HAL open science

Contributions to statistical aspects of extreme value theory and risk assessment

Gilles Stupfler

▶ To cite this version:

Gilles Stupfler. Contributions to statistical aspects of extreme value theory and risk assessment. Statistics [math.ST]. Université de Rennes 1 (UR1), 2020. tel-03142415

HAL Id: tel-03142415
https://theses.hal.science/tel-03142415

Submitted on 16 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Rennes I

Ecole Doctorale Mathématiques et Sciences et Technologies de l'Information et de la Communication (MathSTIC)

Mémoire d'habilitation présenté par

Gilles Stupfler

en vue de l'obtention du diplôme d'Habilitation à Diriger des Recherches de l'Université de Rennes I

Contributions to statistical aspects of extreme value theory and risk assessment

Claudia Klüppelberg	Professeur, TU Munich	Rapporteur
Thomas Mikosch	Professeur, Université de Copenhague	Rapporteur
Olivier Wintenberger	Professeur, Sorbonne Université	Rapporteur

Soutenance le jeudi 3 septembre 2020 devant le jury composé de

Bernard Delyon	Professeur, Université de Rennes 1	Président
Clément Dombry	Professeur, Université de Franche-Comté	Examinateur
Gilles Durrieu	Professeur, Université de Bretagne Sud	Examinateur
Philippe Naveau	Directeur de recherche CNRS, LSCE	Examinateur
Claudia Klüppelberg	Professeur, TU Munich	Rapporteur
Thomas Mikosch	Professeur, Université de Copenhague	Rapporteur
Olivier Wintenberger	Professeur, Sorbonne Université	Rapporteur

Habilitation préparée au sein de l'Ecole Nationale de la Statistique et de l'Analyse de l'Information (ENSAI) et du Center for Research in Economics and Statistics (CREST)

Abstract

This report describes my research activities since I defended my PhD thesis. Chapter 1 outlines my research interests and summarises the background and main contributions of my research. These are organised into strands of work representing my research interests which can very roughly be classified into conditional extreme value analysis, multivariate extreme value theory and its offshoots, extremes in missing data contexts and risk assessment using extreme value theory. This last theme of research is currently my primary area of work. Chapter 2, which is the main chapter of this report, then expands further upon my contributions to this field, mostly through the introduction, study and estimation of risk measures at extreme levels, presenting and discussing the main results of my work in detail and providing some perspectives for future research.

Contents

Chapter 1

Introduction

This first chapter starts by a general description of my research interests and activities since I obtained my PhD, before describing in more detail my main directions of work and the results I obtained. The final section of this chapter contains an outline of the organisation of the remainder of the manuscript.

1.1 An overview of my research activities and collaborative network

The unifying thread of the vast majority of my research work is the statistical analysis of extreme values. An early but excellent example of extreme value problem was set by the Dutch government in the 20th century: given 100 years of flood data, can one determine how high dykes along the coastline should be so that they protect the country against a once-in-10,000-years flood? This problem is crucial to, among others, urban risk planners and insurers, but at first sight, it appears to have no solution. Indeed, if no specific assumption is made about the underlying distribution, one needs much more than 10,000 years (let alone 100 years) of data to carry out inference about a once-in-10,000-years event. This is where extreme value theory comes into play: under a very mild modelling assumption, it can be shown that the right tail of the underlying distribution (here, of water height) can be approximated by a Generalised Pareto distribution. The parametric form of this approximation then allows for an extrapolated estimation of the once-in-10,000-years return level. Modern developments in extreme value theory focus not only on the estimation of extreme quantiles, but also on, among others, the estimation of extreme regions in the multivariate case, the development of extreme value procedures accommodating temporal dependence and the introduction of techniques to deal with non-stationarity.

I got interested in the statistical aspects of extreme value theory and risk assessment when studying for my PhD, which I defended in November 2011 at the University of Strasbourg. My PhD work focused on two distinct topics: (i) the use of hidden Markov models for the description of, and inference about, loss processes in insurance and (ii) the development of extreme value-based procedures for the estimation of a right endpoint, or of a frontier when a conditional endpoint is of interest instead. This body of work directly or indirectly resulted in the six publications [26, 27, 28, 29, 30, 31].

After my PhD, I first worked in the broader context of conditional extreme value analysis, where my main interest was to develop, and analyse the theoretical properties of, conditional extreme value index estimators in nonparametric contexts [9, 21, 24, 25]. Working on asymptotic properties of nonparametric estimators led me to investigate potential convergence results in functional spaces; it became apparent to me that for kernel estimators, such as the one I worked on in [21], this was not a straightforward task. This drove me to work on negative results for the functional convergence of the Parzen-Rosenblatt kernel density estimator in [17, 23]. In parallel, I became interested in multivariate data problems where the interest is in getting a properly multivariate, rather than conditional, understanding of extremes, which led me to start working on the analysis of extremes of a multivariate distribution [15, 19]. This then drove me to learn more about concepts of multivariate extreme value theory; one of my recent lines of work has been to investigate the construction of characteristic functions as an offshoot of such concepts [2, 6, 16]. I had already experimented with this kind of thinking in [20] where a procedure for transformation to distributional symmetry is developed based on the so-called probability weighted empirical characteristic function. Aside from my interest in conditional and multivariate extreme value analysis, one of the more recent orientations of my research has been towards the statistical analysis of extremes in missing data frameworks such as random censoring and truncation [8, 18, 22]. Another of my current efforts is to develop the applied side of my research, in the sense of using existing tools (rather than developing new methods) to solve real data problems [7, 13].

What has, however, arguably constituted my most important strand of work in the last few years is the development and use of extreme value tools for risk assessment [1, 3, 4, 5, 10, 11, 12, 14]. This has stemmed from the increasing understanding, not only within the academic world but also from industrial and regulatory perspectives, that relying solely on quantiles for risk assessment can be an issue in practical setups, not least because the quantile risk measure only depends on the frequency of the tail event and not on the actual values of the risk variable on such an event. Motivated by a number of real data applications in insurance and finance, my research activities in this subfield of extreme value analysis have focused on the design, study, and estimation of extreme risk measures that carry information both about the frequency of tail losses and their values. The distributional framework I have been working in is the family of heavy-tailed distributions, which have been shown to describe the extreme value behaviour of most financial and actuarial data quite well. Three main theoretical tools that have formed the backbone of my research within this context are asymptotic Gaussian approximations of the tail empirical quantile process, convergence theorems for M-estimators defined by minimising convex criteria, and semiparametric constructions of Weissman-type extrapolated estimators. Based on these tools, as well as on simulated and real data illustrations, my work has in particular provided evidence that the use of estimators constructed by minimising asymmetrically weighted L^p criteria has the potential to become part of the standard risk management toolbox.

My research activities have given me the opportunity to work on an array of real data problems coming from insurance, finance, reliability analysis, medicine and law with an international network of colleagues from France, the UK, Germany, Belgium, Denmark, Greece and Canada. I have also recently started forging research collaborations with colleagues from Italy, Switzerland and Japan. Some of my latest published and forthcoming work has originated from my supervision activities: for instance [7] follows the undergraduate research internship of Oliver Church whom I supervised during the 2017/2018 summer term at the University of Nottingham. The submitted article [35] is written with (among others) my PhD student Emily Mitchell, also at the University of Nottingham, while [32, 33] have been written with my post-doctoral research assistant Antoine Usseglio-Carleve, whose 2-year contract at INRIA Grenoble Rhône-Alpes started in October 2018. I have, finally, started to supervise the PhD thesis of Abdul Haris Jameel at the University of Nottingham in October 2019 (on a 4-year project on the applied analysis of medical data) and I have been informally supervising the PhD dissertation of Hibiki Kaibuchi (The Graduate University for Advanced Studies, SOKENDAI, Japan), a former undergraduate student of mine at the University of Nottingham, since early 2019.

1.2 Main strands of work and results

This section gives more information about the main results I have obtained so far in my directions of research as part of my work post-PhD. For each direction, I first give an idea of the context in which my work was carried out, before summarising the results linked to each publication.

1.2.1 Conditional extreme value analysis

Background Understanding the extremes of a random phenomenon is a major question in various areas of statistical application. In addition to the prediction of the intensity of extreme floods, further climate-related examples of extreme value problems are the estimation of extreme daily wind speeds [50] and the estimation of extreme rainfall at a given location [187]. Besides, extreme phenomena may have strong adverse effects on financial institutions or insurance companies, and the investigation of those effects has received good attention in the extreme value literature, see for instance [221]. Such problems typically ask for the estimation of an extreme quantile of a univariate random variable Y (amount of rainfall, wind speed, loss size) based on a small sample of data.

In practical applications, the variable of interest Y can often be linked to a covariate X. In the aforementioned applications linked to climate science, geographical location is certainly an important covariate; in actuarial science, the claim size depends on the sum insured by the policy. In such situations, the extreme value index and quantiles of the random variable Y given $X = x$ are functions of x which are referred to as the conditional extreme value index and conditional quantile functions. Their estimation was at first mostly considered in the "fixed design" case, namely when the covariates are non-random [71, 88, 89, 128, 129, 148, 234]. By contrast and despite the obvious interest in practice, the study of the random covariate case had been initiated only shortly before I started working on the topic of conditional extremes in early 2013. As far as I am aware, available methods were [85] which used a fixed number of nonparametric conditional quantile estimators to estimate the conditional tail index, later generalised in [84] to a regression context with conditional response distributions belonging to the general max-domain of attraction, [130] who introduced a local generalised Pickands-type estimator and [254], based on a maximum likelihood approach in a conditional heavy-tailed context. A different approach had been developed in a quantile regression model: the pioneering paper is [74], subsequently built upon in [252, 253].

My contribution to this area The state of the art at the time when I wrote [25] was that, in the random covariate case, there was only a single estimator of the conditional extreme value index which was not restricted to the Fréchet max-domain of attraction (that is, the space of conditional heavy-tailed distributions). This estimator, introduced in [84], is a local version of a Pickands-type estimator. As such, it relies on a finite number of high order statistics and is thus wasteful of relevant information about the tail, resulting in an estimator having a large variance. In [25], I develop and study a local adaptation of the moment estimator of [92], whose computation relies on an increasing number of high order statistics as the sample size increases, and therefore intuitively makes a more efficient use of the available data. The pointwise consistency and asymptotic normality of the estimators are established. These results are obtained under conditional adaptations of traditional extreme value conditions plus a condition on the local uniform oscillation of the conditional tail quantile function. This latter condition is extensively analysed to reveal its connections with more intuitive assumptions on the regularity of the functions involved and the bandwidth sequence used to compute the estimator. This discussion is in turn used to find the optimal rates of pointwise convergence of the estimator. The proposed methodology is compared to the local Pickands-type estimator of [84] on simulated data and is showcased on a real set of fire insurance data. It is in particular seen that the local moment estimator typically has a lower Mean Squared Error than the local Pickands-type estimator, illustrating the benefit of taking a growing number of high order statistics into account in the calculation.

I observed during my reading of the literature and when writing [25] that the choice of the number of top order statistics $k = k_x$ to be used in local versions of conditional extreme value index estimators, such as the local moment estimator, was an even more difficult problem than in unconditional setups. This is mainly due to the fact that such estimators are constructed at each point x by (i) selecting observations close to x, and then (ii) having to keep only some of the largest observations in this smaller sample. Effective sample sizes in conditional extreme value analysis are therefore typically much lower than in univariate extreme value analysis. Standard selection rules such as those based on finding a "stability region" (formalised in e.g. [106, 123]) are, at those lower sample sizes and with observations whose distribution is only "close" to the target distribution, not guaranteed to perform well. In [24], we tackle this problem by studying an averaged-out version of the local Hill estimator [153] for conditional heavy-tailed distributions; a similar idea in the unconditional case is developed by [216]. The idea is that averaging out local Hill estimators, with respect to $k = k_x$, will produce smoother sample paths and thus will make the selection of k_x easier. Uniform (in x) consistency in probability and pointwise asymptotic normality are established; the uniform consistency result was, to the best of my knowledge, the first one of its kind to be shown for a conditional extreme value index estimator. The conditions of this result are relatively weak: in particular, the only conditions linked to the presence of the covariate are the continuity of the tail index function and Hölder continuity of the density function of the covariate, plus an assumption on the local uniform oscillation of the quantile function. We also construct a selection rule for the choice of k_x. The estimator is compared to several competitors in a finite-sample study, including to a localised version of the standard Hill estimator. I later carried on developing the theoretical side of conditional extreme value analysis by working on the rate of strong uniform consistency of a different, kernel version of the Hill estimator in [21]. The main idea of the proof dates back to [150] and is based on showing strong uniform consistency over a well-selected grid while controlling the oscillation of the estimator.

While investigating the uniform consistency properties of local or kernel versions of extreme value index estimators, I noted that working out the functional convergence of such estimators, and indeed of kernel estimators in general, appeared to be a very difficult question. In the case of functional convergence in a space of continuous functions, the intuition is that while finite-dimensional convergence of kernel estimators is easy to establish using the Cramér-Wold device, the components of the limiting distribution are in fact asymptotically independent. The candidate for a functional limit is thus white noise, which is not an admissible limit since it does not define a continuous process. It is in fact possible to extend these arguments to other functional spaces: it is shown in [211, 222] that the suitably rescaled Parzen-Rosenblatt kernel density estimator does not converge to a Borel measurable limit in the L^2 space on $\mathbb{R}^d$ endowed with its standard topology. In [23], I extended this negative result to weighted L^p spaces, with $p \geq 1$ and finite, and in [17] I tackled the case of the L^∞ space; the latter is more delicate than the former, since L^∞ spaces on $\mathbb{R}^d$ are not separable, but one can still show that a nontrivial functional convergence result for the kernel density estimator in such spaces cannot be written. It is worth noting that those negative results do not contradict the earlier, numerous available results on the uniform or quadratic consistency of kernel estimators, shown among others in [58, 99, 112, 113, 132, 133, 147, 232, 236, 237]. They merely show that, as far as kernel estimators are concerned, rates of uniform or quadratic consistency cannot be obtained as corollary of general functional convergence results in the spirit of those obtained in e.g. [247, 248], and thus that for these estimators, results on uniform or quadratic deviations can only be obtained using adapted tools.

Recently, in [9], I worked on the estimation of the conditional extreme value index (and conditional extreme quantiles) for conditional heavy-tailed distributions when the covariate belongs to a Polish metric space but is otherwise possibly functional. The class of conditional tail index estimators developed therein, coined *integrated functional estimators*, is based on calculating a weighted integral of log-spacings of a local version of the empirical quantile function. We explain how this construction allows the class of estimators to encompass several local versions of well-known semiparametric estimators, such as the Hill and Zipf estimators (for the latter, see e.g. [189, 225]). The pointwise asymptotic distributions of our estimators are derived and the estimators are showcased on a real set of chemometric data linking percentage of fat content in a piece of meat to its absorbance curve.

1.2.2 Multivariate extreme value analysis

Background There are many real data applications in which the problem of interest is inherently multivariate, rather than conditional. If a data set containing the amount of rainfall at several sites across a given region is available, then an interesting question from the extreme value point of view is to estimate the probability of rainfall reaching an extreme level across two or more sites. Similarly, when considering data on the losses of an insurance company across several lines of business, an important question is to estimate the probability of extreme losses across two or more of these lines of business, since these are likely to represent a threat to the company's survival.

With the question of analysing multivariate extremes comes the question of what, exactly, means being extreme in several dimensions. One way to approach this question is to forget about the extreme value context at first, and work on an acceptable definition of a quantile in the multidimensional world; once this is done it becomes possible to declare that an extreme observation in a multivariate setup is one "exceeding" a multivariate extreme quantile in a certain sense. The search for what constitutes a good definition of quantile in $\mathbb{R}^d$ ($d \geq 2$) has been a very fruitful avenue of work; a review of the broad families of available techniques is [228]. My research in the area of multivariate extremes has so far focused on multivariate geometric quantiles, introduced by [70]. Such geometric quantiles are obtained by the minimisation of a loss function, constructed in a way that extends to the multivariate case the well-known absolute loss minimisation framework for univariate quantiles [181, 182]. Geometric quantiles are indexed by a vector $u \in B^d$, where B^d is the unit open Euclidean ball of $\mathbb{R}^d$, and are vectors themselves. As such, a geometric quantile $q(u) \in \mathbb{R}^d$ possesses a direction and a norm. Geometric quantiles are in fact special cases of $M-$quantiles in the sense of [63, 184]. They also benefit from various desirable properties: they are equivariant under any orthogonal transformation [70], characterise the underlying distribution [184] and generalise the well-known L^2-geometric median [233].

My contribution to this area The aforementioned papers almost completely focus on central properties of geometric quantiles and their sample versions. [70] though labels geometric quantiles as "extreme" when $\|u\|$ is close to 1, and later [69] used such extreme geometric quantiles in an application to multivariate outlier detection, but no specific analysis of extreme geometric quantiles had been done to justify how or indeed why they should be used in such a context. In [15, 19] we provide what is to the best of my knowledge the first in-depth study of the behaviour of geometric quantiles as $\|u\| \uparrow 1$. We provide first- and second-order asymptotics for both their direction and norm. In [15], we find in particular that the norm of the geometric quantiles of a given distribution always tends to infinity as $\|u\| \uparrow 1$, even if the distribution has a compact support, and that the norm of the extreme geometric quantiles of a random vector X having a finite covariance matrix grows at a fixed rate and is equivalent to a quantity uniquely determined by this covariance matrix. We also show that, at the sample level, extreme geometric quantile contours generally give a poor idea of the shape of the data cloud: for instance, for elliptically contoured distributions, extreme geometric quantile contour plots and density level plots are in some sense orthogonal. We showed in [19] that, somewhat surprisingly, geometric extreme quantiles of a random vector X having an infinite second moment do contain some information about the multivariate extremes of X, under an assumption of multivariate regular variation introduced in [67] on the (assumed to exist) probability density function of X. This distinction between finite and infinite second moment is arguably unnatural and unwieldy though, and we conclude that, even though central geometric quantiles are certainly an important and valuable notion in multivariate statistics, practitioners should be very cautious when using extreme geometric quantiles for outlier detection or multivariate extreme value analysis.

1.2.3 Characteristic functions as offshoots of multivariate extreme value concepts

Background Besides the need for a sensible definition of what it means to be extreme in several dimensions, there is also the necessity for the construction of a sound theoretical framework for the study of multivariate extremes. This question has been investigated extensively. One particularly interesting framework is obtained by extending the univariate definition of extreme value distributions, found from the convergence of linearly normalised sample maxima, to the multivariate case via the use of the componentwise maximum operator. This gives rise to the class of multivariate max-stable distribution functions. Much is known about such distributions: although they cannot be be written in a simple parametric form (unlike in the univariate case), they can be represented in an exponential form using the concept of angular measure. A theory of multivariate regular variation, at the heart of which the angular measure lies, can also be written. This constitutes a reasonable framework for inference about extreme events in multidimensional settings. Standard references are [118, 215].

Although this is by no means obvious, a common thread through multivariate extreme value theory can be found using the notion of $D-$norm [118]: loosely, there is a simple, explicit one-to-one correspondence between max-stable distributions with negative exponential margins and the set of $D-$norms. When I learnt about this notion in late 2015 at the Winter Course (taught by Michael Falk) preceding the CMStatistics 2015 conference, $D-$norms had mainly been studied for their potential to unify the presentation of multivariate extreme value theory and trivialise certain proofs, such as those of Takahashi's characterisations of complete dependence or independence of the components of a max-stable distribution [239, 240]. As a mathematical object themselves, however, $D-$norms are interesting as well, and their theory remains in full development.

My contribution to this area A $D-$norm is constructed on a generator, which is a $d-$dimensional nonnegative random vector $(Z_1, \ldots, Z_d)$ whose components all have expectation 1. Mathematically, a $D-$norm with generator $(Z_1, \ldots, Z_d)$ is the mapping $(x_1, \ldots, x_d) \mapsto \mathbb{E}(\max\{|x_1|Z_1, \ldots, |x_d|Z_d\})$. It is known that any $D-$norm has infinitely many different generators, such as those obtained by multiplying all of the Z_i by an independent nonnegative random variable X having expectation 1. It is remarkable though that the knowledge of the slightly different $D-$norm $(x_0, x_1, \ldots, x_d) \mapsto \mathbb{E}(\max\{|x_0|, |x_1|Z_1, \ldots, |x_d|Z_d\})$ fully identifies the distribution of $(Z_1, \ldots, Z_d)$. By 1-homogeneity, knowing this function for $x_0 = 1$ is in fact enough to characterise the distribution of $(Z_1, \ldots, Z_d)$, which motivated us to introduce in [16] the *max-characteristic function*

$$(x_1, \ldots, x_d) \mapsto \mathbb{E}(\max\{1, x_1Z_1, \ldots, x_dZ_d\}), \quad x_1, \ldots, x_d > 0,$$

for any nonnegative random vector $(Z_1, \ldots, Z_d)$ having integrable components with nonzero expectations. We show in [16] (see also Section 5.1 in the monograph [117]) that this function indeed characterises the joint distribution of any such random vector. Unlike the standard Fourier transform-based characteristic function, it has a simple closed form for generalised Pareto distributions, which are of prime importance in extreme value theory. We then prove that the convergence of a sequence of max-characteristic functions to a max-characteristic function is equivalent to convergence of the underlying distributions in the Wasserstein metric; in other words, convergence of max-characteristic functions is equivalent to convergence in distribution plus convergence of the sequence of first moments. We also provide some insight into the structure of the set of max-characteristic functions, and we state and prove an inversion formula making it possible to retrieve a distribution from its max-characteristic function. We then noted that taking $x_0 = 1$ above was indeed convenient but destroyed the norm structure. In [6], we focus directly on the mapping

$$(x_0, x_1, \ldots, x_d) \mapsto \mathbb{E}(\max\{|x_0|, |x_1|Z_1, \ldots, |x_d|Z_d\}), \quad x_1, \ldots, x_d \in \mathbb{R},$$

again for any nonnegative random vector $(Z_1, \ldots, Z_d)$ having integrable components with nonzero expectations. This norm, which we call *$F-$norm*, extends the max-characteristic function and therefore characterises the distribution of $(Z_1, \ldots, Z_d)$; we study it as a mathematical object. We provide examples of such $F-$norms and a complete characterisation of $F-$norms when $d = 2$. We also study the geometry of $F-$norms, their estimation based on a sample from $(Z_1, \ldots, Z_d)$, their links with convergence in the Wasserstein and Hausdorff metrics, as well as their algebra and an extension to general random vectors (not necessarily assumed to be nonnegative).

The concept of max-characteristic function applies to nonnegative random vectors having integrable components. The integrability requirement is a substantial restriction, and prevents one from using this max-characteristic function for very heavy-tailed generalised Pareto distributions, for instance, although we know from [16] that lighter-tailed generalised Pareto distributions have a simple max-characteristic function. This appears to be a rather unnatural limitation. In [2], we start by noting that, like linear projections [79], the *max-linear projections* $\{\max(t_1X_1, \ldots, t_dX_d),\ t_1, \ldots, t_d > 0\}$ characterise the distribution of any random vector $(X_1, \ldots, X_d)$ with nonnegative components. We then remark that, in dimension 2, $\min(t_1X_1, t_2X_2) = t_1X_1 + t_2X_2 - \max(t_1X_1, t_2X_2)$, which may suggest that a multivariate distribution with nonnegative components is also characterised by its *min-linear projections* $\{\min(t_1X_1, \ldots, t_dX_d),\ t_1, \ldots, t_d > 0\}$. We prove that this is indeed the case, as long as the X_i are strictly positive with probability 1, and we then study the companion notion of *min-characteristic function*

$$(x_1, \ldots, x_d) \mapsto \mathbb{E}(\max\{1, x_1X_1, \ldots, x_dX_d\}), \quad x_1, \ldots, x_d > 0,$$

for any random vector $(X_1, \ldots, X_d)$ having strictly positive components. We show that the distribution of such a random vector is indeed determined by its associated min-characteristic function, which turns out to be a continuous and concave distribution function itself. We give a number of examples of calculations of the min-characteristic function (including in the generalised Pareto case, where it has a simple explicit form) and we prove that the convergence of a sequence of min-characteristic functions to a min-characteristic function is equivalent to weak convergence of the underlying distributions. We discuss the structure of the set of min-characteristic functions, their estimation, as well as an extension to general random vectors and a couple of applications to $D-$norm theory.

1.2.4 Extremes in missing data frameworks

Background The problem of missing data is frequently encountered in certain fields of statistical applications. A famous example of missing data framework is random right-censoring, found most notably in medicine in the study of the survival times of patients to a given chronic disease in follow-up studies. In mathematical terms, the information then available to the practitioner is the pair (Z, δ), where Z is the minimum between the survival time Y and censoring time T, and δ is the 0-1 variable equal to 1 if and only if the survival time is actually observed. Random right-censoring is also found in, among others, reliability data analysis [204] and non-life insurance [217]. It should not be confused with other types of missing data mechanisms such as right-truncation, where one observes the variable of interest Y if and only if it is less than or equal to the truncation variable T, in which case the full pair (Y, T) is actually observed, with nothing being recorded otherwise. Truncated data may be collected in various cases, for instance when estimating incubation times for a given disease [171, 172, 192], in physics [166, 197], when accounting for reporting lags in insurance data [152, 173, 194] or when dealing with failure and warranty data [161, 162, 202].

In the aforementioned examples, the estimation of extreme parameters of the underlying distribution of the variable of interest Y would ultimately lead to the analysis of survival times of exceptionally strong/weak patients to a given disease, extreme losses/payouts in insurance, or failure times for highly resistant/unreliable devices. When I started working on this topic in 2013, extreme value analysis with missing data was a recent but rapidly developing field. The censored case had been tackled using an adaptation of the Peaks-Over-Threshold framework in [48, 49], the construction of corrected versions of standard extreme value estimators in [110, 139], Huber-type robust estimators [223] and, for heavy-tailed variables, the interpretation of the extreme value index as an average of log-excesses [260]. An adaptation of [110] had also been developed in the conditional heavy-tailed case by [205]. By contrast, the truncated case had only been considered by [37] when the right-truncation point is unknown but non-random.

My contribution to this area I first worked within the framework of random right-truncation. In this situation, the available data is made of independent copies of (Y^*, T^*), where (Y^*, T^*) has the distribution of (Y, T) given that $\{Y \leq T\}$. In the case of a heavy-tailed variable of interest Y truncated by a heavy-tailed T, we provide in [22] the first examples of tail index and extreme quantile estimators. The fundamental idea behind the construction of the extreme quantile estimator is to, based on the heavy-tailed assumption, extrapolate an empirical estimator of a quantile from a so-called intermediate level where it is consistent, using an estimator of the tail index of Y. This follows the classical ideas of [256] for fully observed data. The intermediate quantile estimator is built using an identity showing that the cumulative hazard function can be directly calculated in terms of the marginal distribution functions of the observed pair (Y^*, T^*) [259]. Its consistency and asymptotic normality are shown using analytical results warranted by the heavy-tailed structure and a martingale argument inspired by [238]. The tail index estimator, at the heart of the extrapolation procedure, is obtained after noting that Y^* and T^* are also heavy-tailed with tail indices that are simple functions of those of Y and T. A suitable combination of the two marginal Hill estimators of Y^* and T^* thus provides a consistent estimator of the tail index of Y. The consistency and rate of convergence of our extreme quantile estimator are then established. The finite-sample behaviour of our extreme quantile estimator is examined on a real set of reliability data.

In parallel, I noted that the literature on conditional extreme value analysis with random right-censoring was, at the time I started working in this research area, restricted to [205] and thus limited to the conditional heavy-tailed case. In the context of their real data application to the estimation of survival time to AIDS given age at diagnosis, this seems like a rather strong assumption: there is indeed no guarantee a priori that the distribution of survival time conditional to age has a heavy tail. In [18], I introduce a local version of the moment estimator of the extreme value index adapted to the random right-censoring context provided by [110]. Pointwise consistency, asymptotic normality and the optimal rate of convergence of the estimator are discussed; the proofs combine ideas of [110], which shows the asymptotic normality of a moment estimator adapted to random censoring in the unconditional case, with technical tools from [25] to deal with the conditional context. The estimator is used, along with a local extreme quantile estimator, to revisit the AIDS data set of [110, 205]. It is observed that extreme quantiles of the survival time, which can be interpreted as survival times of very resistant patients, are fairly stable for patients aged between 20 and 53 years and decrease sharply afterwards. This may be interpreted as a consequence of immunosenescence, namely the deterioration of the immune system as age increases. This phenomenon is of course especially critical in the case of AIDS, since HIV targets cells of the immune system; the significant effect of increasing age on survival rates for AIDS has been shown numerous times in the medical literature [45, 87, 196, 219]. It is also apparent that there is indeed heterogeneity in tail behaviour, and in particular it is not clear that the distribution of survival time for older patients has a heavy right tail.

One common feature between my work in [18] and the earlier paper [205] is the assumption of conditional independence between survival time and censoring time. More generally, all the aforementioned papers on random right-censoring work under an independent censoring assumption. This kind of assumption is arguably standard in the context of random right-censoring since the pioneering paper [174] on the product-limit estimator for the survival function. And yet, cases in which there are strong suspicions of dependence between the variable of interest and the censoring time have been reported several times over the last decades. In medical studies especially, a common cause of the probable violation of the independence hypothesis is a sizeable number of patient dropouts [164, 167]. Crucially, using traditional estimators relying on the independence assumption when there is in fact dependence may lead to invalid inferences [108, 178] and even identifiability issues [243]. In [8], using the example of an extreme value copula dependence structure [91, 126, 144] between Y and T, I prove that in such a situation (i) the right-tail behaviour of the observed time Z is essentially that of the variable with the lightest tail in the pair (Y, T) and (ii) the tail censoring probability, which describes how often the extremes of Y are censored, is equal to 1 when the tail of Y is heavier than that of T. This implies that in such a setup, the estimators of [48, 110, 139] are not consistent. I then argue, using further asymptotic considerations on the behaviour of the distribution of the pair (Z, δ) for Z large, that the identifiability of the extreme value index of Y is more generally unclear. This is in stark contrast with the independent censoring case, in which we know from [110] that the problem of inferring this parameter can indeed be solved in a simple way based on the behaviour of the pair (Z, δ) for large Z alone. I also explain why, if the extreme value copula describing the censoring mechanism is assumed to be known, then the extreme value index of Y becomes clearly identifiable, and I outline a couple of possible strategies that may lead to consistent estimators of this extreme value index in future work.

1.2.5 Extreme value theory and risk assessment

Background The class of quantiles is one of the basic tools in risk management. A leading quantile-based risk measure in banking and other financial applications is Value-at-Risk with a confidence level $\tau \in (0, 1)$. It is informally defined as the τth quantile $q(\tau)$ of a non-negative loss distribution, with τ being close to 1, and as $-q(\tau)$ for a real-valued profit-loss distribution, with τ being close to 0. As such, the Value-at-Risk is very easy to interpret. However, a single quantile $q(\tau)$ of a random variable X is only determined by the frequency of the relevant tail event and not by the actual values that X takes on this event; to put it differently, one can readily construct examples of distributions sharing a common quantile but whose actual tail behaviours beyond that quantile are very different. A single Value-at-Risk thus does not and cannot provide a complete picture of tail risk. Furthermore and from the axiomatic point of view, quantiles do not, in general, define coherent risk measures in the sense of [44], because they lack the important subadditivity property, meaning that they do not abide by the intuitive diversification principle.

There has been a substantial effort in the literature towards the definition and study of alternative families of risk measures. My work in this area, which I started in 2016, has mostly focused on the analysis and estimation, at extreme levels, of risk measures either built on weighted integrals of the quantile function or generated from L^p-minimisation. From an inferential point of view, the attention devoted to those risk measures had so far been mostly restricted to ordinary, central risk measures staying away from the tails of the underlying distribution. Notable exceptions are [114] and [249], which respectively studied the estimation of the extreme so-called Conditional Tail Moments in a regression setup and certain extreme Wang distortion risk measures [250] for pricing of reinsurance premiums; Wang distortion risk measures are integrals of the quantile function, which are weighted according to the distortion function, yielding a coherent risk measure when the distortion function is concave. My work was carried out in the context of heavy-tailed distributions, which has been found to describe the tail structure of financial and actuarial data fairly well, see e.g. [115, 215].

My contribution to this area My first contribution in [14] was to provide a couple of methods for the inference of extreme Wang distortion risk measures (DRMs) of a random variable X. In this context, "extreme" means that the focus is on a *fixed* Wang DRM of the variable X given that $X > q(\tau)$, with τ close to 1. The estimators rely on either an asymptotic proportionality relationship linking extreme Wang DRMs with tail quantiles, or a functional plug-in of the tail empirical quantile function, plus an extrapolation method in the spirit of [256]. The asymptotic distribution of each estimator is obtained; in the case of the functional plug-in estimator, the proof uses a Gaussian approximation of the tail empirical quantile function to control the gap between the estimator and the Wang DRM to be estimated. Finite-sample studies, on simulated and real financial and actuarial data, make it possible to assess the behaviour of the estimators in practice. One of the conclusions of the finite-sample study is that the performance of both estimators gets worse as the value of the tail index increases. In the case of the functional plug-in estimator, we argue in [11] that this is a natural consequence of how it is constructed: since this estimator is in general a linear combination of (a power of) the data above a high threshold in the sample, the propensity of heavier-tailed distributions to generate highly variable top order statistics will tend to adversely affect the quality of the estimates. This is illustrated numerically on a simple example in [11], after which we introduce trimmed/Winsorised versions of our functional plug-in estimator in a effort to solve this issue. The methodology is based on (i) trimming/Winsorising the estimator to reduce its variability and (ii) introducing a correction factor to deal with the negative bias thus introduced. The consistency and asymptotic normality of our estimators are shown. We compare these modified versions to the original estimator on simulated and real data, and we highlight to which extent our robustified estimators perform better than the functional plug-in estimator.

It is remarkable that the class of Wang DRMs encompasses the class of quantiles. Indeed, the quantile $q(\tau)$ is exactly obtained by taking the distortion function $g : [0, 1] \to \mathbb{R}$ equal to 0 before $1-\tau$ and 1 otherwise. An entirely different way of looking at quantiles is via the well-known absolute loss minimisation framework for univariate quantiles [181, 182]. Substituting a squared loss for the absolute loss results in the class of expectiles, introduced in [209]; using a more general L^p-loss produces the class of L^p-quantiles introduced in [73]. Quantiles, expectiles and L^p-quantiles more generally are in fact M-quantiles in the sense of [63]. Expectiles in particular have been receiving substantial attention as risk measures in statistical finance and actuarial science since the pioneering paper [190]. They depend on both the tail realisations and their probability, and thus allow to measure extreme risk based on both the frequency of tail losses and their severity. Most importantly, expectiles, for $\tau > 1/2$, induce the only coherent risk measure that is also elicitable, the latter meaning that the quality of expectile forecasts can be evaluated in a straightforward fashion [134, 265]. However, while the estimation of expectiles has been extensively tackled at central levels [157, 188], an inference theory for extreme expectiles remains to be developed. In [12], we study two estimation methods, either relying on an asymptotic proportionality relationship with quantiles or a direct, empirical asymmetric least squares procedure. The limiting distributions of the estimators are established; for the latter estimator, the proof relies on theoretical tools developed by [180] linking the asymptotic behaviour of minimisers of convex random functions to that of this sequence of functions itself. We then use our analysis of tail expectiles to define and study an expectile-based analogue of the Marginal Expected Shortfall studied by [42, 64, 116] and by [68] in an extreme value context specifically. We finally discuss the choice of the expectile level to consider, and examine the behaviour of our estimators on real actuarial and financial data. In [1], we take the idea of estimating tail expectiles one step further by deriving weighted Gaussian approximations of the tail empirical expectile process. This is substantially harder than showing the corresponding results for the tail empirical quantile process (see e.g. [81, 101, 111]) because empirical expectiles only have an implicit formulation. We state and prove an approximation of the tail empirical expectile process by a Gaussian process with drift and we investigate its joint asymptotic behaviour with the tail empirical quantile process. These results are used to introduce and study the asymptotic distributions of new estimators of, among others, extreme expectiles and a coherent expectile-based version of the Expected Shortfall, which are then showcased on real actuarial and financial data as well.

Our work in [10] focuses on the general class of L^p-quantiles. Their existence requires $\mathbb{E}|X|^{p-1} < \infty$, which, when $p < 2$, is a weaker condition than the assumption $\mathbb{E}|X| < \infty$ required for the existence of expectiles. When $p > 1$, L^p-quantiles also take into account the whole tail information about the underlying distribution, unlike quantiles. They do thus steer an advantageous middle course between quantiles ($p = 1$) and expectiles ($p = 2$). We analyse therein the asymptotics of population L^p-quantiles, as their level τ tends to 1, before constructing an estimator of extreme L^p-quantiles based on asymmetrically weighted L^p-minimisation and a competitor using the asymptotic proportionality relationship between L^p-quantiles and standard quantiles. Their asymptotic distributions are investigated in a suitable context of strictly stationary and mixing observations: it is seen in particular that the estimator based on asymmetrically weighted L^p-minimisation is asymptotically normal essentially as soon as $\mathbb{E}|X|^{2(p-1)} < \infty$, which when $1 < p < 2$ is less restrictive than the corresponding condition $\mathbb{E}|X|^2 < \infty$ needed for tail expectile estimation. In other words, when $p < 2$, extreme L^p-quantile estimators can be used in a larger class of heavy-tailed distributions. However, and even though L^p-quantiles with $1 < p < 2$ are more widely applicable than expectiles and take into account the whole tail information, they are neither easy to interpret nor coherent as risk measures. This drove us to work on procedures, based on the asymptotic connection between quantiles, expectiles and L^p-quantiles, making it possible to use L^p-quantiles as vehicles for extreme quantile and expectile estimation. The methodology and its potential for better forecasts of traditional extreme quantiles are illustrated on financial data from the S&P500 index.

In [3] I used this experience of working on L^p-quantiles to construct and study a different notion of risk measure based on L^p-optimisation, whose motivation can be summarised as follows. When the underlying distribution function of X is continuous and has a finite first moment, the Conditional Tail Expectation $\mathrm{CTE}(\tau) := \mathbb{E}[X \,|\, X > q(\tau)]$ defines, unlike the quantile, a coherent risk measure. Besides, since the expectile of a random variable Y is its expectation, the quantity $\mathrm{CTE}(\tau)$ is nothing but the expectile of X given that $X > q(\tau)$. As such, the asymptotic normality of its empirical estimator requires a finite second moment; this can potentially be a strong integrability restriction. Since L^p-quantiles, for $1 < p < 2$, are more widely applicable than expectiles, this led us to study the L^p-quantiles of X given that $X > q(\tau)$, which we call the class of *tail L^p-medians* of X. This class of risk measures encompasses the Conditional Tail Expectation, for $p = 2$, but also the important Median Shortfall [185, 186]. Unlike L^p-quantiles though, tail L^p-medians focus solely on the tail event, and this makes their theoretical analysis substantially different. We study the asymptotics of population tail L^p-medians, as their level τ tends to 1, and we show that they can asymptotically be written as a weighted average (with coefficients depending on p) of the Median Shortfall and the Conditional Tail Expectation. This in turn is instrumental in the development of the interpretation of a tail L^p-median, and can be used to design a rule for the choice of p when a pre-determined compromise between Median Shortfall and the Conditional Tail Expectation is set. We then construct two estimators of tail L^p-medians above extreme levels, whose asymptotic behaviour we study. It is, similarly to the L^p-quantile technique, found that tail L^p-median estimation is, when $1 < p < 2$, applicable to a wider class of heavy-tailed distributions than extreme Conditional Tail Expectation estimation. The methodology is applied to the estimation of a simple middleway between the Median Shortfall and Conditional Tail Expectation on a real fire insurance data set showing very heavy tails.

Tail L^p-quantiles and medians can therefore be used as a way to estimate more appealing risk measures at extreme levels. For instance, extreme expectiles, which define an L^2-risk measure, can be estimated using extreme L^p-quantiles. However, and although expectiles generalise the mean, they do not in general benefit from a simple closed formulation in terms of expectations. This makes them more difficult to interpret than quantiles or Conditional Tail Expectations; while there is a dual representation of expectiles in terms of the gain-loss ratio [52], this is not likely to appeal to practitioners outside finance and insurance. Moreover, expectiles, with $\tau > 1/2$, are not comonotonically additive: this is best seen by combining the results of [191, 265] which imply that the only law-invariant, coherent, elicitable and comonotonically additive risk measure is the expectation. In financial applications, this is a serious problem from the point of view of the regulator, as expectiles may artificially give a sense of diversification when there is in fact none. In an effort to tackle these issues, we introduce in [5] a new notion of L^2-risk measure, called *extremile*. Like expectiles, extremiles are an L^2-version of quantiles as well as a generalisation of the expectation, and define a law-invariant and coherent risk measure. Unlike expectiles, however, they are comonotonically additive and benefit from various equivalent explicit formulations and more intuitive interpretations, most notably as average values of maxima of a (specified) number of losses. Being law-invariant, coherent and comonotonically additive, extremiles can in fact be seen as Wang DRMs with concave distortion functions, which naturally embeds them in a framework of pessimistic decision theory in the sense of [46] where the corresponding distortion function acts to depress the likelihood of the most favourable outcomes and to accentuate the likelihood of the least favourable ones. From an inferential point of view, extremiles can be estimated using L-statistics, M-statistics, linearised M-statistics and PWM-estimators. We estimate extremiles at extreme levels using asymmetric least squares estimation and an asymptotic connection to tail quantiles, and we analyse the convergence in distribution of these estimators. We also compare tail extremiles with tail quantiles and Conditional Tail Expectations, and we find that tail extremiles are more conservative than tail quantiles but less conservative than extreme Conditional Tail Expectations. As such, extremiles constitute a law-invariant, coherent and comonotonically additive risk measure which has the potential to appeal to financial institutions and regulators alike. Being easily interpretable, extremiles are also very likely to be appealing to risk managers outside insurance and finance, for example those working in environmental science or reliability analysis. Our inference methods are illustrated on two real sets of insurance data.

Finally, my contribution in [4] aims towards the multivariate analysis of random vectors with heavy-tailed marginals with applications in view in, among others, multivariate risk management. This piece of work originates in the observation that while the $\sqrt{k}-$asymptotic normality of the Hill estimator (where k denotes the effective sample size) is now a very well-studied question, analysing the joint asymptotic normality of marginal Hill estimators for a random vector with heavy-tailed marginal distributions is a much more difficult problem. This comes from the fact that the Hill estimator is written in terms of log-spacings of top order statistics, which are significantly harder to handle than the original data (assumed here to be independent and identically distributed). As a consequence, how to evaluate the asymptotic dependence between several Hill estimators is far from obvious. Such joint convergence results have only been proven very recently in [94, 156, 177]. The methods of proof therein rely on advanced theoretical methodologies, namely multivariate vague convergence in [94, 177] and multivariate empirical process theory in [156], as well as on various ad hoc technical conditions. And yet, rewriting the Hill estimator $\widehat{\gamma}(k)$ as an average of log-excesses above a high random threshold, one can note that replacing the random threshold by its non-random high quantile counterpart results in a pseudo-estimator $\widetilde{\gamma}(k)$ written in terms of sums of independent random variables, and it is a consequence of a result of [136] that $\widehat{\gamma}(k)$ and $\widetilde{\gamma}(k)$ have the same $\sqrt{k}-$asymptotic distribution. Having observed that, one may wonder whether the stronger relationship $\widehat{\gamma}(k) = \widetilde{\gamma}(k) + \mathrm{o}_{\mathbb{P}}(1/\sqrt{k})$ holds. This asymptotic relationship would make obtaining joint convergence results between Hill estimators a simple corollary of the Cramér-Wold device. The proof of this relationship is the motivation for [4]. More precisely, I embed the Hill estimator in a wider class of average excesses and then provide a simple representation of those empirical average excesses above a high random threshold in terms of their pseudo-estimator versions with a non-random threshold. I show in particular how the results of [94], on the joint asymptotic normality of marginal Hill estimators for a random vector with heavy-tailed marginal distributions, can be recovered and generalised under weaker assumptions and by elementary techniques. I also prove an analogue asymptotic representation for a class of expected shortfalls, which includes the family of Conditional Tail Moments of [114]. I then highlight a couple of applications of these general asymptotic representations, including to the obtention of joint convergence results on empirical Conditional Tail Moments that may be of interest in, for instance, testing whether certain tail moments of two asymptotically dependent variables are equal. Such an idea has already been used in a financial risk assessment context in [156].

1.2.6 Other contributions

This section gathers the research I have had the opportunity to participate in but which does not fit in one of the above main themes. For each piece of work, I briefly outline the context and the paper's contribution.

In [20], we concentrate on the development of a transformation to distributional symmetry. Transformations are often applied to given data sets in order to facilitate statistical inference by inducing finite moments, light tails and/or symmetry, the rationale being that certain statistical procedures are applicable or perform well only under such assumptions. Symmetry, in particular, has definite advantages in the context of regression, see [57, 208] for adaptive and efficient regression estimators under symmetric errors, and quasi-maximum likelihood estimation in GARCH models [142, 210]. The contribution of [20] is, first, to construct a nonparametric analogue of the *probability weighted characteristic function* (PWCF) introduced in [203] in a fully parametric setup. The PWCF is essentially a version of the traditional characteristic function, based on a Fourier transform, where the tails of the underlying distribution are penalised in a suitable way. The empirical counterpart of the PWCF, obtained by plugging in the empirical distribution function, is called the PWECF (where E stands for empirical). We then pair the PWECF with a parametric family of transformations, such as the popular Box-Cox family [61] or a more modern alternative such as the Yeo-Johnson family [262], and define a procedure based on finding the parameter for which the imaginary part of the PWECF of the transformed sample has an L^2-norm closest to 0. This method is inspired by the well-known result that a distribution is symmetric if and only if its characteristic function is real-valued; integral criteria for transformation to symmetry using the empirical characteristic function have been considered in e.g. [263]. The benefit of using the PWECF is that it is square-integrable and therefore can be used directly as part of an L^2-procedure, while to do so with the empirical characteristic function requires choosing a weight function to control its behaviour away from the origin. The strong consistency of our estimator is shown and we examine the finite-sample behaviour of our estimator on a simulation study and a real set of stock market data.

The two articles [7] and [13] are applied contributions to law and finance, respectively. The paper [13] concentrates on CAT bonds, which are part of the wider class of insurance-linked securities. Such bonds pay regular coupons to investors with the principal being contingent on a predetermined catastrophic event defined in the bond indenture. We first provide a new premium principle for such bonds, called the *financial loss premium principle*. The novel feature of this premium principle is that it takes market performance into account, allowing us to improve substantially upon the simpler linear premium principle in the prediction of post-2008 financial crisis CAT bond premiums. The second contribution of the paper is the introduction of simple formulae for the calculation of the expected loss of a bond, which is key to its pricing in both the linear premium principle and our financial loss premium principle. These formulae rely on extreme value theory and in particular on an extended regular variation formulation of the standard domain of attraction condition. We illustrate our work by constructing a fictitious wildfire CAT bond whose behaviour we examine using data from the 2016 Fort McMurray wildfire in Canada.

In [7], I supervised my undergraduate mathematics student Oliver Church in the mining of a data base of law decisions across the European Union in the area of design law, which is part of the broader area of law covering intellectual property rights. Together with Estelle Derclaye, who specialises in intellectual property law, we analysed, among others, the presence of differences in the proportion of designs found valid and infringed as a function of the level of the courts, the type of design right, the dimension of design and the level of specialisation of the judges, using standard techniques such as

χ^2-independence tests and proportion comparison tests. The findings are interpreted in the context of the EU design framework; the conclusions of the paper have been discussed during several meetings in Brussels with the relevant department of the European Commission, which directly reports to lawmakers of the European Parliament.

1.3 Structure of the manuscript

Chapter 2, the main chapter of this report, contains a more detailed presentation of the main results of my contributions in [1, 3, 4, 5, 10, 11, 12, 14] to the field of extreme risk assessment in heavy-tailed models, which currently constitutes my principal area of research. Perspectives for future work related to each of these contributions are also discussed.

Chapter 2

Statistical aspects of extreme risk assessment in heavy-tailed models

The structure of this main chapter is the following. Section 2.1 elaborates on our setting and introduces notation which shall be used throughout. Section 2.2 focuses on Wang distortion risk measures and their estimation at extreme levels. Section 2.3 adopts a different point of view and studies inference on extreme expectile-based risk measures. Section 2.4 then moves to the more general L^p context, with the examples of extreme L^p-quantile and tail L^p-median estimation. Section 2.5 returns to the L^2 world and concentrates on the new class of extremiles. Section 2.6 concludes with the presentation of some technical tools for joint inference about heavy tails, illustrated by a couple of applications to the joint estimation of extreme (marginal) risk measures for a random vector. Extended discussions, details on the simulation studies, further real data examples and all necessary mathematical proofs can be found in each referenced paper and their associated online Supplementary Material documents:

- For Section 2.2, see [11, 14].
- For Section 2.3, see [1, 12].
- For Section 2.4, see [3, 10].
- For Section 2.5, see [5].
- For Section 2.6, see [4].

This is recalled in the title of each section.

2.1 Setting and notation

We work throughout this chapter in a framework of heavy-tailed distributions, which describe the tail structure of actuarial and financial data fairly well; see, among others, p.9 of [115], p.1 of [215], as well as [72] and the references therein. A distribution function F of a real-valued random variable X is said to be *heavy-tailed* if the associated survival function $\overline{F} := 1 - F$ satisfies the following first-order regular variation condition:

$\mathcal{C}_1(\gamma)$ There exists $\gamma > 0$ such that

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

The parameter γ is called the *tail index* of the distribution function F. In other words, the distribution of X has a heavy tail with tail index γ if and only if its survival function $\overline{F}$ is regularly varying with index $-1/\gamma$, in the terminology of [59].

An equivalent, very useful way of rephrasing that condition is found using the tail quantile function U associated to F. This is defined as $U(t) = \inf\{x \in \mathbb{R} \,|\, 1/\overline{F}(x) \geq t\}$ for any $t > 1$. With this notation, condition $\mathcal{C}_1(\gamma)$ is equivalent to the assumption that U is regularly varying with index γ, that is,

$$\forall x > 0,\ \lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^{\gamma}.$$

Condition $\mathcal{C}_1(\gamma)$ can be rewritten either $\overline{F}(x) = x^{-1/\gamma}\ell(x)$ (for large x) or $U(t) = t^{\gamma}L(t)$, where ℓ, L are slowly varying functions (i.e. regularly varying with index 0). The interpretation of $\mathcal{C}_1(\gamma)$ is then that the distribution of X is in some sense close, in its tail, to the Pareto distribution with tail index γ (for which $L \equiv \ell \equiv 1$).

Many of the results we shall see in this chapter are linked to this intuitive proximity between the distribution of the variable of interest X and a Pareto distribution. To derive precise asymptotic results, such as the asymptotic normality of estimators of risk measures at extreme levels, it becomes necessary to quantify the error term in the convergence given by condition $\mathcal{C}_1(\gamma)$. This prompts us to introduce the following second-order regular variation condition:

$\mathcal{C}_2(\gamma, \rho, A)$ The function $\overline{F}$ is second-order regularly varying in a neighbourhood of $+\infty$ with index $-1/\gamma < 0$, second-order parameter $\rho \leq 0$ and an auxiliary measurable function A having constant sign and converging to 0 at infinity, that is,

$$\forall x > 0,\ \lim_{t\to\infty} \frac{1}{A(1/\overline{F}(t))}\left[\frac{\overline{F}(tx)}{\overline{F}(t)} - x^{-1/\gamma}\right] = x^{-1/\gamma}\frac{x^{\rho/\gamma} - 1}{\gamma\rho},$$

where the right-hand side should be read as $x^{-1/\gamma}\log(x)/\gamma^2$ when $\rho = 0$.

This second-order condition on $\overline{F}$ controls the rate of convergence in $\mathcal{C}_1(\gamma)$: the larger $|\rho|$ is, the faster the function $|A|$ converges to 0 (since $|A|$ is regularly varying with index ρ, in virtue of Theorems 2.3.3 and 2.3.9 in [145]) and the smaller the error in the approximation of the right tail of X by a purely Pareto tail will be. Further interpretation of this assumption can be found in [47, 145] along with numerous examples of commonly used continuous distributions satisfying it. More generally, condition $\mathcal{C}_2(\gamma, \rho, A)$ is satisfied by any distribution whose survival function $\overline{F}$ admits an asymptotic power-type expansion of the form

$$\overline{F}(x) = x^{-1/\gamma}\left(a + bx^{-c} + \mathrm{o}(x^{-c})\right) \quad \text{as} \quad x \to \infty,$$

where $a > 0$, $b \in \mathbb{R} \setminus \{0\}$ and $c > 0$ are constants. This contains in particular the Hall-Weiss class of models [163]. Let us finally mention that it is a consequence of Theorem 2.3.9 in [145] that $\mathcal{C}_2(\gamma, \rho, A)$ is actually equivalent to the following, sometimes more convenient extremal assumption on the tail quantile function U:

$$\forall x > 0,\ \lim_{t\to\infty} \frac{1}{A(t)}\left[\frac{U(tx)}{U(t)} - x^{\gamma}\right] = x^{\gamma}\frac{x^{\rho} - 1}{\rho}.$$

Here the right-hand side should be read as $x^{\gamma}\log x$ when $\rho = 0$.

2.2 Estimation of extreme Wang distortion risk measures [11, 14]

2.2.1 Introduction

Extreme quantile estimation is relevant in many areas of statistical application, and especially so in insurance and finance. For example, insurance companies wishing to operate in the European Union should comply with the Solvency II directive, which requires each company to calculate its own solvency capital so as to be able to survive the upcoming calendar year with a probability not less than 0.995. This calculation is precisely an example of extreme quantile estimation. Of course, the estimation of a single extreme quantile only gives incomplete information on extreme risk, since two distributions may well share a quantile at some common level although their respective tail behaviours are different. This is one of the reasons why other quantities, which take into account the whole right tail of the random variable X modelling the risk, were developed and studied. Examples of such indicators include the Tail Value-at-Risk (TVaR), also called Expected Shortfall, and the Stop-loss Premium for reinsurance problems, see [115, 201]. When the survival function of X is continuous, these measures can be obtained by combining the VaR and a Conditional Tail Moment (CTM) introduced by [114].

A way to encompass all these indicators in a unified framework is to consider the flexible class of Wang Distortion Risk Measures (DRMs), introduced by [250]. Wang DRMs are weighted averages of the quantile function, the weighting scheme being specified by the so-called distortion function; the aforementioned VaR, TVaR and CTM actually are particular cases of Wang DRMs, and so are many other interesting objects such as the tail standard deviation premium calculation principle [125], the Wang transform [251], and the recently introduced GlueVaR of [51]. The estimation of Wang DRMs above a *fixed* level of risk has been the subject of a number of papers: in particular, we refer the reader to [95, 96, 169, 206]. Our first objective in [11, 14] is to show how a simple linear transformation allows one to construct an *extreme* analogue of a Wang DRM. In a nutshell, starting with a Wang DRM with distortion function g, we show that the Wang DRM of X given that $X > q(\beta)$ is also a Wang DRM, obtained via a simple and explicit transformation of g, and we focus on this kind of DRM when β is close to 1. We consider the estimation of extreme Wang DRMs under classical conditions in the analysis of heavy tails. We shall consider two estimators: one based on an asymptotic proportionality relationship linking our concept of extreme Wang DRM with the corresponding extreme quantile, and another, called the functional plug-in estimator, constructed upon plugging the empirical quantile function in the weighted integral defining the Wang DRM. Our methods, it appears, provide a unified framework for the study of many frequently used extreme risk metrics, and we shall underline in particular that several results of the literature can be recovered from our own results.

The functional plug-in estimator closely mimics the structure of the true Wang DRM to be estimated. An appealing feature of this estimator in general is that it takes into account all the order statistics above a high threshold in its computation. As such, it uses all the information provided by the sample about the right tail of the underlying distribution. Of course, this is not without drawbacks at the finite-sample level. A particular disadvantage is that the finite-sample performance of the functional plug-in estimator decreases sharply in terms of mean squared error as the tail of the underlying distribution gets heavier. This is due to the propensity of heavier-tailed distributions to generate highly variable top order statistics and, therefore, to increase dramatically the variability of the estimates. We shall then show that robustifying the functional plug-in estimator by deleting certain top order statistics and/or replacing them by lower order statistics, namely trimming or winsorising the estimator, enables one to obtain estimators with improved finite-sample performance in situations of very heavy tails. This is done by a two-stage methodology: trimming/winsorisation is first applied to reduce the variability of the estimator, followed by a bias correction warranted by the heavy-tailed context.

The outline of this section is the following. We first recall the definition of a Wang DRM in Section 2.2.2. In Section 2.2.3, we present a simple way to build extreme analogues of Wang DRMs and we consider their estimation. Section 2.2.4 then introduces and studies our trimmed and winsorised

Risk measure $R_g(X)$	Distortion function g
VaR at level β	$g(x) = \mathbb{I}\{x \geq 1-\beta\}$ where $0 \leq \beta < 1$
TVaR above level β	$g(x) = \min\left\{\dfrac{x}{1-\beta}, 1\right\}$ where $0 \leq \beta < 1$
Proportional Hazard transform	$g(x) = x^\alpha$ where $0 < \alpha < 1$
Dual Power	$g(x) = 1-(1-x)^{1/\alpha}$ where $0 < \alpha < 1$
MAXMINVAR	$g(x) = (1-(1-x)^\alpha)^{1/\alpha}$ where $0 < \alpha < 1$
MINMAXVAR	$g(x) = 1-(1-x^{1/\alpha})^\alpha$ where $0 < \alpha < 1$
Gini's principle	$g(x) = (1+\alpha)x - \alpha x^2$ where $0 < \alpha \leq 1$
Denneberg's absolute deviation	$g(x) = \begin{cases} (1+\alpha)x & \text{if} \quad 0 \leq x \leq 1/2 \\ \alpha + (1-\alpha)x & \text{if} \quad 1/2 \leq x \leq 1 \end{cases}$ where $0 < \alpha \leq 1$
Exponential transform	$g(x) = \begin{cases} (1-\exp(-rx))/(1-\exp(-r)) & \text{if} \quad r > 0 \\ x & \text{if} \quad r = 0 \end{cases}$
Logarithmic transform	$g(x) = \begin{cases} (\log(1+rx))/(\log(1+r)) & \text{if} \quad r > 0 \\ x & \text{if} \quad r = 0 \end{cases}$
Square-root transform	$g(x) = \begin{cases} (\sqrt{1+rx}-1)/(\sqrt{1+r}-1) & \text{if} \quad r > 0 \\ x & \text{if} \quad r = 0 \end{cases}$
S-inverse shaped transform	$g(x) = a\left(\dfrac{x^3}{6} - \dfrac{\delta}{2}x^2 + \left(\dfrac{\delta^2}{2}+\beta\right)x\right)$ where $a = \left(\dfrac{1}{6} - \dfrac{\delta}{2} + \dfrac{\delta^2}{2} + \beta\right)^{-1}$ with $0 \leq \delta \leq 1$ and $\beta \in \mathbb{R}$
Wang's transform	$g(x) = \Phi(\Phi^{-1}(x) + \Phi^{-1}(\alpha))$ where Φ is the standard Gaussian distribution function and $0 \leq \alpha \leq 1$
Beta's transform	$g(x) = \displaystyle\int_0^x \frac{1}{\beta(a,b)} t^{a-1}(1-t)^{b-1}dt$ where $\beta(a,b)$ is the Beta function with parameters a, $b > 0$

Table 2.1: Some risk measures and their distortion functions.

estimators. Section 2.2.5 is devoted to the study of the finite-sample performance of our estimators on simulated and real data. Section 2.2.6 concludes by offering some perspective on future work.

2.2.2 Wang distortion risk measures [14]

We say throughout this section that $g : [0,1] \to [0,1]$ is a *distortion function* if it is right-continuous and nondecreasing with $g(0) = 0$ and $g(1) = 1$. The *Wang DRM* of a positive random variable X with distortion function g is then

$$R_g(X) := \int_0^\infty g(1-F(x))dx,$$

where F is the cumulative distribution function of X. An alternative, easily interpretable expression of $R_g(X)$ can be found. Let $m = \inf\{\alpha \in [0,1] \,|\, g(\alpha) > 0\}$ and $M = \sup\{\alpha \in [0,1] \,|\, g(\alpha) < 1\}$. Assume for the moment that the quantile function q, defined by $q(\tau) = \inf\{x \in \mathbb{R} \,|\, F(x) \geq \tau\}$ for any $\tau \in (0,1)$, is continuous on $V \cap (0,1)$ with V an open interval containing $[1-M, 1-m]$. Noticing that F is the right-continuous inverse of q, a classical change-of-variables formula and an integration by parts then entail that $R_g(X)$, provided it is finite, can be written as a Lebesgue-Stieltjes integral:

$$R_g(X) = \int_0^1 g(\alpha)dq(1-\alpha) = \int_0^1 q(1-\alpha)dg(\alpha).$$

A Wang DRM can thus be understood as a weighted version of the expectation of the random variable X. Specific examples include the quantile at level β or VaR(β), obtained by setting $g(x) = \mathbb{I}\{x \geq 1-\beta\}$, with $\mathbb{I}\{\cdot\}$ denoting the indicator function; the Tail Value-at-Risk TVaR(β) in the worst $100(1-\beta)\%$ of cases, the average of all quantiles exceeding VaR(β), is recovered by taking $g(x) = \min(x/(1-\beta), 1)$. Table 2.1 gives further examples of classical DRMs and their distortion functions. Broadly speaking, the class of Wang DRMs allows almost total flexibility as far as the weighting scheme is considered. Besides, any spectral risk measure of X (see e.g. [78]) is also a Wang DRM.

Risk measure	Expression as a combination of $\mathrm{CTM}_a(\beta)$ and $\mathrm{VaR}(\beta)$
$\mathrm{CTE}(\beta)$	$\mathrm{CTM}_1(\beta)$
$\mathrm{GlueVaR}^{h_1,h_2}_{\beta,\alpha}$	$\omega_1 \mathrm{CTM}_1(\beta) + \omega_2 \mathrm{CTM}_1(\alpha) + \omega_3 \mathrm{VaR}(\alpha)$ where $\omega_1 = h_1 - \dfrac{(h_2-h_1)(1-\beta)}{\beta-\alpha}$, $\omega_2 = \dfrac{(h_2-h_1)(1-\alpha)}{\beta-\alpha}$ and $\omega_3 = 1-\omega_1-\omega_2 = 1-h_2$, with $h_1 \in [0,1]$, $h_2 \in [h_1,1]$ and $\alpha < \beta$
$\mathrm{SP}(\beta)$	$(1-\beta)(\mathrm{CTM}_1(\beta) - \mathrm{VaR}(\beta))$
$\mathrm{CTV}(\beta)$	$\mathrm{CTM}_2(\beta) - \mathrm{CTM}_1^2(\beta)$
$\mathrm{TSD}_\lambda(\beta)$	$\mathrm{CTM}_1(\beta) + \lambda\sqrt{\mathrm{CTM}_2(\beta) - \mathrm{CTM}_1^2(\beta)}$ where $\lambda \geq 0$
$\mathrm{CTS}(\beta)$	$\mathrm{CTM}_3(\beta)/(\mathrm{CTM}_2(\beta) - \mathrm{CTM}_1^2(\beta))^{3/2}$

Table 2.2: Link between the CTM and some risk measures when the distribution function of X is continuous.

Furthermore, we note that if $h : [0,\infty) \to [0,\infty)$ is a strictly increasing, continuously differentiable function, then the Wang DRM of $h(X)$ with distortion function g is

$$R_g(h(X)) = \int_0^1 h \circ q(1-\alpha) dg(\alpha). \tag{2.2.1}$$

For instance, the choices $g(x) = \min(x/(1-\beta), 1)$, $\beta \in (0,1)$ and $h(x) = x^a$, with a a positive real number, yield, after integrating by parts,

$$R_g(X^a) = \mathrm{CTM}_a(\beta) := \mathbb{E}(X^a | X > q(\beta)),$$

provided F is continuous. This is the Conditional Tail Moment (CTM) of order a of the random variable X, as introduced in [114]. Especially, when F is continuous, the TVaR coincides with the Conditional Tail Expectation of X. Table 2.2 gives several examples of risk measures, such as the Stop-loss Premium (SP), that can be obtained by combining a finite number of CTMs and the VaR. In an actuarial context, a DRM is a coherent risk measure (see [44]) if and only if the distortion function g is concave [258]. Coherency of a risk measure reflects in particular on the diversification principle which asserts that aggregating two risks cannot be worse than handling them separately [44]. Especially, while the VaR is not a coherent risk measure in general, the TVaR is, for instance, and this has already been noted several times in the recent literature.

