People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
MOHAMED KHIDER UNIVERSITY, BISKRA
Faculty of Exact Sciences and the Natural Sciences and Life
DEPARTMENT OF MATHEMATICS

A Thesis Presented for the Degree of

DOCTOR OF SCIENCES

In the Field of Statistics

By

Lazhar BENKHELIFA

Title

On the Multivariate Measures of Association and Extreme Risks

Examination Committee Members

Brahim MEZERDI	Professor	Univ. Biskra	Chairman
Abdelhakim NECIR	Professor	Univ. Biskra	Supervisor
Djamel MERAGHNI	Professor	Univ. Biskra	Examinator
Zaher MOHDEB	Professor	Univ. Constantine 1	Examinator
Hocine FELLAG	Professor	Univ. Tizi Ouzou	Examinator
Abderrahmane YOUSFATE	Professor	Univ. Sidi Bel Abbès	Examinator

March 17, 2015

Acknowledgements

First and foremost, I would like to express my deepest gratitude and sincere appreciation to my supervisor, Professor Dr. *Abdelhakim NECIR*, without whose mentorship this thesis never would have been written. I would like to thank him for the useful comments, remarks, excellent guidance, caring, patience and providing me with an excellent atmosphere for doing research.

My appreciation also goes to Dr. *Namir GHORAF* from the Department of Mathematics at Larbi Ben M'Hidi university of Oum El Bouaghi for inviting me to join in the laboratory of Dynamical Systems and Control.

I am extremely grateful to Professor Dr. *Brahim MEZERDI* to have accepted to chair the jury of my thesis. This is a great honor for me.

I also express deep gratitude to my other examination committee members: Professor Dr. *Djamel MERAGHNI*, Professor Dr. *Zaher MOHDEB*, Professor Dr. *Hocine FELLAG* and Professor Dr. *Abderrahmane YOUSFATE*. They generously gave their time and expertise to make my work better. I thank them for their contribution, support, careful reading the entire manuscript and providing valuable comments.

I cannot close these acknowledgements without to thank my parents and the rest of my family who have been my source of strengths and counsel. Thanks to them for providing me with their care, attention and unending support. I would like to give special thanks to my wife. I would not have accomplished this work without her patience and encouragements in very hard moments.

To my baby Meriem

Abstract

The aim of this thesis is to propose new estimators of copula-based measures of multivariate association and extreme risks. The estimation of the distortion risk premiums for heavy-tailed losses was proposed by Necir and Meraghni (2009). Their considerations are based on the Hill estimator (Hill, 1975) of extreme tail index and Weissman's estimator (Weissman, 1978) of the high quantile. It is well known, in the extreme value theory, that Hill's estimator exhibits an important bias which leads to an over/under estimation the aforementioned estimators of the distortion risk premiums. Several reduced biased estimations of the tail index are now available in the literature that solves this problem. In this thesis, we choose the kernel estimation method to derive a new estimator of the distortion risk premiums for large claims and establish its asymptotic normality. From the simulation study, it is clear that the newly estimator has a reduced bias, vis-à-vis to the existing ones, for any choice of the kernel function.

Keywords: Copula, Dependence, Measure of association, Extreme values theory, Tail index, Kernel estimator, Heavy-tailed, Risk premium, Bias reduction, Asymptotic normality.

Résumé

L'objectif de cette thèse est de proposer de nouveaux estimateurs de mesures d'association multivariées et de risques extrêmes. L'estimation des primes de risque de distorsion pour les distributions à queue lourdes a été proposée par Necir et Meraghni (2009). Leurs considérations sont basées sur les estimateurs de Hill (Hill, 1975) de l'indice de queue et des quantiles extrêmes (Weissman, 1978). Il est bien connu, dans la théorie des valeurs extrêmes, que l'estimateur de Hill présente un biais important qui conduit à une sur/sous-estimation des estimateurs des primes de risque de distorsion. Plusieurs estimateurs à biais réduits de l'indice de queue sont maintenant disponibles dans la littérature qui permet de résoudre ce problème. Dans cette thèse, nous choisissons la méthode du noyau pour obtenir un nouvel estimateur des primes de risque de distorsion pour les grandes pertes et établir sa normalité asymptotique. Une simulation, montre que notre estimateur à biais réduit, vis-à-vis ceux qui existent déjà, pour tout choix du noyau.

Mots-clés: Copule, Dépendance, Mesure d'association, Théorie des valeurs extrêmes, Indice de queue, Estimation à noyau, Queue lourde, Prime de risque, Réduction du biais, Normalité asymptotique.

ملخص

الهدف من هذه الأطروحة هو اقتراح مقدرات جديدة لقياسات الارتباط و الأخطار القصوى. التقدير الإحصائي لأقساط التأمين المشوهة للتوزيعات ذات الأذيال الثقيلة المقترح من طرف نصير و مرغني (2009) يعتمد على مقدر هيل (1975) و مقدر ويسمان (1978). من المعروف، في نظرية القيم القصوى، أن مقدر هيل هو مقدر متحيز. لهذا السبب تم اقتراح عدة مقدرات أقل تحيزا. في هذه الأطروحة ، نختار طريقة النواة للحصول على مقدر جديد لأقساط الخطر المشوهة وإثبات توزيعه المقارب. من خلال المحاكاة، أثبتنا أن مقدرنا الجديد أقل تحيزا من المقدرات الموجودة مهما كانت دالة النواة.

الكلمات المفتاحية: دالة الارتباط، الارتباط، نظرية القيم القصوى، مؤشر القيم القصوى، التقدير الاحصائي، التقدير بالنواة، التوزيعات ذات الأذيال الثقيلة، قسط التأمين، تقليل التحيز، التوزيع الطبيعي المقارب.

Achieved Works

1. Kernel-type Estimator of the Reinsurance Premium for Heavy-tailed Loss Distributions.

Insurance: Mathematics and Economics, Vol. 59 (2014), 65-70.

2. Kernel-type Estimators for the Distortion Risk Premiums of Heavy-tailed Distributions.

Scandinavian Actuarial Journal (2014), DOI: 10.1080/03461238.2014.924434.

Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

a.s.	almost sure
CLT	central limit theorem
$Cov(X,Y)$	covariance of random variables X, Y
df	distribution function
$DA\left(H\right)$	domain of attraction
e.g.	for example
EVI	extreme value index
EVT	extreme value theory
$E\left(X\right)$	expectation or mean of X
$E\left(X^p\right)$	pth moment of X $(p = 1, 2, \ldots)$
F	df of X
$\overline{F}$	survival function, tail of F
$F^{\leftarrow}$	generalized inverse of F, quantile function
GEVD	extreme value distribution
i.e.	in other words
iid	independent identically distributed
$\mathbf{1}_A$	indicator function of set A
MSE	mean squared error
$\mathbb{N}$	set of non-negative integers
$\mathcal{N}\left(\mu, \sigma^2\right)$	normal or Gaussian distribution with mean μ and variance σ^2
$\mathcal{N}\left(0, 1\right)$	standard normal or standard Gaussian distribution
$(\Omega, \mathcal{F}, \mathbb{P})$	probability space
$o\left(\cdot\right)$	$f\left(x\right) = o\left(g\left(x\right)\right)$ as $x \to x_0 : f\left(x\right)/g\left(x\right)$ as $x \to x_0$
$O\left(\cdot\right)$	$f\left(x\right) = O\left(g\left(x\right)\right)$ as $x \to x_0 : \exists M > 0,\ \left\vert f\left(x\right)/g\left(x\right)\right\vert \leq M$ as $x \to x_0$

$o_{\mathbb{P}}(\cdot)$ and $O_{\mathbb{P}}(\cdot)$	stochastic order symbols
pdf	probability density function
Q	generalized inverse function of F, quantile function
Q_n	empirical quantile function
$\mathbb{R}$	set of real numbers
$\mathbb{R}_+$	set of positive real numbers
$\mathcal{RV}_\rho$	regular variation at ∞ with index ρ
$\mathcal{RV}_\rho^0$	regular variation at 0 with index ρ
rv	random variable
$U(0,1)$	continuous uniform distribution with support $[0,1]$
$Var(X)$	variance of X
X	rv defined on $(\Omega, \mathcal{F}, \mathbb{P})$, population
$(X_1, \ldots, X_n)$	sample of size n from X
$(X_{1,n}, \ldots, X_{n,n})$	order statistics pertaining to $(X_1, \ldots, X_n)$
$X_{i,n}$	ith order statistic $(i = 1, \ldots, n)$
$\xrightarrow{a.s.}$	a.s. convergence
$\xrightarrow{\mathbb{P}}$	convergence in probability
$\overset{\mathcal{D}}{=}$	equality in distribution
$a \wedge b$	$a \wedge b = \min(a, b)$
$a \vee b$	$a \vee b = \max(a, b)$
$[x]$	integer part of real number x
$\xrightarrow{\mathcal{D}}$	convergence in distribution
#	number of
$\sim$	$f(x) \sim g(x)$ as $x \to x_0 : f(x)/g(x) \to 1$ as $x \to x_0$
$\sim$	$V \sim D : V$ has distribution D

Intorduction

The problem of measuring the amount of association between two or more variables is an important issue in actuarial practices. Generally, the pairwise dependence is measured by the canonical Pearson's correlation coefficient. But this coefficient may not be the best measure of dependence when dealing with extremes (see Joe, 1997) because it does not exist for heavy-tailed variables with infinite variance and only involves a linear kind of dependence. Therefore, the copula models are becoming increasingly popular to measure the relationship of dependences between two or more variables.

The extreme events such as natural disasters, industrial catastrophes and financial crashes are the worst thing that could occur in insurance. These extreme events are responsible for the biggest losses of the insurance and reinsurance companies. For this reason, the statistical methods that deal with extreme losses have become necessary for actuaries. Extreme value theory (EVT) has become one of the main theories in developing statistical models for extreme insurance losses. EVT is also becoming widely used in many other disciplines such as hydrology, finance, structural engineering and biostatistics, see Reiss and Thomas (2007), Beirlant et al. (2004) and Castillo et al. (2005).

Historically, EVT was pioneered by Tippett that was employed by the British Cotton Industry Research Association, where he worked to make cotton thread stronger. In his studies, he realized that the strength of a thread was controlled by the strength of its weakest fibres. With the help of Fisher, Tippet obtained three asymptotic limits describing the distributions of extremes in 1928. In 1958, Gumbel codified the EVT in his book "Statistics of Extremes".

Risk premiums are used to quantify insurance losses. In actuarial insurance literature, there exist several premium calculation principles (see Goovaerts et al., 1984, Denuit et al., 2005 and Furman and Zitikis, 2008). The net premium, the expected value premium, the variance premium, the value at-risk, the conditional tail expectation and the proportional-hazards transform are the most popular premiums. Many of the premiums are a special cases of the distortion premium (see Wang, 1996).

In this thesis, we are interested in the generalization and improvement of the estimator of distortion risk premiums for a heavy-tailed losses by using the EVT. Since, the classical estimators of these premiums are seriously biased under the second order regular variation framework, many authors proposed the use of so-called second order reduced bias estimators for both first order and second order tail parameters to reduce the bias. We have generalized a kernel-type estimator and present a number of results on its distributional behavior and compare its performance with the performance of other estimators.

The organization of the thesis is as follows:

Chapter 1

Copulas are used to evaluate the relationship of dependences between two or more variables. In this chapter, we introduce some basic notions concerning the concepts of copulas and give the most important families of copulas: the elliptical copulas and the Archimedean copulas. Also, we give the empirical copula and weak convergence of the empirical copula process.

Chapter 2

In the literature, several copula-based measures of multivariate association have been proposed. This chapter constitutes a survey on these measures. Wolff (1980) introduced a class of multivariate measures of association which is based on the L_1-, L_2 and L_∞-norms of the difference between the copula and the independence copula. Multivariate extensions of Spearman's rho were considered by Nelsen (1996) and Schmid and Schmidt (2006, 2007a, 2007b). Multivariate version of Blomqvist's beta was proposed by Úbeda-Flores (2005) and Schmid and Schmidt (2007c) whereas a multivariate version of Gini's gamma was proposed by Behboodian et al. (2007). Gaißer et al. (2010) proposed a multivariate version of Hoeffding's phi-square.

Chapter 3

In this chapter, we present the concept of heavy-tailed distributions and different classes of this type of distributions. The heavy tailed distribution allows to model several phenomena encountered in different disciplines such as finance, hydrology and geology. Several definitions were associated with these distributions as a function of classiffication criteria. The characterization the most simple and one based on comparison with the normal distribution. A distribution has a heavy tail if and only if its kurtosis is higher than the normal distribution that is equal to 3. We also provide an overview of the essential definitions and results of univariate EVT.

Chapter 4

The thickness of the tail of a distribution is measured by the so-called tail index or extreme value index (EVI). The estimation of the EVI is a central problem in EVT. Various estimators are available in the literature. In this chapter, we describe analytically four semi-parametric methods of estimation of tail index: Pickands (Pickands, 1975), Hill (Hill, 1975), moment (Dekkers et al., 1989) and kernel-type estimators (Csörgő el al., 1985). High quantile and other related estimations are also presented. The methods of selecting the optimal number of upper order statistics to be used in estimation are given.

Chapter 5

Risk measures and premium calculation principles lie at the heart of actuarial science. This chapter is devoted to the presentation of the concepts of risk measures and premium calculation principles. We start this chapter by giving a definition of risk measures and premium principles. Desirable properties of premium calculation principles are discussed. We also list many well-known premium principles and tabulate which of the properties they satisfy.

Chapter 6

In this chapter, we use the results of previous chapters to introduce a new kernel-type estimators for the distortion risk premiums and reinsurance premium of heavy-tailed loss distributions. Using a least-squares approach, a bias-reduced version of these estimators is proposed. The asymptotic normality of the given estimators is established under suitable assumptions. A simulation study is carried out to illustrate the performance of our method.

Part I

Multivariate Measures of Association

Chapter 1

Copulas

The term copula was first introduced by Sklar (1959) and is derived from the latin word "copũlae", to connect or to join. Initially, copulas were mainly used in the development of the theory of probabilistic metric spaces. Later, it used to define non-parametric measures of dependence between random variables (rv's), and since then, they became an important tools in probability and mathematical statistics. For a comprehensive introduction on the theoretical aspects of copulas, we refer to Joe (1997) and Nelsen (2006) and for a practical approach, we refer to Salvadori et al. (2007) and Jaworski et al. (2010).

1.1 Definitions, properties and examples

Definition 1.1.1 *(Copula)*
A copula or d-dimensional copula $C : [0,1]^d \to [0,1]$ *is a multivariate distribution function (df) on the unit cube with standard uniform marginals.*

Definition 1.1.2 *A copula is a function* $C : [0,1]^d \to [0,1]$ *satisfying the conditions:*

1. *For all* $(u_1, \ldots, u_d)$ *in* $[0,1]^d$*, if at least one component* u_i *is zero, then* $C(u_1, \ldots, u_d) = 0$.

2. *For* $u_i \in [0,1]$, $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$ *for all* $i \in \{1, 2, \ldots, d\}$.

3. *C is d-increasing, for all* $[u_{11}, u_{12}] \times [u_{21}, u_{22}] \times \ldots \times [u_{d1}, u_{d2}]$ *d-dimensional rectangles in* $[0,1]^d$ *, the following inequality holds:*

$$\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\cdots+i_d} C\left(u_{1i_1}, \ldots, u_{di_d}\right) \geq 0.$$

Examples

In the following examples, we give three simplest examples of copulas.

1. Independence Copula

The function $\Pi : [0,1]^d \to [0,1]$, given by

$$\Pi\left(u_1, \ldots, u_d\right) = \prod_{i=1}^{d} u_i, \quad u_1, \ldots, u_d \in [0,1],$$

is called the independence copula. To see that actually is a copula, consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supporting independent identically distributed (iid) rv's $U_1, \ldots, U_d$ with $U_1 \sim U[0,1]$. The random vector $(U_1, \ldots, U_d)$ then has $U[0,1]$-distributed margins and joint df

$$\begin{aligned}\mathbb{P}\left(U_1 \leq u_1, \ldots, U_d \leq u_d\right) &= \prod_{i=1}^{d} \mathbb{P}\left(U_i \leq u_i\right) = \prod_{i=1}^{d} u_i \\ &= \Pi\left(u_1, \ldots, u_d\right), \quad u_1, \ldots, u_d \in [0,1].\end{aligned}$$

2. Comonotonicity Copula

Considering a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supporting a single rv $U \sim U[0,1]$, the random vector $(U_1, \ldots, U_d) = (U, \ldots, U) \in [0,1]^d$ has $U[0,1]$-distributed margins and joint df

$$\begin{aligned}\mathbb{P}\left(U_1 \leq u_1, \ldots, U_d \leq u_d\right) &= \mathbb{P}\left(U \leq \min\left(u_1, \ldots, u_d\right)\right) \\ &= \min\left(u_1, \ldots, u_d\right), \quad u_1, \ldots, u_d \in [0,1].\end{aligned}$$

Consequently, the function $M : [0,1]^d \to [0,1]$, defined by

$$M\left(u_1, \ldots, u_d\right) = \min\left(u_1, \ldots, u_d\right), \quad u_1, \ldots, u_d \in [0,1],$$

is a copula called the copula of complete comonotonicity (also called the upper Fréchet-Hoeffding bound, see theorem 1.3.1).

3. Countermonotonicity Copula

Considering a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supporting a single rv $U \sim U[0,1]$, the bivariate random vector $(U_1, U_2) = (U, 1-U) \in [0,1]^2$ has perfectly negatively associated components, i.e., if U_1 is large then U_2 is small, and vice versa. This random vector has $U[0,1]$-distributed margins and joint df

$$\begin{aligned} \mathbb{P}(U_1 \leq u_1, U_2 \leq u_2) &= \mathbb{P}(1-u_2 \leq U \leq u_1) \\ &= (u_1 + u_2 - 1)\, \mathbf{1}_{\{1-u_2 \leq u_1\}}, \quad u_1, u_2 \in [0,1] \\ &= \max(u_1 + u_2 - 1, 0), \quad u_1, u_2 \in [0,1]. \end{aligned}$$

Consequently, the function $W : [0,1]^2 \to [0,1]$, defined by

$$W(u_1, u_2) = \max(u_1 + u_2 - 1, 0), \quad u_1, u_2 \in [0,1],$$

is a bivariate copula called the copula of complete countermonotonicity (also called the lower Fréchet-Hoeffding bound, see theorem 1.3.1). For any $d \geq 3$, we have the function $W : [0,1]^d \to [0,1]$, defined by

$$W(u_1, \ldots, u_d) = \max(u_1 + \cdots + u_d + d - 1, 0), \quad u_1, \ldots, u_d \in [0,1],$$

is not a copula.

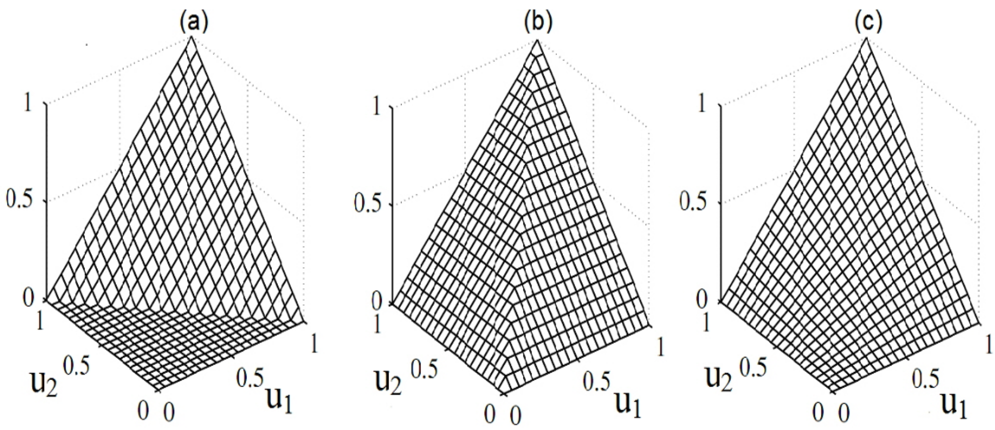

Figure 1.1: Graphs of the bivariate (a) countermonotonicity copula, (b) independence copula and (c) comonotonicity copula.

Proposition 1.1.1 *(Smoothness of a Copula)*
Let $C : [0,1]^d \to [0,1]$ *be a copula.*

1. *For every* $(u_1, \ldots, u_d)$, $(v_1, \ldots, v_d) \in [0,1]^d$ *it holds that*

$$|C(u_1, \ldots, u_d) - C(v_1, \ldots, v_d)| \leq \sum_{i=1}^{d} |u_i - v_i| .$$

   In particular, C *is Lipschitz continuous with the Lipschitz constant equal to 1.*

2. *For* $k = 1, \ldots, d$ *and fixed* $(u_1, \ldots, u_{k-1}, u_{k+1}, \ldots, u_d) \in [0,1]^{d-1}$, *the partial derivative* $u_k \mapsto \frac{\partial}{\partial u_k} C(u_1, \ldots, u_d)$ *exists (Lebesgue) almost everywhere on* $[0,1]$ *and takes values in* $[0,1]$.

1.2 Sklar's theorem

Sklar's theorem is the basic tool of copulas theory and is the foundation of most, of the applications of that theory to statistics. Is show the role that copulas play in the relationship between multivariate df's and their univariate margins.

Theorem 1.2.1 *(Sklar's theorem)*
Let F *be a d-dimensional df with univariate marginal df's* $F_1, \ldots, F_d$. *Then there exists a d-dimensional copula* C *such that for all* $\mathbf{x} = (x_1, \ldots, x_d)$ *in* $\mathbb{R}^d$,

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)) . \tag{1.1}$$

If $F_1, \ldots, F_d$ *are continuous, then* C *is unique. Conversely, if* C *is a d-dimensional copula and* $F_1, \ldots, F_d$ *are univariate df's, then the right-hand side of (1.1) is a d-dimensional df with univariate marginal df's* $F_1, \ldots, F_d$.

The proof is given in Sklar (1959). It follows from Sklar's theorem that a multivariate df can be separated into the univariate (continuous) marginal df's and the multivariate dependence structure, which is represented by the copula. Deheuvels (1978) refers to copulas as "dependence functions".

Corollary 1.2.1 *(Sklar's inversion)*
Let F be a d-dimensional df with univariate marginal df's $F_1, \ldots, F_d$ and corresponding copula C satisfying (1.1). Assuming that $F_1, \ldots, F_d$ are continuous, an explicit representation of C is given by

$$C(\mathbf{u}) = F\left(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right), \quad \mathbf{u} \in [0,1]^d.$$

This result is a direct consequence from theorem 1.2.1 and is important for the construction of copulas from multivariate distributions. If not stated otherwise, we always assume that the univariate marginal df's $F_1, \ldots, F_d$ are continuous.

1.3 Fréchet-Hoeffding bounds

Each copula C is pointwise bounded from above by the comonotonicity copula M. This is intuitive, since the comonotonicity copula implies the strongest positive association possible between components. In dimension $d = 2$, the countermonotonicity copula W is a pointwise lower bound. In contrast, in dimensions $d \geq 3$ there is no "smallest" copula. The explanation for this fact is the following: for $d = 2$ that the two components in a random vector (U_1, U_2), defined on $(\Omega, \mathcal{F}, \mathbb{P})$, are perfectly negatively associated. More clearly, it holds almost surely that $U_1 = 1 - U_2$, i.e. the minus sign implies that if one variable moves in one direction, the other moves precisely in the opposite direction. However, for $d \geq 3$ there are at least three directions and it is not clear how to de ne three directions to be perfect opposites of each other. Nevertheless, the following result provides a sharp lower bound for arbitrary copulas.

Theorem 1.3.1 *(Fréchet-Hoeffding bounds)*
If C is any d-copula, then for every $\mathbf{u} \in [0,1]^d$

$$W^d(\mathbf{u}) \leq C(\mathbf{u}) \leq M^d(\mathbf{u}),$$

where $W^d(\mathbf{u}) = W(u_1, \ldots, u_d)$ and $M^d(\mathbf{u}) = M(u_1, \ldots, u_d)$.

Sometimes it is more convenient to describe the distribution of a random vector $(X_1, \ldots, X_d)$ by means of its survival function instead of its df. Especially when the components X_k are interpreted as lifetimes, this description is more intuitive.

Let $(U_1, \ldots, U_d)$ d-rv's with joint df C. We denote by $\overline{C}$ the joint survival function of C, then

$$\overline{C}(\mathbf{u}) = \mathbb{P}(U_1 > u_1, \ldots, U_d > u_d).$$

Analogously to Sklar's theorem (see theorem 1.2.1), a d-dimensional survival function can be decomposed into a copula and its marginal survival functions.

Theorem 1.3.2 *(Survival analog of Sklar's theorem)*
Let $\overline{F}$ be a d-dimensional df with univariate marginal df's $\overline{F}_1, \ldots, \overline{F}_d$. Then there exists a d-dimensional copula $\overline{C}$ such that for all $x = (x_1, \ldots, x_d)$ in $\mathbb{R}^d$, it holds that

$$\overline{F}(x_1, \ldots, x_d) = \overline{C}\left(\overline{F}_1(x_1), \ldots, \overline{F}_d(x_d)\right). \tag{1.2}$$

If $\overline{F}_1, \ldots, \overline{F}_d$ are continuous, then $\overline{C}$ is unique. Conversely, if $\overline{C}$ is a d-dimensional copula and $\overline{F}_1, \ldots, \overline{F}_d$ are univariate survival functions, then the function defined via (1.2) is a d-dimensional survival function.

1.4 Copulas and random vectors

In this section, we discuss some probabilistic properties of copulas that can be inferred from Sklar's theorem. These results appeared already in Joe (1997) (for the proofs see Nelsen, 2006). First, we state the invariance of the copula of the random vector $\mathbf{X}$ with respect to any increasing rescaling of the components of $\mathbf{X}$.

Proposition 1.4.1 *Let $\mathbf{X} = (X_1, \ldots, X_d)$ be a random vector with continuous joint df F and copula C. Let $T_1, \ldots, T_d$ be strictly increasing functions from $\mathbb{R}$ to $\mathbb{R}$. Then C is also the copula of the rv $(T_1(X_1), \ldots, T_d(X_d))$.*

Thus, copulas that describe the dependence of the components of a random vector are invariant under increasing transformations of each coordinate. The next results characterize some special structures of rv's in terms of the basic copulas $\Pi^d = \Pi(u_1, \ldots, u_d)$, M^d and W^2.

Proposition 1.4.2 *Let $\mathbf{X} = (X_1, \ldots, X_d)$ be a random vector with continuous joint df F. Then the copula of $(X_1, \ldots, X_d)$ is Π^d if and only if $X_1, \ldots, X_d$ are independent.*

Proposition 1.4.3 *Let $(X_1, \ldots, X_d)$ be a random vector with continuous joint df F. Then the copula of $(X_1, \ldots, X_d)$) is M^d if, and only if, there exists a rv Z and increasing functions $T_1, \ldots, T_d$ such that $X = (T_1(Z), \ldots, T_d(Z))$ almost surely.*

Proposition 1.4.4 *Let (X_1, X_2) be a random vector with continuous joint df F. Then (X_1, X_2) has copula W^2 if, and only if, for some strictly decreasing function T, $X_2 = T(X_1)$ almost surely.*

1.5 Families of copula

Copulas play an important role in the construction of multivariate df's and, as a consequence, having at one's disposal a variety of copulas can be very useful for building stochastic models having different properties that are sometimes indispensable in practice such as heavy tails and asymmetries. Therefore, several investigations have been carried out concerning the construction of different families of copulas and their properties. In this section, we present just a few of them, by focusing on the families that seem to be more popular in the literature. Different families or construction methods are discussed in Nelsen (2006).

1.5.1 Elliptical copulas

Elliptical copulas are the copulas of elliptical distributions. The class of elliptical copulas has an unfavorable property when talking about application in the field of finance. The dependence structure in financial data cannot be represented correctly. For instance, the asymmetry of the lower and upper tail of a distribution cannot be described properly by an elliptical copula. This is because elliptical copulas exhibit "radial symmetry" which has the property that

$$C(u,v) = C(v,u) \quad \text{and} \quad C(u,v) = \overline{C}(u,v) = u + v - 1 + C(1-u, 1-v).$$

Another fact that elliptical copulas do not have closed form expressions. The class of elliptical distributions provides a rich source of multivariate distributions which share many of the tractable properties of the multivariate normal distribution and enables modelling of multivariate extremes and other forms of non-normal dependences. Simulation from elliptical distributions is easy, and as a consequence of Sklar's theorem so is simulation from elliptical copulas. Furthermore, we will show that rank correlation and tail dependence coefficients can be easily calculated. For further details on elliptical distributions, we refer to Fang et al. (1990) and Cambanis et al. (1981).

Examples of elliptical copulas

1. **Gaussian or Normal copulas**

   The copula of the d-variate normal distribution with linear correlation matrix $\mathcal{R}$ is
$$C^G(\mathbf{u}) = \Phi^d_{\mathcal{R}}\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right),$$
   where $\Phi^d_{\mathcal{R}}$ denotes the joint df of the d-variate standard normal df with linear correlation matrix $\mathcal{R}$, and Φ^{-1} denotes the inverse of the df of the univariate standard normal distribution. Copulas of the above form are called Gaussian copulas. Equivalently, the following definition is often used:

   The family of the d-dimensional Gaussian copulas is defined as
$$C^G(\mathbf{u}) = \int_{-\infty}^{\Phi^{-1}(u_1)} \cdots \int_{-\infty}^{\Phi^{-1}(u_d)} (2\pi)^{-\frac{d}{2}} |\mathcal{R}|^{-\frac{1}{2}} \exp\left(-\mathbf{x}^T\mathcal{R}^{-1}\mathbf{x}\right) dx_1 \ldots dx_d.$$
   In the bivariate case the copula expression can be written as
$$C^G(\mathbf{u}) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\left(1-\mathcal{R}_{12}^2\right)} \exp\left(-\frac{x_1^2 - 2\mathcal{R}_{12}x_1x_2 + x_2^2}{2\left(1-\mathcal{R}_{12}^2\right)}\right) dx_1 dx_2.$$
   Note that $\mathcal{R}_{12}$ is simply the usual linear correlation coefficient of the corresponding bivariate normal distribution.

2. **t-copulas**

   The family of d-dimensional t-copulas is defined by
$$C^t(\mathbf{u}) = t_{\Sigma,\nu}\left(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_d)\right),$$
   where $t_{\Sigma,\nu}$ denotes the multivariate t-distribution with ν degrees of freedom, location vector zero and correlation matrix and corresponding univariate marginal df t_ν with generalized inverse function t_ν^{-1}. Equivalently, the following definition is often used:
$$C^t(\mathbf{u}) = \int_{-\infty}^{t_\nu^{-1}(u_1)} \cdots \int_{-\infty}^{t_\nu^{-1}(u_d)} \frac{\Gamma\left((\nu+d)/2\right)\left(1+\mathbf{w}^T\Sigma^{-1}\mathbf{w}/\nu\right)}{|\Sigma|^{1/2}\,\Gamma(\nu/2)\,(\nu\pi)^{d/2}} dw_1 \ldots dw_d,$$

where $|\Sigma|$ stands for the determinant of Σ, $\mathbf{w} = (w_1, \ldots, w_d)$ and Γ is the Gamma function. For more details, see Hult and Lindskog (2002).

1.5.2 Archimedean copulas

In this subsection we focus on a very important class of copulas called Archimedean copulas. The adventages of this class are

(i) Ease in construction.

(ii) Rich of great variety of families of copulas belonging to this class.

(iii) Nice properties of copula belonging to this class.

(iv) Reduce the study of a multivariate copula to a single univariate function.

The word Archimedean was employed the first time by Ling in 1965 for Archimedean t-norms (every Archimedean copula is also an Archimedean t-norm). And the term «Archimedean copula» was first appeared in the statistical literature in two papers by Genest and Mackay (1986a, 1986b). Archimedean copulas also appear in Schweizer and Sklar (1983) but without the name. For some background on bivariate Archimedean copulas and a discussion on other statistical questions we refer to Genest and MacKay (1986b), Joe (1997) and Nelsen (2006). Here we present the basic properties and examples of the Archimedean class of copulas. Basically, we follow the approach in Nelsen (2006). First, we introduce some notations.

Definition 1.5.1 *(Pseudo-inverse)*
Let φ be a continuous, strictly decreasing function from $[0,1]$ to $[0,1]$ such that $\varphi(1) = 0$. The pseudo-inverse of 'φ is the function $\varphi^{[-1]}$:$[0,\infty] \to [0,1]$given by

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t), & 0 \le t \le \varphi(0), \\ 0, & \varphi(0) \le t \le \infty. \end{cases}$$

Note that $\varphi^{[-1]}$ is continuous and decreasing on $[0,\infty]$, and strictly decreasing on $[0,\varphi(0)]$. Furthermore, $\varphi^{[-1]}\varphi(0) = u$ on $[0,\infty]$ and

$$\varphi\left(\varphi^{[-1]}(t)\right) = \begin{cases} t, & 0 \le t \le \varphi(0), \\ \varphi(0), & \varphi(0) \le t \le \infty. \end{cases}$$

Finally, if $\varphi(0) = \infty$, then $\varphi^{[-1]} = \varphi^{-1}$.

Theorem 1.5.1 *Let φ be a continuous, strictly decreasing function from $[0,1]$ to $[0,1]$ such that $\varphi(1) = 0$, and let $\varphi^{[-1]}$ be the pseudo-inverse of 'φ. Let C be the function from $[0,1]^2$ to $[0,1]$ given by*

$$C(u,v) = \varphi^{[-1]}(\varphi(u) + \varphi(v)). \tag{1.3}$$

Then C is a copula if and only if φ is convex.

Proof. See Nelsen (2006, p. 111). ■

Definition 1.5.2 *(Archimedean copulas)*
Copulas of the form (1.3) are called bivariate Archimedean copulas.
The function φ is called the generator of the copula C. If $\varphi(0) = \infty$, the generator φ is said to be strict. In this case, $\varphi^{[-1]} = \varphi^{-1}$ and

$$C(u,v) = \varphi^{-1}(\varphi(u) + \varphi(v))$$

is said to be a strict Archimedean copula.

The approach in formula (1.3) can naturally be extended to d ($d \geq 2$) dimensions by imposing additional assumptions on φ. With continuous, strictly decreasing function φ such that $\varphi(1) = 0$ and $\varphi(0) = \infty$, a d-dimensional Archimedean copula is given by

$$C(\mathbf{u}) = \varphi^{-1}\left(\sum_{i=1}^{d} \varphi(u_i)\right), \quad \mathbf{u} \in [0,1]^d,$$

if and only if the inverse φ^{-1} is completely monotone on $[0,\infty)$, i.e., if it has derivatives of all orders which alternate in sign, formally,

$$(-1)^k \frac{d^k}{dt^k}\varphi^{-1}(t) \geq 0,$$

for all $t \geq 0$ and all $k \in \mathbb{N}$.

Examples of Archimedean copulas

1. **Clayton copula**

   The Clayton (1978) copula, originally studied by Kimeldorf and Sampson (1975), takes the form:

$$C^{Cl}(\mathbf{u}) = \left(u_1^{-\theta} + \cdots + u_d^{-\theta} - d + 1\right)^{-1/\theta},$$

where $\theta > 0$ is the dependence parameter. The limiting case $\theta = 0$ corresponds to the independent case. This strict copula family has generator

$$\varphi(t) = \frac{\left(t^{-\theta} - 1\right)}{\theta}.$$

2. Gumbel copula

The d-dimensional Gumbel copula (1960) takes the form:

$$C^{Gu}(\mathbf{u}) = \exp\left\{-\left[(-\ln u_1)^{\theta} + \cdots + (-\ln u_d)^{\theta}\right]^{1/\theta}\right\},$$

where $\theta > 1$ is the dependence parameter. Values of 1 and ∞ correspond to independence and the Fréchet upper bound, but this copula does not attain the Fréchet lower bound for any value of . The Gumbel copulas are strict Archimedean with generator

$$\varphi(t) = (-\ln t)^{\theta}.$$

3. Frank copula

The bivariate Frank (1979) copula function is defined as

$$C^{Fr}(u, v) = -\frac{1}{\theta}\log\left\{1 + \frac{\left(e^{-\theta u} - 1\right)\left(e^{-\theta v} - 1\right)}{e^{-\theta} - 1}\right\}, \quad -\infty < \theta < \infty.$$

For $\theta = 0$, the Frank copula is the independence copula.

In general, there exist several methods to generate random numbers from a given Archimedean copula, if needed, we use the method proposed by Marshall and Olkin (1988). For the discussion of the general class of hierarchical Archimedean copulas and related random number generation, we refer to Hofert (2008).

1.6 Empirical copula

There exist three methods for the estimation of copula functions: parametric, semi-parametric and non-parametric methods. The parametric and semi-parametric estimation methods are usually based on maximum-likelihood methods, see Genest et al. (1995), Joe and Xu (1996), Joe (2005), Malevergne and Sornette (2005) and Kim et al. (2007). In this thesis, we solely consider non-parametric estimation methods for which the joint and the marginal df's are assumed to be unknown. In particular, this method is not exposed to possible misspecifications of the underlying distributions. Non-parametric estimation of copulas was first considered by Rüschendorf (1976) and Deheuvels (1979) proposed the so-called empirical copula as a non-parametric estimator.

1.6.1 Definition

Let $\mathbf{X}$ be a d-dimensional random vector with df F, continuous univariate marginal df's F_i, $i = 1, \ldots, d$, and copula C. Suppose that F, F_i and C are completely unknown and let $\mathbf{X_1}, \ldots, \mathbf{X_n}$ be a random sample from $\mathbf{X}$. We can construct the empirical copula in two steps. First, every univariate marginal df F_i is estimated by its univariate empirical df, i.e.,

$$\widehat{F}_{i,n}(x) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}_{\{X_{ij} \le x\}}, \text{ for } j = 1, \ldots, n \text{ and } x \in \mathbb{R}.$$

Then, the estimated marginal df's are used to obtain the so called pseudo-observations $\widehat{U}_{ij,n} = \widehat{F}_{i,n}(X_{ij})$ with $\widehat{\mathbf{U}}_{j,n} = \left(\widehat{U}_{1j,n}, \ldots, \widehat{U}_{dj,n}\right)$, for $i = 1, \ldots, d$ and $j = 1, \ldots, n$. Second, an estimate of the copula C is given by the empirical df of the sample $\widehat{\mathbf{U}}_{1,n}, \ldots, \widehat{\mathbf{U}}_{n,n}$. This is called the empirical copula and was introduced by Deheuvels (1979) under the name "empirical dependence function". Then, we have the following definition:

Definition 1.6.1 (*Empirical copula*)
Let $\mathbf{X}$ be a d-dimensional random vector with df F, continuous univariate marginal df's F_i, $i = 1, \ldots, d$, and copula C. Based on a random sample $\mathbf{X_1}, \ldots, \mathbf{X_n}$ from $\mathbf{X}$, the empirical copula is defined as

$$\widehat{C}_n(\mathbf{u}) = \frac{1}{n} \sum_{j=1}^{n} \prod_{i=1}^{d} \mathbf{1}_{\{\widehat{U}_{ij,n} \le u_i\}}, \quad \text{for } \mathbf{u} \in [0,1]^d,$$

where $\widehat{U}_{ij,n} = \widehat{F}_{i,n}(X_{ij})$, *for* $i = 1, \ldots, d$ *and* $j = 1, \ldots, n$, *and*

$$\widehat{F}_{i,n}(x) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}_{\{X_{ij} \leq x\}}.$$

Since $\widehat{U}_{ij,n} = 1/n(\text{rank of } X_{ij} \text{ in } X_{i1}, \ldots, X_{in})$, the empirical copula represents a rankbased estimator for the copula C, i.e., only the (normalized) ranks of the observations are included in the estimation. According to definition 1.1.1, the empirical copula itself is a copula. In particular, it is invariant under strictly increasing transformations of the margins (see proposition 1.4.2) due to the invariance property of the ranks with respect to such transformations. According to Genest and Favre (2007), the ranks associated with the random sample $\mathbf{X_1}, \ldots, \mathbf{X_n}$are the statistics that retain the greatest amount of information among all statistics fulfilling this invariance property.

1.6.2 Weak convergence of the empirical copula process

The weak convergence of the empirical copula process $\sqrt{n}\left(\widehat{C}_n - C\right)$ can be established by using the functional delta-method (see van der Vaart and Wellner, 1996, p. 389). This convergence has been investigated, e.g., by Rüschendorf (1976) and Tsukahara (2005). Fermanian et al. (2004) established the weak convergence in the following theorem.

Theorem 1.6.1 *Let* $\mathbf{X_1}, \ldots, \mathbf{X_n}$ *be a random sample from the d-dimensional random vector* $\mathbf{X}$ *with df* F*, continuous univariate marginal df's* $F_1, \ldots, F_d$*, and copula* C*. Under the assumption that the i-th partial derivatives* $D_iC(\mathbf{u})$ *of* C *exist and are continuous for* $i = 1, \ldots, d$*, we have*

$$\sqrt{n}\left(\widehat{C}_n(\mathbf{u}) - C(\mathbf{u})\right) \xrightarrow{\mathcal{D}} \mathbb{G}_C(\mathbf{u}).$$

Weak convergence takes place in $\ell^\infty\left([0,1]^d\right)$ *and*

$$\mathbb{G}_C(\mathbf{u}) = \mathbb{B}_C(\mathbf{u}) - \sum_{i=1}^{d} D_iC(\mathbf{u})\mathbb{B}_C\left(\mathbf{u}^i\right).$$

The vector $\mathbf{u}^i$ *denotes the vector where all coordinates, except the ith coordinate of* u*, are replaced by* 1*. The process* $\mathbb{B}_C$ *is a tight centered Gaussian process on* $[0,1]^d$ *with covariance function*

$$E\left(\mathbb{B}_C(\mathbf{u})\,\mathbb{B}_C(\mathbf{v})\right) = C(\mathbf{u} \wedge \mathbf{v}) - C(\mathbf{u})\,C(\mathbf{v}),$$

i.e., $\mathbb{B}_C$ *is a d-dimensional Brownian bridge.*

The non-parametric estimation of the survival function, see (1.2), can be established analogously. The following result is discussed and proven in Schmid and Schmidt (2007a).

