left(\frac{i}{k_x}, x\right) - (i-1)\Phi_\theta\left(\frac{i-1}{k_x}, x\right) \right| \leq \frac{\theta+1}{\theta} \left( 1 + \left(\frac{i}{k_x}\right)^\theta (\theta-1) \right). \quad (2.23)$$

If  $\theta < 1$ , then for  $i = 1$ ,

$$\left| i\Phi_\theta\left(\frac{i}{k_x}, x\right) - (i-1)\Phi_\theta\left(\frac{i-1}{k_x}, x\right) \right| = \Phi_\theta\left(\frac{1}{k_x}, x\right), \quad (2.24)$$

and for  $i = 2, \dots, k_x$ ,

$$\begin{aligned} \left| i\Phi_\theta\left(\frac{i}{k_x}, x\right) - (i-1)\Phi_\theta\left(\frac{i-1}{k_x}, x\right) \right| &\leq \frac{\theta+1}{\theta} \left( 1 + \left(\frac{i-1}{k_x}\right)^\theta (\theta-1) \right) \\ &\leq \frac{\theta+1}{\theta} \left( 1 + \left(\frac{i}{2k_x}\right)^\theta (\theta-1) \right). \end{aligned} \quad (2.25)$$

Collecting (2.23)-(2.25), leads to

$$\left| i\Phi_\theta\left(\frac{i}{k_x}, x\right) - (i-1)\Phi_\theta\left(\frac{i-1}{k_x}, x\right) \right| \leq h\left(\frac{i}{k_x}, x\right),$$

where

$$h(\alpha, x) = \begin{cases} (\theta+1)(1+(\theta-1)\alpha^\theta)/\theta & \text{if } \theta \geq 1, \\ (\theta+1)(1+(\theta-1)(\alpha/2)^\theta)/\theta & \text{if } \theta < 1. \end{cases}$$

Since  $h(\cdot, x)$  satisfies the condition in **(B.2)**, the proof is complete.

# Chapitre 3

## Nonparametric regression estimation of conditional tails - the random covariate case

### Abstract

We present families of nonparametric estimators for the conditional tail index of a Pareto-type distribution in presence of random covariates. These families are constructed from locally weighted sums of power transformations of excesses over a high threshold. The asymptotic properties of the proposed estimators are derived under some assumptions on the conditional response distribution, the weight function and the density function of the covariates. We also introduce bias-corrected versions of the estimators for the conditional tail index, and propose in this context a consistent estimator for the second order tail parameter. The finite sample performance of some specific examples from our classes of estimators is illustrated with a small simulation experiment.

### 3.1 Introduction

In the area of extreme value statistics, the tail index of a distribution function assumes a central position, and therefore a vast literature has been dedicated to the estimation of this parameter. We refer to Beirlant *et al.* (2004), and de Haan and Ferreira (2006) for recent accounts on univariate tail index estimation. Obtaining an estimate for the tail index is in general a first step in an extreme value analysis. For instance, when interest is in the estimation of extreme quantiles, i.e. quantiles of order  $\alpha_n$ , with  $\alpha_n > 1 - 1/n$ , where  $n$  denotes the sample size, one has to extrapolate beyond the available data by means of an extreme value model depending on the tail index. In this chapter we will consider tail index estimation for heavy tailed models when a random covariate  $X$  is recorded simultaneously with the variable of interest  $Y$ .

Let  $F(y; x)$  denote the conditional distribution function of the response variable  $Y$  given  $X = x$ , and let  $g$  denote the density function of  $X$ . We assume that the conditional response distribution is of Pareto-type, i.e. there exists a positive function  $\gamma(x)$  such that  $\bar{F}(y; x) := 1 - F(y; x)$  is of the form

$$\bar{F}(y; x) = y^{-1/\gamma(x)}\ell(y; x), \quad y > 0, \quad (3.1)$$

where  $\ell$  is a slowly varying function at infinity, so

$$\lim_{y \rightarrow \infty} \frac{\ell(\lambda y; x)}{\ell(y; x)} = 1, \quad \text{for all } \lambda > 0. \quad (3.2)$$

The tail function  $\gamma(x)$  describes the tail heaviness of the conditional response distribution, and is a function that needs to be adequately estimated from the available data.

The estimation of the tail index in a regression context with fixed, nonrandom covariates, has been addressed in the recent extreme value literature, though not to the same

extent as univariate tail index estimation. Parametric models based on the generalized extreme value distribution or the generalized Pareto distribution, extended to regression models by assuming specific functional forms for each of their parameters, were proposed in Smith (1989), and Davison and Smith (1990). Beirlant and Goegebeur (2003) introduced a semi-parametric model by extending the representation for scaled-log spacings of successive order statistics of Beirlant *et al.* (1999) to the regression context. Fully nonparametric methods like local estimation and penalized likelihood estimation can be found in e.g. Davison and Ramesh (2000), Hall and Tajvidi (2000), Pauli and Coles (2001), Beirlant and Goegebeur (2004), Chavez-Demoulin and Davison (2005), Gardes and Girard (2008, 2010), Goegebeur and de Wet (2012). Despite these contributions for the fixed covariate case, conditional tail index estimation with random covariates received only little attention. Wang and Tsai (2009) considered maximum likelihood estimation within the Hall subclass of the Pareto-type models (Hall, 1982) where the tail index is related to the covariates through a log link function. Daouia *et al.* (2011) studied the estimation of extreme quantiles under a conditional Pareto-type model with random covariates, and used a fixed number of such quantile estimates in classical estimators for  $\gamma$ , like the Pickands (Pickands, 1975) and the Hill (Hill, 1975) estimators. The aim of this chapter is to introduce flexible classes of estimators for  $\gamma(x)$  in the random covariate case, by considering weighted sums of power transformations of excesses over a high threshold.

The remainder of this chapter is organized as follows. In Section 3.2 we introduce our main kernel statistic and derive its asymptotic properties under some conditions. Based on these results we introduce two classes of estimators for the conditional tail index  $\gamma(x)$  and establish their limiting distributions. Bias-correction of the estimators for  $\gamma(x)$  is studied in Section 3.3, where we also introduce a consistent estimator for the second order rate parameter. The finite sample behavior of our estimators is illustrated in Section 3.4 by means of a small simulation experiment. The proofs of all results are deferred to

Section 3.5.

## 3.2 Construction of the estimator and asymptotic properties

Let  $(X_i, Y_i)$ ,  $i = 1, \dots, n$ , be independent copies of the random vector  $(X, Y) \in \mathbb{R}^p \times \mathbb{R}_0^+$ , where  $X$  has a distribution with joint density function  $g$ , and the conditional survival function of  $Y$  given  $X = x$  is of the form (3.1).

The idea is to construct estimators for the conditional tail index  $\gamma(x)$  on the basis of statistics of the form

$$T_n^{(s,t)}(x, K) := \frac{1}{n} \sum_{i=1}^n K_h^s(x - X_i) (\ln Y_i - \ln \omega_n)_+^t \mathbf{1}\{Y_i > \omega_n\}, \quad (3.3)$$

where  $s \geq 1$ ,  $t \geq 0$ ,  $K_h(x) := K(x/h)/h^p$ ,  $K$  is a joint density function on  $\mathbb{R}^p$ ,  $h = h_n$  is a non-random sequence with  $h_n \rightarrow 0$  for  $n \rightarrow \infty$ ,  $\mathbf{1}\{A\}$  is the indicator function on the event  $A$ , and  $\omega_n$  is a local non-random threshold sequence for estimation with  $\omega_n \rightarrow \infty$  for  $n \rightarrow \infty$ . Note that for  $t = 0$  the statistic simplifies to

$$T_n^{(s,0)}(x, K) := \frac{1}{n} \sum_{i=1}^n K_h^s(x - X_i) \mathbf{1}\{Y_i > \omega_n\},$$

which is proportional to the kernel estimator for  $\bar{F}(y; x)$ , proposed in Daouia *et al.* (2011).

In order to obtain the limiting distribution of  $T_n^{(s,t)}(x, K)$  we have to impose a further condition on the behavior of the tail of  $\bar{F}(y; x)$ , the second order condition.

**Second order condition ( $\mathcal{R}_2$ ) :** *There exists a positive function  $\gamma(x)$ , a function  $\rho(x) < 0$  and a rate function  $b(\cdot; x)$  with  $b(y; x) \rightarrow 0$  for  $y \rightarrow \infty$ , of constant sign for large values*

of  $y$ , such that for all  $v > 0$

$$\lim_{y \rightarrow \infty} \frac{\frac{\bar{F}(vy;x) - v^{-1/\gamma(x)}}{\bar{F}(y;x)}}{b(y;x)} = v^{-1/\gamma(x)} \frac{v^{\rho(x)/\gamma(x)} - 1}{\gamma(x)\rho(x)}. \quad (3.4)$$

This condition is not too restrictive and is commonly accepted within extreme value statistics. In fact, after some straightforward simplification of (3.4), one easily sees that condition  $(\mathcal{R}_2)$  makes the convergence in (3.2) explicit. We refer to Bingham *et al.* (1987) and de Haan and Ferreira (2006) for further details.

We also need some Lipschitz conditions. In these, for all  $x, z \in \mathbb{R}^p$ , the Euclidean distance between  $x$  and  $z$  is denoted by  $d(x, z)$ .

$(\mathcal{G})$  There exists  $c_g > 0$  such that  $|g(x) - g(z)| \leq c_g d(x, z)$ .

$(\mathcal{F})$  There exists  $c_{\bar{F}} > 0$  and  $y_0 > 1$  such that

$$\sup_{y \geq y_0} \left| \frac{\ln \bar{F}(y; x)}{\ln \bar{F}(y; z)} - 1 \right| \leq c_{\bar{F}} d(x, z).$$

Note that  $(\mathcal{F})$  controls the oscillation of  $\ln \bar{F}(y; x)$  with respect to its second argument  $x$ .

Finally, we impose a condition on the kernel function  $K$ .

$(\mathcal{K})$   $K$  is a bounded density function on  $\mathbb{R}^p$ , with support  $\Omega$  included in the unit hypersphere of  $\mathbb{R}^p$ .

We start by introducing two lemmas that will enable us to obtain the asymptotic expansion of  $\mathbb{E}[T_n^{(s,t)}(x, K)]$  under the above conditions. Let

$$m(\omega_n, t; x) := \mathbb{E}[(\ln Y - \ln \omega_n)_+^t \mathbf{1}\{Y > \omega_n\} | X = x].$$

**Lemma 3.1.** Assume  $(\mathcal{R}_2)$ , then for  $\omega_n \rightarrow \infty$  we have that

$$m(\omega_n, t; x) = \begin{cases} \gamma^t(x)\Gamma(t+1)\bar{F}(\omega_n; x) \left\{ 1 + \frac{b(\omega_n; x)}{\gamma(x)\rho(x)} \left[ \frac{1}{(1-\rho(x))^t} - 1 \right] (1 + o(1)) \right\}, & t > 0 \\ \bar{F}(\omega_n; x), & t = 0. \end{cases}$$

Now consider

$$\tilde{m}_n(x; s, t) := \mathbb{E}[K_h^s(x - X)(\ln Y - \ln \omega_n)_+^t \mathbf{1}\{Y > \omega_n\}].$$

**Lemma 3.2.** Assume  $(\mathcal{R}_2)$ ,  $(\mathcal{G})$ ,  $(\mathcal{F})$  and  $(\mathcal{K})$ . Then, for all  $x \in \mathbb{R}^p$  such that  $g(x) > 0$  we have for  $\omega_n \rightarrow \infty$  and  $h \rightarrow 0$  with  $h \ln \omega_n \rightarrow 0$  that

$$\tilde{m}_n(x; s, t) = \frac{m(\omega_n, t; x)}{h^{(s-1)p}} g(x) \|K\|_s^s (1 + O(h \ln \omega_n)).$$

Note that

$$\begin{aligned} \mathbb{E}[T_n^{(s,t)}(x, K)] &= \tilde{m}_n(x; s, t) \\ &= \mathbb{E}[K_h^s(x - X)m(\omega_n, t; X)]. \end{aligned}$$

From Lemma 3.1 and 3.2 we have that for  $t > 0$ ,

$$\begin{aligned} \mathbb{E}(T_n^{(1,t)}(x, K)) &= \gamma^t(x)g(x)\Gamma(t+1)\bar{F}(\omega_n; x) \\ &\quad \times \left\{ 1 + \frac{b(\omega_n; x)}{\gamma(x)\rho(x)} \left[ \frac{1}{(1-\rho(x))^t} - 1 \right] (1 + o(1)) + O(h \ln \omega_n) \right\} \end{aligned}$$

and

$$\mathbb{E}(T_n^{(1,0)}(x, L)) = g(x)\bar{F}(\omega_n; x)(1 + O(h \ln \omega_n)),$$

where  $L$  is assumed to satisfy  $(\mathcal{K})$ , which motivates the following estimators for  $\gamma(x)$

$$\begin{aligned}\widehat{\gamma}_n^{(1)}(x, t, K, L) &:= \left( \frac{T_n^{(1,t)}(x, K)}{\Gamma(t+1)T_n^{(1,0)}(x, L)} \right)^{1/t}, \\ \widehat{\gamma}_n^{(2)}(x, t, K, L) &:= \frac{T_n^{(1,t+1)}(x, K)}{(t+1)T_n^{(1,t)}(x, L)}.\end{aligned}$$

Note that  $\widehat{\gamma}_n^{(1)}(x, 1, K, L) = \widehat{\gamma}_n^{(2)}(x, 0, K, L)$ .

Now we study the asymptotic behavior of a vector of statistics of the form (3.3), when appropriately normalized. Denote

$$\mathbb{T}'_n := \frac{1}{\overline{F}(\omega_n; x)g(x)} [T_n^{(1,t_1)}(x, K_1), \dots, T_n^{(1,t_J)}(x, K_J)],$$

and let  $\Sigma$  be a  $(J \times J)$  matrix with elements

$$\sigma_{j,k} := \gamma^{t_j+t_k}(x) \|K_j K_k\|_1 \Gamma(t_j + t_k + 1).$$

**Theorem 3.1.** *Let  $(X_1, Y_1), \dots, (X_n, Y_n)$  be a sample of iid random vectors, and assume  $(\mathcal{R}_2)$ ,  $(\mathcal{G})$ ,  $(\mathcal{F})$  and that  $K_1, \dots, K_J$  satisfy  $(\mathcal{K})$ . For all  $x \in \mathbb{R}^p$  where  $g(x) > 0$ , we have that if  $h \rightarrow 0$ ,  $\omega_n \rightarrow \infty$  for  $n \rightarrow \infty$ , with  $h \ln \omega_n \rightarrow 0$  and  $nh^p \overline{F}(\omega_n; x) \rightarrow \infty$ , then*

$$\sqrt{nh^p \overline{F}(\omega_n; x)g(x)} [\mathbb{T}'_n - \mathbb{E}(\mathbb{T}'_n)] \xrightarrow{\mathcal{D}} N_J(0, \Sigma).$$

We now derive the limiting distribution of the estimator  $\widehat{\gamma}_n^{(1)}(x, t, K, L)$  and  $\widehat{\gamma}_n^{(2)}(x, t, K, L)$ .

**Theorem 3.2.** *Let  $(X_1, Y_1), \dots, (X_n, Y_n)$  be a sample of iid random vectors, and assume  $(\mathcal{R}_2)$ ,  $(\mathcal{G})$ ,  $(\mathcal{F})$  and kernel functions  $K$  and  $L$  satisfying  $(\mathcal{K})$ . For all  $x \in \mathbb{R}^p$  where  $g(x) > 0$ , we have that if  $h \rightarrow 0$ ,  $\omega_n \rightarrow \infty$  for  $n \rightarrow \infty$ , with  $nh^p \overline{F}(\omega_n; x) \rightarrow \infty$ ,*



$\sqrt{nh^p \bar{F}(\omega_n; x) b(\omega_n; x)} \rightarrow \lambda$  and  $\sqrt{nh^p \bar{F}(\omega_n; x) h \ln \omega_n} \rightarrow 0$ , then

$$\sqrt{nh^p \bar{F}(\omega_n; x) g(x)} [\hat{\gamma}_n^{(1)}(x, t, K, L) - \gamma(x)] \xrightarrow{\mathcal{D}} N \left( \lambda \frac{\sqrt{g(x)}}{t\rho(x)} \left( \frac{1}{(1-\rho(x))^t} - 1 \right), \frac{\gamma^2(x)}{t^2 \Gamma^2(t+1)} [\Gamma(2t+1) \|K\|_2^2 + \Gamma^2(t+1) \|L\|_2^2 - 2\Gamma^2(t+1) \|KL\|_1] \right).$$

**Theorem 3.3.** Under the same assumptions as in Theorem 3.2, we have

$$\sqrt{nh^p \bar{F}(\omega_n; x) g(x)} [\hat{\gamma}_n^{(2)}(x, t, K, L) - \gamma(x)] \xrightarrow{\mathcal{D}} N \left( \lambda \frac{\sqrt{g(x)}}{(1-\rho(x))^{t+1}}, \frac{\gamma^2(x) \Gamma(2t+1)}{(t+1) \Gamma^2(t+1)} [2(2t+1) \|K\|_2^2 + (t+1) \|L\|_2^2 - 2(2t+1) \|KL\|_1] \right).$$

For  $K = L$  the asymptotic variance expressions simplify and are given by

$$\mathbb{A}\text{Var}(\hat{\gamma}_n^{(1)}(x, t, K, K)) = \frac{\gamma^2(x) \|K\|_2^2}{t^2} \left( \frac{\Gamma(2t+1)}{\Gamma^2(t+1)} - 1 \right) =: \gamma^2(x) \|K\|_2^2 v_1^2,$$

and

$$\mathbb{A}\text{Var}(\hat{\gamma}_n^{(2)}(x, t, K, K)) = \gamma^2(x) \|K\|_2^2 \frac{\Gamma(2t+1)}{\Gamma^2(t+1)} =: \gamma^2(x) \|K\|_2^2 v_2^2.$$

In Figure 3.1 (a) we show the components of the asymptotic standard deviation  $v_1$  (solid line) and  $v_2$  (dashed line) as a function of  $t$ . Clearly, no estimator performs uniformly best in  $t$ , though the estimator  $\hat{\gamma}_n^{(1)}(x, t, K, K)$  shows the better performance over a wide range of values for  $t$ . Note that the asymptotic variance of the estimator  $\hat{\gamma}_n^{(1)}(x, t, K, K)$  is minimized for  $t = 1$  whereas for  $\hat{\gamma}_n^{(2)}(x, t, K, K)$  the minimum variance is reached for  $t = 0$ . For what concerns the asymptotic bias, it can be proven that the estimator  $\hat{\gamma}_n^{(2)}(x, t, K, K)$  performs uniformly better than  $\hat{\gamma}_n^{(1)}(x, t, K, K)$ . This is illustrated in Figure 3.1 (b) where we show the components of the asymptotic bias  $((1-\rho(x))^{-t} - 1)/(t\rho(x))$  (solid line) and  $(1-\rho(x))^{-(t+1)}$  (dashed line) for the case where  $\rho(x) = -1$ , as a function of  $t$ .