Yet another property of risk measures, namely elicitability [134, 265], has gained prominence in recent years since it has been argued to allow for correct forecast performance comparisons. While the VaR is an elicitable (and consistent) risk measure, the TVaR is not; more generally, it has been shown recently by [185, 255] that Wang DRMs different from either the VaR or the simple expectation do not satisfy such a property. An example of a risk measure that is both coherent and elicitable is the expectile [190, 209] when it is larger than the expectation. The estimation of extreme expectiles, which to the best of our knowledge cannot be written as a simple combination of extreme Wang DRMs of X, is tackled in Section 2.3 (see [1, 12]).

2.2.3 Framework and estimation [14]

2.2.3.1 Extreme versions of Wang DRMs

Extreme versions of Wang risk measures may be obtained as follows. Let g be a distortion function and for every $\beta \in (0,1)$, consider the function g_β defined by

$$\forall y \in [0,1],\ g_\beta(y) := g\left(\min\left[1, \frac{y}{1-\beta}\right]\right) = \begin{cases} g\left(\dfrac{y}{1-\beta}\right) & \text{if } \ y \leq 1-\beta \\ 1 & \text{otherwise.} \end{cases}$$

Such a function, which is deduced from g by a simple piecewise linear transform of its argument, is thus constant equal to 1 on $[1-\beta, 1]$. Especially, if g gives rise to a coherent Wang DRM, so does g_β.

We now consider the Wang DRM of X with distortion function g_β:

$$R_{g,\beta}(X) := \int_0^\infty g_\beta(1-F(x))dx.$$

Because the inequality $F(x) \geq \beta$ is equivalent to $x \geq q(\beta)$, we have:

$$R_{g,\beta}(X) = \int_0^\infty g(1-F_\beta(x))dx \;\text{ with }\; F_\beta(x) := \max\left[0, \frac{F(x)-\beta}{1-\beta}\right]. \tag{2.2.2}$$

When moreover q is continuous and strictly increasing in a neighbourhood of β, then

$$F_\beta(x) = \max\left[0, \frac{F(x)-F(q(\beta))}{1-q(\beta)}\right] = \mathbb{P}(X \leq x | X > q(\beta)),$$

which makes the interpretation of the risk measure $R_{g,\beta}(X)$ clear: it is the Wang DRM of X given that it lies above the level $q(\beta)$. In other words, we have shown the following.

Proposition 2.2.1. *Assume that for some $t > 0$, the function q is continuous and strictly increasing on $[t, 1)$. Then for all $\beta > t$ and any strictly increasing and continuously differentiable function h on $(0, \infty)$,*

$$R_{g,\beta}(h(X)) = R_g(h(X_\beta)) \;\text{ with }\; \mathbb{P}(X_\beta \leq x) = \mathbb{P}(X \leq x | X > q(\beta)).$$

When $\beta \uparrow 1$, we may then think of this construction as a way to consider Wang DRMs of the extremes of X. Choosing $h(x) = x$ makes it possible to recover some simple and widely used extreme risk measures: the usual extreme VaR is obtained by setting $g(x) = \mathbb{I}\{x = 1\}$, and an extreme version of the TVaR is obtained by taking $g(x) = x$. The same idea yields extreme analogues of the various risk measures shown in Table 2.1. Furthermore, as highlighted in Section 2.2.2, choosing $g(x) = x$ and $h(x) = x^a$, $a > 0$, yields an extreme version of a CTM of X, and therefore extreme versions of quantities such as those introduced in Table 2.2 can be studied.

It is worth noting at this point that the construction presented here is different from that of [249]. In the latter paper, the authors consider the Wang DRM R_g of $(X - R)\mathbb{I}\{X > R\} = \max(X - R, 0)$ for large R. Their construction is thus adapted to the examination of excess-of-loss reinsurance policies for extreme losses; their work is, by the way, restricted to the case of a concave function g satisfying a regular variation condition in a neighbourhood of 0. It therefore excludes the simple VaR risk measure, for instance, as well as the GlueVaR of [51]. Our idea is rather to consider a conditional construction in the sense that we look at the Wang DRMs of X given that it lies above a high level, with conditions as weak as possible on the function g, in an effort to be able to examine the extremes of X in as unified a way as possible.

2.2.3.2 Estimation using an asymptotic equivalent

We now give a first idea on how to estimate this type of extreme risk measure. Let $(X_1, \ldots, X_n)$ be a sample of independent and identically distributed copies of a random variable X having distribution function F, and let (β_n) be a nondecreasing sequence of real numbers belonging to $(0, 1)$, which converges to 1. Assume for the time being that X is Pareto distributed,

$$\forall x > 1,\; \mathbb{P}(X \leq x) = 1 - x^{-1/\gamma}$$

where $\gamma > 0$ is the so-called tail index of X. In this case, the quantile function of X is $q(\alpha) = (1-\alpha)^{-\gamma}$ for all $\alpha \in (0, 1)$. Using (2.2.1) in Section 2.2.2 and a simple change of variables, we get:

$$\begin{aligned} R_{g,\beta_n}(h(X)) = \int_0^1 h \circ q(1-\alpha)dg_{\beta_n}(\alpha) &= \int_0^1 h \circ q(1-(1-\beta_n)s)dg(s) \\ &= \int_0^1 h(q(\beta_n)s^{-\gamma})dg(s). \end{aligned} \tag{2.2.3}$$

In this case, an estimator of $R_{g,\beta_n}(h(X))$ would then be obtained by plugging estimators of $q(\beta_n)$ and γ in the right-hand side of (2.2.3).

Of course, in general, a strong relationship such as (2.2.3) cannot be expected to hold, but it stays true to some extent when X has a heavy-tailed distribution. We also suppose that the quantile function q of X is continuous and strictly increasing in a neighbourhood of infinity, which makes possible the use of (2.2.1) for n large enough. Finally, we assume that the function h is a positive power of x: $h(x) = x^a$, where $a > 0$. This choice allows us to consider estimators of a large class of risk measures of X, including the aforementioned CTM. In this case, one can show that

$$R_{g,\beta_n}(X^a) = [q(\beta_n)]^a \int_0^1 s^{-a\gamma} dg(s)(1 + \mathrm{o}(1)) \quad \text{as} \quad n \to \infty,$$

provided $\int_0^1 s^{-a\gamma-\eta} dg(s) < \infty$ for some $\eta > 0$. This suggests that the above idea for the construction of the estimator can still be used provided n is large enough. Specifically, if $\widehat{q}_n(\alpha) = X_{\lceil n\alpha \rceil, n}$ denotes the empirical quantile function, in which $X_{1,n} \leq \cdots \leq X_{n,n}$ are the order statistics of the sample $(X_1, \ldots, X_n)$ and $\lceil \cdot \rceil$ is the ceiling function, we set

$$\widehat{R}^{AE}_{g,\beta_n}(X^a) := X^a_{\lceil n\beta_n \rceil, n} \int_0^1 s^{-a\widehat{\gamma}_n} dg(s)$$

where $\widehat{\gamma}_n$ is any consistent estimator of γ. This estimator is called the AE estimator in what follows. Notice that the integrability condition $\int_0^1 s^{-a\gamma-\eta} dg(s) < \infty$, which should be thought of as a condition that guarantees the existence of the considered Wang DRM, makes the estimator introduced here well-defined with probability arbitrarily large when n is large enough, due to the consistency of $\widehat{\gamma}_n$. For a related but different idea, see [249].

An appealing feature of the AE estimator is that it is easy to compute in many cases:

- in the case of the Conditional Tail moment of order a, i.e. $g(x) = x$, the estimator reads

$$\widehat{R}^{AE}_{g,\beta_n}(X^a) = X^a_{\lceil n\beta_n \rceil, n} \int_0^1 s^{-a\widehat{\gamma}_n} ds = \frac{X^a_{\lceil n\beta_n \rceil, n}}{1 - a\widehat{\gamma}_n}$$

  when $a\widehat{\gamma}_n < 1$. In particular, this provides an estimator different from the sample average estimator of [114];

- in the case of the Dual Power risk measure, i.e. $g(x) = 1 - (1 - x)^{1/\alpha}$ where $0 < \alpha < 1$ and $a = 1$, then when $r := 1/\alpha$ is an integer, the estimator is

$$\widehat{R}^{AE}_{g,\beta_n}(X) = X_{\lceil n\beta_n \rceil, n} \int_0^1 r s^{-\widehat{\gamma}_n} (1 - s)^{r-1} ds = \frac{r!\Gamma(1 - \widehat{\gamma}_n)}{\Gamma(1 - \widehat{\gamma}_n + r)} X_{\lceil n\beta_n \rceil, n}$$

  provided $\widehat{\gamma}_n < 1$. Here Γ is Euler's Gamma function, $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$;

- in the case of the Proportional Hazard transform, i.e. $g(x) = x^\alpha$ where $0 < \alpha < 1$ and $a = 1$, the estimator is

$$\widehat{R}^{AE}_{g,\beta_n}(X) = X_{\lceil n\beta_n \rceil, n} \int_0^1 \alpha s^{\alpha - \widehat{\gamma}_n - 1} ds = \frac{\alpha X_{\lceil n\beta_n \rceil, n}}{\alpha - \widehat{\gamma}_n}$$

  provided $\widehat{\gamma}_n < \alpha$.

In order to examine the asymptotic properties of our estimator, it is necessary to compute the order of magnitude of its asymptotic bias. We do so by working under condition $\mathcal{C}_2(\gamma, \rho, A)$, which allows us to write our first main result.

Theorem 2.2.2. *Assume that U is regularly varying with index $\gamma > 0$. Assume further that $\beta_n \to 1$ and $n(1 - \beta_n) \to \infty$.*

1. *Pick a distortion function g and $a > 0$. If there is some $\eta > 0$ such that*

$$\int_0^1 s^{-a\gamma-\eta} dg(s) < \infty$$

and $\widehat{\gamma}_n$ is a consistent estimator of γ, then

$$\frac{\widehat{R}^{AE}_{g,\beta_n}(X^a)}{R_{g,\beta_n}(X^a)} - 1 \xrightarrow{\mathbb{P}} 0 \quad as \quad n \to \infty.$$

2. *Assume moreover that U satisfies condition $\mathcal{C}_2(\gamma, \rho, A)$ and $\sqrt{n(1-\beta_n)}A((1-\beta_n)^{-1}) \to \lambda \in \mathbb{R}$. Pick a $d-$tuple of distortion functions $(g_1, \ldots, g_d)$ and $a_1, \ldots, a_d > 0$. If for some $\eta > 0$,*

$$\forall j \in \{1, \ldots, d\}, \ \int_0^1 s^{-a_j\gamma-1/2-\eta} dg_j(s) < \infty,$$

then, provided we have the joint convergence

$$\sqrt{n(1-\beta_n)} \left(\widehat{\gamma}_n - \gamma, \frac{X_{\lceil n\beta_n \rceil, n}}{q(\beta_n)} - 1 \right) \xrightarrow{d} (\Gamma, \Theta)$$

it holds that the random vector

$$\sqrt{n(1-\beta_n)} \left(\frac{\widehat{R}^{AE}_{g_j,\beta_n}(X^{a_j})}{R_{g_j,\beta_n}(X^{a_j})} - 1 \right)_{1 \le j \le d}$$

asymptotically has the joint distribution of

$$\left(a_j \left[-\lambda \frac{\displaystyle\int_0^1 s^{-a_j\gamma} \frac{s^{-\rho}-1}{\rho} dg_j(s)}{\displaystyle\int_0^1 s^{-a_j\gamma} dg_j(s)} + \frac{\displaystyle\int_0^1 s^{-a_j\gamma} \log(1/s) dg_j(s)}{\displaystyle\int_0^1 s^{-a_j\gamma} dg_j(s)} \Gamma + \Theta \right] \right)_{1 \le j \le d}.$$

Because of the restriction $n(1-\beta_n) \to \infty$, Theorem 2.2.2 only ensures that the estimator consistently estimates so-called intermediate (i.e. not "too extreme") Wang DRMs. This restriction will be lifted in Section 2.2.3.4 by the introduction of an estimator adapted to the extreme value framework.

2.2.3.3 Estimation using a functional plug-in estimator

Our idea here is to introduce an alternative estimator obtained by making a single approximation instead of the two successive ones

$$q(1-(1-\beta_n)s) \approx q(\beta_n)s^{-\gamma} \approx X_{\lceil n\beta_n \rceil, n} s^{-\widehat{\gamma}_n},$$

which we can then expect to perform better than the AE estimator. Recall that

$$R_{g,\beta_n}(h(X)) = \int_0^1 h \circ q(1-(1-\beta_n)s) dg(s).$$

We consider the statistic obtained by replacing the function $s \mapsto q(1-(1-\beta_n)s)$ by its empirical counterpart $s \mapsto \widehat{q}_n(1-(1-\beta_n)s) = X_{\lceil n(1-(1-\beta_n)s) \rceil, n}$. This yields the functional plug-in estimator

$$\widehat{R}^{PL}_{g,\beta_n}(h(X)) = \int_0^1 h \circ \widehat{q}_n(1-(1-\beta_n)s) dg(s)$$

which we call the PL estimator. Contrary to the AE estimator, the PL estimator is well-defined and finite with probability 1, and does not require an external estimator of γ. Its expression is a bit more involved though; in the case when $n(1-\beta_n)$ is actually a positive integer and g is continuous on $[0,1]$, it is easy to show that it takes the simpler form

$$\widehat{R}^{PL}_{g,\beta_n}(h(X)) = h(X_{n\beta_n+1,n}) + \sum_{i=1}^{n(1-\beta_n)-1} g\left(\frac{i}{n(1-\beta_n)}\right)[h(X_{n-i+1,n}) - h(X_{n-i,n})].$$

Our aim is now to examine the asymptotic properties of the PL estimator. While it would be straightforward to obtain the asymptotic properties of this estimator for a fixed order β through L-statistic techniques, see e.g. [169], a difficulty here lies in the fact that $\beta = \beta_n \uparrow 1$. As a consequence, theoretical developments involve knowing the behaviour of the quantile process $s \mapsto \widehat{q}_n(s)$ on $[\beta_n, 1]$. The crucial tool is a corollary of the powerful distributional approximation stated in Theorem 2.1 of [101], relating this tail quantile process to a standard Brownian motion up to a bias term.

Theorem 2.2.3. *Assume that U satisfies condition $\mathcal{C}_2(\gamma,\rho,A)$. Assume further that $\beta_n \to 1$, $n(1-\beta_n) \to \infty$ and $\sqrt{n(1-\beta_n)}A((1-\beta_n)^{-1}) \to \lambda \in \mathbb{R}$. Pick a $d-$tuple of distortion functions $(g_1,\ldots,g_d)$ and $a_1,\ldots,a_d > 0$. If for some $\eta > 0$,*

$$\forall j \in \{1,\ldots,d\},\ \int_0^1 s^{-a_j\gamma-1/2-\eta} dg_j(s) < \infty,$$

then

$$\sqrt{n(1-\beta_n)}\left(\frac{\widehat{R}^{PL}_{g_j,\beta_n}(X^{a_j})}{R_{g_j,\beta_n}(X^{a_j})} - 1\right)_{1\le j\le d} \xrightarrow{d} \mathcal{N}(0,V)$$

with V being the $d\times d$ matrix whose $(i,j)-$th entry is

$$V_{i,j} = a_i a_j \gamma^2 \frac{\int_{[0,1]^2} \min(s,t) s^{-a_i\gamma-1} t^{-a_j\gamma-1} dg_i(s) dg_j(t)}{\int_0^1 s^{-a_i\gamma} dg_i(s) \int_0^1 t^{-a_j\gamma} dg_j(t)}.$$

This asymptotic normality result, unsurprisingly, is also restricted to the case $n(1-\beta_n) \to \infty$, as was Theorem 2.2.2. We can draw an interesting consequence from Theorem 2.2.3: for $b \in \mathbb{R}$, consider the class $\mathcal{E}_b([0,1])$ of those continuously differentiable functions on $(0,1)$ such that $s^{-b}|g'(s)|$ is bounded for s in a neighbourhood of 0. For instance, any polynomial function belongs to $\mathcal{E}_0([0,1])$, and the Proportional Hazard [250] distortion function $g(s) = s^\alpha$, $\alpha \in (0,1)$, belongs to $\mathcal{E}_{\alpha-1}([0,1])$.

Corollary 2.2.4. *Assume that U satisfies condition $\mathcal{C}_2(\gamma,\rho,A)$. Assume further that $\beta_n \to 1$, $n(1-\beta_n) \to \infty$ and $\sqrt{n(1-\beta_n)}A((1-\beta_n)^{-1}) \to \lambda \in \mathbb{R}$. Pick a $d-$tuple of distortion functions $(g_1,\ldots,g_d)$ and $a_1,\ldots,a_d > 0$. Assume there are $b_1,\ldots,b_d \in \mathbb{R}$ such that for all $j \in \{1,\ldots,d\}$, we have $g_j \in \mathcal{E}_{b_j}([0,1])$. If*

$$\forall j \in \{1,\ldots,d\},\ \gamma < \frac{2b_j+1}{2a_j},$$

then

$$\sqrt{n(1-\beta_n)}\left(\frac{\widehat{R}^{PL}_{g_j,\beta_n}(X^{a_j})}{R_{g_j,\beta_n}(X^{a_j})} - 1\right)_{1\le j\le d} \xrightarrow{d} \mathcal{N}(0,V)$$

with V as in Theorem 2.2.3.

In particular, the condition on γ we get for the asymptotic normality of the CTM of order a, obtained with $g(x) = x$ and thus $g \in \mathcal{E}_0([0,1])$, is $\gamma < 1/2a$, which is the condition obtained by [114]. One may also readily check that the asymptotic variance is the same as in Theorem 1 there.

Just like the AE estimator, the PL estimator is only consistent when (β_n) is an intermediate sequence. Our purpose is now to remove this restriction by using the extrapolation methodology of [256].

2.2.3.4 Estimating extreme Wang DRMs of arbitrary order

For any $s \in (0,1)$ and $a > 0$ we have

$$[q(1-(1-\delta_n)s)]^a = \left(\frac{1-\beta_n}{1-\delta_n}\right)^{a\gamma} [q(1-(1-\beta_n)s)]^a(1+\mathrm{o}(1))$$

as $n \to \infty$, as a consequence of the regular variation property of U, and provided that (β_n) is a sequence converging to 1 such that $(1-\delta_n)/(1-\beta_n)$ converges to a positive limit. Integrating this relationship with respect to the distortion measure dg therefore suggests that

$$R_{g,\delta_n}(X^a) = \left(\frac{1-\beta_n}{1-\delta_n}\right)^{a\gamma} R_{g,\beta_n}(X^a)(1+\mathrm{o}(1)).$$

A way to design an adapted estimator of the extreme risk measure $R_{g,\delta_n}(X^a)$, when $n(1-\delta_n) \to c < \infty$, is thus to take a sequence (β_n) such that $n(1-\beta_n) \to \infty$, and to plug in any relatively consistent estimator $\widehat{R}_{g,\beta_n}(X^a)$ of the intermediate Wang DRM $R_{g,\beta_n}(X^a)$. This yields a Weissman-type estimator of $R_{g,\delta_n}(X^a)$ (see [256]):

$$\widehat{R}^W_{g,\delta_n}(X^a;\beta_n) := \left(\frac{1-\beta_n}{1-\delta_n}\right)^{a\widehat{\gamma}_n} \widehat{R}_{g,\beta_n}(X^a).$$

This principle can of course be applied to the AE and PL estimators to obtain two different extrapolated estimators. Our asymptotic result on this class of estimators is the following.

Theorem 2.2.5. *Assume that U satisfies condition $\mathcal{C}_2(\gamma,\rho,A)$, with $\rho < 0$. Assume further that $\beta_n, \delta_n \to 1$, $n(1-\beta_n) \to \infty$, $(1-\delta_n)/(1-\beta_n) \to 0$, $\sqrt{n(1-\beta_n)}/\log([1-\beta_n]/[1-\delta_n]) \to \infty$ and $\sqrt{n(1-\beta_n)}A((1-\beta_n)^{-1}) \to \lambda \in \mathbb{R}$. Pick a $d-$tuple of distortion functions $(g_1,\ldots,g_d)$ and $a_1,\ldots,a_d > 0$. If for some $\eta > 0$,*

$$\forall j \in \{1,\ldots,d\},\ \int_0^1 s^{-a_j\gamma-1/2-\eta} dg_j(s) < \infty$$

and $\sqrt{n(1-\beta_n)}(\widehat{\gamma}_n - \gamma) \xrightarrow{d} \xi$, then provided

$$\forall j \in \{1,\ldots,d\},\ \sqrt{n(1-\beta_n)}\left(\frac{\widehat{R}_{g_j,\beta_n}(X^{a_j})}{R_{g_j,\beta_n}(X^{a_j})} - 1\right) = \mathrm{O}_{\mathbb{P}}(1)$$

we have that

$$\frac{\sqrt{n(1-\beta_n)}}{\log([1-\beta_n]/[1-\delta_n])}\left(\frac{\widehat{R}^W_{g_j,\delta_n}(X^{a_j};\beta_n)}{R_{g_j,\delta_n}(X^{a_j})} - 1\right)_{1\le j\le d} \xrightarrow{d} \begin{pmatrix} a_1\xi \\ \vdots \\ a_d\xi \end{pmatrix}.$$

In the particular case $d=1$, $a=1$ and $g(x)=0$ if $x<1$, we recover the asymptotic result about Weissman's estimator, see Theorem 4.3.8 in [145]. For $g(x)=x$ and $d=1$, we recover a result similar to Theorem 2 of [114] if the intermediate estimator is the PL estimator. In practical situations, the estimation of the parameter γ is of course a central question. Classical tail index estimators (see Section 3 of [145] and the review in Section 5 of [138]), when computed with the top $100(1-\beta_n)\%$ of the data, typically converge at the required rate $\sqrt{n(1-\beta_n)}$.

2.2.4 Trimmed and winsorised versions of the functional plug-in estimator [11]

2.2.4.1 Heavy tails, top order statistics and finite-sample variability

A problem with the use of the PL estimator can arise in practice when g is strictly increasing in a left neighbourhood of 1, which is for instance the case for the Tail-Value-at-Risk, Dual Power and Proportional Hazard risk measures. In that case, the PL estimator takes into account all the data above level $X_{\lceil n\beta_n\rceil,n}$ in the sample, and in any sample where some of the highest order statistics are far from their population counterparts, this will result in inappropriate estimates. Such situations appear regularly: suppose here that X has a Pareto distribution with parameter $1/\gamma$, i.e.

$$\forall x \geq 1,\ F(x) = 1 - x^{-1/\gamma} \ \text{ so that } \ \forall \alpha \in (0,1),\ q(\alpha) = (1-\alpha)^{-\gamma}.$$

The probability that the sample maximum $X_{n,n}$ exceeds a multiple of its population counterpart, namely the quantile $q(1-n^{-1})$, is then

$$\begin{aligned}
\mathbb{P}(X_{n,n} > Kq(1-n^{-1})) &= 1 - \left[1 - \mathbb{P}(X > Kq(1-n^{-1}))\right]^n \\
&= 1 - \left[1 - \frac{K^{-1/\gamma}}{n}\right]^n \\
&\approx 1 - \exp(-K^{-1/\gamma}) \ \text{ for large enough } n.
\end{aligned}$$

This result is, of course, linked to the fact that sample quantiles at extreme levels do not estimate the corresponding true quantiles consistently; a related point is that, sample-wise, the most extreme values tend not to give a fair picture of the extremes of the underlying distribution [131]. Carrying on with this example, it follows that in the case $\gamma = 0.49$, $K = 3$, and $n = 1000$, the sample maximum is larger than 88.54, which is three times the quantile at level 0.999, with probability approximately equal to 0.101. In this sense, approximately 10% of samples feature at least one unusually high value. Besides, as the above calculation shows, the probability that a sample features one or several very large values increases as γ increases, i.e. as the tail gets heavier. The influence of such values on extreme Wang DRM estimates should of course not be underestimated. In the case of the Tail-Value-at-Risk, obtained for $g(s) = s$, namely:

$$\text{TVaR}(\beta) = R_{\text{Id},\beta}(X) = \frac{1}{1-\beta}\int_0^{1-\beta} q(1-\alpha)d\alpha = \frac{(1-\beta)^{-\gamma}}{1-\gamma},$$

a simulation study not reported here shows that, on 5000 replicates of a sample of 1000 independent random copies of the aforementioned Pareto distribution conditioned on the fact that the sample maximum is larger than 88.54, the relative bias of the Tail-Value-at-Risk PL estimator,

$$\widehat{R}^{PL}_{\text{Id},\beta}(X) = \frac{1}{1-\beta}\int_0^{1-\beta} \widehat{q}_n(1-\alpha)d\alpha = \frac{1}{n(1-\beta)}\sum_{j=1}^{n(1-\beta)} X_{n-j+1,n},$$

at level $\beta = 0.95$ is approximately 0.389. In other words, the PL estimator is, on such samples, on average a little less than 40% higher than it should be. Of course, this could have been expected since it is straightforward to see that the above PL estimator is adversely affected by high values of $X_{n,n}$ (just as the sample mean is). The concern here is rather that problematic cases, through the apparition of very high values of the sample maximum and more generally of the highest order statistics, appear more and more frequently as γ increases, even when γ is such that the estimator $\widehat{R}^{PL}_{\text{Id},\beta}(X)$ is asymptotically Gaussian. Our first objective is to introduce estimators which deal with this variability issue.

2.2.4.2 First step of improvement: Reducing finite-sample variability

A simple idea to tackle the problem highlighted in Section 2.2.4.1 is to delete the highest problematic values altogether, namely to trim the PL estimator, by considering the statistic

$$\widehat{R}^{\text{Trim}}_{g,\beta_n,t_n}(h(X)) = \int_0^1 h \circ \widehat{q}_n(t_n - (t_n - \beta_n)s)dg(s),$$

where (t_n) is a sequence of trimming levels, i.e. a sequence such that $\beta_n < t_n \leq 1$. This is the empirical estimator of

$$R^{\text{Trim}}_{g,\beta_n,t_n}(h(X)) = \int_0^1 h \circ q(t_n - (t_n - \beta_n)s)dg(s),$$

which in many cases is actually the Wang DRM of $h(X)$ given that X lies between $q(\beta_n)$ and $q(t_n)$, as the following result shows.

Proposition 2.2.6. *Let $\beta \in (0,1)$ and $t \in (0,1]$ such that $t > \beta$. If q is continuous and strictly increasing on an open interval containing $[\beta, 1)$ then:*

$$R^{\text{Trim}}_{g,\beta,t}(h(X)) = R_g(h(X^{\text{Trim}}_{\beta,t})) \quad \textit{with} \quad X^{\text{Trim}}_{\beta,t} \stackrel{d}{=} X | X \in [q(\beta), q(t)].$$

In practice, it is very often the case that nt_n and $n(t_n - \beta_n)$ are positive integers (see Section 2.2.5). In that particular case, the trimmed estimator $\widehat{R}^{\text{Trim}}_{g,\beta_n,t_n}(h(X))$, which we shall call the Trim-PL estimator, can be conveniently rewritten as a generalised L-statistic:

$$\begin{aligned}
\widehat{R}^{\text{Trim}}_{g,\beta_n,t_n}(h(X)) &= \sum_{i=1}^{n(t_n-\beta_n)} h(X_{nt_n-i+1,n}) \int_0^1 \mathbb{I}\{x_{i-1,n}(\beta_n,t_n) \leq s < x_{i,n}(\beta_n,t_n)\} dg(s) \\
&+ h(X_{n\beta_n,n}) \left[g(1) - \lim_{\substack{s\to 1 \\ s<1}} g(s) \right] \quad \text{with} \quad x_{i,n}(\beta_n,t_n) = \frac{i}{n(t_n-\beta_n)} \\
&= \sum_{i=1}^{n(t_n-\beta_n)} h(X_{nt_n-i+1,n}) \left[\lim_{\substack{s\to x_{i,n}(\beta_n,t_n) \\ s<x_{i,n}(\beta_n,t_n)}} g(s) - \lim_{\substack{s\to x_{i-1,n}(\beta_n,t_n) \\ s<x_{i-1,n}(\beta_n,t_n)}} g(s) \right] \\
&+ h(X_{n\beta_n,n}) \left[1 - \lim_{\substack{s\to 1 \\ s<1}} g(s) \right].
\end{aligned}$$

When the function g is moreover continuous on $[0,1]$, this can be further simplified as

$$\widehat{R}^{\text{Trim}}_{g,\beta_n,t_n}(h(X)) = h(X_{n\beta_n+1,n}) + \sum_{i=1}^{n(t_n-\beta_n)-1} g\left(\frac{i}{n(t_n-\beta_n)}\right) [h(X_{nt_n-i+1,n}) - h(X_{nt_n-i,n})].$$

It should thus be clear at this stage that the Trim-PL estimator $\widehat{R}^{\text{Trim}}_{g,\beta_n,t_n}(h(X))$ is both the empirical counterpart of $R^{\text{Trim}}_{g,\beta_n,t_n}(h(X))$ and a trimmed estimator of $R_{g,\beta_n}(h(X))$ in the sense that the top order statistics $X_{nt_n+1,n}, \ldots, X_{n,n}$ are discarded for the estimation. This amounts to a trimming percentage equal to $100(1-t_n)\%$ in the highest values of the sample. The original, intermediate PL estimator is recovered for $t_n = 1$.

Although the idea of trimming seems appealing because it is expected to curb the estimator's variability, it may not be the best method available in that it effectively reduces the available sample size. The overall bias of the estimator, meanwhile, would be negatively affected as well, since despite their high variability, the highest order statistics in the sample are those who carry the least bias about the extremes of the underlying distribution. One could try reducing the loss of information that trimming entails by winsorising the estimator $\widehat{R}^{PL}_{g,\beta_n}(h(X))$ instead, which amounts to considering the following so-called Wins-PL estimator:

$$\widehat{R}^{\text{Wins}}_{g,\beta_n,t_n}(h(X)) = \int_0^1 h \circ \widehat{q}_n(\min(t_n, 1-(1-\beta_n)s)) dg(s).$$

When nt_n and $n(t_n - \beta_n)$ are positive integers it is easy to see that, contrary to the trimmed estimator, the winsorised estimator replaces the data points $X_{nt_n+1,n}, \ldots, X_{n,n}$ by $X_{nt_n,n}$. This estimator can of course also be written as a generalised L-statistic, viz.

$$\begin{aligned}
\widehat{R}^{\text{Wins}}_{g,\beta_n,t_n}(h(X)) &= h(X_{nt_n,n})\int_0^1 \mathbb{I}\{0 \le s < x_{n(1-t_n),n}(\beta_n,1)\}dg(s) \\
&+ \sum_{i=n(1-t_n)+1}^{n(1-\beta_n)} h(X_{n-i+1,n})\int_0^1 \mathbb{I}\{x_{i-1,n}(\beta_n,1) \le s < x_{i,n}(\beta_n,1)\}dg(s) \\
&+ h(X_{n\beta_n,n})\left[g(1) - \lim_{\substack{s\to 1\\ s<1}} g(s)\right] \\
&= \sum_{i=n(1-t_n)+1}^{n(1-\beta_n)} h(X_{n-i+1,n})\left[\lim_{\substack{s\to x_{i,n}(\beta_n,1)\\ s<x_{i,n}(\beta_n,1)}} g(s) - \lim_{\substack{s\to x_{i-1,n}(\beta_n,1)\\ s<x_{i-1,n}(\beta_n,1)}} g(s)\right] \\
&+ h(X_{n\beta_n,n})\left[g(1) - \lim_{\substack{s\to 1\\ s<1}} g(s)\right] + h(X_{nt_n,n})\lim_{\substack{s\to x_{n(1-t_n),n}(\beta_n,1)\\ s<x_{n(1-t_n),n}(\beta_n,1)}} g(s).
\end{aligned}$$

If g is continuous on $[0,1]$, this reads:

$$\widehat{R}^{\text{Wins}}_{g,\beta_n,t_n}(h(X)) = h(X_{n\beta_n+1,n}) + \sum_{i=n(1-t_n)+1}^{n(1-\beta_n)-1} g\left(\frac{i}{n(1-\beta_n)}\right)[h(X_{n-i+1,n}) - h(X_{n-i,n})].$$

Like the Trim-PL estimator, the Wins-PL estimator is a direct empirical estimator, of the quantity

$$R^{\text{Wins}}_{g,\beta_n,t_n}(h(X)) = \int_0^1 h\circ q(\min(t_n, 1-(1-\beta_n)s))dg(s),$$

which is actually essentially the Wang DRM of $h(X)$ given that X is larger than $q(\beta_n)$ and clipped above level $q(t_n)$:

Proposition 2.2.7. *Let $\beta \in (0,1)$ and $t \in (0,1]$ such that $t > \beta$. If q is continuous and strictly increasing on an open interval containing $[\beta,1)$ then:*

$$R^{\text{Wins}}_{g,\beta,t}(h(X)) = R_g(h(X^{\text{Wins}}_{\beta,t})) \quad \textit{with} \quad X^{\text{Wins}}_{\beta,t} \overset{d}{=} X\mathbb{I}\{q(\beta) \le X < q(t)\} + q(t)\mathbb{I}\{X \ge q(t)\}.$$

The focus in this section is to study the merits of trimming/winsorising in the context of the estimation of extreme Wang DRMs, both theoretically and at the finite-sample level. The crucial tool will, once again, be a Gaussian approximation of the tail quantile process.

Our final step prior to studying the Trim-PL and Wins-PL estimators is to highlight that they are actually part of a common class of estimators. For $0 < \beta < t \le 1$, let $\mathcal{F}(\beta,t)$ be the set of those nonincreasing Borel measurable functions ψ taking values in $[0,1]$ such that

$$\psi(0) = t,\ \psi(1) = \beta \ \text{ and } \ \forall s \in [0,1],\ 0 \le 1-(1-\beta)s - \psi(s) \le 1-t.$$

Let now (ψ_n) be a sequence of functions such that for all n, $\psi_n \in \mathcal{F}(\beta_n, t_n)$, and set

$$R_{g,\beta_n}(h(X);\psi_n) := \int_0^1 h\circ q\circ\psi_n(s)dg(s),$$

whose empirical counterpart is the estimator

$$\widehat{R}_{g,\beta_n}(h(X);\psi_n) = \int_0^1 h\circ\widehat{q}_n\circ\psi_n(s)dg(s) = \int_0^1 h(X_{\lceil n\psi_n(s)\rceil,n})dg(s).$$

All estimators in this class only take into account data points among the $X_{i,n}$, $\lceil n\beta_n\rceil \le i \le \lceil nt_n\rceil$, and can therefore be considered robust with respect to change in the most extreme values in the sample when $n(1-t_n) \geq 1$. The class of estimators $\widehat{R}_{g,\beta_n}(h(X);\psi_n)$ is a reasonable, unifying framework for our purpose: indeed, particular examples of the sequence (ψ_n) are $s \mapsto t_n-(t_n-\beta_n)s$ which appears as the argument of the empirical quantile function in the Trim-PL estimator, and $s \mapsto \min(t_n, 1-(1-\beta_n)s)$ giving rise to the Wins-PL estimator. These two examples should be those coming to mind when reading the asymptotic results below. Finally, the case $\psi_n(s) = 1-(1-\beta_n)s$, corresponding to the original PL estimator, is recovered by setting $t_n = 1$.

The first result below shows that $\widehat{R}_{g,\beta_n}(h(X);\psi_n)$ is a $\sqrt{n(1-\beta_n)}$–asymptotically normal estimator of $R_{g,\beta_n}(h(X))$ when h is a power function, under suitable conditions on β_n and t_n.

Theorem 2.2.8. *Assume that U satisfies condition $\mathcal{C}_2(\gamma,\rho,A)$. Assume further that (ψ_n) is a sequence of functions such that for all n, $\psi_n \in \mathcal{F}(\beta_n, t_n)$, with $0 < \beta_n < t_n \leq 1$, $\beta_n \to 1$, $n(1-\beta_n) \to \infty$ and $(1-t_n)/(1-\beta_n) \to 0$. Pick distortion functions $g_1,\ldots,g_d$ and $a_1,\ldots,a_d > 0$, and assume that for some $\eta > 0$, we have*

$$\forall j \in \{1,\ldots,d\}, \ \int_0^1 s^{-a_j\gamma-1/2-\eta} dg_j(s) < \infty, \quad \text{and} \quad \sqrt{n(1-t_n)}\left(\frac{1-t_n}{1-\beta_n}\right)^{\varepsilon} \to 0,$$

for some $\varepsilon \in (0, \min(1/2, \eta))$. If furthermore $\sqrt{n(1-\beta_n)}A((1-\beta_n)^{-1}) \to \lambda \in \mathbb{R}$, then:

$$\sqrt{n(1-\beta_n)}\left(\frac{\widehat{R}_{g_j,\beta_n}(X^{a_j};\psi_n)}{R_{g_j,\beta_n}(X^{a_j})} - 1\right)_{1\leq j\leq d} \xrightarrow{d} \mathcal{N}(0, V),$$

with the matrix V as in Theorem 2.2.3.

The estimators $\widehat{R}_{g,\beta_n}(X^a;\psi_n)$ thus share the same limiting Gaussian distribution under the classical bias condition $\sqrt{n(1-\beta_n)}A((1-\beta_n)^{-1}) \to \lambda \in \mathbb{R}$, when hypothesis $\sqrt{n(1-t_n)}[(1-t_n)/(1-\beta_n)]^{\varepsilon} \to 0$, relating the order t_n to the intermediate level β_n, holds true. This condition implies that t_n should converge to 1 quickly enough, or, in other words, that not too many values should be deleted from the sample for asymptotic unbiasedness to hold. The necessity of such a condition appears in the earlier works of [80, 83] in the context of mean estimation by the trimmed sample mean: in the former paper, it is shown that discarding a fixed number of order statistics does not create asymptotic bias, while the latter paper states that this may not be true for more severe trimmings. It should be noted that the present assumption is clearly satisfied for $t_n = 1-c/n$, with c being a fixed nonnegative integer, corresponding to the case when the top c order statistics are discarded and the trimming/winsorising percentage across the whole sample is $100c/n\%$. Finally, taking $t_n = 1$ in Theorem 2.2.8 yields Theorem 2.2.3.

The next result, analogously to Corollary 2.2.4, sums up what can be said when g belongs to a space $\mathcal{E}_b[0,1]$ for $b > -1$.

Corollary 2.2.9. *Assume that U satisfies condition $\mathcal{C}_2(\gamma,\rho,A)$. Assume further that (ψ_n) is a sequence of functions such that for all n, $\psi_n \in \mathcal{F}(\beta_n,t_n)$, with $0 < \beta_n < t_n \leq 1$, $\beta_n \to 1$, $n(1-\beta_n) \to \infty$ and $(1-t_n)/(1-\beta_n) \to 0$. Pick distortion functions $g_1,\ldots,g_d$ and $a_1,\ldots,a_d > 0$. Assume there are $b_1,\ldots,b_d > -1$ such that $g_j \in \mathcal{E}_{b_j}[0,1]$ for all $j \in \{1,\ldots,d\}$. If $\gamma < (2b_j+1)/(2a_j)$ for all $j \in \{1,\ldots,d\}$ and*

$$\sqrt{n(1-\beta_n)}A((1-\beta_n)^{-1}) \to \lambda \in \mathbb{R} \ \text{and} \ \sqrt{n(1-t_n)}\left(\frac{1-t_n}{1-\beta_n}\right)^{\varepsilon} \to 0,$$

for some $\varepsilon \in (0, \min(0, b_1-a_1\gamma,\ldots,b_d-a_d\gamma)+1/2)$ then:

$$\sqrt{n(1-\beta_n)}\left(\frac{\widehat{R}_{g_j,\beta_n}(X^{a_j};\psi_n)}{R_{g_j,\beta_n}(X^{a_j})} - 1\right)_{1\leq j\leq d} \xrightarrow{d} \mathcal{N}(0, V),$$

with the matrix V as in Theorem 2.2.3.

2.2.4.3 Second step of improvement: Finite-sample bias correction

The estimators introduced above have been shown to be asymptotically normal estimators of Wang DRMs. It should be noted that on finite-sample situations, such estimators can be expected to carry some (negative) bias, all the more so as the trimming/winsorising order t_n increases. An intuitive justification for this behaviour is that the estimator $\widehat{R}_{g,\beta_n}(X^a;\psi_n)$ is actually the empirical counterpart of $R_{g,\beta_n}(X^a;\psi_n)$, which is in general different from, and especially less than, the target DRM $R_{g,\beta_n}(X^a)$. For instance, in the case of extreme Tail-Value-at-Risk estimation for the Pareto distribution with tail index γ, then the Tail-Value-at-Risk of X in the worst $100(1-\beta_n)\%$ cases is

$$R_{g,\beta_n}(X) = \frac{(1-\beta_n)^{-\gamma}}{1-\gamma},$$

see Section 2.2.4.1. By contrast, the trimmed Tail-Value-at-Risk given that X lies between levels $q(\beta_n)$ and $q(t_n)$ is obtained for $\psi_n(s) = t_n - (t_n-\beta_n)s$ and is

$$\begin{aligned} R^{\text{Trim}}_{g,\beta_n,t_n}(X) = \int_0^1 q(t_n-(t_n-\beta_n)s)dg(s) &= \int_0^1 [1-t_n+(t_n-\beta_n)s]^{-\gamma}ds \\ &= \frac{(1-\beta_n)^{1-\gamma}-(1-t_n)^{1-\gamma}}{(1-\gamma)(t_n-\beta_n)}. \end{aligned}$$

Rewriting this as

$$R^{\text{Trim}}_{g,\beta_n,t_n}(X) = R_{g,\beta_n}(X)\left\{\frac{(1-\beta_n)^{1-\gamma}-(1-t_n)^{1-\gamma}}{(1-\beta_n)^{-\gamma}(t_n-\beta_n)}\right\},$$

results in an expression of $R^{\text{Trim}}_{g,\beta_n,t_n}(X)$ as $R_{g,\beta_n}(X)$ multiplied by a quantity depending on β_n, t_n and γ and smaller than 1. In the case $n = 1000$, $\beta_n = 0.9$, $t_n = 0.99$ and $\gamma = 1/2$, namely the top 100 observations are selected and the top 10 observations among them are eliminated, the reduction factor is actually 0.760, i.e. the expected relative bias is -0.240. When the number of observations removed is halved ($t_n = 0.995$) this factor becomes 0.817 for an expected relative bias of -0.183. The smallest trimming percentage, obtained when $t_n = 0.999$, for removal of the sample maximum only, results in a reduction factor of 0.909, which is still an expected relative bias of -0.091.

To retain the reduction in variability brought by the estimator $\widehat{R}_{g,\beta_n}(X;\psi_n)$ and at the same time obtain an estimator with acceptable finite-sample bias, we design a new estimator based on the previous calculation. More precisely, in the case of Tail-Value-at-Risk estimation, estimating γ by a consistent estimator $\widehat{\gamma}_n$ and plugging in the previous estimator $\widehat{R}_{g,\beta_n}(X;\psi_n) = \widehat{R}^{\text{Trim}}_{g,\beta_n,t_n}(X)$ in the left-hand side of the above equality gives the corrected estimator

$$\widetilde{R}_{g,\beta_n}(X;\psi_n) = \widehat{R}_{g,\beta_n}(X;\psi_n)\left\{\frac{(1-\beta_n)^{1-\widehat{\gamma}_n}-(1-t_n)^{1-\widehat{\gamma}_n}}{(1-\beta_n)^{-\widehat{\gamma}_n}(t_n-\beta_n)}\right\}^{-1}.$$

Note that the correction factor is in fact

$$\left\{\frac{(1-\beta_n)^{1-\widehat{\gamma}_n}-(1-t_n)^{1-\widehat{\gamma}_n}}{(1-\beta_n)^{-\widehat{\gamma}_n}(t_n-\beta_n)}\right\}^{-1} = \frac{R_{g,\beta_n}(Y_{\widehat{\gamma}_n})}{R_{g,\beta_n}(Y_{\widehat{\gamma}_n};\psi_n)},$$

where Y_γ has a Pareto distribution with tail index γ.

Of course, in practice the underlying distribution of X is not known, but in many cases the Pareto distribution (or a multiple of it) still provides a decent approximation for X in its right tail. It can thus be expected that in a wide range of situations and for n large enough,

$$R_{g,\beta_n}(X^a) = R_{g,\beta_n}(X^a;\psi_n)\frac{R_{g,\beta_n}(X^a)}{R_{g,\beta_n}(X^a;\psi_n)} \approx R_{g,\beta_n}(X^a;\psi_n)\frac{R_{g,\beta_n}(Y_{a\gamma})}{R_{g,\beta_n}(Y_{a\gamma};\psi_n)}.$$

This motivates the following class of corrected estimators:

$$\widetilde{R}_{g,\beta_n}(X^a;\psi_n) = \widehat{R}_{g,\beta_n}(X^a;\psi_n)\frac{R_{g,\beta_n}(Y_{a\widehat{\gamma}_n})}{R_{g,\beta_n}(Y_{a\widehat{\gamma}_n};\psi_n)} = \widehat{R}_{g,\beta_n}(X^a;\psi_n)\frac{\int_0^1[(1-\beta_n)s]^{-a\widehat{\gamma}_n}dg(s)}{\int_0^1[1-\psi_n(s)]^{-a\widehat{\gamma}_n}dg(s)}.$$

This estimator should be seen as the result of a two-stage procedure:

- first, compute an estimator of the target extreme Wang DRM using a trimmed/winsorised sample, thus reducing variability;
- then, use what can be found on the tail behaviour of the sample to shift the previous estimate back to an essentially bias-neutral position.

Let us emphasise that this bias-correction procedure is a simple one, much closer in spirit to the construction of the corrected sample variance estimator when the population mean is unknown than to bias-reduction methods based on asymptotic results in a second-order extreme value framework, of which an excellent summary is Section 5.3 in [138]. In particular, the multiplicative correction factor introduced here only depends on the tail index γ, but not on the second-order parameter ρ. Finally, note that the correction factor might depend on the top values in the sample, but can only actually do so through the estimator $\widehat{\gamma}_n$. For instance, the Hill estimator of γ,

$$\widehat{\gamma}_{\beta_n} = \frac{1}{\lceil n(1-\beta_n)\rceil} \sum_{i=1}^{\lceil n(1-\beta_n)\rceil} \log\left(X_{n-i+1,n}\right) - \log\left(X_{n-\lceil n(1-\beta_n)\rceil,n}\right),$$

or its bias-reduced versions considered in the simulation study below depend on the top values only through their logarithms, which sharply reduces their contribution to the variability of our final estimator.

The next result shows that any member of this new class of corrected estimators shares the asymptotic properties of its uncorrected version. Our preference shall thus be driven by finite-sample considerations.

Theorem 2.2.10. *Assume that U satisfies condition $\mathcal{C}_2(\gamma,\rho,A)$. Assume further that (ψ_n) is a sequence of functions such that for all n, $\psi_n \in \mathcal{F}(\beta_n, t_n)$, with $0<\beta_n<t_n\le 1$, $\beta_n\to 1$, $n(1-\beta_n)\to\infty$ and $(1-t_n)/(1-\beta_n)\to 0$. Pick distortion functions $g_1,\dots,g_d$ and $a_1,\dots,a_d>0$, and assume that for some $\eta>0$, we have*

$$\forall j\in\{1,\dots,d\},\ \int_0^1 s^{-a_j\gamma-1/2-\eta}\,dg_j(s)<\infty, \quad \textit{and} \quad \sqrt{n(1-t_n)}\left(\frac{1-t_n}{1-\beta_n}\right)^{\varepsilon}\to 0,$$

for some $\varepsilon\in(0,\min(1/2,\eta))$. If furthermore

$$\sqrt{n(1-\beta_n)}A((1-\beta_n)^{-1})\to\lambda\in\mathbb{R} \quad \textit{and} \quad \sqrt{n(1-\beta_n)}(\widehat{\gamma}_n-\gamma)=\mathrm{O}_{\mathbb{P}}(1),$$

then:

$$\forall j\in\{1,\dots,d\},\ \sqrt{n(1-\beta_n)}\left(\frac{\widetilde{R}_{g_j,\beta_n}(X^{a_j};\psi_n)}{\widehat{R}_{g_j,\beta_n}(X^{a_j};\psi_n)}-1\right)\xrightarrow{\mathbb{P}}0,$$

and therefore

$$\sqrt{n(1-\beta_n)}\left(\frac{\widetilde{R}_{g_j,\beta_n}(X^{a_j};\psi_n)}{R_{g_j,\beta_n}(X^{a_j})}-1\right)_{1\le j\le d}\xrightarrow{d}\mathcal{N}(0,V),$$

with the matrix V as in Theorem 2.2.8.

2.2.4.4 Adaptation to extreme levels

Like in Sections 2.2.3.2 and 2.2.3.3, the empirical estimators developed so far in Section 2.2.4 only work provided β_n is an intermediate level, namely $n(1-\beta_n)\to\infty$. The next and final step is to design an estimator working for arbitrarily extreme levels as well. The construction mimics that of Section 2.2.3.4: suppose that $n(1-\delta_n)\to c<\infty$, take a sequence (β_n) such that $n(1-\beta_n)\to\infty$ and define

$$\widetilde{R}^W_{g,\delta_n}(X^a;\psi_n) := \left(\frac{1-\beta_n}{1-\delta_n}\right)^{a\widehat{\gamma}_n}\widetilde{R}_{g,\beta_n}(X^a;\psi_n),$$

where $\widehat{\gamma}_n$ is the consistent estimator of γ appearing in $\widetilde{R}_{g,\beta_n}(X^a;\psi_n)$. This is again a Weissman-type estimator of $R_{g,\delta_n}(X^a)$, see [256]. Weissman's estimator is actually recovered for $a=1$, $t_n=1$ and $g(s)=0$ if $s<1$, and the extrapolated PL estimator of Section 2.2.3.4 is obtained for $t_n=1$. The third and final main result examines the asymptotic distribution of this class of extrapolated estimators.

Theorem 2.2.11. *Assume that U satisfies condition $\mathcal{C}_2(\gamma,\rho,A)$, with $\rho<0$. Assume further that (ψ_n) is a sequence of functions such that for all n, $\psi_n\in\mathcal{F}(\beta_n,t_n)$, with $0<\beta_n<t_n\le 1$, $\beta_n\to 1$, $n(1-\beta_n)\to\infty$ and $(1-t_n)/(1-\beta_n)\to 0$; let finally a sequence $\delta_n\to 1$ be such that $(1-\delta_n)/(1-\beta_n)\to 0$ and $\sqrt{n(1-\beta_n)}/\log[(1-\beta_n)/(1-\delta_n)]\to\infty$. Pick now distortion functions $g_1,\ldots,g_d$ and $a_1,\ldots,a_d>0$, and assume that for some $\eta>0$, we have*

$$\forall j\in\{1,\ldots,d\},\ \int_0^1 s^{-a_j\gamma-1/2-\eta}dg_j(s)<\infty \quad \text{and} \quad \sqrt{n(1-t_n)}\left(\frac{1-t_n}{1-\beta_n}\right)^{\varepsilon}\to 0,$$

for some $\varepsilon\in(0,\min(1/2,\eta))$. If furthermore

$$\sqrt{n(1-\beta_n)}A((1-\beta_n)^{-1})\to\lambda\in\mathbb{R} \quad \text{and} \quad \sqrt{n(1-\beta_n)}(\widehat{\gamma}_n-\gamma)\xrightarrow{d}\xi,$$

then:

$$\frac{\sqrt{n(1-\beta_n)}}{\log([1-\beta_n]/[1-\delta_n])}\left(\frac{\widetilde{R}^W_{g_j,\delta_n}(X^{a_j};\psi_n)}{R_{g_j,\delta_n}(X^{a_j})}-1\right)_{1\le j\le d}\xrightarrow{d}\begin{pmatrix} a_1\xi \\ \vdots \\ a_d\xi\end{pmatrix}.$$

Again, in the case $t_n=1$, we recover the asymptotic normality result of [14] for the class of extrapolated PL estimators. Our robust extreme risk measure estimators have therefore got the same asymptotic distribution as the original PL estimator, under the same technical conditions. It can thus be concluded that considering trimmed/winsorised estimators results in a generalisation of the existing theory of estimators of extreme Wang DRMs. The next section shall show that the trimmed/winsorised versions are of value when the underlying distribution has a very heavy tail.

2.2.5 Finite-sample study [11, 14]

2.2.5.1 Finite-sample performance on simulated data

The finite-sample performance of our extrapolated estimators is examined on simulation studies in [11, 14], where we consider a couple of classical continuous heavy-tailed distributions and three different distortion functions g:

- the expectation function $g(x)=x$. Under our construction, the associated distortion function g_β generates the Conditional Tail Expectation (CTE, equivalently TVaR) above level $q(\beta)$;
- the Dual Power (DP) function $g(x)=1-(1-x)^{1/\alpha}$ with $\alpha\in(0,1)$. When $r:=1/\alpha$ is a positive integer, the related DRM is the expectation of $\max(X_1,\ldots,X_r)$ for independent copies $X_1,\ldots,X_r$ of X,
- and the Proportional Hazard (PH) transform function $g(x)=x^\alpha$, $\alpha\in(0,1)$. When $c:=1/\alpha$ is a positive integer, the related DRM is the expectation of a random variable Y whose distribution is such that X has the same distribution as $\min(Y_1,\ldots,Y_c)$ for independent copies $Y_1,\ldots,Y_c$ of Y, see [75].

The level β_n, and the truncation level t_n for the CTrim-PL and CWins-PL versions of the PL estimator, are chosen based on stability region arguments whose details can be found in [11, 14]. Our results show that:

- For examples with moderately large values of γ, the PL estimator performs at least as well as the AE estimator. Besides, the PL estimator performs markedly better than the AE estimator for smaller samples or when the risk measure puts more weight on the highest order statistics. See Section 4 in [14].

- For examples with larger values of γ but such that Theorem 2.2.11 holds, the proposed CTrim-PL and CWins-PL estimators perform slightly worse in terms of bias than the original PL estimator. This is not surprising: the correction method, based on an approximation of the upper tails of the underlying distribution by a purely Pareto tail, cannot be expected to recover all the information the (highly variable) top order statistics carry about the extremes of the sample. By contrast, the CTrim-PL and CWins-PL estimators perform essentially comparably to or better than the standard empirical extreme Wang DRM estimator in terms of MSE; for extreme CTE estimation with values of γ close to but less than $1/2$, the improvement is close to up to 40% in the case of the Fréchet distribution. The performance of the trimmed and winsorised estimators relatively to the PL estimator deteriorates, however, when ρ gets closer to 0. This is not surprising either, since the correction step is based on an approximation of the right tail of the underlying distribution by the right tail of (a multiple of) a Pareto distribution, which gets less accurate as $|\rho|$ decreases. See Section 4.1 in [11].

- For examples with yet heavier tails, when Theorem 2.2.11 fails to hold, the CTrim-PL and CWins-PL estimators perform similarly but both represent overall a substantial improvement over the PL estimator especially in samples featuring at least one unusually large observation. See Section 4.2 in [11].

2.2.5.2 Data example 1: The Secura Belgian Re actuarial data set [14]

We consider here the Secura Belgian Re data set on automobile claims from 1998 until 2001, introduced in [47] and further analysed in [249] from the extreme value perspective. The data set consists of $n = 371$ claims which were at least as large as 1.2 million Euros and were corrected for inflation.

We start by estimating the tail index γ. We use the Hill estimator and a bias-reduced version inspired by the work of [212]:

$$\widehat{\gamma}_{\beta}^{RB}(\tau) = \frac{1}{\widehat{\rho}_{\beta_1}(\tau)}\widehat{\gamma}_{\beta} + \left(1 - \frac{1}{\widehat{\rho}_{\beta_1}(\tau)}\right)\frac{\widehat{\gamma}_{\beta}^{S}}{2\widehat{\gamma}_{\beta}},$$

with

$$\widehat{\gamma}_{\beta}^{S} = \frac{1}{\lceil n(1-\beta)\rceil}\sum_{i=1}^{\lceil n(1-\beta)\rceil}\left(\log X_{n-i+1,n} - \log X_{n-\lceil n(1-\beta)\rceil,n}\right)^2$$

and $\widehat{\rho}_{\beta_1}(\tau)$ is the consistent estimator of ρ presented in Equation (2.18) of [122], which depends on a different sample fraction $1-\beta_1$ and a tuning parameter $\tau \geq 0$. Theorem 2.1 in [212] states the asymptotic distribution of $\widehat{\gamma}_{\beta_n}^{RB}$, which can then be used to construct asymptotic confidence intervals. We take $\beta_1 = 1 - \lceil n^{0.975}\rceil/n \approx 0.0882$, as recommended by [66]. Results using the stability region selection procedure presented in Section 5.2 of [14] are given in Table 2.3. Retaining the median estimate of γ yields $\widehat{\gamma} = 0.261$ for $\beta^* = 0.792$ and $\tau = 1/2$, with $\widehat{\rho} = -1.064$. Table 2.4 gives estimates of some risk measures for this data set.

The main example of excess-of-loss reinsurance policy that [249] considered, namely the net premium principle, can actually be recovered from these estimates. Indeed, according to [249], the net premium $\mathrm{NP}(R)$ for a reinsurance policy in excess of a high retention level R is

$$\mathrm{NP}(R) = \int_R^{\infty}[1-F(x)]dx.$$

Rearranging Equation (2.2.2) and setting $g(x) = x$ gives the identity

$$\mathrm{NP}(q(\beta)) = (1-\beta)(R_{g,\beta}(X) - \mathrm{VaR}(\beta))$$

and, in particular, the right-hand side is actually $\mathrm{SP}(\beta)$. When R is equal to 5 million Euros, as considered in [249], it can be seen that the exceedance probability $\mathbb{P}(X > R)$ is estimated to be approximately 0.02, or in other words that R is essentially the estimated VaR at the 98% level. Estimates of our risk measures at this level are provided in Table 2.4; in particular, the net premium is estimated to be approximately 36,000 Euros, which is in line with the 41,798 Euros that [249] obtained, with our estimate being slightly lower partly because a bias-reduced estimate of γ was used in the present work, whereas [249] computed a simple Hill estimate.

Estimator $\widehat{\gamma}$	β^*	Estimate
Hill	0.854	0.292
Bias-reduced, $\tau = 1$	0.782	0.263
Bias-reduced, $\tau = 3/4$	0.792	0.262
Bias-reduced, $\tau = 1/2$	0.792	0.261
Bias-reduced, $\tau = 1/4$	0.792	0.260
Bias-reduced, $\tau = 0$	0.792	0.258

Table 2.3: Secura Belgian Re data set: estimates of the tail index γ.

δ	Estimator	$\widehat{\text{VaR}}$	$\widehat{\text{CTE}}$	$\widehat{\text{SP}}$
0.98	AE	4989 $[3505, 6473]$	6750 $[4742, 8758]$	35.220 $[24.744, 45.696]$
	PL	4989 $[3505, 6473]$	6864 $[4822, 8906]$	37.500 $[26.346, 48.654]$
0.99	AE	5978 $[3673, 8283]$	8087 $[4969, 11205]$	21.092 $[12.960, 29.224]$
	PL	5978 $[3673, 8283]$	8224 $[5053, 11395]$	22.459 $[13.800, 31.118]$
0.995	AE	7163 $[3770, 10556]$	9690 $[5100, 14280]$	12.636 $[6.6506, 18.621]$
	PL	7163 $[3770, 10556]$	9854 $[5186, 14522]$	13.455 $[7.0817, 19.828]$
0.999	AE	10899 $[3506, 18291]$	14744 $[4743, 24745]$	3.8452 $[1.2371, 6.4533]$
	PL	10899 $[3506, 18292]$	14993 $[4823, 25163]$	4.0944 $[1.3172, 6.8716]$

Table 2.4: Secura Belgian Re data set: extreme risk measure estimation (measurement unit: thousands of Euros). Between square brackets: asymptotic 95% confidence intervals.

2.2.5.3 Data example 2: French commercial fire losses [11]

We consider data on $n = 1{,}098$ commercial fire losses recorded between 1 January 1995 and 31 December 1996 by the FFSA (an acronym for the *Fédération Française des Sociétés d'Assurance*), available from the R package `CASdatasets` by prompting `data(frecomfire)`. The data, originally recorded in French francs, is converted into euros, and denoted by $(X_1, \ldots, X_n)$.

The first step in the analysis of this data set is again to estimate the tail index γ. To this end, we use a slightly different bias-reduced version of the Hill estimator, introduced in [65] and implemented in the function `mop` of the R package `evt0` (further discussed in [137]). The sample fraction chosen to compute the tail index, following the stability region procedure presented in Section 4 of [11] is then $1 - \beta^* \approx 0.120$, for an estimate $\widehat{\gamma}_{\beta^*} \approx 0.697$. This suggests a very heavy tail, in the sense that $\widehat{\gamma}_{\beta^*} > 1/2$ and therefore the underlying distribution seems to have an infinite variance. In particular, we know that this may adversely affect the PL estimator of the extreme TVaR and of the extreme DP risk measure, which justifies comparing the PL estimates to those obtained using our CTrim-PL and CWins-PL estimators.

We then compute, at the extreme level $\delta = 0.999 \approx 1 - n^{-1}$, the PL, CTrim-PL and CWins-PL estimators of the extreme TVaR and DP(1/3) risk measures, using the parameter selection procedure of Section 4 of [11]. Results are summarised in Table 2.5. It is not clear, from these results, which estimator should be chosen. Our goal is now to offer some insight into this choice, using the mean

	Estimator	Order t^*	Number of top order statistics cut	Estimate (in euros)
TVaR estimation	PL	N/A	N/A	225,122,925
	CTrim-PL	0.9736	29	219,814,856
	CWins-PL	0.9909	10	208,538,799
DP(1/3) estimation	PL	N/A	N/A	459,285,394
	CTrim-PL	0.9663	37	452,920,888
	CWins-PL	0.9863	15	404,498,511

Table 2.5: French commercial fire losses data set: estimating some risk measures in the case $\delta = 0.999$.

excess plot of the $n(1-\beta^*) = 132$ data points used in the present analysis. The rationale behind the use of the mean excess plot, i.e. the plot of the function

$$u \mapsto \frac{\sum_{i=1}^n (X_i - u)\mathbb{I}\{X_i > u\}}{\sum_{i=1}^n \mathbb{I}\{X_i > u\}},$$

is that its empirical counterpart $u \mapsto \mathbb{E}(X - u|X > u)$ is linearly increasing when $0 < \gamma < 1$ and X has a Generalised Pareto distribution [89]. Therefore, the extremes of the data set should be indicated by a roughly linear part at the right of the mean excess plot. This plot can be tricky to use though: apart from the choice of the lower threshold u above which the mean excess function is computed (which is here chosen to be $X_{n\beta^*,n}$), it has been observed that the mean excess function has very often a non-linear behaviour at the right end of the mean excess plot [131]. This is again because the top order statistics in the sample suffer from a very high variability, and as a consequence the mean excess function is, in its right end, averaging over just a few high-variance values. In other words, the intermediate, roughly linear part of the plot indicates which ones among the top data points can be trusted from the points of view of both bias and variability, and the unstable part at the right end of the mean excess plot represents those highly variable values that may be cut from the analysis using the CTrim-PL and CWins-PL estimators.

We then plot on Figure 2.1 copies of the mean excess plot above the value $u = X_{n\beta^*,n}$ where the values cut from the analysis by the CTrim-PL and CWins-PL estimators are highlighted. The least squares line related to the data points kept for the analysis is also represented. It can be seen on these plots that there is indeed an unstable part at the right end of the plot, which suggests to use either the CTrim-PL or CWins-PL estimator in order to gain some stability. The linear adjustment for the selected data points (using the t_n selection technique presented in Section 4 of [11]) is also reasonable in all cases. It is arguable though that the CTrim-PL estimator is too conservative in the sense that the number of data points it discards is high: in the DP case in particular, the estimator trims 37 top order statistics, which is 29% of the available data above the selected threshold $X_{n\beta^*,n}$. The CWins-PL estimator discards much less data points (less than half of what the CTrim-PL estimator discards, see also Table 2.5), and therefore does not have to compensate for the loss of information this entails as much as the CTrim-PL has to, while the linear adjustment of the least squares line is still perfectly acceptable. It can be argued then that the CWins-PL estimator is preferable here, both for extreme TVaR and DP risk measure estimation. The estimates it yields are appreciably lower (roughly 10% less) than the standard PL estimates, and this makes us think that the extreme TVaR and DP risk measure are actually overestimated by the PL estimator. The conclusion is that, using the CWins-PL estimator, the average loss in the worst 0.1% of cases is estimated to be 208.5 million euros, and the average value of the maximal loss recorded after three extreme fires (i.e. each belonging to the worst 0.1% of fires) to be 404.5 million euros.