Theorem 1.6.2 *Let* $\mathbf{X_1}, \ldots, \mathbf{X_n}$ *be a random sample from the d-dimensional random vector* $\mathbf{X}$ *with df F, continuous univariate marginal df's* $F_1, \ldots, F_d$*, and copula C. Using the same notation as in theorem 1.3.1, a non-parametric estimator for* $\overline{C}$ *is given by*

$$\widehat{\overline{C}}_n(\mathbf{u}) = \frac{1}{n} \sum_{j=1}^{n} \prod_{i=1}^{d} \mathbf{1}_{\left\{\widehat{U}_{ij,n} > u_i\right\}}, \quad \text{for } \mathbf{u} \in [0,1]^d .$$

Under the additional assumption that the i-th partial derivatives $D_i\overline{C}(\mathbf{u})$ *of* $\overline{C}$ *exist and are continuous for* $i = 1, \ldots, d$*, we have*

$$\sqrt{n}\left(\widehat{\overline{C}}_n(\mathbf{u}) - \overline{C}(\mathbf{u})\right) \xrightarrow{\mathcal{D}} \mathbb{G}_{\overline{C}}(\mathbf{u}),$$

in $\ell^\infty\left([0,1]^d\right)$ *and*

$$\mathbb{G}_{\overline{C}}(\mathbf{u}) = \mathbb{B}_{\overline{C}}(\mathbf{u}) - \sum_{i=1}^{d} D_i\overline{C}(\mathbf{u})\mathbb{B}_{\overline{C}}\left(\mathbf{u}^i\right).$$

The vector $\mathbf{u}^i$ *denotes the vector where all coordinates, except the ith coordinate of u, are replaced by 1. The process* $\mathbb{B}_C$ *is a tight centered Gaussian process on* $[0,1]^d$ *with covariance function*

$$E\left(\mathbb{B}_{\overline{C}}(\mathbf{u})\,\mathbb{B}_{\overline{C}}(\mathbf{v})\right) = \overline{C}(\mathbf{u} \wedge \mathbf{v}) - \overline{C}(\mathbf{u})\,C(\mathbf{v}).$$

Chapter 2

Copula-based measures of multivariate association

Copulas are functions that capture the dependence structure between two or more rv's. However, when analyzing the strenght of association in a random vector, this concept does not suffice. Measures of bivariate association, such as Pearson's linear correlation coefficient, fill this gap by aggregating the complete dependence structure to a single numbre. Although measurement of bivariate association has been thoroughly studied and is widely applied, generalizations of the bivariate concepts and measures to dimensions $d \geq 3$ have only recently gained more attention. In this chapter, we introduce and discuss some important measures and concepts of multivariate association. For more details, we refer the interested reader to Joe (1997) and Nelsen (2006).

2.1 Proprieties of measures of multivariate association

A measure of multivariate association quantifies the degree of association between the components of a d-dimensional ($d \geq 2$) random vector $\mathbf{X}$ with df F and copula C. We think of it as a map

$$\delta : \mathcal{C}_d \rightarrow D \subseteq \mathbb{R},$$

which we denote by $\delta(C)$ or equivalently by $\delta(\mathbf{X}) = \delta(X_1, \ldots, X_d)$ where $\mathcal{C}_d$ set of copulas. Desirable properties of bivariate measures of association are well-established and have been discussed, e.g., by Rényi (1959) and Scarsini (1984). However, the extension of those properties to the multivariate case is not always straightforward as the study of multivariate association is generally more complex. An example is the fact that, in contrast to perfect positive dependence, the notion of perfect negative dependence does not generalize to the multivariate case. In this section, we confine ourselves to giving a selection of various properties of measures of multivariate association, which are considered desirable in the literature.

P1. Well-definedness: The measure δ is well-defined for every random vector $\mathbf{X} = (X_1, \dots, X_d)$ with continuous marginals and is a function of the copula $C \in \mathcal{C}_d$, i.e.$\delta(X_1, \dots, X_d) = \delta(C)$.

P2. Invariance with respect to permutations: For every permutation π we have

$$\delta(X_1, \dots, X_d) = \delta\left(X_{\pi(1)}, \dots, X_{\pi(d)}\right).$$

P3. Normalization:

(a) If Π is the copula of $\mathbf{X}$, then $\delta(\mathbf{X}) = \delta(\mathit{\Pi}) = 0$.

(b) If $\delta(\mathbf{X}) = 0$, then $\mathbf{X}$ has copula $\mathit{\Pi}$.

(c) If δ is the copula of $\mathbf{X}$, then $\delta(\mathbf{X}) = \delta(M) = 1$.

(d) If $\delta(\mathbf{X}) = 1$, then $\mathbf{X}$ has copula M or W in dimension $d = 2$. If $\delta(\mathbf{X}) = 1$ then $\mathbf{X}$ has copula M in higher dimension.

Multivariate measures of association further support different notions of orderings in the set of copulas. Here, we consider the partial order $\prec$, where $C_1 \prec C_2$ if and only if $C_1(\mathbf{u}) \leq C_2(\mathbf{u})$ for all $\mathbf{u} \in [0,1]^d$.

P4. Monotonicity and concordance:

If $\mathbf{X}$ has copula $C_{\mathbf{X}}$ and $\mathbf{Y}$ has copula $C_{\mathbf{Y}}$ such that $C_{\mathbf{X}} \prec C_{\mathbf{Y}}$, then $\delta(\mathbf{X}) \leq \delta(\mathbf{Y})$.

P5. Behaviour under transformations:

For strictly increasing (or decreasing) and continuous transformations T_i we have

$$\delta(X_1, \dots, X_d) = \delta\left(T_1(X_1), \dots, T_d(X_d)\right).$$

P6. Continuity:

If $(\mathbf{X}_n)_{n\in\mathbb{N}}$ is a sequence of random vectors with copulas C_n and $\lim_{n\to\infty} C_n(\mathbf{u}) = C(\mathbf{u})$ for all $\mathbf{u} \in [0,1]^d$ and some copula C of the random vector $\mathbf{X}$, then $\lim_{n\to\infty} \delta(\mathbf{X}_n) = \delta(\mathbf{X})$.

P7. Addition of an independent component:

$\delta(X_1, \ldots, X_d) \geq \delta(X_1, \ldots, X_d, X_{d+1})$ if X_{d+1} is independent of $(X_1, \ldots, X_d)$.

P8. If the joint distribution of $\mathbf{X}$ is multivariate normal and all pairwise correlations ρ_{ij} of X_i and X_j are either non-negative or non-positive, then $\delta(\mathbf{X})$ is a strictly increasing function of the absolute value of each of the pairwise correlations.

2.2 Spearman's rho

Spearman's correlation coefficient or Spearman's rho was first studied by Spearman (1904) and represents one of the best-known measures to quantify the degree of association between two rv's. For the two rv's X_1 and X_2 with bivariate df F and continuous univariate margins F_1, F_2, Spearman's rho is defined as

$$\rho(X_1, X_2) = \frac{Cov(F_1(X_1), F_2(X_2))}{\sqrt{Var(F_1(X_1))}\sqrt{Var(F_2(X_2))}}.$$

If X_1 and X_2 have copula C, then this is equivalent to

$$\begin{aligned} \rho(C) &= \frac{\int_{[0,1]^2} u_1 u_2 dC(u_1, u_2) - \left(\frac{1}{2}\right)^2}{\left(\frac{1}{12}\right)} = 12 \int_{[0,1]^2} C(u_1, u_2)\, du_1 du_2 - 3 \\ &= \frac{\int_{[0,1]^2} C(u_1, u_2)\, du_1 du_2 - \int_{[0,1]^2} \Pi(u_1, u_2)\, du_1 du_2}{\int_{[0,1]^2} M(u_1, u_2)\, du_1 du_2 - \int_{[0,1]^2} \Pi(u_1, u_2)\, du_1 du_2}, \end{aligned} \tag{2.1}$$

because $\int_{[0,1]^2} M(u_1, u_2)\, du_1 du_2 = 1/3$ and $\int_{[0,1]^2} \Pi(u_1, u_2)\, du_1 du_2 = 1/4$ where $\int_{[0,1]^2}$ is $\int_0^1 \int_0^1$. Then, $\rho(C)$ can be interpreted as the normalized average difference between the copula C and the independence copula Π.

Multivariate extensions of Spearman's rho and their estimation have been discussed, by Wolff (1980), Nelsen (1996), Joe (1990) and Schmid and Schmidt

(2007a). Motivated by equation (2.1), the following multivariate version of ρ can be derived

$$\begin{aligned}\rho_1(C) &= \frac{\int_{[0,1]^d} C(\mathbf{u})\,d\mathbf{u} - \int_{[0,1]^d} \Pi(\mathbf{u})\,d\mathbf{u}}{\int_{[0,1]^d} M(\mathbf{u})\,d\mathbf{u} - \int_{[0,1]^d} \Pi(\mathbf{u})\,d\mathbf{u}} \\ &= h_\rho(d)\left\{2^d \int_{[0,1]^d} C(\mathbf{u})\,d\mathbf{u} - 1\right\},\end{aligned}$$

where

$$h_\rho(d) = \frac{d+1}{(2^d - d - 1)}.$$

By using

$$\begin{aligned}\rho(C) &= \frac{\int_{[0,1]^2} C(u_1,u_2)\,du_1du_2 - \int_{[0,1]^2} \Pi(u_1,u_2)\,du_1du_2}{\int_{[0,1]^2} M(u_1,u_2)\,du_1du_2 - \int_{[0,1]^2} \Pi(u_1,u_2)\,du_1du_2} \\ &= \frac{\int_{[0,1]^2} u_1u_2dC(u_1,u_2) - \int_{[0,1]^2} u_1u_2d\Pi(u_1,u_2)}{\int_{[0,1]^2} u_1u_2dM(u_1,u_2) - \int_{[0,1]^2} u_1u_2d\Pi(u_1,u_2)},\end{aligned}$$

another multivariate version of Spearman's rho can be similarly defined, which is given by

$$\begin{aligned}\rho_2(C) &= \frac{\int_{[0,1]^d} \mathbf{u}dC(\mathbf{u}) - \int_{[0,1]^d} \mathbf{u}d\Pi(\mathbf{u})}{\int_{[0,1]^2} \mathbf{u}dM(\mathbf{u}) - \int_{[0,1]^d} \mathbf{u}d\Pi(\mathbf{u})} \\ &= h_\rho(d)\left\{2^d \int_{[0,1]^d} \mathbf{u}dC(\mathbf{u}) - 1\right\}.\end{aligned}$$

In particular, $\rho_1(C)$ and $\rho_2(C)$ are the same if the copula C is radially symmetric. Nelsen (1996) further considers the average of the two versions, i.e.,

$$\rho_3(C) = \frac{\rho_1(C) + \rho_2(C)}{2}.$$

All three measures satisfy *P3(a), P3(c), P3(d)*, *P4*, and *P7*. In addition, *P5* can be verified for $\rho_3(C)$, which, thus represents a multivariate measure of concordance according to Taylor (2007).

Statistical inference for $\rho_i(C)$, $i = 1, 2$, based on the empirical copula is investigated in Schmid and Schmidt (2007a). By replacing the copula C with the empirical copula $\widehat{C}_n$, we obtain the following non-parametric estimators for $\rho_i(C)$, $i = 1, 2$:

$$\begin{aligned}\widehat{\rho}_1\left(\widehat{C}_n\right) &= h_\rho(d)\left\{2^d \int_{[0,1]^d} \widehat{C}_n(\mathbf{u})\,d\mathbf{u} - 1\right\} \\ &= h_\rho(d)\left\{\frac{2^d}{n}\sum_{j=1}^{n}\prod_{i=1}^{d}\left(1 - \widehat{U}_{ij,n}\right) - 1\right\},\end{aligned}$$

and

$$\begin{aligned}\widehat{\rho}_2\left(\widehat{C}_n\right) &= h_\rho(d)\left\{2^d \int_{[0,1]^d} \mathbf{u}\,d\widehat{C}_n(\mathbf{u}) - 1\right\} \\ &= h_\rho(d)\left\{\frac{2^d}{n}\sum_{j=1}^{n}\prod_{i=1}^{d}\widehat{U}_{ij,n} - 1\right\}.\end{aligned}$$

Under the assumptions of the theorems 1.6.1 and 1.6.2, Schmid and Schmidt (2007a) proved that

$$\sqrt{n}\left(\widehat{\rho}_i\left(\widehat{C}_n\right) - \rho_i(C)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_i^2\right), \quad i = 1, 2,$$

where

$$\sigma_1^2 = 2^{2d} h_\rho(d)^2 \int_{[0,1]^d}\int_{[0,1]^d} E\left(\mathbb{G}_C(\mathbf{u})\,\mathbb{G}_C(\mathbf{v})\right) d\mathbf{u}\,d\mathbf{v},$$

$$\sigma_2^2 = 2^{2d} h_\rho(d)^2 \int_{[0,1]^d}\int_{[0,1]^d} E\left(\mathbb{G}_{\overline{C}}(\mathbf{u})\,\mathbb{G}_{\overline{C}}(\mathbf{v})\right) d\mathbf{u}\,d\mathbf{v},$$

and Gaussian processes $\mathbb{G}_C$ and $\mathbb{G}_{\overline{C}}$ as defined in the theorems 1.6.1 and 1.6.2. Asymptotic normality of $\widehat{\rho}_3\left(\widehat{C}_n\right)$ can analogously be established based on the joint weak convergence of the process $\sqrt{n}\left(\widehat{C}_n - C,\ \widehat{\overline{C}}_n - C\right)$. If the copula C is radially symmetric, it follows that $\sigma_1^2 = \sigma_2^2$. For a few copulas of simple form, the asymptotic variances can be explicitly computed, e.g. in the case of stochastic independence (i.e. $C = \Pi$) Schmid and Schmidt (2007a) obtain

$$\sigma_1^2 = \sigma_2^2 = \frac{(d+1)^2\left(3\,(4/3)^d - d - 3\right)}{3\left(1 + d - 2^d\right)^2}.$$

The asymptotic variances can consistently be estimated by a non-parametric bootstrap method (see Schmid and Schmidt, 2006). Quessy (2009) investigated statistical hypothesis tests for stochastic independence based on various multivariate versions of Spearman's rho with regard to their asymptotic relative efficiency.

2.3 Spearman's footrule

Spearman's footrule, named after Spearman (1906) is a non-parametric measure of association between two variables X_1 and X_2. This measure has been used in several research areas such as bioinformatics, genomics, information science, management science, litigation and aggregate rankings for search engines. Spearman's footrule is given by

$$\varphi_n = 1 - \frac{3}{n^2-1}\sum_{i=1}^{n} |p_i - q_i| \,,$$

where p_i and q_i denote the ranks of n observed values of two variates X_1 and X_2, respectively.

If X_1 and X_2 have copula C, then

$$\begin{aligned}\varphi(C) &= 1 - 3\int_{[0,1]^2} |u_1 - u_2|\, dC(u_1,u_2) \\ &= -2 + 6\int_0^1 C(t,t)\,dt.\end{aligned}$$

From the identity $|u_1 - u_2| = u_1 + u_2 + 2\min(u_1,u_2)$, Genest et al. (2010) got

$$\varphi_n = \frac{1}{n-1}\sum_{i=1}^{n} J\left(\frac{p_i}{n+1}, \frac{q_i}{n+1}\right) - \frac{2n+1}{n-1},$$

where $J(u_1,u_2) = 6\min(u_1,u_2)$. Genest et al. (2010) showed that φ_n is asymptotically unbiased estimator of $\varphi(C)$.

By assuming that the bivariate copula C admits continuous partial derivatives on $(0,1)$, Genest et al. (2010) showed that

$$\sqrt{n}\left(\varphi_n - \varphi(C)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_\varphi^2\right), \text{ as } n \to \infty,$$

where

$$\sigma_\varphi^2 = 36\int_{[0,1]^2} cov\left\{\mathbb{C}(s,s), \mathbb{C}(t,t)\right\} dsdt, \text{ with } \mathbb{C} = \sqrt{n}\left(\widehat{C}_n - C\right).$$

Various generalisations of Spearman's footrule have been proposed. Cifarelli et al. (1996) considered

$$\varphi_h(C) = h^{-1}\left(\int_0^1 h(|u_1 - u_2|)\, dC(u_1,u_2)\right),$$

where $h : [0,1] \to [0,1]$ is a strictly increasing and continuous function. If $h(t) = t^2$, we obtain Spearman's rho as special case. The asymptotic distribution of the empirical version of $\varphi_h(C)$ is studied by Cifarelli et al. (1996) and proved how to estimate its variance by the jackknife method. But, it is not clear how h should be chosen in practice.

Úbeda-Flores (2005) proposed another multivariate version of Spearman's footrule as follows

$$\varphi_d(C) = \frac{d+1}{d-1}\int_0^1 \left\{C(t,\dots,t) + \overline{C}(t,\dots,t)\right\} dt - \frac{2}{d-1},$$

where $\overline{C}$ is the df of $1-\mathbf{U}$ with $\mathbf{U} = (U_1,\dots,U_d)$ distributed as C. Úbeda-Flores (2005) proved $\varphi_d(C) = 0$ at independence and $\varphi_d(C) = 1$ at the Fréchet–Hoeffding upper bound.

For a random sample $(X_{11},\dots,X_{1d}),\dots,(X_{n1},\dots,X_{nd})$ from some continuous d-variate distribution and $(R_{11},\dots,R_{1d}),\dots,(R_{n1},\dots,R_{nd})$ the associated vectors of componentwise ranks, the empirical version of $\varphi_d(C)$ was given by Úbeda-Flores (2005) as follows

$$\varphi_{dn} = 1 - \frac{d+1}{d-1}\sum_{i=1}^{n}\frac{\max(R_{i1},\dots R_{id}) - \min(R_{i1},\dots R_{id})}{n^2-1}.$$

By assuming that the d-variate copula C admits continuous partial derivatives on $(0,1)^d$, Genest et al. (2010) showed that

$$\sqrt{n}\left(\varphi_{dn} - \varphi_d(C)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2_{\varphi d}\right), \text{ as } n \to \infty,$$

where

$$\sigma^2_{\varphi d} = \left(\frac{d+1}{d-1}\right)^2 \left\{\Gamma(D,D) + \overline{\Gamma}(D,D) + 2\sum_{A\subseteq D}(-1)^{|A|}\Gamma(A,D)\right\},$$

where for arbitrary $A, B \subseteq D$, one has

$$\Gamma(A,B) = \int_{[0,1]^2} cov\left\{\mathbb{C}(\mathbf{s}_A), \mathbb{C}(\mathbf{t}_B)\right\} dsdt,$$

and

$$\overline{\Gamma}(A,B) = \int_{[0,1]^2} cov\left\{\overline{\mathbb{C}}(\mathbf{s}_A), \overline{\mathbb{C}}(\mathbf{t}_B)\right\} dsdt,$$

with the process $\overline{\mathbb{C}}$ is defined in theorems 1.6.2, $|A|$ denote the cardinality of any set $A \subseteq D = \{1,\dots,d\}$ and $\mathbf{t}_A$ is the vector $(t_1,\dots,t_d)$ such that $t_v = t\mathbf{1}_{\{v\in A\})} + t\mathbf{1}_{\{v\notin A\})}$ for all $v \in \{1,\dots,d\}$, for example $\mathbf{t}_D = (t,\dots,t)$.

2.4 Kendall's tau

The non-parametric correlation coefficient or measure of association known as Kendall's tau was first discussed by Fechner and others about 1900, and was rediscovered, independently, by Kendall in 1938.
Let (X_1, X_2) and (Y_1, Y_2) be independent and identically distributed bivariate random vectors with df F. The population version, τ, of Kendall's tau is defined as the probability of concordance minus the probability of discordance. Then

$$\tau = \mathbb{P}\left(\left(X_1 - Y_1\right)\left(X_2 - Y_2\right) > 0\right) - \mathbb{P}\left(\left(X_1 - Y_1\right)\left(X_2 - Y_2\right) < 0\right).$$

If C is the bivariate copula of F, then

$$\tau\left(C\right) = 4\int_{[0,1]^2} C\left(u,v\right) dC\left(u,v\right) - 1. \tag{2.2}$$

Multivariate versions of Kendall's tau are discussed in Nelsen (1996), Joe (1990), and Taylor (2007). Formula (2.2) implies the following multivariate version:

$$\tau\left(C\right) = \frac{1}{2^{d-1}-1}\left\{2^d \int_{[0,1]^d} C\left(\mathbf{u}\right) dC\left(\mathbf{u}\right) - 1\right\}.$$

A natural non-parametric estimator of τ is given by

$$\begin{aligned} \widehat{\tau}\left(\widehat{C}_n\right) &= \frac{1}{2^{d-1}-1}\left\{2^d \int_{[0,1]^d} \widehat{C}_n\left(\mathbf{u}\right) d\widehat{C}_n\left(\mathbf{u}\right) - 1\right\} \\ &= \frac{1}{2^{d-1}-1}\left\{\frac{2^d}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\prod_{i=1}^{d}\mathbf{1}_{\left\{\widehat{U}_{ij,n} \leq \widehat{U}_{ik,n}\right\}} - 1\right\}, \end{aligned}$$

with empirical copula $\widehat{C}_n$. Non-parametric estimation and statistical inference for τ based on the empirical copula process is the focus of an ongoing work. For further non-parametric statistical analysis of Kendall's tau, which is also frequently considered in the context of tests for stochastic independence.

2.5 Blomqvist's beta

Blomqvist's beta or the medial correlation coefficient was introduced by Blomqvist (1950). Let X_1 and X_2 be two continuous rv's having medians $\widetilde{x}_1$ and $\widetilde{x}_2$, then the population version of Blomqvist's beta is given by

$$\beta = \mathbb{P}\left(\left(X_1 - \widetilde{x}_1\right)\left(X_2 - \widetilde{x}_2\right) > 0\right) - \mathbb{P}\left(\left(X_1 - \widetilde{x}_1\right)\left(X_2 - \widetilde{x}_2\right) < 0\right).$$

If X_1 and X_2 have copula C, then β can be expressed in terms of C as follows

$$\begin{aligned}\beta(C) &= 2\mathbb{P}\left(\left(X_1-\widetilde{x}_1\right)\left(X_2-\widetilde{x}_2\right)>0\right)-1=4C(1/2,1/2)\\ &= \frac{C(1/2,1/2)-\Pi(1/2,1/2)+\overline{C}(1/2,1/2)-\overline{\Pi}(1/2,1/2)}{M(1/2,1/2)-\Pi(1/2,1/2)+\overline{M}(1/2,1/2)-\overline{\Pi}(1/2,1/2)}. \end{aligned} \quad (2.3)$$

Blomqvist's beta can be interpreted as a normalized difference between the copula C and the independence copula at $(1/2,1/2)$. Various extensions of Blomqvist's beta to the multivariate case have been considered in Joe (1990), Úbeda-Flores (2005) and Schmid and Schmidt (2007c). The following multivariate version is motivated by equation (2.3)

$$\begin{aligned}\beta(C) &= \frac{C(\mathbf{1/2})-\Pi(\mathbf{1/2})+\overline{C}(\mathbf{1/2})-\overline{\Pi}(\mathbf{1/2})}{M(\mathbf{1/2})-\Pi(\mathbf{1/2})+\overline{M}(\mathbf{1/2})-\overline{\Pi}(\mathbf{1/2})}\\ &= h_\beta(d)\left\{C(\mathbf{1/2})-\overline{C}(\mathbf{1/2})-2^{1-d}\right\}, \end{aligned} \quad (2.4)$$

where $h_\beta(d)=2^{d-1}/\left(2^{d-1}-1\right)$ and $\mathbf{1/2}=(1/2,\ldots,1/2)$. The multivariate version of β satisfies the properties *P3(a), P3(c)* and *P5*. Further, $\beta(C)$ equals the average of pairwise Blomqvist's beta in dimension $d=3$. Note that if the copula C is radially symmetric, i.e. $C=\overline{C}$, the expression in (1.8) reduces to

$$\frac{2^d C(\mathbf{1/2})-1}{2^{d-1}-1},$$

which coincides with the multivariate version originally introduced in Nelsen (1996). Schmid and Schmidt (2007c) studied more general extensions of Blomqvist's beta, which measure the association in the tail region of the copula and which include $\beta(C)$ as defined in (2.4).

A natural estimator for $\beta(C)$ is obtained by replacing the copula C and the survival function $\overline{C}$ in (2.4) with their empirical counterparts, i.e.

$$\widehat{\beta}\left(\widehat{C}_n\right)=h_\beta(d)\left\{\widehat{C}_n(\mathbf{1/2})-\widehat{\overline{C}}_n(\mathbf{1/2})-2^{1-d}\right\}.$$

Under weak assumptions on the copula C and the survival function $\overline{C}$, Schmid and Schmidt (2007c) established the asymptotic normality and consistency of $\widehat{\beta}\left(\widehat{C}_n\right)$. Namely, if the i-th partial derivatives D_iC and $D_i\overline{C}$ exist and are continuous at the point $\mathbf{1/2}$, we have

$$\sqrt{n}\left(\widehat{\beta}\left(\widehat{C}_n\right)-\beta(C)\right)\xrightarrow{\mathcal{D}}\mathcal{N}\left(0,\sigma^2\right),$$

where

$$\sigma^2 = h_\beta (d)^2 E\left(\left\{\mathbb{G}_C (\mathbf{1/2}) + \mathbb{G}_{\overline{C}} (\mathbf{1/2})\right\}^2\right),$$

and Gaussian processes $\mathbb{G}_C$ and $\mathbb{G}_{\overline{C}}$ as defined in the aforementioned theorems. One main advantage of Blomqvist's beta over other copula-based measures is that the asymptotic variance of its estimator can explicitly be calculated whenever the copula and its partial derivatives are of explicit form see Schmid and Schmidt (2007c) for related examples. For example if $C = \Pi$, we have

$$\sigma^2 = \frac{1}{2^{d-1} - 1}.$$

2.6 Gini's gamma

Gini's gamma was introduced by Gini (1910). If p_i and q_i denote the ranks in a sample of size n of two continuous rv's X and Y respectively, then

$$\gamma = \frac{1}{[n^2/2]} \left\{ \sum_{i=1}^{n} |p_i + q_i - n - 1| - \sum_{i=1}^{n} |p_i + q_i| \right\}.$$

If C is the bivariate copula of F, then

$$\begin{aligned} \gamma (C) &= 2 \int_{[0,1]^2} (|u + v - 1| - |u - v|)\, dC (u, v) \\ &= 4 \int_{[0,1]^2} \{M (u, v) + W (u, v)\}\, dC (u, v) - 2. \end{aligned} \tag{2.5}$$

A multivariate extension of Gini's gamma has been considered by Behboodian et al. (2007). By defining the function $A(\mathbf{u}) = \{M(\mathbf{u}) + W(\mathbf{u})\}/2$, $\mathbf{u} \in [0, 1]^d$, with corresponding survival function $\overline{A}$, the expression in (2.5) is equal to

$$\begin{aligned} \gamma (C) = 4 \Bigg(&\int_{[0,1]^2} \left\{A (u, v) + \overline{A} (u, v)\right\} dC (u, v) \\ &- \int_{[0,1]^2} \left\{A (u, v) + \overline{A} (u, v)\right\} d\Pi (u, v) \Bigg), \end{aligned}$$

as $A (u, v) + \overline{A} (u, v) = 1 - u - v + 2A (u, v)$ for every $(u, v) \in [0, 1]^2$. A multivariate version of Gini's gamma is then defined as

$$\gamma (C) = \frac{1}{b (d) - a (d)} \left(\int_{[0,1]^d} \left\{A (\mathbf{u}) + \overline{A} (\mathbf{u})\right\} dC (\mathbf{u}) - a (d) \right), \tag{2.6}$$

where normalization constants $a(d)$ and $b(d)$ of the form

$$\begin{aligned} a(d) &= \int_{[0,1]^d} \left\{ A(\mathbf{u}) + \overline{A}(\mathbf{u}) \right\} d\Pi(\mathbf{u}) \\ &= \frac{1}{d+1} + \frac{1}{2(d+1)!} + \sum_{i=0}^{d} (-1)^d \binom{d}{i} \frac{1}{2(i+1)!}, \end{aligned}$$

and

$$\begin{aligned} b(d) &= \int_{[0,1]^d} \left\{ A(\mathbf{u}) + \overline{A}(\mathbf{u}) \right\} dM(\mathbf{u}) \\ &= 1 - \sum_{i=1}^{d-1} \frac{1}{4i}. \end{aligned}$$

From the above definition we have $\gamma(C) = 0$ if $C = \Pi$ and $\gamma(C) = 1$ if $C = M$, then, *P3(a)* and *P3(c)* hold. In the context of multivariate measures of concordance, Taylor (2007) discussed another multivariate generalization. Behboodian et al. (2007) also provide a sample version for $\gamma(C)$ as defined in (2.6). In the bivariaste case, a sample version based on the empirical copula is considered in Nelsen (2006) which coincides with the traditional sample version of Gini's gamma. The latter plays an important role in the context of tets for stochastic independence and has been discussed by many authors. Under suitable conditions, Cifarelli et al. (1996) established the asymptotic normality of a generalized class of bivariate statistics including Gini's gamma. An asymptotic theory for $d \geq 3$ is not yet available to our knowledge.

2.7 Hoeffding's phi-square

Hoeffding (1940) was the first to consider measures of association based on a L_p-type distance between a copula C and the independence copula Π. His work focuses on $p = 2$ and was extended by Schweizer and Wolff (1980) who introduce L_1- and L_∞-based measures of bivariate association.

The Hoeffding's phi-square (Hoeffding, 1940) measure of association between the components of the two-dimensional random vector $\mathbf{X}$ with copula C is defined by

$$\Phi^2 = 90 \int_{[0,1]^2} \left\{ C(u,v) - uv \right\}^2 dudv.$$

Gaißer et al. (2010) introduced a multivariate version of Hoeffding's Φ^2 defined by

$$\Phi^2(C) = h_2(d) \int_{[0,1]^d} \{C(\mathbf{u}) - \mathit{\Pi}(\mathbf{u})\}^2 \, d\mathbf{u}, \tag{2.7}$$

where

$$h_2(d) = \left(\frac{2}{(d+1)(d+2)} - \frac{1}{2^d} \frac{d!}{\prod_{i=0}^{d} \left(i + \frac{1}{2}\right)} + \left(\frac{1}{3}\right)^d \right)^{-1}.$$

Due to their structure, all L_p-distance-based measures share a set of common properties. Irrespective of the particular choice of p, the measures satisfy *P1* and *P2*. They further possess the strong property that they are zero if and only if $\mathit{\Pi}$ is the copula of $\mathbf{X}$, thus *P3(a)* and *P3(b)* hold. Normalizing by means of the upper Fréchet-Hoeffding bound, *P3(a)* is assured. Consider a multivariate normal random vector $\mathbf{X}$ for which all pairwise correlations ρ_{ij} of X_i and X_j are either non-negative or non-positive. Analogously to Wolff (1980), it can be shown that all L_p-distance-based measures are a strictly increasing function of the absolute value of each of the pairwise correlations. In general, the L_p-distance-based measures further satisfy *P4*, *P6* and *P4*.

In the particular case that an independent component X_{d+1} is added to a d-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_d)$ with copula C, $\Phi^2(X_1, \ldots, X_{d+1})$ can be expressed as a function of the d-dimensional measure:

$$\Phi^2(X_1, \ldots, X_{d+1}) = \frac{1}{3} \frac{h_2(d+1)}{h_2(d)} \Phi^2(X_1, \ldots, X_d) < \Phi^2(X_1, \ldots, X_d).$$

Thus, criterion *P5* is satisfied, meaning that an independent variable X_{d+1} reduces overall association in the enlarged vector.

A non-parametric estimator for Φ^2 is obtained by replacing the copula C, in formula (2.7), by the empirical copula $\widehat{C}_n$, that is

$$\begin{aligned} \Phi^2\left(\widehat{C}_n\right) &= h_2(d) \int_{[0,1]^d} \left\{\widehat{C}_n(\mathbf{u}) - \mathit{\Pi}(\mathbf{u})\right\}^2 d\mathbf{u} \\ &= h_2(d) \left\{ \left(\frac{1}{n}\right)^2 \sum_{j=1}^{n} \sum_{k=1}^{n} \prod_{i=1}^{d} \left(1 - \max\left\{\widehat{U}_{ij,n}, \widehat{U}_{ik,n}\right\}\right) \right. \\ &\qquad \left. - \frac{2}{n} \left(\frac{1}{2}\right)^d \sum_{j=1}^{n} \prod_{i=1}^{d} \left(1 - \widehat{U}_{ij,n}^2\right) + \left(\frac{1}{3}\right)^d \right\}. \end{aligned}$$

The estimate is therefore easy to calculate even for large d. A bias reduction for $\Phi^2\left(\widehat{C}_n\right)$ has been suggested in Gaißer et al. (2010). Simulations have shown that the estimator works well for various copula families. Obviously, we obtain an estimator for the alternative measure Φ by $\Phi\left(\widehat{C}_n\right) = +\sqrt{\Phi^2\left(\widehat{C}_n\right)}$.

The asymptotic theory for $\Phi^2\left(\widehat{C}_n\right)$ is derived from the asymptotic behaviour of the empirical copula process $\sqrt{n}\left(\widehat{C}_n(\mathbf{u}) - C(\mathbf{u})\right)$ as provided by theorem 1.6.1. Then, asymptotic normality of the estimator $\Phi^2\left(\widehat{C}_n\right)$ can be derived by means of the functional delta method. Under the assumptions of theorem 1.6.1 and the additional presumption that $C \neq \Pi$ it follows that

$$\sqrt{n}\left(\Phi^2\left(\widehat{C}_n\right) - \Phi^2(C)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2_{\Phi^2}\right),$$

where

$$\sigma^2_{\Phi^2} = (2h_2(d))^2 \int_{[0,1]^d}\int_{[0,1]^d} E\left(\{C(\mathbf{u}) - \Pi(\mathbf{u})\}\,\mathbb{G}_C(\mathbf{u})\,\mathbb{G}_C(\mathbf{v})\,\{C(\mathbf{v}) - \Pi(\mathbf{v})\}\right) d\mathbf{u}d\mathbf{v}.$$

The proof is given in Gaißer et al. (2010). The above assumption $C \neq \Pi$ guarantees that the limiting rv is non-degenerate as implied by the form of the variance $\sigma^2_{\Phi^2}$, the limiting behaviour of $\Phi^2\left(\widehat{C}_n\right)$ in case $C = \Pi$ is considered in Gaißer et al. (2010).

2.8 Schweizer and Wolff's sigma

Wolff (1980) generalized the L_1-distance-based measure of Schweizer and Wolff (1981) to the multivariate case. It is defined by

$$\sigma(C) = h_1(d) \int_{[0,1]^d} |C(\mathbf{u}) - \Pi(\mathbf{u})|\, d\mathbf{u},$$

where the normalizing factor $h_1(d)$ is given by

$$h_1(d) = \left(\frac{1}{d+1} - \frac{1}{2^d}\right).$$

The measure satisfies *P3*(d). With regard to *P8*, an explicit form of the function is derived in Schweizer and Wolff (1981) for the bivariate case. Except for taking the absolute value, this functional form matches the one that can be derived for

Spearman's ρ, illustrating that the two measures are closely related. A similar calculation shows that σ satisfies *P4*,

$$\sigma\left(X_1,\ldots,X_{d+1}\right)=\frac{1}{2}\frac{h_1\left(d+1\right)}{h_1\left(d\right)}\sigma\left(X_1,\ldots,X_d\right)<\sigma\left(X_1,\ldots,X_d\right).$$

The estimation of $\sigma\left(C\right)$ has not yet been considered in detail. Various estimators for this measure can be obtained by replacing C in the defining formulas with the empirical copula $\widehat{C}_n$. However, no explicit expressions as e.g. for $\Phi^2\left(\widehat{C}_n\right)$ are available and the estimate must be determined numerically, which can be demanding for large dimension d.

2.9 L_∞-distance-based measure

A L_∞-distance-based multivariate measure is derived in Wolff (1980) and investigated in detail by Fernández-Fernández and González-Barrios (2004). This measure is defined by

$$L_\infty(C)=h_\infty\left(d\right)\sup_{\mathbf{u}\in[0,1]^d}\left|C\left(\mathbf{u}\right)-\mathit{\Pi}\left(\mathbf{u}\right)\right|.$$

Fernández-Fernández and González-Barrios (2004) do not normalize the population version of the measure. An addition a normalization factor $h_\infty\left(d\right)$ in order to assure comparability with alternative measures, which is given by

$$h_\infty\left(d\right)=\left(\left(\frac{1}{d}\right)^{\frac{1}{d-1}}\left(1-\frac{1}{d}\right)\right)^{-1}.$$

Wolff (1980) proved that this measure satisfies all normalization criteria except for *P3*(d). This is due to the fact that there exist other copulas than the upper Frechet-Hoeffding bound for which the measure attains its maximal value. With regard to *P8*, an explicit form of the function is derived in Schweizer and Wolff (1981) for the bivariate case. With respect to the addition of further components, the measure behaves differently than the measures discussed before. It generally holds that

$$0\le L_\infty(X_1,X_2)\le L_\infty(X_1,X_2,X_3)\le\cdots\le L_\infty(X_1,\ldots,X_d).$$

In particular, the measure satisfies *P7* if an independent component is added to a d-dimensional random vector $\mathbf{X}$, i.e.

$$L_\infty(X_1,\ldots,X_{d+1})=L_\infty(X_1,\ldots,X_d).$$

The estimation of $L_\infty(C)$ can analogously be performed by replacing all df's with their empirical counterparts

$$L_\infty(\widehat{C}_n) = \frac{\sup_{\mathbf{u}\in[0,1]} \left| \widehat{C}_n(\mathbf{u}) - \prod_{i=1}^d U_n(u_i) \right|}{\max_{0\le i\le n} \left\{ \frac{i}{n} - \left(\frac{i}{n}\right)^d \right\}},$$

where U_n denotes the, univariate, df of a uniformly distributed rv on the set $\{1/n, ..., n/n\}$. In order to reduce bias, the independence copula is replaced by its discretized version $\prod_{i=1}^d U_n(u_i)$. For the unnormalized statistic, Fernández-Fernández and González-Barrios (2004) proved a strong law of large numbers. While, the explicit asymptotic theory for $L_\infty(\widehat{C}_n)$ is not available yet.

2.10 Tail dependence

Tail dependence is used in the modeling and measurement of association between extreme values such as extremely negative asset returns and plays an important role in financial theory, see Joe (1997). Let X_1 and X_2 be two rv's with bivariate df F and continuous univariate margins F_1, F_2 and let u be a threshold value, then the upper tail coefficient, λ_U, is defined as

$$\lambda_U = \lim_{u\uparrow 1} \mathbb{P}\left(F_1(X_1) > u \,|F_2(X_2) > u\right),$$

and the lower tail coefficient, λ_L, is defined as

$$\lambda_L = \lim_{u\downarrow 0} \mathbb{P}\left(F_1(X_1) \le u \,|F_2(X_2) \le u\right),$$

provided that the above limits exists. Note that $0 \le \lambda_L, \lambda_U \le 1$. If $\lambda_U \in (0,1]$, X_1 and X_2 are asymptotically dependent in the upper tail and if $\lambda_U = 0$, X_1 and X_2 are asymptotically independent in the upper tail.
The tail dependence coefficients λ_U and λ_L expressed in terms of copula function C as follows

$$\lambda_U(C) = \lim_{u\uparrow 1} \frac{1 - 2u + C(u,u)}{1-u},$$

and

$$\lambda_L(C) = \lim_{u\downarrow 0} \frac{C(u,u)}{u}.$$

The natural non-parametric estimator for λ_U and λ_L from a random sample $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ of $\mathbf{X}$ is

$$\widehat{\lambda}_{L,k,n} = \frac{n\widehat{C}_n\left(\frac{k}{n}, \frac{k}{n}\right)}{k},$$

with suitably chosen parameter $k = k(n)$. The statistical properties of $\widehat{\lambda}_{L,k,n}$ have been investigated by several authors using techniques from extreme value theory (see chapter 3 in this thesis).

Various generalisations of tail dependence coefficients proposed. Frahm (2006) considered the following multivariate generalization of λ_L

$$\lambda_L(C) = \lim_{u \downarrow 0} \frac{C(u\mathbf{1})}{1 - \overline{C}(u\mathbf{1})}.$$

Schmid and Schmidt (2007a, 2007b) proposed, another multivariate generalizations of λ_L which are based on conditional versions of Spearman's rho as follows

$$\rho_L(C) = \lim_{p \downarrow 0} \rho_p(C) = \lim_{p \downarrow 0} \frac{d+1}{p^{d+1}} \int_{[0,1]^d} C(\mathbf{u})\, d\mathbf{u},$$

if the limit exists. The natural estimator for $\rho_L(C)$ is

$$\widehat{\rho}_L\left(\widehat{C}_n\right) = \widehat{\rho}_{\frac{k}{n}}\left(\widehat{C}_n\right)$$

with appropriate value $k = k(n)$, chosen by the statistician, and

$$\widehat{\rho}_p\left(\widehat{C}_n\right) = \left\{\frac{1}{n}\sum_{j=1}^{n}\prod_{i=1}^{d}\left(p - \widehat{U}_{ij,n}\right)^+ - \left(\frac{p^2}{2}\right)^d\right\} / \left\{\frac{p^{d+1}}{d+1} - \left(\frac{p^2}{2}\right)^d\right\},$$

where, $x^+ = \max(x, 0)$.