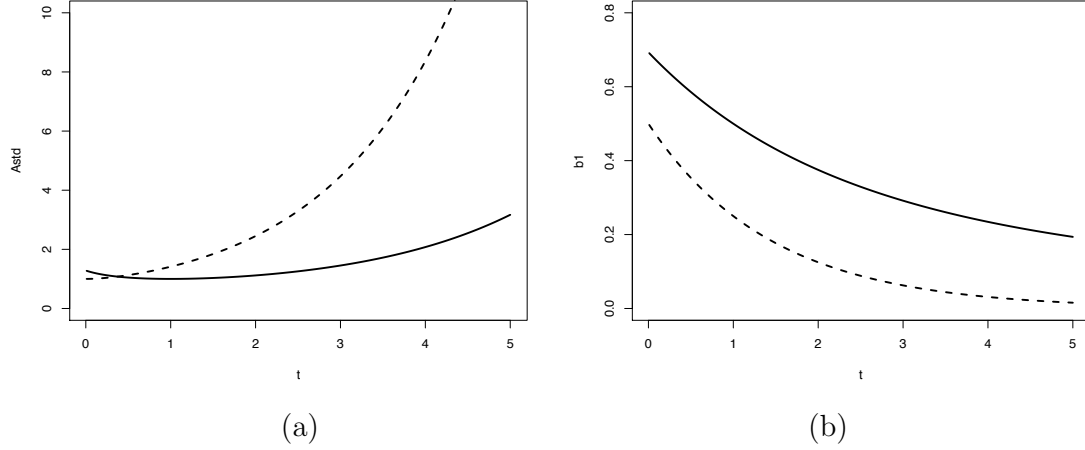


FIGURE 3.1 – (a) Components of the asymptotic standard deviation  $v_1$  (solid line) and  $v_2$  (dashed line) as a function of  $t$ , (b) components of the asymptotic bias  $((1 - \rho(x))^{-t} - 1)/(t\rho(x))$  (solid line) and  $(1 - \rho(x))^{-(t+1)}$  (dashed line), for  $\rho(x) = -1$ , as a function of  $t$ .

As mentioned before,  $\hat{\gamma}_n^{(1)}(x, 1, K, L) = \hat{\gamma}_n^{(2)}(x, 0, K, L)$ . In this case, setting  $K = L = \mathbf{1}\{S\}/|S|$ , where  $S$  is the  $p$  dimensional unit hypersphere with volume  $|S| = 2\pi^{p/2}/(p\Gamma(p/2))$ , one obtains a local Hill-type estimator, denoted  $\hat{\gamma}_{H,n}(x)$ , which distribution is given in the following corollary.

**Corollary 3.1.** *Let  $(X_1, Y_1), \dots, (X_n, Y_n)$  be a sample of iid random vectors, and assume  $(\mathcal{R}_2)$ ,  $(\mathcal{G})$  and  $(\mathcal{F})$ . For all  $x \in \mathbb{R}^p$  where  $g(x) > 0$ , we have that if  $h \rightarrow 0$ ,  $\omega_n \rightarrow \infty$  for  $n \rightarrow \infty$ , with  $nh^p\bar{F}(\omega_n; x) \rightarrow \infty$ ,  $\sqrt{nh^p\bar{F}(\omega_n; x)b(\omega_n; x)} \rightarrow \lambda$  and  $\sqrt{nh^p\bar{F}(\omega_n; x)h \ln \omega_n} \rightarrow 0$ , then*

$$\sqrt{nh^p\bar{F}(\omega_n; x)g(x)} [\hat{\gamma}_{H,n}(x) - \gamma(x)] \xrightarrow{\mathcal{D}} N\left(\lambda \frac{\sqrt{g(x)}}{1 - \rho(x)}, \frac{\gamma^2(x)}{|S|}\right).$$

### 3.3 Bias-corrected estimation of $\gamma(x)$

In this section we will introduce bias-corrected estimators for  $\gamma(x)$  based on  $\widehat{\gamma}_n^{(1)}(x, t, K, L)$  and  $\widehat{\gamma}_n^{(2)}(x, t, K, L)$ , as well as a consistent estimator for the second order parameter  $\rho(x)$ .

A first possibility to obtain a bias-corrected estimator for  $\gamma$  consists in constructing an appropriately chosen weighted sum of two estimators  $\widehat{\gamma}_n^{(1)}(x, t_1, K_1, L_1)$  and  $\widehat{\gamma}_n^{(1)}(x, t_2, K_2, L_2)$ , and therefore we start by establishing the joint asymptotic behavior of the latter.

Let  $\boldsymbol{\mu}^{(1)}$  be a  $(2 \times 1)$  vector with elements

$$\mu_i^{(1)} := \frac{\sqrt{g(x)}}{t_i \rho(x)} \left( \frac{1}{(1 - \rho(x))^{t_i}} - 1 \right), \quad i = 1, 2,$$

and  $\widetilde{\boldsymbol{\Sigma}}^{(1)}$  a  $(2 \times 2)$  matrix with elements

$$\begin{aligned} \widetilde{\sigma}_{i,i}^{(1)} &:= \frac{\gamma^2(x)}{[t_i \Gamma(t_i + 1)]^2} [\Gamma(2t_i + 1) \|K_i\|_2^2 + \Gamma^2(t_i + 1) \|L_i\|_2^2 - 2\Gamma^2(t_i + 1) \|K_i L_i\|_1], \quad i = 1, 2, \\ \widetilde{\sigma}_{1,2}^{(1)} &:= \frac{\gamma^2(x)}{t_1 t_2 \Gamma(t_1 + 1) \Gamma(t_2 + 1)} [\Gamma(t_1 + t_2 + 1) \|K_1 K_2\|_1 + \Gamma(t_1 + 1) \Gamma(t_2 + 1) \|L_1 L_2\|_1 \\ &\quad - \Gamma(t_1 + 1) \Gamma(t_2 + 1) \|K_1 L_2\|_1 - \Gamma(t_1 + 1) \Gamma(t_2 + 1) \|K_2 L_1\|_1]. \end{aligned}$$

**Theorem 3.4.** *Let  $(X_1, Y_1), \dots, (X_n, Y_n)$  be a sample of iid random vectors, and assume  $(\mathcal{R}_2)$ ,  $(\mathcal{G})$ ,  $(\mathcal{F})$  and kernel functions  $K_1, K_2, L_1$  and  $L_2$  satisfying  $(\mathcal{K})$ . For all  $x \in \mathbb{R}^p$  where  $g(x) > 0$ , we have that if  $h \rightarrow 0$ ,  $\omega_n \rightarrow \infty$  for  $n \rightarrow \infty$ , with  $nh^p \overline{F}(\omega_n; x) \rightarrow \infty$ ,  $\sqrt{nh^p \overline{F}(\omega_n; x)} b(\omega_n; x) \rightarrow \lambda$  and  $\sqrt{nh^p \overline{F}(\omega_n; x)} h \ln \omega_n \rightarrow 0$ ,*

$$\sqrt{nh^p \overline{F}(\omega_n; x) g(x)} \begin{bmatrix} \widehat{\gamma}_n^{(1)}(x, t_1, K_1, L_1) - \gamma(x) \\ \widehat{\gamma}_n^{(1)}(x, t_2, K_2, L_2) - \gamma(x) \end{bmatrix} \xrightarrow{\mathcal{D}} N_2(\lambda \boldsymbol{\mu}^{(1)}, \widetilde{\boldsymbol{\Sigma}}^{(1)}).$$

Consider now a weighted sum of two estimators with  $t_1 \neq t_2$  :

$$\widehat{\gamma}_n^{(1)}(x, \alpha) := \alpha \widehat{\gamma}_n^{(1)}(x, t_1, K_1, L_1) + (1 - \alpha) \widehat{\gamma}_n^{(1)}(x, t_2, K_2, L_2).$$

From Theorem 3.4 we obtain the asymptotic bias of such weighted estimators to be

$$\alpha \lambda \frac{\sqrt{g(x)}}{t_1 \rho(x)} \left( \frac{1}{(1 - \rho(x))^{t_1}} - 1 \right) + (1 - \alpha) \lambda \frac{\sqrt{g(x)}}{t_2 \rho(x)} \left( \frac{1}{(1 - \rho(x))^{t_2}} - 1 \right),$$

which can be eliminated by taking the weight  $\alpha$  as

$$\alpha_{BC}^{(1)}(\rho(x)) := \frac{\frac{1}{t_2} \left( \frac{1}{(1 - \rho(x))^{t_2}} - 1 \right)}{\frac{1}{t_2} \left( \frac{1}{(1 - \rho(x))^{t_2}} - 1 \right) - \frac{1}{t_1} \left( \frac{1}{(1 - \rho(x))^{t_1}} - 1 \right)}.$$

The result is formalized in the following corollary. Let  $\boldsymbol{\alpha}'_1 = [\alpha_{BC}^{(1)}(\rho(x)), 1 - \alpha_{BC}^{(1)}(\rho(x))]$ .

**Corollary 3.2.** *Under the conditions of Theorem 3.4 we have that*

$$\sqrt{nh^p \overline{F}(\omega_n; x) g(x)} [\widehat{\gamma}_n^{(1)}(x, \alpha_{BC}^{(1)}(\rho(x))) - \gamma(x)] \xrightarrow{\mathcal{D}} N(0, \boldsymbol{\alpha}'_1 \widetilde{\boldsymbol{\Sigma}}^{(1)} \boldsymbol{\alpha}_1).$$

The elimination of bias by the construction of an appropriately weighted sum of two estimators is also referred to as the generalized Jackknife methodology; see Gray and Schucany (1972) for further details.

We now turn to the construction of a bias-corrected estimator based on a weighted sum of two estimators of the type  $\widehat{\gamma}^{(2)}(x, t, K, L)$ .

Let  $\boldsymbol{\mu}^{(2)}$  be a  $(2 \times 1)$  vector with elements

$$\mu_i^{(2)} := \frac{\sqrt{g(x)}}{(1 - \rho(x))^{t_i+1}}, \quad i = 1, 2,$$

and  $\tilde{\Sigma}^{(2)}$  a  $(2 \times 2)$  matrix with elements

$$\begin{aligned}\tilde{\sigma}_{i,i}^{(2)} &:= \frac{\gamma^2(x)\Gamma(2t_i+1)}{(t_i+1)\Gamma^2(t_i+1)} [2(2t_i+1)\|K_i\|_2^2 + (t_i+1)\|L_i\|_2^2 - 2(2t_i+1)\|K_iL_i\|_1], \quad i=1,2, \\ \tilde{\sigma}_{1,2}^{(2)} &:= \frac{\gamma^2(x)\Gamma(t_1+t_2+1)}{\Gamma(t_1+2)\Gamma(t_2+2)} [(t_1+t_2+1)(t_1+t_2+2)\|K_1K_2\|_1 - (t_2+1)(t_1+t_2+1)\|K_1L_2\|_1 \\ &\quad - (t_1+1)(t_1+t_2+1)\|K_2L_1\|_1 + (t_1+1)(t_2+1)\|L_1L_2\|_1].\end{aligned}$$

**Theorem 3.5.** *Let  $(X_1, Y_1), \dots, (X_n, Y_n)$  be a sample of iid random vectors, and assume  $(\mathcal{R}_2)$ ,  $(\mathcal{G})$ ,  $(\mathcal{F})$  and kernel functions  $K_1, K_2, L_1$  and  $L_2$  satisfying  $(\mathcal{K})$ . For all  $x \in \mathbb{R}^p$  where  $g(x) > 0$ , we have that if  $h \rightarrow 0$ ,  $\omega_n \rightarrow \infty$  for  $n \rightarrow \infty$ , with  $nh^p\bar{F}(\omega_n; x) \rightarrow \infty$ ,  $\sqrt{nh^p\bar{F}(\omega_n; x)b(\omega_n; x)} \rightarrow \lambda$  and  $\sqrt{nh^p\bar{F}(\omega_n; x)h \ln \omega_n} \rightarrow 0$ ,*

$$\sqrt{nh^p\bar{F}(\omega_n; x)g(x)} \begin{bmatrix} \hat{\gamma}_n^{(2)}(x, t_1, K_1, L_1) - \gamma(x) \\ \hat{\gamma}_n^{(2)}(x, t_2, K_2, L_2) - \gamma(x) \end{bmatrix} \xrightarrow{\mathcal{D}} N_2(\lambda\boldsymbol{\mu}^{(2)}, \tilde{\Sigma}^{(2)}).$$

For a weighted estimator

$$\hat{\gamma}_n^{(2)}(x, \alpha) := \alpha\hat{\gamma}_n^{(2)}(x, t_1, K_1, L_1) + (1-\alpha)\hat{\gamma}_n^{(2)}(x, t_2, K_2, L_2),$$

with  $t_1 \neq t_2$ , we obtain from Theorem 3.5 that its asymptotic bias is given by

$$\alpha\lambda\frac{\sqrt{g(x)}}{(1-\rho(x))^{t_1+1}} + (1-\alpha)\lambda\frac{\sqrt{g(x)}}{(1-\rho(x))^{t_2+1}},$$

which vanishes if  $\alpha$  is taken as

$$\alpha_{BC}^{(2)}(\rho(x)) := \frac{1}{1 - (1 - \rho(x))^{t_2 - t_1}}.$$

Let  $\boldsymbol{\alpha}'_2 = [\alpha_{BC}^{(2)}(\rho(x)), 1 - \alpha_{BC}^{(2)}(\rho(x))]$ .

**Corollary 3.3.** *Under the conditions of Theorem 3.5 we have that*

$$\sqrt{nh^p\bar{F}(\omega_n; x)g(x)} [\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(\rho(x))) - \gamma(x)] \xrightarrow{\mathcal{D}} N(0, \boldsymbol{\alpha}'_2\tilde{\Sigma}^{(2)}\boldsymbol{\alpha}_2).$$

From the above discussion it is clear that the bias-correcting weights  $\alpha_{BC}^{(1)}(\rho(x))$  and  $\alpha_{BC}^{(2)}(\rho(x))$  depend on the second order parameter  $\rho(x)$ , which is typically unknown, and therefore we introduce next a consistent estimator for this parameter. In the univariate context, several estimators have been proposed that show good practical performance, for instance, the estimators by Gomes *et al.* (2002), Fraga Alves *et al.* (2003), Ciuperca and Mercadier (2010) and Goegebeur *et al.* (2010).

Let  $\mathbb{K} := (K_1, K_2, K_3)$  and  $\mathbb{L} := (L_1, L_2, L_3)$  be kernel functions satisfying  $(\mathcal{K})$ . Inspired by Fraga Alves *et al.* (2003) we introduce the following class of estimators for  $\rho(x)$ , parameterized in terms of a tuning parameter  $\tau \geq 0$  :

$$\widehat{\rho}_n(x; \mathbb{K}, \mathbb{L}, \tau) := \frac{3(R_n(x; \mathbb{K}, \mathbb{L}, \tau) - 1)}{R_n(x; \mathbb{K}, \mathbb{L}, \tau) - 3},$$

provided  $1 \leq R_n(x; \mathbb{K}, \mathbb{L}, \tau) < 3$ , where

$$R_n(x; \mathbb{K}, \mathbb{L}, \tau) := \frac{\left(\frac{T_n^{(1,1)}(x, K_1)}{T_n^{(1,0)}(x, L_1)}\right)^\tau - \left(\frac{T_n^{(1,2)}(x, K_2)}{2T_n^{(1,0)}(x, L_2)}\right)^{\tau/2}}{\left(\frac{T_n^{(1,2)}(x, K_2)}{2T_n^{(1,0)}(x, L_2)}\right)^{\tau/2} - \left(\frac{T_n^{(1,3)}(x, K_3)}{6T_n^{(1,0)}(x, L_3)}\right)^{\tau/3}}$$

and with the notation  $a^{b\tau} := b \ln a$  whenever  $\tau = 0$ . The consistency of  $\widehat{\rho}_n(x; \mathbb{K}, \mathbb{L}, \tau)$  for  $\rho(x)$  is established in the following theorem.

**Theorem 3.6.** *Let  $(X_1, Y_1), \dots, (X_n, Y_n)$  be a sample of iid random vectors, and assume  $(\mathcal{R}_2)$ ,  $(\mathcal{G})$ ,  $(\mathcal{F})$  and kernel functions  $K_1, K_2, K_3, L_1, L_2$  and  $L_3$  satisfying  $(\mathcal{K})$ . For all  $x \in \mathbb{R}^p$  where  $g(x) > 0$ , if  $h \rightarrow 0$ ,  $\omega_n \rightarrow \infty$  for  $n \rightarrow \infty$ , with  $\sqrt{nh^p \overline{F}(\omega_n; x)} b(\omega_n; x) \rightarrow \infty$  and  $h \ln \omega_n = o(b(\omega_n; x))$ , we have that  $\widehat{\rho}_n(x; \mathbb{K}, \mathbb{L}, \tau) \xrightarrow{\mathbb{P}} \rho(x)$ .*

The next two lemma's state that replacing  $\rho(x)$  in  $\widehat{\gamma}_n^{(1)}(x, \alpha_{BC}^{(1)}(\rho(x)))$  and  $\widehat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(\rho(x)))$  by a consistent estimator  $\widehat{\rho}_n(x)$  does not change the limiting distribution of the bias-corrected estimators.

**Corollary 3.4.** *Under the conditions of Theorem 3.4, and with  $\hat{\rho}_n(x)$  being a consistent estimator for  $\rho(x)$ , we have that*

$$\sqrt{nh^p \bar{F}(\omega_n; x) g(x)} [\hat{\gamma}_n^{(1)}(x, \alpha_{BC}^{(1)}(\hat{\rho}_n(x))) - \gamma(x)] \xrightarrow{\mathcal{D}} N(0, \alpha_1' \tilde{\Sigma}^{(1)} \alpha_1).$$

**Corollary 3.5** *Under the conditions of Theorem 3.5, and with  $\hat{\rho}_n(x)$  being a consistent estimator for  $\rho(x)$ , we have that*

$$\sqrt{nh^p \bar{F}(\omega_n; x) g(x)} [\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(\hat{\rho}_n(x))) - \gamma(x)] \xrightarrow{\mathcal{D}} N(0, \alpha_2' \tilde{\Sigma}^{(2)} \alpha_2).$$

The corollaries follow from a straightforward application of Taylor's theorem to  $\hat{\gamma}_n^{(1)}(x, \alpha_{BC}^{(1)}(\hat{\rho}_n(x)))$  and  $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(\hat{\rho}_n(x)))$ , respectively, and for brevity we omit their proofs.

### 3.4 Simulation results

In this section we illustrate our results by a small simulation study. For the practical implementation of our estimators we have to determine the bandwidth parameter  $h$  and the threshold  $\omega_n$ , where we take, as usual in extreme value statistics, the latter as the  $(k + 1)$ -th largest response observation in the ball  $B(x, h)$ . We propose two algorithms for the purpose of selecting  $(h, k)$  : (i) an oracle strategy and (ii) a completely data driven method.

As proposed in Daouia *et al.* (2011), the oracle strategy consists in selecting  $(h, k)$  as follows

$$(h_o, k_o) := \underset{h, k}{\operatorname{argmin}} \Delta(\hat{\gamma}(\cdot), \gamma(\cdot)), \tag{3.5}$$

with

$$\Delta^2(\widehat{\gamma}(\cdot), \gamma(\cdot)) := \frac{1}{M} \sum_{m=1}^M (\widehat{\gamma}(z_m) - \gamma(z_m))^2,$$

where  $z_1, \dots, z_M$  are regularly spaced in the covariate space. Note that this method requires knowledge of the function  $\gamma(x)$ , which is unknown in practical situations.