2.2.6 Perspectives for future research

Inference theory for very heavy tails Substantial improvements are obtained in practice by using the CTrim-PL and CWins-PL estimators in very heavy-tailed cases, that is, when γ is so large that the estimators are still consistent but Theorem 2.2.11 does not apply. In the case of the Conditional Tail Expectation, this represents the case $1/2 < \gamma < 1$ of an infinite variance but a finite expectation.

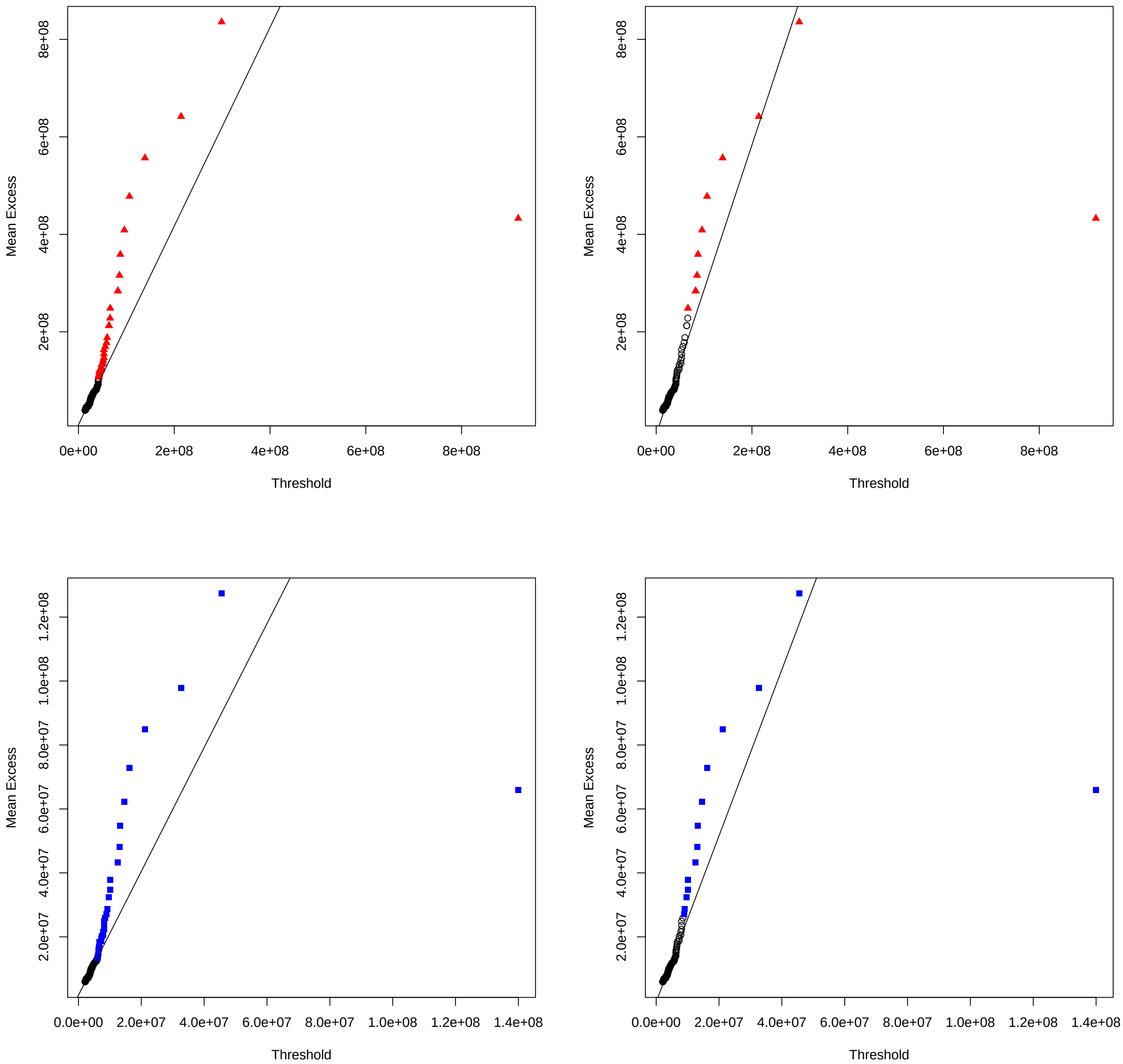

Figure 2.1: French commercial fire losses data set: mean excess plots. Top: the mean excess plot of the top $n(1-\beta^*) = 132$ data points, where data points excluded from the TVaR estimation when using the CTrim-PL and CWins-PL estimators are highlighted using red triangles; left: CTrim-PL estimator, right: CWins-PL estimator. Bottom: the mean excess plot of the top $n(1-\beta^*) = 132$ data points, where data points excluded from the DP(1/3) estimation when using the CTrim-PL and CWins-PL estimators are highlighted using blue squares; left: CTrim-PL estimator, right: CWins-PL estimator. In all four cases the straight line is the least squares line for the set of black data points.

No theoretical results on the asymptotic distribution of our estimators are currently available in such cases. By analogy with the estimation of a mean when the variance of the underlying distribution is infinite, one might expect a reduced rate of convergence and an asymptotic distribution which is a function of a stable distribution (see e.g. Section XVII of [119]). This is worth pursuing, for instance to be able to construct asymptotic confidence intervals for extreme Wang DRMs in cases when the estimate of γ is too high to apply Theorem 2.2.11.

An improved correction step The correction step implemented in the CTrim-PL and CWins-PL estimators is a simple one, based on an asymptotic equivalent of the ratio of a trimmed/winsorised extreme Wang DRM and of its full counterpart. Especially, this correction method does not take into account second-order information. This is why the CTrim-PL and CWins-PL methods should not be expected to show an improved finite-sample performance compared to that of the basic empirical plug-in estimator when ρ is close to 0. It would be very interesting to design another correction factor taking into account second-order information, in order to close the gap between the finite-sample performance of the proposed technique and that of the full PL estimator in standard cases with low $|\rho|$, and retain or even improve its finite-sample performance further in difficult cases. Two reasons why this is a difficult problem are that:

- Estimators of ρ typically have a rate of convergence lower than that of typical tail index estimators, see e.g. p.2638 in [135] and p.298 in [140]. This suggests that estimators of the second-order parameter are in general quite volatile;
- Tail index estimators tend to have a poor finite-sample behaviour for low $|\rho|$, be it because of their bias if they are not bias-corrected, or of their increased asymptotic variance if they are bias-corrected.

Multiplying the Trim-PL or Wins-PL estimators by a correction factor adapted to low values of $|\rho|$ might therefore, if not done carefully, entail multiplying by a highly variable quantity and ultimately wipe out part of or all that was gained in terms of variability from using the trimming/winsorising scheme. The problem of constructing a second-order-adapted correction factor is therefore challenging and worthy of future research.

2.3 Inference methods for extreme expectile-based risk measures [1, 12]

2.3.1 Introduction

The class of Wang DRMs is of course not the only way to encompass the class of quantiles. An alternative way to do so, which has proven very useful for example in regression contexts, is using L^1-minimisation: given an order $\tau \in (0,1)$, it is shown in [182] that the τth quantile of the distribution of a random variable Y, denoted throughout this section by q_τ, can be obtained by minimising asymmetrically weighted mean absolute deviations, as

$$q_\tau \in \underset{\theta \in \mathbb{R}}{\arg\min}\, \mathbb{E}\left\{\rho_\tau(Y-\theta) - \rho_\tau(Y)\right\},$$

with equality if the distribution function of Y is increasing, where $\rho_\tau(y) = |\tau - \mathbb{I}\{y \leq 0\}|\,|y|$ is the so-called quantile check function. This successfully extends the conventional definition of quantiles as left-continuous inverse functions. The concept of expectiles is a least squares analogue of quantiles, which summarises the underlying distribution of a random variable Y in much the same way that quantiles do. It was introduced by [209] by substituting the L^1 loss function with the L^2 loss to define the population expectile of order $\tau \in (0,1)$ as the minimiser

$$\xi_\tau = \underset{\theta \in \mathbb{R}}{\arg\min}\, \mathbb{E}\left\{\eta_\tau(Y-\theta) - \eta_\tau(Y)\right\}, \tag{2.3.1}$$

where $\eta_\tau(y) = |\tau - \mathbb{I}\{y \leq 0\}|\, y^2$. Although formulated using a quadratic loss, problem (2.3.1) is well-defined as soon as $\mathbb{E}|Y|$ is finite, thanks to the presence of the term $\eta_\tau(Y)$. Expectiles are a natural generalisation of the usual mean $\mathbb{E}(Y)$ just as the class of quantiles generalises the median. Both expectiles and quantiles are useful descriptors of the higher and lower regions of the data points in the same way as the mean and median are related to their central behaviour. An advantage of using expectiles is that their estimation makes more efficient use of the available data, since weighted least squares rely on the distance to data points, while empirical quantiles only utilise the information on whether an observation is below or above the predictor [38, 209, 235]. Furthermore, unlike sample quantiles, sample expectiles provide a class of smooth curves as functions of the level τ, see e.g. [226].

Our focus here is to discuss the estimation of extreme expectiles, and more generally expectile-based alternatives to popular risk measures such as the Expected Shortfall (ES) and Marginal Expected Shortfall (MES). Motivating advantages are that expectiles are more alert than quantiles to the magnitude of infrequent catastrophic losses, as they depend on both the tail realisations of Y and their probability, while quantiles only depend on the frequency of tail realisations [190]. Most importantly, from the point of view of the axiomatic theory of risk measures, [53] shows that the only M-quantiles that are coherent risk measures are the expectiles above the mean. Very recently, [265] has proved that expectiles are the only coherent law-invariant measure of risk which is also elicitable. Elicitability corresponds to the existence of a natural backtesting methodology; in comparison, it has been shown that the ES, although a popular coherent risk measure, is not elicitable [134], but jointly elicitable with VaR [120].

In terms of interpretability, the τ-quantile is the point below which $100\tau\%$ of the mass of Y lies, while the τ-expectile specifies the position ξ_τ such that the average distance from the data below ξ_τ to ξ_τ itself is $100\tau\%$ of the average distance between ξ_τ and all the data:

$$\tau = \mathbb{E}\left(|Y - \xi_\tau|\mathbb{I}\{Y \leq \xi_\tau\}\right)/\mathbb{E}\,|Y - \xi_\tau|\,. \tag{2.3.2}$$

Thus the τ-expectile shares an interpretation similar to the τ-quantile, replacing the number of observations by the distance. In a financial context, [52] provide a transparent financial meaning of expectiles in terms of the gain-loss ratio, which is a popular performance measure in portfolio management and is well-known in the literature on no good deal valuation in incomplete markets (see [52] and references therein). Also, [109] shows that expectiles are optimal decision thresholds in certain binary investment problems. The theoretical and numerical results obtained by [52] seem to indicate that expectiles are perfectly viable alternatives to standard VaR and ES.

On the statistical side, and although least asymmetrically weighted squares estimation of expectiles dates back to [209] in the linear regression framework, it has recently regained growing interest in the context of nonparametric, semiparametric and more complex models, see for example [235] and the references therein, as well as the two recent contributions [157, 188] for advanced theoretical developments. In the statistical literature, attention has been, however, restricted to ordinary expectiles of fixed order τ staying away from the tails of the underlying distribution. In the latter two references, for example, several powerful asymptotic results such as uniform consistency and a uniform central limit theorem are shown for expectile estimators, but the order τ is assumed therein to lie within a fixed interval bounded away from 0 and 1. In [1, 12], our focus is to extend the estimation of expectiles and their asymptotic theory far enough into the tails. This translates into considering the expectile level $\tau = \tau_n \to 0$ or $\tau_n \to 1$ as the sample size n goes to infinity. This has been initiated at the population level by [52, 53, 198, 199], which studied the connection of extreme population expectiles with their quantile analogues when Y belongs to the domain of attraction of a Generalised Extreme Value distribution. The literature has not, however, considered the question of their inference. This is precisely our focus here.

Let us point out the main contributions of [1, 12]. First, we estimate the intermediate tail expectiles of order $\tau_n \to 1$ such that $n(1-\tau_n) \to \infty$, and then extrapolate these estimates to the very extreme expectile levels τ_n' which approach one at an arbitrarily fast rate in the sense that $n(1-\tau_n') \to c$, for some nonnegative constant c. Two such estimation methods are considered. One is indirect, based on the use of asymptotic approximations involving intermediate quantiles, and the other relies directly on least asymmetrically weighted squares (LAWS) estimation. Second, we provide adapted extreme expectile-based tools for the estimation of the MES, an important factor when measuring the systemic risk of financial firms. Denoting by X and Y, respectively, the loss on the return of a financial firm and that of the entire market, the MES is equal to $\mathbb{E}(X|Y > t)$, where t is a high threshold reflecting a systemic crisis, i.e. a substantial market decline. For an extreme expectile $t = \xi_{\tau_n'}$ and for a wide nonparametric class of bivariate distributions of (X, Y), we construct two estimators of the MES. A competing procedure introduced by [68] is based on extreme quantiles. Third, and with a motivation to estimate expectile-based ES risk measures at extreme levels, we show that the aforementioned convergence result on the LAWS estimator of an intermediate expectile can be generalised to what we shall call the *tail empirical expectile process*. We will prove that this process can be approximated by a sequence of Gaussian processes with drift and we will also analyse the difference between the tail empirical expectile process and its population counterpart. These powerful asymptotic results will then be used to introduce and study a further class of extreme expectile estimators as well as to estimate a novel expectile-based analogue of the quantile-based ES, called XES, at intermediate and extreme levels. Finally, we unravel the important question of how to select theoretically the extreme level τ_n' so that each expectile-based risk measure (VaR, MES and XES) at this level coincides with its quantile-based analogue at a given tail probability $\alpha_n \to 1$ as $n \to \infty$. The obtained level $\tau_n' = \tau_n'(\alpha_n)$ needs itself to be estimated, which results in two final composite estimators of the risk measure. Our main results establish the asymptotic distributions of all presented estimators.

The organisation of this section is the following. Section 2.3.2 discusses the basic properties of the expectile at high levels, including its connection with the standard quantile-VaR for high levels $\tau_n \to 1$. Section 2.3.3 presents two estimation methods of intermediate and extreme expectiles. Section 2.3.4 considers the problem of estimating the MES when the related variable is extreme. Section 2.3.5 generalises the convergence result of Section 2.3.3 to Gaussian approximations of the tail empirical expectile process and outlines statistical consequences of such results. Section 2.3.6 addresses the important question of how to select the extreme expectile level in the three studied risk measures. A brief discussion of the finite-sample performance of our estimators and concrete applications to medical insurance data and financial data are provided in Section 2.3.7. Section 2.3.8 concludes with a discussion of perspectives for future work.

2.3.2 Basic properties [12]

We consider here the generic financial position Y to be a real-valued random variable, and the available data $\{Y_1, Y_2, \ldots\}$ are considered as the negative of a series of financial returns. The right tail of the distribution of Y then corresponds to the negative of extreme losses. Following [209], the expectile ξ_τ of order $\tau \in (0,1)$ of the variable Y can be defined as the minimiser (2.3.1) or, equivalently, as

$$\xi_\tau = \arg\min_{\theta\in\mathbb{R}} \left\{ \tau\mathbb{E}\left[(Y-\theta)_+^2 - Y_+^2\right] + (1-\tau)\mathbb{E}\left[(Y-\theta)_-^2 - Y_-^2\right] \right\},$$

where $y_+ := \max(y,0)$ and $y_- := \min(y,0)$. The presence of terms Y_+^2 and Y_-^2 makes this problem well-defined indeed as soon as $Y \in L^1$, i.e. $\mathbb{E}|Y| < \infty$. The related first-order necessary condition for optimality can be written in several ways, one of them being

$$\xi_\tau - \mathbb{E}(Y) = \frac{2\tau - 1}{1-\tau}\,\mathbb{E}\left[(Y-\xi_\tau)_+\right].$$

This equation has a unique solution for all $Y \in L^1$. Hence expectiles of a distribution with finite absolute first moment are well-defined, and we will assume in the sequel that $\mathbb{E}|Y| < \infty$. Expectiles summarise the distribution function in much the same way that the quantiles $q_\tau := F_Y^{-1}(\tau) = \inf\{y \in \mathbb{R} : F_Y(y) \geq \tau\}$ do. A justification for their use to describe distributions and their tails, as well as to quantify the "riskiness" implied by the return distribution under consideration, may be based on the following elementary properties (see [38, 53, 209]):

(i) Law invariance: two integrable random variables Y and $\widetilde{Y}$, with continuous densities, have the same distribution if and only if $\xi_{Y,\tau} = \xi_{\widetilde{Y},\tau}$ for every $\tau \in (0,1)$.

(ii) Location and scale equivariance: the τth expectile of the linear transformation $\widetilde{Y} = a + bY$, where $a, b \in \mathbb{R}$, satisfies

$$\xi_{\widetilde{Y},\tau} = \begin{cases} a + b\,\xi_{Y,\tau} & \text{if} \quad b > 0 \\ a + b\,\xi_{Y,1-\tau} & \text{if} \quad b \leq 0. \end{cases}$$

(iii) Constancy: if $Y = c \in \mathbb{R}$ with probability 1, then $\xi_{Y,\tau} = c$ for any τ.

(iv) Strict monotonicity in τ: if $\tau_1 < \tau_2$, with $\tau_1, \tau_2 \in (0,1)$, then $\xi_{\tau_1} < \xi_{\tau_2}$. Also, the function $\tau \mapsto \xi_\tau$ maps $(0,1)$ onto its range $\{y \in \mathbb{R} : 0 < F_Y(y) < 1\}$.

(v) Preserving stochastic order: if $Y \leq \tilde{Y}$ with probability 1, then $\xi_{Y,\tau} \leq \xi_{\tilde{Y},\tau}$ for any τ.

(vi) Subadditivity: for any variables $Y, \tilde{Y} \in L^1$, $\xi_{Y+\tilde{Y},\tau} \leq \xi_{Y,\tau} + \xi_{\tilde{Y},\tau}$ for all $\tau \geq \frac{1}{2}$. Also, $\xi_{Y+\tilde{Y},\tau} \geq \xi_{Y,\tau} + \xi_{\tilde{Y},\tau}$ for all $\tau \leq \frac{1}{2}$.

(vii) Lipschitzianity with respect to the Wasserstein distance: for all $Y, \tilde{Y} \in L^1$ and all $\tau \in (0,1)$, it holds that $|\xi_{Y,\tau} - \xi_{\tilde{Y},\tau}| \leq \tilde{\tau} \cdot d_W(Y, \tilde{Y})$, where $\tilde{\tau} = \max\{\tau/(1-\tau), (1-\tau)/\tau\}$ and

$$d_W(Y,\tilde{Y}) = \int_{-\infty}^{\infty} |F_Y(y) - F_{\tilde{Y}}(y)|dy = \int_0^1 |F_Y^{-1}(t) - F_{\tilde{Y}}^{-1}(t)|dt.$$

(viii) Sensitivity versus resistance: expectiles are very sensitive to the magnitude of extremes since their gross error sensitivity and rejection points are infinite. Whereas they are resistant to systematic rounding and grouping since their local shift sensitivity is bounded.

The convention we have chosen for values of Y as the negative of returns implies that extreme losses correspond to levels τ close to 1. Only [52, 53, 198, 199] have described what happens for large population expectiles ξ_τ and their link with high quantiles q_τ when Y has a heavy-tailed distribution with tail index $0 < \gamma < 1$. Writing $\overline{F}_Y := 1 - F_Y$ for the survival function of Y, [52] have shown in the case $\gamma < 1$ that

$$\frac{\overline{F}_Y(\xi_\tau)}{\overline{F}_Y(q_\tau)} \sim \gamma^{-1} - 1 \quad \text{as} \quad \tau \to 1, \tag{2.3.3}$$

or equivalently $\overline{F}_Y(\xi_\tau)/(1-\tau) \sim \gamma^{-1}-1$ as $\tau \to 1$. It follows that extreme expectiles ξ_τ are larger than extreme quantiles q_τ (i.e. $\xi_\tau > q_\tau$) when $\gamma > \frac{1}{2}$, whereas $\xi_\tau < q_\tau$ when $\gamma < \frac{1}{2}$, for all large τ. The connection (2.3.3) between high expectiles and quantiles can actually be refined under the second-order condition $\mathcal{C}_2(\gamma,\rho,A)$; this is the first main result of this section.

Proposition 2.3.1. *Assume that condition $\mathcal{C}_2(\gamma,\rho,A)$ holds, with $0<\gamma<1$. Then*

$$\begin{aligned}\frac{\overline{F}_Y(\xi_\tau)}{1-\tau} &= (\gamma^{-1}-1)(1+\varepsilon(\tau))\\ \textit{with}\ \ \varepsilon(\tau) &= -\frac{(\gamma^{-1}-1)^{\gamma}}{q_\tau}(\mathbb{E}(Y)+\mathrm{o}(1)) - \frac{(\gamma^{-1}-1)^{-\rho}}{\gamma(1-\gamma-\rho)}A((1-\tau)^{-1})(1+\mathrm{o}(1))\ \ \textit{as}\ \tau\uparrow 1.\end{aligned}$$

One can use this result to quantify precisely the bias term in the asymptotic expansion of ξ_τ/q_τ.

Corollary 2.3.2. *Assume that condition $\mathcal{C}_2(\gamma,\rho,A)$ holds, with $0<\gamma<1$. If F_Y is strictly increasing, then*

$$\begin{aligned}\frac{\xi_\tau}{q_\tau} &= (\gamma^{-1}-1)^{-\gamma}(1+r(\tau))\\ \textit{with}\ \ r(\tau) &= \frac{\gamma(\gamma^{-1}-1)^{\gamma}}{q_\tau}(\mathbb{E}(Y)+\mathrm{o}(1))\\ &\quad+\left(\frac{(\gamma^{-1}-1)^{-\rho}}{1-\rho-\gamma}+\frac{(\gamma^{-1}-1)^{-\rho}-1}{\rho}+\mathrm{o}(1)\right)A((1-\tau)^{-1})\ \ \textit{as}\ \tau\uparrow 1.\end{aligned}$$

Other refinements under similar second order regular variation conditions can also be found in [198, 199]. [Note also that the requirement that F_Y be strictly increasing is in fact unnecessary, see Proposition 2.3.12 in Section 2.3.5.1 below, corresponding to Proposition 1 in [1].]

An extension to a subset of the Gumbel domain of attraction is also derived in Proposition 2.4 in [52]. In practice, the tail quantities ξ_τ, q_τ and γ are unknown and only a sample of random copies $(Y_1,\ldots,Y_n)$ of Y is typically available. While extreme value estimators of high quantiles and of the tail index γ have been widely used in applied statistical analyses and extensively investigated in theoretical statistics, the problem of estimating ξ_τ, when $\tau=\tau_n\to 1$ at an arbitrary rate as $n\to\infty$, has not been addressed yet. This motivated us to construct estimators of large expectiles ξ_{τ_n} and derive their limit distributions when they are located within or beyond the range of the data, where their empirical counterparts usually fail due to data sparseness. We shall assume the second-order regular variation condition $\mathcal{C}_2(\gamma,\rho,A)$ to obtain our convergence results.

2.3.3 Estimation of extreme expectiles [12]

Our main objective in this section is to estimate ξ_{τ_n} for high levels τ_n that may approach 1 at any rate, covering both scenarios of intermediate expectiles with $n(1-\tau_n)\to\infty$ and extreme expectiles with $n(1-\tau_n)\to c$, for some nonnegative finite constant c. We assume that the available data consist of independent copies $(Y_1,\ldots,Y_n)$ of Y, and denote by $Y_{1,n}\le\cdots\le Y_{n,n}$ their ascending order statistics.

2.3.3.1 Intermediate expectile estimation

Estimation based on intermediate quantiles The rationale for this first method relies on the heavy-tailed assumption and on the asymptotic equivalence (2.3.3). Given that Y is heavy-tailed, Corollary 2.3.2 above (or Proposition 2.3 of [52]) entails

$$\frac{\xi_\tau}{q_\tau}\sim(\gamma^{-1}-1)^{-\gamma}\quad\text{as}\quad\tau\uparrow 1. \tag{2.3.4}$$

Therefore, for a suitable estimator $\widehat{\gamma}$ of γ, we suggest estimating the intermediate expectile ξ_{τ_n} by $\widehat{\xi}_{\tau_n} := (\widehat{\gamma}^{-1}-1)^{-\widehat{\gamma}}\,\widehat{q}_{\tau_n}$, where $\widehat{q}_{\tau_n} := Y_{n-\lfloor n(1-\tau_n)\rfloor,n}$ and $\lfloor\cdot\rfloor$ stands for the floor function. This estimator is based on the intermediate quantile-VaR $\widehat{q}_{\tau_n}$ and crucially hinges on the estimated tail index $\widehat{\gamma}$. A simple and widely used estimator $\widehat{\gamma}$ is the popular Hill estimator, written in terms of the k log-excesses above an intermediate order statistic $Y_{n-k,n}$ as

$$\widehat{\gamma}_H = \frac{1}{k}\sum_{i=1}^{k} \log \frac{Y_{n-i+1,n}}{Y_{n-k,n}}, \tag{2.3.5}$$

where $k = k(n)$ is an intermediate sequence, namely $k(n) \to \infty$ and $k(n)/n \to 0$ as $n \to \infty$.

Next, we formulate conditions that lead to asymptotic normality for $\widehat{\xi}_{\tau_n}$.

Theorem 2.3.3. *Assume that F_Y is strictly increasing, that condition $\mathcal{C}_2(\gamma, \rho, A)$ holds with $0 < \gamma < 1$, that $\tau_n \uparrow 1$ and $n(1-\tau_n) \to \infty$. Assume further that*

$$\sqrt{n(1-\tau_n)}\left(\widehat{\gamma} - \gamma, \frac{\widehat{q}_{\tau_n}}{q_{\tau_n}} - 1\right) \xrightarrow{d} (\Gamma, \Theta). \tag{2.3.6}$$

If $\sqrt{n(1-\tau_n)}q_{\tau_n}^{-1} \to \lambda_1 \in \mathbb{R}$ and $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) \to \lambda_2 \in \mathbb{R}$, then

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{\xi}_{\tau_n}}{\xi_{\tau_n}} - 1\right) \xrightarrow{d} m(\gamma)\Gamma + \Theta - \lambda$$

with $m(\gamma) := (1-\gamma)^{-1} - \log(\gamma^{-1} - 1)$ and

$$\lambda := \gamma(\gamma^{-1} - 1)^{\gamma}\mathbb{E}(Y)\lambda_1 + \left(\frac{(\gamma^{-1}-1)^{-\rho}}{1-\gamma-\rho} + \frac{(\gamma^{-1}-1)^{-\rho} - 1}{\rho}\right)\lambda_2.$$

When using the Hill estimator (2.3.5) of γ with $k = \lceil n(1-\tau_n)\rceil$, sufficient regularity conditions for (2.3.6) to hold can be found in Theorems 2.4.1 and 3.2.5 in [145]. Under these conditions, the limit distribution Γ is then Gaussian with mean $\lambda_2/(1-\rho)$ and variance γ^2, while Θ is the centred Gaussian distribution with variance γ^2. Lemma 3.2.3 in [145] shows that these two Gaussian limiting distributions are independent. As an immediate consequence we get the following for $\widehat{\gamma} = \widehat{\gamma}_H$.

Corollary 2.3.4. *Assume that F_Y is strictly increasing, that condition $\mathcal{C}_2(\gamma, \rho, A)$ holds with $0 < \gamma < 1$, that $\tau_n \uparrow 1$ and $n(1-\tau_n) \to \infty$. If $\sqrt{n(1-\tau_n)}q_{\tau_n}^{-1} \to \lambda_1 \in \mathbb{R}$ and $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) \to \lambda_2 \in \mathbb{R}$, then*

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{\xi}_{\tau_n}}{\xi_{\tau_n}} - 1\right) \xrightarrow{d} \mathcal{N}\left(\frac{m(\gamma)}{1-\rho}\lambda_2 - \lambda,\, v(\gamma)\right),$$

with $m(\gamma)$ and λ as in Theorem 2.3.3, and

$$v(\gamma) = \gamma^2\left[1 + \left(\frac{1}{1-\gamma} - \log\left(\frac{1}{\gamma} - 1\right)\right)^2\right].$$

Yet, a drawback to the resulting estimator $\widehat{\xi}_{\tau_n}$ lies in its heavy dependency on the estimated quantile $\widehat{q}_{\tau_n}$ and tail index $\widehat{\gamma}$ in the sense that $\widehat{\xi}_{\tau_n}$ may inherit the vexing defects of both $\widehat{q}_{\tau_n}$ and $\widehat{\gamma}$. Note also that $\widehat{\xi}_{\tau_n}$ is asymptotically biased, which is not the case for $\widehat{q}_{\tau_n}$; it should be pointed out though that one may design a bias-reduced version of the estimator $\widehat{\xi}_{\tau_n}$. Indeed, the bias components λ_1 and λ_2 appearing in Theorem 2.3.3 can be estimated, respectively, by using $\widehat{\lambda}_1 = \sqrt{n(1-\tau_n)}\widehat{q}_{\tau_n}^{-1}$ and by applying the methodology of [65] in conjunction with the Hall-Welsh class of models to get an estimator $\widehat{\lambda}_2$ of λ_2. Plugging these, along with the empirical mean $\overline{Y}$, the estimator $\widehat{\gamma}$, and a consistent estimator $\widehat{\rho}$ of the second-order parameter ρ (a review of possible estimators $\widehat{\rho}$ is given in [138]), into the expression of λ, we get a consistent estimator $\widehat{\lambda}$ of this bias component. This in turn enables one to define a bias-reduced version of $\widehat{\xi}_{\tau_n}$, for instance as

$$\widehat{\xi}_{\tau_n}^{RB} := \widehat{\xi}_{\tau_n}\left(1 - \left[\frac{m(\widehat{\gamma})}{1-\widehat{\rho}}\widehat{\lambda}_2 - \widehat{\lambda}\right]\frac{1}{\sqrt{n(1-\tau_n)}}\right).$$

Of course, one should expect the value of the asymptotic variance of this estimator to be even higher than that of $\widehat{\xi}_{\tau_n}$, similarly to what is observed when bias reduction techniques are applied to the Hill estimator (see e.g. Theorem 3.2 in [65]).

Another efficient way of estimating ξ_{τ_n}, which we turn to now, is by joining together least asymmetrically weighted squares (LAWS) estimation with our heavy-tailed framework.

Asymmetric least squares estimation Here, we consider estimating the expectile ξ_{τ_n} by its empirical counterpart defined through

$$\widetilde{\xi}_{\tau_n} = \arg\min_{u\in\mathbb{R}} \frac{1}{n}\sum_{i=1}^{n} \eta_{\tau_n}(Y_i - u),$$

where we recall that $\eta_\tau(y) = |\tau - \mathbb{I}\{y \le 0\}|y^2$ is the expectile check function. Clearly

$$\sqrt{n(1-\tau_n)}\left(\frac{\widetilde{\xi}_{\tau_n}}{\xi_{\tau_n}} - 1\right) = \arg\min_{u\in\mathbb{R}} \psi_n(u) \tag{2.3.7}$$

$$\text{with} \qquad \psi_n(u) := \frac{1}{2\xi_{\tau_n}^2}\sum_{i=1}^{n}\left[\eta_{\tau_n}\left(Y_i - \xi_{\tau_n} - \frac{u\xi_{\tau_n}}{\sqrt{n(1-\tau_n)}}\right) - \eta_{\tau_n}(Y_i - \xi_{\tau_n})\right].$$

Note that (ψ_n) is a sequence of almost surely continuous and convex random functions. Theorem 5 in [180] then states that to examine the convergence of the left-hand side term of (2.3.7), it is enough to investigate the asymptotic properties of the sequence (ψ_n). Building on this idea, we get the asymptotic normality of the LAWS estimator $\widetilde{\xi}_{\tau_n}$ by applying standard techniques involving sums of independent and identically distributed random variables. Let us recall the notation $Y_- = \min(Y, 0)$ for the negative part of Y.

Theorem 2.3.5. *Assume that there is $\delta > 0$ such that $\mathbb{E}|Y_-|^{2+\delta} < \infty$, that $0 < \gamma < 1/2$ and $\tau_n \uparrow 1$ is such that $n(1-\tau_n) \to \infty$. Then*

$$\sqrt{n(1-\tau_n)}\left(\frac{\widetilde{\xi}_{\tau_n}}{\xi_{\tau_n}} - 1\right) \xrightarrow{d} \mathcal{N}(0, V(\gamma)) \quad \textit{with} \quad V(\gamma) = \frac{2\gamma^3}{1-2\gamma}.$$

In contrast to Theorem 2.3.3 and Corollary 2.3.4, the limit distribution in Theorem 2.3.5 is derived without recourse to either the extended regular variation condition $\mathcal{C}_2(\gamma, \rho, A)$ or any bias condition. A moment assumption and the condition $0 < \gamma < 1/2$ suffice. Most importantly, unlike the indirect expectile estimator $\widehat{\xi}_{\tau_n}$, the new estimator $\widetilde{\xi}_{\tau_n}$ does not hinge by construction on any quantile or tail index estimators. A comparison of the asymptotic variance $V(\gamma)$ of $\widetilde{\xi}_{\tau_n}$ with the asymptotic variance $v(\gamma)$ of $\widehat{\xi}_{\tau_n}$ is provided in Figure 2.2. It can be seen from the left panel that both asymptotic variances are stable and close for values of $\gamma < 0.3$, with an advantage for $V(\gamma)$ in dashed line as visualised more clearly in the right panel. Then $V(\gamma)$ becomes appreciably larger than $v(\gamma)$ for $\gamma > 0.3$ and explodes in a neighbourhood of $1/2$, while $v(\gamma)$ in solid line remains lower than the level 1.25.

2.3.3.2 Extreme expectile estimation

We now discuss the important issue of estimating extreme tail expectiles $\xi_{\tau_n'}$, where $\tau_n' \to 1$ with $n(1-\tau_n') \to c < \infty$ as $n \to \infty$. The basic idea is to extrapolate intermediate expectile estimates of order $\tau_n \to 1$, such that $n(1-\tau_n) \to \infty$, to the very extreme level τ_n'. This is achieved by transferring the device of [256] for estimating an extreme quantile to our expectile setup. Note that the levels τ_n' and τ_n are typically set to be $\tau_n' = 1 - p_n$ for a p_n not greater than $1/n$, and $\tau_n = 1 - k(n)/n$ for an intermediate sequence of integers $k(n)$.

A combination of the heavy-tailed assumption with (2.3.4) suggests that when τ_n and τ_n' satisfy suitable conditions, we may write:

$$\frac{q_{\tau_n'}}{q_{\tau_n}} = \frac{U((1-\tau_n')^{-1})}{U((1-\tau_n)^{-1})} \approx \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\gamma} \quad \text{and thus} \quad \frac{\xi_{\tau_n'}}{\xi_{\tau_n}} \approx \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\gamma}.$$

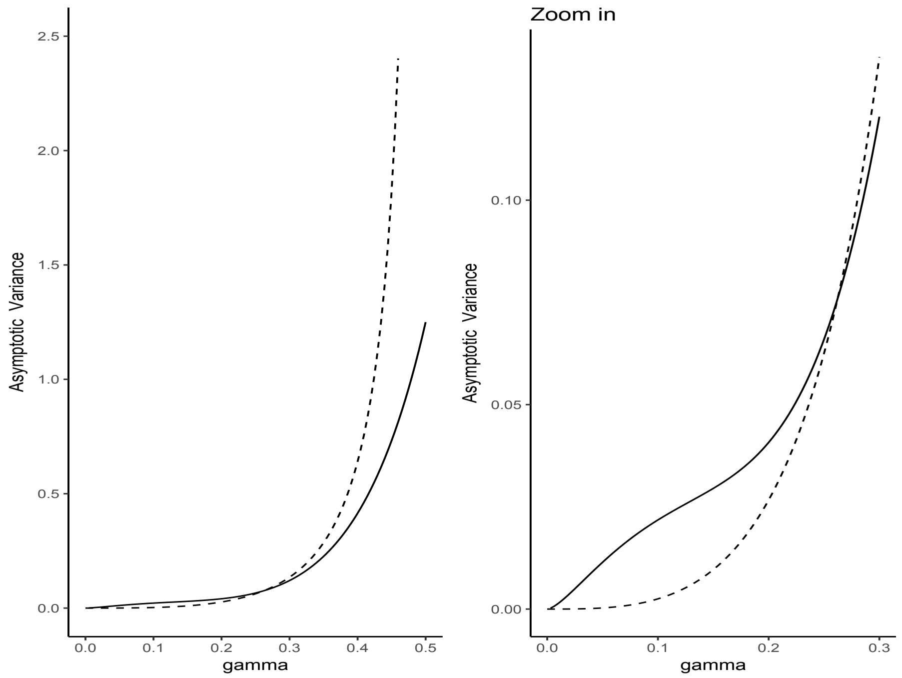

Figure 2.2: Asymptotic variances $V(\gamma)$ of the LAWS estimator $\widetilde{\xi}_{\tau_n}$ in dashed line, and $v(\gamma)$ of the indirect estimator $\widehat{\xi}_{\tau_n}$ in solid line. Left panel: $\gamma \in (0, 1/2)$, right panel: $\gamma < 0.3$.

This approximation motivates the following class of plug-in estimators of $\xi_{\tau'_n}$:

$$\overline{\xi}^{\star}_{\tau'_n} \equiv \overline{\xi}^{\star}_{\tau'_n}(\tau_n) := \left(\frac{1-\tau'_n}{1-\tau_n}\right)^{-\widehat{\gamma}} \overline{\xi}_{\tau_n} \tag{2.3.8}$$

where $\widehat{\gamma}$ is an estimator of γ, and $\overline{\xi}_{\tau_n}$ is either the estimator $\widehat{\xi}_{\tau_n}$ or $\widetilde{\xi}_{\tau_n}$ of the intermediate expectile ξ_{τ_n}. We actually have $\overline{\xi}^{\star}_{\tau'_n}/\overline{\xi}_{\tau_n} = \widehat{q}^{\star}_{\tau'_n}/\widehat{q}_{\tau_n}$, where $\widehat{q}_{\tau_n} = Y_{n-\lfloor n(1-\tau_n)\rfloor,n}$ is the intermediate quantile estimator introduced above, and $\widehat{q}^{\star}_{\tau'_n}$ is the Weissman extreme quantile estimator

$$\widehat{q}^{\star}_{\tau'_n} \equiv \widehat{q}^{\star}_{\tau'_n}(\tau_n) := \left(\frac{1-\tau'_n}{1-\tau_n}\right)^{-\widehat{\gamma}} \widehat{q}_{\tau_n}. \tag{2.3.9}$$

We then show that $\overline{\xi}^{\star}_{\tau'_n}/\xi_{\tau'_n} - 1$ has the same limit distribution as $\widehat{\gamma} - \gamma$ with a different scaling.

Theorem 2.3.6. *Assume that F_Y is strictly increasing, that condition $\mathcal{C}_2(\gamma, \rho, A)$ holds with $0 < \gamma < 1$ and $\rho < 0$, and that τ_n, $\tau'_n \uparrow 1$, with $n(1-\tau_n) \to \infty$ and $n(1-\tau'_n) \to c < \infty$. If moreover*

$$\sqrt{n(1-\tau_n)}\left(\frac{\overline{\xi}_{\tau_n}}{\xi_{\tau_n}} - 1\right) \xrightarrow{d} \Delta \quad \text{and} \quad \sqrt{n(1-\tau_n)}(\widehat{\gamma} - \gamma) \xrightarrow{d} \Gamma,$$

with $\sqrt{n(1-\tau_n)}q^{-1}_{\tau_n} \to \lambda_1 \in \mathbb{R}$, $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) \to \lambda_2 \in \mathbb{R}$ and $\sqrt{n(1-\tau_n)}/\log[(1-\tau_n)/(1-\tau'_n)] \to \infty$, then

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau'_n)]}\left(\frac{\overline{\xi}^{\star}_{\tau'_n}}{\xi_{\tau'_n}} - 1\right) \xrightarrow{d} \Gamma.$$

More specifically, we can choose $\overline{\xi}_{\tau_n}$ in (2.3.8) to be either the indirect intermediate expectile estimator $\widehat{\xi}_{\tau_n}$, the resulting extreme expectile estimator $\widehat{\xi}^{\star}_{\tau'_n} := \overline{\xi}^{\star}_{\tau'_n}$ being

$$\widehat{\xi}^{\star}_{\tau'_n} = \left(\frac{1-\tau'_n}{1-\tau_n}\right)^{-\widehat{\gamma}} \widehat{\xi}_{\tau_n} = \left(\widehat{\gamma}^{-1} - 1\right)^{-\widehat{\gamma}} \widehat{q}^{\star}_{\tau'_n}, \tag{2.3.10}$$

or we may choose $\overline{\xi}_{\tau_n}$ to be the LAWS estimator $\widetilde{\xi}_{\tau_n}$, yielding the extreme expectile estimator

$$\widetilde{\xi}^{\star}_{\tau'_n} = \left(\frac{1-\tau'_n}{1-\tau_n}\right)^{-\widehat{\gamma}} \widetilde{\xi}_{\tau_n}, \tag{2.3.11}$$

Their respective asymptotic properties are given in the next two corollaries of Theorem 2.3.6.

Corollary 2.3.7. *Assume that F_Y is strictly increasing, that condition $\mathcal{C}_2(\gamma,\rho,A)$ holds with $0<\gamma<1$ and $\rho<0$, and that τ_n, $\tau'_n \uparrow 1$ with $n(1-\tau_n)\to\infty$ and $n(1-\tau'_n)\to c<\infty$. Assume further that*

$$\sqrt{n(1-\tau_n)}\left(\widehat{\gamma}-\gamma, \frac{\widehat{q}_{\tau_n}}{q_{\tau_n}}-1\right) \xrightarrow{d} (\Gamma,\Theta).$$

If $\sqrt{n(1-\tau_n)}q_{\tau_n}^{-1}\to\lambda_1\in\mathbb{R}$, $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1})\to\lambda_2\in\mathbb{R}$ and $\sqrt{n(1-\tau_n)}/\log[(1-\tau_n)/(1-\tau'_n)]\to\infty$, then

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau'_n)]}\left(\frac{\widehat{\xi}^{\star}_{\tau'_n}}{\xi_{\tau'_n}}-1\right) \xrightarrow{d} \Gamma.$$

Corollary 2.3.8. *Assume that F_Y is strictly increasing, there is $\delta>0$ such that $\mathbb{E}|Y_-|^{2+\delta}<\infty$, condition $\mathcal{C}_2(\gamma,\rho,A)$ holds with $0<\gamma<1/2$ and $\rho<0$, and that τ_n, $\tau'_n\uparrow 1$ with $n(1-\tau_n)\to\infty$ and $n(1-\tau'_n)\to c<\infty$. If in addition*

$$\sqrt{n(1-\tau_n)}(\widehat{\gamma}-\gamma) \xrightarrow{d} \Gamma$$

and $\sqrt{n(1-\tau_n)}q_{\tau_n}^{-1}\to\lambda_1\in\mathbb{R}$, $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1})\to\lambda_2\in\mathbb{R}$ and $\sqrt{n(1-\tau_n)}/\log[(1-\tau_n)/(1-\tau'_n)]\to\infty$, then

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau'_n)]}\left(\frac{\widetilde{\xi}^{\star}_{\tau'_n}}{\xi_{\tau'_n}}-1\right) \xrightarrow{d} \Gamma.$$

2.3.4 Estimating an extreme expectile-based Marginal Expected Shortfall [12]

2.3.4.1 Setting and objective

With the recent financial crisis and the rising interconnection between financial institutions, interest in the concept of systemic risk has grown. Systemic risk is defined in [42, 64, 116] as the propensity of a financial institution to be undercapitalised when the financial system as a whole is undercapitalised. An important step in constructing a systemic risk measure for a financial firm is to measure the contribution of the firm to a systemic crisis, namely a major stock market decline that happens once or twice a decade. The total risk measured by the expected capital shortfall in the financial system during a systemic crisis is typically decomposed into firm level contributions. Each financial firm's contribution to systemic risk can then be measured as its Marginal Expected Shortfall (MES), i.e. the expected loss on the firm's return X conditional on the loss Y in the aggregated return of the financial market being extreme. More specifically, the MES at probability level $(1-\tau)$ is defined as

$$\mathrm{QMES}(\tau)=\mathbb{E}\{X|Y>q_{Y,\tau}\},\ \tau\in(0,1),$$

where $q_{Y,\tau}$ is the τth quantile of the distribution of Y. Typically, a systemic crisis defined as an extreme tail event corresponds to a probability τ at an extremely high level that can be even larger than $1-1/n$, where n is the sample size of historical data that are used for estimating QMES(τ). The estimation procedure in [42] relies on daily data from only one year and assumes a specific linear relationship between X and Y. A nonparametric kernel estimation method has been performed in [64, 116], but cannot handle extreme events required for systemic risk measures (i.e. $1-\tau=1-\tau_n=\mathrm{O}(1/n)$). Very recently, [68] have proposed adapted extreme value tools for the estimation of QMES(τ) without recourse to any parametric structure on (X,Y). Here, instead of the extreme τth quantile $q_{Y,\tau}$, we shall explore the use of the τth expectile analogue $\xi_{Y,\tau}$ in the marginal expected shortfall

$$\mathrm{XMES}(\tau)=\mathbb{E}\{X|Y>\xi_{Y,\tau}\}$$

at least for the following reason: as claimed by [190, 209, 235], expectiles make a more efficient use of the available data since they rely on the distance of observations from the predictor, while quantile estimation only knows whether an observation is below or above the predictor. It would be awkward to measure extreme risk based only on the frequency of tail losses and not on their values. An interesting asymptotic connection between XMES(τ) and QMES(τ) is provided below in Proposition 2.3.9. The overall objective is to establish estimators of the tail expectile-based MES and unravel their asymptotic behaviour in a general setting.

2.3.4.2 Tail dependence model

Suppose the random vector (X, Y) has a continuous bivariate distribution function $F_{(X,Y)}$ and denote by F_X and F_Y the marginal distribution functions of X and Y, assumed to be increasing in what follows. Given that our goal is to estimate XMES(τ) at an extreme level τ, we adopt the same conditions as [68] on the right-hand tail of X and on the right-hand upper tail dependence of (X, Y). Here, the right-hand upper tail dependence between X and Y is described by the following joint convergence condition:

$\mathcal{JC}(R)$ For all $(x, y) \in [0, \infty]^2$ such that at least x or y is finite, the limit

$$\lim_{t\to\infty} t\,\mathbb{P}(\overline{F}_X(X) \le x/t, \overline{F}_Y(Y) \le y/t) =: R(x, y)$$

exists, with $R(1, 1) > 0$. Here $\overline{F}_X = 1 - F_X$ and $\overline{F}_Y = 1 - F_Y$.

The limit function R completely determines the so-called tail dependence function ℓ (see [107]) via the identity $\ell(x, y) = x + y - R(x, y)$ for all $x, y \ge 0$ (see also Section 8.2 in [47]). Regarding the marginal distributions, we assume that X and Y are heavy-tailed with respective tail indices $\gamma_X, \gamma_Y > 0$, or equivalently, for all $z > 0$,

$$\frac{U_X(tz)}{U_X(t)} \to z^{\gamma_X} \quad \text{and} \quad \frac{U_Y(tz)}{U_Y(t)} \to z^{\gamma_Y} \quad \text{as} \quad t \to \infty,$$

with U_X and U_Y being, respectively, the left-continuous inverse functions of $1/\overline{F}_X$ and $1/\overline{F}_Y$. Compared with the quantile-based MES framework [68], we need the extra condition of heavy-tailedness of Y which is quite natural in the financial setting. Under these regularity conditions, we get the following asymptotic approximations for XMES(τ).

Proposition 2.3.9. *Suppose that condition $\mathcal{JC}(R)$ holds and that X and Y are heavy-tailed with respective tail indices $\gamma_X, \gamma_Y \in (0, 1)$. Then*

$$\lim_{\tau\uparrow 1} \frac{XMES(\tau)}{U_X(1/\overline{F}_Y(\xi_{Y,\tau}))} = \int_0^\infty R(x^{-1/\gamma_X}, 1)dx \tag{2.3.12}$$

$$\text{and} \quad \lim_{\tau\uparrow 1} \frac{XMES(\tau)}{QMES(\tau)} = \left(\gamma_Y^{-1} - 1\right)^{-\gamma_X}. \tag{2.3.13}$$

The first convergence result indicates that XMES(τ) is asymptotically equivalent to the small exceedance probability $U_X(1/\overline{F}_Y(\xi_{Y,\tau}))$ up to a positive multiplicative constant. Since usually in the financial setting $0 < \gamma_X, \gamma_Y < 1/2$, the second result shows that XMES(τ) is less extreme than QMES(τ) as $\tau \to 1$. This is visualised in Figure 2.3 in the case of the restriction of the standard bivariate Student t_ν-distribution to $(0, \infty)^2$, having density

$$f_\nu(x, y) = \frac{2}{\pi}\left(1 + \frac{x^2 + y^2}{\nu}\right)^{-(\nu+2)/2}, \quad x, y > 0,$$

where $\nu = 3$ on the left panel and $\nu = 5$ on the right panel. It can be seen that QMES(τ) becomes overall much more extreme than XMES(τ) as τ approaches 1.

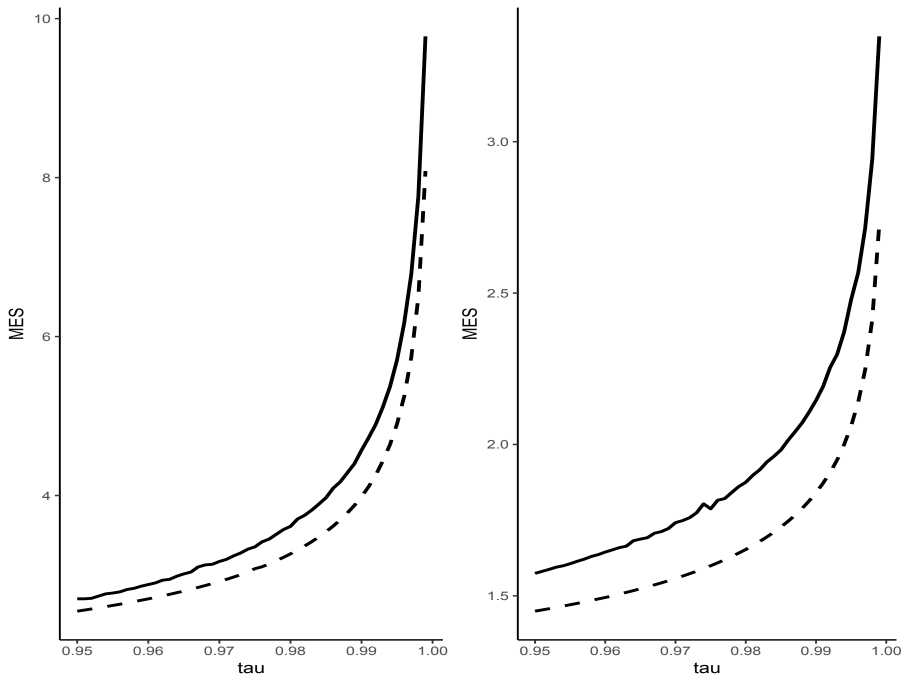

Figure 2.3: QMES(τ) in solid line and XMES(τ) in dashed line, as functions of $\tau \in [0.95, 1)$. Case of the Student t_ν-distribution on $(0, \infty)^2$. Left: $\nu = 3$, right: $\nu = 5$.

2.3.4.3 Estimation and results

The asymptotic equivalences in Proposition 2.3.9 are of particular interest when it comes to proposing estimators for the tail expectile-based MES. Two approaches will be distinguished. We consider first asymmetric least squares estimation by making use of the asymptotic equivalence (2.3.12). Subsequently we shall deal with a nonparametric estimator derived from the asymptotic connection (2.3.13) with the tail quantile-based MES.

Asymmetric least squares estimation On the basis of the limit (2.3.12) and then of the heavy-tailedness assumption on X, we have for $\tau < \tau' < 1$ that, as $\tau \to 1$,

$$\text{XMES}(\tau') \approx \frac{U_X(1/\overline{F}_Y(\xi_{Y,\tau'}))}{U_X(1/\overline{F}_Y(\xi_{Y,\tau}))}\text{XMES}(\tau) \approx \left(\frac{\overline{F}_Y(\xi_{Y,\tau})}{\overline{F}_Y(\xi_{Y,\tau'})}\right)^{\gamma_X} \text{XMES}(\tau).$$

It follows then from Proposition 2.3.1 that

$$\text{XMES}(\tau') \approx \left(\frac{1-\tau'}{1-\tau}\right)^{-\gamma_X} \text{XMES}(\tau). \tag{2.3.14}$$

Hence, to estimate XMES(τ') at an arbitrary extreme level $\tau' = \tau'_n$, we first consider the estimation of XMES(τ) at an intermediate level $\tau = \tau_n$, and then we use the extrapolation technique of [256]. For estimating $\text{XMES}(\tau_n) = \mathbb{E}\{X|Y > \xi_{Y,\tau_n}\}$ at an intermediate level $\tau_n \to 1$ such that $n(1-\tau_n) \to \infty$, as $n \to \infty$, we use the empirical version

$$\widetilde{\text{XMES}}(\tau_n) := \frac{\sum_{i=1}^n X_i \mathbb{I}\{X_i > 0, Y_i > \widetilde{\xi}_{Y,\tau_n}\}}{\sum_{i=1}^n \mathbb{I}\{Y_i > \widetilde{\xi}_{Y,\tau_n}\}},$$

where $\widetilde{\xi}_{Y,\tau_n}$ is the LAWS estimator of ξ_{Y,τ_n}. As a matter of fact, in actuarial settings, we typically have a positive loss variable X, and hence $\mathbb{I}\{X_i > 0\} = 1$. When considering a real-valued profit-loss variable X, the MES is mainly determined by high, and hence positive, values of X as shown in [68].

We shall show under general conditions that the estimator $\widetilde{\mathrm{XMES}}(\tau_n)$ is $\sqrt{n(1-\tau_n)}$-relatively consistent. By plugging this estimator into approximation (2.3.14) together with a $\sqrt{n(1-\tau_n)}$-consistent estimator $\widehat{\gamma}_X$ of γ_X, we obtain the following estimator of $\mathrm{XMES}(\tau_n')$:

$$\widetilde{\mathrm{XMES}}^{\star}(\tau_n') \equiv \widetilde{\mathrm{XMES}}^{\star}(\tau_n';\tau_n) := \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}_X} \widetilde{\mathrm{XMES}}(\tau_n). \tag{2.3.15}$$

To determine the limit distribution of this estimator, we need to quantify the rate of convergence in condition $\mathcal{JC}(R)$ as follows:

$\mathcal{JC}_2(R,\beta,\kappa)$ Condition $\mathcal{JC}(R)$ holds and there exist $\beta > \gamma_X$ and $\kappa < 0$ such that

$$\sup_{\substack{x\in(0,\infty)\\ y\in[1/2,2]}} \left| \frac{t\mathbb{P}(\overline{F}_X(X) \le x/t, \overline{F}_Y(Y) \le y/t) - R(x,y)}{\min(x^\beta, 1)} \right| = \mathrm{O}(t^\kappa) \quad \text{as} \quad t \to \infty.$$

This is exactly condition (a) in [68] under which an extrapolated estimator of $\mathrm{QMES}(\tau_n')$ converges to a normal distribution. See also condition (7.2.8) in [145]. We also need to assume that the tail quantile function U_X (resp. U_Y) satisfies the second-order condition $\mathcal{C}_2(\gamma_X,\rho_X,A_X)$ (resp. $\mathcal{C}_2(\gamma_Y,\rho_Y,A_Y)$). The following generic theorem gives the asymptotic distribution of $\widetilde{\mathrm{XMES}}^{\star}(\tau_n')$. The asymptotic normality follows by using for example the Hill estimator $\widehat{\gamma}_X$ of the tail index γ_X.

Theorem 2.3.10. *Suppose that condition $\mathcal{JC}_2(R,\beta,\kappa)$ holds, that there is $\delta > 0$ such that $\mathbb{E}|Y_-|^{2+\delta} < \infty$, and that U_X and U_Y satisfy conditions $\mathcal{C}_2(\gamma_X,\rho_X,A_X)$ and $\mathcal{C}_2(\gamma_Y,\rho_Y,A_Y)$ with $\gamma_X, \gamma_Y \in (0,1/2)$ and $\rho_X < 0$. Assume further that*

- *(i)* τ_n, $\tau_n' \uparrow 1$, *with* $n(1-\tau_n) \to \infty$, $n(1-\tau_n') \to c < \infty$ *and* $\sqrt{n(1-\tau_n)}/\log[(1-\tau_n)/(1-\tau_n')] \to \infty$ *as* $n \to \infty$;
- *(ii)* $1-\tau_n = \mathrm{O}(n^{\alpha-1})$ *for some* $\alpha < \min\left(\dfrac{-2\kappa}{-2\kappa+1}, \dfrac{2\gamma_X\rho_X}{2\gamma_X\rho_X+\rho_X-1}\right)$;
- *(iii) The bias conditions* $\sqrt{n(1-\tau_n)}q_{Y,\tau_n}^{-1} \to \lambda_1 \in \mathbb{R}$, $\sqrt{n(1-\tau_n)}A_X((1-\tau_n)^{-1}) \to \lambda_2 \in \mathbb{R}$ *and* $\sqrt{n(1-\tau_n)}A_Y((1-\tau_n)^{-1}) \to \lambda_3 \in \mathbb{R}$ *hold;*
- *(iv)* $\sqrt{n(1-\tau_n)}(\widehat{\gamma}_X - \gamma_X) \xrightarrow{d} \Gamma$.

Then, if $X > 0$ almost surely, we have that

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\widetilde{XMES}^{\star}(\tau_n')}{XMES(\tau_n')} - 1\right) \xrightarrow{d} \Gamma.$$

This convergence remains still valid if $X \in \mathbb{R}$ provided

$$\text{(v) } \mathbb{E}|X_-|^{1/\gamma_X} < \infty; \tag{2.3.16}$$

$$\text{(vi) } n(1-\tau_n) = \mathrm{o}\left((1-\tau_n')^{2\kappa(1-\gamma_X)}\right) \quad \text{as } n \to \infty. \tag{2.3.17}$$

Let us point out here that condition (ii), which also appears in Theorem 1 of [68], is a strengthening of the condition $1-\tau_n = \mathrm{o}(1)$. It essentially allows to control additional bias terms that appear in conditions $\mathcal{JC}_2(R,\beta,\kappa)$ and $\mathcal{C}_2(\gamma_X,\rho_X,A_X)$. Condition (vi), which is also utilised in [68], is another bias condition that makes it possible to control the bias coming from the left tail of X.

Estimation based on tail QMES On the basis of the limit (2.3.13), we consider the alternative estimator

$$\widehat{\mathrm{XMES}}^{\star}(\tau_n') := \left(\widehat{\gamma}_Y^{-1} - 1\right)^{-\widehat{\gamma}_X} \widehat{\mathrm{QMES}}^{\star}(\tau_n'), \tag{2.3.18}$$

where $\widehat{\gamma}_X$, $\widehat{\gamma}_Y$ and $\widehat{\mathrm{QMES}}^\star(\tau_n')$ are suitable estimators of γ_X, γ_Y and $\mathrm{QMES}(\tau_n')$, respectively. Here, we use the Weissman-type device

$$\widehat{\mathrm{QMES}}^\star(\tau_n') = \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}_X} \widehat{\mathrm{QMES}}(\tau_n) \tag{2.3.19}$$

of [68] to estimate $\mathrm{QMES}(\tau_n')$, where

$$\widehat{\mathrm{QMES}}(\tau_n) = \frac{1}{\lfloor n(1-\tau_n)\rfloor}\sum_{i=1}^n X_i \mathbb{I}\{X_i > 0, Y_i > \widehat{q}_{Y,\tau_n}\},$$

with $\widehat{q}_{Y,\tau_n} := Y_{n-\lfloor n(1-\tau_n)\rfloor,n}$ being an intermediate quantile-VaR. As a matter of fact, [68] have suggested the use of two intermediate sequences in $\widehat{\gamma}_X$ and $\widehat{\mathrm{QMES}}(\tau_n)$ to be chosen in two steps in practice. To ease the presentation, we use the same intermediate sequence τ_n in both $\widehat{\gamma}_X$ and $\widehat{\mathrm{QMES}}(\tau_n)$. Next, we derive the asymptotic distribution of the new estimator $\widehat{\mathrm{XMES}}^\star(\tau_n')$.

Theorem 2.3.11. *Suppose that condition $\mathcal{JC}_2(R,\beta,\kappa)$ holds, and U_X and U_Y satisfy conditions $\mathcal{C}_2(\gamma_X,\rho_X,A_X)$ and $\mathcal{C}_2(\gamma_Y,\rho_Y,A_Y)$ with $\gamma_X\in(0,1/2)$ and $\rho_X<0$. Assume further that*

(i) τ_n, $\tau_n'\uparrow 1$, *with* $n(1-\tau_n)\to\infty$, $n(1-\tau_n')\to c<\infty$ *and* $\sqrt{n(1-\tau_n)}/\log[(1-\tau_n)/(1-\tau_n')]\to\infty$ *as* $n\to\infty$;

(ii) $1-\tau_n = \mathrm{O}(n^{\alpha-1})$ *for some* $\alpha < \min\left(\dfrac{-2\kappa}{-2\kappa+1}, \dfrac{2\gamma_X\rho_X}{2\gamma_X\rho_X+\rho_X-1}\right)$;

(iii) The bias conditions $\sqrt{n(1-\tau_n)}q_{Y,\tau_n}^{-1}\to\lambda\in\mathbb{R}$ *and* $\sqrt{n(1-\tau_n)}A_X((1-\tau_n)^{-1})\to 0$ *hold;*

(iv) $\sqrt{n(1-\tau_n)}(\widehat{\gamma}_X-\gamma_X)\xrightarrow{d}\Gamma$ *and* $\sqrt{n(1-\tau_n)}(\widehat{\gamma}_Y-\gamma_Y)=\mathrm{O}_{\mathbb{P}}(1)$.

Then, if $X>0$ almost surely, we have that

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\widehat{XMES}^\star(\tau_n')}{XMES(\tau_n')}-1\right)\xrightarrow{d}\Gamma.$$

This convergence remains still valid if $X\in\mathbb{R}$ provided that (2.3.16) *and* (2.3.17) *hold.*

2.3.5 Gaussian approximations for the tail empirical expectile process [1]

Theorem 2.3.5 is, to the best of our knowledge, the first asymptotic result on the direct sample counterpart of an intermediate expectile. It does not, however, allow for simultaneous consideration of several intermediate sample expectiles. By contrast, much more is currently known about the convergence of intermediate empirical quantiles. For example, Gaussian approximations of the tail empirical quantile process have been known for at least three decades; see, among others, [81, 111] as well as their more modern formulations in [101] and Theorem 2.4.8 in [145]. These powerful asymptotic results, and their later generalisations, have been successfully used in the analysis of a number of complex statistical functionals, such as test statistics aimed at checking extreme value conditions [98, 105, 165], bias-corrected extreme value index estimators [146] as well as the results presented in Section 2.2 (i.e. those of [11, 14]) on the estimation of extreme Wang distortion risk measures.

The aim of this section is to fill this gap by showing that the tail empirical expectile process can be approximated by a sequence of Gaussian processes with drift, and deriving its joint asymptotic behaviour with the tail empirical quantile process. We will use this main theoretical result to, first, construct and study a general class of weighted estimators for intermediate expectiles ξ_{τ_n}, by combining the direct and indirect estimators of intermediate expectiles of Section 2.3.3.1. We will then move to the estimation of an expectile-based analogue of the quantile-based ES. We shall show first that the expectile-based form of ES introduced in [242] is not a coherent risk measure. Instead, we will define a coherent alternative form that we call XES_τ, and which is an average of tail expectiles. Its structure will make it easily estimated by an average of the empirical tail expectile process, whose discrepancy from the true XES_{τ_n} can be unravelled thanks to our Gaussian approximations. We will also extrapolate our estimators to the very far tails of the distribution of Y using the argument of [256].