Under suitable conditions, Schmid and Schmidt (2007b) proved that

$$\sqrt{n}\left(\widehat{\rho}_p\left(\widehat{C}_n\right) - \rho_L(C)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_\Lambda^2\right), \text{ as } n \to \infty,$$

where

$$\sigma_\Lambda^2 = (d+1) \int_{[0,1]^d} \mathbb{G}_\Lambda(\mathbf{u})\, d\mathbf{u},$$

with $\mathbb{G}_\Lambda$ is a centered tight continuous Gaussian random field (see Schmid and Schmidt, 2007b). We can estimate the asymptotic variance σ_Λ^2 by using bootstrap techniques. The tail dependence measures verify the following properties *P1, P3(a), P3(c) and P5.*

Part II

Extreme Risks

Chapter 3

Univariate heavy-tailed distributions and extreme value theory

In this chapter, we give the basic definitions and properties of heavy-tailed distributions and EVT. Heavy-tailed distribution are often used when, in a certain phenomenon, the probability of events is relatively big. These distributions was first noticed by Pareto in 1896 in income distribution. After that, many researchers have argued that such distributions can be found not only in economics, but also in other domains such as finance, insurance and network topology. EVT is concerned with probabilistic and statistical questions related to very high or very low values in sequences of rv's and in stochastic processes. The subject has a rich mathematical theory and also a long tradition of applications in a variety of areas. Among many excellent books on the subject, Embrechts et al. (1997) give a comprehensive survey of the mathematical theory with an orientation toward applications in insurance and finance.

3.1 Heavy-tailed distributions

3.1.1 Definitions

Let X be an insured risk, a non-negative rv defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with continuous df $F = \mathbb{P}(X \leq x)$. We define the tail function F

by

$$\overline{F}(x) = 1 - F(x).$$

The tail of a distribution represents probability values for large values of the variable. When large values of the variable appear in a data set, their probabilities of occurrence are not zero.

Definition 3.1.1 *A distribution $F \geq 0$ is heavy-tailed if and only if*

$$\lim_{x\to\infty} e^{\lambda x}\overline{F}(x) = \infty \text{ for all } \lambda > 0.$$

Definition 3.1.2 *A distribution F is heavy-tailed if and only if it has no exponential moment, i.e.,*

$$\int_0^\infty e^{\lambda x} dF(x) = \infty \text{ for all } \lambda > 0,$$

Definition 3.1.3 *A distribution F is said to have a heavy tail if and only if*

$$\overline{F}(x) \sim x^{-\alpha} \text{ as } x \to \infty,$$

where the parameter $\alpha > 0$ is called the tail index.

Remark 3.1.1 *Any heavy-tailed distribution has right-unbounded support $\overline{F}(x) > 0$ for all x.*

3.1.2 Examples of heavy-tailed distributions

We mention, among other, the following distributions:

1. **Pareto Distribution on $\mathbb{R}_+$**

   This distribution has tail function $\overline{F}$ given by

   $$\overline{F}(x) = \left(\frac{\kappa}{x+\kappa}\right)^\alpha,$$

   for some scale parameter $\kappa > 0$ and shape parameter $\alpha > 0$. It is clear that we have $\overline{F}(x) \sim (x/\kappa)^{-\alpha}$ as $x \to \infty$, for this reason the Pareto distributions are sometimes referred to as the power law distributions. The Pareto distribution has all moments of order $p < \alpha$ finite, while all moments of order $p \geq \alpha$ are infinite.

2. **Burr distribution on $\mathbb{R}_+$**

   This has tail function $\overline{F}$ given by

   $$\overline{F}(x) = \left(\frac{\kappa}{x^{\tau}+\kappa}\right)^{\alpha},$$

   for parameters $\alpha, \kappa, \tau > 0$. It is clear that we have $\overline{F}(x) \sim \kappa^{\alpha} x^{-\tau\alpha}$ as $x \to \infty$, then the Burr distribution is similar in its tail to the Pareto distribution, of which it is otherwise a generalisation. The Burr distribution has all moments of order $p < \tau\alpha$ finite, while all moments of order $p \geq \tau\alpha$ are infinite.

3. **Weibull distribution on $\mathbb{R}_+$**

   This distribution has tail function $\overline{F}$ given by

   $$\overline{F}(x) = \exp\left\{-\left(x/\kappa\right)^{\alpha}\right\},$$

   for some scale parameter $\kappa > 0$ and shape parameter $\alpha > 0$. This is a heavy-tailed distribution if and only if $\alpha < 1$. All moments of the Weibull distribution are finite

4. **Cauchy distribution on $\mathbb{R}$**

   This distribution has tail function $\overline{F}$ given by

   $$\overline{F}(x) = \frac{1}{2} - \frac{\arctan x}{\pi},$$

   we have $\overline{F}(x) \sim (\pi x)^{-1}$ as $x \to \infty$, its tail goes to zero like the power function x^{-1}. All moments are infinite.

5. **Log-normal distribution on $\mathbb{R}_+^*$**

   This is most easily given by its density function f where

   $$f(x) = \frac{1}{\sigma x\sqrt{2\pi}} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\},$$

   for parameters σ and $\mu > 0$. Then, the tail of the lognormal distribution is

   $$\overline{F}(x) = \overline{\Phi}\left(\frac{\log x - \mu}{\sigma}\right).$$

   where $\overline{\Phi}$ is the tail of the standard normal rv. All moments of the lognormal distribution are finite.

There exist important classes of heavy-tailed distributions that are regularly varying function, intermediate regularly varying and subexponential distribution. For more detail on these distributions see, Foss et al. (2011).

Definition 3.1.4 *(Dominated-varying distribution)*
We say that F is a dominated-varying distribution if there exists $c > 0$ such that

$$\overline{F}(2x) \geq c\overline{F}(x) \quad \text{for all} \quad x.$$

Definition 3.1.5 *(Intermediate regularly varying)*
A distribution F on $\mathbb{R}$ is called intermediate regularly varying if

$$\lim_{\varepsilon \downarrow 0} \lim_{x \to \infty} \inf \frac{\overline{F}(x(1+\varepsilon))}{\overline{F}(x)} = 1.$$

Definition 3.1.6 *(Subexponential distribution)*
Let (X_n) be iid positive rv's with df F with support $(0, \infty)$, the df F is a subexponential distribution if

$$\lim_{x \to \infty} \frac{\overline{F^{n*}}(x)}{\overline{F}(x)} = n, \quad \text{for } n \geq 2,$$

where $\overline{F^{n*}}(x) = \mathbb{P}(X_1 + \cdots + X_n > x)$, the tail of the n-fold convolution of F.

Remark 3.1.2 *Any regularly varying distribution is intermediate regularly varying and any intermediate regularly varying distribution.*

3.1.3 Regularly varying functions

In the heavy-tail analysis, one of the most important classes of distributions is the regular variation class. This class has been widely used in modeling heavy-tailed phenomena. The notion of regular variation was discovered by Karamata (1930). In this subsection, we introduce these functions with some of their most important properties. For further details on theory of regular variation and Π-variation, we refer to Bingham et al. (1987) and Geluk et al. (1987). For the connection between regular variation and EVT, we refer to de Haan (1970).

Definition 3.1.7 *(Regularly varying function)*

- *A Lebesgue measurable function* $h : \mathbb{R}_+ \to \mathbb{R}_+$ *is regularly varying at infinity with index* $\rho \in \mathbb{R}$, *notation* $h \in \mathcal{RV}_\rho$, *if and only if for any* $x > 0$

$$\lim_{t\to\infty} \frac{h(xt)}{h(t)} = x^\rho. \tag{3.1}$$

The number ρ *in the above definition is called the index of regular variation. A function satisfying (1.1) with* $\rho = 0$ *is called slowly varying at infinity.*

- *We say that* h *is regularly varying at* 0 *with index* $\rho \in \mathbb{R}$, *notation* $h \in \mathcal{RV}_\rho^0$, *if and only if function* $h(1/x)$ *is regularly varying at infinity with index* $(-\rho)$. *In other words, regular variation at* 0 *is defined by replacing* $t \to \infty$ *by* $t \to 0$ *in (3.1).*

- h *is regularly varying at any point* $a > 0$, *notation* $h \in \mathcal{RV}_\rho^a$, *if* $h(a - 1/x) \in \mathcal{RV}_\rho$.

Example 3.1.1 *The functions* x^ρ, $x^\rho \log(1+x)$ *and* $(x \log(1+x))^\rho$ *are* $\mathcal{RV}_\rho$. *The functions positive constants, logarithms and iterated logarithms are slowly varying.at infinity with index* ρ. *The functions* $2 + \sin x$ *and* $\exp(\log x)$ *are not regularly varying.*

If one only assumes that the limit in (3.1) exists and is positive for all $x > 0$, then it can be shown that the limit is necessarily of the form x^ρ for some $\rho \in \mathbb{R}$ (see Embrechts et al., 1997). The following (immediate) result is about writing regularly varying functions in terms of slowly varying ones.

Proposition 3.1.1 *(Regular and slow variations)*
$h \in \mathcal{RV}_\rho$ *if and only if* $h(x) = x^\rho \ell(x)$, *where* $\ell \in \mathcal{RV}_\rho^0$.

The following theorems are used to restate the definition of regular variation and to introduce a new concept called Π-variation.

Theorem 3.1.1 *If* $h : (0, \infty) \to \mathbb{R}_+$ *is measurable such that*

$$\lim_{t\to\infty} \frac{h(tx) - h(t)}{a(t)},$$

exists and is not constant, where $x > 0$ *and* a *is a positive function, then*

$$\lim_{t\to\infty} \frac{h(tx) - h(t)}{a(t)} = c\frac{x^\rho - 1}{\rho},$$

for some $\rho \in \mathbb{R}$ *and* $c \neq 0$*, with the convention that the right hand side reads* $c \log x$ *if* $\rho = 0$.

Theorem 3.1.2 *(Restatement of regular variation)*
Assume that theorem 3.1.1 holds with $\rho \neq \mathbb{R}$ *and* $c > 0$.

1. *If* $\rho > 0$*, then* $f \in \mathcal{RV}_\rho$.
2. *If* $\rho < 0$*, then* $\lim_{t\to\infty} h(t)$ *exists and* $\left(\lim_{t\to\infty} h(t) - h(x)\right) \in \mathcal{RV}_\rho$.

Definition 3.1.8 *(*Π*-varying function)*
A Lebesgue measurable function $h : \mathbb{R}_+ \to \mathbb{R}_+$ *is* Π*-varying at infinity with auxiliary function* $a > 0$*, notation* $h \in \Pi$*, if*

$$\lim_{t\to\infty} \frac{h(tx) - h(t)}{a(t)} = \log x, \ x > 0.$$

h *is said to be* Π*-varying at 0, notation* $h \in \Pi^0$*, if* $h(1/x)$ *is* Π*-varying at infinity.*

Some of the basic properties of regularly varying functions are given in the following results.

Proposition 3.1.2 *(Proprieties of regularly varying functions)*

(i) If $f \in RV_\rho$*, then* $\log f(t) / \log t \to \rho$ *as* $t \to \infty$*. This implies*

$$\lim_{t\to\infty} f(t) = \begin{cases} 0 & \text{if } \rho < 0, \\ \infty & \text{if } \rho > 0. \end{cases}$$

(ii) If $f \in RV_\rho$ *and* $g \in RV_\beta$ *then* $f + g \in RV_\zeta$ *where* $\zeta = max(\rho, \beta)$*.If moreover* $\lim_{t\to\infty} g(t) = \infty$*, then the composition* $f \circ g \in RV_{\rho\beta}$.

(iii) If $f \in RV_\rho$*, then* $f^k \in RV_{\rho k}$.

(iv) If $f \in RV_\rho$ *with* $\rho > 0$ $(\rho < 0)$ *then* f *is asymptotically equivalent to a strictly increasing (decreasing) differentiable function* g *with derivative* $g' \in RV_{\rho-1}$ *if* $\rho > 0$ *and* $-g' \in RV_{\rho-1}$ *if* $\rho < 0$.

(v) *If* $f \in RV_\rho$ *with* $\rho > 0$ *and* f *is increasing, then the inverse function* $f^{-1} \in RV_{1/\rho}$.

(vi) *If* $f \in RV_\rho$ *is integrable on finite intervals of* $\mathbb{R}_+$ *with* $\rho \geq -1$ *then* $\int_0^t f(s)ds \in RV_{\rho+1}$. *If* $f \in RV_\rho$ *with* $\rho < 0$, *then* $\int_t^\infty f(s)ds$ *exists for* t *sufficiently large and is regularly varying with exponent* $\rho + 1$.

(vii) *Suppose that* $f \in RV_\rho$. *If* $\delta_1, \delta_2 > 0$ *are arbitrary, there exists* $t_0 = t_0(\delta_1, \delta_2)$ *such that for* $t \geq t_0, tx \geq t_0$,

$$(1-\delta_1)\, x^\rho \min\left(x^{\delta_2}, x^{-\delta_2}\right) < \frac{f(tx)}{f(t)} < (1+\delta_1)\, x^\rho \max\left(x^{\delta_2}, x^{-\delta_2}\right).$$

Note that conversely, if f satisfies the above property, then $f \in RV_\rho$.

(viii) *If* $f \in RV_\rho$ *with* $\rho \geq 0$ *and* $f(t) = f(t_0) + \int_{t_0}^t \varphi(s)ds$ *for* $t \geq t_0$ *with* φ *is monotone, then*

$$\lim_{t\to\infty} \frac{t\varphi(t)}{f(t)} = \rho.$$

Hence in case $\rho > 0$ *we have* $\varphi \in RV_{\rho-1}$. *Moreover, if* $\rho \leq 0$, *and* $\int_t^\infty \varphi(s)ds < \infty$ *with* φ *is non-increasing, then*

$$\lim_{t\to\infty} \frac{t\varphi(t)}{f(t)} = -\rho.$$

Hence in case $\rho < 0$ *we have* $\varphi \in RV_{\rho-1}$.

Theorem 3.1.3 *(Uniform convergence)*
If $f \in \mathcal{RV}_\rho$, *then for* $0 < a \leq b < \infty$ *relation (3.1) holds uniformly for*

(i) $x \in [a, b]$ *if* $\rho = 0$.
(ii) $x \in [0, b]$ *if* $\rho > 0$.
(iii) $x \in [a, \infty]$ *if* $\rho < 0$.

Theorem 3.1.4 *(Karamata representation)*
If $h \in \mathcal{RV}_\rho$, *then for* $x \geq A$

$$h(x) = c(x) \exp\left\{ \int_A^x \frac{\varepsilon(t)}{t} dt \right\},$$

where $A > 0$, c *and* ε *are measurable functions whith*

$$\lim_{x \to \infty} c(x) = c_0 \in (0, +\infty) \quad \textit{and} \quad \lim_{x \to \infty} \varepsilon(x) = \rho.$$

The following result, known as Karamata's theorem, says that one can take slowly varying functions out of integrals.

Theorem 3.1.5 *(Karamata's theorem)*
Let $l \in \mathcal{RV}_0$.*be locally bounded in* (A, ∞) *for some* $A > 0$. *Then*
(*i*) *for* $\beta > -1$

$$\int_A^x t^\beta l(t) dt \sim (\beta + 1)^{-1} x^{\beta+1} l(x) \quad \textit{as } x \to \infty.$$

(*ii*) *for* $\beta < -1$

$$\int_x^\infty t^\beta l(t) dt \sim -(\beta + 1)^{-1} x^{\beta+1} l(x) \quad \textit{as } x \to \infty.$$

Finally, some of the results that are useful for the theory of extreme values are summarized in the following proposition.

Proposition 3.1.3 *(Regular variation for distribution tails)*
Suppose that F *is a continuous df (with pdf* f*) such that* $F(x) < 1$ *for all* $x \geq 0$.
(*i*) *If* $\lim_{x \to \infty} x f(x) / \overline{F}(x) = \rho > 0$, *then* $f \in \mathcal{RV}_{-1-\rho}$ *and consequently* $\overline{F} \in \mathcal{RV}_{-\rho}$.
(*ii*) *If* $f \in \mathcal{RV}_{-1-\rho}$ $(\rho > 0)$, *then* $\lim_{x \to \infty} x f(x) / \overline{F}(x) = \rho$. *and consequently* $\overline{F} \in \mathcal{RV}_{-\rho}$. *The latter statement also holds if* $\overline{F} \in \mathcal{RV}_{-\rho}$ *and* f *is ultimately monotone.*
(*iii*) *If* X *is a non-negative rv with distribution tail* $\overline{F} \in \mathcal{RV}_{-\rho}$ $(\rho > 0)$, *then*

$$\begin{aligned} E(X^p) &< \infty \textit{ if } p < \rho, \\ E(X^p) &= \infty \textit{ if } p \geq \rho. \end{aligned}$$

(*iv*) *If* $\overline{F} \in \mathcal{RV}_{-\rho}$ $(\rho > 0)$, *then for* $\upsilon \geq \rho$

$$\lim_{x \to \infty} \frac{x^\upsilon \overline{F}(x)}{\int_\upsilon^x t^\upsilon dF(t)} = \frac{\upsilon - \rho}{\rho}.$$

3.2 Extreme value theory

EVT is a statistical and theoretical framework, which deals with modelling the behaviour of sample extremes, such as the sample minimum and the sample maximum. The behaviour of such order statistics may be assessed by their exact df or by their limiting df, the asymptotic df, if we increase the sample size towards infinity.

3.2.1 Order statistics

In EVT, the order statistics are very instrumental because they, more precisely the upper ones, provide information on the distribution (right) tail. Despite the fact that the definition of order statistics does not require common distribution nor independence of the $X_i's$ we will only consider the case where the $X_i's$ is are elements of a sample $(X_1, \ldots, X_n)$ from a rv X with df F. For a comprehensive study, including a list of areas where order statistics might have a significant role, we refer the reader to Arnold et al. (1992).

Definition 3.2.1 *(Order statistics)*
The order statistics of a random sample $(X_1, \ldots, X_n)$ are the sample values placed in ascending order. They are denoted by $X_{1,n}, \ldots, X_{n,n}$ and the rv $X_{n-k+1,n}$ is called the kth upper order statistic for $k = 1, 2, \ldots, n$. Order statistics satisfy $X_{1,n} \leq \cdots \leq X_{n,n}$. Then

$$X_{1,n} = \min(X_1, \ldots, X_n) \text{ and } X_{n,n} = \max(X_1, \ldots, X_n).$$

Definition 3.2.2 *(L-statistics)*
L-statistic is a statistic that is a linear combination of order statistics; the "L" is for "linear". For $(a_1, ..., a_n) \in \mathbb{R}^n$. The L-statistic is

$$T_n = \sum_{i=1}^{n} a_i X_{i,n}.$$

This statistic plays a major role in non-parametric statistics by providing robust estimators for location and scale parameters. For convenience in the study of the asymptotic behavior of T_n, the weights a_i are usually defined as $a_i = \left(\frac{1}{n}\right) J\left(\frac{i}{(n+1)}\right)$, where J is a real application on $(0, 1)$.

By using order statistics, the empirical df of the sample $(X_1, ..., X_n)$ is evaluated as follows

$$F_n(x) = \begin{cases} 0 & x < X_{1,n}, \\ i/n & X_{i,n} \leq x < X_{i+1,n}, \text{ for } i = 1, \ldots, n-1, \\ 1 & x \geq X_{n,n}. \end{cases}$$

Definition 3.2.3 *(Quantile and tail quantile functions)*
The quantile function of df F is the generalized inverse function of F defined by

$$Q(s) = F^{\leftarrow}(s) = \inf\{x \in \mathbb{R} : F(x) \geq s\}, \quad 0 < s < 1,$$

with the convention that the infimum of the empty set is ∞. In EVT, a function, denoted by $\mathbb{U}$ and (sometimes) called tail quantile function, is used quite often. It is defined by

$$\mathbb{U}(t) = Q(1 - 1/t) = \left(1/\overline{F}\right)^{\leftarrow}(t), \quad 1 < t < \infty.$$

Definition 3.2.4 *(Empirical quantile and tail quantile functions)*
The empirical quantile function of the sample $(X_1, \ldots, X_n)$ is defined by

$$Q_n(s) = F_n^{\leftarrow}(s) = \inf\{x \in \mathbb{R}, F_n(x) \geq s\}, \quad 0 < s < 1.$$

The corresponding empirical tail quantile function is

$$\mathbb{U}_n(t) = Q_n(1 - 1/t), \quad 1 < t < \infty.$$

The empirical quantile function Q_n may be expressed as a simple function of the order statistics pertaining to the sample $(X_1, \ldots, X_n)$. Namely, we have, for $0 < s < 1$

$$Q_n(s) = X_{i,n} \quad \text{if} \quad \frac{i-1}{n} < s \leq \frac{i}{n}, \quad i = 1, \ldots, n. \tag{3.2}$$

Then, for $0 < s < 1$,

$$Q_n(s) = X_{[ns]+1,n}.$$

Proposition 3.2.1 *(Quantile transformation)*
Let $(U_1, \ldots, U_n)$ be a sample from the standard uniform rv U and $(U_{1,n}, \ldots, U_{n,n})$ the corresponding ordered sample.

(*i*) *For any df* F*, we have*

$$X_{i,n} = F^{\leftarrow}(U_{i,n}), \quad i = 1, \ldots, n.$$

(*ii*) *When* F *is continuous, we have*

$$F(X_{i,n}) = U_{i,n}, \quad i = 1, \ldots, n.$$

In this case the rv's $F(X_1), \ldots, F(X_n)$ *are iid standard uniform.*

3.2.2 Limit law for maxima and GEVD

In this subsection, we will focus only on the results about the sample maximum, since the corresponding results for the sample minimum can be obtained from those of the sample maximum. Then, consider the sequence of maxima $M_n = X_{n,n} = \max(X_1, \ldots, X_n)$ obtained from the sequence of iid rv's $X_1, \ldots, X_n$. All the results obtained for maxima of course lead to analogous results for minima through the obvious relation

$$\min(X_1, \ldots, X_n) = -\max(-X_1, \ldots, -X_n).$$

The exact distribution of M_n can be obtained from the df F. Indeed, for all $x \in \mathbb{R}$,

$$\begin{aligned} F_{M_n}(x) &= \mathbb{P}(M_n \leq x) \\ &= \mathbb{P}(X_1 \leq x, \ldots, X_n \leq x) \\ &= \prod_{i=1}^{n} \mathbb{P}(X_i \leq x) \\ &= (F(x))^n. \end{aligned}$$

But the interest of this thesis relies on the behaviour of the sample maximum, when the sample size increases towards infinity.

Theorem 3.2.1 *Let* F *be the underlying df of a sequence of rv's and* x_F *its right endpoint, i.e.,* $x_F := \sup\{x : F(x) < 1\} \leq \infty$ *which may be infinite. Then*

$$F_{M_n}(x) \xrightarrow{\mathbb{P}} x_F \quad \text{as } n \to \infty.$$

Therefore, M_n has a degenerate asymptotic distribution. So, in order to do some kind of inference, we need to have a non-degenerate behaviour for M_n. Then, as with the Central Limit Theorem (CLT), a normalization is required. This theorem is concerned with the asymptotic behaviour of the sequence of sum $S_n = \sum_{i=1}^{n} X_i$.

Theorem 3.2.2 *(CLT)*
Consider a sequence of iid. rv's $X_1, \ldots, X_n$, with $\mu = E(X_1)$ and $\sigma^2 = Var(X_1) < \infty$. Therefore,

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad \text{as } n \to \infty.$$

In order to look for an appropriate non-degenerate limiting distribution for the sequence of sample maxima, we need a similar theorem, that is, we look for normalizing sequences $a_n > 0$ and b_n real such that

$$\frac{M_n - b_n}{a_n} \xrightarrow{\mathcal{D}} H \quad \text{as } n \to \infty. \tag{3.3}$$

with H non-degenerate, i.e.,

$$\lim_{n\to\infty} P\left(M_n \le a_n x + b_n\right) = \lim_{n\to\infty} F^n\left(b_n + a_n x\right) = H(x), \; x \in \mathbb{R}. \tag{3.4}$$

The first problem is to determine which df's H may appear on the limit in (3.3). These distributions are called extreme value distributions.

The problem of finding the extreme value distributions has been solved by Fisher and Tippett (1928), completed by Gnedenko (1943) and later revived and streamlined by de Haan (1970). They demonstrate that, if (3.1) holds, the limiting distribution H must be one of just three types. Formally,

Theorem 3.2.3 *(Asymptotic distribution of the sample maximum)*
If df F satisfies assumption (3.3), then df H is the same, up to location and scale, as one of the following distributions:

Type I $\quad \Lambda(x) = \exp\left(-e^{-x}\right), \quad x \in \mathbb{R}, \; \gamma = 0.$

Type II $\quad \Psi_\gamma(x) = \begin{cases} \exp\left(-x^{-\gamma}\right) & x > 0 \\ 0 & x \le 0 \end{cases} \quad \gamma < 0.$

Type III $\quad \Phi_\gamma(x) = \begin{cases} \exp\left(-(-x)^{\gamma}\right) & x \le 0 \\ 1 & x > 0 \end{cases} \quad \gamma > 0.$

Definition 3.2.5 *(Standard extreme value distributions)*
The three df 's of theorem 3.2.3 are called standard extreme value distributions. Λ is known as Gumbel or double exponential type, Ψ_γ as Fréechet or heavy-tailed type and Φ_γ as reverse Weibull type.

Proposition 3.2.2 *(Relations between Λ, Ψ_γ and Φ_γ)*
Let Z be a positive rv $(Z > 0)$, then the following assertions are equivalent:

(i) $Z \sim \Phi_\gamma$
(ii) $\ln Z^\gamma \sim \Lambda$
(iii) $-1/Z \sim \Psi_\gamma$.

Definition 3.2.6 *(GEVD)*
The GEVD is a df H defined, for all $x \in \mathbb{R}$ such that $1 + \gamma x > 0$, as follows

$$H_\gamma(x) = \begin{cases} \exp\left\{-(1+\gamma x)^{-1/\gamma}\right\} & if \quad \gamma \neq 0, \\ \exp\left(-e^{-x}\right) & if \quad \gamma = 0, \end{cases} \tag{3.5}$$

where the shape parameter is known as the extreme value index (EVI).

The parametrization in (3.5) is due to von Mises (1936) and Jenkinson (1955).

The corresponding location-scale family can be derived by replacing x in theorem 3.2.3 by $(x-\mu)/\sigma$, which are given below

$$H_{\gamma,\mu,\sigma}(x) = \begin{cases} \exp\left\{-(1+\gamma\frac{x-\mu}{\sigma})^{-1/\gamma}\right\} & if \quad \gamma \neq 0, \\ \exp\left(-e^{-\left(\frac{x-\mu}{\sigma}\right)}\right) & if \quad \gamma = 0. \end{cases}$$

The corresponding pdf $h_{\gamma,\mu,\sigma}$ and quantile $Q_{\gamma,\mu,\sigma}$ are defined by

$$h_{\gamma,\mu,\sigma}(x) = \begin{cases} H_\gamma\left(\frac{x-\mu}{\sigma}\right)\left(1+\frac{x-\mu}{\sigma}\gamma\right)^{-1/\gamma-1} & if \quad \gamma \neq 0,\ 1+\gamma x > 0, \\ \exp\left(-\left(\frac{x-\mu}{\sigma}\right) - \exp\left(\frac{x-\mu}{\sigma}\right)\right) & if \quad \gamma = 0,\ x \in \mathbb{R}. \end{cases}$$

and

$$Q_{\gamma,\mu,\sigma}(p) = \begin{cases} \mu - \sigma\gamma^{-1}\left(1-(\log p)^{-\gamma}\right) & if \quad \gamma \neq 0, \\ \mu - \sigma\log(-\log p) & if \quad \gamma = 0. \end{cases}$$

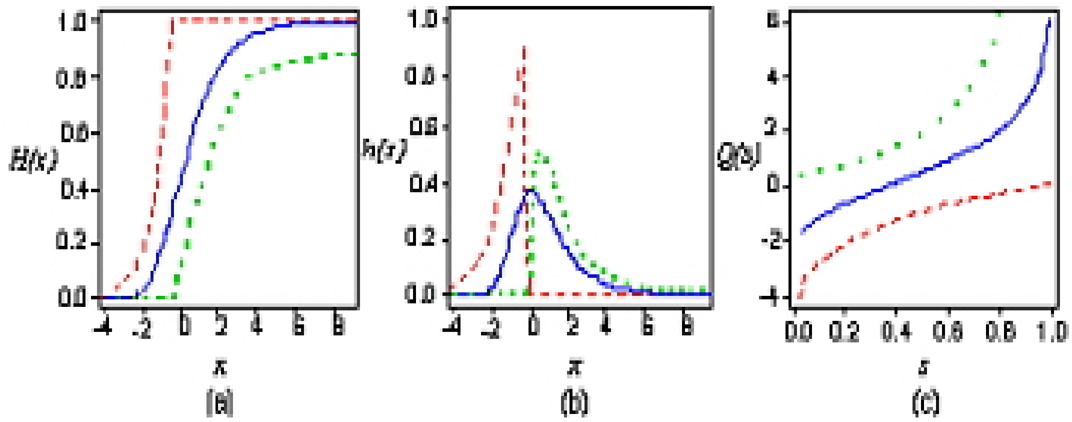

Figure 3.1: (a) df's of the Gumbel (solid line $\gamma = 0$), Fréchet (dotted line $\gamma = 1$) and Weibull (dashed line $\gamma = -1$) distributions, (b) corresponding pdf's and (c) corresponding quantiles.

3.2.3 Domains of attraction

The second problem: assuming H as a possible limiting df for the sequence $(M_n - b_n)/a_n$, what are the necessary and sufficient conditions that F must satisfy in order to belong to the domain of attraction of H? von Mises (1936) provided a set of conditions that ensures that F belongs to the domain of attraction of H. These conditions are known as von Mises' conditions.

Definition 3.2.7 *(Domain of attraction)*
We say that F belongs to the domain of attraction of of a non-degenerate df H, denoted by $F \in DA(H)$, if assumption (3.3) or (3.4) holds.

Proposition 3.2.3 *(Characterization of $DA(H)$)*
The df $F \in DA(H)$ with norming constants $a_n \in \mathbb{R}$ and $b_n > 0$ if and only if

$$\lim_{n\to\infty} n\overline{F}(b_n + a_n x) = -\log H(x), \ x \in \mathbb{R}. \tag{3.6}$$

When $H(x) = 0$, the right hand side is interpreted as ∞.

In addition to formulations (3.3), (3.4) and (3.6) for the domain of attraction assumption, there exist other alternative assertions stated in the following proposition. The first one illustrates the restriction on the upper distribution tail, the second form is in terms of function Q and the third assertion is in terms of function $\mathbb{U}$.

Proposition 3.2.4 *(Characterizations of $DA\,(H_\gamma)$)*
For $\gamma \in \mathbb{R}$ and for all $x > 0$, the following assertions are equivalent.

(a) $F \in DA\,(H_\gamma)$.

(b) *For some positive function g and $(1 + \gamma x) > 0$*

$$\lim_{t \to x_F} \frac{\overline{F}\,(t + xg\,(t))}{\overline{F}\,(t)} = \begin{cases} (1+\gamma x)^{-1/\gamma} & if \;\; \gamma \neq 0, \\ e^{-x} & if \;\; \gamma = 0. \end{cases}$$

(c) *For some positive function $\widetilde{a}$*

$$\lim_{s \to 0} \frac{Q\,(1 - sx) - Q\,(1 - s)}{\widetilde{a}\,(s)} = \begin{cases} \dfrac{x^{-\gamma} - 1}{\gamma} & if \;\; \gamma \neq 0, \\ \log x & if \;\; \gamma = 0. \end{cases}$$

(d) *For some positive function $a\,(t) = \widetilde{a}\,(1/t)$*

$$\lim_{t \to \infty} \frac{\mathbb{U}\,(tx) - \mathbb{U}\,(t)}{a\,(t)} = \begin{cases} \dfrac{x^{\gamma} - 1}{\gamma} & if \;\; \gamma \neq 0, \\ \log x & if \;\; \gamma = 0. \end{cases}$$

(e) *For $y > 0, y \neq 1$*

$$\lim_{s \to 0} \frac{Q\,(1 - sx) - Q\,(1 - s)}{Q\,(1 - sy) - Q\,(1 - s)} = \begin{cases} \dfrac{x^{-\gamma} - 1}{y^{-\gamma} - 1} & if \;\; \gamma \neq 0, \\ \dfrac{\log x}{\log y} & if \;\; \gamma = 0. \end{cases}$$

(f) *For $y > 0, y \neq 1$*

$$\lim_{t \to \infty} \frac{\mathbb{U}\,(tx) - \mathbb{U}\,(t)}{\mathbb{U}\,(ty) - \mathbb{U}\,(t)} = \begin{cases} \dfrac{x^{\gamma} - 1}{y^{\gamma} - 1} & if \;\; \gamma \neq 0, \\ \dfrac{\log x}{\log y} & if \;\; \gamma = 0. \end{cases} \tag{3.7}$$

von Mises (1936) provided a set of conditions that ensures that F belongs to the domain of attraction of H. These conditions are known as von Mises' conditions.

Theorem 3.2.4 *(von Mises' sufficient conditions for $F \in DA\,(H_\gamma)$)*
Let F be an absolutely continuous df. Existing the pdf, $f(x) = F'(x)$, and the second derivative $F''(x)$, let $h(x) = \frac{f(x)}{\overline{F}(x)}$ represent the hazard function or hazard rate from Reliability Theory.

1. *Suppose* $h(x) \neq 0$ *and differentiable for* x *next to* x_F *(or for large* x*, if* $x_F = \infty$*). If*
$$\lim_{x \to x_F} \frac{d}{dx} \left(\frac{1}{h(x)} \right) = 0,$$
then $F \in DA(\Lambda)$.

2. *Suppose* $x_F = \infty$ *and* F' *exists. If, for some* $\gamma > 0$,
$$\lim_{x \to \infty} x h(x) = \frac{1}{\gamma},$$
then $F \in DA(\Phi_\gamma)$.

3. *Suppose* $x_F < \infty$ *and* F' *exists for* $x < x_F$*. If, for some* $\gamma < 0$,
$$\lim_{x \to \infty} (x_F - x) h(x) = -\frac{1}{\gamma},$$
then $F \in DA(\Psi_\gamma)$.

These three conditions may be unified in a unique sufficient condition for F to belong to any of the only three domains of attraction, also derived in von Mises (1936).

Theorem 3.2.5 *(von Mises' sufficient conditions for* $F \in DA(H_\gamma)$*)*
Under the conditions of theorem 3.2.4, if
$$\lim_{x \to x_F} \left(\frac{1}{h(x)} \right)' = \gamma,$$
then $F \in DA(H_\gamma)$.

von Mises' conditions are very easy to check, requiring only the existence of the first or second derivative of F, but are only applicable to absolutely continuous df's F. These conditions are only sufficient conditions, and not necessary. Gnedenko (1943) given a set of necessary and sufficient conditions for maximal attraction to the three types of limit laws:

Theorem 3.2.6 *(Gnedenko's necessary and sufficient conditions for $F \in DA(H_\gamma)$)*

1. *$F \in DA(\Lambda)$ if and only if*

$$x_F \leq \infty \quad \textit{and} \quad \lim_{x \to x_F} \frac{\overline{F}(x + tg(x))}{\overline{F}(x)} = e^{-t}, \ x \in \mathbb{R},$$

   where $g(t)$is a continuous and monotone positive function.

2. *$F \in DA(\Phi_\gamma)$ if and only if, for $\gamma > 0$,*

$$x_F = \infty \quad \textit{and} \quad \lim_{x \to \infty} \frac{\overline{F}(tx)}{\overline{F}(x)} = t^{1/\gamma}, \ t > 0.$$

3. *$F \in DA(\Psi_\gamma)$ if and only if, for $\gamma < 0$,*

$$x_F < \infty \quad \textit{and} \quad \lim_{x \to 0^+} \frac{\overline{F}(x_F - tx)}{\overline{F}(x_F - x)} = t^{-1/\gamma}, \ t > 0.$$

According to theorem 3.2.3, the Fréchet-type distribution Φ_γ only attracts df's where $F(x) < 1$, $\forall x$, i.e., where $x_F = \infty$, and the Weibull-type distribution Ψ_γ only attracts df's where $F(x_F) = 1$, for $x_F < \infty$, and $F(x) < 1$, $\forall x < x_F$. However, Gnedenko refers that the conditions for the Gumbel domain are neither definitive nor convenient for practical use. For this case, the von Mises' condition is better, but not necessary.

3.2.4 Choice of the normalizing sequences

The third and last problem is the choice of suitable normalizing sequences a_n and b_n for the basic limit relation (3.3). This choice is not unique. This is a consequence of Khintchine's theorem whose proof is to be found in, e.g., Resnick (1987). The choice of such sequences depends on the df H that appears on the limit. The most common choices are indicated in the following theorem.

Theorem 3.2.7 *(Convergence to Types, Khintchine's theorem)*

1. *Let W and $\widetilde{W}$ be two rv's with non-degenerate df's G and $\widetilde{G}$ respectively. Suppose that $\{X_n\}_{n \in \mathbb{N}}$ is a sequence of iid rv's with df F and that we have real sequences $a_n, \widetilde{a}_n > 0$ and $b_n, \widetilde{b}_n \in \mathbb{R}$, such that*

$$\frac{X_n - b_n}{a_n} \xrightarrow{\mathcal{D}} W \quad \text{and} \quad \frac{X_n - \widetilde{b}_n}{\widetilde{a}_n} \xrightarrow{\mathcal{D}} \widetilde{W}, \quad \text{as } n \to \infty.$$

Then, there exist constants $A > 0$ *and* $B \in \mathbb{R}$ *such that*

$$\frac{\widetilde{a}_n}{a_n} \to A, \quad \frac{\widetilde{b}_n - b_n}{a_n} \to B \quad \text{as } n \to \infty,$$

and

$$\widetilde{G}(x) = G(Ax + B),$$

for every continuity point x *of* G *and* $\widetilde{G}$.

2. *Conversely,*

$$\text{if } \frac{\widetilde{a}_n}{a_n} \to A, \quad \frac{\widetilde{b}_n - b_n}{a_n} \to B \text{ and } \frac{X_n - b_n}{a_n} \xrightarrow{\mathcal{D}} W, \quad \text{as } n \to \infty,$$

then

$$\frac{X_n - \widetilde{b}_n}{\widetilde{a}_n} \xrightarrow{\mathcal{D}} \frac{W - B}{A}, \quad \text{as } n \to \infty,$$

with $\widetilde{G}(x) = G(Ax + B)$ *and for every continuity point* x *of* G *and* $\widetilde{G}$.

Theorem 3.2.8 *(Normalizing constants)*
If $F \in DA(H_\gamma)$, *then*

1. *For* $\gamma = 0$

$$\lim_{n\to\infty} F^n(b_n + a_n x) = \exp(-\exp(-x)) = \Lambda,$$

holds for all $x \in \mathbb{R}$ *with*

$$a_n = \mathbb{U}(ne) - \mathbb{U}(n) \text{ and } b_n = \mathbb{U}(n).$$

2. *For* $\gamma > 0$

$$\lim_{n\to\infty} F^n(b_n + a_n x) = \exp\left(-x^{-\frac{1}{\gamma}}\right) = \Phi_{\frac{1}{\gamma}},$$

holds for all $x > 0$ *with*

$$a_n = \mathbb{U}(n) \text{ and } b_n = 0.$$

3. *For* $\gamma < 0$

$$\lim_{n\to\infty} F^n(b_n + a_n x) = \exp\left(-(-x)^{-\frac{1}{\gamma}}\right) = \Psi_{-\frac{1}{\gamma}},$$

holds for all $x < 0$ *with*

$$a_n = x_F - \mathbb{U}(n) \text{ and } b_n = x_F.$$

von Mises's sufficient conditions also include normalizing sequences a_n and b_n for the general case $F \in DA(H_\gamma)$, rewritten by de Haan (1970), with the tail quantile function $\mathbb{U}$.

Theorem 3.2.9 *(Normalizing constants for $F \in DA(H_\gamma)$)*

Under the conditions of theorem 3.2.7, we have

$$b_n = Q\left(1 - \frac{1}{n}\right) = \mathbb{U}(n),$$

and

$$a_n = \frac{1}{h(b_n)} = \frac{1}{nf(b_n)} = n\mathbb{U}'(n).$$

It must be emphasized that these latter sequences are distinct from the normalizing sequences presented in theorem 3.2.8. They are used to normalize the sample maximum M_n, when F belongs to the domain of attraction of the GEVD. However, the sequences presented in theorem 3.2.8 are used to normalize the sample maximum when F belongs to the domain of attraction of one of the three standard types of theorem 3.2.3.

Chapter 4

Tail index and high quantile estimation

Tail index or EVI measures the degree of heaviness of the distribution tail: the heavier the tail, the larger γ. Various EVI estimators proposed after the work of Pickands (1975) and Hill (1975). A summary of many existing tail index estimators and their properties can be found in Brazauskas and Serfling (2000), Beirlant et al (2004) and Markovich (2007). In this chapter, we present some of the most well-known estimators with their asymptotic properties. Also, we review some of the classical methods to construct estimators for high quantiles and distribution tails. These estimators are based on the number k of upper statistics. We present some algorithms on how to determine this number. The asymptotic normality of these estimators, require that the underlying df F satisfies another condition, apart from those mentioned. This additional condition is known as the second order regular variation condition.

4.1 Second order regular variation

In a semi-parametric approach, apart from the first order condition, we often need a second order regular variation condition, to guarantee desirable properties for the estimators of the EVI. The only assumption in this approach is that $F \in DA\left(H_{\gamma}\right)$, which is equivalent to assume that the first order extended regular variation property is satisfied, i.e., from proposition 3.2.4, we have

$$F \in DA(H_\gamma) \Longleftrightarrow \lim_{t\to\infty} \frac{\mathbb{U}(tx) - \mathbb{U}(x)}{a(t)} = D_\gamma(x) = \begin{cases} \frac{x^\gamma - 1}{\gamma}, & \gamma \neq 0, \\ \log x, & \gamma = 0, \end{cases}$$

for every $x > 0$ and some positive measurable auxiliary function a, where necessarily we have $a \in \mathcal{RV}_\gamma$, according to definition of extended regular variation.