To solve this issue we introduce our second approach which is completely data driven, and based on a two steps procedure. First we select the bandwidth parameter  $h$  using the following cross-validation criterion

$$h_{cv} := \operatorname{argmin}_{h \in \mathcal{H}} \sum_{i=1}^n \sum_{j=1}^n \left( \mathbf{1}\{Y_i \leq Y_j\} - \widehat{F}_{n,-i}(Y_j|X_i) \right)^2,$$

where  $\mathcal{H}$  is a grid of values for  $h$  and

$$\widehat{F}_{n,-i}(y|x) := \frac{\sum_{j=1, j \neq i}^n K_h(x - X_j) \mathbf{1}\{Y_j \leq y\}}{\sum_{j=1, j \neq i}^n K_h(x - X_j)}.$$

This criterion was introduced in Yao (1999), implemented by Gannoun *et al.* (2002), and considered in an extreme value context by Daouia *et al.* (2011, 2012). Using this bandwidth we determine in the second step the number of extremes  $k$  with a stability criterion of the estimates for  $\gamma(x)$ . The algorithm that is used is as follows :

For all  $x$  under consideration

- compute the estimates for  $\gamma(x)$  with  $k = 5, \dots, m_x - 1$  ( $m_x$  is the number of observations in  $B(x, h)$ ),
- split the range of  $k$  into several blocks of same size,
- calculate the standard deviation of the estimates for  $\gamma(x)$  in each block,
- the block with minimal standard deviation determines the  $k$  to be used (in particular we take the median of the  $k$  values in that block).

Note that in the data driven method  $h$  is selected globally while  $k$  is selected locally.



In the simulations we consider the estimator  $\widehat{\gamma}_n^{(2)}(x, 0, K, L)$  and its bias-corrected version  $\widehat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(\rho(x)))$ , with  $t_1 = 0$  and  $t_2 = 1$ . All kernel functions are taken as the bi-quadratic kernel function

$$K(x) = \frac{15}{16}(1 - x^2)^2 \mathbf{1}\{x \in [-1, 1]\}.$$

We assume that the conditional distribution of  $Y$  given  $X = x$  is the following Burr distribution :

$$1 - F(y; x) = \left(1 + y^{-\rho(x)/\gamma(x)}\right)^{1/\rho(x)},$$

where  $\rho(x) < 0$  is the second order parameter, and  $X$  is uniformly distributed on  $(0, 1)$ .

We fix  $\rho(x) = -1$  and the function  $\gamma(\cdot)$  is, as in Daouia *et al.* (2011), taken as

$$\gamma(x) = \frac{1}{2} \left( \frac{1}{10} + \sin(\pi x) \right) \left( \frac{11}{10} - \frac{1}{2} \exp \left( -64 \left( x - \frac{1}{2} \right)^2 \right) \right).$$

We simulate  $N = 100$  samples of size  $n = 1000$ .

In Figure 3.2 we show the boxplots of the 100 realizations of  $\gamma(x)$  and  $\rho(x)$  for different values of  $x$ , as obtained with the oracle method. The minimization (3.5) is performed on a grid of  $h \in [0.05, 0.5]$  and of  $k \in \{2, \dots, m_x - 1\}$ , with  $M = 35$ . For the bias-corrected estimator for  $\gamma(x)$ , different values for  $\rho(x)$  have been used namely, the true value  $\rho(x) = -1$ , a mis-specification at  $\rho(x) = -2$  and  $-5$ , and an estimation of  $\rho(x)$  using  $\widehat{\rho}_n(x; \mathbb{K}, \mathbb{K}, \tau)$  with  $\tau = 0.5$ . In case  $\rho(x)$  is estimated, we use an oracle strategy, similar to the one discussed above, with a grid of  $h \in [0.05, 0.5]$ , but this time  $k \in \{m_x/4, \dots, m_x - 1\}$ . As expected, the bias-corrected estimator  $\widehat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(\rho(x)))$  outperforms the uncorrected version  $\widehat{\gamma}_n^{(2)}(x, 0, K, K)$  in terms of bias, even with a mis-specification of  $\rho(x)$ . The estimator for  $\rho(x)$  has a reasonable performance, despite the small local sample sizes. In Table 3.1 we compare the errors  $\Delta^2(\widehat{\gamma}(\cdot), \gamma(\cdot))$ , averaged over the 100 simulated data

---

sets. Also in terms of these,  $\widehat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(\rho(x)))$  typically works better than  $\widehat{\gamma}_n^{(2)}(x, 0, K, K)$ .

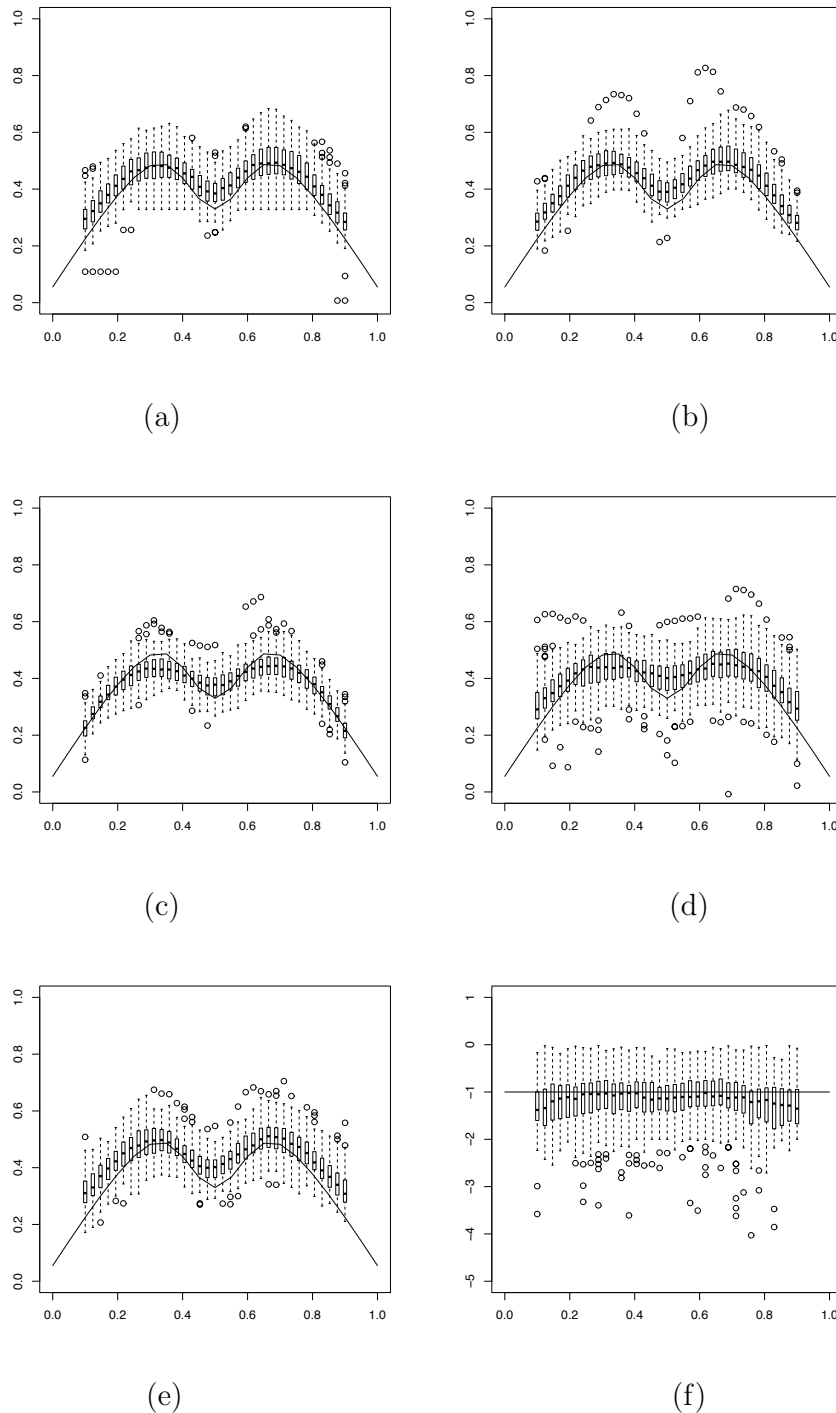


FIGURE 3.2 – Burr simulation, oracle criterion : boxplots of (a)  $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(-5))$ , (b)  $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(-2))$ , (c)  $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(-1))$ , (d)  $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(\hat{\rho}_n(x; \mathbb{K}, \mathbb{K}, 0.5)))$ , (e)  $\hat{\gamma}_n^{(2)}(x, 0, K, K)$  and (f)  $\hat{\rho}_n(x; \mathbb{K}, \mathbb{K}, 0.5)$ .

Estimator	Oracle	Data-driven
$\widehat{\gamma}_n^{(2)}(x, 0, K, K)$	0.00555	0.01930
$\widehat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(-1))$	0.00266	0.00608
$\widehat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(-2))$	0.00468	0.00648
$\widehat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(-5))$	0.00532	0.00729
$\widehat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(\widehat{\rho}_n(x; \mathbb{K}, \mathbb{K}, 0.5)))$	0.00691	0.00621

TABLE 3.1 – Average of  $\Delta^2(\widehat{\gamma}(\cdot), \gamma(\cdot))$  for  $\widehat{\gamma}_n^{(2)}(x, 0, K, K)$  and  $\widehat{\gamma}_n^{(2)}(x, \alpha_{BC}(\rho(x)))$  with oracle and data driven methods.

In Figure 3.3 we show the corresponding results obtained with the data driven method. For what concerns the estimation of  $\gamma(x)$ , we select the optimal value of the bandwidth  $h$  in a set  $\mathcal{H} = \{h_1 \leq \dots \leq h_{10}\}$  where the points  $h_1, \dots, h_{10}$  are regularly spaced in  $[0.05, 0.5]$ , and the selection of  $k$  is along the above described algorithm with a block size of 40. For the estimation of  $\rho(x)$ , we select  $h$  with cross-validation, but we set  $k = m_x^{0.975}$ , see e.g. Caeiro *et al.* (2009). Clearly, all implementations of the bias-corrected estimator outperform the uncorrected estimator in terms of bias, even with a severe mis-specification at  $\rho(x) = -5$ . The superior performance is also reflected in the errors reported in Table 3.1. Note in particular that the error of  $\widehat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(\rho(x)))$  with an estimation of  $\rho(x)$  is close to the one with the true value  $\rho(x) = -1$ . Also, they are only slightly larger in the data driven setting than in the oracle framework.

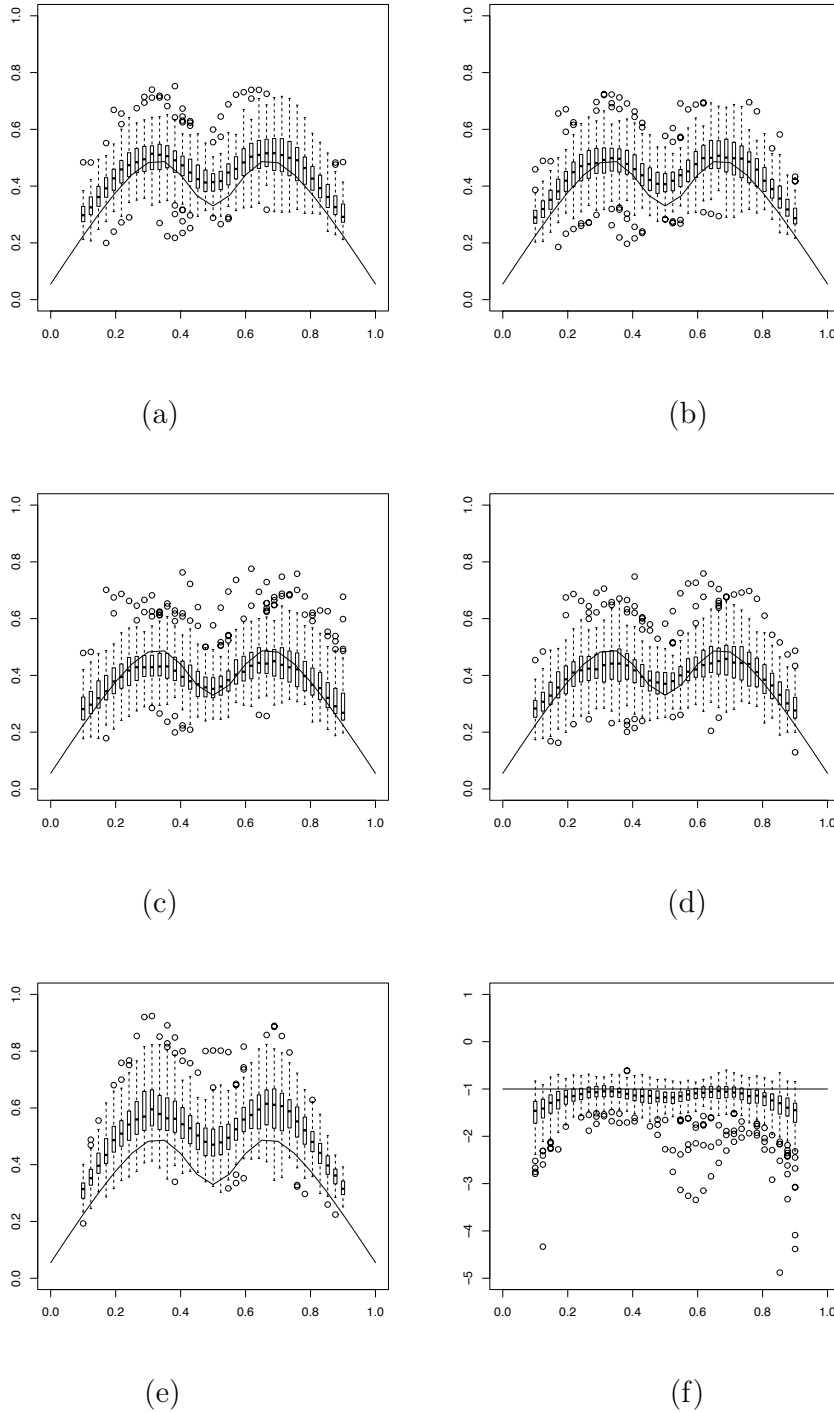


FIGURE 3.3 – Burr simulation, data driven criterion : boxplots of (a)  $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(-5))$ , (b)  $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(-2))$ , (c)  $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(-1))$ , (d)  $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(\hat{\rho}_n(x; \mathbb{K}, \mathbb{K}, 0.5)))$ , (e)  $\hat{\gamma}_n^{(2)}(x, 0, K, K)$  and (f)  $\hat{\rho}_n(x; \mathbb{K}, \mathbb{K}, 0.5)$ .

## 3.5 Proofs

### 3.5.1 Proof of Lemma 3.1

First consider the case  $t > 0$ . From the definition of conditional expectation one easily obtains

$$\begin{aligned} m(\omega_n, t; x) &= t \int_{\omega_n}^{\infty} (\ln y - \ln \omega_n)^{t-1} \frac{1}{y} \bar{F}(y; x) dy \\ &= t \bar{F}(\omega_n; x) \int_{\omega_n}^{\infty} (\ln y - \ln \omega_n)^{t-1} \frac{\bar{F}(y; x)}{y \bar{F}(\omega_n; x)} dy \\ &= t \bar{F}(\omega_n; x) \int_1^{\infty} (\ln z)^{t-1} \frac{\bar{F}(\omega_n z; x)}{z \bar{F}(\omega_n; x)} dz. \end{aligned}$$

For what concerns the ratio  $\bar{F}(\omega_n z; x)/\bar{F}(\omega_n; x)$  we invoke condition  $(\mathcal{R}_2)$ . From Theorem B.2.18 in de Haan and Ferreira (2006), see also Drees (1998), we have that for a function  $b_0$ , possibly different from the function  $b$ , though with  $b_0(\omega_n; x) \sim b(\omega_n; x)$ ,  $\omega_n \rightarrow \infty$ , and for each  $\varepsilon, \delta > 0$  there exists a  $\omega_{n,0}$  such that for  $\omega_n \geq \omega_{n,0}$ ,  $z > 1$ ,

$$\left| \gamma^2(x) \frac{z^{1/\gamma(x)} \frac{\bar{F}(\omega_n z; x)}{\bar{F}(\omega_n; x)} - 1}{b_0(\omega_n; x)} - \frac{z^{\rho(x)/\gamma(x)} - 1}{\rho(x)/\gamma(x)} \right| \leq \varepsilon z^{\rho(x)/\gamma(x) + \delta}. \quad (3.6)$$

Motivated by (3.6), we write

$$\begin{aligned} m(\omega_n, t; x) &= t \bar{F}(\omega_n; x) \left\{ \int_1^{\infty} (\ln z)^{t-1} z^{-1/\gamma(x)-1} dz \right. \\ &\quad + \frac{b_0(\omega_n; x)}{\gamma^2(x)} \int_1^{\infty} (\ln z)^{t-1} z^{-1/\gamma(x)-1} \frac{z^{\rho(x)/\gamma(x)} - 1}{\rho(x)/\gamma(x)} dz \\ &\quad \left. + \frac{b_0(\omega_n; x)}{\gamma^2(x)} \int_1^{\infty} (\ln z)^{t-1} z^{-1/\gamma(x)-1} \left( \gamma^2(x) \frac{z^{1/\gamma(x)} \frac{\bar{F}(\omega_n z; x)}{\bar{F}(\omega_n; x)} - 1}{b_0(\omega_n; x)} - \frac{z^{\rho(x)/\gamma(x)} - 1}{\rho(x)/\gamma(x)} \right) dz \right\} \\ &=: t \bar{F}(\omega_n; x) \left\{ I_1 + \frac{b_0(\omega_n; x)}{\gamma^2(x)} (I_2 + I_3) \right\}. \end{aligned}$$

For  $I_1$  and  $I_2$  one obtains easily

$$\begin{aligned} I_1 &= \gamma^t(x)\Gamma(t), \\ I_2 &= \frac{\gamma^{t+1}(x)\Gamma(t)}{\rho(x)} \left[ \frac{1}{(1-\rho(x))^t} - 1 \right]. \end{aligned}$$

For what concerns  $I_3$  we use (3.6). Hence, for all  $\varepsilon > 0$  we have that there exists an  $\omega_{n,0}$  such that for  $\omega_n > \omega_{n,0}$

$$\begin{aligned} |I_3| &\leq \varepsilon \int_1^\infty (\ln z)^{t-1} z^{-(1-\rho(x))/\gamma(x)+\delta-1} dz \\ &= \varepsilon \int_0^\infty \exp(-((1-\rho(x))/\gamma(x) - \delta)u) u^{t-1} du \\ &= \varepsilon \frac{\Gamma(t)}{\left(\frac{1-\rho(x)}{\gamma(x)} - \delta\right)^t}, \end{aligned}$$

provided  $0 < \delta < (1-\rho(x))/\gamma(x)$ . We have thus that  $I_3 = o(1)$  for  $\omega_n \rightarrow \infty$ . The result for  $t > 0$  is then obtained by collecting the terms.

The case  $t = 0$  follows trivially from the definition of conditional expectation.