2.3.5.1 Setting and main results

It is well-known that, under condition $\mathcal{C}_2(\gamma,\rho,A)$, the tail empirical (intermediate) quantile process

$$(0,1]\to\mathbb{R},\ s\mapsto \widehat{q}_{1-ks/n} := Y_{n-\lfloor ks\rfloor,n},$$

can be approximated by a sequence of scaled Brownian motions with drift. Namely, one can construct a sequence W_n of standard Brownian motions and a suitable measurable function A_0 such that

$$s^{\gamma+1/2+\varepsilon}\left|\sqrt{k}\left(\frac{\widehat{q}_{1-ks/n}}{q_{1-k/n}}-s^{-\gamma}\right)-\gamma s^{-\gamma-1}W_n(s)-\sqrt{k}A_0(n/k)s^{-\gamma}\frac{s^{-\rho}-1}{\rho}\right|$$

converges in probability to 0 uniformly in $s\in(0,1]$ for any sufficiently small $\varepsilon>0$ (see Theorem 2.4.8 in [145]). In addition to satisfying $k\to\infty$ and $k/n\to 0$, the sequence of integers $k=k(n)$ should also satisfy $\sqrt{k}A_0(n/k)=\mathrm{O}(1)$. The proof of this approximation result reveals that it is valid for a suitable version of the quantile process, equal to the original one in distribution, on a rich enough probability space (potentially different from the original space on which the Y_i are defined). We will work in the sequel with this version of the quantile process and thus with the associated Y_i. Our weak convergence results to follow on extreme expectile and Expected Shortfall estimation are of course unaffected by this choice. Besides, the function A_0 is actually asymptotically equivalent to A, see Theorem 2.3.9 in [145]. We may therefore write:

$$\begin{aligned}\frac{\widehat{q}_{1-(1-\tau_n)s}}{q_{\tau_n}} &= s^{-\gamma}\Bigg(1+\frac{1}{\sqrt{n(1-\tau_n)}}\gamma s^{-1}W_n(s)+\frac{s^{-\rho}-1}{\rho}A((1-\tau_n)^{-1})\\ &\qquad +\mathrm{o}_{\mathbb{P}}\left(\frac{s^{-1/2-\varepsilon}}{\sqrt{n(1-\tau_n)}}\right)\Bigg)\quad\text{uniformly in } s\in(0,1],\end{aligned}\tag{2.3.20}$$

where we set $k=n(1-\tau_n)$, with $\tau_n\to1$, $n(1-\tau_n)\to\infty$ and $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1})=\mathrm{O}(1)$.

By analogy with the tail empirical quantile process, we define the *tail empirical expectile process* to be the stochastic process

$$(0,1]\to\mathbb{R},\ s\mapsto\widetilde{\xi}_{1-(1-\tau_n)s}.$$

Here $\widetilde{\xi}$ is the LAWS estimator introduced in Section 2.3.3.1. To show a result analogue to (2.3.20) for this tail empirical expectile process in the greatest possible generality (within the heavy-tailed framework), we first strengthen the asymptotic expansion stated in Corollary 2.3.2 and linking high quantiles and expectiles.

Proposition 2.3.12. *Assume that $\mathbb{E}|Y_-|<\infty$ and condition $\mathcal{C}_2(\gamma,\rho,A)$ holds, with $0<\gamma<1$.*

(i) We have, as $\tau\to1$,

$$\begin{aligned}\frac{\xi_\tau}{q_\tau} &= (\gamma^{-1}-1)^{-\gamma}\Bigg(1+\frac{\gamma(\gamma^{-1}-1)^{\gamma}}{q_\tau}(\mathbb{E}(Y)+\mathrm{o}(1))\\ &\qquad+\left(\frac{(\gamma^{-1}-1)^{-\rho}}{1-\gamma-\rho}+\frac{(\gamma^{-1}-1)^{-\rho}-1}{\rho}+\mathrm{o}(1)\right)A((1-\tau)^{-1})\Bigg).\end{aligned}$$

(ii) Let $\tau_n \to 1$ be such that $n(1-\tau_n) \to \infty$, and pick $s \in (0,1]$. Then

$$\begin{aligned}\frac{\xi_{1-(1-\tau_n)s}}{\xi_{\tau_n}} &= s^{-\gamma}\left(1+(s^{\gamma}-1)\frac{\gamma(\gamma^{-1}-1)^{\gamma}}{q_{\tau_n}}(\mathbb{E}(Y)+\mathrm{o}(1))\right.\\ &+ \left.\frac{(1-\gamma)(\gamma^{-1}-1)^{-\rho}}{1-\gamma-\rho}\times\frac{s^{-\rho}-1}{\rho}A((1-\tau_n)^{-1})(1+\mathrm{o}(1))\right).\end{aligned}$$

Part (i) of this proposition relaxes the conditions in Corollary 2.3.2 by removing the unnecessary assumption of strict monotonicity of F. Part (ii) gives the asymptotic expansion of intermediate expectiles akin to condition $\mathcal{C}_2(\gamma,\rho,A)$ for intermediate quantiles, which also reads as

$$\frac{q_{1-(1-\tau_n)s}}{q_{\tau_n}} = s^{-\gamma}\left(1+\frac{s^{-\rho}-1}{\rho}A((1-\tau_n)^{-1})(1+\mathrm{o}(1))\right).$$

We are now ready to generalise Theorem 2.3.5 to a uniform approximation of the tail expectile process $s \mapsto \widetilde{\xi}_{1-(1-\tau_n)s}$. This is the first main result of this section.

Theorem 2.3.13. *Suppose that $\mathbb{E}|Y_-|^2 < \infty$. Assume further that condition $\mathcal{C}_2(\gamma,\rho,A)$ holds, with $0<\gamma<1/2$. Let $\tau_n \to 1$ be such that $n(1-\tau_n)\to\infty$ and $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) = \mathrm{O}(1)$. Then there exists a sequence W_n of standard Brownian motions such that, for any $\varepsilon>0$ sufficiently small,*

$$\begin{aligned}\frac{\widehat{q}_{1-(1-\tau_n)s}}{q_{\tau_n}} &= s^{-\gamma}\left(1+\frac{1}{\sqrt{n(1-\tau_n)}}\gamma\sqrt{\gamma^{-1}-1}\,s^{-1}W_n\left(\frac{s}{\gamma^{-1}-1}\right)\right.\\ &\left.+\frac{s^{-\rho}-1}{\rho}A((1-\tau_n)^{-1})+\mathrm{o}_{\mathbb{P}}\left(\frac{s^{-1/2-\varepsilon}}{\sqrt{n(1-\tau_n)}}\right)\right)\\ \text{and}\quad \frac{\widetilde{\xi}_{1-(1-\tau_n)s}}{\xi_{\tau_n}} &= s^{-\gamma}\left(1+(s^{\gamma}-1)\frac{\gamma(\gamma^{-1}-1)^{\gamma}}{q_{\tau_n}}(\mathbb{E}(Y)+\mathrm{o}_{\mathbb{P}}(1))\right.\\ &+\frac{1}{\sqrt{n(1-\tau_n)}}\gamma^2\sqrt{\gamma^{-1}-1}\,s^{\gamma-1}\int_0^s W_n(t)\,t^{-\gamma-1}\,dt\\ &+\frac{(1-\gamma)(\gamma^{-1}-1)^{-\rho}}{1-\gamma-\rho}\times\frac{s^{-\rho}-1}{\rho}A((1-\tau_n)^{-1})\\ &\left.+\mathrm{o}_{\mathbb{P}}\left(\frac{s^{-1/2-\varepsilon}}{\sqrt{n(1-\tau_n)}}\right)\right)\quad \text{uniformly in } s\in(0,1].\end{aligned}$$

In the particular case $s=1$, Theorem 2.3.13 entails

$$\sqrt{n(1-\tau_n)}\left(\frac{\widetilde{\xi}_{\tau_n}}{\xi_{\tau_n}}-1\right)\xrightarrow{d}\gamma^2\sqrt{\gamma^{-1}-1}\int_0^1 W(t)\,t^{-\gamma-1}\,dt$$

where W denotes a standard Brownian motion. The right-hand side is a centrred Gaussian random variable, whose variance is

$$\gamma^3(1-\gamma)\int_0^1\int_0^1 \min(s,t)(st)^{-\gamma-1}\,ds\,dt = \frac{2\gamma^3}{1-2\gamma}.$$

We do therefore recover Theorem 2.3.5 (i.e. Theorem 2 in [12]), subject to the additional condition $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) = \mathrm{O}(1)$, but under the reduced moment condition $\mathbb{E}|Y_-|^2<\infty$. Note that the bias condition $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) = \mathrm{O}(1)$ is also required in order to establish the desired approximation (2.3.20) for the tail quantile process, and is instrumental in the evaluation of the bias of extreme value estimators. The conditions $\gamma\in(0,1/2)$ and $\mathbb{E}|Y_-|^2<\infty$, meanwhile, essentially guarantee that the loss variable Y has a finite variance.

Nevertheless, Theorem 2.3.13 does not compare directly the tail expectile process with its population counterpart $s \mapsto \xi_{1-(1-\tau_n)s}$. Our second main result analyses the gap between these two quantities. Such a comparison is particularly relevant when developing the asymptotic theory for integrals of the tail expectile process, as discussed below in the estimation of the extreme, expectile-based Expected Shortfall. This result cannot be obtained as a direct corollary of Theorem 2.3.13, because Proposition 2.3.12(ii) is not a uniform result.

Theorem 2.3.14. *If the conditions of Theorem 2.3.13 hold with $\rho < 0$, then there exists a sequence W_n of standard Brownian motions such that, for any $\varepsilon > 0$ sufficiently small,*

$$\begin{aligned}\frac{\widetilde{\xi}_{1-(1-\tau_n)s}}{\xi_{1-(1-\tau_n)s}} &= 1 + \frac{1}{\sqrt{n(1-\tau_n)}}\gamma^2\sqrt{\gamma^{-1}-1}\, s^{\gamma-1}\int_0^s W_n(t)\, t^{-\gamma-1}\, dt \\ &\quad + o_{\mathbb{P}}\left(\frac{s^{-1/2-\varepsilon}}{\sqrt{n(1-\tau_n)}}\right) \quad \text{uniformly in } s \in (0,1].\end{aligned}$$

Note that the Gaussian term appearing in Theorem 2.3.14 is exactly the same as in the approximation of the tail expectile process in Theorem 2.3.13. Both theorems open the door to the analysis of the asymptotic properties of a vast array of functionals of the tail expectile and quantile processes. We discuss in the next sections particular examples where these results can be used to construct general weighted estimators of extreme expectiles and of an expectile-based analogue for the Expected Shortfall risk measure. Theorems 2.3.13 and 2.3.14 will be the key tools when it comes to unravel the asymptotic behaviour of these estimators.

2.3.5.2 Extreme expectile estimation revisited

In this section, we first return to intermediate expectile estimation by combining nonparametric asymmetric least squares estimates with semiparametric quantile-based estimates to construct a more general class of estimators for high expectiles ξ_{τ_n} such that $\tau_n \to 1$ and $n(1-\tau_n) \to \infty$ as $n \to \infty$. Then we extrapolate the obtained estimators to the very high expectile levels that may approach one at an arbitrarily fast rate.

Instead of using either the indirect or direct intermediate expectile estimators $\widehat{\xi}_{\tau_n}$ and $\widetilde{\xi}_{\tau_n}$ of Section 2.3.3.1, one may actually combine these two estimators to define, for $\beta \in \mathbb{R}$, the weighted estimator

$$\overline{\xi}_{\tau_n}(\beta) := \beta\,\widehat{\xi}_{\tau_n} + (1-\beta)\,\widetilde{\xi}_{\tau_n}.$$

The two special cases $\beta = 0$ and $\beta = 1$ correspond to the estimators $\widetilde{\xi}_{\tau_n}$ and $\widehat{\xi}_{\tau_n}$, respectively. In the sequel we consider that the tail index estimator $\widehat{\gamma} = \widehat{\gamma}_{\tau_n}$ used in the construction of the indirect estimator $\widehat{\xi}_{\tau_n}$ is the Hill estimator (2.3.5) with $k = \lfloor n(1-\tau_n)\rfloor$.

The limiting distribution of the linear combination $\overline{\xi}_{\tau_n}(\beta)$ crucially relies on the asymptotic dependence structure between the tail expectile and quantile processes established in Theorem 2.3.13, since $\overline{\xi}_{\tau_n}(\beta)$ is built on both of these processes. More specifically, it relies on the following asymptotic dependence structure between the Hill estimator $\widehat{\gamma}_{\tau_n}$ and the intermediate sample quantile $\widehat{q}_{\tau_n}$ and expectile $\widetilde{\xi}_{\tau_n}$.

Theorem 2.3.15. *Suppose that $\mathbb{E}|Y_-|^2 < \infty$. Assume further that condition $\mathcal{C}_2(\gamma,\rho,A)$ holds, with $0 < \gamma < 1/2$. Let $\tau_n \to 1$ be such that $n(1-\tau_n) \to \infty$, and suppose that the bias condition $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) \to \lambda_1 \in \mathbb{R}$ is satisfied. Then,*

$$\sqrt{n(1-\tau_n)}\left(\widehat{\gamma}_{\tau_n} - \gamma, \frac{\widehat{q}_{\tau_n}}{q_{\tau_n}} - 1, \frac{\widetilde{\xi}_{\tau_n}}{\xi_{\tau_n}} - 1\right) \xrightarrow{d} \mathcal{N}(\mathfrak{m}, \mathfrak{V})$$

where $\mathfrak{m}$ is the 1×3 vector

$$\mathfrak{m} := \left(\frac{\lambda_1}{1-\rho}, 0, 0\right),$$

and $\mathfrak{V}$ *is the* 3×3 *symmetric matrix with entries*

$$\mathfrak{V}(1,1) = \gamma^2, \quad \mathfrak{V}(1,2) = 0, \quad \mathfrak{V}(1,3) = \frac{\gamma^3}{(1-\gamma)^2}(\gamma^{-1}-1)^{\gamma},$$
$$\mathfrak{V}(2,2) = \gamma^2, \quad \mathfrak{V}(2,3) = \gamma^2\left(\frac{(\gamma^{-1}-1)^{\gamma}}{1-\gamma} - 1\right), \quad \mathfrak{V}(3,3) = \frac{2\gamma^3}{1-2\gamma}.$$

Based on the joint asymptotic normality in Theorem 2.3.15, we obtain the following limit distribution of the weighted estimator $\overline{\xi}_{\tau_n}(\beta)$ for an intermediate level τ_n.

Theorem 2.3.16. *Suppose that the conditions of Theorem 2.3.15 hold with the additional bias condition* $\sqrt{n(1-\tau_n)}/q_{\tau_n} \to \lambda_2 \in \mathbb{R}$. *Then, for any* $\beta \in \mathbb{R}$,

$$\sqrt{n(1-\tau_n)}\left(\frac{\overline{\xi}_{\tau_n}(\beta)}{\xi_{\tau_n}} - 1\right) \xrightarrow{d} \beta\left(\mathfrak{b} + [(1-\gamma)^{-1} - \log(\gamma^{-1}-1)]\Psi + \Theta\right) + (1-\beta)\Xi$$

where the bias component $\mathfrak{b}$ *is* $\mathfrak{b} = \lambda_1\mathfrak{b}_1 + \lambda_2\mathfrak{b}_2$ *with*

$$\begin{aligned}
\mathfrak{b}_1 &= \frac{(1-\gamma)^{-1} - \log(\gamma^{-1}-1)}{1-\rho} - \frac{(\gamma^{-1}-1)^{-\rho}}{1-\gamma-\rho} - \frac{(\gamma^{-1}-1)^{-\rho}-1}{\rho},\\
\mathfrak{b}_2 &= -\gamma(\gamma^{-1}-1)^{\gamma}\mathbb{E}(Y),
\end{aligned}$$

and (Ψ, Θ, Ξ) *is a trivariate Gaussian centred random vector with covariance matrix* $\mathfrak{V}$ *as in Theorem 2.3.15.*

When $\beta = 1$, we recover the convergence of the "indirect" estimator $\widehat{\xi}_{\tau_n}$ obtained in Corollary 2.3.4. When $\beta = 0$, we get the convergence of the "direct" estimator $\widetilde{\xi}_{\tau_n}$ stated in Theorem 2.3.5. [These are respectively Corollary 2 and Theorem 2 in [12].]

In the very far tails where the expectile level $\tau = \tau_n' \to 1$ is such that $n(1-\tau_n') \to c \in [0,\infty)$, just like the initial estimators $\widetilde{\xi}_{\tau_n}$ and $\widehat{\xi}_{\tau_n}$, the estimator $\overline{\xi}_{\tau_n}(\beta)$ becomes unstable and inconsistent due to data sparsity. We use the same Weissman-inspired device as in Section 2.3.3.2 to define the following class of extreme expectile estimators:

$$\overline{\xi}^{\star}_{\tau_n'}(\beta) := \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}_{\tau_n}} \overline{\xi}_{\tau_n}(\beta). \tag{2.3.21}$$

The two special cases $\beta = 0$ and $\beta = 1$ correspond to the previously studied, extrapolated direct and indirect expectile estimators. The next theorem gives the asymptotic behaviour of this generalised extreme expectile estimator $\overline{\xi}^{\star}_{\tau_n'}(\beta)$. Here again $\widehat{\gamma}_{\tau_n}$ is the Hill estimator computed on the $k = \lfloor n(1-\tau_n)\rfloor$ top order statistics in the available data.

Theorem 2.3.17. *Suppose that the conditions of Theorem 2.3.16 hold. Assume also that* $\rho < 0$ *and* $n(1-\tau_n') \to c < \infty$ *with* $\sqrt{n(1-\tau_n)}/\log[(1-\tau_n)/(1-\tau_n')] \to \infty$. *Then, for any* $\beta \in \mathbb{R}$,

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\overline{\xi}^{\star}_{\tau_n'}(\beta)}{\xi_{\tau_n'}} - 1\right) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda_1}{1-\rho}, \gamma^2\right).$$

One can observe that the limiting distribution of $\overline{\xi}^{\star}_{\tau_n'}(\beta)$ is controlled by the asymptotic distribution of $\widehat{\gamma}_{\tau_n}$. In particular, in the cases $\beta = 1$ and $\beta = 0$, we exactly recover Corollaries 2.3.7 and 2.3.8 on the convergence of the extrapolated indirect and direct expectile estimators. [These are respectively Corollary 3 and Corollary 4 in [12].]

2.3.5.3 Estimation of expectile-based tail Expected Shortfall

Background An important alternative to VaR_τ is Expected Shortfall at level τ. Recall that this is defined as

$$\mathrm{QES}_\tau := \frac{1}{1-\tau}\int_\tau^1 q_t\,dt. \tag{2.3.22}$$

When Y is continuous, QES_τ is identical to the Conditional Tail Expectation defined by $\mathrm{QCTE}_\tau := \mathbb{E}[Y|Y > q_\tau]$. Both QES_τ and QCTE_τ can then be interpreted as the average loss incurred in the event of a loss higher than VaR_τ. Note though that QES_τ defines a coherent risk measure but QCTE_τ does not in general (see [41, 257]).

Expectile-based Expected Shortfall Motivated by the merits and good properties of expectiles, [242] has introduced an expectile-based form of Expected Shortfall (ES) as the expectation $\mathrm{XCTE}_\tau := \mathbb{E}[Y|Y > \xi_\tau]$ of exceedances beyond the τth expectile ξ_τ of the distribution of Y. This expectile-based CTE was actually implemented by [242] only as an intermediate instrument for the ultimate interest in estimating the conventional quantile-based form QCTE_τ, or equivalently, the coherent version QES_τ under the continuity assumption on Y. Although the interpretability of XCTE_τ is straightforward, its coherence as a proper risk measure has been an open question so far.

This is now elucidated below in Proposition 2.3.18, showing the failure of XCTE_τ to fulfill the coherence property in general. Instead, by analogy with QES_τ itself, we propose to use the new expectile-based form of ES

$$\mathrm{XES}_\tau := \frac{1}{1-\tau}\int_\tau^1 \xi_t\,dt, \tag{2.3.23}$$

obtained by substituting the expectile ξ_t in place of the quantile q_t in the standard form (2.3.22) of ES. It turns out that, in contrast to XCTE_τ, the new risk measure XES_τ is coherent in general. This is the focus of the following result.

Proposition 2.3.18. *For all $\tau \geq 1/2$,*

(i) XES_τ induces a coherent risk measure;

(ii) $XCTE_\tau$ is neither monotonic nor subadditive in general, and hence does not induce a coherent risk measure.

The coherence property of XES_τ, contrary to that of QES_τ, is actually a straightforward consequence of the coherence of the expectile-based risk measure ξ_τ, for $\tau \geq 1/2$.

Next, we show under the heavy-tailed assumption that XES_τ is asymptotically equivalent to XCTE_τ as $\tau \to 1$, and hence inherits its direct meaning as the conditional expectation $\mathbb{E}[Y|Y > \xi_\tau]$ for all τ large enough.

Proposition 2.3.19. *Assume that $\mathbb{E}|Y_-| < \infty$ and that Y has a heavy-tailed distribution with tail index $0 < \gamma < 1$. Then*

$$\frac{XES_\tau}{QES_\tau} \sim \frac{\xi_\tau}{q_\tau} \sim \frac{XCTE_\tau}{QCTE_\tau} \quad \text{and} \quad \frac{XES_\tau}{\xi_\tau} \sim \frac{1}{1-\gamma} \sim \frac{XCTE_\tau}{\xi_\tau} \quad \text{as} \quad \tau \to 1.$$

Propositions 2.3.18 and 2.3.19 then afford arguments to justify that the new form XES_τ of expectile-based ES provides a better alternative to XCTE_τ not only as a proper risk measure, but also as an intermediate tool for estimating the classical quantile-based version QES_τ. Indeed, XES_τ is coherent and keeps the intuitive meaning of XCTE_τ as a conditional expectation when $\tau \to 1$, since $\mathrm{XES}_\tau \sim \mathrm{XCTE}_\tau$. Most importantly, XES_τ may be adopted as a reasonable alternative to QES_τ itself. As is the case in the duality (2.3.4) between the expectile ξ_τ and the VaR q_τ, the choice in any risk analysis between the expectile-based form XES_τ and its quantile-based analogue QES_τ will then depend on the value at hand of $\gamma \lessgtr \frac{1}{2}$. More precisely, the quantity XES_τ will be more (respectively, less) extreme than QES_τ and hence more (respectively, less) pessimistic or conservative in practice, for all τ large enough, if $\gamma > \frac{1}{2}$ (respectively, $\gamma < \frac{1}{2}$).

The connections in Proposition 2.3.19 are very useful when it comes to interpreting and proposing estimators for XES_τ. Also, by considering the second-order regular variation condition $\mathcal{C}_2(\gamma,\rho,A)$, one may establish a precise control of the remainder term which arises in the asymptotic equivalent $\mathrm{XES}_\tau/\xi_\tau \sim (1-\gamma)^{-1}$.

Proposition 2.3.20. *Assume that $\mathbb{E}|Y_-|<\infty$. Assume further that condition $\mathcal{C}_2(\gamma,\rho,A)$ holds, with $0<\gamma<1$. Then, as $\tau\to 1$,*

$$\begin{aligned}\frac{XES_\tau}{\xi_\tau} &= \frac{1}{1-\gamma}\left(1-\frac{\gamma^2(\gamma^{-1}-1)^\gamma}{q_\tau}(\mathbb{E}(Y)+\mathrm{o}(1))\right.\\ &\qquad \left.+\frac{1-\gamma}{(1-\gamma-\rho)^2}(\gamma^{-1}-1)^{-\rho}A((1-\tau)^{-1})(1+\mathrm{o}(1))\right).\end{aligned}$$

This result will prove instrumental when examining the asymptotic properties of our tail expectile-based ES estimators below.

Estimation and asymptotics Propositions 2.3.12(i) and 2.3.20 indicate that the expectile-based ES satisfies a regular variation property in the same way that quantiles and expectiles do. To estimate an extreme value $\mathrm{XES}_{\tau_n'}$, where $\tau_n'\to 1$ and $n(1-\tau_n')\to c<\infty$, we may therefore start by estimating XES_{τ_n}, with τ_n being an intermediate level, before extrapolating this estimator to properly extreme levels using an estimator of the tail index γ. A natural estimator of XES_{τ_n} is its direct empirical counterpart:

$$\widetilde{\mathrm{XES}}_{\tau_n} := \frac{1}{1-\tau_n}\int_{\tau_n}^1 \widetilde{\xi}_t\,dt,$$

obtained simply by replacing ξ_t in (2.3.23) with its sample version $\widetilde{\xi}_t$. Since this estimator is a linear functional of the tail empirical expectile process, Theorem 2.3.14 is more adapted than Theorem 2.3.13 for the analysis of its asymptotic distribution.

Theorem 2.3.21. *Under the conditions of Theorem 2.3.14,*

$$\sqrt{n(1-\tau_n)}\left(\frac{\widetilde{\mathrm{XES}}_{\tau_n}}{\mathrm{XES}_{\tau_n}}-1\right)\xrightarrow{d}\mathcal{N}\left(0,\frac{2\gamma^3(1-\gamma)(3-4\gamma)}{(1-2\gamma)^3}\right).$$

On the basis of Proposition 2.3.12(ii) and then Proposition 2.3.19, we have for $\tau_n'>\tau_n\to 1$ that

$$\frac{\mathrm{XES}_{\tau_n'}}{\mathrm{XES}_{\tau_n}}\sim\frac{\xi_{\tau_n'}}{\xi_{\tau_n}}\approx\left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\gamma}.$$

Therefore, to estimate $\mathrm{XES}_{\tau_n'}$ at an arbitrary extreme level τ_n', we replace γ by the Hill estimator $\widehat{\gamma}_{\tau_n}$ and XES_{τ_n} at an intermediate level τ_n by the estimator $\widetilde{\mathrm{XES}}_{\tau_n}$ to get

$$\widetilde{\mathrm{XES}}^\star_{\tau_n'} := \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}_{\tau_n}}\widetilde{\mathrm{XES}}_{\tau_n}. \tag{2.3.24}$$

The next result analyses the convergence of this Weissman-type estimator.

Theorem 2.3.22. *Assume that the conditions of Theorem 2.3.17 hold. Then*

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\widetilde{\mathrm{XES}}^\star_{\tau_n'}}{\mathrm{XES}_{\tau_n'}}-1\right)\xrightarrow{d}\mathcal{N}\left(\frac{\lambda_1}{1-\rho},\gamma^2\right).$$

One can also design alternative options for estimating $\mathrm{XES}_{\tau_n'}$ by using the asymptotic connections in Proposition 2.3.19. For example, the asymptotic equivalence $\mathrm{XES}_{\tau_n'}\sim(1-\gamma)^{-1}\xi_{\tau_n'}$, established therein, suggests that $\mathrm{XES}_{\tau_n'}$ can be estimated consistently by substituting the tail quantities γ and $\xi_{\tau_n'}$ with their consistent estimators $\widehat{\gamma}_{\tau_n}$ and $\overline{\xi}^\star_{\tau_n'}(\beta)$ described in (2.3.5) and (2.3.21), respectively. This yields the extrapolated estimator

$$\overline{\mathrm{XES}}^\star_{\tau_n'}(\beta) := [1-\widehat{\gamma}_{\tau_n}]^{-1}\,\overline{\xi}^\star_{\tau_n'}(\beta), \tag{2.3.25}$$

with a weight $\beta\in\mathbb{R}$, whose asymptotic normality is established in the following result.

Theorem 2.3.23. *Assume that the conditions of Theorem 2.3.17 hold. Then, for any $\beta \in \mathbb{R}$,*

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\overline{\mathrm{XES}}^\star_{\tau_n'}(\beta)}{\mathrm{XES}_{\tau_n'}}-1\right)\xrightarrow{d}\mathcal{N}\left(\frac{\lambda_1}{1-\rho},\gamma^2\right).$$

Theorems 2.3.22 and 2.3.23 are, like Theorem 2.3.17, derived by noticing that, on the one hand, the asymptotic behaviours of $\widetilde{\mathrm{XES}}^\star_{\tau_n'}$ and $\overline{\xi}^\star_{\tau_n'}(\beta)$ are controlled by the asymptotic behaviour of the extrapolation factor $\{(1-\tau_n')/(1-\tau_n)\}^{-\widehat{\gamma}_{\tau_n}}$, which is itself governed by that of $\widehat{\gamma}_{\tau_n}$ only. This is why, in Theorems 2.3.22 and 2.3.23, the asymptotic distributions of $\widetilde{\mathrm{XES}}^\star_{\tau_n'}$ and $\overline{\mathrm{XES}}^\star_{\tau_n'}(\beta)$ coincide. On the other hand, the non-random remainder term coming from the use of Proposition 2.3.19 can be controlled thanks to Proposition 2.3.20.

2.3.6 Extreme level selection [1, 12]

An important question that remains to be addressed is the choice of the extreme expectile level τ_n' in the risk measures $\xi_{\tau_n'}$, $\mathrm{XMES}(\tau_n')$ and $\mathrm{XES}_{\tau_n'}$.

In the case of quantile-based risk measures q_{α_n}, $\mathrm{QMES}(\alpha_n)$ and QES_{α_n}, it is customary to choose tail probabilities $\alpha_n \to 1$ with $n(1-\alpha_n) \to c$, a finite nonnegative constant, to allow for more "prudent" risk management. In the case of expectiles, we propose to select τ_n' so that each expectile-based risk measure has the same intuitive interpretation as its quantile-based analogue. This translates into choosing τ_n' such that $\xi_{\tau_n'} \equiv q_{\alpha_n}$ for a given relative frequency α_n. It has been suggested in [52] to pick out τ_n' which satisfies $\xi_{\tau_n'} \equiv q_{\alpha_n}$, but this is restricted to a Gaussian random variable Y. Here, we wish to extend this elegant device to a random variable Y in our heavy-tailed framework.

Thanks to the connection (2.3.2), it is immediate from $\xi_{\tau_n'} \equiv q_{\alpha_n}$ that $\tau_n'(\alpha_n) := \tau_n'$ satisfies

$$1-\tau_n'(\alpha_n) = \frac{\mathbb{E}\left\{|Y-q_{\alpha_n}|\,\mathbb{I}\{Y>q_{\alpha_n}\}\right\}}{\mathbb{E}\,|Y-q_{\alpha_n}|}.$$

As a matter of fact, under the model assumption of Pareto-type tails, it turns out that the expectile level $\tau_n'(\alpha_n)$ depends asymptotically only on the quantile level α_n and on the tail index γ, but not on the quantile q_{α_n} itself.

Proposition 2.3.24. *Suppose Y is heavy-tailed with a tail index $0<\gamma<1$. Then*

$$1-\tau_n'(\alpha_n) \sim (1-\alpha_n)\,\frac{\gamma}{1-\gamma}, \quad \textit{as} \quad n\to\infty.$$

Hence, by substituting the estimated value

$$\widehat{\tau}_n'(\alpha_n) = 1-(1-\alpha_n)\,\frac{\widehat{\gamma}}{1-\widehat{\gamma}} \tag{2.3.26}$$

in place of τ_n', both extreme expectile estimators $\widehat{\xi}^\star_{\tau_n'}$ in (2.3.10) and $\widetilde{\xi}^\star_{\tau_n'}$ in (2.3.11) estimate the same Value at Risk $\xi_{\tau_n'(\alpha_n)} \equiv q_{\alpha_n}$ as the Weissman quantile estimator $\widehat{q}^\star_{\alpha_n}$ in (2.3.9). It is easily seen that the latter estimator is actually identical to the indirect expectile estimator $\widehat{\xi}^\star_{\widehat{\tau}_n'(\alpha_n)}$. Indeed, we have in view of (2.3.9), (2.3.10) and (2.3.26) that

$$\widehat{\xi}^\star_{\widehat{\tau}_n'(\alpha_n)} = \left(\widehat{\gamma}^{-1}-1\right)^{-\widehat{\gamma}}\left(\frac{1-\widehat{\tau}_n'(\alpha_n)}{1-\tau_n}\right)^{-\widehat{\gamma}}\widehat{q}_{\tau_n} = \widehat{q}^\star_{\alpha_n}.$$

This quantile-based estimator $\widehat{q}^\star_{\alpha_n} \equiv \widehat{\xi}^\star_{\widehat{\tau}_n'(\alpha_n)}$ may be criticised for its reliance on a single order statistic $\widehat{q}_{\tau_n} = Y_{n-\lfloor n(1-\tau_n)\rfloor,n}$, and hence because it may not respond properly to the very extreme losses. By contrast, the direct expectile-based estimator $\widetilde{\xi}^\star_{\widehat{\tau}_n'(\alpha_n)}$ relies on the asymmetric least squares estimator $\widetilde{\xi}_{\tau_n}$, and hence bears much better the burden of representing a sensitive risk measure to the magnitude of infrequent catastrophic losses. The next result shows that the asymptotic behaviour of the original extrapolated estimators $\widehat{\xi}^\star_{\tau_n'}$ and $\widetilde{\xi}^\star_{\tau_n'}$, established in Corollaries 2.3.7 and 2.3.8, remains the same for the resulting composite estimators $\widehat{\xi}^\star_{\widehat{\tau}_n'(\alpha_n)}$ and $\widetilde{\xi}^\star_{\widehat{\tau}_n'(\alpha_n)}$, under the same technical conditions.

Theorem 2.3.25. *(i) Suppose the conditions of Corollary 2.3.7 hold with α_n in place of τ_n'. Then*

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\alpha_n)]}\left(\frac{\widehat{\xi}^{\star}_{\widehat{\tau}_n'(\alpha_n)}}{q_{\alpha_n}}-1\right)\xrightarrow{d}\Gamma.$$

(ii) Suppose the conditions of Corollary 2.3.8 hold with α_n in place of τ_n'. Then

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\alpha_n)]}\left(\frac{\widetilde{\xi}^{\star}_{\widehat{\tau}_n'(\alpha_n)}}{q_{\alpha_n}}-1\right)\xrightarrow{d}\Gamma.$$

Let us now turn to $\widetilde{\text{XMES}}^{\star}(\widehat{\tau}_n'(\alpha_n))$ in (2.3.15) and $\widehat{\text{XMES}}^{\star}(\widehat{\tau}_n'(\alpha_n))$ in (2.3.18) that estimate the same marginal expected shortfall $\text{XMES}(\tau_n'(\alpha_n))\equiv\text{QMES}(\alpha_n)$ as the estimator $\widehat{\text{QMES}}^{\star}(\alpha_n)$ defined in (2.3.19). Actually $\widehat{\text{XMES}}^{\star}(\widehat{\tau}_n'(\alpha_n))$ is nothing but $\widehat{\text{QMES}}^{\star}(\alpha_n)$.

Theorem 2.3.26. *(i) Suppose the conditions of Theorem 2.3.10 hold with α_n in place of τ_n'. Then*

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\alpha_n)]}\left(\frac{\widetilde{XMES}^{\star}(\widehat{\tau}_n'(\alpha_n))}{QMES(\alpha_n)}-1\right)\xrightarrow{d}\Gamma.$$

(ii) Suppose the conditions of Theorem 2.3.11 hold with α_n in place of τ_n'. Then

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\alpha_n)]}\left(\frac{\widehat{XMES}^{\star}(\widehat{\tau}_n'(\alpha_n))}{QMES(\alpha_n)}-1\right)\xrightarrow{d}\Gamma.$$

Finally, by substituting $\widehat{\tau}_n'(\alpha_n)$ in place of $\tau_n'\equiv\tau_n'(\alpha_n)$ in the extrapolated estimators $\widetilde{\text{XES}}^{\star}_{\tau_n'}$ and $\overline{\text{XES}}^{\star}_{\tau_n'}(\beta)$ described in (2.3.24) and (2.3.25), we obtain composite estimators that estimate $\text{XES}_{\tau_n'(\alpha_n)}\sim\text{QES}_{\alpha_n}$, by Proposition 2.3.19. The convergence results in Theorems 2.3.22 and 2.3.23 of the extrapolated estimators $\widetilde{\text{XES}}^{\star}_{\tau_n'}$ and $\overline{\text{XES}}^{\star}_{\tau_n'}(\beta)$ still hold true for their composite versions as estimators of QES_{α_n}, with the same technical conditions.

Theorem 2.3.27. *Suppose the conditions of Theorem 2.3.17 hold with α_n in place of τ_n'. Then, for any $\beta\in\mathbb{R}$,*

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\alpha_n)]}\left(\frac{\widetilde{\text{XES}}^{\star}_{\widehat{\tau}_n'(\alpha_n)}}{\text{QES}_{\alpha_n}}-1\right)\xrightarrow{d}\mathcal{N}\left(\frac{\lambda_1}{1-\rho},\gamma^2\right),$$

$$\textit{and}\quad\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\alpha_n)]}\left(\frac{\overline{\text{XES}}^{\star}_{\widehat{\tau}_n'(\alpha_n)}(\beta)}{\text{QES}_{\alpha_n}}-1\right)\xrightarrow{d}\mathcal{N}\left(\frac{\lambda_1}{1-\rho},\gamma^2\right).$$

2.3.7 Finite-sample study [1, 12]

2.3.7.1 Finite-sample performance on simulated data

Detailed simulation results on the finite-sample performance of our estimators can be found in [1, 12]. We obtain that:

- Overall, the direct estimation method (via asymmetric least squares) is more efficient relative to the indirect method in the case of real-valued profit-loss variables, whereas the rival indirect method tends to be superior in the case of non-negative loss distributions. The latter method seems to be also better in the case of extremely heavy tails.
- Taking a weighting parameter β is beneficial in certain cases of non-negative loss distributions.
- These conclusions carry over to when the objective is to use extreme expectile-based risk measures as a way to estimate their extreme quantile-based counterparts.

2.3.7.2 Data example 3: The Society of Actuaries medical data set [12]

The SOA Group Medical Insurance Large Claims Database records all the claim amounts exceeding US \$25,000 over the period 1991-92. As in [47], we only deal here with the $n = 75{,}789$ claims for 1991. The histogram and scatterplot shown in Figure 2.4 (a) give evidence of an important right-skewness. Insurance companies are then interested in estimating the worst tail value of the corresponding loss severity distribution.

One way of measuring this value at risk is by considering the Weissman quantile estimate

$$\widehat{q}^{\star}_{\alpha_n} = Y_{n-k,n}\left(\frac{k}{np_n}\right)^{\widehat{\gamma}_H}$$

as described in (2.3.9), where $\widehat{\gamma}_H$ is the Hill estimator defined in (2.3.5), with $\alpha_n = 1 - p_n$ and $\tau_n = 1 - k/n$. According to the earlier study of [47] (p.123), insurers typically are interested in $p_n = 1/100{,}000 \approx 1/n$ for these medical insurance data, that is, in an estimate of the claim amount that will be exceeded (on average) only once in 100,000 cases. Similar recent studies in the context of the backtesting problem, which is crucial in the current Basel III regulatory framework, are [72, 141], who estimate quantiles exceeded on average once every 100 cases with sample sizes of the order of hundreds. Figure 2.4 (b) shows the quantile-VaR estimates $\widehat{q}^{\star}_{\alpha_n}$ against the sample fraction k (solid line). Here, a stable region appears for k from 150 up to 500, leading to an estimate between 3.73 and 4.12 million. This estimate does not exceed the sample maximum $Y_{n,n} = 4{,}518{,}420$ (indicated by the horizontal line), which is consistent with the earlier analysis of [47] (p.125 and p.159).

The alternative expectile-based estimator $\widetilde{\xi}^{\star}_{\widehat{\tau}'_n(\alpha_n)}$ introduced in Section 2.3.6, which estimates the same VaR $q_{\alpha_n} \equiv \xi_{\tau'_n(\alpha_n)}$ as the quantile-based estimator

$$\widehat{q}^{\star}_{\alpha_n} \equiv \widehat{\xi}^{\star}_{\widehat{\tau}'_n(\alpha_n)},$$

is also graphed in Figure 2.4 (b) in dashed line. As an asymmetric least-squares estimator, it is more affected by the infrequent great claim amounts visualised in the top panel of Figure 2.4. Its plot indicates a more conservative risk measure between 3.92 and 4.33 million, over the stable region $k \in [150, 500]$.

2.3.7.3 Data example 4: Financial data of three large American banks [1, 12]

We consider the same investment banks as in the studies of [64, 68], namely Goldman Sachs, Morgan Stanley and T. Rowe Price. For the three banks, the dataset consists of the loss returns, i.e. the negative log-returns (X_i) on their equity prices at a daily frequency from July 3rd, 2000, to June 30th, 2010. We follow the same setup as in [68] to extract, for the same time period, daily loss returns (Y_i) of a value-weighted market index aggregating three markets: the New York Stock Exchange, American Express Stock Exchange and the National Association of Securities Dealers Automated Quotation system.

Marginal Expected Shortfall estimation given an extreme loss in the global market [12] In [68], the quantile-based marginal expected shortfall $\text{QMES}(\alpha_n) = \mathbb{E}\{X|Y > q_{Y,\alpha_n}\}$ is estimated by $\widehat{\text{QMES}}^{\star}(\alpha_n)$, where $\alpha_n = 1 - 1/n = 1 - 1/2{,}513$, with two intermediate sequences involved in $\widehat{\gamma}_X$ and $\widehat{\text{QMES}}(\tau_n)$ to be chosen in two steps. Instead, we use our expectile-based method to estimate $\text{QMES}(\alpha_n) \equiv \text{XMES}(\tau'_n(\alpha_n)) = \mathbb{E}\{X|Y > \xi_{Y,\tau'_n(\alpha_n)}\}$, with the same extreme relative frequency α_n that corresponds to a once-per-decade systemic event. We employ the rival estimator $\widehat{\text{QMES}}^{\star}(\alpha_n)$ with the same intermediate sequence $\tau_n = 1 - k/n$ in both $\widehat{\gamma}_X$ and $\widehat{\text{QMES}}(\tau_n)$. The conditions required by the procedure were already checked empirically in [68]. It only remains to verify that $\gamma_Y < 1/2$ as it is the case for γ_X. This assumption is confirmed by the plot of the Hill estimates of γ_Y against the sample fraction k (dashed line) in Figure 2.5 (a). Indeed, the first stable region appears for $k \in [70, 100]$ with an averaged estimate $\widehat{\gamma}_Y = 0.35$. Hence, by Proposition 2.3.9, the estimates $\widehat{\text{XMES}}^{\star}(\alpha_n)$ and

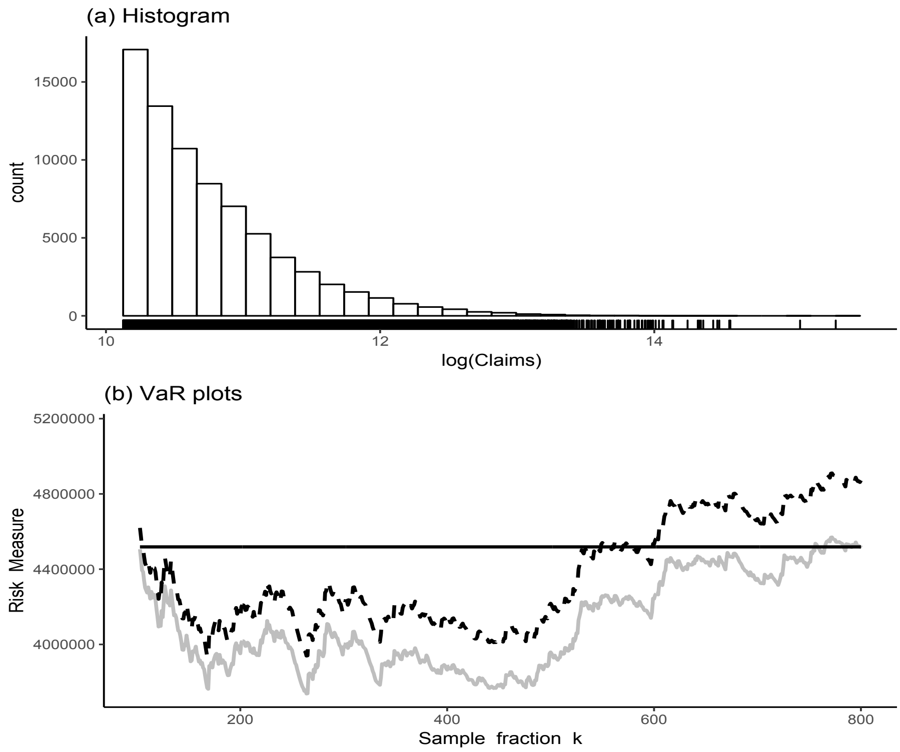

Figure 2.4: SOA Group Medical Insurance data. (a) Histogram and scatterplot of the log-claim amounts. (b) The VaR plots $k \mapsto \widetilde{\xi}^{\star}_{\widehat{\tau}'_n(\alpha_n)}(k)$ in dashed line and $k \mapsto \widehat{q}^{\star}_{\alpha_n}(k)$ in solid line, along with the sample maximum $Y_{n,n}$ in horizontal line.

$\widetilde{\text{XMES}}^{\star}(\alpha_n)$ are expected to be less extreme than the benchmark values $\widehat{\text{QMES}}^{\star}(\alpha_n)$. As a matter of fact, both $\widehat{\text{XMES}}^{\star}(\alpha_n)$ and $\widetilde{\text{XMES}}^{\star}(\alpha_n)$ estimate the less extreme risk measure $\text{XMES}(\alpha_n)$ and not the desired intuitive tail measure $\text{XMES}(\tau'_n(\alpha_n)) \equiv \text{QMES}(\alpha_n)$.

The interest here is rather on the composite estimators $\widehat{\text{XMES}}^{\star}(\widehat{\tau}'_n(\alpha_n))$ and $\widetilde{\text{XMES}}^{\star}(\widehat{\tau}'_n(\alpha_n))$, where $\widehat{\text{XMES}}^{\star}(\widehat{\tau}'_n(\alpha_n))$ is actually nothing but $\widehat{\text{QMES}}^{\star}(\alpha_n)$. The two rival estimates $\widehat{\text{QMES}}^{\star}(\alpha_n)$ and $\widetilde{\text{XMES}}^{\star}(\widehat{\tau}'_n(\alpha_n))$ represent the average daily loss return for a once-per-decade market crisis. They are graphed in Figure 2.5 (b)-(d) as functions of k for each bank: (b) Goldman Sachs; (c) Morgan Stanley; (d) T. Rowe Price. The first stable regions of the plots (b)-(d) appear, respectively, for $k \in [80, 105]$, $k \in [90, 140]$ and $k \in [75, 100]$. The final estimates based on averaging the estimates from these stable regions are reported in the left-hand side of Table 2.6, along with the asymptotic 95% confidence intervals derived from Theorem 2.3.26 with the bias condition $\lambda_2 = 0$ (the asymptotic distribution then being $\mathcal{N}(0, \gamma_X^2)$ due to the use of the Hill estimator of γ_X, see the discussion below Theorem 2.3.3). It may be seen that both expectile- and quantile-based MES levels for Goldman Sachs and T. Rowe Price are almost equal. However, the MES levels for Morgan Stanley are substantially higher than those for Goldman Sachs and T. Rowe Price. It may also be noted that the estimates $\widehat{\text{QMES}}^{\star}(\alpha_n)$, obtained here with a single intermediate sequence, are slightly smaller than those obtained in Table 1 of [68] by using two intermediate sequences. Also, these quantile-based estimates appear to be less conservative than our asymmetric least squares-based estimates, but not by much: this minor difference can already be visualised in Figure 2.5 (b)-(d), where the plots of $\widehat{\text{QMES}}^{\star}(\alpha_n)$, in dashed line, and $\widetilde{\text{XMES}}^{\star}(\widehat{\tau}'_n(\alpha_n))$, in solid line, exhibit a very similar evolution for the three banks.

In our theoretical results we do not enter into the important question of serial dependence. We only consider independent and identically distributed random vectors $(X_1, Y_1), \ldots, (X_n, Y_n)$. One way to reduce substantially the potential serial dependence in this application is by using lower frequency

	Daily loss		**Weekly loss**	
Bank	$\widetilde{\text{XMES}}^\star(\widehat{\tau}_n'(\alpha_n))$	$\widehat{\text{QMES}}^\star(\alpha_n)$	$\widetilde{\text{XMES}}^\star(\widehat{\tau}_n'(\alpha_n))$	$\widehat{\text{QMES}}^\star(\alpha_n)$
Goldman Sachs	0.3123 (0.20,0.42)	0.3077 (0.19,0.41)	0.3423 (0.17,0.51)	0.3375 (0.16,0.50)
Morgan Stanley	0.5622 (0.34,0.78)	0.5552 (0.33,0.77)	0.6495 (0.26,1.03)	0.6641 (0.26,1.05)
T. Rowe Price	0.3308 (0.25,0.52)	0.3098 (0.23,0.50)	0.3407 (0.19,0.48)	0.3405 (0.19,0.48)

Table 2.6: Expectile- and quantile-based MES of the three investment banks. The second and third columns report the results based on daily loss returns ($n = 2{,}513$ and $\alpha_n = 1 - 1/n$). The last two columns report the results based on weekly loss returns from the same sample period ($n = 522$ and $\alpha_n = 1 - 1/n$). Each MES estimate is followed by the 95% asymptotic confidence interval.

Bank	$\widetilde{\text{XES}}^\star_{\widehat{\tau}_n'(\alpha_n)}$	95% C.I.	$\widehat{\text{QES}}^\star_{\alpha_n}$	95% C.I.	$Y_{n,n}$
Goldman Sachs	0.445	(0.194, 0.620)	0.495	(0.226, 0.680)	0.365
Morgan Stanley	0.817	(0.384, 1.305)	0.883	(0.366, 1.478)	0.904
T. Rowe Price	0.386	(0.213, 0.511)	0.407	(0.216, 0.548)	0.305

Table 2.7: ES levels of the three investment banks, with the 95% confidence intervals and the sample maxima. Results based on weekly loss returns, with $n = 522$ and $\alpha_n = 1 - 1/n$.

data. As suggested by [68], we choose weekly loss returns in the same sample period. This results in a sample of size $n = 522$. The final results, obtained following a similar procedure as the one used for daily data, are reported in the right-hand side of Table 2.6. They are very similar to those obtained in [68] by resorting to two intermediate sequences. Both expectile- and quantile-based MES estimates are qualitatively robust to the change from daily to weekly data: they are still almost equal for Goldman Sachs and T. Rowe Price, while almost twice as high for Morgan Stanley.

Extreme Expected Shortfall estimation [1] We now turn to the estimation of the standard quantile-based expected shortfall QES_{α_n}, or equivalently the expectile-based expected shortfall $\text{XES}_{\tau_n'(\alpha_n)}$, for each bank. We work here with weekly loss returns, for a total sample size of $n = 522$.

In this situation of real-valued profit-loss distributions, our experience with simulated data indicates that the composite estimator $\widetilde{\text{XES}}^\star_{\widehat{\tau}_n'(\alpha_n)}$ provides the best QES_{α_n} estimates in terms of MSE and bias. In the estimation, we employ the intermediate sequence $\tau_n = 1 - k/n$ as before. For our comparison purposes, we use as a benchmark the direct quantile-based estimator $\widehat{\text{QES}}^\star_{\alpha_n}$ of [114]. For each bank, we superimpose in Figure 2.6 the plots of the two competing estimates $\widetilde{\text{XES}}^\star_{\widehat{\tau}_n'(\alpha_n)}$ and $\widehat{\text{QES}}^\star_{\alpha_n}$ against k, as rainbow and dashed black curves respectively, along with their associated asymptotic 95% confidence intervals constructed using their limiting Gaussian distribution. The effect of the Hill estimate $\widehat{\gamma}_H$ on $\widetilde{\text{XES}}^\star_{\widehat{\tau}_n'(\alpha_n)}$ is highlighted by a colour-scheme, ranging from dark red (low $\widehat{\gamma}_H$) to dark violet (high $\widehat{\gamma}_H$). The superiority of the composite expectile-based estimator $\widetilde{\text{XES}}^\star_{\widehat{\tau}_n'(\alpha_n)}$ in terms of plots' stability, including confidence intervals, can clearly be visualised in Figure 2.6 for the three banks. The final estimated ES levels are reported in Table 2.7, along with the asymptotic 95% confidence intervals of the ES. Based on the $\widetilde{\text{XES}}^\star_{\widehat{\tau}_n'(\alpha_n)}$ estimates (in the second column), the ES levels for Goldman Sachs and T. Rowe Price seem to be close (around -38% to -44%), whereas the ES level for Morgan Stanley is almost twice as high (around -81%). It is worth noticing that the difference between the $\widetilde{\text{XES}}^\star_{\widehat{\tau}_n'(\alpha_n)}$ levels for Goldman Sachs and T. Rowe Price is very close to the difference between their respective maxima $Y_{n,n}$. The $\widehat{\text{QES}}^\star_{\alpha_n}$ estimates (in the fourth column) point towards slightly more pessimistic risk measures for the three banks.

2.3.8 Perspectives for future research

Inference in the Weibull domain of attraction Our results are derived under the assumption of a heavy right tail, motivated by applications in insurance and finance. The extension of our results

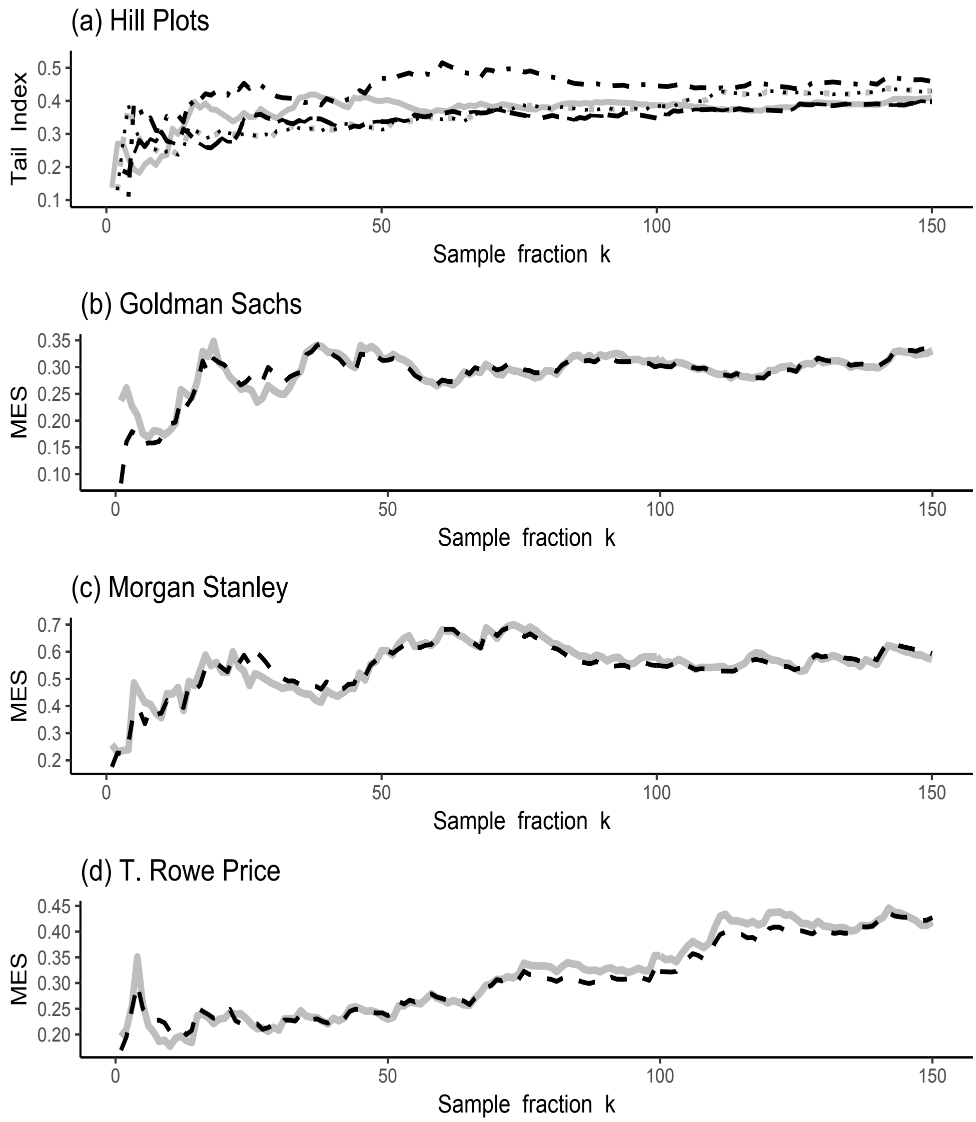

Figure 2.5: (a) Hill estimates $\widehat{\gamma}_Y$ based on daily loss returns of market index (dashed), along with $\widehat{\gamma}_X$ based on daily loss returns of Goldman Sachs (solid), Morgan Stanley (dashed-dotted), and T. Rowe Price (dotted). (b)-(d) The estimates $\widehat{\text{QMES}}^{\star}(\alpha_n)$ in dashed line and $\widetilde{\text{XMES}}^{\star}(\widehat{\tau}_n'(\alpha_n))$ in solid line for the three banks, with $n = 2{,}513$ and $\alpha_n = 1 - 1/n$.

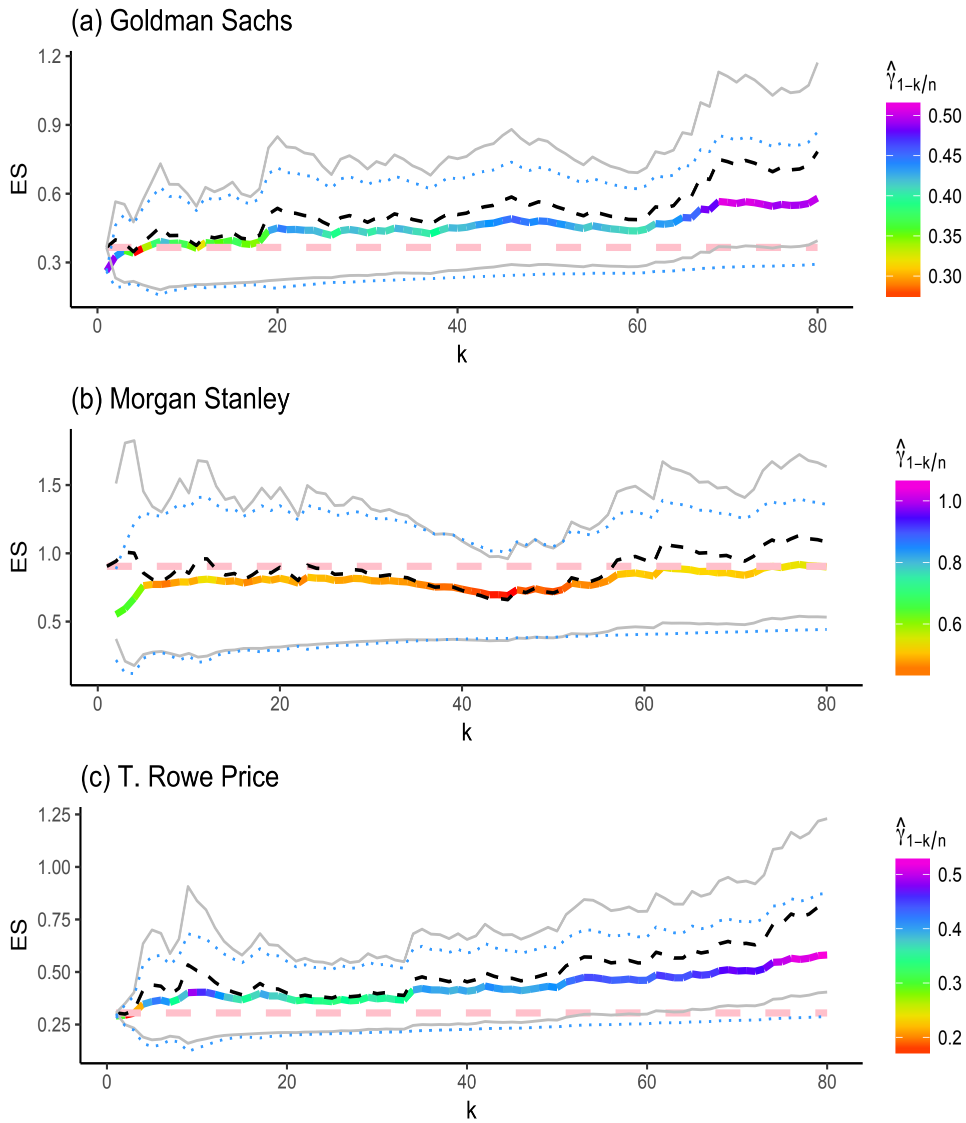

Figure 2.6: Results based on weekly loss returns of the three investment banks: (a) Goldman Sachs, (b) Morgan Stanley, and (c) T. Rowe Price, with $n = 522$ and $\alpha_n = 1 - 1/n$. The estimates $\widetilde{\text{XES}}^{\star}_{\widetilde{\tau}'_n(\alpha_n)}$ as rainbow curve and $\widehat{\text{QES}}^{\star}_{\alpha_n}$ as dashed black curve, along with the associated asymptotic 95% confidence intervals in, respectively, dotted blue lines and solid grey lines. The sample maximum $Y_{n,n}$ is indicated in horizontal dashed pink line.

to the case of an underlying distribution belonging to the Weibull max-domain of attraction is an open problem. A distribution in this domain has a finite right endpoint, making it unlikely to be suitable in applications to the management of catastrophic risk in insurance and finance. In this context, the anticipation is rather that expectiles would allow the development of competitive procedures for right endpoint estimation and thus for monotone frontier estimation, since they bring valuable information about the tail. This problem is of great interest in economics and particularly in the field of economic efficiency analysis, see e.g. [97].

Time series adaptation Our current results are limited to independent and identically distributed observations. Of major interest in finance is the adaptation of our approach to a time-dynamic setting, i.e. the prediction of extreme expectiles in the future conditional on having observed the past. Already [190, 242] have initiated the use of expectiles to estimate VaR and ES in conditional autoregressive expectile models. The use of expectiles to estimate MES may also work by allowing for dynamics in the covariance matrix via a multivariate GARCH model, similarly to the quantile-based method of [64]. From the perspective of extreme values, one way to deal with the heteroscedasticity present in series of financial returns, similarly to e.g. [200] and [201] (see p.263 therein) is by applying our method to residuals standardised by GARCH conditional volatility estimates.

Joint inference for several risk variables Our results focus, for the most part, on a single risk variable Y. Our work on the Marginal Expected Shortfall considers an individual financial risk given an extreme level of risk in the global market, but does not give an idea of the joint, bivariate extreme risk for two companies. An important problem is to develop joint inference procedures for the analysis of extreme individual risks in several actuarial/financial institutions, or several lines of business in a given company. Of particular interest will be the understanding of the joint asymptotic behaviour of several LAWS intermediate expectile estimators attached to different risk variables; this will be crucial to gain insight on how asymptotic dependence between two risk variables influences their extreme expectile estimators. The extrapolation step, meanwhile, can be solved using tools we shall introduce in Section 2.6 if it is carried out using the Hill estimator to estimate the tail index.

2.4 Using L^p-techniques for the estimation of extreme risk measures [3, 10]

2.4.1 Introduction

An intrinsic difficulty with expectiles is that their existence requires $\mathbb{E}|X| < \infty$, which amounts to supposing $\gamma < 1$. Even more seriously, the condition $\gamma < 1/2$ is required to ensure that asymmetric least squares estimators of ξ_{τ_n} are asymptotically Gaussian. Already in the intermediate case, where $n(1-\tau_n) \to \infty$ as $\tau_n \to 1$, good estimates may require in practice $\gamma < 1/4$. Similar concerns occur with the Expected Shortfall (or equivalently, the Conditional Tail Expectation for continuous distributions) and more generally certain extreme Wang distortion risk measures such as those we studied in Section 2.2.5. This restricts appreciably the range of potential applications as may be seen in the actuarial and financial setting from the Danish fire insurance data set considered in Example 4.2 of [215], emerging market stock returns data [154, 195] and exchange rates data [155], as well as from the R package `CASdatasets` where realised values of the tail index γ were found to be larger than $1/4$ in several instances.

Instead of the asymmetric absolute or squared loss, a natural modification of the check function is to use the power loss function $\eta_\tau(x;p) = |\tau - \mathbb{I}\{x \le 0\}||x|^p$, for $p \ge 1$, leading to

$$q_\tau(p) = \underset{q\in\mathbb{R}}{\arg\min}\, \mathbb{E}(\eta_\tau(X-q;p) - \eta_\tau(X;p)).$$

These quantities were called L^p-quantiles by [73] and are special cases of M-quantiles as introduced in [63]. Their existence requires $\mathbb{E}|X|^{p-1} < \infty$. This is a weaker condition, compared with the condition of existence of expectiles, when $p < 2$. The class of L^p-quantiles, with $p \in (1,2)$, steers an advantageous middle course between the robustness of quantiles ($p = 1$) and the sensitivity of expectiles ($p = 2$) to the magnitude of extreme losses. For fixed levels τ staying away from the distribution tails, inference on $q_\tau(p)$ is straightforward using M-estimation theory. The goal of [10], on which we will focus in the first part of this section, is to extend the estimation of $q_\tau(p)$ and its theory far into the upper tail $\tau = \tau_n \to 1$ as $n \to \infty$.