As mentioned before, an increase of the sample size implies an increase of k, the number of intermediate order statistics used to estimate . However, this rise of k may introduce some bias in the EVI-estimators, which can be controlled if we have additional information about the tail of F, in order to control the speed of convergence in the first order condition, i.e., the speed of convergence of the df of the sample maximum, linearly normalized, towards the limit law H. Therefore, the choice of k will also be decided by the second order condition. We must then quantify the speed of convergence, imposing a precise rate. For that, we need to assume that there exists a function A, not changing sign eventually, such that $\lim_{t\to\infty} A(t) = 0$, which measures not only the speed of convergence of the sequence of maximum values to a non-degenerate limit law but also the bias of the estimators. As A measures the speed of convergence of $\frac{\mathbb{U}(tx)-\mathbb{U}(x)}{a(t)}$ towards $D_\gamma(x)$

$$\lim_{t\to\infty} \frac{\frac{\mathbb{U}(tx)-\mathbb{U}(x)}{a(t)} - D_\gamma(x)}{A(t)}, \tag{4.1}$$

must exist, for all $x > 0$. Let be $H(x)$ the limit function of (4.1). We can now define the second order condition as follows:

Definition 4.1.1 *(Second order condition)*
The function $\mathbb{U}$ or the associated df F is said to satisfy the second order condition if, for some positive function a and for some positive or negative function A, with $\lim_{t\to\infty} A(t) = 0$,

$$\lim_{t\to\infty} \frac{\frac{\mathbb{U}(tx)-\mathbb{U}(x)}{a(t)} - D_\gamma(x)}{A(t)} = H(x). \tag{4.2}$$

As for the function a of the first-order condition, we call A the second order auxiliary function.

Definition 4.1.2 *(Second order regular variation assumption)*
We say that, the tail of $F \in DA\left(\Phi_{1/\gamma}\right)$, $\gamma > 0$, is second order regularly varying at infinity if it satisfies one of the following (equivalent) conditions:

(a) *There exist some parameter $\rho \leq 0$ and a function A^* tending to 0 and not changing sign near infinity, such that*

$$\lim_{t\to\infty} \frac{1}{A^*(t)} \left(\frac{1-F(tx)}{1-F(t)} - x^{-1/\gamma} \right) = x^{-1/\gamma} \frac{x^\rho - 1}{\rho}, \text{ for any } x > 0. \quad (4.3)$$

(b) *There exist some parameter $\rho \leq 0$ and a function A^{**} tending to 0 and not changing sign near zero, such that*

$$\lim_{s\to 0} \frac{1}{A^{**}(s)} \left(\frac{Q(1-sx)}{Q(1-s)} - x^{-\gamma} \right) = x^{-\gamma} \frac{x^\rho - 1}{\rho}, \text{ for any } x > 0.$$

(c) *There exist some parameter $\rho \leq 0$ and a function A tending to 0 and not changing sign near infinity, such that*

$$\lim_{t\to\infty} \frac{1}{A(t)} \left(\frac{\mathbb{U}(tx)}{\mathbb{U}(t)} - x^{\gamma} \right) = x^{\gamma} \frac{x^\rho - 1}{\rho}, \text{ for any } x > 0. \quad (4.4)$$

If $\rho = 0$ interpret $\dfrac{x^\rho - 1}{\rho}$ as $\log x$.

We need now to determine which functions $H(x)$ are eligible for the limit relation in (4.2). Following de Haan and Ferreira (2006), we can then state the following result for the function $H(x)$:

Theorem 4.1.1 *(de Haan and Ferreira, 2006)*
Suppose relation (4.2) holds and the function H is not a multiple of $D_\gamma(x)$ and is not identically zero. Then, there exist functions a, positive, and A, positive or negative and a parameter $\rho \leq 0$ such that

$$\begin{aligned} \lim_{t\to\infty} \frac{\frac{\mathbb{U}(tx)-\mathbb{U}(x)}{a(t)} - D_\gamma(x)}{A(t)} &= H_{\rho,\gamma}(x) \\ &= \frac{1}{\rho}\left(\frac{x^{\gamma+\rho}-1}{\gamma+\rho} - \frac{x^\gamma - 1}{\gamma} \right), \text{ for } x > 0. \quad (4.5) \end{aligned}$$

ρ is a second order parameter controlling the speed of convergence of the first order condition. For the cases $\gamma = 0$ and/or $\rho = 0$, $H_{\rho,\gamma}(x)$ is understood to be equal to the respective limit in (4.5), by continuity arguments. Moreover, $A(t)$ is such that

$$\lim_{t\to\infty} \frac{A(tx)}{A(t)} = x^\rho,$$

for $x > 0$, that is, $|A| \in \mathcal{RV}_\rho$.

Note that the second order condition implies the domain of attraction condition. Dekkers and de Haan (1989) showed that the second order condition holds for most of well-known df's: Normal, Gamma, Exponential, Uniform and Cauchy.

Hall's class of df's.
Hall's class of models (Hall and Welsh, 1985) is an example of heavy tailed distributions satisfying the second order condition, where

$$\overline{F}(x) = cx^{-1/\gamma}\left(1 + dx^{\rho/\gamma} + o\left(x^{\rho/\gamma}\right)\right) \quad \text{as } x \to \infty,$$

or equivalently, in terms of functions Q and $\mathbb{U}$,

$$Q(1-s) = c^{\gamma}x^{-\gamma}\left(1 + \gamma dc^{\rho}s^{-\rho} + o\left(x^{-\rho}\right)\right) \quad \text{as } s \to 0,$$

and

$$\mathbb{U}(t) = c^{\gamma}t^{\gamma}\left(1 + \gamma dc^{\rho}t^{\rho} + o(t^{\rho})\right) \quad \text{as } t \to \infty.$$

This class contains most of the heavy-tailed models important in insurance mathematics, like Fréchet, Pareto, Burr and t-Student df's. Straightforward computations show that, in the case of Hall model, functions $A(t)$ and $A^{*}(t)$ are respectively equivalent to $\gamma dc^{\rho}t^{\rho}$ and $\gamma dc^{\rho}t^{\rho/\gamma}$ as $t \to \infty$, whereas function $A^{**}(s)$ is equivalent to $\gamma dc^{\rho}s^{-\rho}$ as $s \to 0$.

4.2 Tail index estimation

The EVI or tail index is an important measure to gauge the heavy-tailed behavior of a distribution. Estimating this index in heavy-tailed distribution plays a central role in EVT.

4.2.1 Pickands' Estimator

Pickands (1975) was the first to present a semi-parametric estimator for a real EVI, $\gamma \in \mathbb{R}$, and thus it can be used to estimate the shape parameter of any one of the three types of extreme value distributions. But, as it is rather unworkable in practise for small or moderate samples, several refinements were introduced mainly by Drees (1996) and Segers (2005). The derivation of this estimator is based on an equivalent condition to $F \in DA\left(H_\gamma\right)$, namely assertion (3.2), which for $x = 2$ and $y = 1/2$ yields

$$\lim_{t\to\infty} \frac{\mathbb{U}(2t) - \mathbb{U}(t)}{\mathbb{U}(t) - \mathbb{U}(t/2)} = 2^\gamma.$$

Furthermore, for any positive function c such that $\lim_{t\to\infty} c(t) = 2$, we have

$$\lim_{t\to\infty} \frac{\mathbb{U}(tc(t)) - \mathbb{U}(t)}{\mathbb{U}(t) - \mathbb{U}(t/c(t))} = 2^\gamma.$$

Combining this with some basic results related to the ordered $(Y_{1,n}, \ldots, Y_{n,n})$ from a standard Pareto rv Y with df $F_Y(y) = 1 - y$, $y \geq 1$, namely $(k/n)/Y_{n-k+1,n} \xrightarrow{\mathbb{P}} 1$ and $Y_{n-k+1,n}/Y_{n-2k+1,n} \xrightarrow{\mathbb{P}} 2$ as $n \to \infty$ see e.g., Fraga Alves (1995), yields

$$\lim_{t\to\infty} \frac{\mathbb{U}(Y_{n-k+1,n}) - \mathbb{U}(Y_{n-2k+1,n})}{\mathbb{U}(Y_{n-2k+1,n}) - \mathbb{U}(Y_{n-4k+1,n})} = 2^\gamma.$$

Finally, we use the distributional identity

$$X_{n-i+1,n} \stackrel{\mathcal{D}}{=} \mathbb{U}(Y_{n-i+1,n}), \quad i = 1, \ldots, n,$$

to get the following definition.

Definition 4.2.1 *(Pickands' estimator)*

$$\widehat{\gamma}^P_{n,k} = \frac{1}{\log 2} \log \frac{X_{n-k+1,n} - X_{n-2k+1,n}}{X_{n-2k+1,n} - X_{n-4k+1,n}}.$$

The properties of $\widehat{\gamma}^P_{n,k}$ were mainly explored by Dekkers and de Haan (1989). They showed, under certain conditions, weak and strong consistency, as well as asymptotic normality.

Theorem 4.2.1 *(Properties of $\widehat{\gamma}^P_{n,k}$)*

Suppose that $F \in DA(H\gamma)$, $\gamma \in \mathbb{R}$, $k \to \infty$ and $k/n \to 0$ as $n \to \infty$.

(a) *Weak Consistency:*

$$\widehat{\gamma}^P_{n,k} \xrightarrow{\mathbb{P}} \gamma \quad as \quad n \to \infty.$$

(b) *Strong consistency: If* $k/\log\log n \to \infty$ *as* $n \to \infty$

$$\widehat{\gamma}^P_{n,k} \xrightarrow{a.s.} \gamma \quad as \quad n \to \infty.$$

(c) *Asymptotic normality: Assume that* $\mathbb{U}$ *has a positive derivative* $\mathbb{U}'$ *and that* $\pm t^{1-\gamma}\mathbb{U}'(t)$ *is* Π*-varying at infinity with auxiliary function* $a > 0$. *If* $k = o(n/g^{\leftarrow}(n))$, *where* $g(t) = t^{3-2\gamma}(\mathbb{U}'(t)/a(t))^2$, *then*

$$\sqrt{k}\left(\widehat{\gamma}_{n,k}^{P} - \gamma\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2\right) \quad \text{as} \quad n \to \infty,$$

where

$$\sigma^2 = \begin{cases} \dfrac{3}{4(\log 2)^4} & if \quad \gamma = 0, \\ \dfrac{\left(1 + 2^{2\gamma+1}\right)\gamma^2}{\left(2\left(2^{\gamma} - 1\right)\log 2\right)^2} & if \quad \gamma \neq 0. \end{cases}$$

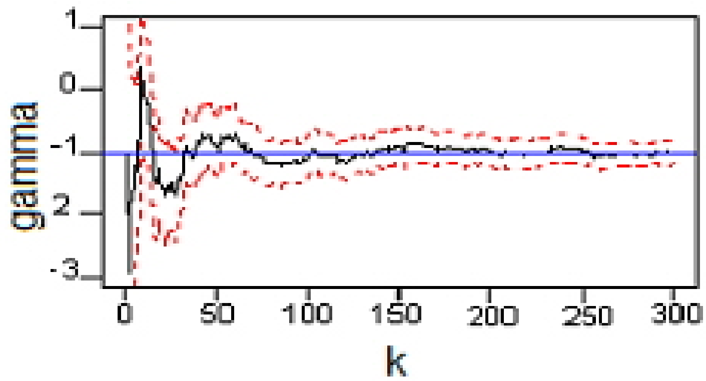

Figure 4.1: Plot of Pickands' estimator, as a function of the number of extremes (solid line) with 95%-confidence bounds (dashed lines), for the EVI of the standard uniform distribution ($\gamma = -1$) based on 100 samples of 3000 observations.

The Pickands estimator is quite volatile as a function of k, and its asymptotic variance is large. Ways of improvement were therefore sought after and discovered by many authors, all of whom realized somehow that this estimator is a linear combination of log-spacings of order statistics. Among improvements, we refer Fraga Alves (1995), Drees (1996) and Segers (2005).

Fraga Alves (1995) proposed a generalization of the Pickands estimator, with the introduction of a tuning or control parameter M, defined as follows

$$\widehat{\gamma}_{n,k,\,M}^{P} = \frac{1}{\log M}\log\frac{X_{l,n} - X_{Ml,n}}{X_{Ml,n} - X_{M^2l,n}}, \quad l = \left[\frac{k+n}{M^2}\right], \quad k = 1, \ldots, n \text{ and } M \in \mathbb{N}\backslash\{0, 1\},$$

which involves $k + 1$ of the top observations, for $k \geq M^2 - 1$. The traditional Pickands estimator corresponds to $\widehat{\gamma}_{n,M}^{P}$.

4.2.2 Hill's estimator

A few months after the publication of Pickands estimator, Hill (Hill, 1975) proposed another estimator restricted to the Fréchet case $\gamma > 0$, i.e., this estimator is applicable only to regular varying df's. In other words, Hill's estimator is only suitable for data exhibiting heavy tails such as those encountered in (re)insurance and finance. It is the most popular tail index estimator.

Definition 4.2.2 *(Hill's estimator: $\gamma > 0$)*

$$\widehat{\gamma}^{H}_{n,k} = \frac{1}{k}\sum_{i=1}^{k} \log X_{n-i+1,n} - \log X_{n-k,n},$$

or, equivalently

$$\widehat{\gamma}^{H}_{n,k} = \frac{1}{k}\sum_{i=1}^{k} i\left(\log X_{n-i+1,n} - \log X_{n-i,n}\right).$$

The original derivation of Hill's estimator relied on the notion of conditional maximum likelihood estimation method. Let $F \in DA\left(\Phi_{1/\gamma}\right)$ and let $Y_u = \left(\frac{X}{u} \mid X > u\right)$ denote the relative excesses over a threshold u. The conditional distribution of Y_u satisfies

$$\overline{F}_{Y_u}(y) = y^{-1/\gamma}\frac{L(u\,\mathrm{y})}{L(u)},\ \ \mathrm{y} \geq 1,$$

where $L \in \mathcal{RV}_0$. Then it immediately follows that

$$F_{Y_u}(y) \to 1 - y^{-1/\gamma} \text{ as } u \to \infty,$$

i.e., Y_u is asymptotically Pareto$(1/\gamma)$ distributed. Assuming that this approximation holds for the k relative excesses $X_{n,n}/u, \ldots, X_{n-k+1,n}/u$ above a high threshold $u = X_{n-k,n}$ leads to the conditional maximum likelihood estimator

$$\frac{1}{k}\sum_{i=1}^{k} \log\left(\frac{X_{n-i+1,n}}{X_{n-k,n}}\right),$$

which is exactly $\widehat{\gamma}^{H}_{n,\,k}$.

One of the appealing features of Hill's estimator is that it can be derived even if we start from very different motivation points. Apart from the likelihood approach described above, the Hill estimator can be derived via a regular variation approach (see Embrechts et al., 1997) as well as via a graphical approach using mean excess function or QQ-plots (see Beirlant et al., 1996).

The regular variation approach is in the same spirit as the construction of Pickands estimator, i.e. we base the inference of γ on a reformulation of $F \in DA(\Phi_\gamma)$ $\gamma > 0$. According to the characterization property of $F \in DA(\Phi_\gamma)$ $\gamma > 0$ if and only if the tail of F, $1 - F$, is regularly varying with index $-\alpha = 1/\gamma > 0$, that means that

$$\lim_{t\to\infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-\alpha}.$$

Partial integration of the above and use of Karamata's theorem (see chapter 1) leads to the relationship

$$\frac{1}{\overline{F}(t)} \int_t^\infty (\log x - \log t)\, dF(x) \to \frac{1}{\alpha}, \quad x \to \infty. \tag{4.6}$$

To turn this into an estimator we replace F by the empirical df $(X_1, \ldots, X_n)$

$$F_n(x) = \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{X_k \leq x\}},$$

and replace t by a high data dependent level $X_{k,n}$. Then

$$\frac{1}{\overline{F_n}(X_{k,n})} \int_{X_{k,n}}^\infty (\log x - \log X_{k,n})\, dF_n(x) = \frac{1}{k} \sum_{i=1}^{k} \log X_{n-i+1,n} - \log X_{n-k,n}.$$

If $k \to \infty$ and $k/n \to 0$ as $n \to \infty$., then $X_{k,n} \overset{a.s.}{\to} \infty$. as $n \to \infty$ and by (4.6)

$$\frac{1}{k} \sum_{i=1}^{k} \log X_{n-i+1,n} - \log X_{n-k,n} \overset{\mathbb{P}}{\to} \frac{1}{\alpha}, \quad n \to \infty.$$

This gives us the Hill's estimator.

Hill did not investigate the asymptotic behavior of $\widehat{\gamma}_{n,k}^H$ in his paper.The weak consistency was was proved by Mason (1982) while Deheuvels et al. (1988) proved the strong consistency. The asymptotic normality was established by several authors such as Csörgő and Mason (1985), Davis and Resnick (1984) and Häusler and Teugels (1985). Beirlant et al. (2006) derived a local asymptotic normality result showing that the asymptotic variance of Hill's estimator attains a lower bound.

Theorem 4.2.2 *(Properties of $\widehat{\gamma}^H_{n,k}$)*

Suppose that $F \in DA(\Phi_\gamma)$ $\gamma > 0$, $k \to \infty$ and $k/n \to 0$ as $n \to \infty$.

(a) *Weak Consistency:*

$$\widehat{\gamma}^H_{n,k} \xrightarrow{\mathbb{P}} \gamma \quad as \quad n \to \infty.$$

(b) *Strong consistency: If $k/\log\log n \to \infty$ as $n \to \infty$*

$$\widehat{\gamma}^H_{n,k} \xrightarrow{a.s.} \gamma \quad as \quad n \to \infty.$$

(c) *Asymptotic normality: Suppose that F satisfies (2.4). If $\lim_{n\to\infty} \sqrt{k}A(n/k) = \lambda$ as $n \to \infty$, then*

$$\sqrt{k}\left(\widehat{\gamma}^H_{n,k} - \gamma\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{\lambda}{1-\rho}, \gamma^2\right) \quad as \quad n \to \infty.$$

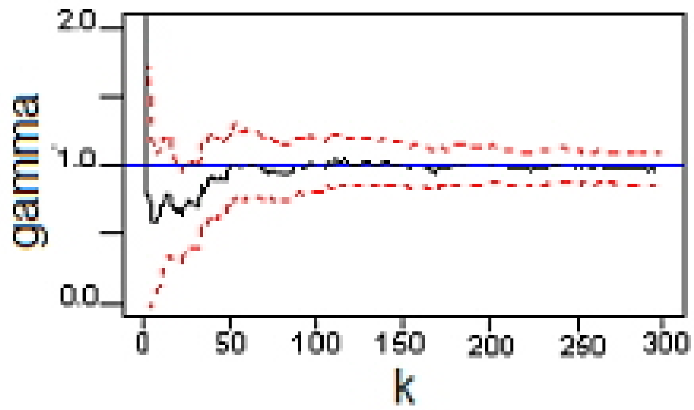

Figure 4.2: Plot of Hill's estimator, as a function of the number of extremes (solid line) with 95%-confidence bounds (dashed lines), for the EVI of the standard Pareto distribution ($\gamma = 1$) based on 100 samples of 3000 observations.

4.2.3 Moment estimator

The moment estimator, denoted by $\widehat{\gamma}^M_{n,k}$, was proposed by Dekkers et al. (1989) as an adaptation of the Hill estimator for $\gamma \in \mathbb{R}$ and not only for $\gamma > 0$.

Definition 4.2.3 *(Moment estimator: $\gamma \in \mathbb{R}$)*

$$\widehat{\gamma}_{n,k}^{M} = M_n^{(1)} + 1 - \frac{1}{2}\left(1 - \frac{\left(M_n^{(1)}\right)^2}{M_n^{(2)}}\right)^{-1},$$

where

$$M_n^{(r)} = M_n^{(r)}(k) := \frac{1}{k}\sum_{i=1}^{n-k}\left(\log X_{n-i+1,n} - \log X_{n-k,n}\right)^r, \; r = 1, \; 2. \tag{4.7}$$

Weak and strong consistency, as well as asymptotic normality of $\widehat{\gamma}_{n,k}^{M}$ have been proven by its creators Dekkers et al. (1989).

Theorem 4.2.3 *(Asymptotic properties of $\widehat{\gamma}_{n,k}^{M}$)*

Suppose that $F \in DA(H_\gamma)$ $\gamma \in \mathbb{R}$, $k \to \infty$ and $k/n \to 0$ as $n \to \infty$.

(a) *Weak Consistency:*

$$\widehat{\gamma}_{n,k}^{M} \xrightarrow{\mathbb{P}} \gamma \quad \text{as} \quad n \to \infty.$$

(b) *Strong consistency: If $k/(\log n)^\delta \to \infty$ as $n \to \infty$,for some $\delta > 0$, then*

$$\widehat{\gamma}_{n,k}^{M} \xrightarrow{a.s.} \gamma \quad \text{as} \quad n \to \infty.$$

(c) *Asymptotic normality: Suppose that F satisfies (4.3). If $\lim_{n\to\infty} \sqrt{k}Q(n/k) = 0$, then*

$$\sqrt{k}\left(\widehat{\gamma}_{n,k}^{M} - \gamma\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \eta^2\right) \quad \text{as} \quad n \to \infty,$$

where

$$\eta^2 = \begin{cases} 1 + \gamma^2 & if \;\; \gamma \geq 0, \\ (1-\gamma)^2(1-2\gamma)\left[4 - 8\dfrac{1-2\gamma}{1-3\gamma} + \dfrac{(5-11\gamma)(1-2\gamma)}{(1-3\gamma)(1-4\gamma)}\right] & if \;\; \gamma < 0. \end{cases}$$

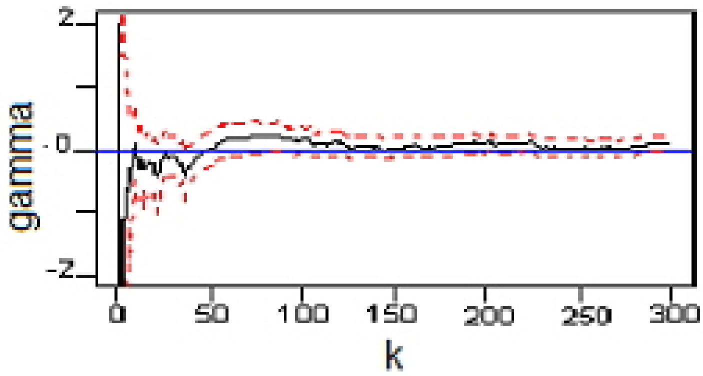

Figure 4.3: Plot of the moment estimator, as a function of the number of extremes (solid line) with 95%-confidence bounds (dashed lines), for the EVI of Gumbel distribution ($\gamma = 0$) based on 100 samples of 3000 observations.

4.2.4 Kernel-type estimators for $\gamma > 0$

All the estimators of EVI are based on the k largest observations. A major drawback of these estimators is the discrete character of the behavior of these estimators: adding a single large-order statistic in the calculation of the estimator, that is, increasing k by 1, can change the actual value of the estimate considerably. Plotting these estimators as a function of the order statistics used therefore often results in a zigzag figure. Using a kernel function K, Csörgő el al. (1985) proposed a smoother version of Hill's estimator, denoted by $\widehat{\gamma}_{n,h}^{K}$ where $h = h\,(n)$ is a bandwidth parameter. In the same paper, it was shown that it is possible to improve on the asymptotic variance of the estimator by choosing appropriate kernels. In this kernel-type estimator, the bandwidth h plays a similar role as the number of order statistics k in the aforementioned estimators: approximately nh order statistics will be used to calculate the estimate. Consequently, the estimator now depends in a continuous way on the fraction of order statistics used. Hence, plotting the estimator as a function of the bandwidth h then yields a smooth figure. Other attempts to construct smoothed versions of the Hill estimator can be found in Schultze and Steinebach (1996), Kratz and Resnick (1996) and Csörgő and Viharos (1998), which consider classical least-squares estimators for the slope $\gamma > 0$ in a Pareto quantile plot.

To define the kernel-type estimators for $\gamma > 0$, we need the following conditions about the kernel K.

Condition ($\mathbb{K}$). Let K be a function defined on $(0, 1]$.

KC1. $K(s) \geq 0$ for $0 < s < \infty$.

KC2. $K(\cdot)$ is non-increasing and right continuous on $(0, \infty)$.

KC3. $\int_0^1 K(u)du = 1$.

KC4. $\int_0^1 u^{-1/2} K(u)du < \infty$.

Then the estimator is defined as

$$\widehat{\gamma}_{n,h}^{K} = \frac{\left(\int_0^{1/h} \log_+ Q_n\left(1 - hu\right) d\left(uK\left(u\right)\right) \right)}{\left(\int_0^{1/h} K\left(u\right) du \right)},$$

where $h > 0$ is called bandwidth, Q_n is the empirical quantile function and $\log_+ x = \log\left(\max\left(x, 1\right)\right)$. Routine manipulation show that $\widehat{\gamma}_{n,h}^{K}$ can be written in the equivalent form

$$\widehat{\gamma}_{n,h}^{K} = \frac{\left(\sum_{i=1}^{n-1} \frac{i}{nh} K_h\left(\frac{i}{nh}\right) \left(\log X_{n-i+1,n} - \log X_{n-i,n}\right) \right)}{\left(\int_0^{1/h} K\left(u\right) du \right)}, \tag{4.8}$$

where $X_{0,1} = 1$. Notice that, using the uniform kernel $K = \mathbf{1}_{(0,1)}$ and $h = k/n$ in (4.8) yields Hill's estimator $\widehat{\gamma}_{n,k}^{H}$ as a special case.

To be able to state the consistency and asymptotic normality of $\widehat{\gamma}_{n,\, k}^{K}$, we will need some additional conditions on the kernel K:

KC5. There exists an $M_1 < \infty$ such that $K\left(u\right) = 0$ for $u > M_1$.

KC6. There exists an $M_2 < \infty$ such that the kernel K has a derivative $k\left(u\right)$ for $u > M_2$ and such that $\lim_{u \to \infty} u^{3/2} k\left(u\right) = 0$.

Also, we will need these conditions on the df F:

FC1. 1. The function $Q(1-\cdot)$ is regularly varying at 0 with index $-\gamma$, i.e., the quantile function satisfies the representation

$$Q(1-s) = s^{-\gamma} c(s) \exp\left(\int_s^1 \frac{b(u)}{u} du\right), \quad 0 < s < 1, \tag{4.9}$$

where c is function with $c(s) \to c \in (0, \infty)$ as $s \to \infty$ and b a function with $b(s) \to 0$ as $s \to \infty$.

2. With out loss of generality, $Q(0) = 1$.

FC2. 1. In the representation (4.9), one has that either KC5 is satisfied and $c(s) = c$ (constant) for $0 < s < \varepsilon$ for some $\varepsilon > 0$ or $c(s) = c$ (constant) for $0 < s \leq 1$.

2. One has either KC6 is satisfied or the function b in (4.9) may be chosen to be bounded on $(0, 1)$.

For a discussion of these conditions on F, we refer to Csörgő et al. (1985).

Theorem 4.2.4 *(Properties of $\widehat{\gamma}_{n,\, h}^{K}$)*
Let KC1–KC4 and FC1 be satisfied. Then, as $h = h(n) \to 0$ and $nh \to \infty$ as $n \to \infty$.

(a) *Weak consistency:*

$$\widehat{\gamma}_{n,h}^{K} \xrightarrow{\mathbb{P}} \gamma \quad as \quad n \to \infty.$$

(b) *Asymptotic normality: Moreover, if in condition FC2 is satisfied, then, for $h = h(n) \to 0$ and $hn \to \infty$ as $n \to \infty$.*

$$\sqrt{nh}\left(\widehat{\gamma}_{n,h}^{K} - \gamma - \beta_C(h)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \gamma^2 \sigma_K^2\right) \quad as \quad n \to \infty,$$

with $\beta_C(h)$ and σ_K^2 given by

$$\beta_C(h) = \left(\int_0^{1/h} b(hu) K(u) du\right) / \left(\int_0^{1/h} K(u) du\right),$$

$$\sigma_K^2 = \int_0^{\infty} K^2(u) du.$$

4.2.5 Kernel-type estimators for $\gamma \in \mathbb{R}$

Unfortunately, the kernel estimator $\widehat{\gamma}_{n,h}^{K}$ is only valid for $\gamma > 0$. Groeneboom et al. (2003) introduced a new kernel type estimator for $\gamma \in \mathbb{R}$, denoted by $\widehat{\gamma}_{n,h}^{W}$, that inherited the smooth behavior of $\widehat{\gamma}_{n,k}^{H}$ as well as the general applicability of $\widehat{\gamma}_{n,k}^{M}$. It should be emphasized that the estimator $\widehat{\gamma}_{n,h}^{W}$ is not a smoothed version of the moment estimator, but is based on the von Mises conditions

$$\lim_{x \to x_F} \frac{d}{dt} \left(\frac{\overline{F}(t)}{F'(t)} \right) = \gamma.$$

The estimator of Groeneboom et al. (2003) for $\gamma \in \mathbb{R}$ is defined as follows

$$\widehat{\gamma}_{n,h}^{W} = \widehat{\gamma}_{n}^{(pos)} - 1 + \frac{q_{n,h}^{(2)}}{q_{n,h}^{(1)}},$$

where

$$\begin{aligned} \widehat{\gamma}_{n}^{(pos)} &= \sum_{i=1}^{n-1} \frac{i}{n} K_h\left(\frac{i}{n}\right) \left(\log X_{n-i+1,n} - \log X_{n-i,n}\right), \\ q_{n,h}^{(1)} &= \sum_{i=1}^{n-1} \left(\frac{i}{n}\right)^{\alpha} K_h\left(\frac{i}{n}\right) \left(\log X_{n-i+1,n} - \log X_{n-i,n}\right), \end{aligned}$$

and

$$q_{n,h}^{(2)} = \sum_{i=1}^{n-1} \frac{d}{du} \left[u^{\alpha+1} K_h(u)\right]_{u=i/n} \left(\log X_{n-i+1,n} - \log X_{n-i,n}\right),$$

with $K_h(u) = K(u/h)/h$ and $\alpha > 0$. Here $K : [0,1] \to \mathbb{R}_+$ is a kernel function satisfying the following conditions:

(CK1) $K(u) = 0$, whenever $u \notin [0,1)$ and $K(u) \geq 0$, whenever $u \in [0,1)$.

(CK2) $K(1) = K'(1) = 0$.

(CK3) $\int_0^1 K(u)du = 1$ and .$\int_0^1 u^{\alpha-1} K(u)du \neq 0$.

(CK4) K, K' and K'' are bounded.

The full description of the way $\widehat{\gamma}_{n,h}^{W}$ is derived, is given by Groeneboom et al. (2003) where weak consistency and asymptotic normality are established. More recently, Necir (2006) proposed a functional law of the iterated logarithm for this estimator and proved its strong consistency.

Theorem 4.2.5 *(Properties of $\widehat{\gamma}_{n,h}^{W}$)*

(a) *Weak consistency: Assume that $F \in DA(H_\gamma)$ for some $\gamma \in \mathbb{R}$. Let K be a kernel satisfying conditions (CK1)–(CK4) and for arbitrary $\alpha > 0$. If $h = h(n)$ is such that $h \to 0$ and $nh \to \infty$ as $n \to \infty$, then*

$$\widehat{\gamma}_{n,h}^{W} \xrightarrow{\mathbb{P}} \gamma \quad as \quad n \to \infty.$$

(b) *Asymptotic normality: Let K be a kernel satisfying conditions (CK1)–(CK4). Then, for any $\alpha > 1/2$ and $h = h(n)$ with $h \to 0$ and $(nh)^{-\alpha} \log n = O\left((nh)^{-1/2}\right)$ as $n \to \infty$,*

$$\sqrt{nh}\left(\widehat{\gamma}_{n,h}^{W} - \gamma\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \gamma_W^2\right),$$

where

$$\gamma_W^2 = \int_0^1 \left(a_0 \tilde{K}(u) + a_1 \tilde{K}^{(2)}(u) - a_2 \tilde{K}^{(1)}(u)\right)^2 du,$$

with

$$\begin{aligned} \tilde{K}(u) &= \int_u^1 t^{-1} d(tK(t)), \quad u \in (0,1], \\ \tilde{K}^{(i)}(u) &= \int_u^1 t^{-1-(\gamma\wedge 0)} dK^{(i)}(t), \quad u \in (0,1], \end{aligned}$$

and

$$\begin{aligned} a_0 &= \gamma \vee 0, \\ a_1 &= 1 / \int_0^1 t^{-1-(\gamma\wedge 0)} K^{(1)}(t)\, dt, \\ a_2 &= (1 + (\gamma \wedge 0))\, a_1. \end{aligned}$$

4.2.6 Automatic bandwidth choice

In the case of kernel-type estimators, the problem becomes a matter of choosing an optimal bandwidth h_{opt} for which a consistent estimate $\widehat{h}_{opt}$ has to be adaptively computed. For instance, Groeneboom et al. (2003) follow a similar approach as in Draisma et al. (1999) and Cheng and Peng (2001). They base their estimators $\widehat{\gamma}_{n,h}^{W_1}$ and $\widehat{\gamma}_{n,h}^{W_2}$ on the following respective kernels:

$$K_1(u) = \frac{15}{8}\left(1 - u^2\right)^2 \mathbf{1}_{(0,1)}(u),$$

and

$$K_2(u) = \frac{35}{16}\left(1-u^2\right)^3 \mathbf{1}_{(0,1)}(u).$$

The five steps algorithm leading to $\widehat{h}_{opt}$ is as follows:

- *Step 1:* Select a bootstrap sample $X_1^*, \ldots, X_{n_1}^*$ of size $n_1 < n$ from the original sample $(X_1, \ldots, X_n)$ and compute $\widehat{\gamma}_{n,h}^{W_1^*}$and $\widehat{\gamma}_{n,h}^{W_2^*}$ in terms of the order statistics $X_{1,n_1,}^* < \cdots < X_{n_1,n_1}^*$pertaining to the bootstrap sample. Next compute

$$\delta_{n_1}^* = \widehat{\gamma}_{n,h}^{W_1^*} - \widehat{\gamma}_{n,h}^{W_2^*}.$$

- *Step 2:* Repeat step 1, r times independently. With the obtained sequence $\delta_{n_1,1}^*, \ldots, \delta_{n_1}^*$ compute

$$\widehat{MSE}^*\left(\delta_{n_1}^*\right) = \frac{1}{r}\sum_{i=1}^{r}\left(\delta_{n_1,i}^*\right)^2.$$

- *Step 3:* Compute

$$h^*(n_1) = \arg\min \widehat{MSE}^*\left(\delta_{n_1}^*\right).$$

- *Step 4:* Repeat steps 1-3 independently with $n_2 = [\frac{n_1^2}{n}]$ instead of n_1.

  Let

$$h^*(n_2) = \arg\min \widehat{MSE}^*(\delta_{n2}^*).$$

- *Step 5:* Estimate the optimal bandwidth h_{opt} by

$$\widehat{h}_{opt} = \psi\frac{\left(h^*(n_1)\right)^2}{h^*(n_2)},$$

  where ψ is a function of $h^*(n_1)$ and $h^*(n_2)$ depending on kernels K_1 and K_2 and sample sizes n_1 and n_2.

4.3 Optimal sample fraction selection

All the estimators of EVI are crucially depend on the choice of the sample fraction k that is used for estimation, i.e., the number of extreme order statistics on which the estimation is based. However, the choice of the optimal threshold is a difficult problem. If k is too large, then the estimators have a large bias, while on the other hand, their variance is large if k is small. In this section, where we mostly concentrate on Hill's estimator $\widehat{\gamma}_{n,k}^{H}$, we present some of the proposed methods to balance between these two vices in order to get an optimal number k which locates where the distribution tail really begins.

It seems reasonable that minimizing the MSE allows for a compromise between the bias and variance components yielding the most accurate estimate possible. That is, the optimal choice of k, denoted by k_{opt}, corresponds to the smallest MSE, i.e,

$$\begin{aligned} k_{opt} &= \arg\min_{k} MSE\left(\widehat{\gamma}_n\right) \\ &= \arg\min_{k} E_{\infty}\left(\widehat{\gamma}_n - \gamma\right)^2, \end{aligned}$$

where E_{∞} denotes the expectation with respect to the limit distribution.

Hall and Welsh (1985) proved that the asymptotic MSE of Hill's estimator is minimal for

$$k_{opt} \sim n^{2\rho/(2\rho+1)} \left(\frac{C^{2\rho}\left(\rho+1\right)^2}{2D^2\rho^3} \right)^{1/(2\rho+1)}.$$

Since parameters $\rho > 0, C > 0$ and $D \neq 0$ are unknown, this result cannot be applied directly to estimate k.

4.3.1 Bootstrap method

The number k_{opt} that are fitted to the tail corresponds to the minimum of MSE

$$\begin{aligned} MSE\left(\widehat{\gamma}_n\right) &= E_{\infty}\left(\widehat{\gamma}_n - \gamma\right)^2 \\ &= bias^2\left(\widehat{\gamma}_n\right) + Var\left(\widehat{\gamma}_n\right), \end{aligned}$$

where the bias is given by

$$bias\left(\widehat{\gamma}_n\right) = E_{\infty}\left(\widehat{\gamma}_n\right) - \gamma,$$

and the variance is determined by

$$Var\left(\widehat{\gamma}_n\right) = E_{\infty}\left(\left(\widehat{\gamma}_n - E_{\infty}\left(\gamma\right)\right)^2\right).$$

Since γ is unknown and MSE cannot be evaluated, the bootstrap approach proposes replacing in the MSE by an average calculated over some amount of resamples. These resamples are drawn from the initial sample $(X_1, \ldots, X_n)$. This implies that some observations from $(X_1, \ldots, X_n)$ will be represented in a resample with repetitions and others will not be represented at all. As a result, in order to estimate k one takes the value that minimizes a bootstrap empirical estimate of the MSE. More precisely, the bootstrap estimate of the bias is given by

$$b^* (n_1, k_1) = E_\infty \left(\widehat{\gamma}^* (n_1, k_1) \mid (X_1, \ldots, X_n) \right) - \gamma,$$

and the bootstrap estimate of the variance is determined by

$$Var \left(\widehat{\gamma}_n \right) = E_\infty \left(\left(\widehat{\gamma}^* (n_1, k_1) - E_\infty \left(\widehat{\gamma}^* (n_1, k_1) \mid (X_1, \ldots, X_n) \right) \right)^2 \mid X^n \right).$$

To construct these estimates, a smaller sample size $n_1 \leq n$ is used. It is determined by the resample $\left(X_1^*, \ldots, X_{n_1}^* \right)$ drawn randomly from $(X_1, \ldots, X_n)$ with replacement, where $X_{1,n_n}^* \leq \cdots \leq X_{n_1,n_1}^*$ are the order statistics of the sample $\left(X_1^*, \ldots, X_{n_1}^* \right)$. In the bootstrap estimates considered $(X_1, \ldots, X_n)$ is fixed and the expectation is calculated among all theoretically possible resamples $\left(X_1^*, \ldots, X_{n_1}^* \right)$ practice, the expectation is replaced by the average over the underlying resamples. The reason for using smaller resamples is that the classical bootstrap with resamples of the same size n as the initial sample leads to underestimates of the bias. Using a smaller sample size $n_1 \leq n$ and k_1 data may help to avoid the situation where the bootstrap estimate of the bias is equal to zero regardless of the true bias of the estimate.

4.3.2 Double bootstrap method

The double bootstrap, proposed in Danielsson et al. (1997), improves the bootstrap method (Hall, 1990). Instead of estimation of the MSE we use the auxiliary statistic

$$MSE \left(z_{n,k} \right) = E \left(\left(z_{n,k} - z_{n,k}^* \right)^2 \right),$$

where

$$z_{n,k} = M_{n,k} - 2 \left(\widehat{\gamma}_n \right)^2, \quad M_{n,k} = \frac{1}{k} \sum_{i=1}^{k} \left(\log X_{n-i+1,n} - \log X_{n-k,n} \right)^2,$$

and and $z_{n,k}^*$ is a bootstrap estimate of $z_{n,k}$. Since $M_{n,k}/2\widehat{\gamma}_n$ and $\widehat{\gamma}_n$ are consistent estimates of γ, then $z_{n,k} \to 0$ as $n \to \infty$. Therefore, the asymptotic

MSE of $z_{n,k}$ is defined by $AMSE\left(z_{n,k}\right) = E\left(z_{n,k}^2\right)$. Hence, the value k_{opt} of k, which minimizes $AMSE\left(z_{n,k}\right)$, has the same order in n as k_{opt} , the value that minimizes $MSE\left(\widehat{\gamma}_n\right)$.

The double bootstrap algorithm is as follows:

- *Step 1:* Draw B bootstrap subsamples of size $n_1 \in (n, \sqrt{n})$ (e.g., $n_1 \sim n^{3/4}$) from the original sample and determine the value $\widehat{k}_{n_1}^*$ that minimizes the MSE of $z_{n_1,k}$.
- *Step 2:* Repeat this for B subsamples of size $n_2 = \lfloor n_1^2/n \rfloor$ and determine the value $\widehat{k}_{n_2}^*$ that minimizes the MSE of $z_{n_2,k}$.
- *Step 3:* Calculate $\widehat{k}_{opt}$ by the formula

$$\widehat{k}_{opt} = \left[\frac{\left(\widehat{k}_{n_1}^*\right)^2}{\widehat{k}_{n_2}^*} \left(1 - \frac{1}{\widehat{\rho}_1}\right)^{\frac{2}{2\widehat{\rho}_1 - 1}} \right],$$

where

$$\widehat{\rho}_1 = \frac{\log \widehat{k}_{n_1}^*}{2 \log \left(\widehat{k}_{n_1}^*/n_1\right)}.$$

4.3.3 Sequential procedure

Drees and Kaufmann (1998) proposed the following algorithm:

- *Step 1:* Obtain an initial estimate $\widehat{\gamma}_0 = \widehat{\gamma}_n^{(H)}\left(2\sqrt{n}\right)$ for the parameter γ by Hill's estimator.
- *Step 2:* For $r_n = 2.5\widehat{\gamma}_0 n^{0.25}$, compute

$$\widehat{k}\left(r_n\right) = \min\left\{k \in 2, \ldots, n-1 : \max_{2 \le i \le k} \sqrt{i} \left| \widehat{\gamma}_n^{(H)}(i) - \widehat{\gamma}_n^{(H)}(k) \right| > r_n \right\}.$$

  If r_n is too large and $\max_{2 \le i \le k} \sqrt{i} \left| \widehat{\gamma}_n^{(H)}(i) - \widehat{\gamma}_n^{(H)}(k) \right| > r_n$ is not satisfied, it is recommended to repeatedly replace r_n by $09 r_n$ until $\widehat{k}\left(r_n\right)$ is well defined.