### 3.5.2 Proof of Lemma 3.2

From the rule of repeated expectations we obtain

$$\begin{aligned} \tilde{m}_n(x; s, t) &= \mathbb{E}[K_h^s(x - X)m(\omega_n, t; X)] \\ &= \int_{\mathbb{R}^p} K_h^s(x - u)m(\omega_n, t; u)g(u)du \\ &= \frac{\|K\|_s^s}{h^{(s-1)p}} \int_{\Omega} \frac{K^s(z)}{\|K\|_s^s} m(\omega_n, t; x - hz)g(x - hz)dz. \end{aligned}$$

In case  $t > 0$  we have

$$\tilde{m}_n(x; s, t) = \frac{t\|K\|_s^s}{h^{(s-1)p}} \int_{\Omega} \frac{K^s(z)}{\|K\|_s^s} \int_{\omega_n}^\infty (\ln y - \ln \omega_n)^{t-1} \frac{\bar{F}(y; x - hz)}{y} dy g(x - hz)dz,$$

and, consequently

$$\begin{aligned}
& \left| \tilde{m}_n(x; s, t) - \frac{m(\omega_n, t; x)}{h^{(s-1)p}} g(x) \|K\|_s^s \right| \\
& \leq \frac{t \|K\|_s^s}{h^{(s-1)p}} \left\{ \int_{\Omega} \frac{K^s(z)}{\|K\|_s^s} \int_{\omega_n}^{\infty} (\ln y - \ln \omega_n)^{t-1} \frac{\bar{F}(y; x)}{y} \left| \frac{\bar{F}(y; x-hz)}{\bar{F}(y; x)} - 1 \right| dy g(x) dz \right. \\
& \quad \left. + \int_{\Omega} \frac{K^s(z)}{\|K\|_s^s} \int_{\omega_n}^{\infty} (\ln y - \ln \omega_n)^{t-1} \frac{\bar{F}(y; x)}{y} dy |g(x-hz) - g(x)| dz \right. \\
& \quad \left. + \int_{\Omega} \frac{K^s(z)}{\|K\|_s^s} \int_{\omega_n}^{\infty} (\ln y - \ln \omega_n)^{t-1} \frac{\bar{F}(y; x)}{y} \left| \frac{\bar{F}(y; x-hz)}{\bar{F}(y; x)} - 1 \right| dy |g(x-hz) - g(x)| dz \right\} \\
& =: \frac{t \|K\|_s^s}{h^{(s-1)p}} \{I_4 + I_5 + I_6\}.
\end{aligned}$$

For what concerns  $I_4$ , let

$$\tilde{I}_4 := \int_{\omega_n}^{\infty} (\ln y - \ln \omega_n)^{t-1} \frac{\bar{F}(y; x)}{y} \left| \frac{\bar{F}(y; x-hz)}{\bar{F}(y; x)} - 1 \right| dy,$$

note that

$$\frac{\bar{F}(y; x-hz)}{\bar{F}(y; x)} = \exp \left[ \ln \bar{F}(y; x) \left( \frac{\ln \bar{F}(y; x-hz)}{\ln \bar{F}(y; x)} - 1 \right) \right],$$

and, by using  $(\mathcal{F})$  and the property that  $\ln \ell(y; x) / \ln y \rightarrow 0$  as  $y \rightarrow \infty$ , we have that if  $\omega_n$  is sufficiently large then

$$|\ln \bar{F}(y; x)| \left| \frac{\ln \bar{F}(y; x-hz)}{\ln \bar{F}(y; x)} - 1 \right| \leq Ch \ln y,$$



for some constant  $C > 0$ . Now use the well known inequality  $|\exp(w) - 1| \leq |w| + \frac{1}{2}(1 + \exp(|w|))w^2$  to obtain for  $\omega_n$  sufficiently large

$$\begin{aligned} \tilde{I}_4 &\leq Ch \int_{\omega_n}^{\infty} (\ln y - \ln \omega_n)^{t-1} \frac{\bar{F}(y; x)}{y} \ln y \, dy \\ &\quad + \frac{C^2 h^2}{2} \int_{\omega_n}^{\infty} (\ln y - \ln \omega_n)^{t-1} \frac{\bar{F}(y; x)}{y} \ln^2 y \, dy \\ &\quad + \frac{C^2 h^2}{2} \int_{\omega_n}^{\infty} (\ln y - \ln \omega_n)^{t-1} \frac{\bar{F}(y; x)}{y} \exp(Ch \ln y) \ln^2 y \, dy \\ &=: \tilde{I}_{4,1} + \tilde{I}_{4,2} + \tilde{I}_{4,3}. \end{aligned}$$

From Lemma 3.1 we have that

$$\begin{aligned} \tilde{I}_{4,1} &= \frac{m(\omega_n, t; x)}{t} O(h \ln \omega_n), \\ \tilde{I}_{4,2} &= \frac{m(\omega_n, t; x)}{t} O(h^2 \ln^2 \omega_n). \end{aligned}$$

Concerning  $\tilde{I}_{4,3}$ , by the substitution  $v = y/\omega_n$  and a straightforward expansion

$$\begin{aligned} \tilde{I}_{4,3} &= \frac{C^2 h^2 \exp(Ch \ln \omega_n)}{2} \left\{ \ln^2 \omega_n \int_1^{\infty} (\ln v)^{t-1} \frac{\bar{F}(\omega_n v; x)}{v} \exp(Ch \ln v) \, dv \right. \\ &\quad + \int_1^{\infty} (\ln v)^{t+1} \frac{\bar{F}(\omega_n v; x)}{v} \exp(Ch \ln v) \, dv \\ &\quad \left. + 2 \ln \omega_n \int_1^{\infty} (\ln v)^t \frac{\bar{F}(\omega_n v; x)}{v} \exp(Ch \ln v) \, dv \right\} \\ &=: \frac{C^2 h^2 \exp(Ch \ln \omega_n)}{2} \left\{ \ln^2 \omega_n \tilde{I}_{4,3,1} + \tilde{I}_{4,3,2} + 2 \ln \omega_n \tilde{I}_{4,3,3} \right\}. \end{aligned}$$

Since all the integrals in the right hand side of the expression are of the same form, we will only consider the first of them. Again, inspired by (3.6), we write

$$\begin{aligned} \tilde{I}_{4,3,1} &= \bar{F}(\omega_n; x) \left\{ \int_1^\infty (\ln v)^{t-1} v^{Ch-1-1/\gamma(x)} dv \right. \\ &\quad + \frac{b_0(\omega_n; x)}{\gamma^2(x)} \int_1^\infty (\ln v)^{t-1} v^{Ch-1-1/\gamma(x)} \frac{v^{\rho(x)/\gamma(x)} - 1}{\rho(x)/\gamma(x)} dv \\ &\quad \left. + \frac{b_0(\omega_n; x)}{\gamma^2(x)} \int_1^\infty (\ln v)^{t-1} v^{Ch-1-1/\gamma(x)} \left( \gamma^2(x) \frac{v^{1/\gamma(x)} \frac{\bar{F}(\omega_n v; x)}{\bar{F}(\omega_n; x)} - 1}{b_0(\omega_n; x)} - \frac{v^{\rho(x)/\gamma(x)} - 1}{\rho(x)/\gamma(x)} \right) dv \right\} \\ &=: \bar{F}(\omega_n; x) \left\{ \tilde{I}_{4,3,1,1} + \frac{b_0(\omega_n; x)}{\gamma^2(x)} (\tilde{I}_{4,3,1,2} + \tilde{I}_{4,3,1,3}) \right\}. \end{aligned}$$

Taking into account that  $Ch$  can be made arbitrarily small we have that

$$\begin{aligned} \tilde{I}_{4,3,1,1} &= \gamma^t(x) \Gamma(t) + o(1), \\ \tilde{I}_{4,3,1,2} &= \frac{\gamma^{t+1}(x) \Gamma(t)}{\rho(x)} \left[ \frac{1}{(1-\rho(x))^t} - 1 \right] + o(1). \end{aligned}$$

For what concerns  $\tilde{I}_{4,3,1,3}$  we assume that  $Ch < \delta'$ . Then, again, by (3.6) we have that for all  $\varepsilon > 0$  there exists an  $\omega_{n,0}$  such that for  $\omega_n > \omega_{n,0}$

$$\begin{aligned} |\tilde{I}_{4,3,1,3}| &\leq \varepsilon \int_1^\infty (\ln v)^{t-1} v^{-(1-\rho(x))/\gamma(x)+\delta'+\delta-1} dz \\ &= \varepsilon \frac{\Gamma(t)}{\left( \frac{1-\rho(x)}{\gamma(x)} - \delta' - \delta \right)^t}, \end{aligned}$$

provided  $0 < \delta + \delta' < (1 - \rho(x))/\gamma(x)$ . Hence  $\tilde{I}_{4,3,1,3} = o(1)$  under the conditions of the lemma. Thus

$$\begin{aligned} \tilde{I}_{4,3,1} &= \bar{F}(\omega_n; x) \gamma^t(x) \Gamma(t) \left\{ 1 + o(1) + \frac{b_0(\omega_n; x)}{\gamma(x) \rho(x)} \left[ \frac{1}{(1-\rho(x))^t} - 1 \right] (1 + o(1)) \right\} \\ &= \frac{m(\omega_n, t; x)}{t} (1 + o(1)). \end{aligned}$$

Collecting the above results gives that

$$\tilde{I}_{4,3} = \frac{m(\omega_n, t; x)}{t} O(h^2 \ln^2 \omega_n).$$

Combined,

$$I_4 = \frac{1}{t}g(x)m(\omega_n, t; x)O(h \ln \omega_n).$$

Concerning  $I_5$ , we use  $(\mathcal{G})$  and  $(\mathcal{K})$  to obtain

$$\begin{aligned} I_5 &\leq c_g \int_{\Omega} \frac{K^s(z)}{\|K\|_s^s} d(0, hz) dz \int_{\omega_n}^{\infty} (\ln y - \ln \omega_n)^{t-1} \frac{\bar{F}(y; x)}{y} dy \\ &= \frac{1}{t}g(x)m(\omega_n, t; x)O(h). \end{aligned}$$

Similarly, by using  $(\mathcal{F})$ ,  $(\mathcal{G})$  and  $(\mathcal{K})$  we have that, for  $\omega_n$  sufficiently large,

$$I_6 = \frac{1}{t}g(x)m(\omega_n, t; x)O(h^2 \ln \omega_n).$$

Combining the above establishes the result for the case  $t > 0$ .

We now turn to the case  $t = 0$ . We have

$$\tilde{m}_n(x; s, 0) = \frac{\|K\|_s^s}{h^{(s-1)p}} \int_{\Omega} \frac{K^s(z)}{\|K\|_s^s} \bar{F}(\omega_n; x - hz) g(x - hz) dz.$$

Further,

$$\begin{aligned} &\left| \tilde{m}_n(x; s, 0) - \frac{m(\omega_n, 0; x)}{h^{(s-1)p}} g(x) \|K\|_s^s \right| \\ &\leq \frac{\|K\|_s^s}{h^{(s-1)p}} \left\{ \bar{F}(\omega_n; x) g(x) \int_{\Omega} \frac{K^s(z)}{\|K\|_s^s} \left| \frac{\bar{F}(\omega_n; x - hz)}{\bar{F}(\omega_n; x)} - 1 \right| dz \right. \\ &\quad + \bar{F}(\omega_n; x) \int_{\Omega} \frac{K^s(z)}{\|K\|_s^s} |g(x - hz) - g(x)| dz \\ &\quad \left. + \bar{F}(\omega_n; x) \int_{\Omega} \frac{K^s(z)}{\|K\|_s^s} \left| \frac{\bar{F}(\omega_n; x - hz)}{\bar{F}(\omega_n; x)} - 1 \right| |g(x - hz) - g(x)| dz \right\}. \end{aligned}$$

The result for the case  $t = 0$  can now be established along a similar line of argumentation as used for the case  $t > 0$ .

### 3.5.3 Proof of Theorem 3.1

To prove Theorem 3.1 we make use of the Cramér-Wold device (see e.g. van der Vaart, 1998, p 16), according to which it is sufficient to prove that for all  $\xi \in \mathbb{R}^J$  we have that

$$\Lambda_n := \sqrt{nh^p \bar{F}(\omega_n; x) g(x)} \xi' [\mathbb{T}_n - \mathbb{E}(\mathbb{T}_n)] \xrightarrow{\mathcal{D}} N(0, \xi' \Sigma \xi).$$

From straightforward calculations we obtain

$$\begin{aligned} \Lambda_n &= \sum_{i=1}^n \sqrt{\frac{h^p}{n \bar{F}(\omega_n; x) g(x)}} \left[ \sum_{j=1}^J \xi_j K_{j,h}(x - X_i) (\ln Y_i - \ln \omega_n)_+^{t_j} \mathbf{1}\{Y_i > \omega_n\} \right. \\ &\quad \left. - \mathbb{E} \left( \sum_{j=1}^J \xi_j K_{j,h}(x - X_i) (\ln Y_i - \ln \omega_n)_+^{t_j} \mathbf{1}\{Y_i > \omega_n\} \right) \right] \\ &=: \sum_{i=1}^n V_i. \end{aligned}$$

Note that  $V_1, \dots, V_n$  are iid random variables, and hence  $\text{Var}(\Lambda_n) = n \text{Var}(V_1)$ . Now

$$\begin{aligned} \text{Var}(V_1) &= \frac{h^p}{n \bar{F}(\omega_n; x) g(x)} \sum_{j=1}^J \sum_{k=1}^J \xi_j \xi_k \text{Cov} \left( K_{j,h}(x - X_1) (\ln Y_1 - \ln \omega_n)_+^{t_j} \mathbf{1}\{Y_1 > \omega_n\}, \right. \\ &\quad \left. K_{k,h}(x - X_1) (\ln Y_1 - \ln \omega_n)_+^{t_k} \mathbf{1}\{Y_1 > \omega_n\} \right) \\ &=: \frac{h^p}{n \bar{F}(\omega_n; x) g(x)} \sum_{j=1}^J \sum_{k=1}^J \xi_j \xi_k \mathbb{C}_{j,k}, \end{aligned}$$

and

$$\begin{aligned} \mathbb{C}_{j,k} &= \mathbb{E} \left[ K_{j,h}(x - X_1) K_{k,h}(x - X_1) (\ln Y_1 - \ln \omega_n)_+^{t_j+t_k} \mathbf{1}\{Y_1 > \omega_n\} \right] \\ &\quad - \mathbb{E} \left[ K_{j,h}(x - X_1) (\ln Y_1 - \ln \omega_n)_+^{t_j} \mathbf{1}\{Y_1 > \omega_n\} \right] \mathbb{E} \left[ K_{k,h}(x - X_1) (\ln Y_1 - \ln \omega_n)_+^{t_k} \mathbf{1}\{Y_1 > \omega_n\} \right] \\ &= \frac{\|K_j K_k\|_1}{h^p} \mathbb{E} \left[ \frac{1}{h^p \|K_j K_k\|_1} K_j \left( \frac{x - X_1}{h} \right) K_k \left( \frac{x - X_1}{h} \right) (\ln Y_1 - \ln \omega_n)_+^{t_j+t_k} \mathbf{1}\{Y_1 > \omega_n\} \right] \\ &\quad - \mathbb{E} \left[ K_{j,h}(x - X_1) (\ln Y_1 - \ln \omega_n)_+^{t_j} \mathbf{1}\{Y_1 > \omega_n\} \right] \mathbb{E} \left[ K_{k,h}(x - X_1) (\ln Y_1 - \ln \omega_n)_+^{t_k} \mathbf{1}\{Y_1 > \omega_n\} \right]. \end{aligned}$$

Using Lemma 3.1 and 3.2 we obtain that

$$\mathbb{C}_{j,k} = \frac{\|K_j K_k\|_1}{h^p} \gamma^{t_j+t_k}(x) \bar{F}(\omega_n; x) g(x) \Gamma(t_j + t_k + 1) (1 + o(1)),$$

and consequently  $\text{Var}(\Lambda_n) = \xi' \Sigma \xi (1 + o(1))$ .

To establish the asymptotic normality of  $\Lambda_n$  we verify Lyapounov's criterion for triangular arrays of random variables, see e.g. Billingsley (1995). In the present context this simplifies to proving that  $\sum_{i=1}^n \mathbb{E}|V_i|^3 = n\mathbb{E}|V_1|^3 \rightarrow 0$ . We have

$$\begin{aligned} \mathbb{E}|V_1|^3 &\leq \left( \frac{h^p}{n\bar{F}(\omega_n; x)g(x)} \right)^{3/2} \left\{ \mathbb{E} \left[ \left( \sum_{j=1}^J |\xi_j| K_{j,h}(x - X_1) (\ln Y_1 - \ln \omega_n)_+^{t_j} \mathbf{1}\{Y_1 > \omega_n\} \right)^3 \right] \right. \\ &\quad + 3\mathbb{E} \left[ \left( \sum_{j=1}^J |\xi_j| K_{j,h}(x - X_1) (\ln Y_1 - \ln \omega_n)_+^{t_j} \mathbf{1}\{Y_1 > \omega_n\} \right)^2 \right] \\ &\quad \times \mathbb{E} \left[ \sum_{j=1}^J |\xi_j| K_{j,h}(x - X_1) (\ln Y_1 - \ln \omega_n)_+^{t_j} \mathbf{1}\{Y_1 > \omega_n\} \right] \\ &\quad \left. + 4 \left[ \mathbb{E} \left( \sum_{j=1}^J |\xi_j| K_{j,h}(x - X_1) (\ln Y_1 - \ln \omega_n)_+^{t_j} \mathbf{1}\{Y_1 > \omega_n\} \right) \right]^3 \right\}. \end{aligned}$$

Again by using Lemma 3.1 and 3.2 we obtain that

$$\mathbb{E}|V_1|^3 = O \left( \left( n \sqrt{nh^p \bar{F}(\omega_n; x)} \right)^{-1} \right), \quad (3.7)$$

and hence, under the conditions of the theorem  $n\mathbb{E}|V_1|^3 \rightarrow 0$ .

### 3.5.4 Proof of Theorem 3.2

Let  $r_n := \sqrt{nh^p \bar{F}(\omega_n; x)g(x)}$ . Write

$$\begin{aligned} &r_n \left( \frac{T_n^{(1,t)}(x, K)}{\Gamma(t+1)T_n^{(1,0)}(x, L)} - \gamma^t(x) \right) \\ &= \frac{\bar{F}(\omega_n; x)g(x)}{T_n^{(1,0)}(x, L)} \left\{ r_n \left[ \frac{T_n^{(1,t)}(x, K)}{\Gamma(t+1)\bar{F}(\omega_n; x)g(x)} - \mathbb{E} \left( \frac{T_n^{(1,t)}(x, K)}{\Gamma(t+1)\bar{F}(\omega_n; x)g(x)} \right) \right] \right. \\ &\quad \left. - \gamma^t(x)r_n \left[ \frac{T_n^{(1,0)}(x, L)}{\bar{F}(\omega_n; x)g(x)} - \mathbb{E} \left( \frac{T_n^{(1,0)}(x, L)}{\bar{F}(\omega_n; x)g(x)} \right) \right] \right\} \end{aligned}$$

$$\begin{aligned}
& + r_n \left[ \frac{\mathbb{E}(T_n^{(1,t)}(x, K)) - \gamma^t(x)\Gamma(t+1)\mathbb{E}(T_n^{(1,0)}(x, L))}{\Gamma(t+1)\bar{F}(\omega_n; x)g(x)} \right] \Big\} \\
& =: \frac{\bar{F}(\omega_n; x)g(x)}{T_n^{(1,0)}(x, L)} \{T_1 + T_2 + T_3\}.
\end{aligned}$$

From Theorem 3.1, we have

$$T_1 + T_2 \xrightarrow{\mathcal{D}} N \left( 0, \frac{\gamma^{2t}(x)}{\Gamma^2(t+1)} [\Gamma(2t+1)\|K\|_2^2 + \Gamma^2(t+1)\|L\|_2^2 - 2\Gamma^2(t+1)\|KL\|_1] \right).$$

Concerning  $T_3$ , by using Lemma 3.1 and 3.2

$$T_3 = r_n \gamma^t(x) \left\{ \frac{b(\omega_n; x)}{\gamma(x)\rho(x)} \left( \frac{1}{(1-\rho(x))^t} - 1 \right) (1 + o(1)) + O(h \ln \omega_n) \right\},$$

and hence under our assumptions,  $T_3 \rightarrow \lambda \sqrt{g(x)} \frac{\gamma^{t-1}(x)}{\rho(x)} \left( \frac{1}{(1-\rho(x))^t} - 1 \right)$ .

Further, from Theorem 3.1 we have that  $T_n^{(1,0)}(x, L)/(\bar{F}(\omega_n; x)g(x)) = 1 + o_{\mathbb{P}}(1)$ .