More specifically, we shall establish two estimators of $q_{\tau_n}(p)$ for a general p and unravel their asymptotic behaviour for τ_n at an extremely high level that can be even larger than $1-1/n$, in a framework of heavy tails motivated by the aforementioned financial and actuarial applications. To do so, just as for quantiles and expectiles, we first estimate the intermediate tail L^p-quantiles of order $\tau_n \to 1$ such that $n(1-\tau_n) \to \infty$, and then extrapolate these estimates to the properly extreme L^p-quantile level τ_n which approaches 1 at an arbitrarily fast rate in the sense that $n(1-\tau_n) \to c$, for some finite nonnegative constant c. The main results state the asymptotic normality of our estimators for distributions with tail index $\gamma < [2(p-1)]^{-1}$. As such, unlike expectiles, extreme L^p-quantile estimators cover a larger class of heavy-tailed distributions for $p < 2$. It should also be clear that, in contrast to standard quantiles, generalised L^p-quantiles take into account the whole tail information about the underlying distribution for $p > 1$. These additional benefits raise the following important question: how to make a choice of p in the interval $[1,2]$? This choice is mainly a practical issue that we first pursue here through some simulation experiments. Although the value of p minimising the Mean Squared Error of empirical L^p-quantiles depends on the tail index γ, Monte Carlo evidence indicates that the choice of $p \in (1.2, 1.6)$ guarantees a good compromise for Pareto-type distributions with $\gamma < 1/2$. In contrast, when the empirical estimates are extrapolated to properly extreme levels τ_n, the underlying tail L^p-quantiles seem to be estimated more accurately for $p \in [1, 1.3]$ or $p \in [1.7, 2]$. We explore further this question from a forecasting perspective, trying to perform extreme L^p-quantile estimation accurately on historical data.

Yet, the L^p-quantile approach is not without disadvantages. It does not have an intuitive interpretation as direct as ordinary L^1-quantiles. More precisely, the generalised quantile $q_\tau(p)$ exists, is unique and satisfies

$$\tau = \frac{\mathbb{E}\left[|X - q_\tau(p)|^{p-1}\mathbb{I}\{X \le q_\tau(p)\}\right]}{\mathbb{E}\left[|X - q_\tau(p)|^{p-1}\right]}. \tag{2.4.1}$$

It can thus be interpreted only in terms of the average distance from X in the (nonconvex when $1 < p < 2$) space L^{p-1}. However, one can recover the usual quantiles $q_{\alpha_n} \equiv q_{\alpha_n}(1)$ of extreme order $\alpha_n \to 1$ and their strong intuitive appeal from tail L^p-quantiles $q_{\tau_n}(p)$, $\tau_n \to 1$, that coincide with $q_{\alpha_n}(1)$. Indeed, given a relative frequency of interest α_n, the level τ_n such that $q_{\tau_n}(p) \equiv q_{\alpha_n}(1)$ can be written in closed form as

$$\tau_n = \frac{\mathbb{E}\left[|X - q_{\alpha_n}(1)|^{p-1}\mathbb{I}\{X \leq q_{\alpha_n}(1)\}\right]}{\mathbb{E}\left[|X - q_{\alpha_n}(1)|^{p-1}\right]} \tag{2.4.2}$$

in view of (2.4.1). One can then estimate τ_n via extrapolation techniques before calculating the corresponding L^p-quantile estimators. In this way, we perform tail L^p-quantile estimation as a main tool when the ultimate interest is in estimating the intuitive L^1-quantiles themselves. We can similarly employ tail L^p-quantiles $q_{\tau_n}(p)$ as a tool for estimating extreme expectiles $\xi_{\alpha_n} \equiv q_{\alpha_n}(2)$ by applying again (2.4.1) in conjunction with similar considerations to the above in extreme quantile estimation. Building on the presented extreme L^p-quantile estimators, we construct three different tail expectile estimators and derive their asymptotic normality. Two among these new estimators appear to be competitive with respect to the extreme expectile estimators we studied in Section 2.3 (namely, the estimators of [12]).

The second part of this section presents the contribution of [3], which uses some of the experience and insight gained in the study of extreme L^p-quantiles in order to design a class of tail L^p-indicators that realise a compromise between the sensitivity of the Conditional Tail Expectation and the robustness of the Median Shortfall of [185, 186]. Recall that the latter, at level $\alpha \in (0,1)$, is nothing but the quantile or VaR at level $(1+\alpha)/2$. We will start by showing that these two quantities can be obtained in a new unified framework, which we call the class of *tail L^p-medians*. A tail L^p-median at level $\alpha \in (0,1)$ is obtained, for $p \geq 1$, precisely by calculating the L^p-quantile of order $1/2$ (or L^p-median) of the variable $X \mid X > q(\alpha)$. In particular, we will note that the Median Shortfall is the tail L^1-median, while the Conditional Tail Expectation is the tail L^2-median. Like L^p-quantiles, tail L^p-medians with $p > 1$ depend on both the frequency of the event $\{X > q(\alpha)\}$ and the actual behaviour of X beyond $q(\alpha)$. At the technical level, a condition for a tail L^p-median to exist is that $\gamma < 1/(p-1)$, and it can be empirically estimated at high levels by an asymptotically Gaussian estimator if $\gamma < 1/[2(p-1)]$. When $p \in (1,2)$, the tail L^p-median and the L^p-quantile therefore simultaneously exist and are accurately estimable in a wider class of models than the CTE is.

However, there are a number of differences between tail L^p-medians and L^p-quantiles. For example, the theoretical analysis of the asymptotic behaviour of tail L^p-medians, as $\alpha \to 1^-$ (where throughout, "$\to 1^-$" denotes taking a left limit at 1), is technically more complex than that of L^p-quantiles. The asymptotic results that arise show that the tail L^p-median at level α is asymptotically proportional to the quantile $q(\alpha)$, as $\alpha \to 1^-$, through a non-explicit but very accurately approximable constant, and the remainder term in the asymptotic relationship is exclusively controlled by extreme value parameters of X. This stands in contrast with L^p-quantiles, which also are asymptotically proportional to the quantile $q(\alpha)$ through a simpler constant, although the remainder term crucially features the expectation and left-tail behaviour of X. The remainder term plays an important role in the estimation, as it determines the bias term in our eventual estimators, and we will then argue that the extreme value behaviour of tail L^p-medians is easier to understand and more natural than that of L^p-quantiles. We will also explain why, for heavy-tailed models, extreme tail L^p-medians are able to interpolate monotonically between extreme MS and extreme CTE, as p varies in $(1,2)$ and for $\gamma < 1$. By contrast, L^p-quantiles are known not to interpolate monotonically between quantiles and expectiles, see Figure 2.7 below. The interpolation property also makes it possible to interpret an extreme tail L^p-median as a weighted average of extreme MS and extreme CTE. This is likely to be helpful as far as the practical applicability of tail L^p-medians is concerned, to the extent that it allows for a simple choice of p reflecting a pre-specified compromise between the extreme MS and CTE.

We shall then examine how to estimate an extreme tail L^p-median. We start, as we shall do in the estimation of extreme L^p-quantiles, by suggesting two estimation methods in the so-called intermediate case of a tail L^p-median level α_n satisfying $n(1-\alpha_n) \to \infty$. Although the design stage of our tail L^p-median estimators has similarities with that of the L^p-quantile estimators, the investigation of the asymptotic properties of the tail L^p-median estimators is more challenging and technically involved. These methods will provide us with basic estimators that we will extrapolate at a proper extreme level α_n, which satisfies $n(1-\alpha_n) \to c \in [0,\infty)$, using the heavy-tailed assumption. We will, as a theoretical byproduct, demonstrate how our results make it possible to recover or improve upon known results in the literature on extreme Conditional Tail Expectation estimation. Our final step will be to assess the finite-sample behaviour of the suggested estimators. Our focus will not be to consider heavy-tailed models with a small value of γ: for these models and more generally in models with finite fourth moment, it is unlikely that any improvement will be brought on the CTE, whether in terms of quality of estimation or interpretability. Our view is rather that the use of tail L^p-medians with $p \in (1,2)$ will be beneficial for very heavy-tailed models, in which γ is higher than the finite fourth moment threshold $\gamma = 1/4$, and possibly higher than the finite variance threshold $\gamma = 1/2$. For such values of γ and with an extreme level set to be $\alpha_n = 1 - 1/n$, we shall then evaluate the finite-sample performance of our estimators on simulated data sets, as well as on the real set of fire insurance data already studied in Section 2.2.5.3, featuring an estimated value of γ larger than $1/2$.

This section is organised as follows. Section 2.4.2 focuses on the study of population and sample extreme L^p-quantiles. Section 2.4.2.1 starts by describing how population L^p-quantiles $q_\tau(p)$ are linked to standard quantiles $q_\tau(1)$ as $\tau \to 1$. Section 2.4.2.2 deals with the estimation of intermediate and extreme L^p-quantiles $q_{\tau_n}(p)$ for $p > 1$. Estimators of the extreme level τ_n in (2.4.2) are discussed in Section 2.4.2.3, with implications for recovering composite estimators of high quantiles $q_{\alpha_n}(1)$. We also discuss extrapolated high expectile ($p = 2$) estimation therein. A brief discussion of the finite-sample performance of our estimators is included in Section 2.4.2.4 and a concrete application to the analysis of data from the S&P500 Index is given in Section 2.4.2.5 to illustrate the usefulness of the extreme L^p-quantile methodology. Section 2.4.2.6 concludes this first part by discussing potential avenues for future research. Section 2.4.3 then turns to the study of the tail L^p-median, of which we give a rigorous definition and elementary properties in Section 2.4.3.1. Section 2.4.3.2 focuses on the analysis of asymptotic properties of the population tail L^p-median. Estimators of an extreme tail L^p-median are introduced and studied in Section 2.4.3.3. A short discussion of the finite-sample behaviour of our estimators is included in Section 2.4.3.4 before they are applied to a set of real fire insurance data in Section 2.4.3.5. Section 2.4.3.6 again discusses perspectives for future work on this topic.

2.4.2 Extreme M-quantiles as risk measures: From L^1 to L^p optimisation [10]

2.4.2.1 Extreme population L^p-quantiles

This section describes in detail what is the behaviour of large population L^p-quantiles and how they are linked to large standard quantiles. We assume for the sake of simplicity throughout Section 2.4.2 that **the distribution function of the random variable X of interest is continuous**. Recalling the notation $X_- = \max(-X, 0)$ for the negative part of X, we first have the following asymptotic connection between $\overline{F}(q_\tau(p))$ and $\overline{F}(q_\tau(1)) \equiv 1-\tau$.

Proposition 2.4.1. *Assume that the survival function $\overline{F}$ satisfies condition $\mathcal{C}_1(\gamma)$. For any $p > 1$, whenever $\mathbb{E}(X_-^{p-1}) < \infty$ and $\gamma < 1/(p-1)$, we have*

$$\lim_{\tau \uparrow 1} \frac{\overline{F}(q_\tau(p))}{1-\tau} = \frac{\gamma}{B(p, \gamma^{-1} - p + 1)}$$

where $B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1} dt$ stands for the Beta function.

Note that when the function $\overline{F}$ satisfies condition $\mathcal{C}_1(\gamma)$ and $\gamma < 1/(p-1)$, we have $\mathbb{E}(X_+^{p-1}) < \infty$ with $X_+ = \max(X, 0)$. This entails together with condition $\mathbb{E}(X_-^{p-1}) < \infty$ that $\mathbb{E}|X|^{p-1} < \infty$, and hence the L^p-quantiles of X are indeed well-defined. Even more strongly, we get the following direct asymptotic connection between $q_\tau(p)$ and $q_\tau(1)$ themselves.

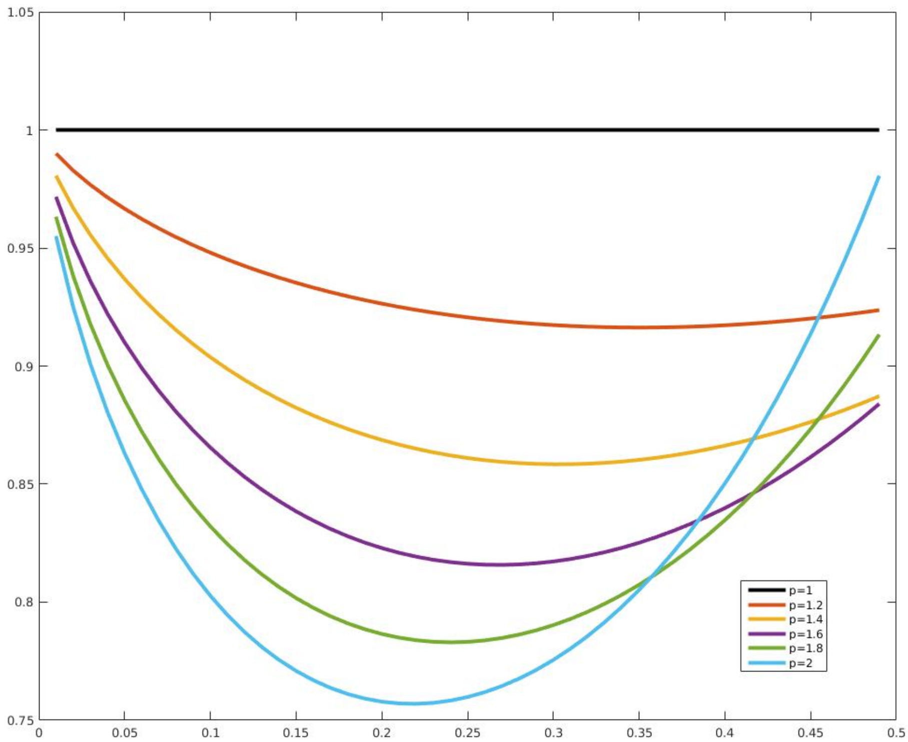

Figure 2.7: Behaviour of $\gamma \in (0, 1/2] \mapsto C(\gamma; p)$ for some values of $p \in [1, 2]$.

Corollary 2.4.2. *Under the conditions of Proposition 2.4.1, we have*

$$\lim_{\tau \uparrow 1} \frac{q_\tau(p)}{q_\tau(1)} = \left[\frac{\gamma}{B(p, \gamma^{-1} - p + 1)}\right]^{-\gamma}.$$

Accordingly, extreme L^p-quantiles are asymptotically proportional to the usual extreme quantiles, for all $p > 1$. The evolution of the proportionality constant

$$C(\gamma; p) := \left[\frac{\gamma}{B(p, \gamma^{-1} - p + 1)}\right]^{-\gamma}$$

with respect to $\gamma \in (0, 1/2]$ is visualised in Figure 2.7, for some values of $p \in [1, 2]$. It can be seen that the usual quantile $q_\tau(1)$ is more spread (conservative) than the L^p-quantile $q_\tau(p)$ as the level $\tau \to 1$.

Next, we shall derive some asymptotic expansions of L^p-quantiles, which shall be very useful when it comes to establish the asymptotic normality of extreme L^p-quantile estimators in the next section. From now on, we denote by $\overline{F}_-$ the survival function of $-X$. Also, a survival function S will be said to be *light-tailed* (and by convention, we shall say it has tail index 0) if it satisfies $x^a S(x) \to 0$ as $x \to +\infty$, for all $a > 0$. The following second-order based refinement of Proposition 2.4.1 is the key element in order to obtain the desired asymptotic expansion of L^p-quantiles.

Proposition 2.4.3. *Assume that $p > 1$ and:*

- *$\overline{F}$ satisfies condition $\mathcal{C}_2(\gamma_r, \rho, A)$;*
- *$\overline{F}_-$ is either light-tailed or satisfies condition $\mathcal{C}_1(\gamma_\ell)$;*
- *$\gamma_r < 1/(p-1)$, and $\gamma_\ell < 1/(p-1)$ in case $\overline{F}_-$ is heavy-tailed.*

Then

$$\frac{\overline{F}(q_\tau(p))}{1-\tau} = \frac{\gamma_r}{B(p,\gamma_r^{-1}-p+1)}(1+R(\tau,p))$$

where

$$\begin{aligned} R(\tau,p) &= -\frac{\gamma_r}{B(p,\gamma_r^{-1}-p+1)}\left((1-\tau)(1+\mathrm{o}(1)) + K(p,\gamma_r,\rho)A\left(\frac{1}{1-\tau}\right)(1+\mathrm{o}(1))\right) \\ &\quad -(p-1)\left(\left[\frac{\gamma_r}{B(p,\gamma_r^{-1}-p+1)}\right]^{\min(\gamma_r,1)} R_r(q_\tau(1),p,\gamma_r)\right. \\ &\quad \left. -\left[\frac{\gamma_r}{B(p,\gamma_r^{-1}-p+1)}\right]^{\gamma_r/\max(\gamma_\ell,1)} R_\ell(q_\tau(1),p,\gamma_\ell)\right) \end{aligned}$$

as $\tau \uparrow 1$, with

$$K(p,\gamma_r,\rho) = \begin{cases} \frac{1}{\gamma_r^2\rho}\left[\frac{\gamma_r}{B(p,\gamma_r^{-1}-p+1)}\right]^{-\rho} & \\ \times\left[(1-\rho)B(p,(1-\rho)\gamma_r^{-1}-p+1) - B(p,\gamma_r^{-1}-p+1)\right] & \text{if } \rho<0, \\ \frac{p-1}{\gamma_r^2}\int_1^{+\infty}(x-1)^{p-2}x^{-1/\gamma_r}\log(x)dx & \text{if } \rho=0, \end{cases}$$

$$R_r(q,p,\gamma_r) = \begin{cases} \frac{\mathbb{E}(X\mathbb{I}\{0<X<q\})}{q}(1+\mathrm{o}(1)) & \text{if } \gamma_r\le 1, \\ \overline{F}(q)B(p-1,1-\gamma_r^{-1})(1+\mathrm{o}(1)) & \text{if } \gamma_r>1, \end{cases}$$

$$\text{and } R_\ell(q,p,\gamma_\ell) = \begin{cases} -\frac{\mathbb{E}(X\mathbb{I}\{-q<X<0\})}{q}(1+\mathrm{o}(1)) & \text{if } \gamma_\ell\le 1 \text{ or } \overline{F}_- \text{ is light-tailed}, \\ F(-q)B(\gamma_\ell^{-1}-p+1,1-\gamma_\ell^{-1})(1+\mathrm{o}(1)) & \text{if } \gamma_\ell>1. \end{cases}$$

When $X \in L^1$, and in particular when expectiles of X can be computed, the asymptotic expansion of L^p-quantiles reduces to the following.

Corollary 2.4.4. *Under the conditions of Proposition 2.4.3, if $\mathbb{E}|X|<\infty$, then*

$$\frac{\overline{F}(q_\tau(p))}{1-\tau} = \frac{\gamma_r}{B(p,\gamma_r^{-1}-p+1)}(1+r(\tau,p))$$

as $\tau \uparrow 1$, where

$$\begin{aligned} r(\tau,p) &= -(p-1)\left[\frac{\gamma_r}{B(p,\gamma_r^{-1}-p+1)}\right]^{\gamma_r}\frac{1}{q_\tau(1)}(\mathbb{E}(X)+\mathrm{o}(1)) \\ &\quad -\frac{\gamma_r}{B(p,\gamma_r^{-1}-p+1)}K(p,\gamma_r,\rho)A\left(\frac{1}{1-\tau}\right)(1+\mathrm{o}(1)). \end{aligned}$$

Finally, we get the following refined asymptotic expansion of $q_\tau(p)$ itself with respect to the ordinary quantile $q_\tau(1)$.

Proposition 2.4.5. *Under the conditions of Proposition 2.4.3, if in addition F is strictly increasing:*

$$\frac{q_\tau(p)}{q_\tau(1)} = C(\gamma_r;p)\left(1-\gamma_r R(\tau,p) + \left\{\frac{1}{\rho}\left[\left[\frac{\gamma_r}{B(p,\gamma_r^{-1}-p+1)}\right]^{-\rho}-1\right]+\mathrm{o}(1)\right\}A\left(\frac{1}{1-\tau}\right)\right)$$

as $\tau \uparrow 1$.

2.4.2.2 Estimation of high L^p-quantiles

Suppose, as will be the case in the remainder of Section 2.4.2, that we observe a random sample $(X_1, \ldots, X_n)$ of independent copies of X and denote by $X_{1,n} \leq \cdots \leq X_{n,n}$ the ascending order statistics of this sample. The overall objective in this section is to estimate extreme L^p-quantiles $q_{\tau_n}(p)$ of X, where $\tau_n \to 1$ as $n \to \infty$. Here τ_n may approach one at any rate, covering the special cases of intermediate L^p-quantiles with $n(1-\tau_n) \to \infty$ and extreme L^p-quantiles with $n(1-\tau_n) \to c$, where c is a finite nonnegative constant.

For the sake of simplicity, and to make it easier to compare our results to those of the other sections of this report, we focus throughout this section on the independent and identically distributed case. The asymptotic theory in [10] is derived in a more general framework of strictly stationary and $\phi-$mixing observations. We shall say more about this, and in particular about the limitations of such a framework, in our discussion of perspectives for future work in Section 2.4.2.6.

Intermediate levels We define the empirical least asymmetrically weighted L^p estimator of $q_{\tau_n}(p)$ as

$$\widehat{q}_{\tau_n}(p) = \arg\min_{u\in\mathbb{R}} \frac{1}{n}\sum_{i=1}^{n} \eta_{\tau_n}(X_i - u; p) = \arg\min_{u\in\mathbb{R}} \frac{1}{n}\sum_{i=1}^{n} |\tau_n - \mathbb{I}\{X_i \leq u\}||X_i - u|^p.$$

Clearly

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{q}_{\tau_n}(p)}{q_{\tau_n}(p)} - 1\right) = \arg\min_{u\in\mathbb{R}} \psi_n(u;p)$$

where

$$\psi_n(u;p) := \frac{1}{p[q_{\tau_n}(p)]^p}\sum_{i=1}^{n} \eta_{\tau_n}\left(X_i - q_{\tau_n}(p) - u q_{\tau_n}(p)/\sqrt{n(1-\tau_n)}; p\right) - \eta_{\tau_n}\left(X_i - q_{\tau_n}(p); p\right).$$

Since this empirical criterion is a convex function of u, the asymptotic properties of the minimiser follow directly from those of the criterion itself by Theorem 5 in [180].

Theorem 2.4.6. *Assume that $p > 1$ and:*

- *There is $\delta > 0$ such that $\mathbb{E}(X_-^{(2+\delta)(p-1)}) < \infty$;*
- *$\overline{F}$ satisfies condition $\mathcal{C}_2(\gamma, \rho, A)$, with $\gamma < 1/[2(p-1)]$;*
- *$\tau_n \uparrow 1$ is such that $n(1-\tau_n) \to \infty$;*
- *We have $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) = \mathrm{O}(1)$.*

Then

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{q}_{\tau_n}(p)}{q_{\tau_n}(p)} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, \gamma^2 V(\gamma;p)\right) \quad \text{as} \quad n \to \infty,$$

with

$$V(\gamma;p) = \frac{B(p-1, \gamma^{-1} - 2p + 2)}{B(p-1, p)} = \frac{\Gamma(2p-1)\Gamma(\gamma^{-1} - 2p + 2)}{\Gamma(p)\Gamma(\gamma^{-1} - p + 1)}.$$

Note that the condition $\gamma < 1/[2(p-1)]$ implies $\gamma < 1/(p-1)$ and hence $\mathbb{E}(X_+^{p-1}) < \infty$. Moreover, the condition $\mathbb{E}(X_-^{(2+\delta)(p-1)}) < \infty$ implies $\mathbb{E}(X_-^{p-1}) < \infty$. Hence $\mathbb{E}|X|^{p-1} < \infty$, and thus the L^p-quantiles exist and are finite. Note also that conditions $\gamma < 1/[2(p-1)]$ and $\mathbb{E}(X_-^{(2+\delta)(p-1)}) < \infty$ ensure the convergence of the (convex) empirical criterion $\psi_n(u;p)$, which entails the convergence of its minimiser.

In the special cases $p \downarrow 1$ and $p = 2$, we recover the asymptotic normality of intermediate sample quantiles and expectiles, respectively, with asymptotic variances

$$V(\gamma;1) = \frac{\Gamma(1)\Gamma(\gamma^{-1})}{\Gamma(1)\Gamma(\gamma^{-1})} = 1 \quad \text{and} \quad V(\gamma;2) = \frac{\Gamma(3)\Gamma(\gamma^{-1} - 2)}{\Gamma(2)\Gamma(\gamma^{-1} - 1)} = \frac{2\gamma}{1-2\gamma}.$$

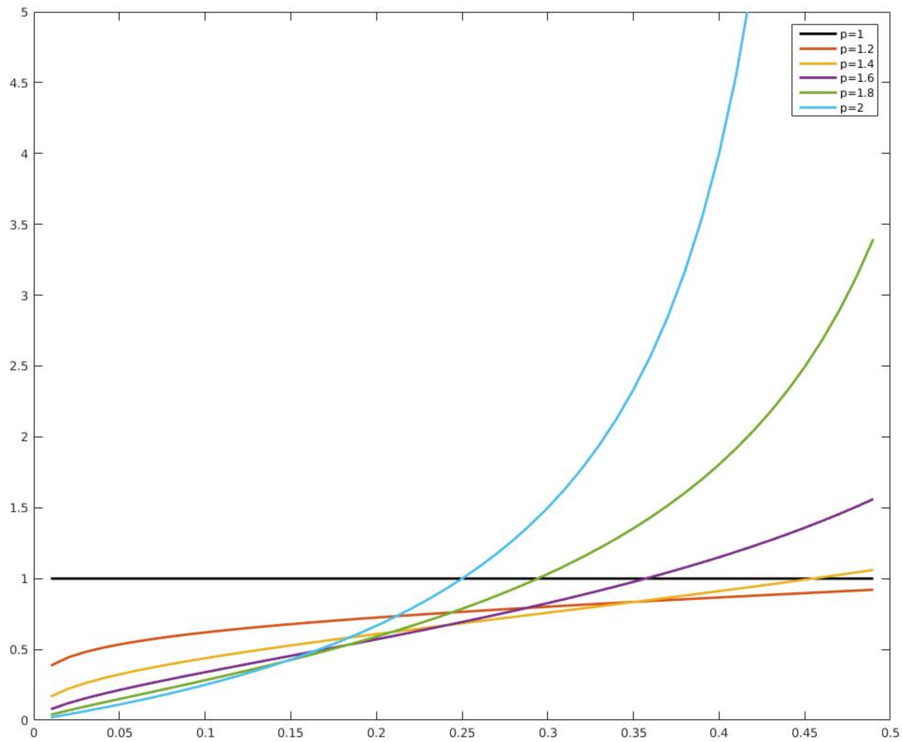

Figure 2.8: Asymptotic variance $\gamma \in (0, 1/2] \mapsto V(\gamma; p)$ for some values of $p \in [1, 2]$. Black line: $p = 1$; Red line: $p = 1.2$; Yellow line: $p = 1.4$; Purple line: $p = 1.6$; Green line: $p = 1.8$; Blue line: $p = 2$.

The behaviour of the variance $\gamma \mapsto V(\gamma; p)$ is visualised in Figure 2.8 for some values of $p \in [1, 2]$, with $\gamma \in (0, 1/2]$. It can be seen in this figure that for values of p close to but larger than 1, the asymptotic variance of the intermediate sample L^p-quantile is appreciably smaller than the asymptotic variance of the traditional sample quantile. In particular, values of p between 1.2 and 1.4 seem to yield estimators who may be more precise than the sample quantile in all applications where the variance of the loss variable is finite (for which $\gamma \in (0, 1/2]$).

Extreme levels We now discuss how to extrapolate intermediate L^p-quantile estimates of order $\tau_n \uparrow 1$, such that $n(1-\tau_n) \to \infty$, to properly extreme levels $\tau_n' \uparrow 1$ with $n(1-\tau_n') \to c < \infty$ as $n \to \infty$. The key argument is to use the asymptotic equivalence

$$q_\tau(p) \sim C(\gamma; p) q_\tau(1) \quad \text{as} \quad \tau \uparrow 1, \tag{2.4.3}$$

shown in Corollary 2.4.2, to get the purely L^p-quantile approximation

$$\frac{q_{\tau_n'}(p)}{q_{\tau_n}(p)} \approx \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\gamma}.$$

This motivates us to define the Weissman-type estimator

$$\widehat{q}^{W}_{\tau_n'}(p) := \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}_n} \widehat{q}_{\tau_n}(p) \tag{2.4.4}$$

for some $\sqrt{n(1-\tau_n)}-$consistent estimator $\widehat{\gamma}_n$ of $\gamma \equiv \gamma_r$, with $\widehat{q}_{\tau_n}(p)$ being the empirical least asymmetrically weighted L^p-estimator of $q_{\tau_n}(p)$.

Theorem 2.4.7. *Assume that $p > 1$ and:*

- $\overline{F}$ *is strictly decreasing and satisfies* $\mathcal{C}_2(\gamma_r, \rho, A)$ *with* $\gamma_r < 1/[2(p-1)]$ *and* $\rho < 0$*;*
- $\overline{F}_-$ *is either light-tailed or satisfies condition* $\mathcal{C}_1(\gamma_\ell)$*;*
- $\gamma_\ell < 1/[2(p-1)]$ *in case* $\overline{F}_-$ *is heavy-tailed.*

Assume further that

- τ_n *and* $\tau_n' \uparrow 1$*, with* $n(1-\tau_n) \to \infty$ *and* $n(1-\tau_n') \to c < \infty$*;*
- $\sqrt{n(1-\tau_n)}(\widehat{\gamma}_n - \gamma_r) \xrightarrow{d} \zeta$*, for a suitable estimator* $\widehat{\gamma}_n$ *of* γ_r *and* ζ *a nondegenerate limiting random variable;*
- $\sqrt{n(1-\tau_n)}\max\left\{1-\tau_n,\, A((1-\tau_n)^{-1}),\, R_r(q_{\tau_n}(1), p, \gamma_r),\, R_\ell(q_{\tau_n}(1), p, \gamma_\ell)\right\} = \mathrm{O}(1)$ *(in this bias condition the notation of Proposition 2.4.3 is used);*
- $\sqrt{n(1-\tau_n)}/\log[(1-\tau_n)/(1-\tau_n')] \to \infty$.

Then

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\widehat{q}^{W}_{\tau_n'}(p)}{q_{\tau_n'}(p)} - 1\right) \xrightarrow{d} \zeta \quad as \quad n \to \infty.$$

Another option for estimating $q_{\tau_n'}(p)$ is by using directly its asymptotic connection (2.4.3) with $q_{\tau_n'}(1)$ to define the plug-in estimator

$$\widetilde{q}^{W}_{\tau_n'}(p) := C(\widehat{\gamma}_n; p)\, \widehat{q}^{W}_{\tau_n'}(1), \tag{2.4.5}$$

obtained by substituting in a $\sqrt{n(1-\tau_n)}$−consistent estimator $\widehat{\gamma}_n$ of γ and the traditional Weissman estimator

$$\widehat{q}^{W}_{\tau_n'}(1) = \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}_n} \widehat{q}_{\tau_n}(1), \tag{2.4.6}$$

of the extreme quantile $q_{\tau_n'}(1)$, where $\widehat{q}_{\tau_n}(1) = X_{n-\lfloor n(1-\tau_n)\rfloor, n}$.

Theorem 2.4.8. *Assume that* $p > 1$ *and:*

- $\overline{F}$ *is strictly decreasing and satisfies* $\mathcal{C}_2(\gamma_r, \rho, A)$ *with* $\gamma_r < 1/(p-1)$ *and* $\rho < 0$*;*
- $\overline{F}_-$ *is either light-tailed or satisfies condition* $\mathcal{C}_1(\gamma_\ell)$*;*
- $\gamma_\ell < 1/(p-1)$ *in case* $\overline{F}_-$ *is heavy-tailed.*

Assume further that

- τ_n *and* $\tau_n' \uparrow 1$*, with* $n(1-\tau_n) \to \infty$ *and* $n(1-\tau_n') \to c < \infty$*;*
- $\sqrt{n(1-\tau_n)}\left(X_{n-\lfloor n(1-\tau_n)\rfloor, n}/q_{\tau_n}(1) - 1\right) = \mathrm{O}_{\mathbb{P}}(1)$*;*
- $\sqrt{n(1-\tau_n)}\,(\widehat{\gamma}_n - \gamma_r) \xrightarrow{d} \zeta$*, for a suitable estimator* $\widehat{\gamma}_n$ *of* γ_r *and* ζ *a nondegenerate limiting random variable;*
- $\sqrt{n(1-\tau_n)}\max\left\{1-\tau_n,\, A((1-\tau_n)^{-1}),\, R_r(q_{\tau_n}(1), p, \gamma_r),\, R_\ell(q_{\tau_n}(1), p, \gamma_\ell)\right\} = \mathrm{O}(1)$ *(in this condition the notation of Proposition 2.4.3 is used);*
- $\sqrt{n(1-\tau_n)}/\log[(1-\tau_n)/(1-\tau_n')] \to \infty$.

Then

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\widetilde{q}^{W}_{\tau_n'}(p)}{q_{\tau_n'}(p)} - 1\right) \xrightarrow{d} \zeta \quad as \quad n \to \infty.$$

A brief comparison of the performance of the plug-in Weissman estimator $\widetilde{q}^{W}_{\tau_n'}(p)$ and the least asymmetrically weighted L^p estimator $\widehat{q}^{W}_{\tau_n'}(p)$ is provided in Section 2.4.2.4. We now turn to the use of L^p−quantiles for the estimation of extreme quantiles and expectiles.

2.4.2.3 Recovering extreme quantiles and expectiles from L^p-quantiles

Extreme L^p-quantile methodology for extreme quantile estimation The L^p-quantiles do not have, for $p > 1$, an intuitive interpretation as direct as the original L^1-quantiles. If the statistician wishes to estimate tail L^p-quantiles $q_{\tau'_n}(p)$ that have the same probabilistic interpretation as a quantile $q_{\alpha_n}(1)$, with a given relative frequency α_n, then the extreme level τ'_n can be specified by the closed form expression (2.4.2), that is,

$$\tau'_n(p, \alpha_n; 1) = \frac{\mathbb{E}\left[|X - q_{\alpha_n}(1)|^{p-1}\mathbb{I}\{X \leq q_{\alpha_n}(1)\}\right]}{\mathbb{E}\left[|X - q_{\alpha_n}(1)|^{p-1}\right]},$$

or equivalently

$$1 - \tau'_n(p, \alpha_n; 1) = \frac{\mathbb{E}\left[|X - q_{\alpha_n}(1)|^{p-1}\mathbb{I}\{X > q_{\alpha_n}(1)\}\right]}{\mathbb{E}\left[|X - q_{\alpha_n}(1)|^{p-1}\right]}. \tag{2.4.7}$$

The statistical problem is now to estimate the unknown extreme level $\tau'_n(p, \alpha_n; 1)$ from the available historical data. To this end, we first note that under condition $\mathcal{C}_1(\gamma_r)$ and if $\gamma_r < 1/(p-1)$, Proposition 2.4.1 entails

$$\frac{\overline{F}(q_{\tau'_n}(p))}{1 - \tau'_n} \sim \frac{\gamma_r}{B(p, \gamma_r^{-1} - p + 1)} \quad \text{as} \quad n \to \infty.$$

It then follows from $q_{\tau'_n}(p) \equiv q_{\alpha_n}(1)$ and $\overline{F}(q_{\alpha_n}(1)) = 1 - \alpha_n$ that

$$\frac{1 - \alpha_n}{1 - \tau'_n} \sim \frac{\gamma_r}{B(p, \gamma_r^{-1} - p + 1)} \quad \text{as} \quad n \to \infty.$$

Therefore $\tau'_n \equiv \tau'_n(p, \alpha_n; 1)$ satisfies the following asymptotic equivalence:

$$1 - \tau'_n(p, \alpha_n; 1) \sim (1 - \alpha_n)\frac{1}{\gamma_r}B\left(p, \frac{1}{\gamma_r} - p + 1\right) \quad \text{as} \quad n \to \infty.$$

Interestingly, and like for the selection of expectile levels in Section 2.3.6, $1 - \tau'_n(p, \alpha_n; 1)$ in (2.4.7) then asymptotically depends on the tail index γ_r but not on the actual value $q_{\alpha_n}(1)$ of the quantile itself. A natural estimator of $1 - \tau'_n(p, \alpha_n; 1)$ can now be defined by replacing, in its asymptotic approximation, the tail index γ_r by a $\sqrt{n(1-\tau_n)}-$consistent estimator $\widehat{\gamma}_n$ as above, to get

$$\widehat{\tau}'_n(p, \alpha_n; 1) = 1 - (1 - \alpha_n)\frac{1}{\widehat{\gamma}_n}B\left(p, \frac{1}{\widehat{\gamma}_n} - p + 1\right). \tag{2.4.8}$$

Next, we derive the limiting distribution of $\widehat{\tau}'_n(p, \alpha_n; 1)$.

Theorem 2.4.9. *Assume that $p > 1$ and:*

- *$\overline{F}$ satisfies condition $\mathcal{C}_2(\gamma_r, \rho, A)$;*
- *$\overline{F}_-$ is either light-tailed or satisfies condition $\mathcal{C}_1(\gamma_\ell)$;*
- *$\gamma_r < 1/(p-1)$, and $\gamma_\ell < 1/(p-1)$ in case $\overline{F}_-$ is heavy-tailed.*

Assume further that

- *τ_n and $\alpha_n \uparrow 1$, with $n(1-\tau_n) \to \infty$;*
- *$\sqrt{n(1-\tau_n)}\,(\widehat{\gamma}_n - \gamma_r) \xrightarrow{d} \zeta$, for a suitable estimator $\widehat{\gamma}_n$ of γ_r and ζ a nondegenerate limiting random variable;*
- *$\sqrt{n(1-\tau_n)}\max\left\{1 - \alpha_n,\, A((1-\alpha_n)^{-1}),\, R_r(q_{\alpha_n}(1), p, \gamma_r),\, R_\ell(q_{\alpha_n}(1), p, \gamma_\ell)\right\} = \mathrm{O}(1)$ (here the notation of Proposition 2.4.3 is used).*

Then:

$$\sqrt{n(1-\tau_n)}\left(\frac{1-\widehat{\tau}_n'(p,\alpha_n;1)}{1-\tau_n'(p,\alpha_n;1)}-1\right)=\mathrm{O}_{\mathbb{P}}(1)$$

as $n\to\infty$. If actually

$$\sqrt{n(1-\tau_n)}\max\left\{1-\alpha_n,\ A((1-\alpha_n)^{-1}),\ R_r(q_{\alpha_n}(1),p,\gamma_r),\ R_\ell(q_{\alpha_n}(1),p,\gamma_\ell)\right\}\to 0$$

then:

$$\sqrt{n(1-\tau_n)}\left(\frac{1-\widehat{\tau}_n'(p,\alpha_n;1)}{1-\tau_n'(p,\alpha_n;1)}-1\right)\xrightarrow{d}-\left\{1+\frac{1}{\gamma_r}\left[F\left(\frac{1}{\gamma_r}-p+1\right)-F\left(\frac{1}{\gamma_r}+1\right)\right]\right\}\frac{\zeta}{\gamma_r}$$

as $n\to\infty$, where $F(x)=\Gamma'(x)/\Gamma(x)$ denotes Euler's digamma function.

In practice, given a tail probability α_n and a power $p\in(1,2]$, the extreme quantile $q_{\alpha_n}(1)$ can be estimated from the generalised L^p-quantile estimators $\widehat{q}^{W}_{\tau_n'}(p)$ and $\widetilde{q}^{W}_{\tau_n'}(p)$ in two steps: first, estimate $\tau_n'\equiv\tau_n'(p,\alpha_n;1)$ by $\widehat{\tau}_n'(p,\alpha_n;1)$ and, second, use the estimators $\widehat{q}^{W}_{\tau_n'}(p)$ and $\widetilde{q}^{W}_{\tau_n'}(p)$ as if τ_n' were known, by substituting the estimated value $\widehat{\tau}_n'(p,\alpha_n;1)$ in place of τ_n', yielding the following two extreme quantile estimators:

$$\begin{aligned}\widehat{q}^{W}_{\widehat{\tau}_n'(p,\alpha_n;1)}(p)&=\left(\frac{1-\widehat{\tau}_n'(p,\alpha_n;1)}{1-\tau_n}\right)^{-\widehat{\gamma}_n}\widehat{q}_{\tau_n}(p)\\ \text{and }\ \widetilde{q}^{W}_{\widehat{\tau}_n'(p,\alpha_n;1)}(p)&=C(\widehat{\gamma}_n;p)\,\widehat{q}^{W}_{\widehat{\tau}_n'(p,\alpha_n;1)}(1).\end{aligned}$$

This is actually a two-stage estimation procedure in the sense that the intermediate level τ_n used in the first stage to compute $\widehat{\tau}_n'(p,\alpha_n;1)$ need not be the same as the intermediate levels used in the second stage to compute the extrapolated L^p-quantile estimators $\widehat{q}^{W}_{\tau_n'}(p)$ and $\widetilde{q}^{W}_{\tau_n'}(p)$. For the sake of simplicity, we do not emphasise in the asymptotic results below the distinction between the intermediate level used in the first stage and those used in the second stage. It should be, however, noted that when the estimation procedure is carried out in one single step instead, i.e. with the same intermediate level in both $\widehat{\tau}_n'(p,\alpha_n;1)$ and the extrapolated L^p-quantile estimators, then the composite version $\widetilde{q}^{W}_{\widehat{\tau}_n'(p,\alpha_n;1)}(p)$ is nothing but the Weissman quantile estimator $\widehat{q}^{W}_{\alpha_n}(1)$. Indeed, in that case, we have by (2.4.6) and the definition of $C(\cdot,\cdot)$ below Corollary 2.4.2 that

$$\begin{aligned}\widetilde{q}^{W}_{\widehat{\tau}_n'(p,\alpha_n;1)}(p)&=C(\widehat{\gamma}_n;p)\,\widehat{q}^{W}_{\widehat{\tau}_n'(p,\alpha_n;1)}(1)\\ &=\left[\frac{\widehat{\gamma}_n}{B(p,\widehat{\gamma}_n^{-1}-p+1)}\cdot\frac{(1-\alpha_n)\frac{1}{\widehat{\gamma}_n}B\left(p,\frac{1}{\widehat{\gamma}_n}-p+1\right)}{1-\tau_n}\right]^{-\widehat{\gamma}_n}\widehat{q}_{\tau_n}(1)\\ &=\left[\frac{1-\alpha_n}{1-\tau_n}\right]^{-\widehat{\gamma}_n}\widehat{q}_{\tau_n}(1)\equiv\widehat{q}^{W}_{\alpha_n}(1).\end{aligned}$$

Our next two convergence results examine the asymptotic properties of the two composite estimators $\widehat{q}^{W}_{\widehat{\tau}_n'(p,\alpha_n;1)}(p)$ and $\widetilde{q}^{W}_{\widehat{\tau}_n'(p,\alpha_n;1)}(p)$. We first consider the estimator $\widehat{q}^{W}_{\widehat{\tau}_n'(p,\alpha_n;1)}(p)$.

Theorem 2.4.10. *Assume that $p>1$ and:*

- *$\overline{F}$ is strictly decreasing and satisfies $\mathcal{C}_2(\gamma_r,\rho,A)$ with $\gamma_r<1/[2(p-1)]$ and $\rho<0$;*
- *$\overline{F}_-$ is either light-tailed or satisfies condition $\mathcal{C}_1(\gamma_\ell)$;*
- *$\gamma_\ell<1/[2(p-1)]$ in case $\overline{F}_-$ is heavy-tailed.*

Assume further that

- *τ_n and $\alpha_n\uparrow 1$, with $n(1-\tau_n)\to\infty$ and $n(1-\alpha_n)\to c<\infty$;*
- *$\sqrt{n(1-\tau_n)}\,(\widehat{\gamma}_n-\gamma_r)\xrightarrow{d}\zeta$, for a suitable estimator $\widehat{\gamma}_n$ of γ_r and ζ a nondegenerate limiting random variable;*

- $\sqrt{n(1-\tau_n)}\max\left\{1-\tau_n,\ A((1-\tau_n)^{-1}),\ R_r(q_{\tau_n}(1),p,\gamma_r),\ R_\ell(q_{\tau_n}(1),p,\gamma_\ell)\right\}=\mathrm{O}(1)$ *(in this condition the notation of Proposition 2.4.3 is used);*
- $\sqrt{n(1-\tau_n)}/\log[(1-\tau_n)/(1-\alpha_n)]\to\infty$.

Then

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\alpha_n)]}\left(\frac{\widehat{q}^{W}_{\widehat{\tau}'_n(p,\alpha_n;1)}(p)}{q_{\alpha_n}(1)}-1\right)\xrightarrow{d}\zeta\quad as\quad n\to\infty.$$

As regards the alternative extrapolated estimator $\widetilde{q}^{W}_{\widetilde{\tau}'_n(p,\alpha_n;1)}(p)$, we have the following asymptotic result.

Theorem 2.4.11. *Assume that $p>1$ and:*

- $\overline{F}$ *is strictly decreasing and satisfies $\mathcal{C}_2(\gamma_r,\rho,A)$ with $\gamma_r<1/(p-1)$ and $\rho<0$;*
- $\overline{F}_-$ *is either light-tailed or satisfies condition $\mathcal{C}_1(\gamma_\ell)$;*
- $\gamma_\ell<1/(p-1)$ *in case $\overline{F}_-$ is heavy-tailed.*

Assume further that

- τ_n *and $\alpha_n\uparrow 1$, with $n(1-\tau_n)\to\infty$ and $n(1-\alpha_n)\to c<\infty$;*
- $\sqrt{n(1-\tau_n)}\,(\widehat{\gamma}_n-\gamma_r)\xrightarrow{d}\zeta$, *for a suitable estimator $\widehat{\gamma}_n$ of γ_r and ζ a nondegenerate limiting random variable;*
- $\sqrt{n(1-\tau_n)}\max\left\{1-\tau_n,\ A((1-\tau_n)^{-1}),\ R_r(q_{\tau_n}(1),p,\gamma_r),\ R_\ell(q_{\tau_n}(1),p,\gamma_\ell)\right\}=\mathrm{O}(1)$ *(in this condition the notation of Proposition 2.4.3 is used);*
- $\sqrt{n(1-\tau_n)}/\log[(1-\tau_n)/(1-\alpha_n)]\to\infty$.

Then

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\alpha_n)]}\left(\frac{\widetilde{q}^{W}_{\widetilde{\tau}'_n(p,\alpha_n;1)}(p)}{q_{\alpha_n}(1)}-1\right)\xrightarrow{d}\zeta\quad as\quad n\to\infty.$$

A comparison and validation on financial data, which we provide in Section 2.4.2.5 below, shows that the two-stage estimation procedure may afford more accurate estimates $\widehat{q}^{W}_{\widehat{\tau}'_n(p,\alpha_n;1)}(p)$ and $\widetilde{q}^{W}_{\widetilde{\tau}'_n(p,\alpha_n;1)}(p)$ of $q_{\alpha_n}(1)$ than the traditional Weissman estimator $\widehat{q}^{W}_{\alpha_n}(1)$ defined in (2.4.6).

Recovering extreme expectiles from L^p-quantiles We focus now on the estimation of extreme L^2-quantiles, or equivalently expectiles, and we assume therefore that $\mathbb{E}(X_-)<\infty$ and $\gamma_r<1$ to guarantee their existence. The first basic tool is the following asymptotic connection between the extreme expectile $q_{\alpha_n}(2)$ and its L^p-quantile analogue $q_{\alpha_n}(p)$:

$$\begin{aligned} q_{\alpha_n}(2) &\sim C(\gamma_r;2)\cdot q_{\alpha_n}(1)\ \text{ as }\ \alpha_n\uparrow 1\\ &\sim C(\gamma_r;2)\cdot C^{-1}(\gamma_r;p)\cdot q_{\alpha_n}(p)\ \text{ as }\ \alpha_n\uparrow 1,\end{aligned}$$

when $p>1$ is such that $\gamma_r<1/(p-1)$ [in particular, this is true for any $p\in(1,2]$ since it is assumed here that $\gamma_r<1$]. This asymptotic equivalence follows immediately by applying Corollary 2.4.2 twice. One may then define the alternative estimator

$$\begin{aligned} \check{q}^{p}_{\alpha_n}(2) &:= C(\widehat{\gamma}_n;2)\cdot C^{-1}(\widehat{\gamma}_n;p)\cdot\widehat{q}^{W}_{\alpha_n}(p) \\ &\equiv (\widehat{\gamma}_n^{-1}-1)^{-\widehat{\gamma}_n}\left[\frac{\widehat{\gamma}_n}{B(p,\widehat{\gamma}_n^{-1}-p+1)}\right]^{\widehat{\gamma}_n}\widehat{q}^{W}_{\alpha_n}(p),\end{aligned}\tag{2.4.9}$$

obtained by substituting in a $\sqrt{n(1-\tau_n)}-$consistent estimator $\widehat{\gamma}_n$ of γ_r and the extrapolated version $\widehat{q}^{W}_{\alpha_n}(p)$ of the least asymmetrically weighted L^p-quantile estimator, given in (2.4.4). The idea is therefore to exploit the accuracy of the asymptotic connection between population L^p-quantiles and traditional quantiles in conjunction with the superiority of sample L^p-quantiles in terms of finite-sample performance. Note that by replacing $\widehat{q}^W_{\alpha_n}(p)$ in (2.4.9) with the plug-in estimator $\widetilde{q}^W_{\alpha_n}(p)$ introduced in (2.4.5), we recover the indirect estimator described in Section 2.3.3.1. As a matter of fact, $\check{q}^p_{\alpha_n}(2)$ approaches $\widehat{q}^W_{\alpha_n}(2)$ when p tends to 2, whereas it approaches $\widetilde{q}^W_{\alpha_n}(2)$ when p tends to 1.

Another way of recovering extreme expectiles from L^p-quantiles is by proceeding as in the case of ordinary quantiles. To estimate the extreme expectile $q_{\alpha_n}(2)$, the idea is to use a tail L^p-quantile $q_{\tau'_n}(p)$ which coincides with (and therefore has the same interpretation as) $q_{\alpha_n}(2)$. Given α_n and the power p, the level τ'_n such that $q_{\tau'_n}(p) \equiv q_{\alpha_n}(2)$ has the explicit expression

$$\tau'_n(p, \alpha_n; 2) = \frac{\mathbb{E}\left[|X - q_{\alpha_n}(2)|^{p-1}\mathbb{I}\{X \le q_{\alpha_n}(2)\}\right]}{\mathbb{E}\left[|X - q_{\alpha_n}(2)|^{p-1}\right]} \tag{2.4.10}$$

in view of (2.4.1). This closed form of $\tau'_n \equiv \tau'_n(p, \alpha_n; 2)$ depends heavily on $q_{\alpha_n}(2)$, but for any $p > 1$ such that $\gamma_r < 1/(p-1)$, condition $\mathcal{C}_1(\gamma_r)$ and Proposition 2.4.1 entail that

$$\frac{\overline{F}(q_{\tau'_n}(p))}{1 - \tau'_n} \sim \frac{\gamma_r}{B(p, \gamma_r^{-1} - p + 1)} \quad \text{as} \quad n \to \infty.$$

It follows from $q_{\tau'_n}(p) \equiv q_{\alpha_n}(2)$ that

$$\frac{\overline{F}(q_{\alpha_n}(2))}{1 - \tau'_n} \sim \frac{\gamma_r}{B(p, \gamma_r^{-1} - p + 1)} \quad \text{as} \quad n \to \infty.$$

We also know (see e.g. Section 2.3.2) that

$$\overline{F}(q_{\alpha_n}(2)) \sim (1 - \alpha_n)(\gamma_r^{-1} - 1) \quad \text{as} \quad n \to \infty.$$

Therefore τ'_n in (2.4.10) satisfies the asymptotic equivalence

$$1 - \tau'_n(p, \alpha_n; 2) \sim (1 - \alpha_n)(\gamma_r^{-1} - 1)\frac{1}{\gamma_r}B\left(p, \frac{1}{\gamma_r} - p + 1\right) \quad \text{as} \quad n \to \infty.$$

By substituting a $\sqrt{n(1-\tau_n)}-$consistent estimator $\widehat{\gamma}_n$ in place of the tail index γ_r, we obtain the following estimator of $\tau'_n(p, \alpha_n; 2)$:

$$\widehat{\tau}'_n(p, \alpha_n; 2) := 1 - (1 - \alpha_n)\left(\widehat{\gamma}_n^{-1} - 1\right)\frac{1}{\widehat{\gamma}_n}B\left(p, \frac{1}{\widehat{\gamma}_n} - p + 1\right).$$

Finally, one may estimate the extreme expectile $q_{\alpha_n}(2) \equiv q_{\tau'_n(p,\alpha_n;2)}(p)$ by the following composite L^p-quantile estimators

$$\widehat{q}^W_{\widehat{\tau}'_n(p,\alpha_n;2)}(p) = \left(\frac{1 - \widehat{\tau}'_n(p, \alpha_n; 2)}{1 - \tau_n}\right)^{-\widehat{\gamma}_n} \widehat{q}_{\tau_n}(p) \tag{2.4.11}$$

$$\text{and } \widetilde{q}^W_{\widehat{\tau}'_n(p,\alpha_n;2)}(p) = C(\widehat{\gamma}_n; p)\,\widetilde{q}^W_{\widehat{\tau}'_n(p,\alpha_n;2)}(1), \tag{2.4.12}$$

obtained by replacing τ'_n in $\widehat{q}^W_{\tau'_n}(p)$ and $\widetilde{q}^W_{\tau'_n}(p)$ with $\widehat{\tau}'_n(p, \alpha_n; 2)$. It is remarkable that these two estimators are intimately linked to the L^p-quantile based extreme L^1-quantile estimators, since

$$\widehat{q}^W_{\widehat{\tau}'_n(p,\alpha_n;2)}(p) = C(\widehat{\gamma}_n; p)\widehat{q}^W_{\widehat{\tau}'_n(p,\alpha_n;1)}(p) \quad \text{and} \quad \widetilde{q}^W_{\widehat{\tau}'_n(p,\alpha_n;2)}(p) = C(\widehat{\gamma}_n; p)\widetilde{q}^W_{\widehat{\tau}'_n(p,\alpha_n;1)}(p).$$

Theory similar to Theorems 2.4.10 and 2.4.11 can be written for these composite extreme expectile estimators; we omit the statements of these results for the sake of brevity.

2.4.2.4 Finite-sample performance on simulated data

A simulation study is carried out in [10] to assess the finite-sample behaviour of our estimators. I describe below its main conclusions.

For which p can L^p-quantiles be accurately estimated? For large values of γ (say $\gamma \geq 0.2$), expectile estimation appears to be the most difficult. By contrast, for these large values of γ (although this seems to be true for smaller γ as well), estimation is more accurate for small values of p, say $p \leq 1.45$. The choice of $p \in (1.2, 1.6)$ seems to be a good global compromise.

This conclusion is, however, no longer valid when it comes to estimate extreme L^p-quantiles $q_{\tau_n'}(p)$ with, for instance, $\tau_n' \in [1-1/n, 1)$, as revealed by a comparison of the extrapolated least asymmetrically weighted L^p-estimators $\widehat{q}^W_{\tau_n'}(p)$ in (2.4.4) and the plug-in Weissman estimators $\widetilde{q}^W_{\tau_n'}(p)$ in (2.4.5). It may be seen that both extreme L^p-quantile estimators are more accurate for $p \in [1, 1.3]$ or $p \in [1.7, 2]$. The best accuracy is generally not achieved at $p = 1$ or $p = 2$. We shall discuss in Section 2.4.2.5 below the important question of how to pick out p in practice in order to get the most accurate estimates from a forecasting perspective.

Which extreme L^p-quantile estimator: $\widehat{q}^W_{\tau_n'}(p)$ or $\widetilde{q}^W_{\tau_n'}(p)$? For non-negative random variables, it appears that $\widetilde{q}^W_{\tau_n'}(p)$ behaves overall better than $\widehat{q}^W_{\tau_n'}(p)$. In the case of the Student distribution, representing a profit-loss distribution, it may be seen that $\widetilde{q}^W_{\tau_n'}(p)$ remains still competitive, but $\widehat{q}^W_{\tau_n'}(p)$ becomes more reliable for large values of p, say, $p \geq 1.9$. Interestingly, for p close to 1 and for positive distributions, the two estimators seem to perform comparably. The important gap in performance which sometimes occurs as p increases is most certainly due to the sensitivity of the least asymmetrically weighted estimator to the top extreme values in the sample. The estimator $\widetilde{q}^W_{\tau_n'}(p)$ does of course benefit from more robustness since it is computed using a single sample quantile.

Extreme expectile estimation using L^p-quantiles Our experience with simulated data suggests to favour the use of $\check{q}^p_{\alpha_n}(2)$ with p very close to 1 for non-negative loss distributions, and with p very close to 2 for real-valued profit-loss random variables. It is in the latter case that $\check{q}^p_{\alpha_n}(2)$ may appear to be appreciably more efficient relatively to both estimators $\widehat{q}^W_{\alpha_n}(2)$ and $\widetilde{q}^W_{\alpha_n}(2)$, especially for profit-loss distributions. We also find that $\widehat{q}^W_{\widehat{\tau}_n'(p,\alpha_n;2)}(p)$ in (2.4.11) behaves very similarly to $\check{q}^p_{\alpha_n}(2)$ in (2.4.9). In particular, $\widehat{q}^W_{\widehat{\tau}_n'(p,\alpha_n;2)}(p)$ exhibits better accuracy relative to both rival estimators $\widehat{q}^W_{\alpha_n}(2)$ and $\widetilde{q}^W_{\alpha_n}(2)$ in the case of profit-loss distributions with heavy tails. By contrast, the second composite estimator $\widetilde{q}^W_{\widehat{\tau}_n'(p,\alpha_n;2)}(p)$ in (2.4.12) does not bring Monte Carlo evidence of any added value with respect to the benchmark estimators $\widehat{q}^W_{\alpha_n}(2)$ and $\widetilde{q}^W_{\alpha_n}(2)$.

Our real data application in Section 2.4.2.5 below highlights the potential of the L^p-quantile methodology for the forecast and validation of extreme quantile estimators.

2.4.2.5 Data example 5: Historical data of the S&P500 index

Here, we focus on the case when the interest is in an estimate of the loss return amount (negative log-return) that will be fallen below (on average) only once in N cases, with N being typically larger than or equal to the sample size n. More specifically, we wish to use $\widetilde{q}^W_{\tau_n'}(p)$ and $\widehat{q}^W_{\tau_n'}(p)$ as estimators of the $(1/n)$th L^1-quantile $q_{1/n}(1) \equiv q_{\tau_n'}(p)$, for which verification and comparison is possible thanks to its elicitability property [134, 265]. Following the ideas of [134, 265], we consider in this section the evaluation and comparison of the competing estimators $\widetilde{q}^W_{\tau_n'}(p)$ and $\widehat{q}^W_{\tau_n'}(p)$ with the standard left tail Weissman quantile estimator $\widehat{q}^W_{1/n}(1)$ from a forecasting perspective, trying to give the best possible point estimate for tomorrow with our knowledge of today. The portfolio under consideration is represented by the S&P500 Index from 4th January 1994 to 30th September 2016, which corresponds to 5,727 trading days. The corresponding log-returns are reported in Figure 2.9.

We now explain how we use the elicitability property. Let the random variable X model the future observation of interest. If the (τ_n')th L^p-quantile $q_{\tau_n'}(p)$ coincides with the $(1/n)$th L^1-quantile $q_{1/n}(1)$,

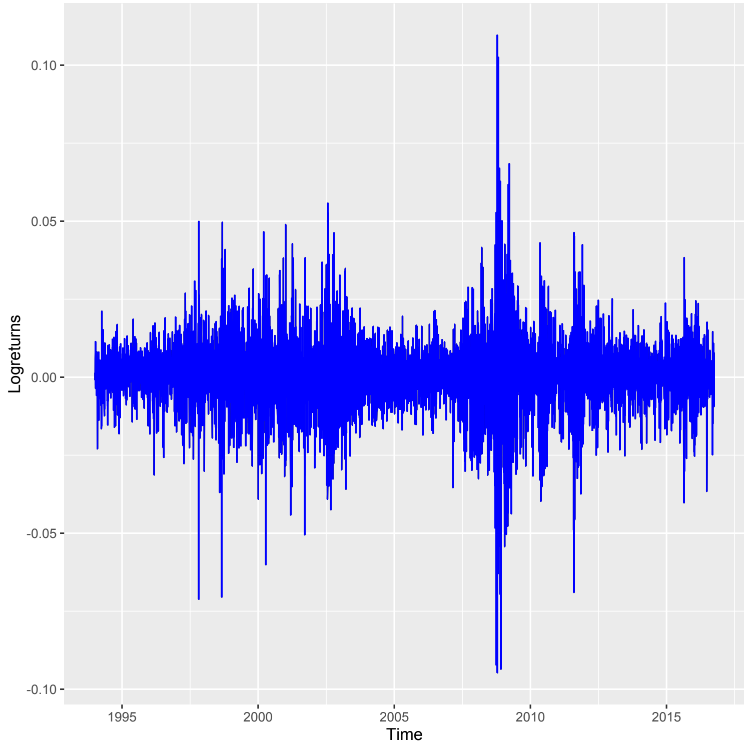

Figure 2.9: Log-returns of the S&P500 Index, from 4th January 1994 to 30th September 2016.

then the definition of a quantile as a minimiser of an absolute error loss exactly means that it equals the optimal point forecast for X given by the Bayes rule

$$q_{\tau_n'}(p) \equiv q_{1/n}(1) = \arg\min_{q\in\mathbb{R}} \mathbb{E}\left[L_n(q, X)\right],$$

under the asymmetric piecewise linear scoring function

$$L_n : \mathbb{R}^2 \to [0, \infty), \quad (q, x) \mapsto \eta_{\frac{1}{n}}(x - q; 1),$$

where $L_n(q, x)$ represents the loss or penalty when the point forecast q is issued and the realisation x of X materialises. Following [134, 265], the point estimates $\widehat{q}^{W}_{\tau_n'}(p)$, $\widetilde{q}^{W}_{\tau_n'}(p)$ and $\widehat{q}^{W}_{1/n}(1)$ of $q_{\tau_n'}(p)$ can then be compared and assessed by means of the scoring function L_n. Suppose that, in T forecast cases, we have point forecasts $\left(q_1^{(m)}, \ldots, q_T^{(m)}\right)$ and realising observations $(x_1, \ldots, x_T)$, where the index m numbers the competing forecasters

$$q_t^{(1)} := \widehat{q}^{W}_{1/n}(1), \quad q_t^{(2)} := \widehat{q}^{W}_{\tau_n'}(p) \quad \text{and} \quad q_t^{(3)} := \widetilde{q}^{W}_{\tau_n'}(p)$$

that are computed at each forecast case $t = 1, \ldots, T$. These purely historical estimates can then be ranked in terms of their average scores (the lower the better):

$$\bar{L}_n^{(m)} = \frac{1}{T}\sum_{t=1}^{T} L_n\left(q_t^{(m)}, x_t\right), \quad m = 1, 2, 3. \tag{2.4.13}$$

In our motivating application concerned with the logarithmic returns of the S&P500 Index, the three estimates were computed on rolling windows of length $n = 2{,}510$, which corresponds to $T = 3{,}217$ forecast cases. Based on the US market, there are on average 251 trading days in a year, and hence each rolling window of size $n = 10 \times 251$ trading days corresponds to a period of 10 years. Therefore, the tail quantity of interest $q_{\tau_n'}(p) \equiv q_{1/n}(1)$ represents the daily loss return (negative log-return) for a once-per-decade market crisis. With the sign convention for values of $Y = -X$ as the negative of returns, the quantile of interest, $q_{1/n}(1)$, can be written as $-Q_{1-1/n}(1)$, where $Q_\tau(p)$ stands for the τth L^p-quantile of Y. The extreme level τ_n' such that $Q_{\tau_n'}(p) = Q_{1-1/n}(1)$ has the closed form expression $\tau_n'(p, \alpha_n; 1)$ described in (2.4.7), with $\alpha_n = 1-1/n$, and can be estimated by $\widehat{\tau}_n'(p, 1-1/n; 1)$ in (2.4.8). Alternatively, without recourse to this sign convention as the negative of returns, it is not hard to check that the level τ_n' such that $q_{\tau_n'}(p) = q_{1/n}(1)$ can directly be estimated by

$$\widehat{\tau}_n'(p, 1/n) := \frac{1}{n}\frac{1}{\widehat{\gamma}_n} B\left(p, \frac{1}{\widehat{\gamma}_n} - p + 1\right)$$

where $\widehat{\gamma}_n$ is an estimator of the left tail index of X. This suggests the following strategy at each forecast case $t = 1, \ldots, T$:

(a) Calculate the first competing estimate $q_t^{(1)} = \widehat{q}^W_{1/n}(1)$;

(b) For a given value of $p \in (1, 2]$, calculate the τ_n' estimate $\widehat{\tau}_n'(t) = \widehat{\tau}_n'(p, 1/n)$;

(c) Calculate the other competing estimates $q_t^{(2)} = \widehat{q}^W_{\tau_n'}(p)$ and $q_t^{(3)} = \widetilde{q}^W_{\tau_n'}(p)$ by substituting the estimated value $\widehat{\tau}_n'(t)$ in place of τ_n'.