- *Step 3:* Compute, for $\varepsilon = 0.7$,

$$\widehat{k}_{opt} = \frac{1}{3}\left(2\widehat{\gamma}_0\right)^{1/3}\left(\frac{\widehat{k}\left(r_n^{\varepsilon}\right)}{\left(\widehat{k}\left(r_n\right)\right)^{\varepsilon}}\right)^{1/(1-\xi)}.$$

The method is sensitive to the choice of r_n. For a complete description of this sequential approach, we refer to Danielsson et al. (2001) where it is mentioned that the procedure also works for a much broader class of tail index estimators including Pickands' estimator, the moment estimator.

4.3.4 Cheng and Peng procedure

By minimizing the absolute coverage error for confidence intervals of level $(1-\alpha)$, Cheng and Peng (2001) proposed an optimal sample fraction as follows:

$$\widehat{k}_{opt} = \begin{cases} n^{-\widehat{\rho}/(1-\widehat{\rho})}\left(\dfrac{(1+2z_\alpha^2)}{3\widehat{\delta}\,(1-2\widehat{\rho})}\right)^{1/(1-\widehat{\rho})} & if\ \ \widehat{\delta} > 0, \\ n^{-\widehat{\rho}/(1-\widehat{\rho})}\left(\dfrac{(1+2z_\alpha^2)}{-3\widehat{\delta}}\right)^{1/(1-\widehat{\rho})} & if\ \ \widehat{\delta} < 0, \end{cases}$$

where z_α is defined by $\mathbb{P}\left(\mathcal{N}\left(0,1\right) \leq z\right) = 1-\alpha$, i.e., z_α is the $(1-\alpha)$-quantile of the standard normal distribution,

$$\widehat{\rho} = -\frac{1}{\log 2}\left(\left|\frac{M_n^{(2)}\left(n/\left(2\sqrt{\log n}\right)\right) - 2\left\{M_n^{(1)}\left(n/\left(2\sqrt{\log n}\right)\right)\right\}^2}{M_n^{(2)}\left(n/\left(\sqrt{\log n}\right)\right) - 2\left\{M_n^{(1)}\left(n/\sqrt{\log n}\right)\right\}^2}\right|\right),$$

and

$$\widehat{\delta} := \left(1+\widehat{\rho}\right)\left(\log n\right)^{-\widehat{\rho}}\frac{M_n^{(2)}\left(n/\left(\sqrt{\log n}\right)\right) - 2\left\{M_n^{(1)}\left(n/\sqrt{\log n}\right)\right\}^2}{-2\widehat{\rho}\left\{M_n^{(1)}\left(n/\sqrt{\log n}\right)\right\}^2},$$

with, $M_n^{(r)}\left(k\right), \left(r=1,2\right)$, as defined by (4.7).

4.3.5 Reiss and Thomas approach

Reiss and Thomas (1997) proposed a heuristic method of choosing the optimal number of upper extremes used in the computation of the tail index estimate.

For Hall's model, this methodology selects the value $\widehat{k}_{opt}$ of k_{opt} which minimizes the quantity

$$\frac{1}{k}\sum_{i=1}^{k} i^{\theta}\left|\widehat{\gamma}_n(i) - \text{median}\left(\widehat{\gamma}_n(1), \ldots, \widehat{\gamma}_n(k)\right)\right|, \quad 0 \le \theta \le 0.5.$$

Considering Hill's and the moment estimators, Neves and Fraga Alves (2004) discuss and evaluate the performance of this methodology by substantially reducing the domain of variation of the weight θ of the penalty term i^{θ}. Depending on the prior information one might have about the value of the EVI, the authors provide, for each estimator, suitable values of θ and indicate which expression, out of the two above, to minimize. On the light of a thorough simulation study they come up with the overall conclusion that the most proper choice for θ is 0.

4.3.6 Example

We apply the algorithms of Cheng and Peng (2001) and Reiss and Thomas (1997) on 2000 simulated observations from the standard Pareto distribution. The results are summarized in Table 3.1 and illustrated by Figure 3.4.

Algorithm	# of extremes	% of extremes	Estimate EVI
Cheng and Peng	123	6.15	1.05
Reiss and Thomas	212	10.60	0.99

Table 4.1: Optimal numbers of upper order statistics used in the computation of Hill's estimate of the EVI of Pareto(1) distribution, based on 3000 observations.

4.4 High quantile estimation

This section discusses estimators of the high quantiles for heavy-tailed distributions, i.e, quantiles of order $(1-p)$ with $0 < p < 1$ and p tending to zero as the sample size increases. The quantile of order $(1-p)$ or $(1-p)$-quantile of df F is defined to be the solution, denoted by x_p, of the equation $1 - F(x) = p$. We define, by using the functions introduced in definition 3.2.3, the $(1-p)$-quantile of F as

$$x_p = Q(1-p) = \mathbb{U}(1/p).$$

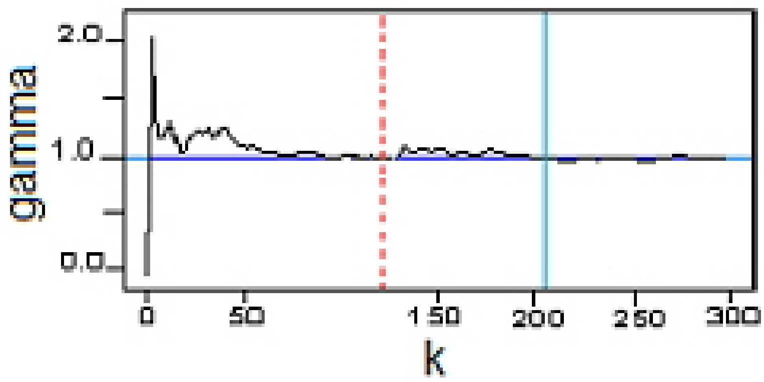

Figure 4.4: Hill's estimator of the EVI of the standard Pareto distribution, based on 3000 observations. The horizontal line represents the true value of the tail index whereas the vertical lines correspond to the optimal numbers of extremes of Cheng and Peng (dashed) and Reiss and Thomas (solid).

The estimation of high quantiles great interest in many applications such as insurance, finance, hydrology and statistical quality when the parametric form of the underlying distribution is not available. The main difficulty of this estimation is due to the fact that when p is very small, the point x_p is beyond the range of the sample $(X_1, \ldots, X_n)$ withdrawn from an unknown df F.

As we use asymptotic theory, p must depend on the sample size n, i.e., $p = p_{n.}$ Two cases are possible for x_p, within and outside the sample. If $p_n \to 0$ with $np_n \to c \in [1, \infty]$ as $n \to \infty$, the $(1 - p_n)$-quantile is within the sample and if $p_n \to 0$ with $np_n \to c \in [0, 1)$ as $n \to \infty$, the $(1 - p_n)$-quantile is outside the sample. In other words, the within-sample estimation is possible up to the $(1/n)$-quantile whereas for $p < 1/n$, quantile estimates are beyond the range of the data. The latter case is the most relevant for purposes of real-life applications.

For the first situation, we apply result (3.2) with $s = 1 - p = 1 - (i - 1)/n$ for $i = 2, \ldots, n$ and we get

$$Q_n \left(1 - \frac{i-1}{n} \right) = X_{n-i+1,n}.$$

In the second case, we have to infer beyond the limits of the sample by extrapolating from intermediate quantiles. Obviously, this cannot be done without some kind of information on the tails of the distribution. An accurate modelling of the distribution tails is then needed. In other words, a good estimate of the tail index is essential to the process of extreme quantile estimation. Since estimating high quantiles is directly linked to estimating the EVI, one would expect to find, in the literature, as many quantile estimators as there are tail index estimators. Moreover, confidence intervals for the quantile estimates are easily constructed since the proposed estimators are asymptotically normal. Finally, endpoints in case they are finite are estimated as quantiles of order 1.

Here, we discuss the simple approach proposed by Weissman (1978) based on the Hill estimator. If $F \in DA(H_\gamma)$ with $\gamma > 0$, then

$$1 - F(x) = x^{-1/\gamma}\mathbb{L}(x),$$

where $\mathbb{L}$ is a slowly varying function at infinity, that is, $\mathbb{L}(tx)/\mathbb{L}(t) \to 1$ as $t \to \infty$, $x > 0$. In terms of the quantile, we have

$$Q(1-s) = s^{\gamma}\mathbb{L}_1(s),$$

where $\mathbb{L}_1$ is a slowly varying function at infinity, that is, $\mathbb{L}(\lambda s)/\mathbb{L}(s) \to 1$ as $s \to 0$, $\lambda > 0$. Which allows to obtain

$$\frac{Q(1-\lambda s)}{Q(1-s)} = \lambda^{\gamma},$$

then

$$Q(1-\lambda s) = \lambda^{\gamma} Q(1-s).$$

By setting $\lambda s = p$ and $s = k/n$, we obtain

$$Q(1-p) = Q\left(1 - \frac{k}{n}\right)\left(\frac{np}{k}\right)^{\gamma}.$$

Then, the quantile estimator is

$$\begin{aligned} \widehat{x}_p &= \widehat{Q}(1-p) \\ &= \widehat{Q}_n\left(1 - \frac{k}{n}\right)\left(\frac{np}{k}\right)^{\widehat{\gamma}^H_{n,k}}. \end{aligned}$$

Finally, the Weissman estimator is

$$\widehat{x}_p = X_{n-k,n}\left(\frac{np}{k}\right)^{\widehat{\gamma}^H_{n,k}}.$$

4.5 Bias reduction

The problem with Hill's estimator of the EVI, is that they usually exhibit a substantial bias for large values of the number k of upper order statistics used in the estimation. This bias arises from taking asymptotic approximations as exact equalities, which is sometimes overoptimistic. To solve this problem, several bias-reduced estimators have been introduced in recent literature mainly for heavy-tailed distributions. The advantage of such estimators is that, not only they reduce the bias to a large extent but they preserve the variance level with respect to the classical estimators as well (see Feuerverger and Hall, 1999). Furthermore, the following asymptotic distributional representation holds true for any one of them (denoted by $\widehat{\gamma}_n^{BC}$)

$$\widehat{\gamma}_n^{BC} \overset{\mathcal{D}}{=} \gamma + \frac{\eta}{\sqrt{k}} Z_k + o_{\mathbb{P}} \left(A \left(n/k \right) \right) \quad \text{as} \quad n \to \infty,$$

where Z_k is an asymptotically standard normal rv, A is the function defined in (4.3) and $\eta^2/\sqrt{k}$ is the asymptotic variance of the bias-corrected estimator $\widehat{\gamma}_n^{BC}$. Hence, it is readily checked that if $\sqrt{k} A \left(n/k \right) \to \lambda$, then $\sqrt{k} \left(\widehat{\gamma}_n^{BC} - \gamma \right)$ is asymptotically normal with null mean value whatever λ, is (null or not). For a review of some explicit bias-corrected estimators, one may consult Gomes and Figueiredo (2006). As an example of such bias-corrected estimators, we consider the so-called (refined) exponential regression estimator, first introduced in Beirlant et al. (1999) and Feuerverger and Hall (1999).

Let, for intermediate k,

$$U_{ik} = \log \frac{X_{n-i+1,n}}{X_{n-k,n}} \quad \text{and} \quad V_i = i \left(\log \frac{X_{n-i+1,n}}{X_{n-i,n}} \right), \quad i = 1, \ldots, k,$$

respectively be the log-excesses and the scaled log-spacings. It is shown in, e.g., Beirlant et al. (1999) that, under the second order framework (2.4), both U_{ik} and V_i approximately follow exponential regression models. More precisely, we have

$$U_{ik} \sim \gamma E_{k-i+1,n} + A \left(n/k \right) \frac{Y_{n-i+1,n}^{\rho} - 1}{\rho} \quad \text{and} \quad V_i \sim \gamma \left(1 + \frac{A \left(n/k \right) \left(k/i \right)^{\rho}}{\gamma} \right) E_i,$$

where $E_{1,n} \leq \cdots \leq E_{n,n}$ and $Y_{1,n} \leq \cdots \leq Y_{n,n}$ are the respective order statistics pertaining to the standard exponential sample $(E_1, \ldots, E_n)$ and the unit Pareto sample $(Y_1, \ldots, Y_n)$. Beirlant et al. (1999) and Feuerverger and Hall (1999) obtain simultaneous MLE's for ρ and $A(n/k)$, whereas Gomes and Martins (2002) start with an external estimation of the second order parameter ρ and then deduce an estimate for the first order parameter γ.

Weissman's estimator of high quantiles for heavy-tailed distributions, known to be largely biased. As a better alternative to Weissman's estimators, several estimators of extreme quantiles with reduced biases are proposed in the literature, see e.g., Gomes and Martins (2004), Caeiro et al. (2009), Gomes and Figueiredo (2006) and the references therein. We derive a new estimator for $\widehat{\gamma}^H_{n,k}$ with reduced bias by applying the results of Feuerverger and Hall (1999) and Beirlant et al. (1999, 2002) who proposed, under (4.4), the following exponential regression model for the log-spacings of order statistics

$$Z_i \sim \gamma + A\left(\frac{n}{k}\right)\left(\frac{i}{k+1}\right)^{-\rho} + \varepsilon_i, \quad 1 \leq i \leq k, \tag{4.10}$$

where the ε_i are zero-centered error terms. We get the Hill estimator $\widehat{\gamma}^H$ when we ignore the term $A(n/k)$ in (4.10) and by taking the mean of the left-hand side of (4.10). We can exploit (4.10), using a least-squares approach, to propose a reduced-bias estimator for γ in which ρ is substituted by a consistent estimator $\widehat{\rho} = \widehat{\rho}(n,k)$ (see for instance Beirlant et al.,2002 and Fraga Alves et al., 2003) or by a canonical choice, such as $\rho = -1$ (see e.g., .Feuerverger and Hall, 1999 or Beirlant et al., 1999). The least-squares estimators for γ and $A(n/k)$ are then given by

$$\widehat{\gamma}^{LS}_{n,k}(\widehat{\rho}) = \frac{1}{k}\sum_{i=1}^{k} Z_i - \frac{\widehat{A}^{LS}(\widehat{\rho})}{1-\widehat{\rho}} = \widehat{\gamma}^H_{n,k} - \frac{\widehat{A}^{LS}(\widehat{\rho})}{1-\widehat{\rho}},$$

$$\widehat{A}^{LS}(\widehat{\rho}) = \frac{(1-2\widehat{\rho})(1-\widehat{\rho})^2}{\widehat{\rho}^2}\frac{1}{k}\sum_{i=1}^{k}\left(\left(\frac{i}{k+1}\right)^{-\widehat{\rho}} - \frac{1}{1-\widehat{\rho}}\right) Z_i.$$

The main asymptotic properties of $\widehat{\gamma}^{LS}_{n,k}(\widehat{\rho})$ and $\widehat{A}^{LS}(\widehat{\rho})$ as a function of Brownian bridges have been established in Deme et al. (2013, Lemma 5). Note that $\widehat{\gamma}^{LS}_{n,k}(\widehat{\rho})$ can be viewed as the kernel estimator

$$\widehat{\gamma}^{LS}_{n,k}(\widehat{\rho}) = \widehat{\gamma}^{K_\rho} = \frac{1}{k}\sum_{i=1}^{k} K_\rho\left(\frac{i}{k+1}\right) Z_i,$$

where for $0 < u \leq 1$,

$$K_\rho(u) = \frac{1-\rho}{\rho}\underline{K}(u) + \left(1 - \frac{1-\rho}{\rho}\right)\underline{K}_\rho(u),$$

with $\underline{K}(u) = \mathbf{1}_{(0<u<1)}$ and $\underline{K}_\rho(u) = \left(\frac{1-\rho}{\rho}\right)(u^{-\rho} - 1)\,\mathbf{1}_{(0<u<1)}$, both kernels satisfying condition $(\mathbb{K})$. On the contrary K_ρ does not satisfy statement KC1 in condition $(\mathbb{K})$.

Now, using the second order refinements of assumption, we can construct the following asymptotically unbiased estimator of the quantile

$$Q_n^{LS}(1-s) = (ns/k)^{-\widehat{\gamma}^{LS}(\widehat{\rho})} X_{n-k,n}\left(1 - \widehat{\rho}^{-1}\widehat{A}^{LS}(\widehat{\rho})\left(1 - (ns/k)^{-\widehat{\rho}}\right)\right).$$

The effect of the bias reduction on the MSE is illustrated in Figure 3.5 where we show the MSE computed over 100 samples of size $n = 1000$ from the Burr(1,0.5,2) distribution and $s = 0.0002$. We can observe that the MSE of the bias reduced estimator $Q_n^{LS}(1-s)$ is almost constant with respect to k.

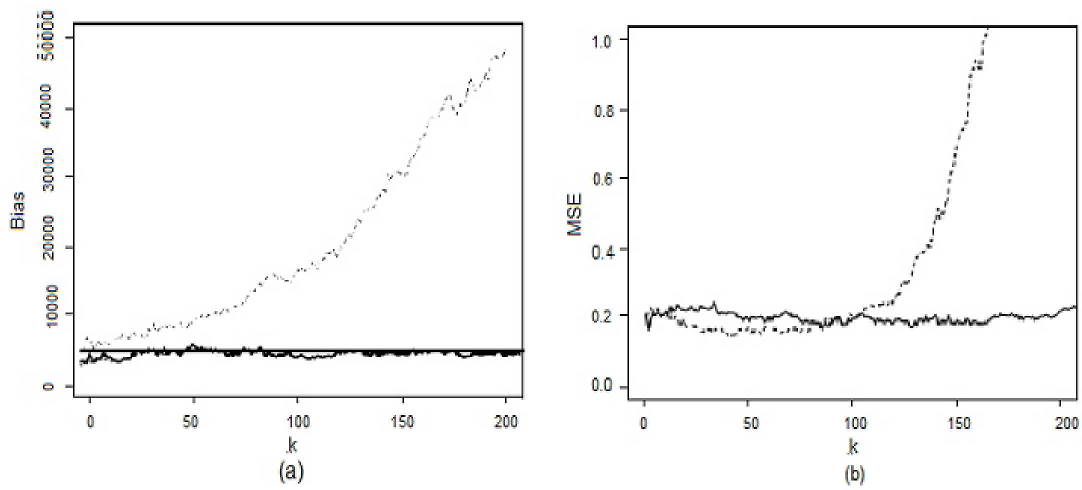

Figure 4.5: (a) Bias and (b) MSE of $Q_n^W(1-s)$ (dashed line) $Q_n^{LS}(1-s)$ (solid line) with $s = 0.0002$ for 100 simulated samples of size n = 1000 from the Burr(1,0.5,2) distribution. The horizontal line indicates the true value of $Q(1-s)$.

Chapter 5

Risk measures and premium principles

The concept of risk measures and premium principles have been studied from various angles in the actuarial literature. For an insurance companies, any contract of insurance brings a risk. A claim may occur some time in the future and the amount of the claim is a non-negative rv which is called a risk. One of the main tasks of actuaries is to calculate the risk premiums. The basis of insurance is the hypothesis that claims can be compensated by fixed payments called premiums. Premiums are calculated by a premium calculation principle. In order to quantify risk, it is necessary to specify the probability distributions of the risks involved and to apply a preference function to these probability distributions. The first use of risk measures in actuarial science was the development of premium principles. These were applied to a loss distribution to determine an appropriate premium to charge for the risk. In this chapter, we focus on the premium that accounts for the monetary payout by the insurer in connection with insurable losses plus the risk loading that the insurer imposes to reflect the fact that experienced losses rarely, if ever, equal expected losses. We give the definition of risk measures and premium principle. We discuss some desirable properties of premium calculation principles. We also list many well-known premium principles and tabulate which of the properties they satisfy.

5.1 Risk measures

5.1.1 Definition

Risk measures are used to determine the amount of an asset or set of assets, traditionally currency, to be kept in reserve. The aim of this reserve is to make the risks taken by financial institutions, such as insurance companies, acceptable to the regulator. Because in actuarial applications a risk is represented by a non-negative rv, measuring risk is equivalent to establishing a correspondence between the space of rv's and non-negative real numbers $\mathbb{R}_+$. Thus, a risk measure is a functional that assigns a non-negative real number to a risk. It is essential to understand which aspect of the riskiness associated with the uncertain outcome the risk measure attempts to quantify.

Definition 5.1.1 *(Risk measure)*
Let $\mathcal{X}$ be a set of a risks that is set of a non-negative rv's on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. A risk measure R is a functional mapping from $\mathcal{X}$ to the non-negative real numbers $\mathbb{R}_+$, possibly infinite, i.e., $R : \mathcal{X} \rightarrow \mathbb{R}_+ \cup \{\infty\}$.

5.2 Premium principles

The premium calculation principle is one of the main objectives of study for actuaries. Premium principles are the most important risk measures in actuarial sciences and frequently the insurers are also interested in measuring the upper tails of df's. There are different methods that actuaries use to develop premium principles (see, Denuit et al., 2005). In this section, we list some premium principles and discuss their desirable properties.

Definition 5.2.1 *(Premium principle)*
A premium principle Π, is a risk measure, is a function from a set of insurance risks $\mathcal{X}$ to the set of non-negative real numbers $\mathbb{R}_+$. It is possible that Π takes the value ∞, in this case we say that the risk is unacceptable or non-insurable.

5.2.1 Properties for premium principles

In this subsection, we list various properties that premium principles $\Pi[X]$ may or may not satisfy.

1. *Law invariance (independence):* For all $X, Y \in X$ and $\forall x \in \mathbb{R}$, a premium principle Π is law invariant if

   $$\Pi[X] = \Pi[Y] \text{ when } \mathbb{P}(X \leq x) = \mathbb{P}(Y \leq x).$$

   This property means that for a given risk X the premium $\Pi[X]$ depends on X only via the df $\mathbb{P}(X \leq x)$. It states that the premium depends only on the monetary loss of the insurable event and the probability that a given monetary loss occurs, not the cause of the monetary loss.

2. *Risk loading*: A premium principle Π induces a risk loading if and only if

   $$\Pi[X] \geq E(X) \text{ for all } X \in \mathcal{X}.$$

   This property is desirable because one generally requires a premium rule to charge at least the expected payout of the risk X, namely $E(X)$, in exchange for insuring the risk. Otherwise, the insurer will lose money on average.

3. *No unjustified risk loading:*

   If a risk $X \in \mathcal{X}$ is identically equal to a constant $c \geq 0$, almost everywhere, then $\Pi[X] = c$.

   This property means that, in contrast to property 2 (risk loading), if we know for certain (with probability 1) that the insurance payout is c, then we have no reason to charge a risk loading because there is no uncertainty as to the payout.

4. *Maximal loss (no rip-off):*

   $$\Pi[X] \leq \max(X), \text{ for all } X \in \mathcal{X}.$$

5. *Translation equivariance (or translation invariance):* For all $X \in \mathcal{X}$ and for all $a \geq 0$, Π is called translution invariunt if and only if

   $$\Pi[X + a] = \Pi[X] + a.$$

If we increase a risk X by a fixed amount a, then property 5 states that the premium for $X + a$ should be the premium for X increased by that fixed amount a.

6. *Scale equivariance (scale invariance)*: Π is called scale invariunt if and only if
$$\Pi[bX] = b\Pi[X] \text{ for all } X \in \mathcal{X} \text{ and for all } b \geq 0.$$
Note that properties 5 and 6 imply property 3 as long as there exists a risk Y such that $\Pi[Y] < \infty$. This property is also known as homogeneity of degree one in the economics literature. This property essentially states that the premium for doubling a risk is twice the premium of the single risk. One usually uses a noarbitrage argument to justify this rule. Indeed, if the premium for $2X$ were greater than twice the premium of X, then one could buy insurance for $2X$ by buying insurance for X with two different insurers, or with the same insurer under two policies.

7. *Additivity:* For all $X, Y \in \mathcal{X}$, a premium principle Π is called additive if and only if
$$\Pi[X + Y] = \Pi[X] + \Pi[Y].$$
This is a stronger form of property 6 (scale equivariance). One can use a similar no-arbitrage argument to justify the additivity property.

8. *Subadditivity:* A premium principle Π is called subadditive if and only if
$$\Pi[X + Y] \leq \Pi[X] + \Pi[Y] \text{ for all } X, Y \in \mathcal{X}.$$
Subadditivity is a reasonable property because the no-arbitrage argument works well to ensure that the premium for the sum of two risks is not greater than the sum of the individual premiums; otherwise, the buyer of insurance would simply insure the two risks separately. However, the no-arbitrage argument that asserts that $\Pi[X + Y]$ cannot be less than $\Pi[X] + \Pi[Y]$ fails because it is generally not possible for the buyer of insurance to sell insurance for the two risks separately.

9. *Superadditivity:*
$$\Pi[X + Y] \geq \Pi[X] + \Pi[Y] \text{ for all } X, Y \in \mathcal{X}.$$

Superadditivity might be a reasonable property of a premium principle if there are surplus constraints that require that an insurer charge a greater risk load for insuring larger risks. For example, we might observe in the market that $\Pi[2X] > 2\Pi[X]$ because of such surplus constraints. Note that both properties 8 and 9 can be weakened by requiring only $\Pi[bX] \leq b\Pi[X]$ or $\Pi[bX] \geq b\Pi[X]$ for $b > 0$, respectively. Next, we weaken the additivity property by requiring additivity only for certain insurance risks.

10. *Additivity for independent risks:*

    $\Pi[X + Y] = \Pi[X] + \Pi[Y]$ for all $X, Y \in \mathcal{X}$, such that X and Y are independent.

    Some actuaries might feel that property 7 (additivity) is too strong and that the no-arbitrage argument only applies to risks that are independent. They avoid the problem of surplus constraints for dependent risks.

11. *Additivity for comonotonic risks:*

    $\Pi[X + Y] = \Pi[X] + \Pi[Y]$ for all $X, Y \in \mathcal{X}$, such that X and$Y$ are comonotonic.

    Additivity for comonotonic risks is desirable because if one adopts sub-additivity as a general rule, then it is unreasonable to have $\Pi[X + Y] < \Pi[X] + \Pi[Y]$ because neither risk is a hedge against the other, that is, they move together (see Yaari, 1987),. If a premium principle is additive for comonotonic risks, then is it layer additive (see Wang, 1996). Note that property 11 implies property 6, (scale equivariance), if Π additionally satisfies a continuity condition. Next, we consider properties of premium rules that require that they preserve common orderings of risks.

12. *Monotonicity:*

    Let $X, Y \in \mathcal{X}$, If $X(\omega) \leq Y(\omega)$ for all $\omega \in \Omega$, then $\Pi[X] \leq \Pi[Y]$.

13. *Preserves first stochastic dominance (FSD) ordering*

    If $S_X(t) \leq S_Y(t)$ for all $t \geq 0$, then $\Pi[X] \leq \Pi[Y]$.

14. *Preserves stop-loss ordering (SL) ordering:*

$$\text{If } E\left(X-d\right)_{+} \leq E\left(Y-d\right)_{+} \text{ for all } d \geq 0, \text{ then } \Pi[X] \leq \Pi[Y].$$

Property 1, (Independence), together with property 12, (Monotonicity), imply property 13, (preserves FSD ordering). Also, if Π preserves SL ordering, then Π preserves FSD ordering because stop-loss ordering is weaker. These orderings are commonly used in actuarial science to order risks because they represent the common orderings of groups of decision makers (see, Kaas et al. 1994), for example.

15. *Continuity:*

Let $X \in \mathcal{X}$, then,

$$\lim_{a \to 0^{+}} \Pi\left[\max\left(X-a\right), 0\right] = \Pi[X] \text{ and } \lim_{a \to \infty} \Pi\left[\min\left(X, a\right)\right] = \Pi[X].$$

16. *Iterativity:*

$$\Pi[X+Y] = \Pi\left[\Pi\left[X|Y\right]\right] \quad \text{for all} \quad X, Y \in \mathcal{X}.$$

The premium for X can be calculated in two steps. First, apply $\Pi[X]$ to the conditional distribution of X, given $Y = y$. The resulting premium is a function $h(y)$, say, of y. Then, apply the same premium principle to the rv $\Pi[X|Y] = h(Y)$.

Coherence

Several authors have selected some of these conditions to form a set of requirements that any risk measure should satisfy. The following definition is taken from the seminal paper of Artzner et al. (1999).

Definition 5.2.2 *(Coherence)*
A coherent risk measure is a function that satisfies properties of monotonicity, subadditivity, homogeneity and translational invariance.

5.2.2 Various premium principles

We give some premium principles. For more details, we refer to Young (2004).

A. *Net Premium Principle:* The net premium principle is given by

$$\Pi[X] = E(X).$$

This premium does not load for risk. It is the first premium principle that many actuaries learn. It is widely applied in the literature because actuaries often assume that risk is essentially non-existent if the insurer sells enough identically distributed and independent policies.

B. *Expected Value Premium Principle:* The expected value principle is given by

$$\Pi[X] = (1+\theta)E(X), \ \theta > 0.$$

This premium principle builds on the net premium principle, by including a proportional risk load. It is commonly used in insurance economics and in risk theory. The expected value principle is easy to understand and to explain to policyholders.

C. *Variance Premium Principle:* The variance principle is given by

$$\Pi[X] = E(X) + \alpha Var(X), \ \alpha > 0.$$

The loading is proportional to $Var(X)$. This principle counts two characteristics of the risk - the mean value and the variance and is more sensible to higher risks. The premium principle also build on the net premium principle by including a risk load that is proportional to the variance (standard deviation) of the risk. Bühlmann (1970) studied this premium principle in detail.

D. *Standard Deviation Premium Principle:* This premium is given by

$$\Pi[X] = E(X) + \beta\sqrt{Var(X)}, \ \beta > 0.$$

The loading is proportional to the standard deviation of X. The loss can be written as

$$\Pi[X] - X = \sqrt{Var(X)}\left(\beta - \frac{X - E(X)}{\sqrt{Var(X)}}\right),$$

or the loss is equal to the loading parameter minus a rv with mean value 0 and variance 1.

E. *Exponential Premium Principle:* This premium is defined by the equation

$$\Pi[X] = \frac{1}{\theta}\log(E\left(\exp(\theta X)\right)),\ \theta > 0.$$

This premium principle arises from the principle of equivalent utility when the utility function is exponential.

F. *Esscher Premium Principle:* For some $\theta > 0$ and for some rv Z, Bühlmann (1980) derived this premium principle when he studied risk exchanges

$$\Pi[X] = \frac{E(X exp(\theta Z))}{E(exp(\theta Z))}.$$

G. *Principle of Equivalent Utility:* The equivalent utility premium Π is derived by solving the following equation

$$u(w) = E\left[u\left(w - X + \Pi\right)\right],$$

where u is an increasing, concave utility of wealth (of the insurer), and w is the initial wealth of the insurer.

H. *Swiss Premium Principle:* For a given non-negative and non-decreasing function u on $\mathbb{R}$ and a given parameter $p \in [0, 1]$, the Swiss premium Π is the root of

$$E\left(u\left(X - p\Pi\right)\right) = u\left(\left(1 - p\right)\Pi\right).$$

I. *Proportional Hazards Premium Principle:*

$$\Pi[X] = \int_0^{\infty} \left(\overline{F}\left(x\right)\right)^{\rho} dx,$$

where $\rho \geq 0$ is called a risk index or distortion parameter. This parameter controls the amount of the risk loading included in the premium for given riskiness of the loss variable X. Wang (1996) studied the many nice properties of this premium principle.

J. *Distortion Risk Premium Principle*:

$$\Pi[X] = \int_0^\infty g\left(\overline{F}(x)\right) dx,$$

where g is an increasing function that maps $[0,1]$ onto $[0,1]$. The function g is called a distortion and $g\left(\overline{F}(x)\right)$ is called a distorted probability.

Distortion risk premium principles have their origin in Yaari's (1987) dual theory of choice under risk that consists in measuring the risks by applying a distortion function g on the df F. The net premium principle and proportional hazards premium principle are a special cases of distortion risk premium principle with the distortions g given by $g(s) = s$ and $g(s) = s^\rho$ respectively. We give some distortion functions in section 3 of chapter 6. For the other distortions, see Wang (1996).

In the Table 5.1, we list the properties of precedents premium principles where 'Y' indicates that the premium principle satisfies the given property whereas 'N' indicates that the premium principle does not satisfy.

Premium principle letter ⟶ Property number ↓	A	B	C	D	E	F	G	H	I	J
1	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
2	Y	Y	Y	Y	Y	N	Y	Y	Y	Y
3	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
4	Y	N	N	N	Y	Y	Y	Y	Y	Y
5	Y	N	Y	Y	N	Y	Y	N	Y	Y
6	Y	Y	N	Y	N	N	N	N	Y	Y
7	Y	Y	N	N	N	N	N	N	N	N
8	Y	Y	N	N	N	N	N	N	Y	Y
9	Y	Y	N	N	Y	N	N	N	N	N
10	Y	Y	Y	N	N	N	N	N	N	N
11	Y	Y	N	N	Y	N	N	N	Y	Y
12	Y	Y	N	N	Y	N	Y	N	Y	Y
13	Y	Y	N	N	Y	N	Y	Y	Y	Y
14	Y	Y	N	N	Y	N	Y	Y	Y	Y
15	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
16	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Table 5.1: Properties of premium principles.

Chapter 6

Empirical estimation of the distortion risk premiums for heavy-tailed losses

In this chapter, we deal with the empirical estimation of the distortion risk premiums for heavy tailed losses by making use of the EVT. However, this approach has often a strong asymptotic bias in the estimation. Therefore we look at this framework here and propose a bias-reduced of the classical estimators already suggested in the literature. A simulation study is provided in order to prove the efficiency of our method.

6.1 Semi-parametric estimator for the distortion risk premiums

As mentioned in chapter 5, the risk premiums are used to quantify insurance losses. One of the most commonly used is the net premium

$$\Pi = E\left(X\right) = \int_0^\infty \overline{F}\left(X\right)dx.$$

In order to avoid that the insurer loses money on average, premiums are required to be greater than or equal to the net premium Π (see property 2, risk loading). To achieve this goal Wang (1996) proposed the distortion risk premium, which is defined by

$$\Pi\left(g\right) = \int_0^\infty g\left(\overline{F}\left(X\right)\right)dx,$$

where $g : [0,1] \to [0,1]$ is a non-decreasing function, called distortion function, satisfying $g(0) = 0$ and $g(1) = 1$. In general, the function g is parametrized by a one dimentional parameter called the distortion parameter or the risk aversion index. This parameter controls the amount of the risk loading included in the premium. Along this chapter, we assume that F is a continous loss distribution.

Changing variables and integrating by parts yield the following expression for $\Pi(g)$, in terms of the quantile function Q,

$$\Pi(g) = \int_0^1 g'(s)\, Q(1-s)\, ds, \tag{6.1}$$

where, g' denotes the Lebesgue derivative of g.

As mentioned in chapter 5, an example of such a distortion function g is $g(x) = x^{1/r}$, for $r \geq 1$. Therefore, we obtain the Proportional Hazard Premium (PHP) defined as follows

$$\Pi_r(X) = \int_0^\infty \left(\overline{F}(x)\right)^r dx.$$

Note that this quantity is not necessarily finite. The parameter $r \geq 1$ represents the distortion coefficient. When the parameter is at its minimal value, $r = 1$, then $\Pi_r(X)$ is the net premium Π, and thus there is no loading. The risk loading increases when r increases. This class of premium has been extensively studied in the literature in particular as it can also be grounded in economics via Yaari's (1987) dual theory of expected utility.

For a high layer with retention level $R > 0$, the corresponding PHP or the reinsurance premium is defined as

$$\Pi_{r,R} = \int_R^\infty \left(\overline{F}(x)\right)^{1/r} dx.$$

For recent literature on statistical inference for distortion premiums, we refer to Peng et al. (2001) Jones and Zitikis (2003), Necir and Boukhetala (2004), Centeno and Andrade (2005), Necir et al. (2007), Jones and Zitikis (2003, 2007), Brazauskas et al. (2008), Necir and Meraghni (2009), Necir et al. (2010) Necir and Zitikis (2011), Brahimi et al. (2012), Deme et al. (2013), Deme and Lo (2013), Benkhelifa (2014a, 2014b) and the references therein.

To estimate $\Pi(g)$, we consider a random sample given by iid rv's whose the common df is that of the risk X and let $X_{1,n}, \ldots, X_{n,n}$ be the corresponding order statistics. The empirical estimator of $\Pi(g)$ is obtained by replacing the true quantile Q on the right-hand side of formula (6.1) by the sample quantiles Q_n, see (3.2). Then, we obtain the following natural estimator

$$\widehat{\Pi}(g) = \sum_{i=k+1}^{n} a_{i,n}(g) X_{n-i+1,n}, \tag{6.2}$$

whose right-hand side is an L-statistic with the coefficients

$$a_{i,n}(g) = g\left(\frac{i}{n}\right) - g\left(\frac{i-1}{n}\right), \ i = 1, \ldots, n.$$

Jones and Zitikis (2003) showed that, using the asymptotic theory for L-statistics (see, e.g., Shorack and Wellner, 1986) for underlying distributions with a sufficient number of finite moments and under certain regularity conditions on the distortion function g, the asymptotic normality of $\widehat{\Pi}(g)$

$$\sqrt{n}\left(\widehat{\Pi}(g) - \Pi(g)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_g^2\right), \tag{6.3}$$

provided that the variance

$$\sigma_g^2 = \int_0^1 \int_0^1 (\min(s,t) - st) g'(s) g'(t) dQ(1-s) dQ(1-t) < \infty.$$

The same authors are also discussed the PHP estimators. In this case the asymptotic normality (6.3) holds for any $1 < \gamma < 2$, provided that

$$E(X^\eta) < \infty \text{ for some } \eta > 2r/(2-r).$$

We assume that the distortion functions g are such that $t \mapsto g(t)$ are regularly varying at zero with index $1/\beta \in (0,1]$, i.e.,

$$g(t) = t^{1/\beta} \ell_g(t), \tag{6.4}$$

where ℓ_g is a slowly varying function at zero satisfying $\ell_g(\lambda t)/\ell_g(t) \to 1$ as $t \to 0$ for $\lambda > 0$. In this section, we present an asymptotically normal semi-parametric estimator for the distortion risk premiums $\Pi(g)$ of heavy-tailed claim amounts. Moreover, we focus on the case $\gamma \in (1/2, 1)$ and $\beta \in [1, 1/\gamma)$ in order to ensure that the $\Pi(g)$ is finite and since in that case the results (6.3) of Jones and Zitikis (2003) cannot be applied, the second moment of X, $E(X^2)$, being infinite and thus we need to seek another approach to handle this situation. By making use of the EVT, Necir and Meraghni (2009) proposed an alternative estimator for $\Pi(g)$ and established its asymptotic normality for any $\gamma \in (1/2, 1)$.

6.1.1 Defining the estimator

As above mentioned, the use of empirical quantiles to estimate risk premiums $\Pi(g)$ does not garantee the asymptotic normality when losses have a heavy-tailed distribution. We have

$$\Pi(g) = \int_0^1 g'(s) Q(1-s) ds,$$

then

$$\Pi(g) = \left\{ \int_{k/n}^1 g'(s) Q(1-s) ds \right\} + \left\{ g(k/n) Q(1-k/n) - \int_0^{k/n} g(s) dQ(1-s) \right\}$$

$$= \Pi_1(g) + \Pi_2(g).$$

Note that $X_{n-k,n}$ is the simple estimator of $Q(1-k/n)$. Hence, coming back to the quantile $Q(1-s)$ we estimate it by using the empirical estimator $Q_n(1-s)$ when $s \in (k/n, 1)$ and by using the Weissman's estimator $Q_n^W(1-s)$ when $s \in (k/n, 1)$. Therefore, the estimator of $\Pi_1(g)$ is

$$\widetilde{\Pi}_1(g) = \sum_{i=k+1}^{n} a_{i,n}(g) X_{n-i+1,n},$$

where the coefficients $a_{i,n}(g)$ are those of the L-statistic $\widehat{\Pi}(g)$ defined in (6.2). Since the distortion functions g satisfy the condition (6.4), with $\beta \in [1, 1/\gamma)$ and since $\gamma_{n,k}^H$ is a consistent estimator of γ (see Masson, 1982) then we have, for all large values of n and $\mathbb{P}\left(\gamma_{n,k}^H > 1/\beta\right) = o(1)$,

$$\begin{aligned} -\int_0^{k/n} g(s) dQ^W(1-s) &= \gamma_{n,k}^H \left(\frac{k}{n}\right)^{\gamma_{n,k}^H} X_{n-k,n} \int_0^{k/n} s^{-1-\gamma_{n,k}^H} g(s) ds \\ &= \frac{\gamma_{n,k}^H}{\frac{1}{\beta} - \gamma_{n,k}^H} g(k/n) X_{n-k,n} (1+o(1)). \end{aligned}$$

Then, we may estimate $\Pi_2(g)$ by

$$\begin{aligned} \widetilde{\Pi}_2(g) &= g(k/n) X_{n-k,n} + \frac{\gamma_{n,k}^H}{\frac{1}{\beta} - \gamma_{n,k}^H} g(k/n) X_{n-k,n} \\ &= g(k/n) \frac{X_{n-k,n}}{1 - \beta\gamma_{n,k}^H}. \end{aligned}$$

Finally, the estimator of $\Pi(g)$ is

$$\widetilde{\Pi}(g) = g(k/n)\frac{X_{n-k,n}}{1-\beta\gamma_{n,k}^{H}} + \sum_{i=k+1}^{n} a_{i,n}(g)\, X_{n-i+1,n}, \text{ for } \gamma_{n,k}^{H} < 1/\beta. \tag{6.5}$$

In a similar way, for a fixed $r \geq 1$ and at an optimal retention level $R = R_{opt} = Q(1-k/n)$, Necir et al. (2007) proposed the following semi-parametric estimator for $\Pi_{r,R}$

$$\widetilde{\Pi}_{r,\widehat{R}_{opt}} = (k/n)^{1/r}\frac{r}{1/\widehat{\gamma}_{n,\,k}^{H} - r}X_{n-k,n}, \text{ for } \widehat{\gamma}_{n,\,k}^{H} < 1/r.$$

Deme and Lo (2013) gave an universal estimator of the distortion risk premiums $\Pi(g)$ as follows

$$\widetilde{\Pi}^{*}(g) = \widetilde{\Pi}(g)\,\mathbf{1}_{\{\sigma_g^2=\infty\}} + \widetilde{\Pi}(g)\,\mathbf{1}_{\{\sigma_g^2<\infty\}},$$

where $\widetilde{\Pi}(g)$ is as in (6.5). More precisely

$$\widetilde{\Pi}^{*}(g) = \widetilde{\Pi}(g)\,\mathbf{1}_{\{S(\gamma,\beta)\}} + \widetilde{\Pi}(g)\,\mathbf{1}_{\{\overline{S}(\gamma,\beta)\}},$$

where

$$S(\gamma,\beta) = \left\{(\gamma,\beta) \in (0,\infty)\times[1,\infty),\ \gamma \in (1/2,1) \text{ and } \beta < \frac{1}{\gamma}\right\},$$

and $\overline{S}(\gamma,\beta)$ is its complementary in $(0,\infty)\times[1,\infty)$.