Combined, the above gives that

$$\begin{aligned}
r_n \left( \frac{T_n^{(1,t)}(x, K)}{\Gamma(t+1)T_n^{(1,0)}(x, L)} - \gamma^t(x) \right) & \xrightarrow{\mathcal{D}} N \left( \lambda \sqrt{g(x)} \frac{\gamma^{t-1}(x)}{\rho(x)} \left( \frac{1}{(1-\rho(x))^t} - 1 \right), \right. \\
& \left. \frac{\gamma^{2t}(x)}{\Gamma^2(t+1)} [\Gamma(2t+1)\|K\|_2^2 + \Gamma^2(t+1)\|L\|_2^2 - 2\Gamma^2(t+1)\|KL\|_1] \right).
\end{aligned}$$

Finally, after a straightforward application of the  $\delta$ -method, Theorem 3.2 follows.

### 3.5.5 Proof of Theorem 3.3

First write

$$\begin{aligned}
& r_n \left( \frac{T_n^{(1,t+1)}(x, K)}{(t+1)T_n^{(1,t)}(x, L)} - \gamma(x) \right) \\
& = \frac{\bar{F}(\omega_n; x)g(x)}{T_n^{(1,t)}(x, L)} \left\{ r_n \left[ \frac{T_n^{(1,t+1)}(x, K)}{(t+1)\bar{F}(\omega_n; x)g(x)} - \mathbb{E} \left( \frac{T_n^{(1,t+1)}(x, K)}{(t+1)\bar{F}(\omega_n; x)g(x)} \right) \right] \right. \\
& \quad \left. - \gamma(x) r_n \left[ \frac{T_n^{(1,t)}(x, L)}{\bar{F}(\omega_n; x)g(x)} - \mathbb{E} \left( \frac{T_n^{(1,t)}(x, L)}{\bar{F}(\omega_n; x)g(x)} \right) \right] \right\}
\end{aligned}$$

$$\begin{aligned}
 & + r_n \left[ \frac{\mathbb{E}(T_n^{(1,t+1)}(x, K)) - \gamma(x)(t+1)\mathbb{E}(T_n^{(1,t)}(x, L))}{(t+1)\bar{F}(\omega_n; x)g(x)} \right] \Big\} \\
 & =: \frac{\bar{F}(\omega_n; x)g(x)}{T_n^{(1,t)}(x, L)} \{T_4 + T_5 + T_6\}.
 \end{aligned}$$

From Theorem 3.1, we have

$$T_4 + T_5 \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, \frac{\gamma^{2t+2}(x)\Gamma(2t+1)}{t+1} [2(2t+1)\|K\|_2^2 + (t+1)\|L\|_2^2 - 2(2t+1)\|KL\|_1] \right).$$

Concerning  $T_6$ , using Lemma 3.1 and 3.2, we have

$$T_6 = r_n \left\{ \frac{\gamma^t(x)\Gamma(t+1)b(\omega_n; x)}{(1-\rho(x))^{t+1}}(1+o(1)) + O(h \ln \omega_n) \right\},$$

and hence under our assumptions, we have  $T_6 \rightarrow \lambda \sqrt{g(x)} \frac{\gamma^t(x)\Gamma(t+1)}{(1-\rho(x))^{t+1}}$ .

Now, according to Theorem 3.1,  $T_n^{(1,t)}(x, L)/(\bar{F}(\omega_n; x)g(x)) = \gamma^t(x)\Gamma(t+1) + o_{\mathbb{P}}(1)$ . This achieves the proof of Theorem 3.3.

### 3.5.6 Proof of Theorem 3.4

First we will show that under the conditions of the theorem

$$r_n \left[ \begin{array}{c} \frac{T_n^{(1,t_1)}(x, K_1)}{\Gamma(t_1+1)T_n^{(1,0)}(x, L_1)} - \gamma^{t_1}(x) \\ \frac{T_n^{(1,t_2)}(x, K_2)}{\Gamma(t_2+1)T_n^{(1,0)}(x, L_2)} - \gamma^{t_2}(x) \end{array} \right] \xrightarrow{\mathcal{D}} N_2(\boldsymbol{\delta}, \boldsymbol{\Omega}) \quad (3.8)$$

whereafter the result will follow by a straightforward application of the  $\delta$ -method. Using the Cramér-Wold device, to establish (3.8) it is sufficient to show that for all  $\xi \in \mathbb{R}^2$  we have that

$$\xi' r_n \left[ \begin{array}{c} \frac{T_n^{(1,t_1)}(x, K_1)}{\Gamma(t_1+1)T_n^{(1,0)}(x, L_1)} - \gamma^{t_1}(x) \\ \frac{T_n^{(1,t_2)}(x, K_2)}{\Gamma(t_2+1)T_n^{(1,0)}(x, L_2)} - \gamma^{t_2}(x) \end{array} \right] \xrightarrow{\mathcal{D}} N(\xi' \boldsymbol{\delta}, \xi' \boldsymbol{\Omega} \xi). \quad (3.9)$$

From Lemma 3.1 and 3.2, and Theorem 3.1 we have that  $T_n^{(1,0)}(x, L_i)/(\bar{F}(\omega_n; x)g(x)) = 1 + o_{\mathbb{P}}(1)$ ,  $i = 1, 2$ , and consequently, by the line of argumentation of Theorem 3.2 we

obtain

$$\begin{aligned} \xi' r_n \left[ \begin{array}{l} \frac{T_n^{(1,t_1)}(x, K_1)}{\Gamma(t_1+1)T_n^{(1,0)}(x, L_1)} - \gamma^{t_1}(x) \\ \frac{T_n^{(1,t_2)}(x, K_2)}{\Gamma(t_2+1)T_n^{(1,0)}(x, L_2)} - \gamma^{t_2}(x) \end{array} \right] &= \xi_1 r_n \frac{T_n^{(1,t_1)}(x, K_1) - \gamma^{t_1}(x)\Gamma(t_1+1)T_n^{(1,0)}(x, L_1)}{\Gamma(t_1+1)\bar{F}(\omega_n; x)g(x)} \\ &+ \xi_2 r_n \frac{T_n^{(1,t_2)}(x, K_2) - \gamma^{t_2}(x)\Gamma(t_2+1)T_n^{(1,0)}(x, L_2)}{\Gamma(t_2+1)\bar{F}(\omega_n; x)g(x)} \\ &+ o_{\mathbb{P}}(1), \end{aligned}$$

which can be further decomposed as

$$\begin{aligned} \xi' r_n \left[ \begin{array}{l} \frac{T_n^{(1,t_1)}(x, K_1)}{\Gamma(t_1+1)T_n^{(1,0)}(x, L_1)} - \gamma^{t_1}(x) \\ \frac{T_n^{(1,t_2)}(x, K_2)}{\Gamma(t_2+1)T_n^{(1,0)}(x, L_2)} - \gamma^{t_2}(x) \end{array} \right] &= \frac{\xi_1}{\Gamma(t_1+1)} r_n \left[ \frac{T_n^{(1,t_1)}(x, K_1)}{\bar{F}(\omega_n; x)g(x)} - \mathbb{E} \left( \frac{T_n^{(1,t_1)}(x, K_1)}{\bar{F}(\omega_n; x)g(x)} \right) \right] \\ &- \xi_1 \gamma^{t_1}(x) r_n \left[ \frac{T_n^{(1,0)}(x, L_1)}{\bar{F}(\omega_n; x)g(x)} - \mathbb{E} \left( \frac{T_n^{(1,0)}(x, L_1)}{\bar{F}(\omega_n; x)g(x)} \right) \right] \\ &+ \frac{\xi_2}{\Gamma(t_2+1)} r_n \left[ \frac{T_n^{(1,t_2)}(x, K_2)}{\bar{F}(\omega_n; x)g(x)} - \mathbb{E} \left( \frac{T_n^{(1,t_2)}(x, K_2)}{\bar{F}(\omega_n; x)g(x)} \right) \right] \\ &- \xi_2 \gamma^{t_2}(x) r_n \left[ \frac{T_n^{(1,0)}(x, L_2)}{\bar{F}(\omega_n; x)g(x)} - \mathbb{E} \left( \frac{T_n^{(1,0)}(x, L_2)}{\bar{F}(\omega_n; x)g(x)} \right) \right] \\ &+ \xi_1 r_n \frac{\mathbb{E}(T_n^{(1,t_1)}(x, K_1)) - \gamma^{t_1}(x)\Gamma(t_1+1)\mathbb{E}(T_n^{(1,0)}(x, L_1))}{\Gamma(t_1+1)\bar{F}(\omega_n; x)g(x)} \\ &+ \xi_2 r_n \frac{\mathbb{E}(T_n^{(1,t_2)}(x, K_2)) - \gamma^{t_2}(x)\Gamma(t_2+1)\mathbb{E}(T_n^{(1,0)}(x, L_2))}{\Gamma(t_2+1)\bar{F}(\omega_n; x)g(x)} \\ &+ o_{\mathbb{P}}(1) \\ &=: T_7 + T_8 + T_9 + T_{10} + T_{11} + T_{12} + o_{\mathbb{P}}(1). \end{aligned}$$

From Theorem 3.1 we have that

$$T_7 + T_8 + T_9 + T_{10} \xrightarrow{\mathcal{D}} N(0, \xi' \Omega \xi),$$



where  $\mathbf{\Omega}$  is a  $(2 \times 2)$  matrix with elements

$$\begin{aligned}\omega_{i,i} &:= \frac{\gamma^{2t_i}(x)}{\Gamma^2(t_i+1)} [\Gamma(2t_i+1)\|K_i\|_2^2 + \Gamma^2(t_i+1)\|L_i\|_2^2 - 2\Gamma^2(t_i+1)\|K_iL_i\|_1], \quad i = 1, 2, \\ \omega_{1,2} &:= \frac{\gamma^{t_1+t_2}(x)}{\Gamma(t_1+1)\Gamma(t_2+1)} [\Gamma(t_1+t_2+1)\|K_1K_2\|_1 + \Gamma(t_1+1)\Gamma(t_2+1)\|L_1L_2\|_1 \\ &\quad - \Gamma(t_1+1)\Gamma(t_2+1)\|K_1L_2\|_1 - \Gamma(t_1+1)\Gamma(t_2+1)\|K_2L_1\|_1],\end{aligned}$$

and by Lemma 3.1 and 3.2

$$\begin{aligned}T_{11} &\rightarrow \xi_1\delta_1 := \xi_1\lambda \frac{\sqrt{g(x)}\gamma^{t_1-1}(x)}{\rho(x)} \left( \frac{1}{(1-\rho(x))^{t_1}} - 1 \right) \\ T_{12} &\rightarrow \xi_2\delta_2 := \xi_2\lambda \frac{\sqrt{g(x)}\gamma^{t_2-1}(x)}{\rho(x)} \left( \frac{1}{(1-\rho(x))^{t_2}} - 1 \right).\end{aligned}$$

Thus, we have established (3.9), with  $\boldsymbol{\delta}' = (\delta_1, \delta_2)$ , and, by the Cramér-Wold device, also (3.8). Now, denoting

$$\mathbf{D} := \begin{bmatrix} \gamma^{1-t_1}(x)/t_1 & 0 \\ 0 & \gamma^{1-t_2}(x)/t_2 \end{bmatrix},$$

a straightforward application of the  $\delta$ -method gives

$$r_n \begin{bmatrix} \widehat{\gamma}_n^{(1)}(x, t_1, K_1, L_1) - \gamma(x) \\ \widehat{\gamma}_n^{(1)}(x, t_2, K_2, L_2) - \gamma(x) \end{bmatrix} \xrightarrow{\mathcal{D}} N_2(\mathbf{D}\boldsymbol{\delta}, \mathbf{D}\mathbf{\Omega}\mathbf{D}),$$

which is the desired result.

### 3.5.7 Proof of Theorem 3.5

To establish the bivariate asymptotic normality we invoke the Cramér-Wold device, according to which it is sufficient that for all  $\boldsymbol{\xi} \in \mathbb{R}^2$  we have that

$$\boldsymbol{\xi}' r_n \begin{bmatrix} \widehat{\gamma}_n^{(2)}(x, t_1, K_1, L_1) - \gamma(x) \\ \widehat{\gamma}_n^{(2)}(x, t_2, K_2, L_2) - \gamma(x) \end{bmatrix} \xrightarrow{\mathcal{D}} N_2(\lambda \boldsymbol{\xi}' \boldsymbol{\mu}^{(2)}, \boldsymbol{\xi}' \widetilde{\boldsymbol{\Sigma}}^{(2)} \boldsymbol{\xi}).$$

First write

$$\begin{aligned} & \xi' r_n \begin{bmatrix} \widehat{\gamma}_n^{(2)}(x, t_1, K_1, L_1) - \gamma(x) \\ \widehat{\gamma}_n^{(2)}(x, t_2, K_2, L_2) - \gamma(x) \end{bmatrix} \\ &= \frac{\overline{F}(\omega_n; x)g(x)}{T_n^{(1, t_1)}(x, L_1)} \xi_1 r_n \frac{T_n^{(1, t_1+1)}(x, K_1) - \gamma(x)(t_1 + 1)T_n^{(1, t_1)}(x, L_1)}{(t_1 + 1)\overline{F}(\omega_n; x)g(x)} \\ &+ \frac{\overline{F}(\omega_n; x)g(x)}{T_n^{(1, t_2)}(x, L_2)} \xi_2 r_n \frac{T_n^{(1, t_2+1)}(x, K_2) - \gamma(x)(t_2 + 1)T_n^{(1, t_2)}(x, L_2)}{(t_2 + 1)\overline{F}(\omega_n; x)g(x)}. \end{aligned}$$

From Lemma 3.1 and 3.2, and Theorem 3.1 we have that

$$\frac{T_n^{(1, t_i)}(x, L_i)}{\overline{F}(\omega_n; x)g(x)} = \gamma^{t_i}(x)\Gamma(t_i + 1)(1 + o_{\mathbb{P}}(1)), \quad i = 1, 2,$$

and hence, by using the line of argumentation of Theorem 3.3, we obtain that

$$\begin{aligned} \xi' r_n \begin{bmatrix} \widehat{\gamma}_n^{(2)}(x, t_1, K_1, L_1) - \gamma(x) \\ \widehat{\gamma}_n^{(2)}(x, t_2, K_2, L_2) - \gamma(x) \end{bmatrix} &= \xi_1 r_n \frac{T_n^{(1, t_1+1)}(x, K_1) - \gamma(x)(t_1 + 1)T_n^{(1, t_1)}(x, L_1)}{\gamma^{t_1}(x)\Gamma(t_1 + 2)\overline{F}(\omega_n; x)g(x)} \\ &+ \xi_2 r_n \frac{T_n^{(1, t_2+1)}(x, K_2) - \gamma(x)(t_2 + 1)T_n^{(1, t_2)}(x, L_2)}{\gamma^{t_2}(x)\Gamma(t_2 + 2)\overline{F}(\omega_n; x)g(x)} \\ &+ o_{\mathbb{P}}(1), \end{aligned}$$

which can be rewritten as

$$\begin{aligned}
& \xi' r_n \begin{bmatrix} \widehat{\gamma}_n^{(2)}(x, t_1, K_1, L_1) - \gamma(x) \\ \widehat{\gamma}_n^{(2)}(x, t_2, K_2, L_2) - \gamma(x) \end{bmatrix} \\
&= \frac{\xi_1}{\gamma^{t_1}(x)\Gamma(t_1+2)} r_n \left[ \frac{T_n^{(1,t_1+1)}(x, K_1)}{\overline{F}(\omega_n; x)g(x)} - \mathbb{E} \left( \frac{T_n^{(1,t_1+1)}(x, K_1)}{\overline{F}(\omega_n; x)g(x)} \right) \right] \\
&\quad - \frac{\xi_1 \gamma(x)(t_1+1)}{\gamma^{t_1}(x)\Gamma(t_1+2)} r_n \left[ \frac{T_n^{(1,t_1)}(x, L_1)}{\overline{F}(\omega_n; x)g(x)} - \mathbb{E} \left( \frac{T_n^{(1,t_1)}(x, L_1)}{\overline{F}(\omega_n; x)g(x)} \right) \right] \\
&\quad + \frac{\xi_2}{\gamma^{t_2}(x)\Gamma(t_2+2)} r_n \left[ \frac{T_n^{(1,t_2+1)}(x, K_2)}{\overline{F}(\omega_n; x)g(x)} - \mathbb{E} \left( \frac{T_n^{(1,t_2+1)}(x, K_2)}{\overline{F}(\omega_n; x)g(x)} \right) \right] \\
&\quad - \frac{\xi_2 \gamma(x)(t_2+1)}{\gamma^{t_2}(x)\Gamma(t_2+2)} r_n \left[ \frac{T_n^{(1,t_2)}(x, L_2)}{\overline{F}(\omega_n; x)g(x)} - \mathbb{E} \left( \frac{T_n^{(1,t_2)}(x, L_2)}{\overline{F}(\omega_n; x)g(x)} \right) \right] \\
&\quad + \xi_1 r_n \frac{\mathbb{E}(T_n^{(1,t_1+1)}(x, K_1)) - \gamma(x)(t_1+1)\mathbb{E}(T_n^{(1,t_1)}(x, L_1))}{\gamma^{t_1}(x)\Gamma(t_1+2)\overline{F}(\omega_n; x)g(x)} \\
&\quad + \xi_2 r_n \frac{\mathbb{E}(T_n^{(1,t_2+1)}(x, K_2)) - \gamma(x)(t_2+1)\mathbb{E}(T_n^{(1,t_2)}(x, L_2))}{\gamma^{t_2}(x)\Gamma(t_2+2)\overline{F}(\omega_n; x)g(x)} + o_{\mathbb{P}}(1) \\
&=: T_{13} + T_{14} + T_{15} + T_{16} + T_{17} + T_{18} + o_{\mathbb{P}}(1).
\end{aligned}$$

From Theorem 3.1 we have that

$$T_{13} + T_{14} + T_{15} + T_{16} \xrightarrow{\mathcal{D}} N(0, \xi' \widetilde{\Sigma}^{(2)} \xi),$$

and, by Lemma 3.1 and 3.2,

$$T_{17} \rightarrow \xi_1 \lambda \frac{\sqrt{g(x)}}{(1-\rho(x))^{t_1+1}}, \quad \text{and} \quad T_{18} \rightarrow \xi_2 \lambda \frac{\sqrt{g(x)}}{(1-\rho(x))^{t_2+1}}.$$

By combining the above results we have that

$$\xi' r_n \begin{bmatrix} \widehat{\gamma}_n^{(2)}(x, t_1, K_1, L_1) - \gamma(x) \\ \widehat{\gamma}_n^{(2)}(x, t_2, K_2, L_2) - \gamma(x) \end{bmatrix} \xrightarrow{\mathcal{D}} N_2(\lambda \xi' \boldsymbol{\mu}^{(2)}, \xi' \widetilde{\Sigma}^{(2)} \xi),$$

from which the theorem follows.

### 3.5.8 Proof of Theorem 3.6

We only give the proof of the result for the case where  $\tau > 0$ . The case  $\tau = 0$  can be obtained in an analogous way and is therefore omitted.