As a matter of fact, we use in step (c) the two-stage estimators $q_t^{(2)} = \widehat{q}^W_{\widehat{\tau}_n'(t)}(p)$ and $q_t^{(3)} = \widetilde{q}^W_{\widehat{\tau}_n'(t)}(p)$: first, we estimate τ_n' by $\widehat{\tau}_n'(t)$ in step (b) and, second, we use the estimators $\widehat{q}^W_{\tau_n'}(p)$ and $\widetilde{q}^W_{\tau_n'}(p)$, as if τ_n' were known, by substituting $\widehat{\tau}_n'(t)$ in place of τ_n'. [Of course, the computation of the different point estimates in steps (a), (b) and (c) requires the determination of the values of the sample fraction k involved in the intermediate levels τ_n of these estimates. The selection procedure is explained in detail in Section 6.3 of [10].]

In order to decide on the global accuracy of the three competing methods, we rank the values of their realised losses $\bar{L}_n^{(m)}$ by making use of the T forecasts and realising observations, as described in (2.4.13). The plots of the realised losses versus k are graphed in Figure 2.10 (a) for $\widehat{q}^W_{1/n}(1)$ and $\widehat{q}^W_{\tau_n'}(p)$, with various values of $p \in \{1.1, 1.2, \ldots, 1.9, 2\}$, and in Figure 2.10 (b) for the $\widehat{q}^W_{1/n}(1)$ benchmark and $\widetilde{q}^W_{\tau_n'}(p)$ with the same values of p.

The optimal values of the realised losses for the three methods (the lower the better), displayed in Table 2.8, indicate that the popular Weissman quantile estimator $\widehat{q}^W_{1/n}(1)$ does not ensure the most accurate forecasts of the classical risk measure $q_{1/n}(1)$.

$$\bar{L}_n^{(1)} = 5.758 \cdot 10^{-5}$$

	$p = 1.1$	$p = 1.2$	$p = 1.3$	$p = 1.4$	$p = 1.5$
$\bar{L}_n^{(2)}$	$5.856 \cdot 10^{-5}$	$5.842 \cdot 10^{-5}$	$5.632 \cdot 10^{-5}$	$5.861 \cdot 10^{-5}$	$5.727 \cdot 10^{-5}$
$\bar{L}_n^{(3)}$	$5.755 \cdot 10^{-5}$	$5.752 \cdot 10^{-5}$	$5.748 \cdot 10^{-5}$	$5.745 \cdot 10^{-5}$	$5.742 \cdot 10^{-5}$

	$p = 1.6$	$p = 1.7$	$p = 1.8$	$p = 1.9$	$p = 2$
$\bar{L}_n^{(2)}$	$5.712 \cdot 10^{-5}$	$5.847 \cdot 10^{-5}$	$5.918 \cdot 10^{-5}$	$6.024 \cdot 10^{-5}$	$6.118 \cdot 10^{-5}$
$\bar{L}_n^{(3)}$	$5.739 \cdot 10^{-5}$	$5.735 \cdot 10^{-5}$	$5.919 \cdot 10^{-5}$	$6.030 \cdot 10^{-5}$	$6.025 \cdot 10^{-5}$

Table 2.8: Optimal values $\bar{L}_n^{(1)}$, $\bar{L}_n^{(2)}$ and $\bar{L}_n^{(3)}$ of the realised loss for the three forecasters $\widehat{q}^W_{1/n}(1)$, $\widehat{q}^W_{\tau_n'}(p)$ and $\widetilde{q}^W_{\tau_n'}(p)$, respectively. Results based on daily loss returns.

The top forecaster is $\widehat{q}^W_{\tau_n'}(p)$ for $p = 1.3, 1.6, 1.5$, then $\widetilde{q}^W_{\tau_n'}(p)$ for $p = 1.7, 1.6, 1.5, 1.4, 1.3, 1.2, 1.1$, and finally the basic Weissman estimator $\widehat{q}^W_{1/n}(1)$. Their optimal values obtained in the first and last forecast cases are shown in Table 2.9. In the forecast case $t = 1$, based on the loss returns observed during the first decade, no forecast of the VaR $q_{1/n}(1)$ succeeds in falling below the worst recorded loss return $X_{1,n}$. Yet, all of the generalised L^p-quantiles $\widehat{q}^W_{\tau_n'}(p)$ and $\widetilde{q}^W_{\tau_n'}(p)$ appear to be smaller and

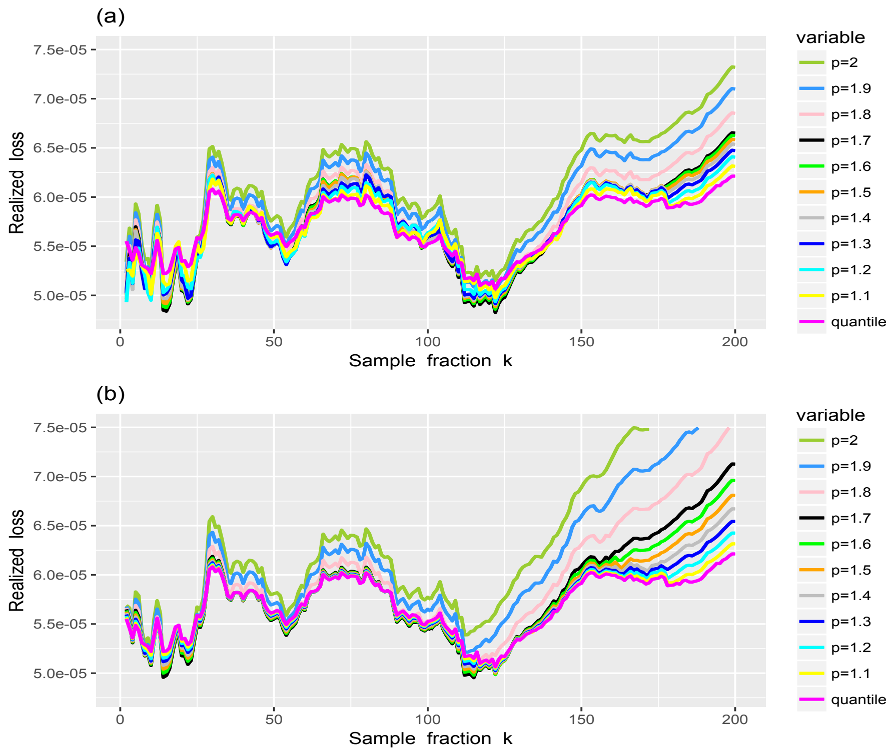

Figure 2.10: (a) Plots of the realised loss $k \mapsto \bar{L}_n^{(m)}(k)$ for $\widehat{q}^W_{1/n}(1)$ in magenta and $\widehat{q}^W_{\tau'_n}(p)$ with different values of p. (b) Results with $\widetilde{q}^W_{\tau'_n}(p)$ in place of $\widehat{q}^W_{\tau'_n}(p)$. Results based on daily loss returns.

hence more conservative than the usual L^1-quantile $\widehat{q}^W_{1/n}(1)$. Here, the tail index estimate is found to be $\widehat{\gamma} = 0.256$. In the forecast case $t = T$, based on the last decade of observations, all forecasts of the VaR were capable of extrapolating outside the minimal loss return $X_{1,n}$. This is due to the turbulent episodes that have been experienced by financial markets during 2007-2008, see Figure 2.9. In particular, the tail index estimate becomes $\widehat{\gamma} = 0.359$. Yet, the top forecasters $\widehat{q}^W_{\tau'_n}(p)$ and $\widetilde{q}^W_{\tau'_n}(p)$ appear to be larger and hence less pessimistic than the extrapolated L^1-quantile estimate $\widehat{q}^W_{1/n}(1)$.

Forecaster	$\widehat{q}^W_{\tau'_n}(1.3)$	$\widehat{q}^W_{\tau'_n}(1.6)$	$\widehat{q}^W_{\tau'_n}(1.5)$	$\widetilde{q}^W_{\tau'_n}(1.7)$	$\widetilde{q}^W_{\tau'_n}(1.6)$	$\widetilde{q}^W_{\tau'_n}(1.5)$
$t = 1$	-0.067901	-0.067169	-0.067288	-0.066846	-0.066841	-0.066837
$t = T$	-0.105717	-0.103084	-0.104613	-0.101254	-0.102211	-0.103131

Forecaster	$\widetilde{q}^W_{\tau'_n}(1.4)$	$\widetilde{q}^W_{\tau'_n}(1.3)$	$\widetilde{q}^W_{\tau'_n}(1.2)$	$\widetilde{q}^W_{\tau'_n}(1.1)$	$\widehat{q}^W_{1/n}(1)$	$X_{1,n}$
$t = 1$	-0.066832	-0.066828	-0.066824	-0.066820	-0.066816	-0.071127
$t = T$	-0.104016	-0.104870	-0.105695	-0.106493	-0.107266	-0.094695

Table 2.9: Optimal values of the top forecasters, obtained in the first and last forecast cases, along with the sample minimum $X_{1,n}$. Results based on daily loss returns.

Finally, we would like to comment on the evolution of the extreme L^p-quantile level $\widehat{\tau}'_n(t)$ with t. The optimal estimates $t \mapsto \widehat{\tau}'_n(t)$, obtained for the different values of p, are graphed in Figure 2.11. It can be seen that $\widehat{\tau}'_n(t)$ decreases, uniformly in t, as p increases. Also, it may be seen that the curve corresponding to the best choice $p = 1.3$ (dark blue) exhibits two different trends before and after the severe losses of 2007-2008. Both trends appear to produce much more extreme levels than the quantile level $1/n \approx 4 \cdot 10^{-4}$.

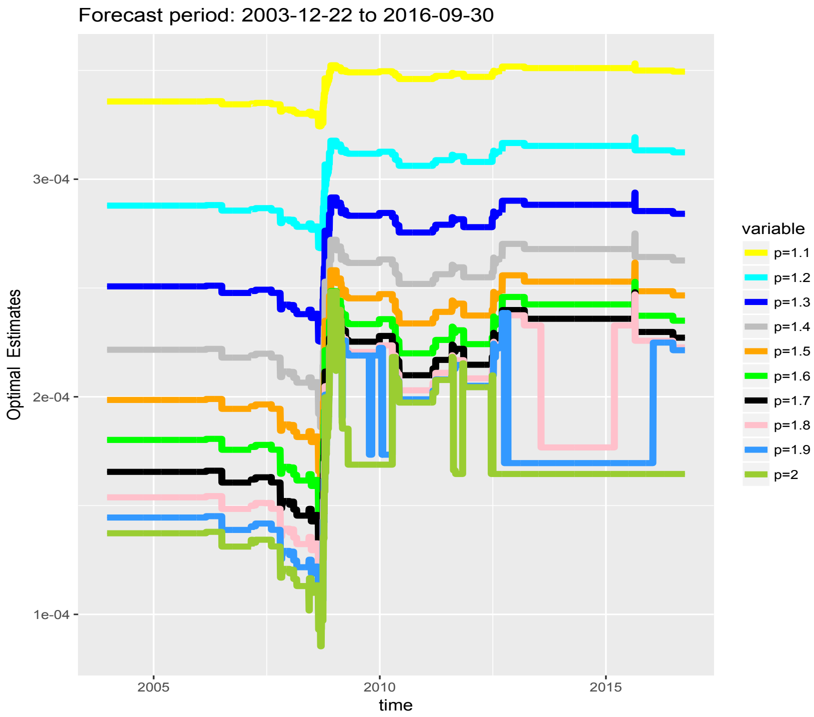

Figure 2.11: The final estimates $t \mapsto \widehat{\tau}'_n(t)$, obtained for $p \in \{1.1, 1.2, \ldots, 1.9, 2\}$.

2.4.2.6 Perspectives for future research

Adaptation to a general weak dependence framework The theory in [10] is actually derived in a context of $\phi-$mixing observations, see [62]. This is theoretically convenient because it allows the use of limit theorems from [245] to prove the convergence of intermediate empirical L^p-quantiles. The $\phi-$mixing framework allows to handle $m-$dependent sequences as well as certain processes featuring Markovian dependence. This condition is, however, quite restrictive: for instance, it is only satisfied for a very small class of autoregressive processes (see the Introduction in [218]). The work of [102, 103, 104] provides tools making it possible to examine the asymptotic properties of a wide class of statistical indicators of extremes of strictly stationary and dependent observations in a much more general $\beta-$mixing context. As [104] shows, this assumption covers, among others, processes obtained by solving certain stochastic recurrence equations (including ARCH processes) and ARMA models under reasonably general conditions. Writing an inference theory for this $\beta-$mixing context would greatly enhance the applicability of the L^p-quantile methodology. A possible strategy, at the intermediate step, is to note that the criterion $\psi_n(u; p)$ can be rewritten as the combination of a non-random term and tail array sums in the sense of [220], and then apply central limit theory developed therein. The extrapolation step can be handled using results on tail index estimators in the $\beta-$mixing framework [104].

In-p-functional convergence We analyse the convergence of intermediate and extreme estimators of $q_{\tau_n}(p)$. This is done pointwise in $p \in [1, 2]$, as our ultimate interest is to use extreme L^p-quantiles with a fixed p to derive a simple methodology for the estimation of extreme quantiles and expectiles. A further interesting question is to prove the weak convergence of the suitably normalised process $p \mapsto \widehat{q}_{\tau_n}(p)$ (for p in a compact subinterval of $(1, 2)$, or potentially $p \in [1, 2]$ itself). This would provide an appealing theoretical ground for plotting $p \mapsto \widehat{q}_{\tau_n}(p)$ and the estimators we derive from extremes L^p-quantiles. Such a functional result would also open the door to the analysis of further estimators of extreme quantiles and expectiles, constructed by e.g. averaging our composite estimators over a suitable range of values of p. Potentially useful tools for this problem are Theorem 1 in [176] and its Pollard-type corollary (see [213]). These results are dedicated to the study of stochastic processes obtained by minimising convex functions, of which L^p-quantiles are part.

2.4.3 Beyond tail median and conditional tail expectation: extreme risk estimation using tail L^p−optimisation [3]

2.4.3.1 Definition and first properties

Let X be a real-valued random variable. **It is assumed throughout this section that the distribution function F of X is continuous**, so that $F(q(\alpha)) = \alpha$ for any $\alpha \in (0,1)$. Our construction is motivated by the following two observations. Firstly, the median and mean of X can respectively be obtained by minimising an expected absolute deviation and expected squared deviation:

$$\begin{aligned} q(1/2) &= \arg\min_{m\in\mathbb{R}} \mathbb{E}\left(|X-m| - |X|\right) \\ \text{and } \mathbb{E}(X) &= \arg\min_{m\in\mathbb{R}} \mathbb{E}\left(|X-m|^2 - |X|^2\right) \text{ (provided } \mathbb{E}|X| < \infty). \end{aligned}$$

The minimiser on the right-hand side of the first identity may actually not be unique, although it is if F is strictly increasing. Our convention throughout this section will be that in such a situation, the minimiser is taken as the smallest possible minimiser, making the identity valid in any case. Note also that subtracting $|X|$ (resp. $|X|^2$) within the expectation in the cost function for $q(1/2)$ (resp. $\mathbb{E}(X)$) makes it possible to define this cost function, and therefore its minimiser, without assuming any integrability condition on X (resp. by assuming only $\mathbb{E}|X| < \infty$), as a consequence of the triangle inequality (resp. the identity $|X-m|^2 - |X|^2 = m(m-2X)$). These optimisation problems extend their arguably better-known formulations

$$q(1/2) = \arg\min_{m\in\mathbb{R}} \mathbb{E}|X-m| \text{ and } \mathbb{E}(X) = \arg\min_{m\in\mathbb{R}} \mathbb{E}(X-m)^2$$

which are only well-defined when $\mathbb{E}|X| < \infty$ and $\mathbb{E}(X^2) < \infty$ respectively. Secondly, since F is continuous, the Median Shortfall $\mathrm{MS}(\alpha) = q([1+\alpha]/2)$ is the median of X given $X > q(\alpha)$, see Example 3 in [185]. Since $\mathrm{CTE}(\alpha)$ is the expectation of X given $X > q(\alpha)$, we find that

$$\begin{aligned} \mathrm{MS}(\alpha) &= \arg\min_{m\in\mathbb{R}} \mathbb{E}\left(|X-m| - |X| \,|\, X > q(\alpha)\right) \\ \text{and } \mathrm{CTE}(\alpha) &= \arg\min_{m\in\mathbb{R}} \mathbb{E}\left(|X-m|^2 - |X|^2 \,|\, X > q(\alpha)\right) \text{ (provided } \mathbb{E}(|X| \,|\, X > q(\alpha)) < \infty). \end{aligned}$$

Our construction now encompasses these two quantities by replacing the absolute or squared deviations by power deviations.

Definition 2.4.1. The *tail L^p−median* of X, of order $\alpha \in (0,1)$, is (when it exists)

$$m_p(\alpha) = \arg\min_{m\in\mathbb{R}} \mathbb{E}\left(|X-m|^p - |X|^p \,|\, X > q(\alpha)\right).$$

Let us highlight the following important connection between the tail L^p−median and the notion of L^p−quantiles: recall from [10, 73] that an L^p−quantile of order $\tau \in (0,1)$ of a univariate random variable Y, with $\mathbb{E}|Y|^{p-1} < \infty$, is defined as

$$q_\tau(p) = \arg\min_{q\in\mathbb{R}} \mathbb{E}\left(|\tau - \mathbb{I}\{Y \le q\}||Y-q|^p - |\tau - \mathbb{I}\{Y \le 0\}||Y|^p\right).$$

Consequently, the L^p−median of Y, obtained for $\tau = 1/2$, is

$$q_{1/2}(p) = \arg\min_{q\in\mathbb{R}} \mathbb{E}\left(|Y-q|^p - |Y|^p\right).$$

For an arbitrary $p \ge 1$, the tail L^p−median $m_p(\alpha)$ of X is then exactly an L^p−median of X given that $X > q(\alpha)$. This construction of $m_p(\alpha)$ as an L^p−median given the tail event $\{X > q(\alpha)\}$ is what motivated the name "tail L^p−median" for $m_p(\alpha)$.

We underline again that subtracting $|X|^p$ inside the expectation in Definition 2.4.1 above makes the cost function well-defined whenever $\mathbb{E}(|X|^{p-1} \,|\, X > q(\alpha))$ is finite (or equivalently $\mathbb{E}(X_+^{p-1}) < \infty$, where $X_+ = \max(X, 0)$). This is a straightforward consequence of the triangle inequality when $p = 1$; when $p > 1$, this is a consequence of the fact that the function $x \mapsto |x|^p$ is continuously differentiable with derivative $x \mapsto p|x|^{p-1}\operatorname{sign}(x)$, together with the mean value theorem. If X is moreover assumed to satisfy $\mathbb{E}(X_+^p) < \infty$, then the definition of $m_p(\alpha)$ is equivalently and perhaps more intuitively

$$m_p(\alpha) = \underset{m \in \mathbb{R}}{\arg\min}\, \mathbb{E}\left(|X - m|^p \mid X > q(\alpha)\right).$$

The following result shows in particular that if $\mathbb{E}(X_+^{p-1}) < \infty$ for some $p \geq 1$, then the tail L^p-median always exists and is characterised by a simple equation. Especially, for $p \in (1, 2)$, the tail L^p-median $m_p(\alpha)$ exists and is unique under a weaker integrability condition than the assumption of a finite first tail moment which is necessary for the existence of $\mathrm{CTE}(\alpha)$.

Proposition 2.4.12. *Let $p \geq 1$. Pick $\alpha \in (0, 1)$ and assume that $\mathbb{E}(X_+^{p-1}) < \infty$. Then:*

(i) The tail L^p-median $m_p(\alpha)$ exists and is such that $m_p(\alpha) > q(\alpha)$.

(ii) The tail L^p-median is equivariant with respect to increasing affine transformations: if $m_p^X(\alpha)$ is the tail L^p-median of X and $Y = aX + b$ with $a > 0$ and $b \in \mathbb{R}$, then the tail L^p-median $m_p^Y(\alpha)$ of Y is $m_p^Y(\alpha) = a\, m_p^X(\alpha) + b$.

(iii) When $p > 1$, the tail L^p-median $m_p(\alpha)$ is the unique $m \in \mathbb{R}$ solution of the equation

$$\mathbb{E}[(m - X)^{p-1}\mathbb{I}\{q(\alpha) < X < m\}] = \mathbb{E}[(X - m)^{p-1}\mathbb{I}\{X > m\}].$$

(iv) The tail L^p-median defines a monotonic functional with respect to first-order stochastic dominance: if Y is another random variable such that $\mathbb{E}(Y_+^{p-1}) < \infty$ then

$$(\forall t \in \mathbb{R},\ \mathbb{P}(X > t) \leq \mathbb{P}(Y > t)) \Rightarrow m_p^X(\alpha) \leq m_p^Y(\alpha).$$

(v) When $\mathbb{E}|X|^{p-1} < \infty$, the function $\alpha \mapsto m_p(\alpha)$ is nondecreasing on $(0, 1)$.

Let us highlight that, in addition to these properties, we have most importantly

$$\mathrm{MS}(\alpha) = m_1(\alpha) \quad \text{and} \quad \mathrm{CTE}(\alpha) = m_2(\alpha). \tag{2.4.14}$$

In other words, the Median Shortfall is the tail L^1-median and the Conditional Tail Expectation is the tail L^2-median: the class of tail L^p-medians encompasses the notions of MS and CTE. We conclude this section by noting that, like MS, the tail L^p-median is not subadditive (for $1 < p < 2$). Our objective here is not, however, to construct an alternative class of risk measures which have perfect axiomatic properties. This work is, rather, primarily intended to provide an interpretable middleway between the risk measures $\mathrm{MS}(\alpha)$ and $\mathrm{CTE}(\alpha)$, for a level α close to 1. This will be useful when the number of finite moments of X is low (typically, less than 4) because the estimation of $\mathrm{CTE}(\alpha)$, for high α, is then a difficult task in practice.

2.4.3.2 Asymptotic properties of an extreme tail L^p-median

Our first focus is to discuss whether $m_p(\alpha)$, for $p \in (1, 2)$, does indeed provide a middle ground between $\mathrm{MS}(\alpha)$ and $\mathrm{CTE}(\alpha)$, as $\alpha \to 1^-$, when X has a heavy-tailed distribution with tail index γ. Since the existence of a tail L^p-median requires finite conditional tail moments of order $p - 1$, our minimal working condition on the pair (p, γ) will be $\gamma < 1/(p - 1)$ throughout. We start by providing some insight into what asymptotic result on $m_p(\alpha)$ we should aim for under condition $\mathcal{C}_1(\gamma)$. A consequence of this condition is that above a high threshold u, the variable X/u is approximately Pareto distributed with tail index γ, or, in other words:

$$\forall x > 1,\ \mathbb{P}\left(\frac{X}{q(\alpha)} > x \,\middle|\, \frac{X}{q(\alpha)} > 1\right) \to x^{-1/\gamma} \quad \text{as} \quad \alpha \to 1^-.$$

This is exactly the first statement of Theorem 1.2.1 in [145]. The above conditional Pareto approximation then suggests that when α is close enough to 1, the optimisation criterion in Definition 2.4.1 can be approximately written as follows:

$$\frac{m_p(\alpha)}{q(\alpha)} \approx \arg\min_{M\in\mathbb{R}} \mathbb{E}\left[|Z_\gamma - M|^p - |Z_\gamma|^p\right]$$

where Z_γ has a Pareto distribution with tail index γ. For the variable Z_γ, we can differentiate the cost function and use change-of-variables formulae to get that the minimiser M of the right-hand side should satisfy the equation $g_{p,\gamma}(1/M) = B(p, \gamma^{-1} - p + 1)$, where $g_{p,\gamma}(t) := \int_t^1 (1-u)^{p-1} u^{-1/\gamma - 1} du$ for $t \in (0,1)$ and $B(x,y) = \int_0^1 v^{x-1}(1-v)^{y-1} dv$ denotes the Beta function. In other words, and for a Pareto random variable, we have

$$\frac{m_p(\alpha)}{q(\alpha)} = \frac{1}{\kappa(p,\gamma)} \quad \text{where} \quad \kappa(p,\gamma) := g_{p,\gamma}^{-1}(B(p, \gamma^{-1} - p + 1))$$

and $g_{p,\gamma}^{-1}$ denotes the inverse of the decreasing function $g_{p,\gamma}$ on $(0,1)$. Our first asymptotic result states that this proportionality relationship is still valid asymptotically under condition $\mathcal{C}_1(\gamma)$.

Proposition 2.4.13. *Suppose that $p \geq 1$ and $\mathcal{C}_1(\gamma)$ holds with $\gamma < 1/(p-1)$. Then:*

$$\lim_{\alpha \to 1^-} \frac{m_p(\alpha)}{q(\alpha)} = \frac{1}{\kappa(p,\gamma)}.$$

This first-order result is similar in spirit to other asymptotic proportionality relationships linking extreme risk measures to extreme quantiles; in particular, similarly to extreme L^p-quantiles, an extreme tail L^p-median $m_p(\alpha)$ contains both the information contained in the corresponding quantile $q(\alpha)$ plus the information on tail heaviness provided by the tail index γ. However, the asymptotic proportionality constant $\kappa(p,\gamma) \in (0,1)$ does not have a simple closed form in general, due to the complicated expression of the function $g_{p,\gamma}$. It does have a nice explicit expression though in the two particular cases $p = 1$ and $p = 2$:

- For $p = 1$, one can show that $\kappa(1,\gamma) = 2^{-\gamma}$. This clearly yields the same equivalent as the one obtained using the regular variation of the tail quantile function U with index γ:

$$\frac{m_1(\alpha)}{q(\alpha)} = \frac{q((1+\alpha)/2)}{q(\alpha)} = \frac{U(2(1-\alpha)^{-1})}{U((1-\alpha)^{-1})} \to 2^\gamma \text{ as } \alpha \to 1^-.$$

- For $p = 2$ and $\gamma \in (0,1)$, one has $\kappa(2,\gamma) = 1 - \gamma$. Since in this case, $m_2(\alpha)$ is nothing but $\mathrm{CTE}(\alpha)$, Proposition 2.4.13 agrees here with the asymptotic equivalent of $\mathrm{CTE}(\alpha)$ in terms of the exceedance level $q(\alpha)$, see e.g. [163].

For other values of p, an accurate numerical computation of the constant $\kappa(p,\gamma)$ can be carried out instead. Such numerical computations, on the domain $(p,\gamma) \in [1,2] \times (0,1)$, show that $p \mapsto \kappa(p,\gamma)$ and $\gamma \mapsto \kappa(p,\gamma)$ appear to be both decreasing. This entails in particular that, for all $p_1, p_2 \in (1,2)$ such that $p_1 < p_2$, we have, for α close enough to 1:

$$\mathrm{MS}(\alpha) = m_1(\alpha) < m_{p_1}(\alpha) < m_{p_2}(\alpha) < m_2(\alpha) = \mathrm{CTE}(\alpha).$$

The tail L^p-median $m_p(\alpha)$ can therefore be seen as a risk measure interpolating monotonically between $\mathrm{MS}(\alpha)$ and $\mathrm{CTE}(\alpha)$, at a high enough level α. Actually, Proposition 2.4.13 yields, for all $p \in [1,2]$ and $\gamma < 1/(p-1)$ that, as $\alpha \to 1^-$,

$$m_p(\alpha) \approx \lambda(p,\gamma)\mathrm{MS}(\alpha) + [1 - \lambda(p,\gamma)]\mathrm{CTE}(\alpha), \tag{2.4.15}$$

where the weighting constant $\lambda(p,\gamma) \in [0,1]$ is defined by

$$\lambda(p,\gamma) := \lim_{\alpha \to 1^-} \frac{m_p(\alpha) - \mathrm{CTE}(\alpha)}{\mathrm{MS}(\alpha) - \mathrm{CTE}(\alpha)} = \frac{1 - (1-\gamma)/\kappa(p,\gamma)}{1 - 2^\gamma(1-\gamma)}. \tag{2.4.16}$$

Extreme tail L^p-medians of heavy-tailed models can then be interpreted, for $p \in (1,2)$, as weighted averages of extreme Median Shortfall and extreme Conditional Tail Expectation at the same level. It should be noted that, by contrast, the monotonic interpolation property (and hence the weighted average interpretation) is demonstrably false in general for L^p-quantiles, as is most easily seen from Figure 2.7 in Section 2.4.2.1: this figure suggests that high L^p-quantiles define, for γ close to $1/2$, a decreasing function of p when it is close to 1 and an increasing function of p when it is close to 2.

The fact that $\gamma \mapsto \kappa(p,\gamma)$ is decreasing, meanwhile, can be proven rigorously by noting that its partial derivative $\partial\kappa/\partial\gamma$ is negative (see Theorem 2.4.16 below). More intuitively, the monotonicity of $\gamma \mapsto \kappa(p,\gamma)$ can be seen as a consequence of the heavy-tailedness of the distribution function F. Indeed, we saw that heuristically, as $\alpha \to 1^-$,

$$\frac{m_p(\alpha)}{q(\alpha)} \approx \underset{M\in\mathbb{R}}{\arg\min}\, \mathbb{E}\left[|Z_\gamma - M|^p - |Z_\gamma|^p\right]$$

where Z_γ is a Pareto random variable with tail index γ. When γ increases, the random variable Z_γ tends to return higher values because its survival function $\mathbb{P}(Z_\gamma > z) = z^{-1/\gamma}$ (for $z > 1$) is an increasing function of γ. We can therefore expect that, as γ increases, a higher value of $M = m_p(\alpha)/q(\alpha)$ will be needed in order to minimise the above cost function.

Our next goal is to derive an asymptotic expansion of the tail L^p-median $m_p(\alpha)$, relatively to the high exceedance level $q(\alpha)$. This will be the key theoretical tool making it possible to analyse the asymptotic properties of estimators of an extreme tail L^p-median. For this, we need to quantify precisely the error term in the convergence given by Proposition 2.4.13. This is the focus of the next result.

Proposition 2.4.14. *Suppose that $p \geq 1$ and $\mathcal{C}_2(\gamma,\rho,A)$ holds with $\gamma < 1/(p-1)$. Then, as $\alpha \to 1^-$,*

$$\frac{m_p(\alpha)}{q(\alpha)} = \frac{1}{\kappa(p,\gamma)}\left(1 + [R(p,\gamma,\rho) + \mathrm{o}(1)]A\left((1-\alpha)^{-1}\right)\right)$$

with

$$R(p,\gamma,\rho) = \frac{[\kappa(p,\gamma)]^{-(\rho-1)/\gamma}}{(1-\kappa(p,\gamma))^{p-1}} \times \frac{1-\rho}{\gamma\rho}\left[B(p,(1-\rho)\gamma^{-1}-p+1) - g_{p,\gamma/(1-\rho)}(\kappa(p,\gamma))\right].$$

This result is, again, similar in spirit to second-order results that have been shown for other extreme risk measures: see e.g. Proposition 2.4.5 in Section 2.4.2.1 (corresponding to Proposition 3 of [10]) for an analogue result used as a basis to carry out extreme value-based inference on extreme L^p-quantiles. It should, however, be underlined that the asymptotic expansion of an extreme tail L^p-median depends solely on the extreme parameters γ, ρ and A, along with the power p. By contrast, the asymptotic expansion of an extreme L^p-quantile depends on the expectation and left-tail behaviour of X, which are typically considered to be irrelevant to the understanding of the right tail of X. From an extreme value point of view, the asymptotic expansion of extreme tail L^p-medians is therefore easier to understand than that of L^p-quantiles. Statistically speaking, it also implies that there are less sources of potential bias in extreme tail L^p-median estimation than in extreme L^p-quantile estimation. Both of these statements can be explained by the fact that the tail L^p-median is constructed exclusively on the event $\{X > q(\alpha)\}$, while the equation defining an L^p-quantile $q_p(\alpha)$ is (see [73]):

$$(1-\alpha)\mathbb{E}((q_p(\alpha)-X)^{p-1}\mathbb{I}\{X < q_p(\alpha)\}) = \alpha\mathbb{E}((X-q_p(\alpha))^{p-1}\mathbb{I}\{X > q_p(\alpha)\}).$$

The left-hand side term ensures that the central and left-tail behaviour of X will necessarily have an influence on the value of any L^p-quantile, even at an extreme level. The construction of a tail L^p-median as an L^p-median in the right tail of X removes this issue.

Like the asymptotic proportionality constant $\kappa(p,\gamma)$ on which it depends, the remainder term $R(p,\gamma,\rho)$ does not have an explicit form in general. That being said, $R(1,\gamma,\rho)$ and $R(2,\gamma,\rho)$ have simple explicit values:

- For $p = 1$, one can show that $R(1, \gamma, \rho) = (2^\rho - 1)/\rho$ (recall the convention $(x^\rho - 1)/\rho = \log x$ for $\rho = 0$). We then find back the result that comes as an immediate consequence of (2.4.14) and our second-order condition $\mathcal{C}_2(\gamma, \rho, A)$:

$$\frac{m_1(\alpha)}{q(\alpha)} = \frac{U(2(1-\alpha)^{-1})}{U((1-\alpha)^{-1})} = 2^\gamma \left(1 + A((1-\alpha)^{-1})\frac{2^\rho - 1}{\rho}(1 + \mathrm{o}(1))\right).$$

- For $p = 2$ and $\gamma \in (0, 1)$, (2.4.14) and the second-order condition suggest that:

$$\frac{m_2(\alpha)}{q(\alpha)} = \int_1^\infty \frac{U((1-\alpha)^{-1}x)}{U((1-\alpha)^{-1})}\frac{dx}{x^2} = \frac{1}{1-\gamma}\left(1 + \frac{1}{1-\gamma-\rho}A((1-\alpha)^{-1})(1 + \mathrm{o}(1))\right)$$

  which coincides with Proposition 2.4.14, given that $R(2, \gamma, \rho) = 1/(1 - \gamma - \rho)$.

We close this section by noting that all our results, and indeed the practical use of the tail L^p-median more generally, depend on the fixed value of the constant p. Just as when using L^p-quantiles, the choice of p in practice is a difficult but important question. Although [73] introduced L^p-quantiles in the context of testing for symmetry in non-parametric regression, it did not investigate the question of the choice of p. For extreme tail L^p-medians, which unlike extreme L^p-quantiles satisfy an interpolation property, we may suggest a simple and (potentially) intuitive way to choose p. Recall the weighted average relationship (2.4.15):

$$m_p(\alpha) \approx \lambda(p, \gamma)\mathrm{MS}(\alpha) + [1 - \lambda(p, \gamma)]\mathrm{CTE}(\alpha), \text{ with } \lambda(p, \gamma) = \frac{1 - (1-\gamma)/\kappa(p, \gamma)}{1 - 2^\gamma(1-\gamma)}$$

for α close to 1. In practice, given the (estimated) value of γ, and a pre-specified weighting constant λ_0 indicating a compromise between robustness of MS and sensitivity of CTE, one can choose $p = p_0$ as the unique root of the equation $\lambda(p, \gamma) = \lambda_0$ with unknown p. Although $\lambda(p, \gamma)$ does not have a simple closed form, our experience shows that this equation can be solved very quickly and accurately with standard numerical solvers. This results in a tail L^p-median $m_{p_0}(\alpha)$ satisfying, for α close to 1,

$$m_{p_0}(\alpha) \approx \lambda_0\mathrm{MS}(\alpha) + (1 - \lambda_0)\mathrm{CTE}(\alpha).$$

The interpretation of $m_{p_0}(\alpha)$ is easier than that of a generic tail L^p-median $m_p(\alpha)$ due to its explicit and fully-determined connection with the two well-understood quantities $\mathrm{MS}(\alpha)$ and $\mathrm{CTE}(\alpha)$. The question of which weighting constant λ_0 should be chosen is itself difficult and depends on the requirements of the situation at hand; our real data application in Section 2.4.3.5 provides an illustration with $\lambda_0 = 1/2$, corresponding to a simple average of extreme MS and CTE.

2.4.3.3 Estimation of an extreme tail L^p-median

Suppose that we observe a random sample $(X_1, \ldots, X_n)$ of independent copies of X, and denote by $X_{1,n} \leq \cdots \leq X_{n,n}$ the corresponding set of order statistics arranged in increasing order. Our goal in this section is to estimate an extreme tail L^p-median $m_p(\alpha_n)$, where $\alpha_n \to 1^-$ as $n \to \infty$. The final aim is to allow α_n to approach 1 at any rate, covering the cases of an intermediate tail L^p-median with $n(1 - \alpha_n) \to \infty$ and of a proper extreme tail L^p-median with $n(1 - \alpha_n) \to c$, where c is some finite nonnegative constant.

Intermediate case: direct estimation by empirical L^p-optimisation Recall that the tail L^p-median $m_p(\alpha_n)$ is, by Definition 2.4.1,

$$m_p(\alpha_n) = \underset{m \in \mathbb{R}}{\arg\min}\, \mathbb{E}\left(|X - m|^p - |X|^p \mid X > q(\alpha_n)\right).$$

Assume here that $n(1-\alpha_n) \to \infty$, so that $\lfloor n(1-\alpha_n)\rfloor > 0$ eventually. We can therefore define a direct empirical tail L^p-median estimator of $m_p(\alpha_n)$ by minimising the above empirical cost function:

$$\widehat{m}_p(\alpha_n) = \underset{m \in \mathbb{R}}{\arg\min}\, \frac{1}{\lfloor n(1-\alpha_n)\rfloor} \sum_{i=1}^{\lfloor n(1-\alpha_n)\rfloor} \left(|X_{n-i+1,n} - m|^p - |X_{n-i+1,n}|^p\right).$$

We now pave the way for a theoretical study of this estimator. We clearly have the equivalent definition

$$\psi_n(u;p) = \frac{1}{p[m_p(\alpha_n)]^p} \sum_{i=1}^{\lfloor n(1-\alpha_n)\rfloor} \left(\left| X_{n-i+1,n} - m_p(\alpha_n) - \frac{u m_p(\alpha_n)}{\sqrt{n(1-\alpha_n)}} \right|^p - |X_{n-i+1,n} - m_p(\alpha_n)|^p \right).$$

Note that the empirical criterion $\psi_n(u;p)$ is a continuous and convex function of u, so that the asymptotic properties of the minimiser follow directly from those of the criterion itself by the convexity results of [180]. The empirical criterion, then, is analysed by using its continuous differentiability (for $p > 1$) in order to formulate an L^p-analogue of Knight's identity [179] and divide the work between, on the one hand, the study of a $\sqrt{n(1-\alpha_n)}-$consistent and asymptotically Gaussian term which is an affine function of u and, on the other hand, a bias term which converges to a non-random multiple of u^2.

This programme of work is broadly similar to that of the proof of the convergence of the direct intermediate L^p-quantile estimator. The difficulty in this particular case, however, is twofold: first, the affine function of u is a generalised $L-$statistic, in the sense of for instance [60], whose analysis requires delicate arguments relying on the asymptotic behaviour of the tail empirical quantile process. For L^p-quantiles, this is not necessary because the affine term is actually a sum of independent, identically distributed and centred variables. Second, the bias term is essentially a doubly integrated oscillation of a power function with generally noninteger exponent. The examination of its convergence requires certain precise real analysis arguments which do not follow from those developed in [10] for the asymptotic analysis of intermediate L^p-quantiles.

With this in mind, the asymptotic normality result for the direct intermediate tail L^p-median estimator $\widehat{m}_p(\alpha_n)$ is the following.

Theorem 2.4.15. *Suppose that $p \geq 1$ and $\mathcal{C}_2(\gamma,\rho,A)$ holds with $\gamma < 1/[2(p-1)]$. Assume further that $\alpha_n \to 1^-$ is such that $n(1-\alpha_n) \to \infty$ and $\sqrt{n(1-\alpha_n)}A((1-\alpha_n)^{-1}) = \mathrm{O}(1)$. Then we have, as $n \to \infty$:*

$$\sqrt{n(1-\alpha_n)} \left(\frac{\widehat{m}_p(\alpha_n)}{m_p(\alpha_n)} - 1 \right) \xrightarrow{d} \mathcal{N}\left(0, V(p,\gamma)\right).$$

Here $V(p,\gamma) = V_1(p,\gamma)/V_2(p,\gamma)$ with

$$V_1(p,\gamma) = \frac{[\kappa(p,\gamma)]^{1/\gamma}}{\gamma} \left(B(2p-1,\gamma^{-1}-2(p-1)) + g_{2p-1,\gamma}\left(\kappa(p,\gamma)\right) \right) + [1-\kappa(p,\gamma)]^{2(p-1)}$$

while $V_2(p,\gamma)$ is defined by: $V_2(1,\gamma) = 1/\gamma^2$ and, for $p > 1$,

$$V_2(p,\gamma) = \left(\frac{p-1}{\gamma} [\kappa(p,\gamma)]^{1/\gamma} \left[B(p-1,\gamma^{-1}-p+2) + g_{p-1,\gamma}\left(\kappa(p,\gamma)\right) \right] \right)^2.$$

Moreover, the functions $V_2(\cdot,\gamma)$ and $V(\cdot,\gamma)$ defined this way are right-continuous at 1.

The asymptotic variance $V(p,\gamma)$ has a rather involved expression. Numerical studies show that $\gamma \mapsto V(p,\gamma)$ appears to be increasing. This reflects the increasing tendency of the underlying distribution to generate extremely high observations when the tail index increases, see Section 2.2.4.1, thus increasing the variability of the empirical criterion $\psi_n(\cdot;p)$ and consequently that of its minimiser. Besides, it turns out that, somewhat surprisingly, the function $p \mapsto V(p,\gamma)$ is not in general a monotonic function of p. One can actually remark that we have $V(p,\gamma) < V(1,\gamma)$ for any $(p,\gamma) \in (1,1.2] \times [0.25,0.5]$. This suggests that for all heavy-tailed distributions having only a second moment (an already difficult case as far as estimation in heavy-tailed models is concerned), a direct L^p-tail median estimator with $p \in (1,1.2]$ will have a smaller asymptotic variance than the empirical L^1-tail median estimator, or, in other words, the empirical Median Shortfall.

We conclude this section by noting that, like the constants appearing in our previous asymptotic results, the variance term $V(p,\gamma)$ has a simple expression when $p = 1$ or $p = 2$. In the case $p = 1$, we can prove that $V(1,\gamma) = 2\gamma^2$. Statement (2.4.14) suggests that this should be identical to the asymptotic variance of the high quantile estimator $\widehat{\mathrm{MS}}(\alpha_n) = X_{n-\lfloor n(1-\alpha_n)/2 \rfloor, n}$, and indeed

$$\sqrt{n(1-\alpha_n)}\left(\frac{\widehat{\mathrm{MS}}(\alpha_n)}{\mathrm{MS}(\alpha_n)} - 1\right) = \sqrt{2}\sqrt{\frac{n(1-\alpha_n)}{2}}\left(\frac{X_{n-\lfloor n(1-\alpha_n)/2 \rfloor, n}}{q(1-(1-\alpha_n)/2)} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, 2\gamma^2\right)$$

by Theorem 2.4.8 p.52 in [145]. For $p = 2$, we have $V(2,\gamma) = 2\gamma^2(1-\gamma)/(1-2\gamma)$. This constant should coincide with the asymptotic variance of the empirical counterpart of the Conditional Tail Expectation, namely

$$\widehat{\mathrm{CTE}}(\alpha_n) = \frac{1}{\lfloor n(1-\alpha_n) \rfloor} \sum_{i=1}^{\lfloor n(1-\alpha_n) \rfloor} X_{n-i+1,n}.$$

This is indeed the case as Corollary 1 in [114] shows. In particular, the function $\gamma \mapsto V(2,\gamma)$ tends to infinity as $\gamma \uparrow 1/2$, reflecting the increasing difficulty of estimating a high Conditional Tail Expectation by its direct empirical counterpart as the right tail of X gets heavier.

Intermediate case: indirect quantile-based estimation We can also design an estimator of $m_p(\alpha_n)$ based on the asymptotic equivalence between $m_p(\alpha_n)$ and $q(\alpha_n)$ that is provided by Proposition 2.4.13. Indeed, since this result suggests that $m_p(\alpha)/q(\alpha) \sim 1/\kappa(p,\gamma)$ when $\alpha \to 1^-$, it makes sense to build a plug-in estimator of $m_p(\alpha_n)$ by setting $\widetilde{m}_p(\alpha_n) = \widehat{q}(\alpha_n)/\kappa(p,\widehat{\gamma}_n)$, where $\widehat{q}(\alpha_n)$ and $\widehat{\gamma}_n$ are respectively two consistent estimators of the high quantile $q(\alpha_n)$ and of the tail index γ. Since we work here in the intermediate case $n(1-\alpha_n) \to \infty$, we know that the sample counterpart $X_{\lceil n\alpha_n \rceil, n}$ of $q(\alpha_n)$ is a relatively consistent estimator of $q(\alpha_n)$, see Theorem 2.4.1 in [145]. This suggests to use the estimator

$$\widetilde{m}_p(\alpha_n) := \frac{X_{\lceil n\alpha_n \rceil, n}}{\kappa(p,\widehat{\gamma}_n)}.$$

Our next result analyses the asymptotic distribution of this estimator, when the pair $(\widehat{\gamma}_n, X_{\lceil n\alpha_n \rceil, n})$ is jointly $\sqrt{n(1-\alpha_n)}$−consistent.

Theorem 2.4.16. *Suppose that $p \geq 1$ and $\mathcal{C}_2(\gamma,\rho,A)$ holds with $\gamma < 1/(p-1)$. Assume further that $\alpha_n \to 1^-$ is such that $n(1-\alpha_n) \to \infty$ and $\sqrt{n(1-\alpha_n)}A((1-\alpha_n)^{-1}) \to \lambda \in \mathbb{R}$, and that*

$$\sqrt{n(1-\alpha_n)}\left(\widehat{\gamma}_n - \gamma, \frac{X_{\lceil n\alpha_n \rceil, n}}{q(\alpha_n)} - 1\right) \xrightarrow{d} (\xi_1, \xi_2)$$

where (ξ_1,ξ_2) is a pair of nondegenerate random variables. Then we have, as $n \to \infty$:

$$\sqrt{n(1-\alpha_n)}\left(\frac{\widetilde{m}_p(\alpha_n)}{m_p(\alpha_n)} - 1\right) \xrightarrow{d} \sigma(p,\gamma)\xi_1 + \xi_2 - \lambda R(p,\gamma,\rho),$$

with the positive constant $\sigma(p,\gamma)$ being $\sigma(p,\gamma) = -\dfrac{1}{\kappa(p,\gamma)}\dfrac{\partial \kappa}{\partial \gamma}(p,\gamma)$. In other words,

$$\sigma(p,\gamma) = \frac{B(p,\gamma^{-1}-p+1)[\digamma(\gamma^{-1}+1) - \digamma(\gamma^{-1}-p+1)] - \int_{\kappa(p,\gamma)}^{1}(1-u)^{p-1}u^{-1/\gamma-1}\log(u)\,du}{\gamma^2[1-\kappa(p,\gamma)]^{p-1}[\kappa(p,\gamma)]^{-1/\gamma}}$$

where $\digamma(x) = \Gamma'(x)/\Gamma(x)$ is Euler's digamma function.

Again, the constant $\sigma(p,\gamma)$ does not generally have a simple explicit form, but we can compute it when $p = 1$ or $p = 2$. For $p = 1$, $\sigma(1,\gamma) = \log 2$, while $\sigma(2,\gamma) = 1/(1-\gamma)$. Contrary to our previous analyses, it is more difficult to relate these constants to pre-existing results in high quantile or high CTE estimation because Theorem 2.4.16 is a general result that applies to a wide range of estimators $\widehat{\gamma}_n$. To the best of our knowledge, there is no general analogue of this result in the literature for the case $p = 1$. In the case $p = 2$, we find back the asymptotic distribution result in Theorem 2.2.2 in Section 2.2.3.2 (corresponding to Theorem 1 in [14]):

$$\sqrt{n(1-\alpha_n)}\left(\frac{\widetilde{m}_2(\alpha_n)}{\mathrm{CTE}(\alpha_n)} - 1\right) \xrightarrow{d} \frac{1}{1-\gamma}\xi_1 + \xi_2 - \frac{\lambda}{1-\gamma-\rho}.$$

It should be highlighted that, for $p = 2$, Theorem 2.4.16 is a stronger result than Theorem 2.2.2 in Section 2.2.3.2, since the condition on γ is less stringent. As the above convergence is valid for any $\gamma < 1$, one may therefore think that the estimator $\widetilde{m}_2(\alpha_n)$ is a widely applicable estimator of the CTE at high levels. Since it is also more robust than the direct empirical CTE estimator due to its reliance on the sample quantile $X_{\lceil n\alpha_n\rceil,n}$, this would defeat the point of looking for a middle ground solution between the sensitivity of CTE to high values and the robustness of VaR-type measures in very heavy-tailed models. The simulation study in [14] shows however that in general, the estimator $\widetilde{m}_2$ fares worse than the direct CTE estimator $\widehat{m}_2$, and increasingly so as γ increases within the range $(0, 1/4]$. The benefit of using the indirect estimator $\widetilde{m}_p$ will rather be found for values of p away from 2, when a genuine compromise between sensitivity and robustness is sought.

Theorem 2.4.16 applies whenever $\widehat{\gamma}_n$ is a consistent estimator of γ that satisfies a joint convergence condition together with the intermediate order statistic $X_{\lceil n\alpha_n\rceil,n}$. As an example, we get the following asymptotic result on $\widetilde{m}_p(\alpha_n)$ when the estimator $\widehat{\gamma}_n$ is the Hill estimator, which we write in this section as $\widehat{\gamma}_n^H$:

$$\widehat{\gamma}_n^H := \frac{1}{\lfloor n(1-\alpha_n)\rfloor} \sum_{i=1}^{\lfloor n(1-\alpha_n)\rfloor} \log\left(X_{n-i+1,n}\right) - \log\left(X_{n-\lfloor n(1-\alpha_n)\rfloor,n}\right).$$

Corollary 2.4.17. *Suppose that $p \geq 1$ and $\mathcal{C}_2(\gamma, \rho, A)$ holds with $\gamma < 1/(p-1)$. Assume further that $\alpha_n \to 1^-$ is such that $n(1-\alpha_n) \to \infty$ and $\sqrt{n(1-\alpha_n)}A((1-\alpha_n)^{-1}) \to \lambda \in \mathbb{R}$. Then, as $n \to \infty$:*

$$\sqrt{n(1-\alpha_n)}\left(\frac{\widetilde{m}_p(\alpha_n)}{m_p(\alpha_n)} - 1\right) \xrightarrow{d} \mathcal{N}\left(\lambda\left(\frac{\sigma(p,\gamma)}{1-\rho} - R(p,\gamma,\rho)\right), \mathfrak{v}(p,\gamma)\right),$$

where $\mathfrak{v}(p,\gamma) = \gamma^2\left([\sigma(p,\gamma)]^2 + 1\right)$.

A numerical study shows that the indirect estimator has a lower variance than the direct one. The difference between the two variances becomes sizeable when the quantity $2\gamma(p-1)$ gets closer to 1, as should be expected since the asymptotic variance of the direct estimator asymptotically increases to infinity (see Theorem 2.4.15), while the asymptotic variance of the indirect estimator is kept under control (see Corollary 2.4.17). This seems to indicate that the indirect estimator should be preferred to the direct estimator in terms of variability. However, the indirect estimator is asymptotically biased (due to the use of the approximation $m_p(\alpha)/q(\alpha) \sim 1/\kappa(p,\gamma)$ in its construction), while the direct estimator is not. We will see that this can make one prefer the direct estimator in terms of mean squared error on finite samples, even for large values of γ and p.

Extreme case: a Weissman-type extrapolation device Both the direct and indirect estimators constructed so far are only consistent for intermediate sequences α_n such that $n(1-\alpha_n) \to \infty$. Our purpose is now to extrapolate these intermediate tail L^p-median estimators to proper extreme levels $\beta_n \to 1^-$ with $n(1-\beta_n) \to c < \infty$ as $n \to \infty$. The key point is then that, when $\gamma < 1/(p-1)$, the quantity $m_p(\alpha)$ is asymptotically proportional to $q(\alpha)$, by Proposition 2.4.13. Combining this with the standard Weissman approximation on ratios of high quantiles suggests the following extrapolation formula:

$$m_p(\beta_n) \approx \left(\frac{1-\beta_n}{1-\alpha_n}\right)^{-\gamma} m_p(\alpha_n).$$

An estimator of the extreme tail L^p-median $m_p(\beta_n)$ is obtained from this approximation by plugging in a consistent estimator $\widehat{\gamma}_n$ of γ and a consistent estimator of $m_p(\alpha_n)$. In our context, the latter can be the direct, empirical L^p-estimator $\widehat{m}_p(\alpha_n)$, or the indirect, intermediate quantile-based estimator $\widetilde{m}_p(\alpha_n)$, yielding the extrapolated estimators

$$\widehat{m}_p^W(\beta_n) := \left(\frac{1-\beta_n}{1-\alpha_n}\right)^{-\widehat{\gamma}_n} \widehat{m}_p(\alpha_n) \text{ and } \widetilde{m}_p^W(\beta_n) := \left(\frac{1-\beta_n}{1-\alpha_n}\right)^{-\widehat{\gamma}_n} \widetilde{m}_p(\alpha_n).$$

We note moreover that the latter estimator is precisely the estimator deduced by plugging the Weissman extreme quantile estimator $\widehat{q}^W(\beta_n)$ in the relationship $m_p(\beta_n)/q(\beta_n) \sim 1/\kappa(p,\gamma)$, since

$$\widetilde{m}_p^W(\beta_n) = \left(\frac{1-\beta_n}{1-\alpha_n}\right)^{-\widehat{\gamma}_n} \times \left\{\frac{X_{\lceil n\alpha_n\rceil,n}}{\kappa(p,\widehat{\gamma}_n)}\right\} = \frac{\widehat{q}^W(\beta_n)}{\kappa(p,\widehat{\gamma}_n)}.$$

The asymptotic behaviour of the two extrapolated estimators $\widehat{m}_p^W(\beta_n)$ or $\widetilde{m}_p^W(\beta_n)$ is analysed in our next main result.

Theorem 2.4.18. *Suppose that $p \geq 1$ and $\mathcal{C}_2(\gamma,\rho,A)$ holds with $\rho < 0$. Assume also that $\alpha_n, \beta_n \to 1^-$ are such that $n(1-\alpha_n) \to \infty$ and $n(1-\beta_n) \to c < \infty$, with $\sqrt{n(1-\alpha_n)}/\log([1-\alpha_n]/[1-\beta_n]) \to \infty$. Assume finally that $\sqrt{n(1-\alpha_n)}(\widehat{\gamma}_n - \gamma) \xrightarrow{d} \xi$, where ξ is a nondegenerate limiting random variable.*

(i) If $\gamma < 1/[2(p-1)]$ and $\sqrt{n(1-\alpha_n)}A((1-\alpha_n)^{-1}) = \mathrm{O}(1)$ then, as $n \to \infty$:

$$\frac{\sqrt{n(1-\alpha_n)}}{\log[(1-\alpha_n)/(1-\beta_n)]}\left(\frac{\widehat{m}_p^W(\beta_n)}{m_p(\beta_n)} - 1\right) \xrightarrow{d} \xi.$$

(ii) If $\gamma < 1/(p-1)$ and $\sqrt{n(1-\alpha_n)}A((1-\alpha_n)^{-1}) \to \lambda \in \mathbb{R}$ then, as $n \to \infty$:

$$\frac{\sqrt{n(1-\alpha_n)}}{\log[(1-\alpha_n)/(1-\beta_n)]}\left(\frac{\widetilde{m}_p^W(\beta_n)}{m_p(\beta_n)} - 1\right) \xrightarrow{d} \xi.$$

This result shows that both of the estimators $\widehat{m}_p^W(\beta_n)$ and $\widetilde{m}_p^W(\beta_n)$ have their asymptotic properties governed by those of the tail index estimator $\widehat{\gamma}_n$. This is of course not an unusual phenomenon for extrapolated estimators; it was, in our context, already observed below Theorems 2.3.17 and 2.3.23.

Let us highlight though that while the asymptotic behaviour of $\widehat{\gamma}_n$ is crucial, we should anticipate that in finite-sample situations, an accurate estimation of the intermediate tail L^p-median $m_p(\alpha_n)$ is also important. A mathematical reason for this is that in the typical situation when $1-\beta_n = 1/n$, the logarithmic term $\log[(1-\alpha_n)/(1-\beta_n)]$ has order at most $\log(n)$, and thus for a moderately high sample size n, the quantity $\sqrt{n(1-\alpha_n)}/\log[(1-\alpha_n)/(1-\beta_n)]$ representing the rate of convergence of the extrapolation factor may only be slightly lower than the quantity $\sqrt{n(1-\alpha_n)}$ representing the rate of convergence of the estimator at the intermediate step. Hence the idea that, while for n very large the difference in finite-sample behaviour between any two estimators of the tail L^p-median at the basic intermediate level α_n will be eventually wiped out by the performance of the estimator $\widehat{\gamma}_n$, there may still be a significant impact of the quality of the intermediate tail L^p-median estimator used on the overall accuracy of the extrapolated estimator when n is moderately large.

2.4.3.4 Finite-sample performance on simulated data

It was highlighted in (2.4.15) and (2.4.16) that, for $p \in [1,2]$ and $\gamma < 1/(p-1)$, an extreme tail L^p-median $m_p(\alpha)$ can be understood asymptotically as a weighted average of $\mathrm{MS}(\alpha)$ and $\mathrm{CTE}(\alpha)$. In other words, defining the interpolating risk measure

$$R_\lambda(\alpha) := \lambda \mathrm{MS}(\alpha) + (1-\lambda)\mathrm{CTE}(\alpha),$$

we have $m_p(\alpha) \approx R_\lambda(\alpha)$ as $\alpha \to 1^-$ with $\lambda = \lambda(p,\gamma)$ as in (2.4.16). It then turns out that at the population level, we have two distinct (but asymptotically equivalent) possibilities to interpolate, and thus create a compromise, between extreme Median Shortfall and extreme Conditional Tail Expectation:

- Consider the family of measures $m_p(\alpha)$, $p \in [1,2]$;
- Consider the family of measures $R_\lambda(\alpha)$, $\lambda \in [0,1]$.

The simulation study of [3] considers the estimation of the tail L^p-median $m_p(\alpha_n)$ for both intermediate and extreme levels α_n, and how this estimation compares with direct estimation of the interpolating measure $R_\lambda(\alpha_n)$. The latter is done at the intermediate level using the estimator

$$\widehat{R}_\lambda(\alpha_n) := \lambda X_{n-\lfloor n(1-\alpha_n)/2\rfloor,n} + (1-\lambda)\widehat{\mathrm{CTE}}(\alpha_n)$$

$$\text{with } \widehat{\mathrm{CTE}}(\alpha_n) = \widehat{m}_2(\alpha_n) = \frac{1}{\lfloor n(1-\alpha_n)\rfloor}\sum_{i=1}^{\lfloor n(1-\alpha_n)\rfloor} X_{n-i+1,n}.$$

An extrapolated estimator at a level $\beta_n \to 1$ such that $n(1-\beta_n) = \mathrm{O}(1)$ is then

$$\widehat{R}_{\lambda}^{W}(\beta_n, k) := \left(\frac{n(1-\beta_n)}{k}\right)^{-\widehat{\gamma}_n^H(k)} \left[\lambda X_{n-\lfloor k/2 \rfloor, n} + (1-\lambda)\widehat{\mathrm{CTE}}(1-k/n)\right]$$

where $k = k(n)$ is the effective sample size of the Hill estimator, satisfying $k \to \infty$ and $k/n \to 0$. The two paragraphs below give a summary of the conclusions of the simulation study of [3].

Intermediate case We examine the cases $\gamma \in \{1/4, 1/2, 3/4\}$, corresponding to, respectively, a borderline case for finite fourth moment, a borderline case for finite variance, and a case where there is no finite variance. It can be seen that, on the Burr distribution, the indirect estimator $\widetilde{m}_p(\alpha_n)$ has a lower Mean Log-Squared Error (MLSE) than the direct estimator, except for γ close to 1 and p close to 2. By contrast, the direct estimator is generally more accurate than the indirect one when the underlying distribution is a Student distribution, especially for $p \geq 1.6$. Furthermore, it should be noted that the direct estimator $\widehat{m}_p(\alpha_n)$ performs overall noticeably better than the estimator $\widehat{R}_{\lambda(p,\gamma)}(\alpha_n)$. This confirms our theoretical expectation that estimation of $m_p(\alpha_n)$ should be easier than estimation of $R_{\lambda(p,\gamma)}(\alpha_n)$, since $\widehat{m}_p(\alpha_n)$ is relatively $\sqrt{n(1-\tau_n)}-$consistent as soon as $\gamma < 1/[2(p-1)]$, which is a weaker condition than the assumption $\gamma < 1/2$ needed to ensure the relative $\sqrt{n(1-\tau_n)}-$consistency of $\widehat{R}_{\lambda(p,\gamma)}(\alpha_n)$ (due to its reliance on the estimator $\widehat{\mathrm{CTE}}(\alpha_n)$). In other words, on finite samples and at intermediate levels, it is preferable to interpolate between extreme MS and extreme CTE via tail L^p-medians rather than using direct linear interpolation, even though these two ideas are asymptotically equivalent at the population level.

Extreme case The conclusions reached in the intermediate case remain true in the extreme case $\beta_n = 1 - 1/n$: on the Burr distribution, the extrapolated indirect estimator of $m_p(\beta_n)$ performs comparably to or better than the extrapolated direct estimator, except for γ close to 1 and p close to 2. The reverse conclusion holds true on the Student distribution. The extrapolated direct estimator also performs generally better than the extrapolated linear interpolation estimator $\widehat{R}_{\lambda(p,\gamma)}^{W}(\beta_n, \hat{k}_{\mathrm{opt}})$. [Here $\hat{k}_{\mathrm{opt}}$ is a selected value of k obtained following a procedure presented in Section 5.2 of [3].] There is a noticeable improvement for $p \in [1.25, 1.75]$ in the case $\gamma = 1/2$ and even more so for $\gamma = 3/4$, which are the most relevant cases for our purpose. The accuracy of the extrapolated direct estimator is also comparable overall to that of $\widehat{R}_{\lambda(p,\gamma)}^{W}(\beta_n, \hat{k}_{\mathrm{opt}})$ for $\gamma = 1/4$, although this is not the case we originally constructed the tail L^p-median for.

2.4.3.5 Data example 2 revisited: French commercial fire losses

We revisit here the data set of Section 2.2.5.3, made of $n = 1{,}098$ commercial fire losses recorded between 1st January 1995 and 31st December 1996 by the *Fédération Française des Sociétés d'Assurance*, and available from the R package `CASdatasets` as `data(frecomfire)`.

Our first step in the analysis of the extreme losses in this data set is to estimate the tail index γ. Using the Hill estimator, and a selection procedure of the effective sample size which is explained in Section 5.2 of [3], we find an estimate $\widehat{\gamma}^H = 0.67$. This estimated value of γ is in line with the findings of Section 2.2.5.3, and it suggests that there is evidence for an infinite second moment of the underlying distribution. We know, in this context, that the estimation of the extreme CTE is going to be a difficult problem, although it would give a better understanding of risk in this data set than a single quantile such as the VaR or MS would do. It therefore makes sense, on this data set, to use the class of tail L^p-medians to find a middle ground between MS and CTE estimation by interpolation.