In the literature, a number of special cases that are covered by statistical inferential theory for $\Pi(g)$ have been investigated based on EVT. We refer to Peng et al. (2001), Necir and Boukhetala (2004), Necir et al. (2010), Necir and Zitikis (2011), Brahimi et al. (2012), Deme et al. (2013) and Benkhelifa (2014a). The estimator $\widetilde{\Pi}(g)$ is also used by Necir and Zitikis (2011) in order to introduce an estimator of a coupled risk premiums for heavy-tailed losses.

6.1.2 Asymptotic normality

In the following theorems, Necir and Meraghni (2009) and Necir et al. (2007, 2010) established the asymptotic normality of the estimators $\widetilde{\Pi}(g)$ and $\widetilde{\Pi}_{r,\widehat{R}_{opt}}$ respectively.

Theorem 6.1.1 *(Asymptotic normality of $\widetilde{\Pi}(g)$: Necir and Meraghni, 2009) Assume that F satisfies* (2.4) *with $\gamma \in (1/2, 1)$ and its corresponding quantile function $Q(\cdot)$ is continuously differentiable on $[0,1)$. Let For any differentiable distortion function g satisfying (6.4) with $1 \leq \beta < 1/\gamma$ and any sequence of integer $k = k_n$ satisfying $k \to \infty$, $k/n \to 0$ and $\sqrt{k}A(n/k) \to \lambda \in \mathbb{R}$ as $n \to \infty$, we have*

$$\frac{\sqrt{k}\left(\widetilde{\Pi}(g) - \Pi(g)\right)}{g(k/n)\,Q(1-k/n)} \xrightarrow{\mathcal{D}} \mathcal{N}\left(\lambda \mathcal{AB}(\gamma,\beta,\rho), \mathcal{AV}(\gamma,\beta)\right), \ \ as\ n \to \infty,$$

where

$$\mathcal{AB}(\gamma,\beta,\rho) = \frac{\beta\rho(\gamma\beta+\beta-1)}{(1-\rho)(\beta\gamma+\rho\beta-1)(1-\beta\gamma)^2},$$

and

$$\mathcal{AV}(\gamma,\beta) = \frac{\gamma^2\beta(\gamma\beta+\beta-1)^2}{(2\gamma\beta+\beta-2)(1-\gamma\beta)^4}.$$

Theorem 6.1.2 *(Asymptotic normality of $\widetilde{\Pi}_{r,\widehat{R}_{opt}}$: Necir et al., 2007, 2010) Fix $r \geq 1$ and assume hat F satisfies* (2.4) *with $t^{-1/r}\mathbb{U}(t)$ as $t \to \infty$. If $k \to \infty$, $k/n \to 0$ and $\sqrt{k}A(n/k) \to \lambda \in \mathbb{R}$ as $n \to \infty$, then for $1 \leq r < 1/\gamma$, we have as $n \to \infty$,*

$$\frac{(k/n)^{-1/r}\sqrt{k}}{X_{n-k,n}}\left(\widetilde{\Pi}_{r,\widehat{R}_{opt}} - \Pi_{r,R_{opt}}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\lambda\frac{r\rho(r\gamma+r-1)}{(1-\rho)(r\gamma+r\rho-1)(1-r\gamma)^2}, \frac{r^2\gamma^2(\gamma^2+r^2\gamma^4-2r\gamma^3+1)}{(1-r\gamma)^4}\right).$$

6.2 Kernel-type estimators for reinsurance premium [11]

Abstract[1]**.** We generalize the classical estimator of the reinsurance premium for heavy-tailed loss distributions with a kernel-type estimator. Because this estimator exhibits a bias, we propose its bias-reduced version by using a least-squares method. The asymptotic normality of the proposed estimators is established under suitable assumptions. A small simulation study is carried out to prove the performance of our approach.

6.2.1 Introduction

The major worry for the insurance and reinsurance companies is to determine the adequate premium. In the insurance literature, there exist several premium principles such as: expected value, variance and value-at-risk. For more details on premium principles and their properties, we refer to Goovaerts et al. (1984). Wang (1996) proposed a premium principle named proportional hazard premium (PHP) of an insured risk X, a non-negative rv defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with continuous df F, depends on the hazard function $\overline{F}$ and a parameter $r \geq 1$ called the risk aversion index or the distortion parameter. The PHP is defined as follows

$$\Pi_r = \int_0^\infty \left(\overline{F}(x)\right)^{1/r} dx.$$

In some actuarial problems, as in the reinsurance treaty, one is interested in the estimation of a premium for a given retention level $R > 0$ notation $\Pi_{r,R}$, that is, a reinsurance premium of the high layer $[R, \infty)$. This type of problem can be found whenever the insured represents a dangerous level of risk for the insurance company, and decides to give a part of this loss to another reinsurance company, because it may not have sufficient capital to cover the total risk. The reinsurance premium of the high layer is defined as follows

$$\Pi_{r,R} = \int_R^\infty \left(\overline{F}(x)\right)^{1/r} dx.$$

[1]This section is a paper appeared in Journal: *Insurance: Mathematics and Economics* 59, 65-70.

For heavy-tailed distributions, Beirlant al. (2001), Necir and Boukhetala (2004), Vandewalle and Beirlant (2006) and Necir et al. (2007) have introduced and studied different estimators for $\Pi_{r,R}$, in the case of high-excess loss layers ($R \to \infty$).

A df F is said to be heavy-tailed whenever the tail function $1 - F$ is a regularly varying function with index $(-1/\gamma) < 0$, i.e., for any $x > 0$,

$$\overline{F}(x) = x^{-1/\gamma}\mathbb{L}(x),$$

where $\mathbb{L}$ is a slowly varying function at infinity, that is, $\mathbb{L}(tx)/\mathbb{L}(t) \to 1$ as $t \to \infty$. The class of regularly varying functions includes popular distributions such those Pareto's, Burr's, Student's, Fréchet's, α-stable ($0 < \alpha < 2$), and log-gamma, which are known to be appropriate models of fitting large insurance claims, large fluctuations of prices, log-returns, and so on (see Beirlant al., 2001).

Let $X_{1,n} \leq \cdots \leq X_{n,n}$, $n \geq 1$, be the order statistics pertaining to a sample $X_1, \ldots, X_n$ from X and let $k = k_n$ be an integer sequence satisfying

$$1 < k < n, \ k \to \infty, \text{ and } k/n \to 0 \text{ as } n \to \infty.$$

Let, for $0 < s < 1$, $Q(s) = \inf\{x : F(x) \geq s\}$ be the quantile function pertaining to F. At an optimal retention level $R = R_{opt} = Q(1 - k/n)$, the semi-parametric estimator for $\Pi_{r,R}$ that proposed by Necir et al. (2007) is

$$\widetilde{\Pi}_{r,\widehat{R}_{opt}} = (k/n)^{1/r} \frac{r}{1/\widehat{\gamma}_{n,\,k}^{H} - r} X_{n-k,n}, \text{ for } \widehat{\gamma}_{n,\,k}^{H} < 1/r,$$

where $\widehat{R}_{opt} = X_{n-k,n}$ and $\widehat{\gamma}_{n,\,k}^{H}$ is the classical Hill estimator (Hill , 1975) of the tail index γ, defined by

$$\widehat{\gamma}_{n,\,k}^{H} = \frac{1}{k}\sum_{i=1}^{k} i\left(\log X_{n-i+1,n} - \log X_{n-i,n}\right).$$

A major drawback of the Hill estimator is the discrete character of its behavior in the sense that increasing k by 1, can change the actual value of the estimate considerably. Using a kernel function K, Csörgő et al. (1985) proposed a smoother version of Hill's estimator defined by

$$\widehat{\gamma}_{n,\,k}^{K} = \frac{1}{k}\sum_{i=1}^{k} K\left(\frac{i}{k+1}\right) Z_i,$$

where $Z_i = i(\log X_{n-i+1,n} - \log X_{n-i,n})$. The class of kernel estimators $\widehat{\gamma}_{n,\,k}^{K}$ generalizes the Hill estimator. Note that, using the uniform kernel $K = \underline{K} = \mathbf{1}_{(0,1)}$ yields Hill's estimator $\widehat{\gamma}_{n,\,k}^{H}$ as a special case.

We propose a kernel-type estimator for the reinsurance premium $\Pi_{r,R_{opt}}$ of a heavy-tailed distribution. Thus, $\Pi_{r,R_{opt}}$ can be estimated by

$$\widehat{\Pi}_{r,\widehat{R}_{opt}}^{K} = (k/n)^{1/r} \frac{r}{1/\widehat{\gamma}_{n,\,k}^{K} - r} X_{n-k,n}, \text{ for } \widehat{\gamma}_{n,\,k}^{K} < 1/r. \tag{6.6}$$

In subsection 6.2.2,we study the asymptotic properties of $\widehat{\Pi}_{r,\widehat{R}_{opt}}^{K}$ and propose its bias-reduced version whose asymptotic normality is also obtained. In subsection 6.2.3, we perform a small simulation study, by sampling from Fréchet distribution, to compare these estimators. All proofs are given in subsection 6.2.4.

6.2.2 Main results

Firstly, in this subsection, we study the asymptotic normality of $\widehat{\Pi}_{r,\widehat{R}_{opt}}^{K}$.

6.2.2.1 Asymptotic normality of $\widehat{\Pi}_{r,\widehat{R}_{opt}}^{K}$

From (6.6), it is clear that the asymptotic normality of $\widehat{\Pi}_{r,\widehat{R}_{opt}}^{K}$ is related to $\widehat{\gamma}_{n,\,k}^{K}$. To establish such a type of result, as usual in the extreme value theory, we need a second-order condition on the tail quantile function $\mathbb{U}$ defined, for $1 < t < \infty$, as

$$\mathbb{U}(t) = (1/(1-F))^{-1}(t) = Q(1 - 1/t).$$

We say that the function $\mathbb{U}$ fulfills the second-order regular variation condition with second-order parameter $\rho < 0$ if there exists a function $A(t)$ tending to 0 and not changing sign near infinity, such that for all $x > 0$

$$\lim_{t\to\infty} \frac{\log \mathbb{U}(tx) - \log \mathbb{U}(t) - \gamma \log x}{A(t)} = \frac{x^{\rho} - 1}{\rho}. \tag{6.7}$$

We also need the following classical conditions about the kernel K.

Condition ($\mathbb{K}$). Let K be a function defined on $(0,1]$.

(i) $K(s)\geq 0$, whenever, $0<s\leq 1$ and $K(1)=K'(1)=0$.

(ii) $K(\cdot)$ is differentiable, non-increasing and right continuous on $(0,1]$.

(iii) K and K' are bounded.

(iv) $\int_0^1 K(u)du=1$.

(v) $\int_0^1 u^{-1/2}K(u)du<\infty$.

Theorem 6.2.1 *Let F be a df satisfying (6.7) with $\gamma\in(1/2,1)$ and suppose that ($\mathbb{K}$) holds. Let $k=k_n$ be an integer sequence satisfying $k\to\infty$, $k/n\to 0$ and $\sqrt{k}A(n/k)=O(1)$ as $n\to\infty$. For any $1\leq r<1/\gamma$, on the probability space $(\Omega,\mathcal{A},\mathbb{P})$, there exists a sequence of Brownian bridges $\{\mathbb{B}_n(s);\ 0\leq s\leq 1\}$ such that, as $n\to\infty$,*

$$\frac{(k/n)^{-1/r}\sqrt{k}}{\mathbb{U}(n/k)}\left(\widehat{\Pi}^{K}_{r,\widehat{R}_{opt}}-\Pi_{r,R_{opt}}\right)=\sqrt{k}A(n/k)\mathcal{AB}_K(\gamma,r,\rho)+\mathcal{W}_{1,n}+\mathcal{W}_{2,n}(K)+o_{\mathbb{P}}(1),$$

where

$$\mathcal{AB}_K(\gamma,r,\rho)=\frac{r}{1-r\gamma}\left(\frac{1}{r\gamma+r\rho-1}+\frac{1}{1-r\gamma}\int_0^1 s^{-\rho}K(s)\,ds\right),$$

and

$$\begin{cases}\mathcal{W}_{1,n}=-\dfrac{r\gamma^2}{1-r\gamma}\sqrt{\dfrac{n}{k}}\mathbb{B}_n\left(1-\dfrac{k}{n}\right),\\ \mathcal{W}_{2,n}(K)=\dfrac{r\gamma}{(1-r\gamma)^2}\sqrt{\dfrac{n}{k}}\displaystyle\int_0^1 s^{-1}\mathbb{B}_n\left(1-\dfrac{sk}{n}\right)d(sK(s)).\end{cases}$$

Corollary 6.2.1 *Under the assumptions of theorem 6.2.1, if $\sqrt{k}A(n/k)\to\lambda\in\mathbb{R}$, we have*

$$\frac{(k/n)^{-1/r}\sqrt{k}}{\mathbb{U}(n/k)}\left(\widehat{\Pi}^{K}_{r,\widehat{R}_{opt}}-\Pi_{r,R_{opt}}\right)\xrightarrow{\mathcal{D}}\mathcal{N}\left(\lambda\mathcal{AB}_K(\gamma,r,\rho),\mathcal{AV}_K(\gamma,r)\right),\ \text{as } n\to\infty,$$

where

$$\mathcal{AV}_K(\gamma,r)=\frac{r^2\gamma^4}{(1-r\gamma)^2}+\frac{r^2\gamma^2}{(1-r\gamma)^4}\int_0^1 K^2(s)ds.$$

Corollary 6.2.1 generalizes theorem 2 in Necir et al. (2007, 2010) when $\lambda\neq 0$ and when we use a general kernel instead of $\underline{K}$ (see theorem 6.1.2). In view of these results, $\widehat{\Pi}^{K}_{r,\widehat{R}_{opt}}$ is an estimator of $\Pi_{r,R_{opt}}$ with an asymptotic bias given by

$$(k/n)^{1/r}\mathbb{U}(n/k)\,A(n/k)\mathcal{AB}_K(\gamma,r,\rho).$$

For any kernel K, we can compute the asymptotic bias and variance. If $K = \underline{K}$, we have the following corollary (similar of theorem Necir et al., 2007, 2010 see theorem 6.1.2).

Corollary 6.2.2 *Under the assumptions of corollary 6.2.1, and in the special case where $K = \underline{K}$, we have as $n \to \infty$*

$$\frac{(k/n)^{-1/r}\sqrt{k}}{\mathbb{U}(n/k)}\left(\widehat{\Pi}^{\underline{K}}_{r,\widehat{R}_{opt}} - \Pi_{r,R_{opt}}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\lambda\frac{r\rho\,(r\gamma+r-1)}{(1-\rho)\,(r\gamma+r\rho-1)\,(1-r\gamma)^2},\right.$$
$$\left.\frac{r^2\gamma^2\,(\gamma^2+r^2\gamma^4-2r\gamma^3+1)}{(1-r\gamma)^4}\right).$$

Next, in the following subsection, we propose a bias-reduced estimator for $\Pi_{r,R_{opt}}$.

6.2.2.2 Bias-reduced estimator for $\Pi_{r,R_{opt}}$

From theorem 6.2.1, we have

$$\widehat{\Pi}^{K}_{r,\widehat{R}_{opt}} - (k/n)^{1/r}\mathbb{U}(n/k)\,A(n/k)\mathcal{AB}_K(\gamma,r,\rho) \tag{6.8}$$

is an asymptotically unbiased estimator for $\Pi_{r,R_{opt}}$. Note that γ, ρ, $\mathbb{U}(n/k)$ and $A(n/k)$ are unknown quantities that we have to estimate.

Feuerverger and Hall (1999) and Beirlant et al. (1999, 2002), using (6.7), proposed the following exponential regression model for the log-spacings of order statistics

$$Z_i \sim \gamma + A\left(\frac{n}{k}\right)\left(\frac{i}{k+1}\right)^{-\rho} + \varepsilon_i,\ 1 \le i \le k, \tag{6.9}$$

where the ε_i are zero-centered error terms. We get the Hill estimator $\widehat{\gamma}^H_{n,\,k}$ when we ignore the term $A(n/k)$ in (6.9) and by taking the mean of the left-hand side of (6.9). We can exploit (6.9), using a least-squares approach, to propose a bias-reduced estimator for γ in which ρ is substituted by a consistent estimator $\widehat{\rho} = \widehat{\rho}(n,k)$ (see for instance Beirlant et al., 2002 and Fraga Alves et al., 2003) or by a canonical choice, such as $\rho = -1$ (see e.g., Feuerverger and Hall , 1999). Then, the least-squares estimators for γ and $A(n/k)$ are given by

$$\widehat{\gamma}^{LS}_{n,k}(\widehat{\rho}) = \frac{1}{k}\sum_{i=1}^{k} Z_i - \frac{\widehat{A}^{LS}_{n,k}(\widehat{\rho})}{1-\widehat{\rho}} = \widehat{\gamma}^H_{n,\,k} - \frac{\widehat{A}^{LS}_{n,k}(\widehat{\rho})}{1-\widehat{\rho}},$$

$$\widehat{A}^{LS}_{n,k}(\widehat{\rho}) = \frac{(1-2\widehat{\rho})\,(1-\widehat{\rho})^2}{\widehat{\rho}^2}\frac{1}{k}\sum_{i=1}^{k}\left(\left(\frac{i}{k+1}\right)^{-\widehat{\rho}} - \frac{1}{1-\widehat{\rho}}\right)Z_i.$$

We can view $\widehat{\gamma}_{n,k}^{LS}(\rho)$ as the kernel estimator

$$\widehat{\gamma}_{n,k}^{K_\rho} = \frac{1}{k}\sum_{i=1}^{k} K_\rho\left(\frac{i}{k+1}\right) Z_i,$$

where for $0 < u \leq 1$

$$K_\rho(u) = \frac{1-\rho}{\rho}\underline{K}(u) + \left(1 - \frac{1-\rho}{\rho}\right)\underline{K}_\rho(u), \tag{6.10}$$

with $\underline{K}(u) = \mathbf{1}_{(0,1)}$ and $\underline{K}_\rho(u) = \left(\frac{1-\rho}{\rho}\right)\left(u^{-\rho} - 1\right)\mathbf{1}_{(0,1)}$, both kernels satisfying condition $(\mathbb{K})$. On the contrary K_ρ does not satisfy statement (i) in $(\mathbb{K})$. We refer to Gomes and Martins (2004) and Gomes et al. (2007) for other techniques of bias reduction based on the estimation of the second-order parameter. Then, from (6.8) and using the above estimators for the different unknown quantities, we obtain the following bias-reduced estimator for $\Pi_{r,R_{opt}}$

$$\widetilde{\Pi}_{r,\widehat{R}_{opt}}^{K} = \widehat{\Pi}_{r,\widehat{R}_{opt}}^{K} - \left(\frac{k}{n}\right)^{1/r} X_{n-k,n}\widehat{A}_{n,k}^{LS}(\widehat{\rho})\,\mathcal{AB}_K\left(\widehat{\gamma}_{n,k}^{LS}(\widehat{\rho}), r, \widehat{\rho}\right).$$

The asymptotic normality of $\widetilde{\Pi}_{r,\widehat{R}_{opt}}^{K}$ is established in the following theorem.

Theorem 6.2.2 *Under the assumptions of theorem 6.2.1, if $\widehat{\rho}$ is a consistent estimator for ρ, then we have*

$$\frac{(k/n)^{-1/r}\sqrt{k}}{\mathbb{U}(n/k)}\left(\widetilde{\Pi}_{r,\widehat{R}_{opt}}^{K} - \Pi_{r,R_{opt}}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \widetilde{\mathcal{AV}}_K(\gamma, r, \rho)\right), \text{ as } n \to \infty,$$

where

$$\widetilde{\mathcal{AV}}_K(\gamma, r, \rho) = \mathcal{AV}_K(\gamma, r) + \frac{\gamma^2(1-2\rho)(1-\rho)^2}{\rho^2}\mathcal{AB}_K^2(\gamma, r, \rho)$$
$$+ \frac{2r\gamma^2(1-2\rho)(1-\rho)}{\rho^2(1-r\gamma)^2}\left(1 - (1-\rho)\int_0^1 s^{-\rho}K(s)ds\right)\mathcal{AB}_K(\gamma, r, \rho).$$

We observe that $\widetilde{\Pi}_{r,\widehat{R}_{opt}}^{K}$ has a null asymptotic bias, which was not the case for $\widehat{\Pi}_{r,\widehat{R}_{opt}}^{K}$ (see corollary 6.2.1).

Corollary 6.2.3 *Under the same assumptions as in theorem 6.2.2, and in the special case where $K = \underline{K}$, we have*

$$\frac{(k/n)^{-1/r}\sqrt{k}}{\mathbb{U}(n/k)}\left(\widetilde{\Pi}_{r,\widehat{R}_{opt}}^{\underline{K}} - \Pi_{r,R_{opt}}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \widetilde{\mathcal{AV}}_{\underline{K}}(\gamma, r, \rho)\right), \text{ as } n \to \infty,$$

where

$$\widetilde{\mathcal{AV}}_{\underline{K}}(\gamma,r,\rho)=\frac{r^2\gamma^2\left(\gamma^2+r^2\gamma^4-2r\gamma^3+1\right)}{\left(1-r\gamma\right)^4}+\frac{r^2\gamma^2\left(1-2\rho\right)\left(r\gamma+r-1\right)^2}{\left(1-r\gamma\right)^4\left(r\gamma+r\rho-1\right)^2}.$$

In the special case where $K=K_\rho$, we have the estimator $\widehat{\gamma}_{n,k}^{LS}(\rho)$ coincides with $\widehat{\gamma}_{n,k}^{K_\rho}$. The goal of the next corollary is to establish the asymptotic normality of the resulting reinsurance premium estimator $\widetilde{\Pi}_{r,\widehat{R}_{opt}}^{K_\rho}$, denoted by $\widetilde{\Pi}_{r,\widehat{R}_{opt}}^{LS}$, when the least-squares method is adopted.

Corollary 6.2.4 *Under the same assumptions as in theorem 6.2.2, and in the special case where $K=K_\rho$, we have*

$$\frac{(k/n)^{-1/r}\sqrt{k}}{\mathbb{U}(n/k)}\left(\widetilde{\Pi}_{r,\widehat{R}_{opt}}^{LS}-\Pi_{r,R_{opt}}\right)\overset{\mathcal{D}}{\longrightarrow}\mathcal{N}\left(0,\widetilde{\mathcal{AV}}_{K_\rho}(\gamma,r,\rho)\right),\ \ as\ n\to\infty,$$

where

$$\widetilde{\mathcal{AV}}_{K_\rho}(\gamma,r,\rho)=\frac{r^2\gamma^4}{\left(1-r\gamma\right)^2}+\frac{r^2\gamma^2\left(1-\rho\right)\left(1-2\rho\right)\left(r\gamma\rho+r\gamma+2r\rho-\rho-1\right)}{\rho^2\left(1-r\gamma\right)^3\left(r\gamma+r\rho-1\right)^2}.$$

6.2.3 A small simulation study

We use the statistical software **R**, see Ihaka and Gentleman (1996), to compare, in terms of bias and root of the mean squared error (RMSE), the performances of the kernel-type estimator $\widehat{\Pi}_{r,\widehat{R}_{opt}}^{K}$ and least-squares estimator $\widetilde{\Pi}_{r,\widehat{R}_{opt}}^{LS}$. We generate 1000 samples of different sizes $n=1000$, 2000 and 5000 from a Fréchet distribution with hazard function $\overline{F}(x)=1-\exp\left(-x^{-1/\gamma}\right)$, $x>0$, $\gamma=3/4$ and the second-order parameter $\rho=-1$. For the kernel function K, we choose the uniform kernel $K=\underline{K}=\mathbf{1}_{(0,1)}$. Note that $R_{opt}=\mathbb{U}(n/k^*)$, where k^* is the optimal value of k. Several methods are available for the choice of k^*, see e.g. Danielsson et al. (2001), Cheng and Peng (2001), Neves and Fraga Alves (2004) and the references therein. In our simulation study, we use the method of Neves and Fraga Alves (2004). The simulation results are summarized in Table 6.1. We conclude that $\widetilde{\Pi}_{r,\widehat{R}_{opt}}^{LS}$ has smaller bias and RMSE and consequently it performs better than $\widehat{\Pi}_{r,\widehat{R}_{opt}}^{\underline{K}}$.

n	1000		2000		5000	
r	1.10	1.20	1.10	1.20	1.10	1.20
$\Pi_{r,R_{opt}}$	3.803	8.153	3.635	8.044	3.478	7.772
$\widehat{\Pi}^{K}_{r,\widehat{R}_{opt}}$	4.071	8.364	3.867	8.208	3.521	7.807
Bias	0.268	0.211	0.232	0.164	0.043	0.035
RMSE	0.569	0.699	0.464	0.587	0.229	0.282
$\widetilde{\Pi}^{LS}_{r,\widehat{R}_{opt}}$	3.862	8.206	3.683	8.084	3.507	7.781
Bias	0.059	0.053	0.048	0.040	0.029	0.009
RMSE	0.421	0.589	0.376	0.437	0.187	0.248

Table 6.1: Comparison of $\widehat{\Pi}^{K}_{r,\widehat{R}_{opt}}$ and $\widetilde{\Pi}^{LS}_{r,\widehat{R}_{opt}}$ for 1000 samples of size $n \in \{1000, 2000, 5000\}$ of a Fréchet distribution with $\gamma = 3/4$.

6.2.4 Proofs

For each integer n, let $Y_{1,n} \leq \cdots \leq Y_{n,n}$ be the order statistics pertaining to a sample $Y_1, \ldots, Y_n$ of independent identically distributed rv's, defined on the same probability space as the $X_i's$, with df $G(y) = 1 - y^{-1}$, $y > 1$. Note that

$$\{X_{j,n}\}_{j=1}^{n} \stackrel{\mathcal{D}}{=} \{\mathbb{U}(Y_{j,n})\}_{j=1}^{n}. \tag{6.11}$$

Let $\xi_1, \xi_2, \ldots$ be a sequence of independent rv's, defined on probability space $(\Omega, \mathcal{A}, \mathbb{P})$, uniformly distributed on $(0,1)$ in such a way that $Y_i = G^{-1}(\xi_i)$, for all $1 \leq i \leq n$. Consequently, we have $Y_{i,n} = \left(1 - \xi_{i,n}\right)^{-1}$ for all $1 \leq i \leq n$ and $n \geq 1$, where $\xi_{1,n} \leq \cdots \leq \xi_{n,n}$ denote the order statistics of $\xi_1, \ldots, \xi_n$ and G^{-1} is the quantile function pertaining to G.

We will use in this section the Csörgő et al. (1986) weak approximations. On the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, there exists a sequence of Brownian bridges $\{\mathbb{B}_n(s);\ 0 \leq s \leq 1\}_{n \geq 1}$, such that for every $0 \leq v < 1/2$ and for all n

$$\sup_{1/n \leq s \leq 1-1/n} \frac{|\beta_n(s) - \mathbb{B}_n(s)|}{(s(1-s))^{1/2-v}} = O_{\mathbb{P}}\left(n^{-v}\right), \tag{6.12}$$

where the resulting uniform empirical quantile process, is denoted by

$$\beta_n(t) = \sqrt{n}(t - \mathbb{V}_n(t)),\ 0 \leq t \leq 1, \tag{6.13}$$

with $\mathbb{V}_n$ is the empirical quantile function pertaining to the sample $\xi_1, \ldots, \xi_n$ which is defined by

$$\mathbb{V}_n(s) = \xi_{j,n},\ \frac{j-1}{n} < s \leq \frac{j}{n},\ j = 1, \ldots, n \text{ and } \mathbb{V}_n(0) = \xi_{1,n}.$$

Proof of Theorem 6.2.1. From (6.11), we may rewrite $\widehat{\Pi}^K_{r,\widehat{R}_{opt}}$ as follows

$$\widehat{\Pi}^K_{r,\widehat{R}_{opt}} = \frac{r\,(k/n)^{1/r}}{1/\widehat{\gamma}^K_{n,k} - r}\mathbb{U}\left(Y_{n-k,n}\right), \text{ for } \widehat{\gamma}^K_{n,\,k} < 1/r.$$

It is easy to verify that

$$\frac{(k/n)^{-1/r}\sqrt{k}}{\mathbb{U}\left(n/k\right)}\left(\widehat{\Pi}^K_{r,\widehat{R}_{opt}} - \Pi_{r,R_{opt}}\right) = \sum_{i=1}^{4} T_{i,n},$$

where

$$T_{1,n} = \frac{r\sqrt{k}}{1/\widehat{\gamma}^K_{n,\,k} - r}\left(\frac{\mathbb{U}\left(Y_{n-k,n}\right)}{\mathbb{U}\left(n/k\right)} - \left(\frac{k}{n}Y_{n-k,n}\right)^{\gamma}\right),$$

$$T_{2,n} = \frac{r\sqrt{k}}{1/\widehat{\gamma}^K_{n,k} - r}\left(\left(\frac{k}{n}Y_{n-k,n}\right)^{\gamma} - 1\right),$$

$$T_{3,n} = \frac{r}{\left(1 - r\widehat{\gamma}^K_{n,k}\right)\left(1 - r\gamma\right)}\sqrt{k}\left(\widehat{\gamma}^K_{n,\,k} - \gamma\right),$$

and

$$T_{4,n} = \frac{(k/n)^{-1/r}\sqrt{k}}{\mathbb{U}\left(n/k\right)}\left(\frac{r\,(k/n)^{1/r}}{1/\gamma - r}\mathbb{U}\left(n/k\right) - \Pi_{r,R_{opt}}\right).$$

We start with the term $T_{1,n}$, according to de Haan (2006 and theorem 2.3.9, page 48), for any $\delta > 0$, we have

$$\frac{\mathbb{U}\left(Y_{n-k,n}\right)}{\mathbb{U}\left(n/k\right)} - \left(\frac{k}{n}Y_{n-k,n}\right)^{\gamma} = A_0\left(\frac{n}{k}\right)\left\{\left(\frac{k}{n}Y_{n-k,n}\right)^{\gamma}\frac{\left(\frac{k}{n}Y_{n-k,n}\right)^{\rho} - 1}{\rho} + o_{\mathbb{P}}\left(1\right)\left(\frac{k}{n}Y_{n-k,n}\right)^{\gamma+\rho\pm\delta}\right\},$$

where $A_0\left(t\right) \sim A\left(t\right)$ as $t \to \infty$. Since $\dfrac{k}{n}Y_{n-k,n} = 1 + o_{\mathbb{P}}\left(1\right)$ and $\sqrt{k}A(n/k) = O(1)$, as $n \to \infty$, we have

$$\sqrt{k}\left(\frac{\mathbb{U}\left(Y_{n-k,n}\right)}{\mathbb{U}\left(n/k\right)} - \left(\frac{k}{n}Y_{n-k,n}\right)^{\gamma}\right) = o_{\mathbb{P}}\left(1\right),$$

and since $\widehat{\gamma}^K_{n,\,k} \xrightarrow{\mathbb{P}} \gamma$ (see Csörgő et al., 1985), then we obtain as $n \to \infty$

$$T_{1,n} = o_{\mathbb{P}}\left(1\right). \tag{6.14}$$

For the term $T_{2,n}$, the equality $Y_{n-k,n} = \left(1 - \xi_{n-k,n}\right)^{-1}$ yields

$$\sqrt{k}\left(\left(\frac{k}{n}Y_{n-k,n}\right)^{\gamma} - 1\right) = \sqrt{k}\left(\left(\frac{n}{k}\left(1 - \xi_{n-k,n}\right)\right)^{-\gamma} - 1\right).$$

Using a Taylor expansion, we get

$$\sqrt{k}\left(\left(\frac{n}{k}\left(1 - \xi_{n-k,n}\right)\right)^{-\gamma} - 1\right) = -\gamma\left(\lambda_n(k)\right)^{-\gamma-1}\sqrt{k}\left(\frac{n}{k}\left(1 - \xi_{n-k,n}\right) - 1\right),$$

where $\lambda_n(k)$ is a sequence of rv's with values in the open interval of endpoints 1 and $\dfrac{n}{k}\left(1 - \xi_{n-k,n}\right)$. From Balkema and de Haan (1975), we have

$$\frac{n}{k}\left(1 - \xi_{n-k,n}\right) \xrightarrow{\mathbb{P}} 1, \text{ as } n \to \infty.$$

It follows that, $\lambda_n(k) \xrightarrow{\mathbb{P}} 1$ as $n \to \infty$. Therefore, as $n \to \infty$,

$$\sqrt{k}\left(\left(\frac{n}{k}\left(1 - \xi_{n-k,n}\right)\right)^{-\gamma} - 1\right) = -\gamma\sqrt{k}\left(\frac{n}{k}\left(1 - \xi_{n-k,n}\right) - 1\right)\left(1 + o_{\mathbb{P}}(1)\right).$$

On the other hand we have

$$\sqrt{k}\left(\frac{n}{k}\left(1 - \xi_{n-k,n}\right) - 1\right) = \sqrt{\frac{n}{k}}\left\{\sqrt{n}\left(\left(1 - \frac{k}{n}\right) - \mathbb{V}_n\left(1 - \frac{k}{n}\right)\right)\right\}.$$

Using the uniform empirical quantile process, defined in (6.13), we obtain

$$\sqrt{k}\left(\left(\frac{n}{k}\left(1 - \xi_{n-k,n}\right)\right)^{-\gamma} - 1\right) = -\gamma\sqrt{\frac{n}{k}}\beta_n\left(1 - \frac{k}{n}\right)\left(1 + o_{\mathbb{P}}(1)\right), \text{ as } n \to \infty.$$

Using the asymptotic approximation (6.12), we get for all large n

$$\begin{aligned}\sqrt{k}\left(\left(\frac{n}{k}\left(1 - \xi_{n-k,n}\right)\right)^{-\gamma} - 1\right) &= -\gamma\sqrt{\frac{n}{k}}\left\{\mathbb{B}_n\left(1 - \frac{k}{n}\right)\right. \\ &\quad \left. + O_{\mathbb{P}}\left(n^{-v}\right)\left(\frac{k}{n}\right)^{1/2-v}\right\}\left(1 + o_{\mathbb{P}}(1)\right), \\ &= -\gamma\sqrt{\frac{n}{k}}\mathbb{B}_n\left(1 - \frac{k}{n}\right)\left(1 + o_{\mathbb{P}}(1)\right).\end{aligned}$$

Consequently, since $\widehat{\gamma}_{n,\,k}^{K} \xrightarrow{\mathbb{P}} \gamma$, we obtain for all large n

$$\begin{aligned}T_{2,n} &= -\frac{r\gamma^2}{1 - r\gamma}\sqrt{\frac{n}{k}}\mathbb{B}_n\left(1 - \frac{k}{n}\right)\left(1 + o_{\mathbb{P}}(1)\right), \\ &= \mathcal{W}_{1,n} + o_{\mathbb{P}}(1). \qquad (6.15)\end{aligned}$$

For the term $T_{3,n}$, from theorem 1 of Deme et al. (2013), we have for all large n

$$\sqrt{k}\left(\widehat{\gamma}_{n,\,k}^{K}-\gamma\right)=\sqrt{k}A\left(n/k\right)\int_{0}^{1}s^{-\rho}K\left(s\right)ds$$
$$+\gamma\sqrt{\frac{n}{k}}\int_{0}^{1}s^{-1}\mathbb{B}_{n}\left(1-\frac{sk}{n}\right)d\left(sK\left(s\right)\right)+o_{\mathbb{P}}\left(1\right).$$

Then, since $\widehat{\gamma}_{n,\,k}^{K}\overset{\mathbb{P}}{\to}\gamma$, we get as $n\to\infty$

$$T_{3,n}=\frac{r}{\left(1-r\gamma\right)^{2}}\left\{\sqrt{k}A\left(\frac{n}{k}\right)\int_{0}^{1}s^{-\rho}K\left(s\right)ds\right.$$
$$\left.+\gamma\sqrt{\frac{n}{k}}\int_{0}^{1}s^{-1}\mathbb{B}_{n}\left(1-\frac{sk}{n}\right)d\left(sK\left(s\right)\right)\right\}+o_{\mathbb{P}}\left(1\right)$$
$$=\frac{r}{\left(1-r\gamma\right)^{2}}\sqrt{k}A\left(\frac{n}{k}\right)\int_{0}^{1}s^{-\rho}K\left(s\right)ds+\mathcal{W}_{2,n}\left(K\right)+o_{\mathbb{P}}\left(1\right). \quad (6.16)$$

For the term $T_{4,n}$, we have

$$T_{4,n}=\left(k/n\right)^{-1/r}\sqrt{k}\left(\frac{r\left(k/n\right)^{1/r}}{1/\gamma-r}-\frac{\Pi_{r,R_{opt}}}{\mathbb{U}\left(n/k\right)}\right),$$

where

$$\Pi_{r,R_{opt}}=\int_{\mathbb{U}(n/k)}^{\infty}\left(S(x)\right)^{1/r}dx.$$

Since $x^{-1/r}\mathbb{U}(x)\to0$ as $x\to\infty$, then an integration by parts with a change of variables yields

$$\Pi_{r,R_{opt}}=\left(k/n\right)^{1/r}\left\{\frac{1}{r}\int_{1}^{\infty}x^{-1-1/r}\mathbb{U}\left(nx/k\right)dx-\mathbb{U}\left(n/k\right)\right\}.$$

Therefore

$$T_{4,n}=\sqrt{k}\left\{\frac{1}{1-r\gamma}-\frac{1}{r}\int_{1}^{\infty}x^{-1-1/r}\frac{\mathbb{U}\left(nx/k\right)}{\mathbb{U}\left(n/k\right)}dx\right\}$$
$$=-\frac{1}{r}\sqrt{k}\int_{1}^{\infty}x^{-1-1/r}\left(\frac{\mathbb{U}\left(nx/k\right)}{\mathbb{U}\left(n/k\right)}-x^{\gamma}\right)dx.$$

From theorem 2.3.9 of de Haan (2006), for $\gamma\in(1/2,1)$ and $r\in[1,1/\gamma)$, we obtain as $n\to\infty$

$$T_{4,n}=-\frac{1}{r}\sqrt{k}A\left(\frac{n}{k}\right)\int_{1}^{\infty}x^{\gamma-1-1/r}\frac{x^{\rho}-1}{\rho}dx\left(1+o\left(1\right)\right)$$
$$=\sqrt{k}A\left(\frac{n}{k}\right)\frac{r}{\left(1-r\gamma\right)\left(r\gamma+r\rho-1\right)}\left(1+o\left(1\right)\right). \quad (6.17)$$

Combining (6.14)-(6.17), we get as $n \to \infty$

$$\frac{(k/n)^{-1/r}\sqrt{k}}{\mathbb{U}(n/k)}\left(\widehat{\Pi}^{K}_{r,\widehat{R}_{opt}} - \Pi_{r,R_{opt}}\right) = \sqrt{k}A(n/k)\mathcal{A}\mathcal{B}_K(\gamma,r,\rho)+\mathcal{W}_{1,n}+\mathcal{W}_{2,n}(K)+o_{\mathbb{P}}(1).$$

This finishes the proof of theorem 6.2.1. ■

Proof of Corollary 6.2.1. Since $\{\mathbb{B}_n(s);\ 0 \le s \le 1\}_{n\ge1}$, is a sequence of Brownian bridges, then

$$\frac{(k/n)^{-1/r}\sqrt{k}}{\mathbb{U}(n/k)}\left(\widehat{\Pi}^{K}_{r,\widehat{R}_{opt}} - \Pi_{r,R_{opt}}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\mathcal{AV}_K(\gamma,r)\right),\ \text{as } n\to\infty,$$

with

$$\begin{aligned}\mathcal{AV}_K(\gamma,r) &= \lim_{n\to\infty} E\left(\left(\mathcal{W}_{1,n}+\mathcal{W}_{2,n}(K)\right)^2\right)\\ &= \lim_{n\to\infty}\left(E\left(\mathcal{W}_{1,n}^2\right)+E\left(\mathcal{W}_{2,n}^2(K)\right)+2E\left(\mathcal{W}_{1,n}\mathcal{W}_{2,n}(K)\right)\right).\end{aligned}$$