Using Lemma 3.1 and 3.2, and Theorem 3.1, we obtain

$$\frac{T_n^{(1,1)}(x, K_1)}{T_n^{(1,0)}(x, L_1)} = \gamma(x) \left\{ 1 + \frac{b(\omega_n; x)}{\gamma(x)\rho(x)} \left[ \frac{1}{1 - \rho(x)} - 1 \right] (1 + o(1)) + O(h \ln \omega_n) + O_{\mathbb{P}} \left( \frac{1}{r_n} \right) \right\},$$

and, by Taylor's theorem,

$$\left( \frac{T_n^{(1,1)}(x, K_1)}{T_n^{(1,0)}(x, L_1)} \right)^{\tau} = \gamma^{\tau}(x) \left\{ 1 + \tau \frac{b(\omega_n; x)}{\gamma(x)\rho(x)} \left[ \frac{1}{1 - \rho(x)} - 1 \right] (1 + o(1)) + O(h \ln \omega_n) + O_{\mathbb{P}} \left( \frac{1}{r_n} \right) \right\}.$$

Similarly

$$\begin{aligned} \left( \frac{T_n^{(1,2)}(x, K_2)}{2T_n^{(1,0)}(x, L_2)} \right)^{\tau/2} &= \gamma^{\tau}(x) \left\{ 1 + \frac{\tau b(\omega_n; x)}{2\gamma(x)\rho(x)} \left[ \frac{1}{(1 - \rho(x))^2} - 1 \right] (1 + o(1)) \right. \\ &\quad \left. + O(h \ln \omega_n) + O_{\mathbb{P}} \left( \frac{1}{r_n} \right) \right\}, \\ \left( \frac{T_n^{(1,3)}(x, K_3)}{6T_n^{(1,0)}(x, L_3)} \right)^{\tau/3} &= \gamma^{\tau}(x) \left\{ 1 + \frac{\tau b(\omega_n; x)}{3\gamma(x)\rho(x)} \left[ \frac{1}{(1 - \rho(x))^3} - 1 \right] (1 + o(1)) \right. \\ &\quad \left. + O(h \ln \omega_n) + O_{\mathbb{P}} \left( \frac{1}{r_n} \right) \right\}. \end{aligned}$$

We now easily obtain that under the conditions of the theorem

$$\begin{aligned} \frac{\left( \frac{T_n^{(1,1)}(x, K_1)}{T_n^{(1,0)}(x, L_1)} \right)^{\tau} - \left( \frac{T_n^{(1,2)}(x, K_2)}{2T_n^{(1,0)}(x, L_2)} \right)^{\tau/2}}{b(\omega_n; x)} &\xrightarrow{\mathbb{P}} -\tau \gamma^{\tau-1}(x) \frac{\rho(x)}{2(1 - \rho(x))^2}, \\ \frac{\left( \frac{T_n^{(1,2)}(x, K_2)}{2T_n^{(1,0)}(x, L_2)} \right)^{\tau/2} - \left( \frac{T_n^{(1,3)}(x, K_3)}{6T_n^{(1,0)}(x, L_3)} \right)^{\tau/3}}{b(\omega_n; x)} &\xrightarrow{\mathbb{P}} -\tau \gamma^{\tau-1}(x) \frac{\rho(x)(3 - \rho(x))}{6(1 - \rho(x))^3}, \end{aligned}$$

and hence

$$R_n(x; \mathbb{K}, \mathbb{L}, \tau) \xrightarrow{\mathbb{P}} \frac{3(1 - \rho(x))}{3 - \rho(x)}.$$

Finally, the consistency of  $\widehat{\rho}_n(x; \mathbb{K}, \mathbb{L}, \tau)$  follows from a straightforward application of the continuous mapping theorem.

# Chapitre 4

## Madogram and asymptotic independence among maxima

### Abstract

A strong statistical research effort has been devoted in multivariate extreme value theory in order to assess the strength of dependence among extremes. This topic is particularly difficult in the case where block maxima are near independence. In this chapter, we adapt and study a simple inference tool inspired from geostatistics, the madogram, to the context of asymptotic independence between pairwise block maxima. In particular, we introduce an extremal coefficient and study the theoretical properties of its estimator. Its behaviour is also illustrated on a small simulation study and a real data set.

### 4.1 Introduction

One recurrent question in multivariate extreme value theory (MEVT) is how to infer the strength of dependence among maxima (e.g. de Haan and Ferreira, 2006). To illustrate this inquiry by an example, monthly maxima of hourly precipitation mea-

sured at two french stations from February 1987 to December 2002 are displayed in Figure 4.1. The two stations belong to the same hydrological basin of Orgeval (<https://gisoracle.cemagref.fr/>) that is located in France, west of Paris. For each season, a scatterplot between the two stations shows the original 45 (15 years  $\times$  3 months per season) monthly maxima recorded in millimeters. The dependence structure seems to vary with season and it is not clear if the largest summer values are closed to independence.

This concept of asymptotic independence has been studied by many authors. In this chapter, we follow the approach introduced by Ledford and Tawn (1996) and extended by Ramos and Ledford (2009). Before explaining the details of our method, we need to recall a few basic concepts about MEVT and to introduce some notations. Suppose that we have at our disposal  $n$  iid pairs  $(X_i, Y_i)$  with unit-Fréchet margins ( $\mathbb{P}(X_i \leq x) = \exp(-1/x)$  for  $x > 0$ ) and that the component-wise maxima vector  $(M_{X,n}, M_{Y,n}) = (\max(X_1, \dots, X_n), \max(Y_1, \dots, Y_n))$  converges in the following way

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \frac{M_{X,n}}{n} \leq x, \frac{M_{Y,n}}{n} \leq y \right) = G(x, y), \text{ for } x, y > 0. \quad (4.1)$$

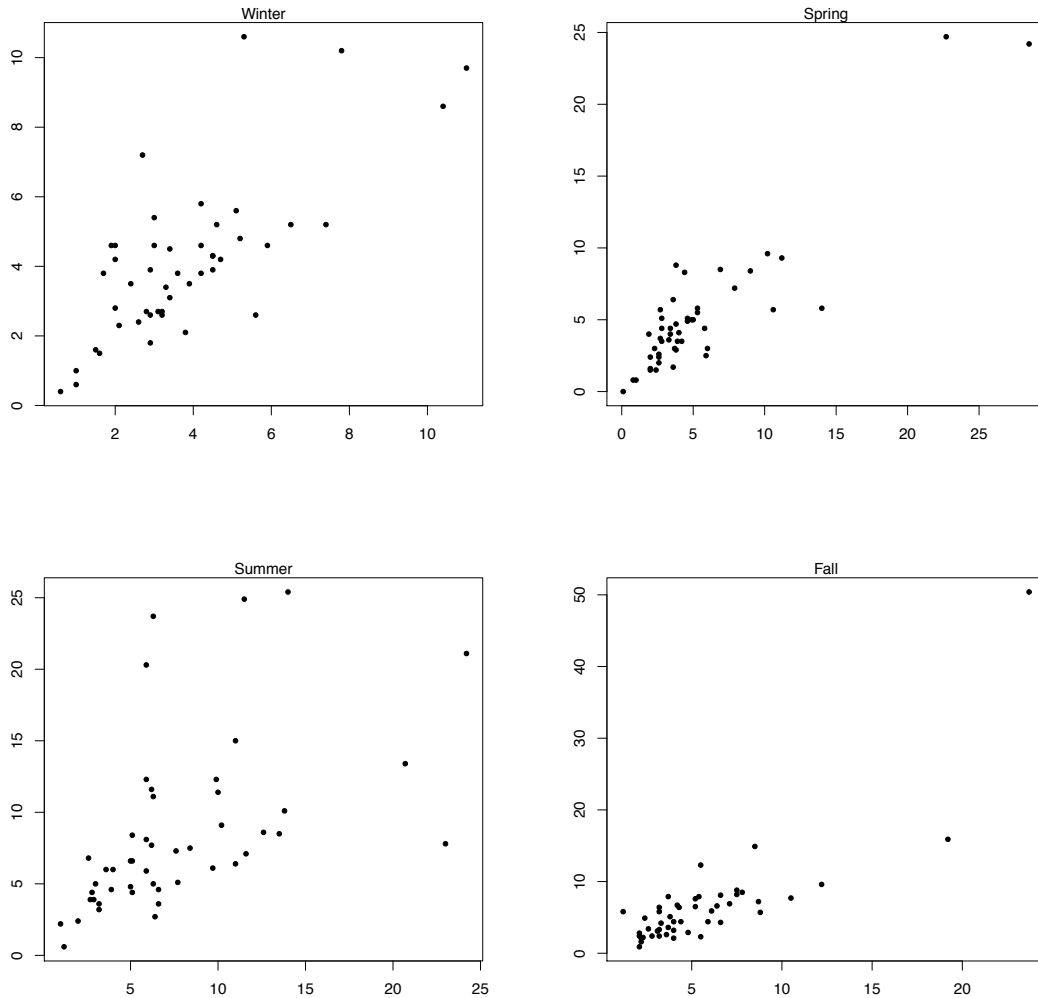


FIGURE 4.1 – Monthly maxima of hourly precipitation for each season, measured at two stations in the basin of Orgeval (near Paris) during 1987-2002.

The limiting distribution function  $G(.,.)$  is called the bivariate extreme value distribution and it can be written as  $G(x, y) = \exp\{-V(x, y)\}$ , with

$$V(x, y) = \int_0^1 \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right) dH(\omega),$$

where  $H(.)$  is a finite non-negative measure on  $[0, 1]$  such that  $\int_0^1 \omega dH(\omega) = \int_0^1 (1 - \omega) dH(\omega) = 1$ . This latter condition on  $H$  ensures that the margins  $G(x, \infty)$  and  $G(\infty, y)$



are unit-Fréchet distributed. The function  $V(.,.)$  is called the pairwise extremal dependence function. It is homogeneous of order  $-1$ , ie  $V(tx, ty) = t^{-1}V(x, y)$  for any positive  $t$  and  $G$  is max-stable, ie  $G^t(tx, ty) = G(x, y)$ . By definition of  $H(.,.)$ , the function  $V(.,.)$  has no explicit form and various non-parametric estimators of  $V$  have been studied (e.g. Capéraà *et al.*, 1997). As an example, an approach based on a classical geostatistical tool, the madogram (Matheron, 1987), was proposed by Naveau *et al.* (2009). Its simplest version (Cooley *et al.*, 2006) focused on the estimation of the extremal coefficient  $\theta = V(1, 1)$ . This coefficient provides a quick summary of the dependence between maxima. It belongs to the interval  $[1, 2]$ . If  $\theta$  equals two, the pairwise maxima are independent, and if  $\theta$  equals one, it is the complete dependence case. Cooley *et al.* (2006) defined the so-called  $F$ -madogram

$$\nu = \frac{1}{2} \mathbb{E} |F(M_{X,n}) - F(M_{Y,n})| \quad (4.2)$$

where  $F$  denotes the distribution function of  $M_{X,n}$  and  $M_{Y,n}$ , in order to express the extremal coefficient as

$$\theta = \frac{1 + 2\nu}{1 - 2\nu}. \quad (4.3)$$

Going back to the maxima displayed in Figure 4.1, one may wonder if convergence (4.1) provides an appropriate probabilistic framework to study the near independence seen in our summer rainfall data. Convergence (4.1) implies that  $\lim_{n \rightarrow \infty} n\mathbb{P}\left(\frac{X_i}{n} > x \text{ or } \frac{Y_i}{n} > y\right) = -\ln G(x, y)$ . Hence

$$\lim_{n \rightarrow \infty} n\mathbb{P}\left(\frac{X_i}{n} > x \text{ and } \frac{Y_i}{n} > y\right) = \ln G(x, y) - \ln G(x, \infty) - \ln G(\infty, y).$$

If we are in the asymptotically independent case, i.e.  $G(x, y)$  can be written as the product  $G(x, y) = G(x, \infty)G(\infty, y)$ , the right-hand side of the last convergence is nothing else than zero. This result is uninformative about the degree of independence among our rainfall maxima. A conceptual extension is needed to improve our understanding of the probability

of having joint extremes. To fill in this gap, Ledford and Tawn in a series of papers (1996, 1997, 1998) introduced a new tail tail model of the distribution which has been simplified by Ramos and Ledford (2009) as follows

$$\mathbb{P}(X > x, Y > y) = (xy)^{-\frac{1}{2\eta}} \mathcal{L}(x, y), \text{ for some } \eta \in (0, 1], \quad (4.4)$$

with  $\mathcal{L}$  a bivariate slowly varying function at infinity. The coefficient of tail dependence,  $\eta$ , is a measure of asymptotic independence. It is equal to one in the asymptotic dependence case and less than one in the asymptotic independence one. Condition (4.4) is tailored to analyze large excesses in the asymptotic independent case but it needs a reformulation in order to be used with pairs of maxima, as the ones pictured in Figure 4.1. This reformulation has been recently proposed by Ramos and Ledford (2011) who studied an extension of (4.1) by providing under the tail model (4.4) that, for  $x, y > 0$

$$\lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow \infty} \mathbb{P} \left[ \frac{M_{X,n,\varepsilon b_n}}{b_n} \leq x, \frac{M_{Y,n,\varepsilon b_n}}{b_n} \leq y \right] = G_\eta(x, y) = \exp \left[ -V_\eta(x, y) \right] \quad (4.5)$$

where the normalising constant  $b_n$  is defined implicitly as  $n\mathbb{P}(X > b_n, Y > b_n) = 1$ ,  $M_{\bullet,n,\varepsilon}$  corresponds to the component-wise maxima such that  $(X_i, Y_i)$  occur within the set  $R_\varepsilon = \{(x, y) : x > \varepsilon, y > \varepsilon\}$  and

$$V_\eta(x, y) = \eta \int_0^1 \left[ \max \left( \frac{\omega}{x}, \frac{1-\omega}{y} \right) \right]^{\frac{1}{\eta}} dH_\eta(\omega) \quad (4.6)$$

with  $H_\eta$  a finite and non-negative measure satisfying the constraint

$$\eta^{-1} = \int_0^{\frac{1}{2}} \omega^{\frac{1}{\eta}} dH_\eta(\omega) + \int_{\frac{1}{2}}^1 (1-\omega)^{\frac{1}{\eta}} dH_\eta(\omega).$$

The new dependence function  $V_\eta(\cdot, \cdot)$  is homogeneous of order  $-\frac{1}{\eta}$ ,

$$V_\eta(tx, ty) = t^{-\frac{1}{\eta}} V_\eta(x, y),$$

and the distribution  $G_\eta(x, y)$  obeys an extended max-stable property,

$$G_\eta^{t^{1/n}}(tx, ty) = G_\eta(x, y).$$

In (4.1), a normalisation of  $n^{-1}$  is required in order to stabilize the component-wise maxima, whereas in (4.5)  $b_n$  is of order  $O(n^\eta)$ .

The main goal of this chapter is to adapt the concept of madogram to this framework of asymptotic independence. The asymptotic properties of our estimators are also derived. A simple simulation study allows us to compare our inference scheme with the maximum likelihood estimation approach. All these estimators are applied to our rainfall data set.

## 4.2 The $F$ -madogram in the asymptotic independence case

Denote  $(M_X^*, M_Y^*)$  the bivariate vector that follows the distribution  $G_\eta(x, y)$ , i.e.

$$\mathbb{P}(M_X^* \leq x, M_Y^* \leq y) = \exp\{-V_\eta(x, y)\}, \quad (4.7)$$

with  $V_\eta(x, y)$  of the form (4.6).

Concerning the marginals, we denote

$$F_X^*(x) := \mathbb{P}(M_X^* \leq x) = \exp\left[-\sigma_X x^{-\frac{1}{\eta}}\right] \text{ and } F_Y^*(y) := \exp\left[-\sigma_Y y^{-\frac{1}{\eta}}\right] \quad (4.8)$$

with  $\sigma_X = V_\eta(1, \infty)$  and  $\sigma_Y = V_\eta(\infty, 1)$ . As the scaling coefficients  $\sigma_X$  and  $\sigma_Y$  are not necessarily equal, the Fréchet margins of  $M_X^*$  and  $M_Y^*$  differs by a multiplicative factor. In the classical MEVT setup defined by (4.1), the extremal coefficient  $\theta = V(1, 1)$  was simple to explain. It always varied between one (dependence) and two (independence). Having different marginals in (4.8) makes it difficult to find simple and interpretable summaries like the extremal coefficient. One possible way around this interpretability issue is to go

back to the madogram distance because it is trivial to interpret as a metric and it is marginal free. The  $F$ -madogram for the pair  $(M_X^*, M_Y^*)$  can be defined as

$$\nu_\eta = \frac{1}{2} \mathbb{E} |F_X^*(M_X^*) - F_Y^*(M_Y^*)|, \quad (4.9)$$

and we can derive from (4.6) and (4.8) the relationship (see Section 4.5)

$$\theta_\eta = \frac{1 + 2\nu_\eta}{1 - 2\nu_\eta}, \quad (4.10)$$

where  $\theta_\eta := V_\eta(\sigma_X^\eta, \sigma_Y^\eta)$  could be viewed as an analog of the classical extremal coefficient, compare equations (4.3) and (4.10). If  $\nu_\eta$  equals zero, then  $\theta_\eta$  equals one. As the distance  $\nu_\eta$  increases, the coefficient  $\theta_\eta$  also increases. If  $M_X^*$  and  $M_Y^*$  are independent, then  $F_X^*(M_X^*)$  and  $F_Y^*(M_Y^*)$  are independent and uniformly distributed random variables. It follows that  $\nu_\eta = 1/6$ . From (4.10), we deduce  $\theta_\eta = 2$ .

The only difference between equations (4.3) and (4.10) resides in the fact that the pairwise maxima vector belongs now to the larger family  $G_\eta(\cdot, \cdot)$  instead of the classical  $G(\cdot, \cdot)$ . It is also essential to emphasize that the  $F$ -madogram should not be interpreted alone. The coefficient  $\eta$  is paramount to explore the asymptotic independence domain.

## 4.3 Inference

### 4.3.1 A method-of-moment approach

Our main result is the following theorem that deals with the convergence of the empirical estimator deduced from Equation (4.9).

**Theorem 4.1.** *Let  $(M_{X_i,n}^*, M_{Y_i,n}^*)$  be a sample of  $N$  bivariate maxima vectors of block size  $n$  that converges in distribution, as  $n \rightarrow \infty$ , to a bivariate extreme value distribution*

with an  $\eta$ -dependence function defined as in (4.6). Let

$$\widehat{\nu}_\eta = \frac{1}{2N} \sum_{i=1}^N \left| \widehat{F}_X^*(M_{X_i,n}^*) - \widehat{F}_Y^*(M_{Y_i,n}^*) \right|, \quad (4.11)$$

where  $\widehat{F}_X^*$ , resp.  $\widehat{F}_Y^*$ , denotes the empirical distribution function of the sample  $M_{X_i,n}^*$ , resp.  $M_{Y_i,n}^*$ . Then, as  $n \rightarrow \infty$  and  $N \rightarrow \infty$

$$\sqrt{N}(\widehat{\nu}_\eta - \nu_\eta) \xrightarrow{d} \int_{[0,1]^2} N_C(u, v) dJ(u, v)$$

where  $J(x, y) = \frac{1}{2}|x - y|$  and  $N_C$  is a Gaussian process defined by

$$N_C(u, v) = B_C(u, v) - B_C(u, 1) \frac{\partial C}{\partial u}(u, v) - B_C(1, v) \frac{\partial C}{\partial v}(u, v), \quad (4.12)$$

and  $B_C$  is a Brownian bridge on  $[0, 1]^2$  with covariance function

$$\mathbb{E}\{B_C(u, v) \cdot B_C(u', v')\} = C(u \wedge u', v \wedge v') - C(u, v) \cdot C(u', v'),$$

with  $u \wedge u' = \min(u, u')$  and  $C$  the copula function with respect to (4.6).

From (4.10), we introduce the following estimator for the extremal coefficient

$$\widehat{\theta}_\eta = \frac{1 + 2\widehat{\nu}_\eta}{1 - 2\widehat{\nu}_\eta}. \quad (4.13)$$

Applying the delta method, the following corollary follows.