Estimates of the tail L^p-median $m_p(1-1/n)$, obtained through our extrapolated direct and extrapolated indirect estimators, are depicted on Figure 2.12. These estimates are fairly close overall on the range $p \in [1, 2]$. There is a difference for p close to 2, where the increased theoretical sensitivity of the direct estimator to the highest values in the sample makes it exceed the more robust indirect estimator; note that here γ is estimated to be 0.67, which is sufficiently far from 1 to ensure that the instability of the indirect estimator for very high γ is not an issue, and the estimate returned by the indirect extrapolated method can be considered to be a reasonable one.

With these estimates at hand comes the question of how to choose p. This is of course a difficult question that depends on the objective of the analysis from the perspective of risk assessment: should the analysis be conservative (i.e. return a higher, more prudent estimation) or not? We do not wish to enter into such a debate here, which would be better held within the financial and actuarial communities. Let us, however, illustrate a simple way to choose p based on our discussion at the very end of Section 2.4.3.2. Recall from (2.4.15) and (2.4.16) that a tail L^p-median has an asymptotic interpretation as a weighted average of MS and CTE:

$$\begin{aligned} m_p(1-1/n) &\approx \lambda(p,\gamma)\mathrm{MS}(1-1/n) + (1-\lambda(p,\gamma))\mathrm{CTE}(1-1/n) \\ \text{with } \lambda(p,\gamma) &= \frac{1-(1-\gamma)/\kappa(p,\gamma)}{1-2^{\gamma}(1-\gamma)}. \end{aligned}$$

We also know from our simulation study that it is generally more accurate to estimate $m_p(1-1/n)$ rather than the corresponding linear combination of $\mathrm{MS}(1-1/n)$ and $\mathrm{CTE}(1-1/n)$. With $\lambda_0 = 1/2$, representing the simple average between $\mathrm{MS}(1-1/n)$ and $\mathrm{CTE}(1-1/n)$, choosing $p = \widehat{p}$ as the unique root of the equation $\lambda(\widehat{p},\widehat{\gamma}) = \lambda_0$ yields $\widehat{p} = 1.711$. The corresponding estimates, in million euros, of the linear combination

$$\lambda_0\mathrm{MS}(1-1/n) + (1-\lambda_0)\mathrm{CTE}(1-1/n)$$

are $\widehat{m}_{\widehat{p}}^W(1-1/n) = 160.8$ and $\widetilde{m}_{\widehat{p}}^W(1-1/n) = 155.4$. It is interesting to note that, although these quantities estimate the average of $\mathrm{MS}(1-1/n) = m_1(1-1/n)$ and $\mathrm{CTE}(1-1/n) = m_2(1-1/n)$, we also have $\widehat{m}_1^W(1-1/n) = 106.3$ and $\widehat{m}_2^W(1-1/n) = 225.2$ so that

$$\widetilde{m}_{\widehat{p}}^W(1-1/n) = 155.4 < \widehat{m}_{\widehat{p}}^W(1-1/n) = 160.8 < \frac{1}{2}\left[\widehat{m}_1^W(1-1/n) + \widehat{m}_2^W(1-1/n)\right] = 165.8.$$

The estimate $\widehat{m}_{\widehat{p}}^W(1-1/n)$ of the simple average between $\mathrm{MS}(1-1/n)$ and $\mathrm{CTE}(1-1/n)$ obtained through extrapolating the direct tail L^p-median estimator is therefore itself a middleway between the indirect estimator $\widetilde{m}_{\widehat{p}}^W(1-1/n)$, which relies on a VaR estimator, and the direct estimator of this average which depends on a highly variable estimator of $\mathrm{CTE}(1-1/n)$.

2.4.3.6 Perspectives for future research

Uncertainty quantification for very heavy tails The straightforward estimator of the interpolating measure $R_\lambda(\alpha_n)$ generally fares worse than our estimators and, for $\gamma > 1/2$, its asymptotic distribution is not clear. For nonnegative distributions, the indirect estimator seems to give more accurate point estimates and, for any $p \in [1,2]$ and $\gamma \in (0,1)$, its asymptotic normality is provided in Section 2.4.3.3. The variance of this distribution depends only on γ, whose estimation is possible through e.g. the Hill estimator, and thus asymptotic confidence intervals for tail L^p-medians and the interpolating measure can be constructed. Our simulation results did not, however, investigate the accuracy of these asymptotic confidence intervals (in terms of coverage probability) suggested by our theoretical findings. An interesting further question in this direction would be to look for the asymptotic distribution of the direct estimator of $R_\lambda(\alpha_n)$, for $\gamma > 1/2$, first at intermediate and then extreme levels. This will presumably involve a reduced rate of convergence (of the order of $[n(1-\alpha_n)]^{1-\gamma}$) and an asymptotic stable distribution. Such a result would rigorously establish that the indirect estimator converges at a faster rate than the direct estimator of the measure $R_\lambda(\alpha_n)$.

Composite estimation of Median Shortfall and Conditional Tail Expectation The asymptotic proportionality relationship in Proposition 2.4.13 can be used to connect the extreme Median Shortfall and Conditional Tail Expectation to the tail L^p-median. In the case of the Conditional Tail Expectation, for instance, one has, for any $p \in [1,2)$ and $\gamma \in (0,1)$, $\mathrm{CTE}(\alpha)/m_p(\alpha) \approx \kappa(p,\gamma)/(1-\gamma)$ for α close to 1. This might suggest to estimate $\mathrm{CTE}(\alpha)$ by estimating first $m_p(\alpha)$ and then multiplying by an estimator of the proportionality constant on the right-hand side. Early numerical investigations suggest that such an estimator is not competitive. However, since $\alpha \mapsto m_p(\alpha)$ is regularly varying at 1, one has

$$\mathrm{CTE}(\alpha) \approx \frac{\kappa(p,\gamma)}{1-\gamma} m_p(\alpha) \approx m_p(\beta) \quad \text{with} \quad \beta = 1 - \left[\frac{1-\gamma}{\kappa(p,\gamma)}\right]^{1/\gamma}(1-\alpha).$$

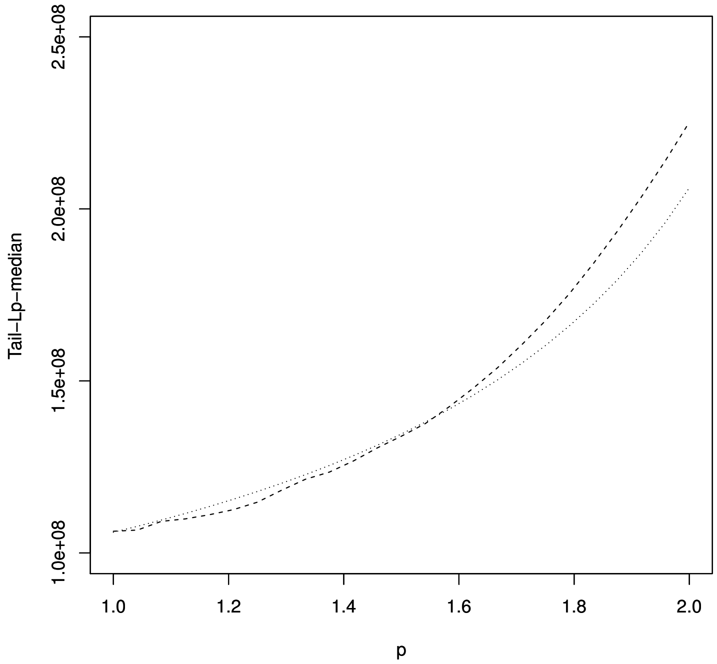

Figure 2.12: Real fire insurance data set. Extrapolated tail L^p-median estimators $\widehat{m}_p^W(1-1/n)$ (dashed line) and $\widetilde{m}_p^W(1-1/n)$ (dotted line) as functions of $p \in [1,2]$.

This indicates that, just like extreme quantiles and expectiles can be estimated using L^p-quantiles at other levels, an extreme Conditional Tail Expectation at level α can be estimated using an estimator of $m_p(\beta)$. The same principle applies to the Median Shortfall. One may then, like in Section 2.4.2.3, define composite, tail L^p-median based estimators of the Median Shortfall and Conditional Tail Expectation at extreme levels. The investigation of their practical properties is worthy of further research.

2.5 Extremiles: A new perspective on asymmetric least squares [5]

2.5.1 Introduction

Even though the risk measures we have worked on so far are valuable tools, they have all been criticised in the literature for either axiomatic or practical reasons. For example, the expectile terminology coined for e_τ is frustrated by the absence of any closed form expression in terms of tail expectations in the same way that the quantile q_τ is explicitly determined by the generalised inverse of the distribution function. This makes them more difficult to interpret than quantiles or Conditional Tail Expectations. Moreover, expectiles, with $\tau > 1/2$, are not comonotonically additive. This is a serious problem from a regulatory point of view, as expectiles can then give a false sense of diversification. The Conditional Tail Expectation, meanwhile, is (at least for continuous distributions) law-invariant, coherent and comonotonically additive, but relies exclusively on the tail event it is calculated on, and as a result tends to induce a conservative risk measure. This makes it appealing to regulators but less so to individual financial institutions.

The present section, based on [5], proposes a new L^2 analogue of quantiles, called extremiles, which defines a valuable alternative option for general statistical diagnoses. As shown in the following section, the new class is a generalisation of the usual central moment $\mathbb{E}(Y)$, which summarises the distribution of Y in the same way as quantiles do. Extremiles are by construction more alert/sensitive to the magnitude of extreme values than quantiles and make more efficient use of the available data since they rely on both the distance to observations and their probability, while quantiles only depend on the frequency of observations below or above the predictor and not on their values. Unlike expectiles, extremiles benefit from various equivalent explicit formulations and more intuitive interpretations. They are proportional to specific probability-weighted moments (PWMs) and can be estimated by L-statistics, M-statistics, linearised M-statistics and PWM-estimators. In addition, inference on extremiles is much easier than inference on expectiles, since their various estimators have closed form expressions and are straightforward to compute. Also, these estimators steer an advantageous middle course between the robustness of ordinary quantile estimators and sensitivity of extreme quantile estimators in the sense that the extremile estimators of order τ tend to be more tail sensitive than the τth quantile estimators for ordinary levels τ, but become more resistant for extremely high/low levels τ. Finally, extremiles are very closely connected to quantiles from an extreme value perspective, and provide an appropriate theory that better displays the interesting features of long-tailed population distributions. Although extremiles have been used in [86] in the context of estimating a frontier cost function (involving a conditional expectation), the paper [5] is the first to introduce the notion of extremiles as a new class of least squares analogues of quantiles, and to give a full study of this class, including relationships with expectiles and quantiles, as well as extensive statistical inference within this class.

The presented M- and L-estimates of extremiles, as well as their linearised M- and PWM-estimates, suffer from instability at very far tails due to data sparseness, especially for heavy-tailed distributions. This motivated us to extend their estimation and the underlying asymptotic theory far enough into the tails, which translates into considering the level $\tau = \tau_n \to 1$ as the sample size n goes to infinity. We show that these high τ_nth extremiles are asymptotically connected to extreme τ_nth quantiles by a constant depending on the tail index of the underlying distribution. Our first estimation method is based on this asymptotic connection, so that extreme value estimators of large quantiles and of the tail index can be used for extrapolation beyond the range of the data. Our second method relies directly on asymmetric least squares estimation. Although both approaches work quite well and either might be used in practice, we have a particular preference for the second due to some theoretical findings and simulation evidence.

The extremile and the quantile of Y with the same level τ are actually identical to the mean and the median, respectively, of a common asymmetric distribution $K_\tau(F)$, where F stands for the distribution function of Y and K_τ is a well-specified power transformation. The use of extremiles appears then naturally in the context of any decision theory where "optimistic" and "pessimistic" judgements are contrasted such as, for instance, extreme risk analysis, survival analysis and medical decision making. We focus here more specifically on the actuarial and financial risk management context, where extremiles bear much better than quantiles the burden of representing an alert risk measure to the magnitude of infrequent catastrophic losses. In particular, they are comonotonically additive, coherent spectral risk measures of law–invariant type. Also, they belong to the class of Wang DRMs with a concave distortion function that acts to downweight the likelihood of the most favourable outcomes, yielding a risk measure appropriate for the pessimistic decision theory formalised in [46]. To illustrate these ideas, we consider real data examples on trended hurricane losses and large claims in medical insurance.

This section is organised as follows. Section 2.5.2 presents a detailed description of the proposed new class, including its motivation, interpretation and basic properties. Section 2.5.3 deals with estimation of extremiles within the range of the data and beyond the sample maximum. Section 2.5.4 provides an interesting connection between extremiles and coherent law-invariant measures of risk in actuarial and financial management. Section 2.5.5 discusses the behaviour of the proposed notion of extremiles on simulated data and on two real data examples. Section 2.5.6 discusses potential avenues for further research.

2.5.2 The class of extremiles

2.5.2.1 Definition and motivation

For ease of presentation, **we assume throughout that F is continuous**. It is not hard to verify that q_τ is identical to the median of a random variable Z_τ having cumulative distribution function $F_{Z_\tau} = K_\tau(F)$, where

$$K_\tau(t) = \begin{cases} 1-(1-t)^{s(\tau)} & \text{if} \quad 0<\tau\le 1/2, \\ t^{r(\tau)} & \text{if} \quad 1/2\le\tau<1 \end{cases}$$

is a distribution function with support $[0,1]$, and $r(\tau) = s(1-\tau) = \log(1/2)/\log(\tau)$. Hence, the τth quantile can be derived from the alternative minimisation problem

$$q_\tau \in \underset{\theta\in\mathbb{R}}{\arg\min}\, \mathbb{E}\left\{J_\tau\left(F(Y)\right)\cdot\left[|Y-\theta|-|Y|\right]\right\},$$

with the special weight-generating function $J_\tau(\cdot) = K'_\tau(\cdot)$ on $(0,1)$. This does not seem to have been appreciated in the literature before. The parallel concept to the quantile q_τ, which we call *extremile of order τ of Y*, is then defined in a similar way by substituting the squared deviations in place of the absolute deviations:

$$\xi_\tau = \underset{\theta\in\mathbb{R}}{\arg\min}\, \mathbb{E}\left\{J_\tau\left(F(Y)\right)\cdot\left[|Y-\theta|^2-|Y|^2\right]\right\}. \tag{2.5.1}$$

As a matter of fact, while q_τ coincides with the median of the transformation Z_τ, it is easily seen that, whenever $\mathbb{E}|Z_\tau|<\infty$,

$$\xi_\tau = \mathbb{E}(Z_\tau). \tag{2.5.2}$$

We shall see in Proposition 2.5.2 below that the condition $\mathbb{E}|Z_\tau|<\infty$ is implied by $\mathbb{E}|Y|<\infty$, and therefore extremiles of any order exist as soon as Y has a finite first moment. Denote by $y_\ell = \inf\{y : F(y) > 0\}$ and $y_u = \sup\{y : F(y) < 1\}$ the lower and, respectively, upper endpoint of the support of F. One way of defining the extremile ξ_τ, for $0\le\tau\le 1$, is as the explicit quantity

$$\xi_\tau = \begin{cases} -\displaystyle\int_{y_\ell}^{0}\left\{1-[1-F(y)]^{s(\tau)}\right\}dy + \int_0^{y_u}[1-F(y)]^{s(\tau)}\,dy & \text{for} \quad 0\le\tau\le 1/2, \\ -\displaystyle\int_{y_\ell}^{0}[F(y)]^{r(\tau)}\,dy + \int_0^{y_u}\left(1-[F(y)]^{r(\tau)}\right)dy & \text{for} \quad 1/2\le\tau\le 1. \end{cases} \tag{2.5.3}$$

This follows from a general property of expectations (see p.117 of [231]). Clearly, F_{Z_τ} reduces to F when $\tau = 1/2$, and $\xi_{1/2}$ is just the expectation of Y, while the endpoints y_ℓ and y_u of the support of F

coincide respectively with the lower and upper extremiles ξ_0 and ξ_1, since $s(0) = r(1) = \infty$ in (2.5.3). A thorough description of basic properties of ξ_τ is given in Section 2.5.2.3.

Extremiles can be of considerable importance for modelling extremes of natural phenomena. When $\tau \geq 1/2$ with $r(\tau) = 1, 2, \ldots$ the τth extremile has a nice interpretation: it equals the expectation of the maximum of $r(\tau)$ independent copies $Y^1, \ldots, Y^{r(\tau)}$ of Y, i.e. $\xi_\tau = \mathbb{E}[\max(Y^1, \ldots, Y^{r(\tau)})]$. Of interest is the case $\tau \uparrow 1$, or equivalently $r(\tau) \to \infty$, which leads to access the upper endpoint y_u of the support of F. Likewise, if $\tau \leq 1/2$ with $s(\tau) = 1, 2, \ldots$ we have $\xi_\tau = \mathbb{E}[\min(Y^1, \ldots, Y^{s(\tau)})]$, the expectation of the minimum of $s(\tau)$ i.i.d. observations from Y. Of interest is also the case $\tau \downarrow 0$, or equivalently $s(\tau) \to \infty$, which leads to access the lower bound y_ℓ of the support of Y. For a general order τ, we have

$$\begin{aligned}
\mathbb{E}\left[\max\left(Y^1, \ldots, Y^{\lfloor r(\tau) \rfloor}\right)\right] \leq \quad \xi_\tau \quad &\leq \mathbb{E}\left[\max\left(Y^1, \ldots, Y^{\lfloor r(\tau) \rfloor + 1}\right)\right] \quad \text{if} \quad \frac{1}{2} \leq \tau < 1, \\
\mathbb{E}\left[\min\left(Y^1, \ldots, Y^{\lfloor s(\tau) \rfloor + 1}\right)\right] \leq \quad \xi_\tau \quad &\leq \mathbb{E}\left[\min\left(Y^1, \ldots, Y^{\lfloor s(\tau) \rfloor}\right)\right] \quad \text{if} \quad 0 < \tau \leq \frac{1}{2},
\end{aligned}$$

where $Y^1, Y^2, \ldots$ are i.i.d. observations from Y. The bracketing of ξ_τ becomes narrower when $\tau \uparrow 1$ or $\tau \downarrow 0$.

Yet, there is still another way of looking at ξ_τ for orders τ in $(0, 1)$. These extremiles are likely to be most useful when the quantile function q can be written in closed form, for then we have

$$\xi_\tau = \int_0^1 F_{Z_\tau}^{-1}(u)du = \int_0^1 q_t \, dK_\tau(t) = \int_0^1 J_\tau(t) q_t \, dt. \tag{2.5.4}$$

These expressions are key when it comes to proposing an estimator for an extremile. They clearly link the τth extremile of the random variable Y to its quantile function as a weighted average of q, which is the most convenient way of evaluating ξ_τ.

2.5.2.2 Relating extremiles and probability-weighted moments

Extremiles are closely related to the concept of probability-weighted moments (PWMs) introduced by [143] and defined by the quantities

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

where $p, r, s \geq 0$. These moments have been extensively utilised in extreme value procedures (see e.g. [47, 145]). Generally, a distribution is characterized either by the moments $M_{1,0,s}$ ($s = 0, 1, 2, \ldots$) or by the moments $M_{1,r,0}$ ($r = 0, 1, 2, \ldots$). In our approach, the moments $M_{1,0,s}$ (with $s \geq 0$) are favoured for describing the distribution of the population in the left tail (i.e. for $\tau \leq 1/2$), but a more convenient definition for ξ_τ in the right tail ($\tau \geq 1/2$) is formulated by considering the moments $M_{1,r,0}$ (with $r \geq 0$). Indeed, for $0 < \tau < 1$, we can rewrite (2.5.4) as

$$\xi_\tau = \int_0^1 J_\tau(t) q_t \, dt = \mathbb{E}\left[Y J_\tau\left(F(Y)\right)\right], \tag{2.5.5}$$

or equivalently

$$\xi_\tau = \begin{cases} s(\tau) M_{1,0,s(\tau)-1} & \text{for} \quad 0 < \tau \leq 1/2 \\ r(\tau) M_{1,r(\tau)-1,0} & \text{for} \quad 1/2 \leq \tau < 1. \end{cases} \tag{2.5.6}$$

Thus, extremiles are proportional to specific PWMs. The weight-generating function $J_\tau(\cdot)$ being monotonically increasing for $\tau \geq 1/2$ and decreasing for $\tau \leq 1/2$, the extremile ξ_τ in (2.5.5) depends by construction on all feasible values of Y, putting more weight to the high values for $\tau \geq 1/2$ and more weight to the low values for $\tau \leq 1/2$. Therefore ξ_τ is sensitive to the magnitude of extreme values for any order $\tau \in (0, 1)$. In contrast, the τth quantile q_τ is determined solely by the probability level (relative frequency) τ, and so it may be unaffected by extreme values whatever the shape of the tail distribution, unless τ is very extreme. In addition, when sample quantiles break down at $\tau \downarrow 0$

or $\tau \uparrow 1$, the various extremile estimators discussed below in Section 2.5.3 remain more resistant to extreme values thanks to their formulation as L-functionals whose weighting function J_τ converges to 0 pointwise on $(1/2, 1)$ at a geometrically fast rate as $\tau \uparrow 1$. Hence, the new class represents a compromise between the robustness of ordinary quantiles and the sensitivity of the extreme ones.

The interpretation in terms of expected minimum and expected maximum of a sample from Y, the impact on the lower and upper tails of the underlying distribution, the sensitivity and resistance properties to extremes and the proportionality to PWMs are of particular interest in extreme value theory. This inspired the name *extremiles* for this class. As a matter of taste, we prefer to regard equation (2.5.5), or equivalently (2.5.4), as the most convenient definition of extremiles from a mathematical perspective. It should be clear that the first-order necessary condition for optimality related to the initial definition of extremiles as an L^2 risk measure in (2.5.1) leads to

$$\xi_\tau = \frac{\mathbb{E}\left[Y J_\tau\left(F(Y)\right)\right]}{\mathbb{E}\left[J_\tau\left(F(Y)\right)\right]},$$

which in turn gives the identity (2.5.5), since $\mathbb{E}\left[J_\tau\left(F(Y)\right)\right] = 1$ for all $\tau \in (0, 1)$ by continuity of F. Note also that the transformed random variable Z_τ in (2.5.2) has the same expectation ξ_τ as the random variable $Y J_\tau\left(F(Y)\right)$, but not necessarily the same continuous distribution. In the particular case where $\tau \geq 1/2$ and $r(\tau)$ is an integer, it is easily seen that Z_τ is the maximum of $r(\tau)$ independent copies of Y; similarly, when $\tau \leq 1/2$ and $s(\tau)$ is an integer, Z_τ is the minimum of $s(\tau)$ independent copies of Y. In the general setting, we have the following characterization of Z_τ.

Proposition 2.5.1. *For a random variable Y with continuous distribution function F and quantile function q, and for $\tau \in (0, 1)$, define*

$$\phi_\tau = q \circ K_\tau^{-1} \circ F \quad i.e. \quad \phi_\tau(y) = q_{K_\tau^{-1}(F(y))}.$$

Then $Z_\tau \stackrel{d}{=} \phi_\tau(Y)$.

2.5.2.3 Basic properties

An alternative justification for the use of extremiles to describe probability distributions may be based on the following propositions. As established in Proposition 2.5.2, extremiles have basic properties similar to those of quantiles and expectiles.

Proposition 2.5.2. *(i) If $\mathbb{E}|Y| < \infty$ then ξ_τ exists for any $\tau \in (0, 1)$ and (if Y is not a constant) defines a continuous increasing function which maps $(0, 1)$ onto the set $\{y \in \mathbb{R} \,|\, 0 < F(y) < 1\}$.*

(ii) Two integrable random variables Y and $\widetilde{Y}$ have the same distribution if and only if $\xi_{Y,\tau} = \xi_{\widetilde{Y},\tau}$ for every $\tau \in (0, 1)$.

(iii) The τth extremile of the linear transformation $\widetilde{Y} = a + bY$, $a, b \in \mathbb{R}$, is given by

$$\xi_{\widetilde{Y},\tau} = \begin{cases} a + b\,\xi_{Y,\tau} & \text{if} \quad b > 0, \\ a + b\,\xi_{Y,1-\tau} & \text{if} \quad b \leq 0. \end{cases}$$

(iv) If Y has a symmetric distribution with mean μ, then $\xi_{1-\tau} = 2\mu - \xi_\tau$ for any $\tau \in (0, 1)$.

(v) If Y and $\widetilde{Y}$ are comonotone [i.e. there exists a third random variable $\bar{Y}$ and increasing functions u and v such that $Y = u(\bar{Y})$ and $\widetilde{Y} = v(\bar{Y})$], then $\xi_{Y+\widetilde{Y},\tau} = \xi_{Y,\tau} + \xi_{\widetilde{Y},\tau}$.

The wide range of distributions covered by property *(i)* can be extended further by relaxing the condition of finite "absolute" first moments. By *(ii)*, a distribution with finite absolute first moment is uniquely defined by its class of extremiles. Expectiles satisfy this law invariance property as well, but (to the best of our knowledge) for distributions with continuous densities, see Theorem 1 in [209]. Extremiles are also location and scale equivariant by property *(iii)* in the same way as quantiles and expectiles are. The desirable properties *(iv)* for symmetric distributions and *(v)* on comonotonic additivity are also shared by quantiles. Expectiles also satisfy property *(iv)*, but they are not comonotonically additive. This is a serious problem for a regulatory risk standard.

We now describe what happens for large extremiles ξ_τ and how they are linked to extreme quantiles q_τ when F is attracted to the Fisher-Tippett distributions of extreme value types:

(Fréchet) $\Phi_\gamma(y) = \exp\{-y^{-1/\gamma}\}$ with support $[0, \infty)$ and $\gamma > 0$;

(Weibull) $\Psi_\gamma(y) = \exp\{-(-y)^{-1/\gamma}\}$ with support $(-\infty, 0]$ and $\gamma < 0$;

(Gumbel) $\Lambda(y) = \exp\{-e^{-y}\}$ with support $\mathbb{R}$.

Let $\mathrm{DA}(\cdot)$ denote the maximum domain of attraction of an extreme value distribution, i.e. the set of distribution functions whose asymptotic distributions of suitably normalised maxima are of an extreme value type. Recall the notation $\Gamma(\cdot)$ for Euler's Gamma function.

Proposition 2.5.3. *Suppose that* $\mathbb{E}|Y| < \infty$.

(i) If $F \in DA(\Phi_\gamma)$ *with* $\gamma < 1$, *then*

$$\frac{\xi_\tau}{q_\tau} \sim \Gamma(1-\gamma)\{\log 2\}^\gamma \quad \textit{as} \quad \tau \uparrow 1.$$

(ii) If $F \in DA(\Psi_\gamma)$, *then* $y_u < \infty$ *and*

$$\frac{y_u - \xi_\tau}{y_u - q_\tau} \sim \Gamma(1-\gamma)\{\log 2\}^\gamma \ \textit{as}\ \tau \uparrow 1.$$

(iii) If $F \in DA(\Lambda)$ *and* $y_u = \infty$, *then* $\xi_\tau \sim q_\tau$ *as* $\tau \uparrow 1$. *If on the contrary* $y_u < \infty$, *then* $y_u - \xi_\tau \sim y_u - q_\tau$ *as* $\tau \uparrow 1$.

Note that the moments of $F \in \mathrm{DA}(\Phi_\gamma)$ do not exist when $\gamma > 1$. The index γ tunes the tail heaviness of the distribution function F, with higher positive values indicating heavier tails. A consequence of Proposition 2.5.3 is that the extremiles of distributions with heavy tails of index $\gamma < 1$ are asymptotically more spread than the quantiles since $\Gamma(1-\gamma)\{\log 2\}^\gamma > 1$. Indeed, the function $\varphi : y \mapsto \log\left(\Gamma(1-y)\{\log 2\}^y\right)$ has derivative

$$\varphi'(y) = -\digamma(1-y) + \log(\log 2), \quad \text{for all } y \in (0,1),$$

where $\digamma(x) = \Gamma'(x)/\Gamma(x)$ denotes Euler's digamma function. Because $\digamma$ is increasing and $\digamma(1) \approx -0.577$ (see Formulae 6.3.2 and 6.3.21 in [39]), the function φ is increasing. Hence, for any $\gamma \in (0,1)$, $\Gamma(1-\gamma)\{\log 2\}^\gamma > \exp(\varphi(0)) = 1$ as announced. A similar property holds for short-tailed distributions $F \in \mathrm{DA}(\Psi_\gamma)$ depending on the value of the extreme value index: numerically, when $-0.2907 < \gamma < 0$, extremiles are asymptotically closer to the right endpoint than quantiles are. By contrast, when $\gamma < -0.2907$, we rather have that quantiles are asymptotically closer to the right endpoint than extremiles are. Finally, ξ_τ and q_τ are asymptotically equivalent for light-tailed distributions $F \in \mathrm{DA}(\Lambda)$. A similar proposition can of course be given when F is rather in the minimum domain of attraction of a Fisher-Tippett extreme value distribution.

2.5.3 Estimation of extremiles

This section shows that results for ordinary and trimmed extremiles are easily obtained by means of L-statistics theory. By contrast, asymmetric least squares estimation of high extremiles leads to non-trivial developments from the perspective of extreme value theory. Our theorems are derived for independent and identically distributed random variables.

2.5.3.1 Ordinary extremiles

Given a random sample $Y_1, Y_2, \ldots, Y_n$ from Y, a natural estimator for the extremile of fixed order $\tau \in (0,1)$ is easily obtained by replacing F with its empirical version $\widehat{F}_n$ in (2.5.3), or equivalently, by replacing q_t with its empirical analogue $\widehat{q}_t$ in (2.5.4), leading to an L-statistic generated by the measure dK_τ:

$$\widehat{\xi}_\tau^L = \int_0^1 \widehat{q}_t \, dK_\tau(t) = \sum_{i=1}^n \left\{ K_\tau\left(\frac{i}{n}\right) - K_\tau\left(\frac{i-1}{n}\right) \right\} Y_{i,n},$$

where $Y_{1,n} \le Y_{2,n} \le \cdots \le Y_{n,n}$ denotes the ordered sample. Thanks to this special closed-form expression for $\widehat{\xi}_\tau^L$, we can rely for example on tools of [227, 231] for proving its consistency, asymptotic normality, and deriving the Berry-Esséen rate of uniform convergence $\mathrm{O}\left(n^{-1/2}\right)$.

Theorem 2.5.4. *For any index $\tau \in (0,1)$,*

(i) if $\mathbb{E}|Y|^\kappa < \infty$ for some $\kappa > 1$, then $\widehat{\xi}_\tau^L \xrightarrow{a.s.} \xi_\tau$ as $n \to \infty$.

(ii) if $\mathbb{E}|Y|^\kappa < \infty$ for some $\kappa > 2$, then $\sqrt{n}\left(\widehat{\xi}_\tau^L - \xi_\tau\right)$ has an asymptotic normal distribution with mean zero and variance $\sigma_\tau^2 = \int_0^1 \int_0^1 (s \wedge t - s\,t)\, J_\tau(s)\, J_\tau(t)\, dF^{-1}(s)\, dF^{-1}(t)$.

(iii) If $\mathbb{E}|Y|^3 < \infty$, then $\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left(\frac{\sqrt{n}}{\sigma_\tau} \left(\widehat{\xi}_\tau^L - \xi_\tau \right) \le t \right) - \Phi(t) \right| = \mathrm{O}\left(n^{-1/2}\right)$, for any $\tau \in [1 - (1/2)^{1/3}, (1/2)^{1/3}]$, where Φ stands for the standard normal distribution function.

Alternatively, one may estimate ξ_τ from its formulation (2.5.5) as a probability-weighted moment. Estimating the expectation by the empirical moment and replacing F by its empirical version $\widehat{F}_n$ leads to the τth sample extremile

$$\widehat{\xi}_\tau^{LM} = \xi_\tau(\widehat{F}_n) = \frac{1}{n} \sum_{i=1}^n J_\tau\left(\frac{i}{n}\right) Y_{i,n}.$$

This estimator is another L-statistic, whose asymptotic properties are closely linked to those of $\widehat{\xi}_\tau^L$ since the finite differences built on the function K_τ in the estimator $\widehat{\xi}_\tau^L$ can be approximated by the derivative J_τ when n is large. Note further that an M-estimator $\widehat{\xi}_\tau^M$ of ξ_τ is provided by solving the empirical least squares problem $\min_{\theta\in\mathbb{R}} \sum_{i=1}^n J_\tau\left(\frac{i}{n}\right) |Y_{i,n} - \theta|^2$, which yields the closed form expression

$$\widehat{\xi}_\tau^M = \frac{\sum_{i=1}^n J_\tau\left(\frac{i}{n}\right) Y_{i,n}}{\sum_{i=1}^n J_\tau\left(\frac{i}{n}\right)} \equiv \frac{\widehat{\xi}_\tau^{LM}}{\frac{1}{n}\sum_{i=1}^n J_\tau\left(\frac{i}{n}\right)}.$$

Since the denominator $\frac{1}{n}\sum_{i=1}^n J_\tau\left(\frac{i}{n}\right)$ in the last equality converges to 1 as $n \to \infty$, the L-statistic $\widehat{\xi}_\tau^{LM}$ in the numerator can thus be viewed as a *Linearised* variant of the M-estimator $\widehat{\xi}_\tau^M$. Both $\widehat{\xi}_\tau^{LM}$ and $\widehat{\xi}_\tau^M$ are first-order equivalent to $\widehat{\xi}_\tau^L$. It is also easily seen that the asymptotic distributions of $\widehat{\xi}_\tau^L$, $\widehat{\xi}_\tau^{LM}$ and $\widehat{\xi}_\tau^M$ are identical (see Example A pp. 277–278 in [227] and Example 1 in [230]).

Of particular interest are the expected minimum $\xi_\tau = \mathbb{E}[\min(Y^1, \ldots, Y^{s(\tau)})]$ and expected maximum $\xi_\tau = \mathbb{E}[\max(Y^1, \ldots, Y^{r(\tau)})]$ which correspond respectively to the cases where $s(\tau)$ and $r(\tau)$ in (2.5.6) are positive integers. In these special cases, estimation of ξ_τ might be most conveniently based on unbiased estimators of the probability-weighted moments $M_{1,0,s}$ ($s = 0, 1, 2, \ldots$) and $M_{1,r,0}$ ($r = 0, 1, 2, \ldots$) given respectively by (see [193])

$$\widehat{M}_{1,0,s} = \frac{1}{n} \sum_{i=1}^{n-s} \left(\prod_{j=1}^{s} \frac{n-i+1-j}{n-j} \right) Y_{i,n} \quad \text{and} \quad \widehat{M}_{1,r,0} = \frac{1}{n} \sum_{i=r+1}^{n} \left(\prod_{j=1}^{r} \frac{i-j}{n-j} \right) Y_{i,n}.$$

Then, when $s(\tau)$ and $r(\tau)$ in (2.5.6) are positive integers, the statistic

$$\widehat{\xi}_\tau^{\mathrm{PWM}} = \begin{cases} s(\tau)\widehat{M}_{1,0,s(\tau)-1} & \text{for} \quad 0 < \tau \le \frac{1}{2} \\ r(\tau)\widehat{M}_{1,r(\tau)-1,0} & \text{for} \quad \frac{1}{2} \le \tau < 1 \end{cases}$$

is an unbiased estimator of the extremile ξ_τ. It is straightforward to see from the asymptotic normality of $\widehat{M}_{1,0,s(\tau)-1}$ and $\widehat{M}_{1,r(\tau)-1,0}$ (see [159]) that $\widehat{\xi}_\tau^{\mathrm{PWM}}$ converges to the same normal distribution as the estimators $\widehat{\xi}_\tau^{L}$, $\widehat{\xi}_\tau^{LM}$ and $\widehat{\xi}_\tau^{M}$ do.

2.5.3.2 Large extremiles

In the domain of attraction of heavy-tailed distributions with tail index $\gamma < 1$, which plays in particular a crucial role in financial and actuarial considerations, extremiles ξ_τ are more sensitive to the magnitude of heavy right tails than quantiles q_τ are, as $\tau \uparrow 1$. However, the use of sample counterparts such as, for instance, $\widehat{\xi}_\tau^{L}$ to estimate such large population extremiles is typically not appropriate in the extreme region $\tau = \tau_n' \in [1-1/n, 1)$. We construct below two extreme value type estimators for $\xi_{\tau_n'}$ when $\tau_n' \uparrow 1$ at an arbitrary rate as $n \to \infty$. The first one is based on the use of the asymptotic connection between extremiles and quantiles, while the second one relies directly on asymmetric least squares estimation.

Estimation based on extreme quantiles A first option to estimate the extremile $\xi_{\tau_n'}$ is by using its asymptotic equivalence $\xi_{\tau_n'} \sim q_{\tau_n'}\,\mathcal{G}\,(\gamma)$ obtained in Proposition 2.5.3 *(i)*, where $\mathcal{G}(s) := \Gamma(1-s)\{\log 2\}^s$, $s < 1$, and γ is the tail index of Y. This suggests to define the quantile-based estimator

$$\widehat{\xi}_{\tau_n'}^{Q,\star} := \widehat{q}_{\tau_n'}^{\star}\,\mathcal{G}\,(\widehat{\gamma}) \tag{2.5.7}$$

by substituting in suitable estimators $\widehat{q}_{\tau_n'}^{\star}$ of $q_{\tau_n'}$ and $\widehat{\gamma}$ of γ. Thenceforth, $\widehat{\xi}_{\tau_n'}^{Q,\star}$ and $\widehat{q}_{\tau_n'}^{\star}$ inherit the same property about the spread of their population counterparts. On the other hand, the naive sample maximum $Y_{n,n}$ will not be a consistent estimator of the extreme quantile $q_{\tau_n'}$. In order to extrapolate outside the range of the available observations, one may use the traditional Weissman estimator $\widehat{q}_{\tau_n'}^{\star}$ already seen in (2.3.9):

$$\widehat{q}_{\tau_n'}^{\star} \equiv \widehat{q}_{\tau_n'}^{\star}(\tau_n) := \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}} \widehat{q}_{\tau_n}, \quad \text{where} \quad \widehat{q}_{\tau_n} := Y_{n-\lfloor n(1-\tau_n)\rfloor,n}$$

with $\tau_n \to 1$ such that $n(1-\tau_n) \to \infty$ as $n \to \infty$.

Our first step in order to establish the limiting distribution of $\widehat{\xi}_{\tau_n'}^{Q,\star}/\xi_{\tau_n'}$ is to analyse the bias incurred by using the estimator $\widehat{\xi}_{\tau_n'}^{Q,\star}$, i.e. the approximation error in Proposition 2.5.3 *(i)*.

Proposition 2.5.5. *Suppose that $\mathbb{E}|Y| < \infty$, and that condition $\mathcal{C}_2(\gamma,\rho,A)$ holds with $\gamma < 1$. Recall the notation $\mathcal{G}(s) = \Gamma(1-s)\{\log 2\}^s$, for $s < 1$. Then, as $\tau \uparrow 1$:*

$$\frac{\xi_\tau}{q_\tau} - \mathcal{G}(\gamma) = C_1(\gamma,\rho)A((1-\tau)^{-1}) + C_2(\gamma)(1-\tau) + \mathrm{o}(A((1-\tau)^{-1})) + \mathrm{o}(1-\tau).$$

Here

$$C_1(\gamma,\rho) = \begin{cases} \dfrac{\mathcal{G}(\gamma+\rho)-\mathcal{G}(\gamma)}{\rho} & \text{if } \rho < 0 \\ (\log 2)^\gamma \int_0^\infty e^{-t}t^{-\gamma}\left(\log(\log 2) - \log(t)\right)dt & \text{otherwise,} \end{cases}$$

and

$$C_2(\gamma) = -\frac{1}{2}\mathcal{G}(\gamma) + \left[1 + \frac{\log 2}{2}\right]\mathcal{G}(\gamma-1) - \frac{\log 2}{2}\mathcal{G}(\gamma-2).$$

This result makes it possible to obtain the rate of convergence of the estimator $\widehat{\xi}_{\tau_n'}^{Q,\star}$ via standard extrapolation arguments in the spirit of Theorem 4.3.8 in [145]. The next result goes in this sense.

Theorem 2.5.6. *Suppose that $\mathbb{E}|Y| < \infty$ and:*

(i) condition $\mathcal{C}_2(\gamma,\rho,A)$ holds with $\gamma < 1$ and $\rho < 0$;

(ii) $\tau_n \to 1$, $n(1-\tau_n) \to \infty$ and $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) \to \lambda \in \mathbb{R}$ as $n \to \infty$;

(iii) $\sqrt{n(1-\tau_n)}\,(\widehat{\gamma}-\gamma) \xrightarrow{d} Z$*, for a suitable estimator* $\widehat{\gamma}$ *of* γ*, where* Z *is a nondegenerate limiting random variable;*

(iv) $n(1-\tau_n') \to c < \infty$ *and* $\sqrt{n(1-\tau_n)}/\log[(1-\tau_n)/(1-\tau_n')] \to \infty$*.*

Then

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\widehat{\xi}^{Q,\star}_{\tau_n'}}{\xi_{\tau_n'}}-1\right) \xrightarrow{d} Z \quad as \quad n\to\infty.$$

We shall discuss in Section 2.5.5 below a concrete application to medical insurance data using $\widehat{\xi}^{Q,\star}_{\tau_n'}$ in conjunction with the Hill estimator $\widehat{\gamma}_H$ of γ and the Weissman estimator $\widehat{q}^{\star}_{\tau_n'}$ of $q_{\tau_n'}$. Other estimation methods of $q_{\tau_n'}$ and γ, such as those of the Peaks-Over-Threshold approach and maximum likelihood techniques, can of course be used as well in (2.5.7) to define $\widehat{\xi}^{Q,\star}_{\tau_n'}$.

Estimation based on intermediate extremiles Here, we first consider estimating a high extremile ξ_{τ_n} of intermediate level τ_n satisfying $\tau_n \to 1$ and $n(1-\tau_n)\to\infty$ as $n\to\infty$. Then, we extrapolate the resulting estimate to the very extreme level τ_n' which approaches 1 at an arbitrarily fast rate in the sense that $n(1-\tau_n') \to c$, for some finite nonnegative constant c.

A natural estimator of the intermediate extremile ξ_{τ_n} follows from the solution of the asymmetric least squares problem as

$$\widehat{\xi}^{M}_{\tau_n} = \frac{\sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right)Y_{i,n}}{\sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right)} = \frac{\int_0^1 J_{\tau_n}\left(\frac{\lceil nt\rceil}{n}\right)Y_{\lceil nt\rceil,n}\,dt}{\frac{1}{n}\sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right)}.$$

This is the M-estimator $\widehat{\xi}^{M}_{\tau}$ presented in Section 2.5.3.1 when $\tau=\tau_n$. Alternative estimators can be defined by plugging $\tau=\tau_n$ in our L-estimator and LM-estimator, resulting in the following estimators:

$$\widehat{\xi}^{L}_{\tau_n} = \sum_{i=1}^n\left\{K_{\tau_n}\left(\frac{i}{n}\right)-K_{\tau_n}\left(\frac{i-1}{n}\right)\right\}Y_{i,n} = \int_0^1 J_{\tau_n}(t)Y_{\lceil nt\rceil,n}\,dt$$

$$\text{and}\quad \widehat{\xi}^{LM}_{\tau_n} = \frac{1}{n}\sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right)Y_{i,n} = \int_0^1 J_{\tau_n}\left(\frac{\lceil nt\rceil}{n}\right)Y_{\lceil nt\rceil,n}\,dt.$$

The asymptotic normality of these estimators requires, in addition to condition $\mathcal{C}_2(\gamma,\rho,A)$, to control the empirical quantile function $t\mapsto Y_{\lceil nt\rceil,n}$ in the central part of the distribution of Y. This is why we introduce the following extra condition:

($\mathcal{H}$) The support of Y is an interval, and on its interior F is twice differentiable with positive probability density function f and

$$\sup_{0<t<1} t(1-t)\frac{f'(q_t)}{[f(q_t)]^2} < \infty.$$

Condition ($\mathcal{H}$) makes it possible to approximate the empirical quantile process $t\mapsto Y_{\lceil nt\rceil,n}$ by a sequence of standard Brownian bridges in the central part of the interval $(0,1)$, as well as in the far left tail due to the geometrically strong penalisation of left tail quantiles by the weighting function J_{τ_n}. A rigorous statement can be found in Theorem 6.2.1 of [82] and Proposition 2.4.9 in [145], among others. Finally, we shall slightly strengthen the heavy-tailed assumption by assuming:

$$\lim_{t\to+\infty} t\frac{f(t)}{1-F(t)} = \gamma. \tag{2.5.8}$$

This von Mises condition indeed implies the heavy-tailed assumption, as shown in Theorem 1.11 p.17 of [145] and Proposition 2.1 p.60 of [47]. All examples of commonly used Pareto-type distributions satisfy (2.5.8); see pp.59–60 of [47]. Under these assumptions, we unravel the common limit distribution of the normalised estimators $\widehat{\xi}^{L}_{\tau_n}/\xi_{\tau_n}$, $\widehat{\xi}^{LM}_{\tau_n}/\xi_{\tau_n}$ and $\widehat{\xi}^{M}_{\tau_n}/\xi_{\tau_n}$.

Theorem 2.5.7. *Suppose that* $\mathbb{E}|Y|<\infty$ *and:*

- *conditions* $\mathcal{C}_2(\gamma,\rho,A)$*,* $(\mathcal{H})$ *and* (2.5.8) *hold with* $\gamma < 1/2$*;*
- *the sequence* $\tau_n \uparrow 1$ *is such that* $n(1-\tau_n) \to \infty$ *and* $\sqrt{n(1-\tau_n)}\,A((1-\tau_n)^{-1}) = \mathrm{O}(1)$*.*

Then, if $\widehat{\xi}_{\tau_n}$ *is either* $\widehat{\xi}^{L}_{\tau_n}$*,* $\widehat{\xi}^{LM}_{\tau_n}$ *or* $\widehat{\xi}^{M}_{\tau_n}$*, we have*

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{\xi}_{\tau_n}}{\xi_{\tau_n}} - 1\right) \xrightarrow{d} \frac{\gamma\sqrt{\log 2}}{\Gamma(1-\gamma)}\int_0^\infty e^{-s}s^{-\gamma-1}W(s)ds$$

where W *is a standard Brownian motion. In other words, the above limit distribution is Gaussian centered with variance*

$$V(\gamma) = \left(\frac{\gamma}{\Gamma(1-\gamma)}\right)^2 (\log 2)\int_0^\infty\int_0^\infty e^{-s-t}(st)^{-\gamma-1}(s\wedge t)\,ds\,dt.$$

Turning now to the extreme level τ_n', we have under the model assumption of heavy-tailed distributions that

$$\frac{\xi_{\tau_n'}}{q_{\tau_n'}} \sim \mathcal{G}(\gamma) \sim \frac{\xi_{\tau_n}}{q_{\tau_n}} \quad\text{and hence}\quad \frac{\xi_{\tau_n'}}{\xi_{\tau_n}} \sim \frac{q_{\tau_n'}}{q_{\tau_n}} \quad\text{as}\quad n\to\infty,$$

in view of Proposition 2.5.3 *(i)*. This motivates the alternative purely extremile-based estimator

$$\widehat{\xi}^{M,\star}_{\tau_n'} := \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}}\widehat{\xi}^{M}_{\tau_n}. \tag{2.5.9}$$

This is still a Weissman-type device which, in contrast to $\widehat{\xi}^{Q,\star}_{\tau_n'}$ in (2.5.7), relies crucially on our intermediate asymmetric least squares estimator $\widehat{\xi}^{M}_{\tau_n}$. Another option would be to replace the intermediate M-estimator $\widehat{\xi}^{M}_{\tau_n}$ in (2.5.9) by either the L-estimator $\widehat{\xi}^{L}_{\tau_n}$ or the LM-estimator $\widehat{\xi}^{LM}_{\tau_n}$. The resulting extrapolated L- and LM-estimators can easily be shown to share the asymptotic properties of $\widehat{\xi}^{M,\star}_{\tau_n'}$ stated below. However, experiments with simulated data indicate that these estimators perform no better than $\widehat{\xi}^{M,\star}_{\tau_n'}$. We therefore restrict our attention to the latter estimator.

Theorem 2.5.8. *Suppose that* $\mathbb{E}|Y| < \infty$ *and:*

- *conditions* $\mathcal{C}_2(\gamma,\rho,A)$*,* $(\mathcal{H})$ *and* (2.5.8) *hold with* $\gamma < 1/2$ *and* $\rho < 0$*;*
- *the sequence* $\tau_n \uparrow 1$ *is such that* $n(1-\tau_n) \to \infty$ *and*
$$\sqrt{n(1-\tau_n)}\max(A((1-\tau_n)^{-1}), 1-\tau_n) = \mathrm{O}(1);$$
- $\sqrt{n(1-\tau_n)}\,(\widehat{\gamma}-\gamma) \xrightarrow{d} Z$*, for a suitable estimator* $\widehat{\gamma}$ *of* γ*, where* Z *is a nondegenerate limiting random variable;*
- $n(1-\tau_n') \to c < \infty$ *and* $\sqrt{n(1-\tau_n)}/\log[(1-\tau_n)/(1-\tau_n')] \to \infty$*.*

Then

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\widehat{\xi}^{M,\star}_{\tau_n'}}{\xi_{\tau_n'}} - 1\right) \xrightarrow{d} Z \quad \text{as} \quad n\to\infty.$$

Like the quantile-based estimator $\widehat{\xi}^{Q,\star}_{\tau_n'}$, the extrapolated M-estimator $\widehat{\xi}^{M,\star}_{\tau_n'}$ inherits the limit distribution of $\widehat{\gamma}$ with a slightly slower rate of convergence.

2.5.4 Extremiles as risk measures

We investigate below the properties of the extremile functional from the point of view of the axiomatic theory of risk measures. The discussion in Section 2.5.4.1 pertains to non-negative loss distributions, while Section 2.5.4.2 focuses on real-valued profit-loss random variables. Section 2.5.4.3 examines the connection between tail extremiles, expectiles and ES.

2.5.4.1 Coherency, regularity and pessimism

Taking ξ_τ as a margin (amount of capital as a hedge against extreme risks), a larger ξ_τ is then a more prudential margin requirement and results in larger τ. As such, the index τ reflects the level of prudentiality to be typically set by regulators and/or the management level. When $\tau = 1/2$, ξ_τ is the net expected loss. When $\tau < 1/2$, the values of the risk measure ξ_τ are less than $\mathbb{E}(Y)$, which is not appropriate when $Y \geq 0$ is a loss random variable. Hence, a natural range for the security level τ is given by $[1/2, 1]$. The use of ξ_τ as a proper risk measure can be justified further as follows. Recall the class of (coherent) Wang distortion risk measures,

$$\Pi[Y] = \int_0^\infty g(1 - F(y))dy, \tag{2.5.10}$$

where g is a nondecreasing and concave function with $g(0) = 0$ and $g(1) = 1$. The review article by [257] discusses a number of important distortion functions including those we considered in Section 2.2.5: the Proportional Hazard transform $g(t) = \mathrm{PH}_\lambda(t) := t^\lambda$, for the so-called risk-aversion index $\lambda \in (0, 1]$, and the Dual Power transform $g(t) = \mathrm{DP}_r(t) := 1-(1-t)^r$ for $r \geq 1$. The ES (or equivalently CTE, in our context of continuous distribution functions) can be expressed in terms of the distortion function $t \mapsto (t/(1-\tau))\mathbb{I}\{0 \leq t < 1-\tau\} + \mathbb{I}\{1-\tau \leq t \leq 1\}$. While the VaR and ES use only a small part of the loss distribution, both PH and DP distortion functions utilise the whole loss distribution and are more reliable for the purpose of differentiating between more and less risky distributions. Extremiles themselves clearly belong to the class of risk measures of the form (2.5.10), with the alternative concave distortion function

$$g_\tau(t) := 1 - K_\tau(1-t) = 1 - (1-t)^{r(\tau)}, \ \frac{1}{2} \leq \tau \leq 1.$$

It turns out that this function is very closely related to the DP transform in the sense that $g_\tau = \mathrm{DP}_{r(\tau)}$. Its formulation in terms of the asymmetry parameter $\tau \in [1/2, 1]$ allows, like the tail probability τ in the VaR q_τ and ES $\nu_\tau = \mathbb{E}[Y|Y > q_\tau]$, for better interpretability compared to the distortion parameter $r \in [1, \infty)$ in the DP transform. Let us highlight in particular that the asymmetric least squares formulation makes the comparison of extremiles (and, incidentally, of DP transforms) with both VaR and ES easier and more insightful than the purely distortion-based interpretation: being least squares analogues of quantiles, extremiles rely on the distance to observations and depend by construction on both the tail losses and their probability, making thus more efficient use of the available data, whereas VaR only depends on the frequency of tail losses and ES only depends on the tail event. Besides, the extremile distortion function g_τ results in a more widely applicable risk measure than the PH-measure, at least in the following respect: the empirical estimators of both measures allow one to determine the price of an insurance risk without recourse to any fitting of a parametric model. However, as shown by [169], the asymptotic normality of the empirical PH-measure, $\int_0^\infty (1 - \widehat{F}_Y(y))^\lambda dy$, does not cover the range $0 < \lambda \leq 1/2$, and requires the assumption that $\mathbb{E}|Y|^\kappa < \infty$, for some $\kappa > 1/(\lambda - 1/2)$, when $1/2 < \lambda \leq 1$. Hence, the number κ of required finite moments tends to $+\infty$ as $\lambda \downarrow 1/2$. By contrast, as shown in Theorem 2.5.4, inference on ξ_τ is feasible for any index τ, under the weaker assumption that $\mathbb{E}|Y|^\kappa < \infty$ for some $\kappa > 2$.

It is straightforward to see that the extremile-based risk measure $\Pi_\tau[Y] := \xi_{Y,\tau}$, for $\tau \in (1/2, 1)$, satisfies the following two natural properties related to risk loading:

(A1) Positive loading and no ripoff: $\mathbb{E}(Y) < \Pi_\tau[Y] < y_u$, $\lim_{\tau\to1/2} \Pi_\tau[Y] = \mathbb{E}(Y)$, and $\lim_{\tau\to1} \Pi_\tau[Y] = y_u$.

(A2) No unjustified risk-loading: if $P(Y = b) = 1$ for some constant b then $\Pi_\tau[Y] = b$.

More importantly, the extremile risk measure also satisfies the requirements for being a coherent risk measure:

(A3) Translation and scale invariance: $\Pi_\tau[a + bY] = a + b\Pi_\tau[Y]$, for any $a \in \mathbb{R}$ and $b > 0$.

(A4) Subadditivity: $\Pi_\tau[Y + \widetilde{Y}] \leq \Pi_\tau[Y] + \Pi_\tau[\widetilde{Y}]$, for any loss variables Y and $\widetilde{Y}$.

(A5) Preserving of stochastic order: $\Pi_\tau[Y] \le \Pi_\tau[\widetilde{Y}]$ if $Y \le \widetilde{Y}$ with probability 1.

Finally, $\Pi_\tau[\cdot]$ is a regular risk measure since it fulfills the following additional fundamental conditions (as imposed e.g. in Definition 5 of [46]):

(A6) Law invariance: $\Pi_\tau[Y] = \Pi_\tau[\widetilde{Y}]$ if Y and $\widetilde{Y}$ have the same distribution.

(A7) Comonotonic additivity: $\Pi_\tau[Y + \widetilde{Y}] = \Pi_\tau[Y] + \Pi_\tau[\widetilde{Y}]$ if Y and $\widetilde{Y}$ are comonotone.

According to Theorem 1 in [46], being coherent and regular, $\Pi_\tau[\cdot]$ is then a pessimistic risk measure in the sense that the corresponding distortion function g_τ acts to depress the implicit likelihood of the most favourable outcomes, and to accentuate the likelihood of the least favourable ones (see Definition 4 in [46] for a formal characterisation of pessimistic risk measures). As such, the index τ becomes a natural measure of the degree of pessimism. Note also that equation (2.5.4) makes $\Pi_\tau[\cdot]$ a spectral risk measure in the sense of [40].

2.5.4.2 Real-valued profit-loss random variables

The extremile-based risk measure $\Pi_\tau[Y]$ can still be defined when Y is an asset return, which can take any real value, by using the more general expression (2.5.3), that is,

$$\xi_\tau = \int_{-\infty}^{0} \left[g_\tau(1 - F(y)) - 1\right] dy + \int_{0}^{\infty} g_\tau(1 - F(y))dy,$$

or its various equivalent formulations described in Sections 2.5.2.1-2.5.2.2, where

$$g_\tau(t) := t^{s(\tau)}\mathbb{I}\{0 \le \tau < 1/2\} + [1 - (1 - t)^{r(\tau)}]\mathbb{I}\{1/2 \le \tau \le 1\}.$$

A number of studies, including [46, 169] and the references therein, have recognised the usefulness of such Choquet integrals in actuarial and financial applications as well as in the area of measuring economic inequality. The coherency axioms of the risk measure $\Pi_\tau[Y] := -\xi_{Y,\tau}$ for asset returns can easily be checked by making use of the alternative formulation (2.5.5) of ξ_τ as a probability-weighted moment. The key argument is that the special weight-generating function $J_\tau(\cdot)$ is an admissible risk spectrum in the sense of [40].

Proposition 2.5.9. *For the profit-loss Y of a given portfolio and for any $0 < \tau < 1/2$,*

$$\Pi_\tau[Y] = -\xi_{Y,\tau} \equiv -\mathbb{E}\left[Y J_\tau\left(F(Y)\right)\right]$$

is a coherent spectral risk measure.

In the special case where the power $s(\tau)$ in the distortion function g_τ is an integer, we recover the particularly attractive and very intuitive interpretation of the so-called *MINVAR* risk measure introduced in [75], namely the negative of the expected minimum of $s(\tau)$ independent observations from Y. See also [121].

Despite their statistical virtues and all their nice axiomatic properties as spectral risk measures and concave distortion risk measures, extremiles have an apparent limitation when applied to distributions with infinite mean. This should not be considered to be a serious disadvantage however, at least in financial and actuarial applications, since the definition of a coherent risk measure for distributions with an infinite first moment is not clear, see the discussion in Section 3 of [207]. Whether operational risk models in which losses have an infinite mean make sense in the first place has also recently been questioned by [76].

2.5.4.3 Connection between extremiles, expectiles and Expected Shortfall

We consider both cases of non-negative loss distributions and real-valued profit-loss distributions with heavy right tails. In the case of a profit-loss distribution, the financial position Y stands for the negative of the asset return so that the right tail of F corresponds to the negative of extreme losses. Accordingly, as established in Proposition 2.5.3, the corresponding extremile-based risk measure ξ_τ is more pessimistic, from the risk management viewpoint, than the traditional quantile-VaR q_τ for large values of τ. Next, we show that ξ_τ is more pessimistic than the expectile-based VaR e_τ as well, in the standard case of finite-variance distributions, but it is always less pessimistic than the ES ν_τ, for large τ. The key ingredients to show this are the following asymptotic connections.

Proposition 2.5.10. *Suppose that $\mathbb{E}|Y| < \infty$ and Y has a heavy-tailed distribution with tail index $0 < \gamma < 1$. Then, as $\tau \to 1$,*

$$\frac{\xi_\tau}{e_\tau} \sim \Gamma(1-\gamma)\{(\gamma^{-1}-1)\log 2\}^\gamma \quad \textit{and} \quad \frac{\xi_\tau}{\nu_\tau} \sim \Gamma(2-\gamma)\{\log 2\}^\gamma.$$

We now show how Proposition 2.5.10 entails that high extremiles are more conservative than high expectiles at the same level, when $\gamma \in (0, 1/2)$. The idea is simply to note that for $\gamma \in (0, 1/2)$, we have $\gamma^{-1}-1 > 1$, and therefore $\Gamma(1-\gamma)\{(\gamma^{-1}-1)\log 2\}^\gamma > \Gamma(1-\gamma)\{\log 2\}^\gamma > 1$, establishing thus that $\xi_\tau > e_\tau$ for τ large enough when the underlying distribution has a finite variance. This is, however, no longer valid for the heaviest tails since the proportionality constant $\Gamma(1-\gamma)\{(\gamma^{-1}-1)\log 2\}^\gamma$ tends to $\log 2 < 1$ as $\gamma \to 1$. More precisely, a numerical study shows that high extremiles shall be more conservative than high expectiles if and only if $0 < \gamma < \gamma_0$, with $\gamma_0 \approx 0.8729$. By contrast, that high extremiles are always less conservative than their ES analogues is a consequence of the inequality $\Gamma(2-\gamma)\{\log 2\}^\gamma < \{\log 2\}^\gamma < 1$, for all $\gamma \in (0, 1)$.

Finally, we would like to stress why our finding in Proposition 2.5.10 that $q_\tau < \xi_\tau < \nu_\tau$, as $\tau \to 1$, is not a contradiction to Theorem 13 in [93]. This theorem states that any coherent, law-invariant risk measure satisfying the Fatou property and greater than or equal to q_τ (for every $Y \in L^\infty$) must also be greater than or equal to ν_τ. The extremile ξ_τ defines a law-invariant, coherent, and hence convex risk measure, satisfying thus the Fatou property, see Theorem 2.1 in [170] and Theorem 2.2 in [244]. However, the key condition of Delbaen's theorem that $q_\tau \le \xi_\tau$ is not fulfilled for short-tailed distributions with a large $|\gamma|$, as discussed below our Proposition 2.5.3.

2.5.5 Finite-sample study

2.5.5.1 Finite-sample performance on simulated data

The finite-sample performance of the two rival estimators $\widehat{\xi}^{Q,\star}_{\tau'_n}$ in (2.5.7) and $\widehat{\xi}^{M,\star}_{\tau'_n}$ in (2.5.9) is considered in Section 3.3.3 of [5] on the Student $t_{1/\gamma}$ distribution, the Pareto distribution and the Fréchet distribution, for a tail index $\gamma \in \{1/4, 1/3, 2/5\}$. In terms of the relative MSE, it may be seen that $\widehat{\xi}^{M,\star}_{\tau'_n}$ has a very similar behaviour to $\widehat{\xi}^{Q,\star}_{\tau'_n}$ in the Fréchet model, but performs better in both the Pareto and Student models, for all considered values of γ. It may also be seen that $\widehat{\xi}^{M,\star}_{\tau'_n}$ is superior in terms of bias in all scenarios. Although either might be used in practice, we would therefore have a particular preference for the purely extremile-based estimator $\widehat{\xi}^{M,\star}_{\tau'_n}$.

2.5.5.2 Data example 6: Trended hurricane losses

We first consider a dataset on trended or, equivalently, inflation-adjusted (to 1981 using the U.S. Residential Construction Index) hurricane losses that occurred between 1949 and 1980. Figure 2.13 (left panel) displays the histogram and scatterplot of the recorded $n = 35$ trended hurricane losses in excess of \$5 million (in units of \$1,000). An analysis of this data set conducted from a central point of view is presented in [169]; our focus here is rather on the right tail of the observations from a descriptive point of view. We treat the 35 amounts as the outcomes of i.i.d. non-negative loss random variables $Y_1, \ldots, Y_{35}$. The corresponding sample mean and standard deviation are 199,900

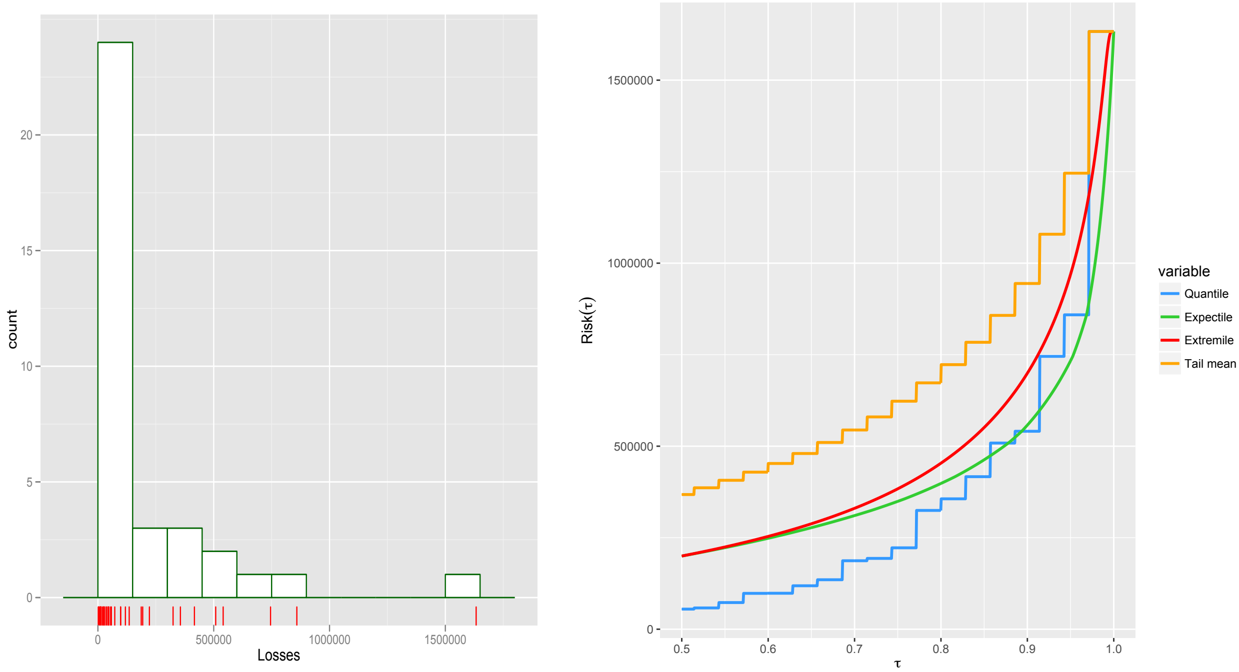

Figure 2.13: Trended Hurricane Losses data. Left panel: histogram and scatterplot, right panel: empirical expected shortfall, extremile, expectile and quantile τ-risk measures.

and 325,807, respectively. The empirical estimates of the expected shortfall, quantile, expectile and extremile risk measures are graphed in Figure 2.13 (right panel) against the security level τ. The first impression to be gained from this figure is the lack of smoothness and stability of both sample quantile-VaR $\widehat{q}_\tau$ (blue) and ES $\widehat{\nu}_\tau$ (orange): their discreteness as piecewise constant functions of the argument τ is a serious defect, especially in the upper tail. Indeed, a small change in τ can trigger a (severe) jump in the values of the estimated VaR and ES, while the "steps" result in the same or similar measures for significantly different risk levels. By contrast, the sample extremile $\widehat{\xi}_\tau = \widehat{\xi}_\tau^L$ (red) and expectile $\widehat{e}_\tau$ (green) are very stable and change continuously and increasingly without recourse to any smoothness procedure.