Elementary computation gives, as $n \to \infty$

$$\begin{aligned}E\left(\mathcal{W}_{1,n}^2\right) &= \frac{r^2\gamma^4}{(1-r\gamma)^2}+o(1),\\ E\left(\mathcal{W}_{2,n}^2(K)\right) &= \frac{r^2\gamma^2}{(1-r\gamma)^4}\int_0^1 K^2(s)ds+o(1),\end{aligned}$$

and

$$E\left(\mathcal{W}_{1,n}\mathcal{W}_{2,n}(K)\right)=o(1).$$

Then, we get

$$\mathcal{AV}_K(\gamma,r)=\frac{r^2\gamma^4}{(1-r\gamma)^2}+\frac{r^2\gamma^2}{(1-r\gamma)^4}\int_0^1 K^2(s)ds.$$

We complete the proof of corollary 6.2.1. ■

Proof of Corollary 6.2.2. The proof is a direct result of corollary 6.2.1 with the kernel $K=\underline{K}=1_{(0,1)}$. ■

Proof of Theorem 6.2.2. According to theorem 6.2.1 and (6.8), we have

$$\frac{(k/n)^{-1/r}\sqrt{k}}{\mathbb{U}(n/k)}\left(\widetilde{\Pi}^{K}_{r,\widehat{R}_{opt}} - \Pi_{r,R_{opt}}\right) = \mathcal{W}_{1,n}+\mathcal{W}_{2,n}(K)+\mathcal{W}_{3,n}(K)+o_{\mathbb{P}}(1),$$

where

$$\mathcal{W}_{3,n}(K) = \sqrt{k}\left(A(n/k)\mathcal{AB}_K(\gamma,r,\rho) - \widehat{A}_{n,k}^{LS}(\widehat{\rho})\,\mathcal{AB}_K\left(\widehat{\gamma}_{n,k}^{LS}(\widehat{\rho}),r,\widehat{\rho}\right)\frac{X_{n-k,n}}{\mathbb{U}(n/k)}\right)$$
$$= -\mathcal{AB}_K(\gamma,r,\rho)\sqrt{k}\left(\widehat{A}_{n,k}^{LS}(\widehat{\rho}) - A(n/k)\right)$$
$$- \sqrt{k}\widehat{A}_{n,k}^{LS}(\widehat{\rho})\left(\mathcal{AB}_K\left(\widehat{\gamma}_{n,k}^{LS}(\widehat{\rho}),r,\widehat{\rho}\right) - \mathcal{AB}_K(\gamma,r,\rho)\right)$$
$$- \sqrt{k}\widehat{A}_{n,k}^{LS}(\widehat{\rho})\,\mathcal{AB}_K\left(\widehat{\gamma}_{n,k}^{LS}(\widehat{\rho}),r,\widehat{\rho}\right)\left(\frac{X_{n-k,n}}{\mathbb{U}(n/k)} - 1\right).$$

From Lemma 5 of Deme et al. (2013), we have

$$\sqrt{k}\left(\widehat{A}_{n,k}^{LS}(\widehat{\rho}) - A(n/k)\right) = \gamma(1-\rho)$$
$$\times\sqrt{\frac{n}{k}}\int_0^1 s^{-1}\mathbb{B}_n\left(1-\frac{sk}{n}\right)d\left(s\left(\underline{K}(s) - K_\rho(s)\right)\right) + o_{\mathbb{P}}(1).$$

Since $\widehat{\rho}$ is a consistent estimator for ρ, then we get as $n\to\infty$

$$\sqrt{k}\widehat{A}_{n,k}^{LS}(\widehat{\rho})\left(\mathcal{AB}_K\left(\widehat{\gamma}_{n,k}^{LS}(\widehat{\rho}),r,\widehat{\rho}\right) - \mathcal{AB}_K(\gamma,r,\rho)\right) = o_{\mathbb{P}}(1).$$

Making use of Potter's inequalities (see 5th assertion of proposition B.1.9 in de Haan, 2006), we obtain as $n\to\infty$

$$\sqrt{k}\widehat{A}_{n,k}^{LS}(\widehat{\rho})\,\mathcal{AB}_K\left(\widehat{\gamma}_{n,k}^{LS}(\widehat{\rho}),r,\widehat{\rho}\right)\left(\frac{X_{n-k,n}}{\mathbb{U}(n/k)} - 1\right) = o_{\mathbb{P}}(1).$$

Therefore

$$\mathcal{W}_{3,n}(K) = -\gamma(1-\rho)\,\mathcal{AB}_K(\gamma,\beta,\rho)$$
$$\times\sqrt{\frac{n}{k}}\int_0^1 s^{-1}\mathbb{B}_n\left(1-\frac{sk}{n}\right)d\left(s\left(\underline{K}(s) - K_\rho(s)\right)\right) + o_{\mathbb{P}}(1),\ \text{as } n\to\infty.$$

It is clear that $\mathcal{W}_{1,n} + \mathcal{W}_{2,n}(K) + \mathcal{W}_{3,n}(K)$ is a Gaussian rv with mean zero and variance

$$\widetilde{\mathcal{AV}}_K(\gamma,r,\rho) = \lim_{n\to\infty} E\left(\left(\mathcal{W}_{1,n} + \mathcal{W}_{2,n}(K) + \mathcal{W}_{3,n}(K)\right)^2\right).$$

After elementary computations, we get

$$\widetilde{\mathcal{AV}}_K(\gamma,r,\rho) = \mathcal{AV}_K(\gamma,r) + \frac{\gamma^2(1-2\rho)(1-\rho)^2}{\rho^2}\mathcal{AB}_K^2(\gamma,r,\rho)$$
$$+\frac{2r\gamma^2(1-2\rho)(1-\rho)}{\rho^2(1-r\gamma)^2}\left(1-(1-\rho)\int_0^1 s^{-\rho}K(s)ds\right)\mathcal{AB}_K(\gamma,r,\rho).$$

This achieves the proof of theorem 6.2.2. ■

Proof of Corollary 6.2.3. The proof is a direct result of theorem 6.2.1 with the kernel $K = \underline{K} = \mathbf{1}_{(0,1)}$. ■

Proof of Corollary 6.2.4. The proof is a direct result of theorem 6.2.1 with the kernel $K = K_\rho$ defined in (6.10). ■

6.3 Kernel-type estimators for the distortion risk premiums [12]

Abstract.[2]A new kernel-type estimator for the distortion risk premiums of heavy-tailed losses is introduced. Using a least-squares approach, a bias-reduced version of this estimator is proposed. Under suitable assumptions, the asymptotic normality of the given estimators is established. A small simulation study, to illustrate the performance of our method, is carried out.

6.3.1 Introduction

To determine an adequate price or premium for an insured risk, one should use an appropriate pricing principle. For a presentation of the existing variants of premium principles, we refer to Goovaerts et al. (1984), Rolski et al. (1999), Denuit et al. (2005), and the references therein. Some premium principles are special cases of the distortion risk premium (see Wang, 1996). For an insured risk X, a non-negative rv defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with continuous df F, the distortion risk premium is defined by

$$\Pi(g) = \int_0^\infty g\left(\overline{F}(x)\right) dx,$$

where $g : [0, 1] \to [0, 1]$ is a non-decreasing function, called distortion function, satisfying $g(0) = 0$ and $g(1) = 1$.

[2]This section is a paper appeared in: *Scandinavian Actuarial Journal*, DOI: 10.1080/03461238.2014.924434.

By changing of variables and integrating by parts, we get the following expression for $\Pi(g)$

$$\Pi(g) = \int_0^1 g'(s)\, Q(1-s)\, ds,$$

where, g' denotes the Lebesgue derivative of g and $Q(s) = \inf\{x : F(x) \geq s\}$, $0 < s < 1$ is the quantile function of F. Note in passing that, for $t \downarrow 0$, the quantile $Q(1-t)$ is called high or extreme quantile.

For heavy-tailed distributions, some authors have introduced and studied different estimators for $\Pi(g)$ by making use of the EVT, see e.g.,Beirlant et al. (2001), Necir et Boukhetala (2004), Necir et al. (2007), Necir & Meraghni (2009), Brahimi et al. (2012), Deme et al. (2013) and the references therein.

A df F is called heavy-tailed if its tail $1 - F$ is a regularly varying at infinity with index $(-1/\gamma) < 0$, i.e.

$$\overline{F}(x) = x^{-1/\gamma} \ell_F(x), \tag{6.18}$$

where ℓ_F is a slowly varying function at infinity, that is, $\ell_F(tx)/\ell_F(t) \to 1$ as $t \to \infty$ for any $x > 0$. The class of regularly varying functions includes popular distributions such those Pareto's, Burr's, Student's, Fréchet's, α-stable $(0 < \alpha < 2)$, and log-gamma, which are known to be appropriate models of fitting large insurance claims, large fluctuations of prices, log-returns and other data (see Beirlant et al., 2001). For more details on these distributions, we refer to Bingham et al. (1987) and Rolski et al. (1999).

We restrict ourselves to this class of distributions. Then it is quite natural to suppose that the distortion functions g is such that $t \mapsto g(t)$ is regularly varying at zero with index $1/\beta \in (0, 1]$, that is

$$g(t) = t^{1/\beta} \ell_g(t), \tag{6.19}$$

where ℓ_g is a slowly varying function at zero satisfying $\ell_g(\lambda t)/\ell_g(t) \to 1$ as $t \to 0$ for $\lambda > 0$. There exist several examples of distortion functions satisfying (6.19), we mention

- Net premium,

$$g(t) = t \text{ with } \beta = 1 \text{ and } \ell_g(t) = 1.$$

- Tail value-at-risk, $0 < p < 1$,

$$g(t) = \min\left(\frac{t}{p}, 1\right) \text{ with } \beta = 1 \text{ and } \ell_g(t) = 1 \text{ if } t \leq p.$$

- Proportional hazard transform, $\theta \geq 1$,

$$g(t) = t^{1/\theta} \text{ with } \beta = \theta \text{ and } \ell_g(t) = 1.$$

- Gini principle, $0 < \omega \leq 1$,

$$g(t) = (1+\omega)t - \omega t^2 \text{ with } \beta = 1 \text{ and } \ell_g(t) = (1+\omega) - \omega t.$$

- Dual-power function principle, $\alpha \geq 1$,

$$g(t) = 1 - (1-t)^{\alpha} = t\left(\alpha - \frac{\alpha(\alpha-1)}{2}t + o(t)\right)$$
$$\text{with } \beta = 1 \text{ and } \ell_g(t) = \alpha - \frac{\alpha(\alpha-1)}{2}t + o(t) \text{ as } t \downarrow 0. \tag{6.20}$$

- Beta-distortion risk premium, $a \leq 1 \leq b$,

$$g(t) = t^a\left\{\frac{1}{a\beta(a,b)}\right\} \text{ with } \beta = \frac{1}{a} \text{ and } \ell_g(t) = \frac{1}{a\beta(a,b)},$$

where $\beta(a,b) = \int_0^1 v^{a-1}(1-v)^{b-1}\,dv$.

- MINMAXVAR2 risk premium, $\eta > 0$, $\tau > 0$,

$$g(t) = 1 - \left(1 - t^{\frac{1}{1+\eta}}\right)^{1+\tau}$$
$$\text{with } \beta = \frac{1+\eta}{1+\tau} \text{ and } \ell_g(t) = t^{-\frac{1+\tau}{1+\eta}} - \left(t^{-\frac{1}{1+\eta}} - 1\right)^{1+\tau}.$$

For each $n \geq 1$, let $X_1, \ldots, X_n$ be a sample of X with df F satisfying (6.18) and denote by $X_{1,n} \leq \cdots \leq X_{n,n}$ the corresponding order statistics. Let $k = k_n$ be a sequence of positive integers, such that

$$1 < k < n, \; k \to \infty, \text{ and } k/n \to 0 \text{ as } n \to \infty. \tag{6.21}$$

Based on Weissman's estimator of high quantiles (Weissman, 1978), Necir and Meraghni (2009) proposed the following estimator for $\Pi(g)$

$$\widehat{\Pi}(g) = g(k/n)\frac{X_{n-k,n}}{1-\beta\widehat{\gamma}^H} + \sum_{i=k+1}^{n} a_{i,n}(g)\,X_{n-i+1,n}, \text{ for } \widehat{\gamma}^H < 1/\beta,$$

where

$$a_{i,n}(g) = g\left(\frac{i}{n}\right) - g\left(\frac{i-1}{n}\right), \; i = 1, \ldots, n,$$

and $\widehat{\gamma}^H$ is the classical Hill estimator (Hill, 1975) of the tail index γ, defined by

$$\widehat{\gamma}^H = \frac{1}{k}\sum_{i=1}^{k} i\left(\log X_{n-i+1,n} - \log X_{n-i,n}\right).$$

Based on $\widehat{\gamma}^H$, Deme and Lo (2013) proposed a reduced-bias estimator for the distortion risk premium $\Pi(g)$ of heavy-tailed distributions. A major drawback of $\widehat{\gamma}^H$ is the discrete character of its behavior in the sense that increasing k by 1, can change the actual value of the estimate considerably. Plotting $\widehat{\gamma}^H$ as a function of k used therefore often results in a zig-zag figure. Using a kernel function K, Csörgő et al. (1985) proposed a smoother version of Hill's estimator defined by

$$\widehat{\gamma}^K = \frac{1}{k}\sum_{i=1}^{k} K\left(\frac{i}{k+1}\right) Z_i,$$

where $Z_i = i\left(\log X_{n-i+1,n} - \log X_{n-i,n}\right)$. The class of kernel estimators $\widehat{\gamma}^K$ generalizes and includes the Hill estimator. Notice that, using the uniform kernel $K = \underline{K} = \mathbf{1}_{(0,1)}$ yields Hill's estimator $\widehat{\gamma}^H$, with $\mathbf{1}_{(\cdot)}$ being the indicator function.

We propose a kernel-type estimator for $\Pi(g)$ of heavy-tailed distribution, that is

$$\widehat{\Pi}^K(g) = g(k/n)\frac{X_{n-k,n}}{1-\beta\widehat{\gamma}^K} + \sum_{i=k+1}^{n} a_{i,n}(g) X_{n-i+1,n}, \text{ for } \widehat{\gamma}^K < 1/\beta. \tag{6.22}$$

In subsection 6.3.2, we study the asymptotic properties of $\widehat{\Pi}^K(g)$ and propose a new reduced-biased estimator for $\Pi(g)$ whose asymptotic results is also given. The performance of our approach is shown on a small simulation study in section subsection 6.3.3. The proofs are postponed until section subsection 6.3.4.

6.3.2 Main results

Firstly, in this subsection, we study the asymptotic properties of $\widehat{\Pi}^K(g)$.

6.3.2.1 Asymptotic distribution of $\widehat{\Pi}^K(g)$

From (6.22), it is clear that the asymptotic normality of $\widehat{\Pi}^K(g)$ is related to $\widehat{\gamma}^K$. To establish such a type of result, the regular condition (6.18) itself is not sufficient. For this reason, we strengthen the condition (6.18) into the following one: the df F is said to satisfy the second-order regular variation with second-order parameter $\rho \leq 0$ if there exists a function $t \mapsto A(t)$ that converges to zero when t tends to infinity, does not change its sign for all sufficiently large t, and such that

$$\lim_{t\to\infty} \frac{1}{A(t)} \left(\frac{1-F(tx)}{1-F(t)} - x^{-1/\gamma} \right) = x^{-1/\gamma} \frac{x^{\rho/\gamma}-1}{\rho/\gamma}, \text{ for any } x > 0. \tag{6.23}$$

If $\rho = 0$ interpret $\dfrac{x^{\rho/\gamma}-1}{\rho/\gamma}$ as $\log x$. In terms of the quantile function Q, this is equivalent to (see theorem 2.3.9 in de Haan and Ferreira, 2006, page 48)

$$\lim_{s\downarrow 0} \frac{1}{A(1/s)} \left(\frac{Q(1-sx)}{Q(1-s)} - x^{-\gamma} \right) = x^{-\gamma} \frac{x^{-\rho}-1}{\rho}, \text{ for any } x > 0.$$

Moerever, we need the following classical conditions about the kernel K.

Condition $(\mathcal{K})$. Let K be a function defined on $(0,1]$.

(i) $K(s) \geq 0$, whenever, $0 < s \leq 1$ and $K(1) = K'(1) = 0$.
(ii) $K(\cdot)$ is differentiable, non-increasing and right continuous on $(0,1]$.
(iii) K and K' are bounded.
(iv) $\int_0^1 K(u)\,du = 1$.
(v) $\int_0^1 u^{-1/2} K(u)\,du < \infty$.

Theorem 6.3.1 *Let F be a df satisfying (6.23) with $\gamma \in (1/2, 1)$ and suppose that the corresponding quantile function $Q(\cdot)$ is continuously differentiable on $[0,1)$. Let $k = k_n$ be an integer sequence satisfying (6.21) with $\sqrt{k}A(n/k) = O(1)$ as $n \to \infty$. If the condition $(\mathcal{K})$ holds and for any differentiable distortion function g satisfying (6.19) with $1 \leq \beta < 1/\gamma$, then there exists a sequence of Brownian bridges $\{\mathbb{B}_n(s);\ 0 \leq s \leq 1\}$ such that for all large n*

$$\frac{\sqrt{k}\left(\widehat{\Pi}^K(g) - \Pi(g)\right)}{g(k/n)\,Q(1-k/n)} = \sqrt{k}A(n/k)\,\mathcal{AB}_K(\gamma,\beta,\rho) + \mathcal{W}_{1n} + \mathcal{W}_{2n} + \mathcal{W}_{3n}(K) + o_{\mathbb{P}}(1),$$

where

$$\mathcal{AB}_K(\gamma,\beta,\rho) = \frac{\beta}{(1-\gamma\beta)(\gamma\beta+\beta\rho-1)} + \frac{\beta}{(1-\gamma\beta)^2} \int_0^1 s^{-\rho} K(s)\,ds,$$

and

$$\begin{cases} \mathcal{W}_{1n} = -\dfrac{\displaystyle\int_{k/n}^{1} g'(s)\,\mathbb{B}_n(1-s)\,Q'(1-s)\,ds}{(n/k)^{1/2}\,g(k/n)\,Q(1-k/n)}, \\ \mathcal{W}_{2n} = -\dfrac{\gamma}{1-\beta\gamma}\sqrt{\dfrac{n}{k}}\mathbb{B}_n\left(1-\dfrac{k}{n}\right), \\ \mathcal{W}_{3n}(K) = \dfrac{\beta\gamma}{(1-\beta\gamma)^2}\sqrt{\dfrac{n}{k}}\displaystyle\int_0^1 s^{-1}\mathbb{B}_n\left(1-\dfrac{sk}{n}\right)d\left(sK(s)\right). \end{cases}$$

Corollary 6.3.1 *Under the assumptions of theorem 6.3.1, if* $\sqrt{k}A(n/k) \to \lambda \in \mathbb{R}$*, then we have*

$$\frac{\sqrt{k}}{g(k/n)\,Q(1-k/n)}\left(\widehat{\Pi}^{K}(g) - \Pi(g)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\lambda\mathcal{AB}_K(\gamma,\beta,\rho), \mathcal{AV}_K(\gamma,\beta)\right), \text{ as } n \to \infty,$$

where

$$\mathcal{AV}_K(\gamma,\beta) = \frac{\gamma^2\beta}{(2\gamma\beta+\beta-2)(1-\gamma\beta)^2} + \frac{\gamma^2\beta^2}{(1-\gamma\beta)^4}\int_0^1 K^2(s)\,ds.$$

Corollary 6.3.1 generalizes theorem 2 in Necir and Meraghni (2009) when $\lambda \neq 0$, when we use a general kernel K instead of $\underline{K}$ and when we use a general regularly varying distortion function g. For any kernel K, we can compute the asymptotic bias and variance. If $K = \underline{K}$, we have the following corollary.

Corollary 6.3.2 *Under the assumptions of corollary 6.3.1, and in the special case where* $K = \underline{K}$*, we have as* $n \to \infty$,

$$\frac{\sqrt{k}}{g(k/n)\,Q(1-k/n)}\left(\widehat{\Pi}^{\underline{K}}(g) - \Pi(g)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\lambda\frac{\beta\rho(\gamma\beta+\beta-1)}{(1-\rho)(\beta\gamma+\rho\beta-1)(1-\beta\gamma)^2}, \frac{\gamma^2\beta(\gamma\beta+\beta-1)^2}{(2\gamma\beta+\beta-2)(1-\gamma\beta)^4}\right).$$

Note that this corollary 6.3.2 corrects a mistake in the asymptotic variance of theorem 2 in Necir and Meraghni (2009).

Next we propose a bias-reduced estimator for $\Pi(g)$.

6.3.2.2 Bias-reduced estimator for $\Pi(g)$

From theorem 6.3.1, we have

$$\widehat{\Pi}^K(g) - g(k/n)\,Q(1-k/n)\,A(n/k)\,\mathcal{A}\mathcal{B}_K(\gamma,\beta,\rho) \tag{6.24}$$

is an asymptotically unbiased estimator for $\Pi(g)$ with γ, ρ, $Q(1-k/n)$ and $A(n/k)$ are unknown quantities that we have to estimate. The estimator $\widehat{\Pi}^K(g)$ exhibit a bias because it is based on Weissman's estimator of high quantiles for heavy-tailed distributions, known to be largely biased. As a better alternative to Weissman's estimators, several estimators of extreme quantiles with reduced biases are proposed in the literature, see e.g., Gomes and Martins (2004), Caeiro et al. (2009), Gomes and Figueiredo (2006) and the references therein. Here, we derive a new estimator for $\Pi(g)$ with reduced bias by applying the results of Feuerverger and Hall (1999) and Beirlant et al. (1999, 2002) who proposed, under (6.23), the following exponential regression model for the log-spacings of order statistics

$$Z_i \sim \left(\gamma + A\left(\frac{n}{k}\right)\left(\frac{i}{k+1}\right)^{-\rho}\right) + \varepsilon_i, \quad 1 \le i \le k, \tag{6.25}$$

where the ε_i are zero-centered error terms. We get the Hill estimator $\widehat{\gamma}^H$ when we ignore the term $A(n/k)$ in (6.25) and by taking the mean of the left-hand side of (6.25). We can exploit (6.25), using a least-squares approach, to propose a reduced-bias estimator for γ in which ρ is substituted by a consistent estimator $\widehat{\rho} = \widehat{\rho}(n,k)$ (see for instance Beirlant et al., 2002 and Fraga Alves et al., 2003) or by a canonical choice, such as $\rho = -1$ (see e.g., Feuerverger and Hall, 1999 or Beirlant et al., 1999). The least-squares estimators for γ and $A(n/k)$ are then given by

$$\widehat{\gamma}^{LS}(\widehat{\rho}) = \frac{1}{k}\sum_{i=1}^{k} Z_i - \frac{\widehat{A}^{LS}(\widehat{\rho})}{1-\widehat{\rho}} = \widehat{\gamma}^H - \frac{\widehat{A}^{LS}(\widehat{\rho})}{1-\widehat{\rho}},$$

$$\widehat{A}^{LS}(\widehat{\rho}) = \frac{(1-2\widehat{\rho})(1-\widehat{\rho})^2}{\widehat{\rho}^2}\frac{1}{k}\sum_{i=1}^{k}\left(\left(\frac{i}{k+1}\right)^{-\widehat{\rho}} - \frac{1}{1-\widehat{\rho}}\right) Z_i.$$

Note that $\widehat{\gamma}^{LS}(\widehat{\rho})$ can be viewed as the kernel estimator

$$\widehat{\gamma}^{LS}(\rho) = \widehat{\gamma}^{K_\rho} = \frac{1}{k}\sum_{i=1}^{k} K_\rho\left(\frac{i}{k+1}\right) Z_i,$$

where for $0 < u \leq 1$,

$$K_\rho(u) = \frac{1-\rho}{\rho}\underline{K}(u) + \left(1 - \frac{1-\rho}{\rho}\right)\underline{K}_\rho(u),$$

with $\underline{K}(u) = \mathbf{1}_{(0<u<1)}$ and $\underline{K}_\rho(u) = \left(\frac{1-\rho}{\rho}\right)(u^{-\rho} - 1)\,\mathbf{1}_{(0<u<1)}$, both kernels satisfying condition $(\mathcal{K})$. On the contrary K_ρ does not satisfy statement (i) in condition $(\mathcal{K})$. Then, from (6.24) and using the above estimators for the different unknown quantities, we obtain the following bias-reduced estimator for $\Pi(g)$

$$\widetilde{\Pi}^K(g) = \widehat{\Pi}^K(g) - g(k/n)\,X_{n-k,n}\widehat{A}^{LS}(\widehat{\rho})\,\mathcal{AB}_K\left(\widehat{\gamma}^{LS}(\widehat{\rho}), \beta, \widehat{\rho}\right).$$

The asymptotic normality of $\widetilde{\Pi}^K(g)$ is established in the following theorem.

Theorem 6.3.2 *Under the assumptions of theorem 6.3.1, if $\widehat{\rho}$ is a consistent estimator for ρ, then we have*

$$\frac{\sqrt{k}}{g(k/n)\,Q(1-k/n)}\left(\widetilde{\Pi}^K(g) - \Pi(g)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \widetilde{\mathcal{AV}}_K(\gamma,\beta,\rho)\right), \text{ as } n \to \infty,$$

where

$$\widetilde{\mathcal{AV}}_K(\gamma,\beta,\rho) = \mathcal{AV}_K(\gamma,\beta) + \frac{\gamma^2(1-2\rho)(1-\rho)^2}{\rho^2}\mathcal{AB}_K^2(\gamma,\beta,\rho) + \frac{2\gamma^2\beta(1-2\rho)(1-\rho)^2}{\rho^2(1-\gamma\beta)^2}\left(1 - (1-\rho)\int_0^1 s^{-\rho}K(s)ds\right)\mathcal{AB}_K(\gamma,\beta,\rho).$$

We observe that $\widetilde{\Pi}^K(g)$ has a null asymptotic bias, which was not the case for $\widehat{\Pi}^K(g)$ (see corollary 6.3.3).

Corollary 6.3.3 *Under the same assumptions as in theorem 6.3.1 and in the special case where $K = \underline{K}$, we have*

$$\frac{\sqrt{k}}{g(k/n)\,Q(1-k/n)}\left(\widetilde{\Pi}^{\underline{K}}(g) - \Pi(g)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \widetilde{\mathcal{AV}}_{\underline{K}}(\gamma,\beta,\rho)\right), \text{ as } n \to \infty,$$

where

$$\widetilde{\mathcal{AV}}_{\underline{K}}(\gamma,\beta,\rho) = \frac{\gamma^2\beta(\gamma\beta+\beta-\beta\rho-1)^2(\gamma\beta+\beta-1)^2}{(1-\gamma\beta)^4(\gamma\beta+\beta\rho-1)^2(2\gamma\beta+\beta-2)}.$$

In the special case where $K = K_\rho$, we have the estimator $\widehat{\gamma}^{K_\rho}$ coincides with $\widehat{\gamma}^{LS}(\rho)$. The aim of the next corollary is to establish the asymptotic normality of the resulting estimator $\widetilde{\Pi}^{K_\rho}(g)$, denoted by $\widetilde{\Pi}^{LS}(g)$, when the least-squares approach is adopted.

Corollary 6.3.4 *Under the same assumptions as in theorem 6.3.2, and in the special case where $K = K_\rho$, we have*

$$\frac{\sqrt{k}}{g(k/n)\,Q(1-k/n)}\left(\widetilde{\Pi}^{LS}(g) - \Pi(g)\right) \overset{\mathcal{D}}{\to} \mathcal{N}\left(0, \widetilde{\mathcal{AV}}_{K_\rho}(\gamma,\beta,\rho)\right), \quad \text{as } n \to \infty,$$

where

$$\widetilde{\mathcal{AV}}_{K_\rho}(\gamma,\beta,\rho) = \frac{\gamma^2\beta}{(2\gamma\beta+\beta-2)(1-\gamma\beta)^2} + \frac{\gamma^2\beta^2(1-\rho)^2}{\rho^2(1-\gamma\beta)^4} + \frac{\gamma^2\beta^2(1-2\rho)(1-\rho)(\gamma\beta\rho+2\beta\rho+\gamma\beta-\rho-1)}{\rho^2(\gamma\beta+\beta\rho-1)^2(1-\gamma\beta)^3}.$$

6.3.3 Simulation study

By means of the statistical software **R**, see Ihaka and Gentleman (1996), we carry out a small simulation study to compare, in terms of bias and root of the mean squared error (RMSE), the performances of the kernel-type estimator $\widehat{\Pi}^{\underline{K}}(g)$ and least-squares estimator $\widetilde{\Pi}^{LS}(g)$. To this aim, 1000 samples of different sizes $n = 1000, 2000$ and 5000 are simulated from a Fréchet distribution defined as: $F(x) = \exp\left(-x^{-1/\gamma}\right)$, $x > 0$ with $\gamma = 2/3$ and $\gamma = 3/4$ while the second-order parameter $\rho = -1$. Concerning the premium calculation principles, we choose the dual-power premium principle (6.20) where, as in Wang (1996), we have set the loading parameter α at 1.366. The extreme value theory based estimators rely on the number k of upper order statistics involved in estimate computations. Several procedures are available for choosing the optimal values of k, see e.g., Danielsson et al. (2001), Cheng and Peng (2001), Neves and Fraga Alves (2004), and the references therein. In our simulation study, we apply the Reiss and Thomas (2007) method whose performance is discussed by Neves and

Fraga Alves (2004). The results are presented in Table 6.2. We conclude that $\widetilde{\Pi}^{LS}(g)$ has smaller bias and RMSE and consequently it performs better than $\widehat{\Pi}^{\underline{K}}(g)$.

n	1000		2000		5000	
γ	2/3	3/4	2/3	3/4	2/3	3/4
$\Pi(g)$	3.2981	4.5811	3.2981	4.5811	3.2981	4.5811
$\widehat{\Pi}^{\underline{K}}(g)$	3.7554	5.0127	3.5126	4.7867	3.3778	4.6317
Bias	0.4573	0.4316	0.2145	0.2056	0.0797	0.0506
RMSE	0.5744	0.4918	0.5072	0.4368	0.1827	0.1208
$\widetilde{\Pi}^{LS}(g)$	3.5416	4.8132	3.4207	4.6693	3.3380	4.5902
Bias	0.2435	0.2321	0.1226	0.0882	0.0399	0.0091
RMSE	0.3394	0.2662	0.2711	0.2103	0.1081	0.0906

Table 6.2: Comparison of $\widehat{\Pi}^{\underline{K}}(g)$ and $\widetilde{\Pi}^{LS}(g)$ for 1000 samples of size $n \in \{1000, 2000, 5000\}$ of a Fréchet distribution with $\gamma = 2/3$ and $\gamma = 3/4$, where g is the dual-power function principle.

6.3.4 Proofs

Let $\xi_1, \xi_2, \ldots$ be a sequence of independent rv's, defined on the same probability space as the $X_i's$, uniformly distributed on $(0,1)$. For each integer n, the uniform empirical df is defined by

$$\mathbb{G}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{(\xi_i \leq t)},\ 0 \leq t \leq 1,$$

and the corresponding uniform empirical quantile function is given by

$$\mathbb{V}_n(t) = \inf\{s : \mathbb{G}_n(s) \geq t\}, 0 \leq s \leq 1, \mathbb{V}_n(0) = \mathbb{V}_n(0+).$$

In terms of the order statistics $\xi_{1,n} \leq \cdots \leq \xi_{n,n}$ pertaining to the sample $\xi_1, \ldots, \xi_n$, we have

$$\mathbb{V}_n(s) = \xi_{j,n},\ \frac{j-1}{n} < s \leq \frac{j}{n},\ j = 1, \ldots, n \text{ and } \mathbb{V}_n(0) = \xi_{1,n}.$$

The corresponding uniform quantile process is defined by

$$\beta_n(t) = \sqrt{n}\,(t - \mathbb{V}_n(t)),\ 0 \leq t \leq 1.$$

The two sequences of order statistics $X_{1,n} \leq \cdots \leq X_{n,n}$ and $\xi_{1,n} \leq \cdots \leq \xi_{n,n}$ are linked via the following equality

$$\{X_{n-j+1,n}\}_{j=1}^{n} \overset{\mathcal{D}}{=} \left\{Q\left(1-\xi_{n-j+1,n}\right)\right\}_{j=1}^{n}. \tag{6.26}$$

In this section, we use the well-known Gaussian approximation given by Csörgő et al. (1986). On the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, there exists a sequence of Brownian bridges $\{\mathbb{B}_n(s);\ 0 \leq s \leq 1\}_{n\geq 1}$ such that for every $0 \leq \nu < 1/2$

$$\sup_{1/n\leq s\leq 1-1/n} \frac{|\beta_n(s) - \mathbb{B}_n(s)|}{\left(s(1-s)\right)^{1/2-\nu}} = O_{\mathbb{P}}\left(n^{-\nu}\right), \text{ as } n\to\infty. \tag{6.27}$$

Proof of Theorem 6.3.1. We have

$$\begin{aligned}\Pi(g) &= \int_{k/n}^{1} g'(s)\,Q(1-s)\,ds + \int_{0}^{k/n} g'(s)\,Q(1-s)\,ds \\ &= \Pi_1(g) + \Pi_2(g),\end{aligned}$$

and

$$\begin{aligned}\widehat{\Pi}^{K}(g) &= \sum_{i=k+1}^{n} a_{i,n}(g)\,X_{n-i+1,n} + \frac{g(k/n)}{1-\beta\widehat{\gamma}^{K}}X_{n-k,n} \\ &= \widehat{\Pi}_1(g) + \widehat{\Pi}_2^{K}(g).\end{aligned}$$

It is clear that

$$\frac{\sqrt{k}\left(\widehat{\Pi}_1(g) - \Pi_1(g)\right)}{g(k/n)\,Q(1-k/n)} = \frac{\sqrt{n}\left(\widehat{\Pi}_1(g) - \Pi_1(g)\right)}{(n/k)^{1/2}\,g(k/n)\,Q(1-k/n)}.$$

Necir and Meraghni (2009)„ by using the second-order condition (6.23), have shown that, as $n\to\infty$,

$$\frac{\sqrt{n}\left(\widehat{\Pi}_1(g) - \Pi_1(g)\right)}{(n/k)^{1/2}\,g(k/n)\,Q(1-k/n)} = \mathcal{W}_{1n} + o_{\mathbb{P}}(1),$$

this implies that

$$\frac{\sqrt{k}\left(\widehat{\Pi}_1(g) - \Pi_1(g)\right)}{g(k/n)\,Q(1-k/n)} = \mathcal{W}_{1n} + o_{\mathbb{P}}(1), \text{ as } n\to\infty, \tag{6.28}$$

where

$$\mathcal{W}_{1n} = -\frac{\int_{k/n}^{1} g'(s)\,\mathbb{B}_n(1-s)\,Q'(1-s)\,ds}{(n/k)^{1/2}\,g(k/n)\,Q(1-k/n)}.$$

From (6.26), we can rewritte $\widehat{\Pi}_2^K(g)$ as follows

$$\widehat{\Pi}_2^K(g) \overset{\mathcal{D}}{=} \frac{g(k/n)}{1-\beta\widehat{\gamma}^K} Q\left(1-\xi_{n-k,n}\right).$$

It is easy to verify that

$$\frac{\sqrt{k}}{g(k/n)\,Q(1-k/n)}\left(\widehat{\Pi}_2^K(g) - \Pi_2(g)\right) = \sum_{i=1}^{4} T_{in},$$

where

$$T_{1n} = \frac{\sqrt{k}}{1-\beta\widehat{\gamma}^K}\left[\frac{Q\left(1-\xi_{n-k,n}\right)}{Q(1-k/n)} - \left(\frac{n}{k}\left(1-\xi_{n-k,n}\right)\right)^{-\gamma}\right],$$

$$T_{2n} = \frac{\sqrt{k}}{1-\beta\widehat{\gamma}^K}\left[\left(\frac{n}{k}\left(1-\xi_{n-k,n}\right)\right)^{-\gamma} - 1\right],$$

$$T_{3n} = \frac{\beta}{\left(1-\beta\widehat{\gamma}^K\right)(1-\beta\gamma)}\sqrt{k}\left(\widehat{\gamma}^K - \gamma\right),$$

and

$$T_{4n} = \sqrt{k}\left[\frac{1}{1-\beta\gamma} - \frac{\int_0^{k/n} g'(s)\,Q(1-s)}{g(k/n)\,Q(1-k/n)}ds\right].$$

We begin with T_{1n}, according to de Haan and Ferreira (2006, and theorem 2.3.9, page 48), for any $\delta > 0$, we have

$$\begin{aligned}
&\frac{Q\left(1-\xi_{n-k,n}\right)}{Q(1-k/n)} - \left(\frac{n}{k}\left(1-\xi_{n-k,n}\right)\right)^{-\gamma} \\
&= A_0\left(\frac{n}{k}\right)\left\{\left(\frac{n}{k}\left(1-\xi_{n-k,n}\right)\right)^{-\gamma}\frac{\left(\frac{n}{k}\left(1-\xi_{n-k,n}\right)\right)^{-\rho}-1}{\rho}\right. \\
&\qquad\left. + o_{\mathbb{P}}(1)\left(\frac{n}{k}\left(1-\xi_{n-k,n}\right)\right)^{-\gamma-\rho\pm\delta}\right\},
\end{aligned}$$

where $A_0(t) \sim A(t)$ as $t \to \infty$. Since $\frac{n}{k}\left(1-\xi_{n-k,n}\right) = 1 + o_{\mathbb{P}}(1)$ (see Balkema and de Haan , 1975) and $\sqrt{k}A(n/k) = O(1)$ as $n \to \infty$, then we have

$$\sqrt{k}\left(\frac{Q\left(1-\xi_{n-k,n}\right)}{Q(1-k/n)} - \left(\frac{n}{k}\left(1-\xi_{n-k,n}\right)\right)^{-\gamma}\right) = o_{\mathbb{P}}(1).$$

From Csörgő et al. (1985), $\widehat{\gamma}^K \xrightarrow{\mathbb{P}} \gamma$, we obtain for all large n

$$T_{1n} = o_{\mathbb{P}}(1). \tag{6.29}$$

For T_{2n}, Taylor's expansion yields

$$\sqrt{k}\left\{\left(\frac{n}{k}\left(1-\xi_{n-k,n}\right)\right)^{-\gamma}-1\right\} = -\gamma\left(\lambda_n(k)\right)^{-\gamma-1}\sqrt{k}\left\{\frac{n}{k}\left(1-\xi_{n-k,n}\right)-1\right\},$$

where $\lambda_n(k)$ is a sequence of rv's with values in the open interval of endpoints 1 and $\frac{n}{k}\left(1-\xi_{n-k,n}\right)$ which, from Balkema and de Haan (1975), converges in probability to 1 as $n\to\infty$. It follows that, $\lambda_n(k)\xrightarrow{\mathbb{P}}1$, as $n\to\infty$. Therefore, as $n\to\infty$,

$$\sqrt{k}\left\{\left(\frac{n}{k}\left(1-\xi_{n-k,n}\right)\right)^{-\gamma}-1\right\} = -\gamma\sqrt{k}\left\{\frac{n}{k}\left(1-\xi_{n-k,n}\right)-1\right\}\left(1+o_{\mathbb{P}}(1)\right).$$

On the other hand we have

$$\sqrt{k}\left\{\frac{n}{k}\left(1-\xi_{n-k,n}\right)-1\right\} = \sqrt{\frac{n}{k}}\left\{\sqrt{n}\left(\left(1-\frac{k}{n}\right)-\mathbb{V}_n\left(1-\frac{k}{n}\right)\right)\right\}.$$

Then, for all large values of n

$$\sqrt{k}\left\{\left(\frac{n}{k}\left(1-\xi_{n-k,n}\right)\right)^{-\gamma}-1\right\} = -\gamma\sqrt{\frac{n}{k}}\beta_n\left(1-\frac{k}{n}\right)\left(1+o_{\mathbb{P}}(1)\right).$$

By using the Gaussian approximation (6.27), we get as $n\to\infty$,

$$\begin{aligned}\sqrt{k}\left\{\left(\frac{n}{k}\left(1-\xi_{n-k,n}\right)\right)^{-\gamma}-1\right\} &= -\gamma\sqrt{\frac{n}{k}}\left\{\mathbb{B}_n\left(1-\frac{k}{n}\right)\right.\\ &\quad\left.+O_{\mathbb{P}}\left(n^{-\nu}\right)\left(\frac{k}{n}\right)^{1/2-\nu}\right\}\left(1+o_{\mathbb{P}}(1)\right)\\ &= -\gamma\sqrt{\frac{n}{k}}\mathbb{B}_n\left(1-\frac{k}{n}\right)\left(1+o_{\mathbb{P}}(1)\right).\end{aligned}$$

Consequently, since $\widehat{\gamma}^K \xrightarrow{\mathbb{P}} \gamma$, we obtain for all large n

$$\begin{aligned}T_{2n} &= -\frac{\gamma}{1-\beta\gamma}\sqrt{\frac{n}{k}}\mathbb{B}_n\left(1-\frac{k}{n}\right)+o_{\mathbb{P}}(1)\\ &= \mathcal{W}_{2n}+o_{\mathbb{P}}(1).\end{aligned} \tag{6.30}$$

For T_{3n}, from theorem 1 of Deme et al. (2013), we have for all large n

$$\sqrt{k}\left(\widehat{\gamma}^{K}-\gamma\right)=\sqrt{k}A\left(n/k\right)\int_0^1 s^{-\rho}K\left(s\right)ds + \gamma\sqrt{\frac{n}{k}}\int_0^1 s^{-1}\mathbb{B}_n\left(1-\frac{sk}{n}\right)d\left(sK\left(s\right)\right)+o_{\mathbb{P}}\left(1\right).$$