**Corollary 4.1.** *Under the assumption of Theorem 4.1, we have*

$$\sqrt{N}(\widehat{\theta}_\eta - \theta_\eta) \xrightarrow{d} (1 + \theta_\eta)^2 \int_{[0,1]^2} N_C(u, v) dJ(u, v).$$

To infer the value of  $\eta$ , we complement our method-of-moment via a Generalized Probability Weighted Moment (GPWM) approach (Diebolt *et al.*, 2008b) based on the following

moment equality

$$\mu_\omega = \mathbb{E}(M\omega(F)) = \int_{-\infty}^{+\infty} x \omega(F(x)) dF(x),$$

for any variable  $M$  with a distribution function  $F$  and  $\omega$  a suitable continuous function satisfying

$$\begin{cases} \omega(t) = O((1-t)^b) & \text{for } t \text{ close to } 1, b \geq 0 \\ \omega(t) = O(t^{a'}) & \text{for } t \text{ close to } 0, a' > 0. \end{cases} \quad (4.14)$$

Set  $M^* = \max(M_{X,n}^*, M_{Y,n}^*)$ . Its distribution function equals  $F_{M^*}(x) = \exp\{-V_\eta(1,1)x^{-\frac{1}{\eta}}\}$  (see Equation (4.7)), i.e. a  $\text{GEV}_{\eta, \eta V_\eta^\eta(1,1), V_\eta^\eta(1,1)}$  distribution. Thus, if  $\omega_{a,b}(t) = t^a(-\ln t)^b, a > a'$ , Diebolt *et al.* (2008b) proved that

$$\mu_{a,b} := \mu_\omega = \frac{V_\eta^\eta(1,1)}{(a+1)^{b-\eta+1}} \Gamma(b-\eta+1).$$

A natural estimator for  $\mu_{a,b}$  can be obtained by replacing  $F$  by its empirical version  $\mathbb{F}_n$ ,

$$\hat{\mu}_{a,b} = \int_0^1 \mathbb{F}_n^{-1}(u) u^a (-\ln u)^b du.$$

Selecting suitable values for  $(a, b)$  allows us to obtain an estimator for  $\eta$  in function of  $\hat{\mu}_{a,b}$

$$\hat{\eta}_{gpwm} = 2 \left( 1 - \frac{\hat{\mu}_{1,2}}{\hat{\mu}_{1,1}} \right). \quad (4.15)$$

The asymptotic normality of  $\hat{\eta}_{gpwm}$  can then be deduced from the asymptotic properties of the GPWM estimators, see Section 4.5.

**Proposition 4.1** *Let  $(M_{X_i,n}^*, M_{Y_i,n}^*)$  be a sample of  $N$  bivariate maxima vectors of block size  $n$  that follows a bivariate extreme value distribution with an  $\eta$ -dependence function defined as in (4.6). Then the GPWM estimator for  $\eta$  defined by  $\hat{\eta}_{gpwm}$  converges in the following way*

$$\sqrt{N} (\hat{\eta}_{gpwm} - \eta) \xrightarrow{d} \frac{\eta 2^{3-\eta}}{\Gamma(2-\eta)} [I_1 - (1-\eta/2)I_2]$$

with  $I_1 = \int_0^1 B(t)(-\ln t)^{-\eta+1} dt$ ,  $I_2 = \int_0^1 B(t)(-\ln t)^{-\eta} dt$  and  $B$  a Brownian bridge.

### 4.3.2 The maximum likelihood approach

Besides our aforementioned method-of-moment approach, a Maximum Likelihood (ML) method can also be implemented. Our ML method is based on the normalized sample  $\{M_i\}_{i=1,\dots,N} = \left\{ \max \left( \frac{M_{X_i,n}^*}{\sigma_X^\eta}, \frac{M_{Y_i,n}^*}{\sigma_Y^\eta} \right) \right\}_{i=1,\dots,N}$  which admits the following log-likelihood

$$\ln L(M_1, \dots, M_N; \theta_\eta, \eta) = N \ln \left( \frac{\theta_\eta}{\eta} \right) - (1/\eta + 1) \sum_{i=1}^N \ln(M_i) - \theta_\eta \sum_{i=1}^N M_i^{-1/\eta}.$$

If  $\hat{\eta}_{mle}$  denotes the ML estimator for  $\eta$  based on the univariate sample  $\{\max(M_{X_i,n}^*, M_{Y_i,n}^*)\}$  with a distribution function given by (4.7), it allows us to write

$$\hat{\theta}_{\eta,mle} = \left[ \frac{1}{N} \sum_{i=1}^N \min \left( \hat{\sigma}_{X,mle} (M_{X_i,n}^*)^{-1/\hat{\eta}_{mle}}, \hat{\sigma}_{Y,mle} (M_{Y_i,n}^*)^{-1/\hat{\eta}_{mle}} \right) \right]^{-1}.$$

The estimates for  $\sigma_X$  and  $\sigma_Y$  in the above equality can be derived from (4.8) as  $\hat{\sigma}_{X,mle} = \left[ \frac{1}{N} \sum_{i=1}^N (M_{X_i,n}^*)^{-1/\hat{\eta}_{mle}} \right]^{-1}$  and a similar expression for  $\hat{\sigma}_{Y,mle}$ . Thus we can define

$$\hat{\theta}_{\eta,mle} = \left[ \frac{1}{N} \sum_{i=1}^N \min \left( \frac{(M_{X_i,n}^*)^{-1/\hat{\eta}_{mle}}}{\frac{1}{N} \sum_{j=1}^N (M_{X_j,n}^*)^{-1/\hat{\eta}_{mle}}}, \frac{(M_{Y_i,n}^*)^{-1/\hat{\eta}_{mle}}}{\frac{1}{N} \sum_{j=1}^N (M_{Y_j,n}^*)^{-1/\hat{\eta}_{mle}}} \right) \right]^{-1}. \quad (4.16)$$

## 4.4 Examples

### 4.4.1 A small simulation

To compare our estimators with the ML approach, we simulate 300 samples of 500 pairs of maxima from the  $\eta$ -asymmetric logistic dependence model (see Ramos and Ledford, 2011)

$$V_\eta(x, y) = \frac{1}{2 - 2^{\alpha/\eta}} (x^{-1/\alpha} + y^{-1/\alpha})^{\alpha/\eta}, \quad \text{for } x, y > 0,$$

with  $\alpha \in \{0.1, 0.3, 0.5, 0.6\}$  and  $\eta = 0.7$ . This specific value of  $\eta$  corresponds a case of asymptotic independence ( $\eta < 1$ ) and provides  $\theta_\eta = 2^{\alpha/\eta}$ .

The boxplots in Figure 4.2 indicate that the estimate  $\hat{\theta}_{\eta,mle}$  from (4.16) has a higher

variability than  $\hat{\theta}_\eta$  from (4.13). This is particularly true when  $\alpha$  is close to  $\eta$ , i.e.  $\hat{\theta}_\eta$  near two.

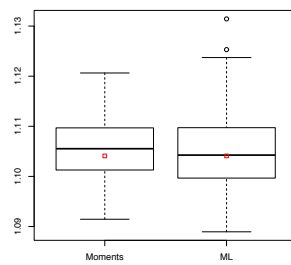
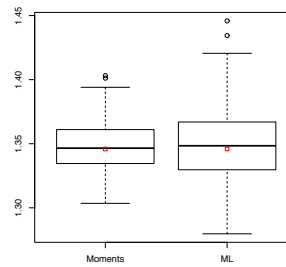
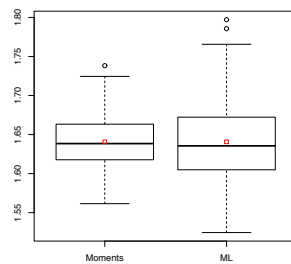
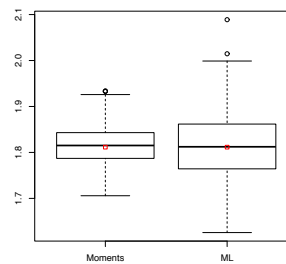
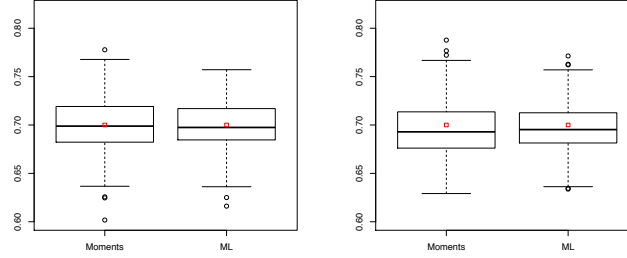
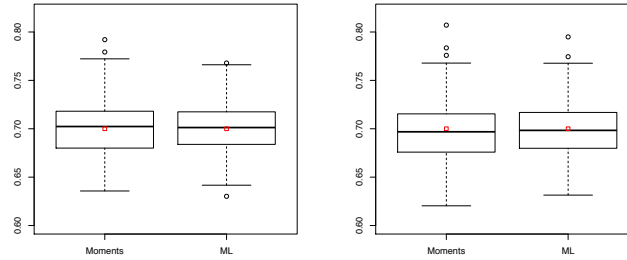
(a) Case  $\alpha = 0.1, \eta = 0.7$ (b) Case  $\alpha = 0.3, \eta = 0.7$ (c) Case  $\alpha = 0.5, \eta = 0.7$ (d) Case  $\alpha = 0.6, \eta = 0.7$ 

FIGURE 4.2 – Simulation : comparing  $\hat{\theta}_\eta$  from (4.13) with  $\hat{\theta}_{\eta,mle}$  from (4.16)

Concerning the estimation of  $\eta$ , Figure 4.3 basically tells the opposite story. The ML approach appears slightly better than a method-of-moment. This small simulation study advocates for not restricting one inference approach but rather to combine or at least compare different inference techniques.



(a) Case  $\alpha = 0.1, \eta = 0.7$ (b) Case  $\alpha = 0.3, \eta = 0.7$ (c) Case  $\alpha = 0.5, \eta = 0.7$ (d) Case  $\alpha = 0.6, \eta = 0.7$ FIGURE 4.3 – Simulation : comparing  $\hat{\eta}_{gpwm}$  from (4.15) with  $\hat{\eta}_{mle}$ .

#### 4.4.2 Orgeval Rainfall data

The table below summarizes our inference with respect to the maxima plotted in Figure 4.1.

	$\hat{\theta}_\eta$	$\hat{\theta}_{\eta,mle}$	$\hat{\eta}_{gpwm}$	$\hat{\eta}_{mle}$
winter	1.44	1.26	0.44	0.71
spring	1.33	1.22	0.50	0.70
summer	1.45	1.47	0.56	0.72
fall	1.36	1.60	0.49	0.51

TABLE 4.1 – Estimates with GPWM and ML approaches for the Orgeval rainfall data

The estimates of  $\eta$  are all below one. This suggests asymptotic independence for all seasons. The parameter  $\theta_\eta$  gives a second order measure of the dependence inside asymptotic independence : the estimates  $\hat{\theta}_\eta$  obtained by the madogram are similar for all seasons with

a maximal value in summer. This corroborates the particular pattern noticed in Figure 4.1. This study is based on a small sample size and general conclusions about rainfall cannot be made here.

## 4.5 Proof

**Proof of (4.10).** Applying the equality  $|a - b|/2 = \max(a, b) - (a + b)/2$  to  $\nu_\eta$ , we get :

$$\frac{1}{2}\mathbb{E} |F_X^*(M_X^*) - F_Y^*(M_Y^*)| = \mathbb{E} \max\{F_X^*(M_X^*), F_Y^*(M_Y^*)\} - \frac{1}{2}.$$

Then we calculate

$$\begin{aligned} \mathbb{P}[\max\{F_X^*(M_X^*), F_Y^*(M_Y^*)\} \leq u] &= \mathbb{P}[M_X^* \leq F_X^{*\leftarrow}(u), M_Y^* \leq F_Y^{*\leftarrow}(u)] \\ &= \exp\{-V_\eta(F_X^{*\leftarrow}(u), F_Y^{*\leftarrow}(u))\} \\ &= \exp\{\ln(u)V_\eta(\sigma_X^\eta, \sigma_Y^\eta)\} = u^{V_\eta(\sigma_X^\eta, \sigma_Y^\eta)} \end{aligned}$$

because from the margin model (4.8) we have  $F_X^{*\leftarrow}(u) = (-\ln(u)/\sigma_X)^\eta$  and  $F_Y^{*\leftarrow}(u) = (-\ln(u)/\sigma_Y)^\eta$ . Therefore,  $\mathbb{E} \max\{F_X^*(M_X^*), F_Y^*(M_Y^*)\} = \frac{V_\eta(\sigma_X^\eta, \sigma_Y^\eta)}{1+V_\eta(\sigma_X^\eta, \sigma_Y^\eta)}$  from which (4.10) follows.

**Proof of Theorem 4.1.** First, we introduce the ‘normalized’ empirical distribution functions

$$\tilde{F}_{n,N,X}(u) := \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{[b_n^{-1}M_{X_i,n}^* \leq u]}, \quad \tilde{F}_{n,N,Y}(u) := \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{[b_n^{-1}M_{Y_i,n}^* \leq u]},$$

and we rewrite the estimator of the madogram as

$$\hat{\nu}_\eta = \frac{1}{2N} \sum_{i=1}^N \left| \tilde{F}_{n,N,X}(b_n^{-1}M_{X_i,n}^*) - \tilde{F}_{n,N,Y}(b_n^{-1}M_{Y_i,n}^*) \right|.$$

Before starting the proof, we need to introduce a series of definitions linked to the copula function  $C$ . Although very similar, these definitions represent slightly different estimators of the same copula function. One difficulty of the proof is to show how close these versions

are :

$$\begin{aligned}\tilde{C}_{n,N}(u, v) &:= \frac{1}{N} \sum_{i=1}^N \mathbb{1} \left\{ \tilde{F}_{n,N,X}(b_n^{-1} M_{X_i,n}^*) \leq u, \tilde{F}_{n,N,Y}(b_n^{-1} M_{Y_i,n}^*) \leq v \right\}, \\ C_{n,N}(u, v) &:= \frac{1}{N} \sum_{i=1}^N \mathbb{1} \left\{ b_n^{-1} M_{X_i,n}^* \leq \tilde{F}_{n,N,X}^{\leftarrow}(u), b_n^{-1} M_{Y_i,n}^* \leq \tilde{F}_{n,N,Y}^{\leftarrow}(v) \right\}, \\ \tilde{C}_{n,N}^*(u, v) &:= \frac{1}{N} \sum_{i=1}^N \mathbb{1} \left\{ U_{X_i,n} \leq \tilde{F}_{n,N,X}^{\leftarrow}(u), V_{Y_i,n} \leq \tilde{F}_{n,N,Y}^{\leftarrow}(v) \right\}\end{aligned}$$

where

$$\begin{aligned}\tilde{F}_{X,n}(z) &:= \mathbb{P}(b_n^{-1} M_{X_i,n}^* \leq z), \quad \tilde{F}_{Y,n}(z) := \mathbb{P}(b_n^{-1} M_{Y_i,n}^* \leq z), \\ U_{X_i,n} &:= \tilde{F}_{X,n}(b_n^{-1} M_{X_i,n}^*), \quad V_{Y_i,n} := \tilde{F}_{Y,n}(b_n^{-1} M_{Y_i,n}^*), \\ \tilde{F}_{n,N,X}^*(u) &:= \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\{U_{X_i,n} \leq u\}}, \quad \tilde{F}_{n,N,Y}^*(v) := \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\{V_{Y_i,n} \leq v\}}.\end{aligned}$$

The proof of Theorem 4.1 is divided into the following five steps.

*Step 1.* The function  $\tilde{C}_{n,N}(u, v)$  is very similar to  $C_{n,N}(u, v)$ , i.e.  $\sup_{0 \leq u, v \leq 1} |\tilde{C}_{n,N}(u, v) - C_{n,N}(u, v)| \leq 2/N$ .

*Step 2.* We have  $C_{n,N}(u, v) = \tilde{C}_{n,N}^*(u, v)$ .

*Step 3.* Define now the empirical distribution function of  $(U_{X_i,n}, V_{Y_i,n})$  as

$$\tilde{H}_{n,N}^*(u, v) = \frac{1}{N} \sum_{i=1}^N \mathbb{1} \left\{ U_{X_i,n} \leq u, V_{Y_i,n} \leq v \right\}$$

and its non-empirical version as

$$\tilde{H}_n^*(u, v) = \mathbb{P} \left( \tilde{F}_{X,n}(b_n^{-1} M_{X_i,n}^*) \leq u, \tilde{F}_{Y,n}(b_n^{-1} M_{Y_i,n}^*) \leq v \right).$$

We establish that the process  $\sqrt{N}(\tilde{H}_{n,N}^* - \tilde{H}_n^*)$  tends in distribution to a Brownian bridge  $B_C$ . To this end, we prove the convergence of the finite-dimensional distributions and the tightness of the process.

*Step 4.* The process  $\sqrt{N}(\tilde{C}_{n,N}^* - \tilde{H}_n^*)$  tends in distribution to a Gaussian process  $N_C$ .

*Step 5.* We conclude the proof of our theorem using the integration by parts.

This proof is sketched here as it is a slightly modified version of the one of Proposition 4 in Naveau *et al.* (2009) which is detailed in <http://sama.ipsl.jussieu.fr/Documents/articles/NaveauBiometrika07DetailedProofs.pdf>.

**Remark about our Theorem 4.1.** The limiting process is such that

$$\int_{[0,1]^2} N_C(u, v) dJ(u, v) = \frac{1}{2} \int_0^1 N_C(0, v) dv + \frac{1}{2} \int_0^1 N_C(u, 0) du - \int_0^1 N_C(u, u) du.$$

This limiting process cannot be precised without specifying the copula function and in special cases where these integrals can be computed. For instance, consider the Product copula, also called the independent copula, defined as  $C(u, v) = uv$ . In that case

$$N_C(u, v) = B_C(u, v) - vB_C(u, 1) - uB_C(1, v),$$

from which direct computations lead to

$$\text{Var} \left( \int_{[0,1]^2} N_C(u, v) dJ(u, v) \right) = \frac{1}{90}.$$

**Proof of Proposition 4.1.** As  $\eta \in (0, 1]$ , we have according to Theorem 2.1 in Diebolt *et al.* (2008b) that

$$\sqrt{N} \begin{pmatrix} \hat{\mu}_{1,1} - \mu_{1,1} \\ \hat{\mu}_{1,2} - \mu_{1,2} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \eta V_\eta^\eta(1, 1) \int_0^1 \frac{B(t)}{t} (-\ln t)^{-\eta-1} t(-\ln t) dt \\ \eta V_\eta^\eta(1, 1) \int_0^1 \frac{B(t)}{t} (-\ln t)^{-\eta-1} t(-\ln t)^2 dt \end{pmatrix} \quad (4.17)$$

where  $B$  denotes a Brownian bridge and  $n \rightarrow \infty$ . It follows

$$\begin{aligned} \sqrt{N} (\hat{\eta}_{gpwm} - \eta) &= -\frac{2}{\hat{\mu}_{1,1}\mu_{1,1}} \sqrt{N} (\mu_{1,1}\hat{\mu}_{1,2} - \mu_{1,2}\hat{\mu}_{1,1}) \\ &= -\frac{2}{\hat{\mu}_{1,1}\mu_{1,1}} \left[ \mu_{1,1}\sqrt{N}(\hat{\mu}_{1,2} - \mu_{1,2}) - \mu_{1,2}\sqrt{N}(\hat{\mu}_{1,1} - \mu_{1,1}) \right]. \end{aligned}$$

An application of Slutsky's theorem leads to

$$\sqrt{N} (\widehat{\eta}_{gpwm} - \eta) \xrightarrow{d} -\frac{2\eta V_{\eta}^{\eta}(1, 1)}{\mu_{1,1}^2} \int_0^1 \frac{B(t)}{t} (-\ln t)^{-\eta-1} [\mu_{1,1}\omega_{1,2}(t) - \mu_{1,2}\omega_{1,1}(t)] dt,$$

from which Proposition 4.1 follows.