Then, since $\widehat{\gamma}^{K}\xrightarrow{\mathbb{P}}\gamma$, we get as $n\to\infty$,

$$\begin{aligned} T_{3n} &= \frac{\beta}{\left(1-\beta\gamma\right)^2}\left\{\sqrt{k}A\left(\frac{n}{k}\right)\int_0^1 s^{-\rho}K\left(s\right)ds \right. \\ &\left. +\gamma\sqrt{\frac{n}{k}}\int_0^1 s^{-1}\mathbb{B}_n\left(1-\frac{sk}{n}\right)d\left(sK\left(s\right)\right)\right\}+o_{\mathbb{P}}\left(1\right) \\ &= \frac{\beta}{\left(1-\beta\gamma\right)^2}\sqrt{k}A\left(\frac{n}{k}\right)\int_0^1 s^{-\rho}K\left(s\right)ds+\mathcal{W}_{3,n}\left(K\right)+o_{\mathbb{P}}\left(1\right). \end{aligned} \tag{6.31}$$

For T_{4n}, a change of variables and an integration by parts yield

$$\begin{aligned} T_{4n} &= \sqrt{k}\left[\frac{1}{1-\beta\gamma}-\frac{k}{n}\int_0^1\frac{g'\left(sk/n\right)Q\left(1-sk/n\right)}{g\left(k/n\right)Q\left(1-k/n\right)}ds\right] \\ &= \sqrt{k}\left[\frac{1}{1-\beta\gamma}-\frac{k}{n}\int_0^1 s^{-\gamma}\frac{g'\left(sk/n\right)}{g\left(k/n\right)}ds \right. \\ &\left. -\frac{k}{n}\int_0^1\frac{g'\left(sk/n\right)}{g\left(k/n\right)}\left(\frac{Q\left(1-ks/n\right)}{Q\left(1-k/n\right)}-s^{-\gamma}\right)ds\right]. \end{aligned}$$

Since g is regularly varying function at zero with index $1/\beta>0$ and with $g\left(0\right)=0$, then by using the 11th assertion of proposition B.1.9 in de Haan and Ferreira, 2006, yields that for $t\downarrow 0$

$$g'\left(t\right)=\frac{1}{\beta}t^{-1}g\left(t\right)+o\left(1\right). \tag{6.32}$$

Then, by using (6.32) and (6.19), we get as $n\to\infty$,

$$\begin{aligned} \frac{k}{n}\int_0^1 s^{-\gamma}\frac{g'\left(sk/n\right)}{g\left(k/n\right)}ds &= \frac{1}{\beta}\int_0^1 s^{-\gamma-1}\frac{g\left(sk/n\right)}{g\left(k/n\right)}ds+o\left(1\right) \\ &= \frac{1}{\beta}\int_0^1 s^{\frac{1}{\beta}-\gamma-1}\frac{\ell_g\left(sk/n\right)}{\ell_g\left(k/n\right)}ds+o\left(1\right) \\ &= \frac{1}{1-\beta\gamma}+o\left(1\right). \end{aligned}$$

Finally, we get

$$T_{4n} = -\sqrt{k}\frac{k}{n}\int_0^1 \frac{g'(sk/n)}{g(k/n)}\left(\frac{Q(1-ks/n)}{Q(1-k/n)} - s^{-\gamma}\right)ds + o(1), \text{ as } n\to\infty.$$

Next, we apply the uniform inequality of regularly varying functions (see theorem 2.3.9 in de Haan and Ferreira, 2006). For a possibly different function A_0, with $A_0(x) \sim A(x)$, $tx \to 1$, and for any $\delta > 0$, there exists a thresholds $s_\delta \in (0,1)$ such that for all t, $ts \le s_\delta$

$$\left|\frac{1}{A_0(1/t)}\left(\frac{Q(1-ts)}{Q(1-t)} - s^{-\gamma}\right) - s^{-\gamma}\frac{s^{-\rho}-1}{\rho}\right| \le \delta s^{-\rho-\gamma}\max\left(s^{\delta}, s^{-\delta}\right). \quad (6.33)$$

We have $g'(s) \ge 0$ with $s \in (0,1)$, because g is non-decreasing and differentiable function. Then, by using the previous inequality (6.33), we have

$$\left|\frac{1}{A_0(n/k)}\frac{k}{n}\int_0^1 \frac{g'(sk/n)}{g(k/n)}\left(\frac{Q(1-ks/n)}{Q1(-k/n)} - s^{-\gamma}\right)ds \right.$$
$$\left. -\frac{1}{\rho}\frac{k}{n}\int_0^1 s^{-\gamma}\left(s^{-\rho}-1\right)\frac{g'(sk/n)}{g(k/n)}ds\right| \le \delta\frac{k}{n}\int_0^1 s^{-\rho-\gamma-\delta}\frac{g'(sk/n)}{g(k/n)}ds.$$

By using (6.32) and (6.19), we obtain as $n\to\infty$,

$$\frac{k}{n}\int_0^1 s^{-\rho-\gamma-\delta}\frac{g'(sk/n)}{g(k/n)}ds = O(1).$$

Therefore, as $n\to\infty$,

$$T_{4n} = -\frac{1}{\rho}\sqrt{k}A\left(\frac{n}{k}\right)\frac{k}{n}\int_0^1 s^{-\gamma}\left(s^{-\rho}-1\right)\frac{g'(sk/n)}{g(k/n)}ds + o(1).$$

By using again (6.32) and (6.19), and after easy calculation, we get as $n\to\infty$,

$$T_{4n} = \sqrt{k}A\left(\frac{n}{k}\right)\frac{\beta}{(1-\beta\gamma)(\beta\gamma+\beta\rho-1)} + o(1). \quad (6.34)$$

Combining (6.29)–(6.31) and (6.34), we get

$$\frac{\sqrt{k}\left(\widehat{\Pi}_2^K(g) - \Pi_2(g)\right)}{g(k/n)Q(1-k/n)} = \sqrt{k}A(n/k)\mathcal{AB}_K(\gamma,\beta,\rho) + \mathcal{W}_{2n} + \mathcal{W}_{3n}(K) + o_{\mathbb{P}}(1). \quad (6.35)$$

Finally, combining (6.28) and (6.35), theorem 6.3.1 follows. ■

Proof of Corollary 6.3.1. Since $\{\mathbb{B}_n(s);\ 0 \leq s \leq 1\}_{n\geq 1}$, is a sequence of Brownian bridges, then

$$\frac{\sqrt{k}}{g(k/n)\,Q(1-k/n)}\left(\widehat{\Pi}^{\underline{K}}(g) - \Pi(g)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \mathcal{AV}_K(\gamma,\beta)\right), \text{ as } n \to \infty,$$

with

$$\begin{aligned}\mathcal{AV}_K(\gamma,\beta) &= \lim_{n\to\infty} E\left((\mathcal{W}_{1n} + \mathcal{W}_{2n} + \mathcal{W}_{3n}(K))^2\right),\\ &= \lim_{n\to\infty}\left(E\left(\mathcal{W}_{1n}^2\right) + E\left(\mathcal{W}_{2n}^2\right) + E\left(\mathcal{W}_{3n}^2(K)\right)\right.\\ &\quad \left.+\, 2E(\mathcal{W}_{1n}\mathcal{W}_{2n}) + 2E(\mathcal{W}_{1n}\mathcal{W}_{3n}(K)) + 2E(\mathcal{W}_{2n}\mathcal{W}_{3n}(K))\right).\end{aligned}$$

After elementary but tedious computations, we obtain as $n \to \infty$,

$$\begin{aligned} E\left(\mathcal{W}_{1n}^2\right) &= \frac{2\gamma^2}{(\gamma\beta+\beta-1)(2\gamma\beta+\beta-2)} + o(1),\\ E\left(\mathcal{W}_{2n}^2\right) &= \frac{\gamma^2}{(1-\beta\gamma)^2} + o(1),\\ E\left(\mathcal{W}_{3n}^2(K)\right) &= \frac{\gamma^2\beta^2}{(1-\gamma\beta)^4}\int_0^1 K^2(s)ds + o(1),\\ E(\mathcal{W}_{1n}\mathcal{W}_{2n}(K)) &= \frac{\gamma^2}{(\gamma\beta+\beta-1)(1-\gamma\beta)} + o(1),\\ E(\mathcal{W}_{1n}\mathcal{W}_{3n}(K)) &= o(1),\end{aligned}$$

and

$$E(\mathcal{W}_{2n}\mathcal{W}_{3n}(K)) = o(1).$$

Then, we get

$$\mathcal{AV}_K(\gamma,\beta) = \frac{\gamma^2\beta}{(2\gamma\beta+\beta-2)(1-\gamma\beta)^2} + \frac{\gamma^2\beta^2}{(1-\gamma\beta)^4}\int_0^1 K^2(s)ds.$$

We complete the proof of corollary 6.3.1. ■

Proof of Corollary 6.3.2. The proof is a direct result of corollary 6.3.1 with the kernel $K = \underline{K} = \mathbf{1}_{(0,1)}$. ■

Proof of Theorem 6.3.2. According to theorem 6.3.1 and (6.22), we have as $n \to \infty$,

$$\frac{\sqrt{k}}{g(k/n)\,Q(1-k/n)}\left(\widetilde{\Pi}^{K}(g)-\Pi(g)\right)=\mathcal{W}_{1n}+\mathcal{W}_{2n}+\mathcal{W}_{3n}(K)+\mathcal{W}_{4n}(K)+o_{\mathbb{P}}(1),$$

where

$$\begin{aligned}\mathcal{W}_{4n}(K)&=\sqrt{k}\left(A(n/k)\mathcal{AB}_K(\gamma,\beta,\rho)-\widehat{A}^{LS}(\widehat{\rho})\,\mathcal{AB}_K\left(\widehat{\gamma}^{LS}(\widehat{\rho}),\beta,\widehat{\rho}\right)\frac{X_{n-k,n}}{Q(1-k/n)}\right)\\&=-\mathcal{AB}_K(\gamma,\beta,\rho)\sqrt{k}\left(\widehat{A}^{LS}(\widehat{\rho})-A(n/k)\right)\\&\quad-\sqrt{k}\widehat{A}^{LS}(\widehat{\rho})\left(\mathcal{AB}_K\left(\widehat{\gamma}^{LS}(\widehat{\rho}),\beta,\widehat{\rho}\right)-\mathcal{AB}_K(\gamma,\beta,\rho)\right)\\&\quad-\sqrt{k}\widehat{A}^{LS}(\widehat{\rho})\,\mathcal{AB}_K\left(\widehat{\gamma}^{LS}(\widehat{\rho}),\beta,\widehat{\rho}\right)\left(\frac{X_{n-k,n}}{Q(1-k/n)}-1\right).\end{aligned}$$

From Lemma 5 of Deme et al. (2013), we have

$$\begin{aligned}\sqrt{k}\left(\widehat{A}^{LS}(\widehat{\rho})-A(n/k)\right)&=\gamma(1-\rho)\\&\quad\times\sqrt{\frac{n}{k}}\int_0^1 s^{-1}\mathbb{B}_n\left(1-\frac{sk}{n}\right)d\left(s\left(\underline{K}(s)-K_\rho(s)\right)\right)+o_{\mathbb{P}}(1).\end{aligned}$$

Since $\widehat{\rho}$ is a consistent estimator for ρ, then we get as $n \to \infty$

$$\sqrt{k}\widehat{A}^{LS}(\widehat{\rho})\left(\mathcal{AB}_K\left(\widehat{\gamma}^{LS}(\widehat{\rho}),r,\widehat{\rho}\right)-\mathcal{AB}_K(\gamma,r,\rho)\right)=o_{\mathbb{P}}(1).$$

Making use of Potter's inequalities (see 5th assertion of proposition B.1.9 in de Haan, 2006), we obtain as $n \to \infty$

$$\sqrt{k}\widehat{A}^{LS}(\widehat{\rho})\,\mathcal{AB}_K\left(\widehat{\gamma}^{LS}(\widehat{\rho}),r,\widehat{\rho}\right)\left(\frac{X_{n-k,n}}{\mathbb{U}(n/k)}-1\right)=o_{\mathbb{P}}(1).$$

Therefore

$$\begin{aligned}\mathcal{W}_{4n}(K)&=-\gamma(1-\rho)\,\mathcal{AB}_K(\gamma,\beta,\rho)\\&\quad\times\sqrt{\frac{n}{k}}\int_0^1 s^{-1}\mathbb{B}_n\left(1-\frac{sk}{n}\right)d\left(s\left(\underline{K}(s)-K_\rho(s)\right)\right)+o_{\mathbb{P}}(1),\ \text{as } n\to\infty.\end{aligned}$$

It is clear that $\mathcal{W}_{1n}+\mathcal{W}_{2n}+\mathcal{W}_{3n}(K)+\mathcal{W}_{4n}(K)$ is a Gaussian rv with mean zero and variance

$$\widetilde{\mathcal{AV}}_K(\gamma,\beta,\rho)=\lim_{n\to\infty}E\left(\left(\mathcal{W}_{1n}+\mathcal{W}_{2n}+\mathcal{W}_{3n}(K)+\mathcal{W}_{4n}(K)\right)^2\right).$$

Elementary calculation gives

$$\widetilde{\mathcal{AV}}_K(\gamma,\beta,\rho) = \mathcal{AV}_K(\gamma,\beta) + \frac{\gamma^2(1-2\rho)(1-\rho)^2}{\rho^2}\mathcal{AB}_K^2(\gamma,\beta,\rho)$$
$$+ \frac{2\gamma^2\beta(1-2\rho)(1-\rho)^2}{\rho^2(1-\gamma\beta)^2}\left(1-(1-\rho)\int_0^1 s^{-\rho}K(s)ds\right)\mathcal{AB}_K(\gamma,\beta,\rho).$$

This achieves the proof of theorem 6.3.2. ■

Proof of Corollary 6.3.3. The proof is a direct result of theorem 6.3.2 with the kernel $K = \underline{K} = \mathbf{1}_{(0,1)}$. ■

Proof of Corollary 6.3.4. From the proof of theorem 6.3.2, we deduce the corollary 6.3.4. ■

Bibliography

[1] Artzner, P., Delbaen, F., Eber, J.M. and Heath, D. (1999). Coherent Measures of Risk. Mathematical Finance 9, 203-228.

[2] Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N. (1992). A First Course in Order Statistics. Wiley, New York.

[3] Balkema, A. and de Haan, L. (1975). Limit Laws of Order Statistics. In: Limit Theorems of Probability Theory (Colloquium, Keszthely, 1974), vol. 11 of Colloquia Mathematica Societatis Janos Bolyai, pp.17-22. North Holland, Amsterdam.

[4] Beirlant, J., Matthys, G. and Dierckx, G. (2001). Heavy-tailed Distributions and Rating. ASTIN Bulletin 31, 37-58.

[5] Beirlant, J., Dierckx, G., Goegebeur, M. and Matthys, G. (1999). Tail Index Estimation and an Exponential Regression Model. Extremes 2, 177-200.

[6] Beirlant, J., Dierckx, G., Guillou, A. and Starica, C. (2002). On Exponential Representations of Log-spacings of Extreme Order Statistics. Extremes 5, 157-180.

[7] Beirlant, J., Bouquiaux, C. and Werker, B.J.M. (2006). Semi-parametric Lower Bounds for Tail Index Estimation. Journal of Statistical Planning and Inference 136, 705-729.

[8] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). Statistics of Extremes: Theory and Applications. Wiley.

[9] Behboodian, J., Dolati, A. and Úbeda-Flores, M. (2007). A Multivariate Version of Gini's Rank Association Coefficient. Statistical Papers 48, 295-304.

[10] Benkhelifa, L. (2009). Estimation du Parametre de Distortion pour les Mesures de Risques. Mémoire présenté en vue de l'obtention du diplôme de Magister en Mathématiques, Université Mohamed Khider, Biskra.

[11] Benkhelifa, L. (2014). Kernel-type Estimator of the Reinsurance Premium for Heavy-tailed Loss Distributions. Insurance: Mathematics and Economics 59, 65-70.

[12] Benkhelifa, L. (2014). Kernel-type Estimators for the Distortion Risk Premiums of Heavy-tailed Distributions. Scandinavian Actuarial Journal, DOI: 10.1080/03461238.2014.924434.

[13] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). Regular Variation. Cambridge University Press, Cambridge.

[14] Blomqvist, N. (1950). On a Measure of Dependence Between Two Random Variables. Ann Math Stat 21, 593-600.

[15] Brahimi, B., Meddi, F. and Necir A. (2012). Bias-corrected Estimation in Distortion Risk Premiums for Heavy-tailed Losses. Journal Afrika Statistika 7, 474-490.

[16] Brazauskas, V. and Serfling, R. (2000). Robust and Efficient Estimation of the Tail Index of a Single-Parameter Pareto Distribution. North American Actuarial Journal 4, 12-27.

[17] Brazauskas, V., Jones, B., Puri, M. and Zitikis, R. (2008). Estimating Conditional Tail Expectation with Actuarial Applications in View. Journal of Statistical Planning and Inference 138, 3590-3604.

[18] Bühlmann, H. (1970). Mathematical Models in Risk Theory. Springer, New York.

[19] Bühlmann, H. (1980). An Economic Premium Principle. ASTIN Bulletin 11, 52-60.

[20] Caeiro, F., Gomes, M.I. and Rodrigues, L.H. (2009). Reduced-bias Tail Index Estimators under a Third-order Framework. Communications in Statistics-Theory and Methods 38, 1019-1040.

[21] Cambanis, S., Huang, S. and Simons G. (1981). On the Theory of Elliptically Contoured Distributions. Journal of Multivariate Analysis 1, 368-385.

[22] Castillo, E., Hadi, A., Balakrishnan, N. and Sarabia, J.M. (2005). Extreme Value and Related Models with Applications in Engineering and Science. Wiley, Hoboke, New Jersey.

[23] Centeno, M.L. and Andrade e Silva, J. (2005). Applying the Proportional Hazard Premium Calculation Principle. ASTIN Bulletin 35, 409-425.

[24] Cheng, S. and Peng, L. (2001). Confidence Intervals for the Tail Index. Bernoulli 7, 751-760.

[25] Cifarelli, D.M., Conti, P.L. and Regazzini, E. (1996). On the asymptotic distribution of a general measure of monotone dependence. Annals of Statistics 24, 1386-1399.

[26] Clayton, D.G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika 65, 141-151.

[27] Christopeit, N. (1994). Estimating Parameters of an Extreme Value Distribution by the Method of Moments. Journal of Statistical Planning and Inference 41, 173-186.

[28] Csörgő, S., Deheuvels, P. and Mason, D.M. (1985). Kernel Estimates of the Tail Index of a Distribution. Annals of Statistics 13, 1050-1077.

[29] Csörgő, M., Csörgő, S., Horváth, L. and Mason, D. (1986). Weighted Empirical and Quantile Processes. Annals of Probability 14, 31-85.

[30] Csörgő, S. and Mason, D.M. (1985). Central Limit Theorems for Sums of Extreme Values. Mathematical Proceedings of the Cambridge Philosophical Society 98, 547-558.

[31] Csörgő, S. and Viharos, L. (1998). Estimating the Tail Index. Asymptotic Methods in Probability and Statistics. B. Szyszkowicz, ed., North Holland, 833-881.

[32] Danielsson, J., de Haan, L., Peng, L. and de Vries, C.G. (2001). Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation. Journal of Multivariate Analysis 76, 226-248.

[33] Draisma, G., de Haan, L., Peng, L. and Perreira, T.T. (1999). A Bootstrap-Based Method to Achieve Optimality in Estimating the Extereme-Value Index. Extremes 2, 367-404.

[34] Davis, R. and Resnick, S.I. (1984). Tail Estimates Motivated by Extreme Value Theory. Annals of Statistics 12, 1467-1487.

[35] Davison, A.C. and Smith R.L. (1990). Models for Exceedances over High Thresholds (with discussion). Journal of the Royal Statistical Society, Series B 52, 393-442.

[36] Deheuvels, P., (1978). Caractérisation Complète des lois Extrêmes Multivariées et de la Convergence des Types Extrêmes. Publications de l'Institut de Statistique de Paris 23, 1-37.

[37] Deheuvels, P. (1979). La fonction de Dépendance Empirique et ses Propriétés. Académie Royale de Belgique Bulletin de la Classe des Sciences 65, 274-292.

[38] Deheuvels, P., Häusler, E. and Mason, D.M. (1988). Almost Sure Convergence of the Hill Estimator. Mathematical Proceedings of the Cambridge Philosophical Society 104, 371-381.

[39] Dekkers, A.L.M. and de Haan, L. (1989). On the Estimation of the Extreme-Value Index and Large Quantile Estimation. Annals of Statistics 17, 1795-1832.

[40] Dekkers, A. L. M. and de Haan, L. (1993). Optimal Choice of Sample Fraction in Extreme Value Estimation. Journal of Multivariate Analysis 47, 173-195.

[41] Dekkers, A.L.M., Einmahl, J.H.J. and de Haan, L. (1989). A Moment Estimator for the Index of an Extreme Value Index. Annals of Statistics 17, 1833-1855.

[42] Deme, E. and Lo, G.S. (2013). Reduced-bias Estimators for the Distortion risk Premiums for Heavy-tailed Distributions. Available from http://hal.inria.fr/hal-00868624.

[43] Deme, E., Girard, S. and Guillou, A. (2013). Reduced-bias Estimator of the Proportional Hazard Premium for Heavy-tailed Distributions. Insurance: Mathematics and Economics 52, 550-559.

[44] Denuit, M., Dhaene, J., Goovaerts, M.J. and Kaas, R. (2005). Actuarial Theory for Dependent Risk: Measures, Orders and Models. Wiley, New York.

[45] Drees, H. (1996). Refined Pickands Estimator of the Extreme Value Index. Annals of Statistics 23, 2059-2080.

[46] Drees, H. and Kaufmann, E. (1998). Selection of the Optimal Sample Fraction in Univariate Extreme Value Estimation. Stochastic Processes and their Applications 75, 149-195.

[47] Dolati, A. and Úbeda-Flores, M. (2006). On Measures of Multivariate Concordance. Journal of Probability and Statistical Science 4, 147-164.

[48] Embrechts, P., Hoeing, A. and Juri., A. (2003). Using Copulae to Bound the Value-at-Risk for Functions of Dependent Risks. Finance & Stochastics 7, 145-167.

[49] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). Modelling Extremal Events for Insurance and Finance. Springer Verlag, Berlin.

[50] Fang, K.T., Kotz, S. and Ng, K.W. (1990). Symmetric Multivariate and Related Distributions. London, Chapman and Hall.

[51] Frahm, G. (2006). On the Extremal Dependence Coefficient of Multivariate Distributions. Statistics and Probability Letters 76, 1470-1481.

[52] Fermanian, J-D., Radulovic, D. and Wegkamp, M. (2004). Weak Convergence of Empirical Copula Processes. Bernoulli 10, 847–860.

[53] Fernández-Fernández, B. and González-Barrios, J.M. (2004). Multidimensional Dependency Measures. Journal of Multivariate Analysis 89, 351–370.

[54] Ferreira, A., de Haan, L. and Peng, L. (2003). On Optimizing the Estimation of High Quantiles of a Probability Distribution. Statistics 37, 401-434.

[55] Feuerverger, A. and Hall, P. (1999). Estimating a Tail Exponent by Modelling Departure from a Pareto Distribution. Annals of Statistics 27, 760-781.

[56] Foss, S., Korshunov D. and Zachary S. (2011). An Introduction to Heavy-tailed and Subexponential Distributions. Springer, New York.

[57] Fraga Alves, M.I., de Haan, L. and Lin, T. (2003). Estimation of the Parameter Controlling the Speed of Convergence in Extreme Value Theory. Mathematical Methods of Statistics 12, 155-176.

[58] Furman, E., Zitikis, R. (2008). Weighted Premium Calculation Principles. Insurance: Mathematics and Economics, 42, 459-465.

[59] Fisher, R.A. and Tippett, L.H.C. (1928). Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample. Proceedings of the Cambridge Philosophical Society 24, 180-190.

[60] Fraga Alves, M. I. (1995). Estimation of the Tail Parameter in the Domain of Attraction of an Extremal Distribution. Journal of Statistical Planning and Inference 45, 143-173.

[61] Gaißer, S.; Ruppert, M. and Schmid, F.(2010). A multivariate version of Hoeffding's Phi-Square. Journal of Multivariate Analysis 101, 2571-2586.

[62] Geluk, J.L. and de Haan, L. (1987). Regular Variation, Extensions and Tauberian Theorems. Mathematical Centre Tract 40, Centre for Mathematics and Computer Science, Amsterdam.

[63] Genest, C., and Favre, A.-C. (2007). Everything you Always Wanted to Know About Copula Modeling but were Afraid to Ask. Journal of Hydrologic Engineering 12, 347-368.

[64] Genest, C. and MacKay, R., (1986a). The joy of copulas: Bivariate distributions with uniform marginals. The American Statistician 40, 280-283.

[65] Genest, C. and MacKay, R., (1986b). Copules Archimédiennes et Familles de lois Bidimensionell dont les Marges sont Données. The Canadian Journal of Statistics 14, 145-159.

[66] Genest, C., Ghoudi, K. and Rivest, L.-P. (1995) . A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions. Biometrika 82, 543-552.

[67] Genest, C., Nešlehová, J. and Ben Ghorbal, N. (2010). Spearman's Footrule and Gini's Gamma: A Review with Complements. Journal of Nonparametric Statistics 22, 105-119.

[68] Gnedenko, B.V. (1943). Sur la Distribution Limite du Terme Maximum d'une Série Aléatoire. Annales de Mathématiques 44, 423-453.

[69] Gomes, M.I. and Figueiredo, F. (2006). Bias Reduction in Risk Modelling: Semi-parametric Quantile Estimation. Test 15, 375-396.

[70] Gomes, M.I. and Martins, M.J. (2002). "Asymptotically Unbiased" Estimators of the Tail Index Based on External Estimation of the Second Order Parameter. Extremes 5, 5-31.

[71] Gomes, M.I. and Martins, M.J. (2004). Bias Reduction and Explicit Semi-parametric Estimation of the Tail Index. Journal of Statistical Planning and Inference 124, 361-378.

[72] Goovaerts, M.J., de Vylder, F. and Haezendonck, J. (1984). Insurance Premiums: Theory and Applications. North Holland, Amsterdam.

[73] Groeneboom, P., Lopuhaä, H.P. and de Wolf, P.P. (2003). Kernel-type Estimators for the Extreme Value Index. Annals of Statistics 31, 1956-1995.

[74] Gumbel, E.J. (1935). Les Valeurs Extrêmes des Distribution Statistiques. Annales de l'Institut Henri Poincaré 5, 115-158.

[75] Gumbel, E.J. (1958). Statistics of Extremes. Columbia University Press, New York.

[76] Gumbel, E.J. (1960). Bivariate Exponential Distributions. Journal of the American Statistical Association 55, 698-707.

[77] de Haan, L. (1970). On Regular Variation and Its Application to the Weak Convergence of Sample Extremes. Mathematical Centre Tract 32, Centre for Mathematics and Computer Science, Amsterdam.

[78] de Haan, L. and Ferreira, A. (2006). Extreme Value Theory - An Introduction. Springer, New York.

[79] Hall, P. (1990). Using the Bootstrap to Estimate Mean Squared Error and Select Smoothing Parameter in Nonparametric Problems. Journal of Multivariate Analysis 32, 177-203.

[80] Hall, P. and Welsh, A.H. (1985). Adaptive Estimates of Parameters of Regular Variation. Annals of Statistics 13, 331-341.

[81] Häusler, E. and Teugels, J.L. (1985). On Asymptotic Normality of Hill's estimator for the Exponent of Regular Variation. Annals of Statistics 13, 743-756.

[82] Hill, B. (1975). A Simple Approach to Inference About the Tail of a Distribution. Annals of Statistics 3, 1163-1174.

[83] Hoeffding, W. (1940). Masstabinvariante Korrelationstheorie, Schrift. Math. Seminars Inst. Angew. Math. Univ. Berlin 5, 181-233.

[84] Hofert, M. (2008). Sampling Archimedean Copulas. Computational Statistics and Data Analysis 52, 5163-5174.

[85] Hosking, J.R.M.,Wallis, J.R. and Wood, E.F. (1985). Estimation of the Generalized Extreme Value Distribution by the Method of Probability Weighted Moments. Technometrics 27, 251-261.

[86] Hosking, J.R.M. and Wallis, J.R. (1987). Parameter and Quantile Estimation for the Generalized Pareto Distribution. Technometrics 29, 339-349.

[87] Hult, H. and Lindskog F. (2002). Multivariate Extremes, Aggregation and Dependence in Elliptical Distributions. Advances in Applied Probability 34, 587-608.

[88] Ihaka, R. and Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. Journal of Computational and Graphical Statistics 5, 299-314.

[89] Jaworski, P., Durante, F., Haerdle, W. and Rychlik T. (2010). Copula theory and its applications. Lecture notes in statistics – proceedings, vol 198. Springer, Berlin.

[90] Jenkinson, A. F. (1955). The Frequency Distribution of the Annual Maximum (Minimum) Values of Meteorological Events. Quarterly Journal of the Royal Meteorological Society 81, 158-172.

[91] Joe, H., (1990). Multivariate Concordance. Journal of Multivariate Analysis 35, 12-30.

[92] Joe, H. (1997). Multivariate Models and Dependence Concepts. Chapman & Hall, London.

[93] Joe, H., (2005). Asymptotic efficiency of the two-stage estimation method for copulabased models. Journal of Multivariate Analysis 94, 401-419.

[94] Joe, H., and Xu, J.J. (1996). The Estimation Method of Inference Function for Margins for Multivariate Models. British Columbia, Dept. of Statistics, Technical Report 166.

[95] Jones, B.L., Puri, M.L. and Zitikis, R. (2006). Testing Hypotheses About the Equality of Several Risk Measure Values with Applications in Insurance. Insurance: Mathematics and Economics 38, 253-270.

[96] Jones, B.L. and Zitikis, R. (2003). Empirical Estimation of Risk Measures and Related Quantities. North American Actuarial Journal 7, 44-54.

[97] Jones, B.L. and Zitikis, R. (2005). Testing for the Order of Risk Measures: An application of L-statistics in Actuarial Science. Metron-International, Journal of Statistics 63, 193-211.

[98] Jones, B.L. and Zitikis, R. (2007). Risk measures, Distortion Parameters, and their Empirical Estimation. Insurance: Mathematics and Economics 41, 279-297.

[99] Kaas, R., Goovaerts, M., Dhaene, J. and Denuit, M. (2001). Modern Actuarial Risk Theory. Kluwer Academic Publishers, Dordrecht.

[100] Kaas, R., van Heerwaarden, A.E. and Goovaerts, M.J. (1994). Ordering of Actuarial Risks. Education Series 1, CAIRE, Brussels.

[101] Karamata, J. (1930). Sur un Mode de Croissance Régulire des Fonctions. Mathematica (Cluj) 4, 38-53.

[102] Kendall, M.G. (1938). A New Measure of Rank Correlation. Biometrika 30, 81-93.

[103] Kimeldorf, G. and Sampson, A. (1975). Uniform Representations of Bivariate Distributions. Communications in Statistics 4, 617–627

[104] Kratz, M.F. and Resnick, S.I. (1996). The QQ-Estimator of the Index of Regular variation. Communications in Statistics - Stochastic Models 12, 699-724.

[105] Landsman, Z. and Sherris, M. (2001). Risk Measures and Insurance Premium Principles. Insurance: Mathematics and Economics 29, 103-115.

[106] Malevergne, Y. and Sornette, D. (2005). Extreme Financial Risks - From Dependence to Risk Management. Springer, New York.

[107] Markovich, N. (2007). Nonparametric Analysis of Univariate Heavy-tailed Data: Research and Practice. Wiley.

[108] Marshall, A.W. and Olkin, I. (1988). Families of Multivariate Distributions. Journal of the American Statistical Association 83, 834–841.

[109] Matthys, G. and Beirlant, J. (2003). Estimating the Extreme Value Index and High Quantiles with Exponential Regression Models. Statistica Sinica 13, 853-880.

[110] Meraghni, D. (2008). Modeling Distribution Tails. Thèse de Doctorat, Université de Biskra, Algérie.

[111] von Mises, R. (1936). La Distribution de la Plus Grande des n Valeurs. Selected Papers, American Mathematical Society, 271-294.

[112] Necir, A. (2006). A Functional Law of the Iterated Logarithm for Kernel-type Estimators of the Tail Index. Journal of Statistical Planning and Inference 136, 780-802.

[113] Necir, A. and Boukhetala, K. (2004). Estimating the Risk Adjusted Premium of the Largest Reinsurance Covers. Proceeding in Computational Statistics (Edited by Jaromir Antoch), 1577-1584. Physica Verlag, Springer, ISBN: 3-7908-1554-3.

[114] Necir, A. and Meraghni, D. (2009). Empirical estimation of the Proportional Hazard Premium for Heavy-tailed Claim Amounts. Insurance: Mathematics and economics 45, 49-58.

[115] Necir, A. and Zitikis, R. (2011). Coupled Risk Measures and their Empirical Estimation When Losses Follow Heavy-tailed Distributions. Available at: http://ssrn.com/abstract=1855623.

[116] Necir, A., Meraghni, D. and Meddi, F. (2007). Statistical Estimate of the Proportional Hazard Premium of Loss. Scandinavian Actuarial Journal 3, 147-161.

[117] Necir, A., Rassoul, A. and Zitikis, R. (2010). Estimating the Conditional Tail Expectation in the Case of Heavy-tailed Losses. Journal of Probability and Statistics 2010, ID 956838.

[118] Nelsen, R.B. (1996). Nonparametric measures of multivariate association, in: Distribution with Fixed Marginals and Related Topics. IMS Lecture Notes - Monograph Series, Vol. 28, 223–232.

[119] Nelsen, R. (2006). An Introduction to Copulas. 2nd ed. Springer, New York.

[120] Neves, C. and Fraga Alves, M.I. (2004). Reiss and Thomas' Automatic Selection of the Number of Extremes. Computational Statistics and Data Analysis 47, 689-704.

[121] Peng, L. (2001). Estimating the Mean of a Heavy Tailed Distribution. Statistics and Probability Letters 52, 255-264.

[122] Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. Annals of Statistics 3, 119-131.

[123] Quessy, J.-F., (2009). Theoretical Efficiency Comparisons of Independence Tests Based on Multivariate Versions of Spearman's rho. Metrika 70, 315-338.

[124] Reiss, R.D. and Thomas, M. (2007). Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields, 3rd edition. Birkhäuser, Basel.

[125] Rényi, A. (1959). On Measures of Dependence. Acta Mathematica Hungarica 10, 441–451.

[126] Resnick, S.I. (1987). Extreme Values, Regular Variation and Point Processes. Springer, New York.

[127] Rolski, T., Schimidli, H., Schmidt, V. and Teugels, J.L. (1999). Stochastic Processes for Insurance and Finance. John Wiley and Sons, New York.

[128] Rüschendorf, L. (1976). Asymptotic Distributions of Multivariate Rank Order Statistics. Annals of Statistics 4, 912-923.

[129] Salvadori, G., De Michele, C., Kottegoda, N. and Rosso, R. (2007). Extremes in Nature: An Approach Using Copulas. Water Science and Technology Library, vol 56. Springer, Dordrecht.

[130] Scarsini, M. (1984). On Measures of Concordance. Stochastica 8, 201–218.

[131] Schmid, F., and Schmidt R. (2006). Bootstrapping Spearman's Multivariate Rho, in: A. Rizzi, M. Vichi (eds.), COMPSTAT-Proceedings in Computational Statistics, Physika-Verlag.

[132] Schmid, F., and Schmidt R. (2007a). Multivariate Extensions of Spearman's Rho and Related Statistics. Statistics and Probability Letters 77, 407–416.

[133] Schmid, F., and Schmidt R. (2007*b*). Multivariate Conditional Versions of Spearman's Rho and Related Measures of Tail Dependence. Journal of Multivariate Analysis 98, 1123-1140.

[134] Schmid, F., and Schmidt R. (2007*c*). Nonparametric Inference on Multivariate Versions of Blomqvist's Beta and Related Measures of Tail Dependence. Metrika 66, 323-354.

[135] Schmid F., Schmidt, R., Blumentritt, T., Gaißer, S. and Ruppert, M. (2010). Copula-Based Measures of Multivariate Association. In: P. Jaworski, F. Durante, W. Hardle, and T. Rychlik (Eds.), Copula theory and

its applications - Proceedings of the workshop held in Warsaw, Springer Verlag, 209–235.

[136] Schultze, J. and Steinebach, J. (1996). On Least Squares Estimates of an Exponential Tail Coefficient. Statistical Decisions 14, 353-372.

[137] Segers, J. (2005). Generalized Pickands Estimators for the Extreme Value Index. Journal of Statistical Planning and Inference 128, 381-396.

[138] Schweizer, B. and Sklar, A. (1983). Probabilistic Metric Spaces. North-Holland, New York.

[139] Schweizer, B. and Wolff, E.F. (1981). On Nonparametric Measures of Dependence for Random Variables. Annals of Statistics 9, 879-885.

[140] Shorack, G. and Wellner, J. (1986). Empirical Processes with Applications to Statistics. Wiley, New York.

[141] Sklar, A. (1959). Fonctions de Répartition à n Dimensions et leur Marges. Publications de l'Institut de Statistique de Paris 8, 229-231.

[142] Smith, R.L. (1985). Maximum Likelihood Estimation in a Class of Non Regular Cases. Biometrika 72, 67-90.

[143] Smith, R.L. (1987). Estimating Tails of Probability Distributions. Annals of Statistics 15, 1174-1207.

[144] Spearman, C. (1904). The Proof and Measurement of Association Between two Things. The American Journal of Psychiatry 15, 72-101.

[145] Spearman, C. (1906). Footrule for Measuring Correlation. The British Journal of Psychiatry 2, 89-108.

[146] Taylor, M.D. (2007). Multivariate Measures of Concordance. Annals of the Institute of Statistical Mathematics 59, 789-806.

[147] Tsukahara, H. (2005). Semiparametric Estimation in Copula Model. The Canadian Journal of Statistics 33, 357-375.

[148] Úbeda-Flores, M. (2005). Multivariate Versions of Blomqvist's Beta and Spearman's Footrule. Annals of the Institute of Statistical Mathematics 57, 781-788.

[149] van der Vaart, A.W. (1998). Asymptotic Statistics. Cambridge University Press, Cambridge.

[150] Vandewalle, B. and Beirlant, J. (2006). On Univariate Extreme Value Statistics and the Estimation of Reinsurance Premiums. Insurance: Mathematics and Economics 38, 441-459.

[151] Wang, S.S. (1996). Premium Calculation by Transforming the Layer Premium Density. ASTIN Bulletin 26, 71-92.

[152] Weissman, I. (1978). Estimation of Parameters and Large Quantiles Based on the k Largest Observations. Journal of the American Statistical Association 73, 812-815.

[153] Wirch, J.L., and Hardy, M.R. (1999). A Synthesis of Risk Measures for Capital Adequacy. Insurance: Mathematics and Economics 25, 337-347.

[154] Wolff, E.F. (1980). N-dimensional Measures of Dependence. Stochastica 4, 175-188.

[155] Yaari, M.E. (1987). The Dual Theory of Choice under Risk. Econometrica 55, 95-115.

[156] Young, V. (2004). Premium Principles. Encyclopedia of Actuarial Science, Wiley 4, 1322-1331.

Abstract

The aim of this thesis is to propose new estimators of copula-based measures of multivariate association and extreme risks. The estimation of the distortion risk premiums for heavy-tailed losses is proposed by Necir and Meraghni (2009). Their considerations are based on the Hill estimator (Hill, 1975) of extreme tail index and Weissman's estimator (Weissman, 1978) of the high quantile. It is well known, in the extreme value theory, that Hill's estimator exhibits an important bias which leads to an over/under estimation the aforementioned estimators of the distortion risk premiums. Several reduced biased estimations of the tail index are now available in the literature that solves this problem. In this thesis, we choose the kernel estimation method to derive a new estimator of the distortion risk premiums for large claims and establish its asymptotic normality. From the simulation study, it is clear that the newly estimator has a reduced bias, vis-à-vis to the existing ones, for any choice of the kernel function.

Résumé

L’objectif de cette thèse est de proposer de nouveaux estimateurs de mesures d’association multivariées et de risques extrêmes. L'estimation des primes de risque de distorsion pour les distributions à queue lourdes a été proposée par Necir et Meraghni (2009). Leurs considérations sont basées sur les estimateurs de Hill (Hill, 1975) de l'indice de queue et des quantiles extrêmes (Weissman, 1978). Il est bien connu, dans la théorie des valeurs extrêmes, que l'estimateur de Hill présente un biais important qui conduit à une sur/sous-estimation des estimateurs des primes de risque de distorsion. Plusieurs estimateurs à biais réduits de l'indice de queue sont maintenant disponibles dans la littérature qui permet de résoudre ce problème. Dans cette thèse, nous choisissons la méthode du noyau pour obtenir un nouvel estimateur des primes de risque de distorsion pour les grandes pertes et établir sa normalité asymptotique. Une simulation, montre que notre estimateur à biais réduit, vis-à-vis ceux qui existent déjà, pour tout choix du noyau.

ملخص

الهدف من هذه الأطروحة هو اقتراح مقدرات جديدة لقياسات الارتباط و الأخطار القصوى. التقدير الإحصائي لأقساط الخطر المشوهة للتوزيعات ذات الأذيال الثقيلة المقترح من طرف نصير و مرغني (2009) يعتمد على مقدر هيل (1975) و مقدر ويسمان (1978). من المعروف، في نظرية القيم القصوى، أن مقدر هيل هو مقدر متحيز. لهذا السبب تم اقتراح عدة مقدرات أقل تحيزا. في هذه الأطروحة ، نختار طريقة النواة للحصول على مقدر جديد لأقساط الخطر المشوهة وإثبات توزيعه المقارب. من خلال المحاكاة، أثبتنا أن مقدرنا الجديد أقل تحيزا من المقدرات الموجودة مهما كانت دالة النواة.