# Conclusions et Perspectives

Nous avons souhaité à travers cette thèse, aborder un certain nombre de sujets peu traités jusqu'à présent dans la littérature concernant la théorie des valeurs extrêmes. Pour cela, nous avons rappelé, dans le premier chapitre, les résultats et outils essentiels à la compréhension de nos contributions.

Puis, dans les chapitres 2 et 3, nous avons considéré le cas où la variable d'intérêt  $Y$  est observée simultanément avec une covariable, pouvant être fixe (chapitre 2) ou aléatoire (chapitre 3). Dans ce contexte, l'indice de queue dépend de la covariable et nous avons proposé des estimateurs de ce paramètre dont nous avons étudié les propriétés asymptotiques. Leurs comportements à distance finie ont été validés par simulations.

Enfin, dans le chapitre 4, nous nous sommes intéressés aux extrêmes multivariés et plus particulièrement à mesurer la dépendance entre les extrêmes. Dans une situation proche de l'indépendance asymptotique, il est très difficile de mesurer cette dépendance et de nouveaux modèles doivent être introduits. Dans ce contexte, nous avons adapté un outil de géostatistique, le madogramme, et nous avons étudié ses propriétés asymptotiques. Ses performances sur simulations et données réelles ont également été exhibées.

Parmi ces différents travaux de recherche, certains sont motivés par une problématique appliquée, d'autres sont plus à caractère mathématique, mais tous ont en commun de

soulever des perspectives que nous envisageons d'ores et déjà d'explorer. Une liste non exhaustive est donnée ci-dessous :

- Les estimateurs que nous avons proposés requièrent le choix de plusieurs paramètres : noyau, fenêtre, paramètre de second ordre. Leur estimation ou leur choix n'est pas toujours un problème simple. Il serait intéressant de développer une méthode rigoureuse de sélection et le cas échéant d'établir la normalité asymptotique des estimateurs proposés (par exemple pour le paramètre de second ordre) ;
- L'analyse des valeurs extrêmes en présence de covariable à été traitée principalement dans le cas de distributions appartenant au domaine de Fréchet. En vu d'applications pratiques, il serait bon de pouvoir étendre cette approche aux autres domaines d'attraction et également à l'estimation de quantités d'intérêt autres que l'indice de queue conditionnel, par exemple l'estimation de quantiles extrêmes ou de petites probabilités ;
- Bien que des efforts considérables aient été dévolus au développement d'estimateurs pour des paramètres / fonctions décrivant des dépendances de queue, peu d'attention a été consacrée à la réduction de biais dans un cadre multivarié. À notre connaissance, les seuls papiers existants sur le sujet sont ceux de Beirlant *et al.* (2011) et Goegebeur et Guillou (2012) qui, comme dans le cas univarié, font appel à une condition du second ordre faisant intervenir de nouveaux paramètres inconnus à estimer. L'estimation de ces paramètres de second ordre est un problème complètement inexploré dans le cadre multivarié ;
- Comme mentionné dans le Chapitre 4, dans le cas où les marges sont les mêmes ( $F^* := F_X^* = F_Y^*$ ),  $\theta_\eta$  donne uniquement de l'information sur  $V_\eta(1, 1)$  mais pas la moindre sur  $V_\eta(u, v)$  pour  $u \neq v$ . Pour résoudre ce problème, une idée naturelle serait d'utiliser le

$\lambda$ -madogramme introduit par Naveau *et al.* (2009) :

$$\begin{aligned} \nu_\eta(\lambda) &= \frac{1}{2} \mathbb{E} |F^{*\lambda}(M_X^*) - F^{*1-\lambda}(M_Y^*)| \\ &= \frac{\frac{V_\eta(\lambda^\eta, (1-\lambda)^\eta)}{\sigma}}{1 + \frac{V_\eta(\lambda^\eta, (1-\lambda)^\eta)}{\sigma}} - \frac{3}{2(1+\lambda)(2-\lambda)} \end{aligned}$$

où  $\sigma := V_\eta(1, \infty) = V_\eta(\infty, 1)$ .

Ceci conduit donc au nouveau coefficient extremal :

$$\theta_\eta(\lambda) := \frac{V_\eta(\lambda^\eta, (1-\lambda)^\eta)}{\sigma}.$$

La procédure d'inférence développée dans le Chapitre 4 pour  $\nu_\eta$  peut alors être utilisée pour  $\nu_\eta(\lambda)$  et une normalité asymptotique similaire à celle du Théorème 4.1 peut être obtenue. Malheureusement cela ne permet que d'estimer la fonction  $V_\eta$  au point  $(\lambda^\eta, (1-\lambda)^\eta)$  avec  $\eta$  inconnu. Pour résoudre ce problème, une petite modification du  $\lambda$ -madogramme peut être introduite

$$\begin{aligned} \nu_{\eta,\lambda} &= \frac{1}{2} \mathbb{E} \left| F^{*\lambda^{1/\eta}}(M_X^*) - F^{*(1-\lambda)^{1/\eta}}(M_Y^*) \right| \\ &= \frac{\theta_{\eta,\lambda}}{1 + \theta_{\eta,\lambda}} - \frac{1}{2(1 + \lambda^{1/\eta})} - \frac{1}{2(1 + (1-\lambda)^{1/\eta})} \end{aligned}$$

permettant ainsi l'utilisation d'un nouveau coefficient

$$\theta_{\eta,\lambda} := \frac{V_\eta(\lambda, 1-\lambda)}{\sigma}.$$

Si  $\hat{\eta}$  est un estimateur consistant de  $\eta$  (par exemple celui obtenu par GPWM), nous pouvons alors estimer  $\nu_{\eta,\lambda}$  par

$$\hat{\nu}_{\eta,\lambda} = \frac{1}{2N} \sum_{i=1}^N \left| \hat{F}_X^{*\lambda^{1/\hat{\eta}}}(M_{X_i,n}^*) - \hat{F}_Y^{*(1-\lambda)^{1/\hat{\eta}}}(M_{Y_i,n}^*) \right|$$

et déduire ainsi l'estimateur suivant du coefficient extremal

$$\hat{\theta}_{\eta,\lambda} = \frac{\hat{\nu}_{\eta,\lambda} + \frac{1}{2(1+\lambda^{1/\hat{\eta}})} + \frac{1}{2(1+(1-\lambda)^{1/\hat{\eta}})}}{1 - \hat{\nu}_{\eta,\lambda} - \frac{1}{2(1+\lambda^{1/\hat{\eta}})} - \frac{1}{2(1+(1-\lambda)^{1/\hat{\eta}})}}.$$



Cette version du madogramme nous permet d'estimer la fonction complète  $V_\eta$ . Toutefois la normalité asymptotique de  $\hat{\theta}_{\eta,\lambda}$  est très complexe à obtenir.

# Bibliographie

- Balkema, A. & de Haan, L. (1974). Residual life time at great age, *Annals of Probability*, **2**, 792-804.
- Beirlant, J., Dierckx, G., Goegebeur, Y. & Matthys, G. (1999). Tail index estimation and an exponential regression model, *Extremes*, **2**, 177–200.
- Beirlant, J., Dierckx, G. & Guillou, A. (2011). Bias-reduced estimators for bivariate tail modelling, *Insurance : Mathematics and Economics*, **49**, 18–26.
- Beirlant, J., Dierckx, G., Guillou, A. & Stărică, C. (2002). On exponential representations of log-spacings of extreme order statistics, *Extremes*, **5**, 157–180.
- Beirlant, J. & Goegebeur, Y. (2003). Regression with response distributions of Pareto-type, *Computational Statistics and Data Analysis*, **42**, 595–619.
- Beirlant, J. & Goegebeur, Y. (2004). Local polynomial maximum likelihood estimation for Pareto-type distributions, *Journal of Multivariate Analysis*, **89**, 97–118.
- Beirlant, J., Goegebeur, Y., Segers, J. & Teugels, J. (2004). *Statistics of Extremes - Theory and Applications*, Wiley Series in Probability and Statistics.
- Beirlant, J. & Teugels, J.L. (1989). Asymptotic normality of Hill's estimator, In *Extreme Value Theory, Oberwolfach 1987*, (J. Hüsler and R.D. Reiss, Eds.), Springer-Verlag, New York.
- Billingsley, P. (1995). *Probability and Measure*, Wiley Series in Probability and Mathematical Statistics.
- Bingham, N.H., Goldie, C.M. & Teugels, J.L. (1987). *Regular Variation*, Cambridge University Press, Cambridge.
- Caeiro, F., Gomes, M.I. & Rodrigues, L.H. (2009). Reduced-bias tail index estimators under a third order framework, *Communications in Statistics – Theory and Methods*, **38**, 1019–1040.
- Capéraà, P., Fougères, A.-L. & Genest, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas, *Biometrika*, **84**, 567–577.
- Chavez-Demoulin, V. & Davison, A.C. (2005). Generalized additive modelling of sample

- extremes, *Journal of the Royal Statistical Society Series C*, **54**, 207–222.
- Ciuperca, G. & Mercadier, C. (2010). Semi-parametric estimation for heavy tailed distributions, *Extremes*, **13**, 55–87.
- Coles, S.G., Hefferman, J.E. & Tawn, J.A. (2000). Dependence measures for extreme value analysis, *Extremes*, **2**, 339–365.
- Cooley, D., Naveau, P. & Poncet, P. (2006). Variograms for spatial max-stable random fields, In *Dependence in Probability and Statistics*, Lecture Notes In Statistics, Vol. 187, Ed. P. Bertail, P. Doukhan and P. Soulier, pp. 373–90. New York : Springer.
- Csörgő, S., Deheuvels, P. & Mason, D.M. (1985). Kernel estimates of the tail index of a distribution, *Annals of Statistics*, **13**, 1050–1077.
- Daouia, A., Gardes, L. & Girard, S. (2012). On kernel smoothing for extremal quantile regression, *Technical report*.
- Daouia, A., Gardes, L., Girard, S. & Lekina, A. (2011). Kernel estimators of extreme level curves, *Test*, **20**, 311–333.
- Davison, A.C. & Ramesh, N.I. (2000). Local likelihood smoothing of sample extremes, *Journal of the Royal Statistical Society Series B*, **62**, 191–208.
- Davison, A.C. & Smith, R.L. (1990). Models for exceedances over high thresholds, *Journal of the Royal Statistical Society Series B*, **52**, 393–442.
- de Haan, L. & Ferreira, A. (2006). *Extreme Value Theory : An Introduction*, Springer.
- Deheuvels, P., Haeusler, E. & Mason, D.M. (1988). Almost sure convergence of the Hill estimator, *Mathematical Proceedings of the Cambridge Philosophical Society*, **104**, 371–381.
- Dekkers, A., Einmahl, J. & de Haan, L. (1989). A moment estimation for the index of an extreme-value distribution, *Annals of Statistics*, **17**, 1833–1855.
- Diebolt, J., Gardes, L., Girard, S. & Guillou, A. (2008a). Bias-reduced estimators of the Weibull tail-coefficient, *Test*, **17**, 311–331.
- Diebolt, J., Guillou, A., Naveau, P. & Ribereau, P. (2008b). Improving probability-

weighted moment methods for the generalized extreme value distribution, *Revstat - Statistical Journal*, **6**, 1, 33-50.

Diebolt, J., Guillou, A. & Rached, I. (2007). Approximation of the distribution of excesses through a generalized probability-weighted moments method, *Journal of the Statistical Planning and Inference*, **137**, 841–857.

Drees, H. (1998). On smooth statistical tail functionals, *Scandinavian Journal of Statistics*, **25**, 187–210.

Ferraty, F. & Vieu, P. (2006). *Nonparametric Functional Data Analysis : Theory and Practice*, Springer Series in Statistics, Springer.

Feuerverger, A. & Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution, *Annals of Statistics*, **27**, 760–781.

Fraga Alves, M.I., Gomes, M.I. & de Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter, *Portugaliae Mathematica*, **60**, 193–213.

Gannoun, A., Girard, S., Guinot, C. & Saracco, J. (2002). Reference ranges based on nonparametric quantile regression, *Statistics in Medicine*, **21**, 3119–3135.

Gardes, L. & Girard, S. (2008). A moving window approach for nonparametric estimation of the conditional tail index, *Journal of Multivariate Analysis*, **99**, 2368–2388.

Gardes, L. & Girard, S. (2010). Conditional extremes from heavy-tailed distributions : an application to the estimation of extreme rainfall return levels, *Extremes*, **13**, 177–204.

Goegebeur, Y. & Guillou, A. (2012). Asymptotically unbiased estimation of the coefficient of tail dependence, *to appear in Scandinavian Journal of Statistics*.

Goegebeur, Y., Beirlant, J. & de Wet, T. (2010). Kernel estimators for the second order parameter in extreme value statistics, *Journal of the Statistical Planning and Inference*, **140**, 2632–2652.

Goegebeur, Y. & de Wet, T. (2012). Local estimation of the second order parameter in extreme value statistics and local unbiased estimation of the tail index, *Communications in Statistics - Theory and Methods*, **41**, 3575-3607.

- Gomes, M.I., de Haan, L. & Peng, L. (2002). Semi-parametric estimation of the second order parameter in statistics of extremes, *Extremes*, **5**, 387–414.
- Gray, H.L. & Schucany, W.R. (1972). *The Generalized Jackknife Statistic*, Marcel Dekker.
- Hall, P. (1982). On some simple estimates of an exponent of regular variation, *Journal of the Royal Statistical Society Series B*, **44**, 37–42.
- Hall, P. & Tajvidi, N. (2000). Nonparametric analysis of temporal trend when fitting parametric models to extreme value data, *Statistical Science*, **15**, 153–167.
- Heffernan, J. E. (2000). A directory of coefficients of tail dependence, *Extremes*, **3**, 279–290.
- Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *Annals of Statistics*, **3**, 1163–1174.
- Hosking, J. & Wallis, J. (1987). Parameter and quantile estimation for the generalized Pareto distribution, *Technometrics*, **29**, 339–349.
- Huang, X. (1992). *Statistics and bivariate extremes*, PhD thesis, Erasmus University, Rotterdam, Tinbergen Institute Research series No 22.
- Joe, H. (1993). Parametric family of multivariate distributions with given margins, *Journal of Multivariate Analysis*, **46**, 262–82.
- Kratz, M. & Resnick, S. (1996). The QQ-estimator and heavy tails, *Stochastic Models*, **12**, 699–724.
- Ledford, A.W. & Tawn, J.A. (1996). Statistics for near independence in multivariate extreme values, *Biometrika*, **83**, 169–187.
- Ledford, A.W. & Tawn, J.A. (1997). Modelling dependence within joint tail regions, *Journal of the Royal Statistical Society Series B*, **59**, 475–99.
- Ledford, A.W. & Tawn, J.A. (1998). Concomitant tail behaviour for extremes, *Advances in Applied Probability*, **30**, 197–215.
- Mason, D.M. (1982). Laws of large numbers for sums of extreme values, *Annals of Probability*, **10**, 754–764.

- Matheron, G. (1987). Suffit-il, pour une covariance, d'être de type positif?, *Sciences de la Terre, Série informatique géologique*, **26**, 51–66.
- Naveau, P., Guillou, A., Cooley, D. & Diebolt, J. (2009). Modeling pairwise dependence of maxima in space, *Biometrika*, **96**, 1–17.
- Pauli, F. & Coles, S. (2001). Penalized likelihood inference in extreme value analysis, *Journal of Applied Statistics*, **28**, 547–560.
- Pickands, J. (1975). Statistical inference using extreme order statistics, *Annals of Statistics*, **3**, 119–131.
- Pickands, J. (1981). Multivariate extreme value distributions, *Bulletin of the International Statistical Institute*, **49**, 859–878.
- Ramos, A. & Ledford, A. (2009). A new class of models for bivariate joint tails, *Journal of the Royal Statistical Society Series B*, **71**, 219–241.
- Ramos, A. & Ledford, A. (2011). Alternative point process framework for modeling multivariate extreme values, *Communications in Statistics - Theory and Methods*, **40**, 2205–2224.
- Schlather, M. (2002). Models for stationary max-stable random fields, *Extremes*, **5**, 33–44.
- Schlather, M. & Tawn, J. (2003). A dependence measure for multivariate and spatial extreme values : properties and inference, *Biometrika*, **90**, 139–56.
- Schultze, J. & Steinebach, J. (1996). On least squares estimates of an exponential tail coefficient, *Statistics and Decisions*, **14**, 353–372.
- Smith, R.L. (1987). Estimating tails of probability distributions, *Annals of Statistics*, **15**, 1174–1207.
- Smith, R.L. (1989). Extreme value analysis of environmental time series : an application to trend detection in ground-level ozone, *Statistical Science*, **4**, 367–393.
- van der Vaart, A.W. (1998). *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics.
- Wang, H. & Tsai, C.L. (2009). Tail index regression, *Journal of the American Statistical*

*Association*, **104**, 1233–1240.

Yao, Q. (1999). Conditional predictive regions for stochastic processes, *Technical report*, University of Kent at Canterbury.



# Antoine SCHORGEN

## Valeurs Extrêmes : covariables et cadre bivarié

### Résumé

Cette thèse aborde deux sujets peu traités dans la littérature concernant le théorie des valeurs extrêmes : celui des observations en présence de covariables et celui des mesures de dépendance pour des paires d'observations.

Dans la première partie de cette thèse, nous avons considéré le cas où la variable d'intérêt est observée simultanément avec une covariable, pouvant être fixe ou aléatoire. Dans ce contexte, l'indice de queue dépend de la covariable et nous avons proposé des estimateurs de ce paramètre dont nous avons étudié les propriétés asymptotiques. Leurs comportements à distance finie ont été validés par simulations.

Puis, dans la deuxième partie, nous nous sommes intéressés aux extrêmes multivariés et plus particulièrement à mesurer la dépendance entre les extrêmes. Dans une situation proche de l'indépendance asymptotique, il est très difficile de mesurer cette dépendance et de nouveaux modèles doivent être introduits. Dans ce contexte, nous avons adapté un outil de géostatistique, le madogramme, et nous avons étudié ses propriétés asymptotiques. Ses performances sur simulations et données réelles ont également été exhibées.

Cette thèse offre de nombreuses perspectives, tant sur le plan pratique que théorique dont une liste non exhaustive est présentée en conclusion de la thèse.

Mots-clés : statistiques, valeurs extrêmes, covariables, indépendance asymptotique, madogramme, simulations

### Résumé en anglais

This thesis presents a study of the extreme value theory and is focused on two subjects rarely analyzed: observations associated with covariates and dependence measures for pairs of observations.

In the first part, we considered the case where the variable of interest is simultaneously recorded with a covariate which can be either fixed or random. The conditional tail index then depends on the covariate and we proposed several estimators with their asymptotic properties. Their behavior have been approved by simulations.

In the second part, we were interested in multivariate extremes and more particularly in measuring the dependence between them. In a case of near asymptotic independence, we have to introduce new models in order to measure the dependence properly. In this context, we adapted a geostatistical tool, the madogram, and studied its asymptotic properties. We completed the study with simulations and real data of precipitations.

Keywords: statistics, extreme values, covariates, asymptotic independence, madogram, simulations