When comparing the four estimated risk measures at the same level τ, it can be seen that the ES $\widehat{\nu}_\tau$, in orange, is much larger and hence more conservative than the extremile $\widehat{\xi}_\tau$ in red. By contrast, the quantile $\widehat{q}_\tau$, in blue, remains less alert to extreme risks than $\widehat{\xi}_\tau$ until it breaks down at $\tau = (n-1)/n = 0.9714$. Thenceforth, for all $\tau > 0.9714$, $\widehat{q}_\tau$ becomes identical to $\widehat{\nu}_\tau$ which in turn coincides with the maximum catastrophic loss $Y_{n,n}$ = 1,633,000, whereas $\widehat{\xi}_\tau$ provides less pessimistic values. Finally, although the expectile $\widehat{e}_\tau$, in green, exhibits a smooth evolution, it diverges from $\widehat{\xi}_\tau$ in the region $\tau \in [0.8, 0.975]$ and becomes less alert to infrequent disasters. The extremile in red seems to provide a compromise between the pessimistic ES in orange and the optimistic expectile in green.

Of particular interest is the deviation between the estimated extremile $\widehat{\xi}_\tau$ and quantile $\widehat{q}_\tau$. Being the estimates, respectively, of the mean and the median of the same asymmetric distribution $K_\tau(F)$ of the transformation $\phi_\tau(Y)$, a significant deviation between their values diagnoses a heavier right tail of Y. Thereby the comparison above with the same level τ may be viewed as an explanatory tool for quantifying the "riskiness" implied by the distribution of Y, rather than as a method for final analysis, especially since we know that non-extrapolated sample extremiles and quantiles will be inconsistent in the typical extreme range $\tau \geq 1 - 1/n$.

2.5.5.3 Data example 3 revisited: The Society of Actuaries medical data set

We revisit here the data set studied in Section 2.3.7.2. A histogram of the data is given again in Figure 2.14 (a). As in Section 2.3.7.2, we deal here only with the 75,789 claims for 1991 and we compare risk measures estimates at the extreme level $\tau_n' = 1 - p_n$, with $p_n = 1/100{,}000 < 1/n$.

Figure 2.14 (b) shows the Weissman estimate $\widehat{q}^{\star}_{\tau_n'}$ against the effective sample size $k = n(1-\tau_n)$ (solid black curve). A stable region appears for k from 150 up to 500, leading to estimates between 3.73 and 4.12 million, with an averaged estimate around 3.90 million. This tail risk estimate does not exceed the sample maximum $Y_{n,n} = 4.51$ million (indicated by the horizontal pink line).

An alternative estimator of the same extreme quantile-VaR $q_{\tau_n'}$ is found from Section 2.3.6:

$$\widehat{e}^{\star}_{\widehat{\tau}_n'} = \left(\frac{1-\widehat{\tau}_n'}{1-\tau_n}\right)^{-\widehat{\gamma}_H} \widehat{e}_{\tau_n} \quad \text{where} \quad \widehat{\tau}_n' = 1 - \frac{1-\tau_n'}{\widehat{\gamma}_H^{-1} - 1} = 1 - \frac{p_n}{\widehat{\gamma}_H^{-1} - 1}$$

and $\widehat{e}_{\tau_n}$ stands for the empirical counterpart of the expectile e_{τ_n}, with $\tau_n = 1 - k/n$. The plot of this composite expectile estimator $\widehat{e}^{\star}_{\widehat{\tau}_n'}$ against k (dashed grey curve) indicates an averaged estimate of around 4.13 million for $k \in [150, 500]$. In contrast to $\widehat{q}^{\star}_{\tau_n'}$ which relies on a single order statistic $Y_{n-k,n}$, the extrapolated expectile estimator $\widehat{e}^{\star}_{\widehat{\tau}_n'}$ is based on the least asymmetrically weighted squares estimator $\widehat{e}_{\tau_n}$, and hence is more sensitive to the magnitude of infrequent large claims. Yet, both $\widehat{e}^{\star}_{\widehat{\tau}_n'}$ and $\widehat{q}^{\star}_{\tau_n'}$ actually estimate the median $q_{\tau_n'}$ of the asymmetric heavy-tailed distribution $K_{\tau_n'}(F)$ of $Z_{\tau_n'} \stackrel{d}{=} \phi_{\tau_n'}(Y)$, see Proposition 2.5.1. The burden of representing a pessimistic risk measure is thwarted by the robustness properties of the median, especially in a heavy-tailed context. The mean of $Z_{\tau_n'}$, which is nothing but the extremile $\xi_{\tau_n'}$, bears naturally much better this burden. The plots of its two estimators $\widehat{\xi}^{Q,\star}_{\tau_n'}$ (rainbow curve) and $\widehat{\xi}^{M,\star}_{\tau_n'}$ (dotted black curve), defined respectively in (2.5.7) and (2.5.9), clearly afford more pessimistic risk information than the plots of $\widehat{q}^{\star}_{\tau_n'}$ and $\widehat{e}^{\star}_{\widehat{\tau}_n'}$. The final results based on averaging these extremile estimates from the stable region $k \in [150, 500]$ are 4.83 million for $\widehat{\xi}^{Q,\star}_{\tau_n'}$ and 4.78 million for $\widehat{\xi}^{M,\star}_{\tau_n'}$. These estimates deserve indeed to be qualified as "pessimistic" since they do lie beyond the range of the data, but not by much. This might be good news to practitioners whose concern is to contrast "pessimistic" and "optimistic" judgments as in the duality between the mean and the median. Besides this duality, it should also be clear that the resulting extremile estimates have their own intuitive interpretation. Indeed, since $\tau_n' = (1/2)^{1/r}$ with $r \approx 69{,}314$, the quantity $\xi_{\tau_n'} \equiv \mathbb{E}[\max(Y^1, \ldots, Y^r)]$ gives the expected maximum claim amount among a fixed number of 69,314 potential claims.

The risk estimates $\widehat{q}^{\star}_{\tau_n'}$, $\widehat{e}^{\star}_{\widehat{\tau}_n'}$, $\widehat{\xi}^{M,\star}_{\tau_n'}$ and $\widehat{\xi}^{Q,\star}_{\tau_n'}$, graphed in Figure 2.14 (b), are all based on the Hill estimator $\widehat{\gamma}_H$ of γ and the Weissman estimator $\widehat{q}^{\star}_{\tau_n'}$ of $q_{\tau_n'}$. Instead, one may choose to use the maximum likelihood (ML) and Peaks-Over-Threshold (POT) estimators of γ and $q_{\tau_n'}$. The POT estimator of $q_{\tau_n'}$ has the form

$$\widetilde{q}^{\star}_{\tau_n'} = Y_{n-k,n} + \frac{\widetilde{\sigma}_{\text{ML}}}{\widetilde{\gamma}_{\text{ML}}}\left(\left(\frac{np_n}{k}\right)^{-\widetilde{\gamma}_{\text{ML}}} - 1\right),$$

where $\widetilde{\sigma}_{\text{ML}}$ and $\widetilde{\gamma}_{\text{ML}}$ are chosen here to be the ML estimates for the parameters σ and γ of the Generalised Pareto approximation (see e.g. p.158 of [47]). Replacing $\widehat{\gamma}_H$ and $\widehat{q}^{\star}_{\tau_n'}$ by their respective analogues $\widetilde{\gamma}_{\text{ML}}$ and $\widetilde{q}^{\star}_{\tau_n'}$ in the expectile and extremile risk estimates $\widehat{e}^{\star}_{\widehat{\tau}_n'}$, $\widehat{\xi}^{M,\star}_{\tau_n'}$ and $\widehat{\xi}^{Q,\star}_{\tau_n'}$, we get the alternative versions

$$\widetilde{e}^{\star}_{\widetilde{\tau}_n'} := \left(\frac{1-\widetilde{\tau}_n'}{1-\tau_n}\right)^{-\widetilde{\gamma}_{\text{ML}}} \widehat{e}_{\tau_n} \quad \text{where} \quad \widetilde{\tau}_n' = 1 - \frac{1-\tau_n'}{\widetilde{\gamma}_{\text{ML}}^{-1} - 1} = 1 - \frac{p_n}{\widetilde{\gamma}_{\text{ML}}^{-1} - 1},$$

$$\widetilde{\xi}^{M,\star}_{\tau_n'} := \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widetilde{\gamma}_{\text{ML}}} \widehat{\xi}^{M}_{\tau_n} \quad \text{and} \quad \widetilde{\xi}^{Q,\star}_{\tau_n'} := \widetilde{q}^{\star}_{\tau_n'}\,\mathcal{G}\left(\widetilde{\gamma}_{\text{ML}}\right).$$

The plots of the alternative risk estimates $\widetilde{q}^{\star}_{\tau_n'}$, $\widetilde{e}^{\star}_{\widetilde{\tau}_n'}$, $\widetilde{\xi}^{M,\star}_{\tau_n'}$ and $\widetilde{\xi}^{Q,\star}_{\tau_n'}$ are graphed in Figure 2.14 (c). These plots seem to be more volatile than those obtained via the Weissman extrapolation method in

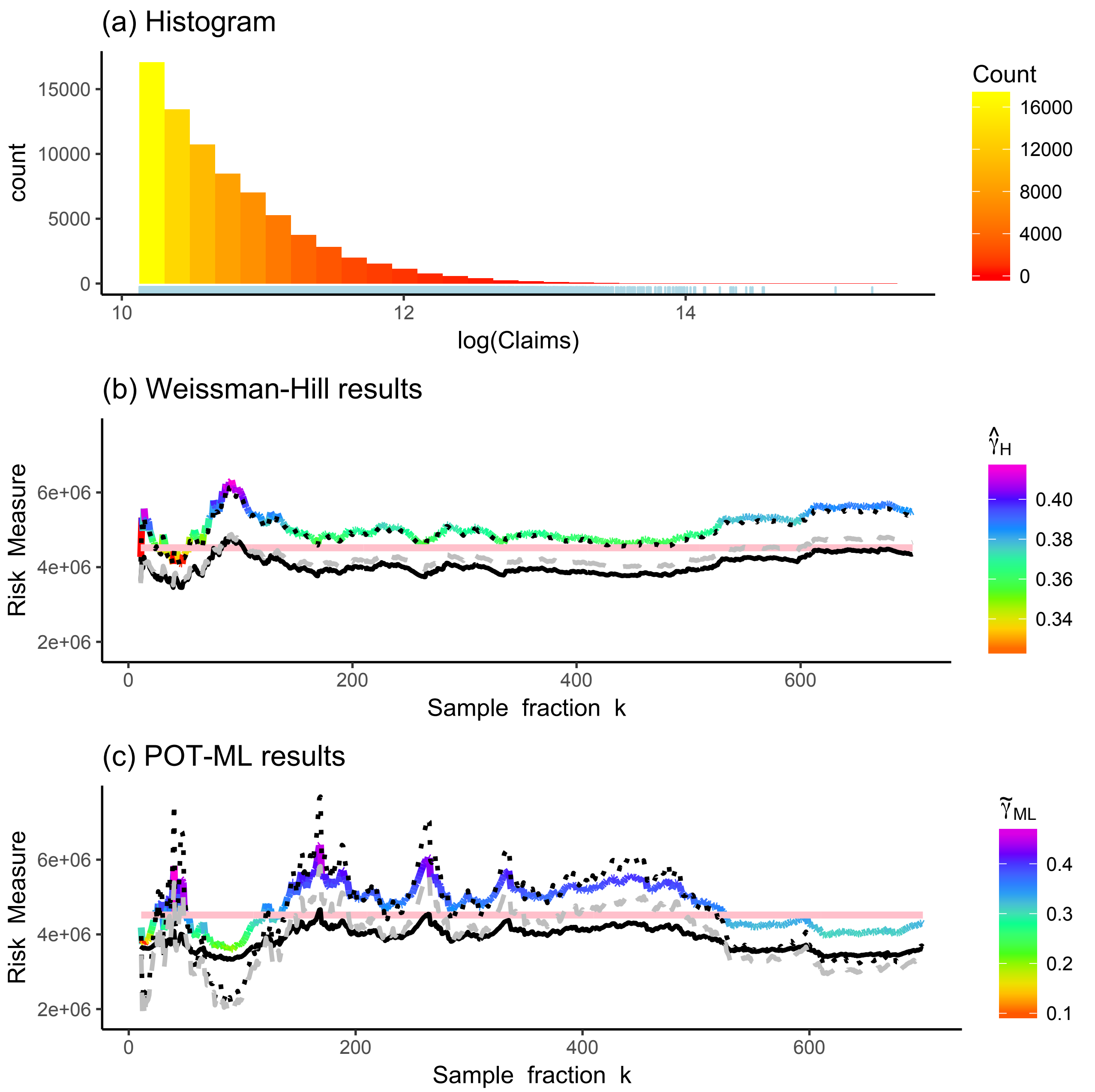

Figure 2.14: SOA Group Medical Insurance data. (a) Histogram and scatterplot of the log-claim amounts. (b) Using the Weissman extrapolation method and the Hill estimator: the extremile plots $k \mapsto \widehat{\xi}^{Q,\star}_{\tau'_n}$ (rainbow) and $k \mapsto \widehat{\xi}^{M,\star}_{\tau'_n}$ (dotted black), along with the quantile plot $k \mapsto \widehat{q}^{\star}_{\tau'_n}$ (solid black), the expectile plot $k \mapsto \widehat{e}^{\star}_{\widehat{\tau}'_n}$ (dashed grey) and the sample maximum $Y_{n,n}$ (pink line). (c) Using the POT approach and the maximum likelihood estimator of γ: the extremile plots $k \mapsto \widetilde{\xi}^{Q,\star}_{\tau'_n}$ (rainbow) and $k \mapsto \widetilde{\xi}^{M,\star}_{\tau'_n}$ (dotted black), along with the quantile plot $k \mapsto \widetilde{q}^{\star}_{\tau'_n}$ (solid black), the expectile plot $k \mapsto \widetilde{e}^{\star}_{\widetilde{\tau}'_n}$ (dashed grey) and the maximum $Y_{n,n}$ (pink line).

Figure 2.14 (b). A stable region appears, however, for k from 200 up to 500, leading to the averaged estimates $\widetilde{q}^{\star}_{\tau'_n} = 4.10$ million, $\widetilde{e}^{\star}_{\widetilde{\tau}'_n} = 4.52$ million, $\widetilde{\xi}^{M,\star}_{\tau'_n} = 5.32$ million and $\widetilde{\xi}^{Q,\star}_{\tau'_n} = 5.17$ million. These POT and ML-based risk values are larger and hence more conservative than their Weissman and Hill-based analogues $\widehat{q}^{\star}_{\tau'_n} = 3.90$ million, $\widehat{e}^{\star}_{\widehat{\tau}'_n} = 4.13$ million, $\widehat{\xi}^{M,\star}_{\tau'_n} = 4.78$ million and $\widehat{\xi}^{Q,\star}_{\tau'_n} = 4.83$ million. The most important difference is 0.54 million, between the two extremile estimators $\widetilde{\xi}^{M,\star}_{\tau'_n}$ and $\widehat{\xi}^{M,\star}_{\tau'_n}$, followed by a difference of 0.39 million between the expectile estimators $\widetilde{e}^{\star}_{\widetilde{\tau}'_n}$ and $\widehat{e}^{\star}_{\widehat{\tau}'_n}$. Taking a closer look to the construction of these competing asymmetric least squares estimators, we see that their substantial deviation is mainly due to the use of the Hill estimator $\widehat{\gamma}_H$ in the Weissman extrapolation method and the ML estimator $\widetilde{\gamma}_{\mathrm{ML}}$ in the POT method. The volatility of the POT plots can thus be explained, on the one hand, by an averaged estimate $\widetilde{\gamma}_{\mathrm{ML}} = 0.38$ slightly higher than $\widehat{\gamma}_H = 0.36$ (these estimates were averaged over $k \in [200, 500]$ and $k \in [150, 500]$, respectively), and most importantly, on the other hand, by ML estimates of γ being far more volatile than their Hill counterparts as functions of k. The advantageous stability of $\widehat{\gamma}_H$ relative to $\widetilde{\gamma}_{\mathrm{ML}}$ becomes clear by comparing the colour-schemes in Figures 2.14 (b) and (c).

2.5.6 Perspectives for future research

Forecast evaluation and validation We showed that extremiles define a law-invariant, coherent, and comonotonically additive risk measure. It follows that, for $\tau \neq 1/2$, they are not elicitable [191, 265], contrary to expectiles (which are not comonotonically additive). Elicitability is a key part of the appeal of expectiles to actuarial and financial risk managers, because it corresponds to the existence of a natural forecast verification and comparison methodology. Yet, it is known that certain spectral risk measures are jointly elicitable with quantiles (see [120]). Although this result in its current form focuses on finite convex combinations of Expected Shortfall at multiple levels and therefore does not apply to extremiles, it is reasonable to investigate whether a similar result dedicated to this class can be shown. This would then pave the way for a validation procedure of extremile forecasts.

Extreme value inference outside the class of heavy-tailed distributions Our results on extremile estimation at high levels are, like the results we obtained for expectiles in Section 2.3, derived under the assumption of a heavy right tail, motivated by applications in insurance and finance. Unlike the situation for expectiles, an extension of our results to the Weibull and Gumbel domains of attraction has an interest in itself, since extremiles benefit from several interpretations, in particular in terms of expected maxima and minima, and are therefore likely to appeal to risk analysts in fields such as environmental science, where distributions from the Weibull and Gumbel domains of attraction either naturally appear or are used as modelling tools [77, 124, 246]. A crucial first step before developing an inference method for high extremiles adapted to such setups will be to generalise the asymptotic expansion in Proposition 2.5.5 outside of the class of heavy-tailed distributions.

Adaptation to the regression context Like quantiles and expectiles, extremiles can be formulated as the minimiser of a convex criterion. For quantiles and expectiles, this is most useful in regression contexts; as a matter of fact, expectiles were introduced by [209] specifically in a regression context. The extension of extremile estimation to a regression context is therefore a very natural step. Already the implications of Proposition 2.5.2 *(iii)* and *(iv)* on regression extremiles are clear. For example, conditional extremile curves will be parallel to each other if the conditional distributions of the response are homogeneous. Also, for any τ, the lower and upper τth conditional extremile curves will be symmetric about the mean curve if the conditional distributions of the response are symmetric. From an inferential point of view, the closed form expression of extremiles as weighted averages of the quantiles suggests that regression extremiles may be estimated by an integral of the generalised inverse of a nonparametric kernel estimator of the conditional distribution function; the functional convergence of the latter has been extensively investigated, see e.g. [158]. At extreme levels, the connection between extremiles and quantiles at high levels through an asymptotic proportionality constant only depending on the tail index hints at the possibility of constructing extreme conditional extremile estimators by combining conditional tail index estimators (such as the averaged local Hill estimator in [24]) and extreme conditional quantile estimators (such as the one of [85]).

2.6 Towards multivariate risk management: technical tools on the joint behaviour of extreme value estimators for heavy tails [4]

2.6.1 Introduction

All the risk measures we have worked on so far in Chapter 2 (Wang DRMs, expectiles, L^p-quantiles, tail L^p-medians, extremiles) are, in heavy-tailed models, asymptotically proportional to tail quantiles. As such, their estimators at extreme levels can be extrapolated from empirical estimators at intermediate levels via straightforward adaptations of Weissman's method by relying on tail index estimators. Tail index estimation is therefore a central question in the statistical analysis of extreme risk measures, as it is in the development of inference procedures for the extremes of heavy-tailed distributions in general. The most popular and well-known tail index estimator in heavy-tailed models is arguably the Hill estimator [153], which we have worked with several times throughout this report. Since this estimator is central to the motivation and writing of this section, we recall once again its expression: assuming that $X_1, \ldots, X_n$ is a random sample of copies of a variable X having a heavy-tailed distribution with tail index γ, the Hill estimator of γ is

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

Here, $k = k(n)$ is such that $k \to \infty$ and $k/n \to 0$, and $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$ denote the order statistics related to $X_1, \ldots, X_n$. It is well-known that, under an appropriate second-order condition on X controlling the gap between the underlying distribution and the Pareto distribution, this estimator is $\sqrt{k}-$asymptotically normal. A variety of proofs of this result are available, including arguments based on a representation of log-spacings in terms of independent exponential random variables (see Section 4.4 in [47]), Rényi's representation of order statistics (see pp.74–75 in [145]) or the tail empirical process (see Proposition 9.3 p.302 in [215]). A common aim of these approaches is to deal with the structure of the Hill estimator in terms of top order statistics, which are significantly harder to handle than the original $X_1, \ldots, X_n$. Although this theoretical difficulty is now well-understood, it is further compounded if one wants to show the joint asymptotic normality of marginal Hill estimators for a random vector with heavy-tailed marginal distributions, since one then also needs to take into account the asymptotic dependence structure within the underlying multivariate distribution. Such joint convergence results have only been proven very recently by [94, 156, 177]. The methods of proof therein rely on advanced theoretical methodologies, namely multivariate vague convergence in [94, 177] and multivariate empirical process theory in [156], as well as on various *ad hoc* technical conditions. These joint asymptotic results have been found useful for testing the tail homogeneity assumption across marginals of a multivariate heavy-tailed distribution in environmental or financial contexts [156, 177] and constructing improved tail index estimators by pooling [94].

The Hill estimator can be written in several different ways. When there are no ties in the sample $X_1, \ldots, X_n$ (this is for example the case when the distribution function of X is continuous), one has $k = \sum_{i=1}^{n} \mathbb{I}\{X_i > X_{n-k,n}\}$. Noting then that

$$\widehat{\gamma}(k) = \frac{\sum_{i=1}^{n} [\log X_i - \log X_{n-k,n}] \mathbb{I}\{X_i > X_{n-k,n}\}}{\sum_{i=1}^{n} \mathbb{I}\{X_i > X_{n-k,n}\}}$$

suggests that a major difficulty in the theoretical analysis of the Hill estimator lies in the fact that the high threshold $X_{n-k,n}$, used to guarantee the consistency of the estimator by retaining only the high values in the sample, is *random*. Indeed, if we could, in the asymptotic analysis of $\widehat{\gamma}(k)$, replace the random quantity $X_{n-k,n}$, which is nothing but the empirical quantile at level $1-k/n$, by the unknown but *non-random* population quantile $q(1-k/n)$, we would find the pseudo-estimator

$$\widetilde{\gamma}(k) = \frac{\sum_{i=1}^{n} [\log X_i - \log q(1-k/n)] \mathbb{I}\{X_i > q(1-k/n)\}}{\sum_{i=1}^{n} \mathbb{I}\{X_i > q(1-k/n)\}}$$

which is conceptually far easier to analyse than $\widehat{\gamma}(k)$: for independent $X_1, \ldots, X_n$, this pseudo-estimator is a ratio of sums of independent variables constructed on the X_i, and thus can easily be handled by a combination of Lyapunov's central limit theorem and the Cramér-Wold device. The striking and perhaps unexpected point here is that the asymptotic distributions of $\widehat{\gamma}(k)$ and $\widetilde{\gamma}(k)$ are identical, as can be seen by comparing Theorem 3.2.5 in [145] and Theorem 4.3.1 in [136]. On this basis, one may therefore ask whether a relationship of the form

$$\widehat{\gamma}(k) = \widetilde{\gamma}(k) + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right) \tag{2.6.1}$$

can be shown. Such a relationship has not, to the best of our knowledge, been proven so far in the literature. Its validity is not obvious either, since $\widetilde{\gamma}(k)$ is obtained from $\widehat{\gamma}(k)$ by replacing $X_{n-k,n}$ with $q(1-k/n)$, and we know that $X_{n-k,n}/q(1-k/n) - 1$ only converges to 0 at the rate $\sqrt{k}$ (see Theorem 2.4.1 in [145]), which is the common rate of convergence of $\widetilde{\gamma}(k)$ and $\widehat{\gamma}(k)$. Let us also point out straightaway that although Equation (2.6.1) would give an additional proof of the convergence of the Hill estimator $\widehat{\gamma}(k)$, this is not where its full value lies, since the asymptotic properties of the Hill estimator are well-known. It really becomes useful when, for instance, analysing the joint convergence of marginal Hill estimators since, in contrast to the joint convergence of the randomly thresholded versions, the joint convergence of the non-randomly thresholded versions is easy to obtain (under an appropriate upper tail dependence condition) by a combination of standard central limit theory and the Cramér-Wold device.

The proof of Equation (2.6.1) is the motivation for the paper [4], whose contribution we present in this section. More precisely, the objective of the paper is to embed the Hill estimator in a wider class of average excesses and then provide a simple representation of those empirical average excesses above a high random threshold in terms of their pseudo-estimator versions with a non-random threshold. The class of average excesses we consider includes the Hill estimator, and can be modified in a very simple way to encompass Conditional Tail Moments (CTMs). We show in particular how the results of [94], on the joint asymptotic normality of marginal Hill estimators for a random vector with heavy-tailed marginal distributions, can be recovered and generalised under weaker assumptions and by elementary techniques. We shall then highlight a couple of applications of this joint asymptotic normality result, including to the obtention of the asymptotic properties of the tail index estimator introduced by [22] when the variable of interest is randomly right-truncated. The question of the convergence of this estimator was considered first by [22] under restrictive assumptions, and then by [54] using a delicate theoretical argument based on the weighted tail copula process and a joint tail assumption on the observed pair. Our results will make it possible to unify and extend the results of [22, 54], without either resorting to the former's advanced methodology and joint dependence condition or to the latter's technical conditions. In doing so, we will also be able to give a very simple expression of the asymptotic variance of the limiting normal distribution. Finally, and motivated by the ideas of [156], we shall apply our results to find joint convergence results on empirical CTMs that may be of independent interest, for instance in testing whether certain tail moments of two asymptotically dependent variables are equal. We will highlight how such results nicely complement results of [14, 114].

The outline of this section is the following. Our framework and main results are stated in Section 2.6.2. Applications of our results, to the joint convergence of Hill and CTM estimators and to the convergence of a tail index estimator under random right-truncation, are presented in Section 2.6.3. Section 2.6.4 concludes by discussing perspectives for future work.

2.6.2 Framework and main results

We assume in this section that the data is made of independent copies $X_1, \ldots, X_n$ of a random variable X. We assume for ease of presentation that **the distribution function F is continuous**, so that, with probability 1, there are no ties in the sample $X_1, \ldots, X_n$. This is not restrictive for our purposes, as our applications of the main results presented here focus on the obtention of joint convergence results for several estimators of extreme value indicators pertaining to a multivariate random vector with heavy-tailed marginal distributions. Such an endeavour indeed typically requires the standardisation of the marginal distributions to uniform distributions (made convenient by the continuity assumption), in order to obtain a meaningful description of the relevant extremal dependence structure. Our results can still be shown, at the expense of extra technical details, if F is only continuous in a neighbourhood of infinity.

Let us now define the central concepts we will focus on in this section. Again, let us recall our motivation, which is to link the Hill estimator $\widehat{\gamma}(k)$ to its pseudo-estimator version $\widetilde{\gamma}(k)$. One interpretation of the Hill estimator $\widehat{\gamma}(k)$ is that it is a sample version of the average log-excess $\mathbb{E}(\log X - \log U(n/k) \,|\, X > U(n/k))$ (see p.104 of [47], and p.69 of [145]); it should also be clear that $\widetilde{\gamma}(k)$ is a pseudo-estimator of this average log-excess. This motivates us to carry forward the notion of average excess with the following definition.

Definition 2.6.1. Let X be a heavy-tailed random variable and f be a continuous function on a neighbourhood of infinity such that for some t_0, the quantity $\mathbb{E}(|f(X)| \,|\, X > t_0)$ is well-defined and finite. For any $t \geq t_0$, we define the *average f-excess of X above level t* to be

$$\mathrm{AE}_f(t) := \mathbb{E}(f(X) - f(t) \,|\, X > t) = \frac{\mathbb{E}([f(X) - f(t)]\mathbb{I}\{X > t\})}{\overline{F}(t)}$$

and the *empirical average f-excess of X above level t* to be

$$\widehat{\mathrm{AE}}_f(t) := \frac{n^{-1}\sum_{i=1}^n [f(X_i) - f(t)]\mathbb{I}\{X_i > t\}}{\widehat{\overline{F}}_n(t)}, \quad \text{where} \quad \widehat{\overline{F}}_n(t) := \frac{1}{n}\sum_{i=1}^n \mathbb{I}\{X_i > t\}.$$

We also define the *expected f-shortfall above level t* as

$$\mathrm{ES}_f(t) := \mathbb{E}(f(X) \,|\, X > t) = \mathrm{AE}_f(t) + f(t)$$

and the *empirical expected f-shortfall above level t* as

$$\widehat{\mathrm{ES}}_f(t) := \frac{n^{-1}\sum_{i=1}^n f(X_i)\mathbb{I}\{X_i > t\}}{\widehat{\overline{F}}_n(t)} = \widehat{\mathrm{AE}}_f(t) + f(t).$$

With this definition, letting $f = \log$, we find

$$\begin{aligned} \widehat{\mathrm{AE}}_{\log}(X_{n-k,n}) &= \widehat{\gamma}(k), \quad \widehat{\mathrm{AE}}_{\log}(U(n/k)) = \widetilde{\gamma}(k) \\ \text{and} \quad \mathrm{AE}_{\log}(U(n/k)) &= \mathbb{E}(\log X - \log U(n/k) \,|\, X > U(n/k)). \end{aligned}$$

The other main example allowed by Definition 2.6.1 which we will consider in this section is obtained by choosing $f = f_a : x \mapsto x^a$ for some $a > 0$. We then get, for any $t > 0$,

$$\mathrm{ES}_{f_a}(t) = \mathbb{E}(X^a \,|\, X > t) =: \mathrm{CTM}_a(t),$$

where CTM_a is the Conditional Tail Moment of order a introduced in [114]. This definition makes sense for any $a < 1/\gamma$, since all conditional tail moments of X of order smaller than $1/\gamma$ are finite (a rigorous statement is Exercise 1.16 in [145]). An empirical counterpart of $\mathrm{CTM}_a(U(n/k)) = \mathrm{ES}_{f_a}(U(n/k))$ is then

$$\frac{1}{k}\sum_{i=1}^k X_{n-i+1,n}^a = \widehat{\mathrm{ES}}_{f_a}(X_{n-k,n}) = \widehat{\mathrm{AE}}_{f_a}(X_{n-k,n}) + X_{n-k,n}^a.$$

Common features of the above two examples are that they crucially hinge on the notion of average $f-$excess, and most importantly that the derivative of the function f involved in their construction is a power function. This observation provides the motivation for our first main result, which provides an asymptotic relationship between $\widehat{\mathrm{AE}}_f(X_{n-k,n})$ and $\widehat{\mathrm{AE}}_f(U(n/k))$ when f' is regularly varying at infinity.

Theorem 2.6.1. *Suppose that X satisfies condition $\mathcal{C}_2(\gamma, \rho, A)$. Let $k = k(n) \to \infty$ be such that $k/n \to 0$ and $\sqrt{k}A(n/k) = \mathrm{O}(1)$. Assume finally that f is continuously differentiable in a neighbourhood of infinity, ultimately increasing, and that f' is regularly varying with index $a-1$, where $0 \leq 2a\gamma < 1$. Then we have:*

$$\frac{\widehat{\mathrm{AE}}_f(X_{n-k,n})}{\mathrm{AE}_f(U(n/k))} - 1 = \left(\frac{\widehat{\mathrm{AE}}_f(U(n/k))}{\mathrm{AE}_f(U(n/k))} - 1\right) + a\gamma\left(\frac{\widehat{\overline{F}}_n(U(n/k))}{\overline{F}(U(n/k))} - 1\right) + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right).$$

In particular

$$\sqrt{k}\left(\frac{\widehat{\mathrm{AE}}_f(X_{n-k,n})}{\mathrm{AE}_f(U(n/k))} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{1-2a\gamma} + a^2\gamma^2\right).$$

Theorem 2.6.1 provides an asymptotic representation of the empirical average $f-$excess $\widehat{\mathrm{AE}}_f(X_{n-k,n})$, whose expression hinges on top order statistics, in terms of sums of independent and identically distributed random variables constructed on the X_i in a simple way. Note that, due to the continuity of F, we have $\overline{F}(U(n/k)) = k/n$; writing $\overline{F}(U(n/k))$ instead of k/n, as we will do in this section, is to emphasise that the second term on the right-hand side is centred. Assumptions $k = k(n) \to \infty$, $k/n \to 0$ and $\sqrt{k}A(n/k) = \mathrm{O}(1)$ are standard for the asymptotic analysis of extreme value estimators; the assumption $0 \leq 2a\gamma < 1$, meanwhile, ensures that $\widehat{\mathrm{AE}}_f(U(n/k))$ is $\sqrt{k}-$asymptotically normal, and therefore that Theorem 2.6.1 provides an expression of $\widehat{\mathrm{AE}}_f(X_{n-k,n})$ in terms of $\widehat{\mathrm{AE}}_f(U(n/k))$ and $\widehat{\overline{F}}_n(U(n/k))$ that is meaningful for the asymptotic analysis.

In particular, setting $f = \log$ (and therefore $a = 0$) in Theorem 2.6.1 allows us to show that the motivating representation (2.6.1), of the Hill estimator in terms of sums of independent and identically distributed random variables, holds indeed. This is the focus of the following corollary.

Corollary 2.6.2. *Suppose that X satisfies condition $\mathcal{C}_2(\gamma, \rho, A)$. Let $k = k(n) \to \infty$ be such that $k/n \to 0$ and $\sqrt{k}A(n/k) = \mathrm{O}(1)$. Then we have*

$$\widehat{\gamma}(k) = \widetilde{\gamma}(k) + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right)$$

where

$$\widehat{\gamma}(k) = \frac{1}{k}\sum_{i=1}^{k} \log\frac{X_{n-i+1,n}}{X_{n-k,n}} \text{ and } \widetilde{\gamma}(k) = \frac{\sum_{i=1}^n [\log X_i - \log U(n/k)]\mathbb{I}\{X_i > U(n/k)\}}{\sum_{i=1}^n \mathbb{I}\{X_i > U(n/k)\}}.$$

Corollary 2.6.2 can of course be used to provide yet another proof of the asymptotic normality of the Hill estimator, via the Lyapunov central limit theorem and the Cramér-Wold device applied to get the $\sqrt{k}-$asymptotic normality of $\widetilde{\gamma}(k)$. This is not, however, where the value of Corollary 2.6.2 lies, if only because its proof uses an approximation of the tail empirical process $x \mapsto \widehat{\overline{F}}_n(xU(n/k))$ by a Gaussian process, which can itself be used to provide a direct proof of the asymptotic normality of the Hill estimator (see pp.162–163 of [145]). A much more relevant impact of Corollary 2.6.2 lies in its potential for the analysis of the joint convergence of several Hill estimators. For instance, if $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ are independent copies of random variables X and Y, which are both heavy-tailed and satisfy a second-order condition, then we have under suitable assumptions on k that:

$$\begin{pmatrix} \widehat{\gamma}_X(k) \\ \widehat{\gamma}_Y(k) \end{pmatrix} = \begin{pmatrix} \widetilde{\gamma}_X(k) \\ \widetilde{\gamma}_Y(k) \end{pmatrix} + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right)$$

with obvious notation. The benefit of writing this is that while showing directly the joint convergence of the random pair on the left-hand side is difficult and appears to require advanced theoretical arguments (see [94, 156]), the convergence of the right-hand side is much easier to obtain since it is nothing but a pair of (ratios of) sums of independent and identically distributed random variables. We will return to this in Section 2.6.3.1 to show how this observation leads to conceptually simple proofs of the joint asymptotic normality of several Hill estimators.

Theorem 2.6.1 is somewhat tedious to apply if the focus is on an average shortfall, such as in the case of a CTM. The following theorem provides an analogue of Theorem 2.6.1 specifically dedicated to the analysis of $\widehat{\mathrm{ES}}_f(X_{n-k,n})$, under a slightly stronger condition on f.

Theorem 2.6.3. *Work under the conditions of Theorem 2.6.1, under the additional assumption that $xf'(x)/f(x) \to a > 0$ as $x \to \infty$. Then:*

$$\frac{\widehat{\mathrm{ES}}_f(X_{n-k,n})}{\mathrm{ES}_f(U(n/k))} - 1 = \left(\frac{\widehat{\mathrm{ES}}_f(U(n/k))}{\mathrm{ES}_f(U(n/k))} - 1\right) + a\gamma\left(\frac{\widehat{\overline{F}}_n(U(n/k))}{\overline{F}(U(n/k))} - 1\right) + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right).$$

In particular

$$\sqrt{k}\left(\frac{\widehat{\mathrm{ES}}_f(X_{n-k,n})}{\mathrm{ES}_f(U(n/k))} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{2a^2\gamma^2(1-a\gamma)}{1-2a\gamma}\right).$$

Setting $f = f_a$, we find back the asymptotic variance of the empirical estimator $\widehat{\mathrm{ES}}_{f_a}(X_{n-k,n})$ of $\mathrm{CTM}_a(U(n/k))$; see Theorem 1 of [114] in the regression case. The condition $2a\gamma < 1$, part of Theorems 2.6.1 and 2.6.3, also naturally appears in the asymptotic normality results of [114] and ensures in particular that the conditional tail variance of X^a is finite. Again, the main value of the above result does not lie in the fact that it gives the asymptotic distribution of the empirical average expected shortfall, but rather in that it allows one to establish the joint convergence of several of those estimators with no conceptual difficulty. With an eye on the latter, we state the following corollary of Theorem 2.6.3, which is easier to use in practice.

Corollary 2.6.4. *Work under the conditions of Theorem 2.6.3. Then:*

$$\frac{\widehat{\mathrm{ES}}_f(X_{n-k,n})}{\mathrm{ES}_f(U(n/k))} - 1 = a\gamma\left(\frac{n^{-1}\sum_{i=1}^n [f(X_i) - f(U(n/k))]\mathbb{I}\{X_i > U(n/k)\}}{\mathbb{E}([f(X) - f(U(n/k))]\mathbb{I}\{X > U(n/k)\})} - 1\right) + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right).$$

Let us point out that Corollary 2.6.4 gives an asymptotic representation of $\widehat{\mathrm{ES}}_f(X_{n-k,n})$ as a single sum of independent, identically distributed and centred random variables. Its use does not even involve any linearisation (unlike Theorems 2.6.1 and 2.6.3), making it particularly simple to apply.

Our objective in the rest of this section is to show how the main theoretical results of this section can be applied to finding a solution to two theoretical questions: the joint convergence of marginal Hill estimators, and the joint convergence of marginal Conditional Tail Moments. We shall also explore how our result on the joint convergence of Hill estimators can be fruitfully applied to solve the question of the convergence of a specific tail index estimator, introduced by [22] to tackle the case when the heavy-tailed variable of interest is randomly right-truncated. Before that, we conclude this section on a generalisation of Theorem 2.6.1 and Corollary 2.6.2 to the case when the order statistic $X_{n-k,n}$ is replaced by an *arbitrary* $\sqrt{k}-$consistent estimator $\widehat{U}(n/k)$ of the quantile $U(n/k)$. In this context, $\widehat{\mathrm{AE}}_f(\widehat{U}(n/k))$ cannot be represented by sums of independent, identically distributed and centred random variables anymore, but interestingly Corollary 2.6.2 still stands.

Theorem 2.6.5. *Work under the conditions of Theorem 2.6.1. Then we have:*

$$\frac{\widehat{\mathrm{AE}}_f(\widehat{U}(n/k))}{\mathrm{AE}_f(U(n/k))} - 1 = \left(\frac{\widehat{\mathrm{AE}}_f(U(n/k))}{\mathrm{AE}_f(U(n/k))} - 1\right) + a\gamma\left(\frac{\widehat{\overline{F}}_n(U(n/k))}{\widehat{\overline{F}}_n(\widehat{U}(n/k))} - 1\right) + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right)$$

for any estimator $\widehat{U}(n/k)$ of $U(n/k)$ such that $\widehat{U}(n/k)/U(n/k) - 1 = \mathrm{O}_{\mathbb{P}}(1/\sqrt{k})$.

Corollary 2.6.6. *Suppose that X satisfies condition $\mathcal{C}_2(\gamma, \rho, A)$. Let $k = k(n) \to \infty$ be such that $k/n \to 0$ and $\sqrt{k}A(n/k) = \mathrm{O}(1)$. Assume finally that $\widehat{U}(n/k)$ is an estimator of $U(n/k)$ such that $\widehat{U}(n/k)/U(n/k) - 1 = \mathrm{O}_{\mathbb{P}}(1/\sqrt{k})$. Then we have*

$$\overline{\gamma}(k) = \widetilde{\gamma}(k) + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right)$$

where

$$\overline{\gamma}(k) = \frac{\sum_{i=1}^n [\log X_i - \log \widehat{U}(n/k)]\mathbb{I}\{X_i > \widehat{U}(n/k)\}}{\sum_{i=1}^n \mathbb{I}\{X_i > \widehat{U}(n/k)\}}$$

$$\textit{and}\ \ \widetilde{\gamma}(k) = \frac{\sum_{i=1}^n [\log X_i - \log U(n/k)]\mathbb{I}\{X_i > U(n/k)\}}{\sum_{i=1}^n \mathbb{I}\{X_i > U(n/k)\}}.$$

We observe that Theorem 2.6.5 and Corollary 2.6.6 are indeed generalisations of Theorem 2.6.1 and Corollary 2.6.2: when $\widehat{U}(n/k) = X_{n-k,n}$, one has $\widehat{\overline{F}}_n(\widehat{U}(n/k)) = \widehat{\overline{F}}_n(X_{n-k,n}) = k/n = \overline{F}(U(n/k))$ by continuity of F and $\overline{\gamma}(k) = \widehat{\gamma}(k)$.

2.6.3 Applications

2.6.3.1 Joint convergence of marginal Hill estimators

Let $(\mathbf{X}_i)_{1\leq i\leq n}$, with $\mathbf{X}_i = (X_i^{(1)}, X_i^{(2)}, \ldots, X_i^{(d)})$, be a sample of independent copies of a random vector $\mathbf{X} = (X^{(1)}, X^{(2)}, \ldots, X^{(d)})$ such that each component $X^{(j)}$ has a continuous distribution function F_j and satisfies condition $\mathcal{C}_2(\gamma_j, \rho_j, A_j)$. Our first goal is to establish a joint convergence result for the Hill estimators of the γ_j built on this sample, that is:

$$\widehat{\gamma}_j(k_j) = \frac{1}{k_j}\sum_{i=1}^{k_j} \log(X_{n-i+1,n}^{(j)}) - \log(X_{n-k_j,n}^{(j)})$$

where $k_j = k_j(n) \to \infty$, with $k_j/n \to 0$. This theoretical question is addressed in [94] and further discussed in [177] under the assumption $\gamma_j = \gamma$ for all $j \in \{1, \ldots, d\}$.

Since each $\widehat{\gamma}_j(k_j)$ is built on top order statistics from the corresponding $X^{(j)}$, our objective of analysing the joint convergence of the $\widehat{\gamma}_j(k_j)$ calls for some sort of extremal dependence assumption between the $X^{(j)}$. We shall work under the following condition on the pairwise upper tail dependence between any two components $X^{(j)}$ and $X^{(l)}$:

$\boldsymbol{\mathcal{J}(R)}$ For any (j,l) with $1 \leq j < l \leq d$, there is a function $R_{j,l}$ on $[0,\infty]^2 \setminus \{(\infty,\infty)\}$ such that we have the convergence

$$\lim_{t\to\infty} t\,\mathbb{P}\left(\overline{F}_j(X^{(j)}) \leq \frac{x_j}{t},\ \overline{F}_l(X^{(l)}) \leq \frac{x_l}{t}\right) = R_{j,l}(x_j, x_l)$$

for any $(x_j, x_l) \in [0,\infty]^2 \setminus \{(\infty,\infty)\}$.

This condition appears, among others, in [68] as well as, in a slightly different form, in [156], and is a generalisation of condition $\mathcal{JC}(R)$ in Section 2.3.4.2. It is a convenient way to describe the asymptotic dependence structure of the bivariate random vector $(X^{(j)}, X^{(l)})$, while being weaker than a pairwise bivariate regular variation assumption (in the sense of e.g. [214]) and a fortiori weaker than a multivariate regular variation assumption on the random vector $\mathbf{X}$ such as the one of [94]. The function $R_{j,l}$ is sometimes called the tail copula of $(X^{(j)}, X^{(l)})$ (see [224]).

Second-order regular variation of each marginal distribution and the above pairwise upper tail dependence assumption turn out to be sufficient to obtain the joint convergence of the $\widehat{\gamma}_j(k_j)$, as the next result shows.

Theorem 2.6.7. *Assume that, for each $1 \leq j \leq d$, $X^{(j)}$ satisfies condition $\mathcal{C}_2(\gamma_j, \rho_j, A_j)$, and suppose that condition $\boldsymbol{\mathcal{J}(R)}$ holds. Let, for $1 \leq j \leq d$, $k_j = k_j(n) \to \infty$ be such that:*

- *$k_j/n \to 0$ and $\sqrt{k_j}A_j(n/k_j) \to \lambda_j \in \mathbb{R}$;*
- *$k_1/k_j \to c_j \in (0,\infty)$ (with then $c_1 = 1$).*

Set $\mathbf{c} = (c_1, c_2, \dots, c_d)$*. Then we have:*

$$\sqrt{k_1}\left(\widehat{\gamma}_j(k_j) - \gamma_j\right)_{1\le j\le d} \xrightarrow{d} \mathcal{N}_d\left(\mathbf{b}(\mathbf{c}),\, \mathbf{V}(\mathbf{c})\right)$$

with

$$\mathbf{b}_j(\mathbf{c}) = \sqrt{c_j}\frac{\lambda_j}{1-\rho_j} \quad \textit{and} \quad \mathbf{V}_{j,l}(\mathbf{c}) = \begin{cases} c_j\gamma_j^2 & \textit{if } j = l, \\ \gamma_j\gamma_l R_{j,l}(c_l, c_j) & \textit{if } j < l. \end{cases}$$

Let us briefly highlight that the structure of the asymptotic bias component $\mathbf{b}(\mathbf{c})$ is quite strongly constrained by the values of the second-order parameters ρ_j: for all its components to be non-zero (or equivalently, for all the constants λ_j to be non-zero), all the second-order parameters ρ_j should be equal. This is due to the fact that the k_j are proportional and each function $|A_j|$ is regularly varying with index ρ_j. For the same reason, if $\rho^* = \max_{1\le j\le d}\rho_j$, then $\mathbf{b}_j(\mathbf{c}) = 0$ whenever j is such that $\rho_j < \rho^*$.

Our Theorem 2.6.7 contains Theorem 3.3 and Corollary 3.4 in [94], which are stated under more restrictive conditions, including equality of all tail indices γ_j, a multivariate regular variation assumption on $\mathbf{X}$, von Mises conditions on each marginal distribution of $\mathbf{X}$, and a uniform analogue of condition $\boldsymbol{\mathcal{J}(R)}$ [note also that Corollary 3.4 in [94] should, with their notation, read $\Gamma_{i,j} = c_ic_j\nu_{i,j}(c_i^{1/\alpha}, c_j^{1/\alpha})$ in the case $i < j$; as stated, their covariance matrix may fail to be positive semi-definite, for instance for $d = 2$ and small c_2. Compare with Proposition 1 in [177]]. Theorem 2.6.7 also contains Proposition 1 in [168], which is limited to the case $d = 2$ and $X^{(2)} = -X^{(1)}$. Another result related to Theorem 2.6.7 is Proposition 3 in [156], although the present result and that of [156] are more difficult to compare since the latter is stated within the particular context of time series analysis. Our proof of Theorem 2.6.7 is also conceptually less involved than that of [94], which rests upon a multivariate functional central limit theorem (see Theorem 7.1 and Corollary 7.2 therein). Finally, and without taking the time series framework into account, the result of [156] is based on delicate arguments involving a *multivariate*, joint Skorokhod construction of Gaussian approximations for marginal tail empirical processes. Our proof, meanwhile, rests on the standard Lyapunov central limit theorem and Corollary 2.6.2, whose proof is based only on a *univariate* Gaussian approximation of the tail empirical process. It should nonetheless be made clear once again that Proposition 3 in [156] holds in a framework of $\beta-$mixing time series, and as such is not restricted to independent and identically distributed observations, unlike our Theorem 2.6.7.

Results such as Theorem 2.6.7 may be applied to define a test of tail homogeneity, that is, equality of tail indices across marginals. If there is no evidence to reject this assumption, one may then define improved estimators of the common value of the tail index by pooling together the marginal Hill estimators. These ideas are considered in [94, 177]. To illustrate how our results can be applied to such problems, we state a corollary of Theorem 2.6.7 in the case $d = 2$ of a bivariate distribution with heavy-tailed marginals.

Corollary 2.6.8. *Suppose that X and Y have continuous distribution functions F_X and F_Y, which satisfy conditions $\mathcal{C}_2(\gamma_X, \rho_X, A_X)$ and $\mathcal{C}_2(\gamma_Y, \rho_Y, A_Y)$, respectively. Assume that there is a function R on $[0,\infty]^2 \setminus \{(\infty,\infty)\}$ such that we have the convergence*

$$\lim_{t\to\infty} t\,\mathbb{P}\left(\overline{F}_X(X) \le \frac{x}{t},\ \overline{F}_Y(Y) \le \frac{y}{t}\right) = R(x,y) \quad \textit{for any} \quad (x,y) \in [0,\infty]^2 \setminus \{(\infty,\infty)\}.$$

Let $k_X = k_X(n), k_Y = k_Y(n)$ be such that $k_X, k_Y \to \infty$ and:

- $\sqrt{k_X}A_X(n/k_X) \to 0$ *and* $\sqrt{k_Y}A_Y(n/k_Y) \to 0$*;*
- $k_X/k_Y \to c \in (0,\infty)$*.*

Let also

$$R_i^X := \sum_{j=1}^n \mathbb{I}\{X_j \le X_i\} \quad \textit{and} \quad R_i^Y := \sum_{j=1}^n \mathbb{I}\{Y_j \le Y_i\}$$

denote the ranks of observations X_i and Y_i and $\widehat{R}_k(x,y)$ be the nonparametric estimator of $R(x,y)$ defined by

$$\widehat{R}_k(x,y) := \frac{1}{k}\sum_{i=1}^{n} \mathbb{I}\{R_i^X \geq n-kx,\, R_i^Y \geq n-ky\}$$

(see [107]). Then the following hold:

(i) Unless $c=1$ and (X,Y) are asymptotically perfectly dependent in the sense that $R(x,y) = \min(x,y)$, we have:

$$k_X \frac{[\widehat{\gamma}_X(k_X) - \widehat{\gamma}_Y(k_Y)]^2}{\widehat{\gamma}_X^2(k_X) + c\widehat{\gamma}_Y^2(k_Y) - 2\widehat{\gamma}_X(k_X)\widehat{\gamma}_Y(k_Y)\widehat{R}_{k_X}(c,1)} \xrightarrow{d} \begin{cases} \chi_1^2 & \text{if } \gamma_X = \gamma_Y, \\ +\infty & \text{if } \gamma_X \neq \gamma_Y. \end{cases}$$

(ii) If moreover $\gamma_X = \gamma_Y = \gamma$, then the estimator

$$\widehat{\gamma}(k_X,k_Y) := \frac{c - \widehat{R}_{k_X}(c,1)}{1+c-2\widehat{R}_{k_X}(c,1)}\widehat{\gamma}_X(k_X) + \frac{1 - \widehat{R}_{k_X}(c,1)}{1+c-2\widehat{R}_{k_X}(c,1)}\widehat{\gamma}_Y(k_Y)$$

satisfies

$$\sqrt{k_X}\,(\widehat{\gamma}(k_X,k_Y) - \gamma) \xrightarrow{d} \mathcal{N}\left(0, \gamma^2 \frac{c - R^2(c,1)}{1+c-2R(c,1)}\right)$$

and has the minimal possible asymptotic variance among the class of convex combinations of $\widehat{\gamma}_X(k_X)$ and $\widehat{\gamma}_Y(k_Y)$.

Corollary 2.6.8(i) complements Proposition 2 in [177] in the case $d=2$: the latter is stated under significantly stronger assumptions, although it includes the possibility of different sample sizes for the X and Y samples, which may be relevant in specific applied setups. Corollary 2.6.8(ii) provides, under the tail homogeneity condition $\gamma_X = \gamma_Y = \gamma$, a simple expression of the convex combination of $\widehat{\gamma}_X(k_X)$ and $\widehat{\gamma}_Y(k_Y)$ that is optimal for the estimation of γ in terms of asymptotic variance. It therefore constitutes, in the case $d=2$, an explicit version of the BEAR estimator given on pp.159–160 of [94]. With our notation, this estimator can be written, for a general value of d, as

$$\widehat{\gamma}_{\text{BEAR}}(k_1,\ldots,k_d) := \sum_{j=1}^{d} \widehat{\mu}_j \widehat{\gamma}_j(k_j), \text{ with } (\widehat{\mu}_1,\ldots,\widehat{\mu}_d) = \mathop{\arg\min}_{\substack{\mathbf{u}\in[0,1]^d \\ u_1+\cdots+u_d=1}} \mathbf{u}^\top \widehat{\mathbf{V}}\mathbf{u}$$

and $\widehat{\mathbf{V}}$ is a consistent estimator of the asymptotic covariance matrix $\mathbf{V}$ defined in Theorem 2.6.7. Let us highlight that the estimator $\widehat{\gamma}(k_X,k_Y)$ is analysed here under weaker conditions than those of [94] and, due to the particular choice $d=2$, without having to resort to a numerical optimisation routine for the calculation of the estimator. It is worth noting that since

$$\frac{c - R^2(c,1)}{1+c-2R(c,1)} - \min(c,1) = -\frac{[\min(c,1) - R(c,1)]^2}{1+c-2R(c,1)} \leq 0,$$

the estimator analysed in Corollary 2.6.8(ii) indeed has an asymptotic variance which is lower than each of the asymptotic variances of $\widehat{\gamma}_X(k_X)$ and $\widehat{\gamma}_Y(k_Y)$, and we can quantify this improvement. More precisely, since the function

$$r \mapsto \frac{[\min(c,1) - r]^2}{1+c-2r}$$

is decreasing on $[0,\min(c,1)]$, the improvement in asymptotic variance brought by the use of the pooled estimator $\widehat{\gamma}(k_X,k_Y)$ gets stronger as the asymptotic dependence structure of (X,Y) gets closer to asymptotic independence. In the case of asymptotic independence, we have

$$\sqrt{k_X}\,(\widehat{\gamma}(k_X,k_Y) - \gamma) \xrightarrow{d} \mathcal{N}\left(0, \gamma^2 \frac{c}{1+c}\right).$$

This is rather intuitive: since $\widehat{\gamma}(k_X, k_Y)$ is then essentially the weighted average of two independent quantities, made respectively of k_X independent log-excesses from X (with weighting $c/(1+c)$) and $k_Y \approx k_X/c$ independent log-excesses from Y (with weighting $1/(1+c)$), each with individual variance γ^2 by Corollary 2.6.2, we should expect the total asymptotic variance to be

$$\gamma^2 \left[\frac{c^2}{(1+c)^2} + c\frac{1}{(1+c)^2} \right] = \gamma^2 \frac{c}{1+c},$$

just as we found through our rigorous asymptotic analysis.

2.6.3.2 Convergence of a tail index estimator for right-truncated samples

We now illustrate how our results can be used to complete the analysis of the convergence of a tail index estimator when the available data is subject to random right-truncation. The context we consider is the following: let $(Y_1, T_1), \ldots, (Y_n, T_n)$ be n independent copies of a random pair (Y, T), where Y and T are independent, nonnegative, and have continuous marginal distribution functions F_Y and F_T. Assume also that Y and T have heavy-tailed distributions with tail indices γ_Y and γ_T. In the random right-truncation problem considered here, it is assumed that the pair (Y_i, T_i) is observed if and only if $Y_i \leq T_i$; otherwise, no information on this pair is available at all. The objective is to estimate γ_Y. This problem has only been considered very recently, starting with [22], where the focus was the estimation of extreme quantiles of Y. Several studies have since then proposed alternative techniques for tail index estimation in this context; we refer to [55, 56, 149, 261]. The random right-truncation context should not be mistaken for random right-censoring, where the available information is made of the pairs $(\min(Y_i, T_i), \mathbb{I}\{Y_i \leq T_i\})$, $1 \leq i \leq n$.

Our focus in this section is to revisit the asymptotic properties of the estimator of [22] using the tools we have developed in Sections 2.6.2 and 2.6.3.1. Let us introduce some notation beforehand: let N be the total (random) number of observed pairs (Y_i, T_i) such that $Y_i \leq T_i$. Such pairs shall be denoted as (Y_i^*, T_i^*), $1 \leq i \leq N$. It is straightforward to show the following:

- N has a binomial distribution with parameters n and $p := \mathbb{P}(Y \leq T)$, where we assume throughout this section that $p > 0$;
- Given N, the pairs (Y_i^*, T_i^*), $1 \leq i \leq N$ are independent copies of a random pair (Y^*, T^*) having joint distribution function $H^*(y, t) = \mathbb{P}(Y \leq y,\ T \leq t \,|\, Y \leq T)$, and corresponding marginal survival functions

$$\overline{F}_Y^*(y) := \frac{1}{p} \int_y^\infty \overline{F}_T(z) dF_Y(z) \quad \text{and} \quad \overline{F}_T^*(t) := \frac{1}{p} \int_t^\infty F_Y(z) dF_T(z).$$

It is also reasonably easy to show that in this context, $\overline{F}_Y^*$ and $\overline{F}_T^*$ are heavy-tailed, with tail indices $\gamma_Y^* := \gamma_Y \gamma_T / (\gamma_Y + \gamma_T)$ and γ_T (see Lemma 3 in [22] as well as the Introduction of [54]). Rewriting this equality as $\gamma_Y = \gamma_Y^* \gamma_T / (\gamma_T - \gamma_Y^*)$ motivates the following estimator of γ_Y:

$$\widehat{\gamma}_Y(k_N, k_N') := \frac{\widehat{\gamma}_Y^*(k_N) \widehat{\gamma}_T(k_N')}{\widehat{\gamma}_T(k_N') - \widehat{\gamma}_Y^*(k_N)}$$

$$\text{with} \quad \widehat{\gamma}_Y^*(k_N) := \frac{1}{k_N} \sum_{i=1}^{k_N} \log \frac{Y_{N-i+1,N}^*}{Y_{N-k_N,N}^*} \quad \text{and} \quad \widehat{\gamma}_T(k_N') := \frac{1}{k_N'} \sum_{i=1}^{k_N'} \log \frac{T_{N-i+1,N}^*}{T_{N-k_N',N}^*}.$$

Here (k_m) and (k_m') are two non-random sequences of integers and $(k_N, k_N') := (k_m, k_m')$ given $N = m$. The asymptotic distribution of this estimator, under the condition that $k_m/k_m' \to 0$ or $k_m'/k_m \to 0$ as $m \to \infty$, is examined in [22]. This technical restriction was imposed because the analysis of the dependence between the two Hill estimators $\widehat{\gamma}_Y^*(k_N)$ and $\widehat{\gamma}_T(k_N')$ is difficult; assuming that either $k_m/k_m' \to 0$ or $k_m'/k_m \to 0$ ensures that one of the estimators converges at a slower rate than the other, and therefore imposes its asymptotic distribution to $\widehat{\gamma}_Y(k_N, k_N')$. The asymptotic distribution of this estimator was subsequently studied in [54] when $k_m = k'_m$ and under a condition on the asymptotic dependence between Y^* and T^*. The main result of this section, which we state now, examines the general case $k_m/k'_m \to c \in [0,\infty]$.

Theorem 2.6.9. *Assume that Y^* and T^* satisfy conditions $\mathcal{C}_2(\gamma_Y^*, \rho_Y^*, A_Y^*)$ and $\mathcal{C}_2(\gamma_T, \rho_T^*, A_T^*)$, respectively. Let (k_m) and (k'_m) be two sequences of integers tending to infinity such that*

- $\max(k_m, k'_m)/m \to 0$, $\sqrt{k_m} A_Y^*(m/k_m) \to \lambda_1 \in \mathbb{R}$ *and* $\sqrt{k'_m} A_T^*(m/k'_m) \to \lambda_2 \in \mathbb{R}$*;*
- $k_m/k'_m \to c \in [0,\infty]$.

Then the following hold, as $n \to \infty$:

(i) If $c = 0$,

$$\sqrt{k_N}(\widehat{\gamma}_Y(k_N, k'_N) - \gamma_Y) \xrightarrow{d} \mathcal{N}\left(\frac{\gamma_Y^2}{\gamma_T^2}\frac{\lambda_1}{1-\rho_Y^*}\left[1+\frac{\gamma_T}{\gamma_Y}\right]^2, \frac{\gamma_Y^4}{\gamma_T^2}\left[1+\frac{\gamma_T}{\gamma_Y}\right]^2\right).$$

(ii) If $c = \infty$,

$$\sqrt{k'_N}(\widehat{\gamma}_Y(k_N, k'_N) - \gamma_Y) \xrightarrow{d} \mathcal{N}\left(-\frac{\gamma_Y^2}{\gamma_T^2}\frac{\lambda_2}{1-\rho_T^*}, \frac{\gamma_Y^4}{\gamma_T^2}\right).$$

(iii) If $0 < c < \infty$,

$$\sqrt{k_N}(\widehat{\gamma}_Y(k_N, k'_N) - \gamma_Y) \xrightarrow{d} \mathcal{N}\left(\frac{\gamma_Y^2}{\gamma_T^2}\left\{\frac{\lambda_1}{1-\rho_Y^*}\left[1+\frac{\gamma_T}{\gamma_Y}\right]^2 - \sqrt{c}\frac{\lambda_2}{1-\rho_T^*}\right\}, \frac{\gamma_Y^4}{\gamma_T^2}\left\{c + \left[1+\frac{\gamma_T}{\gamma_Y}\right]^2\right\}\right).$$

With convergences (i) and (ii), we essentially find back Theorem 3 in [22], which was stated under a slightly different set of second-order extreme value conditions. The value of our result, however, mainly resides in the general convergence result (iii), whose proof, unlike that of Theorem 2.1 in [54], does not hinge on a Gaussian construction for the weighted tail copula process. Most importantly for practical setups, and contrary to Theorem 2.1 in [54], Theorem 2.6.9(iii) does not rely on any assumption on the form of asymptotic dependence between Y^* and T^*. The reason for this is that the combination of the independence assumption between Y and T with the heavy-tailed framework is actually sufficient to ensure that Y^* and T^* are asymptotically independent, in the sense that

$$\lim_{t\to\infty} t\,\mathbb{P}\left(\overline{F}_Y^*(Y^*) \le \frac{x_1}{t},\ \overline{F}_T^*(T^*) \le \frac{x_2}{t}\right) = 0$$

for any $(x_1, x_2) \in [0,\infty)^2$. With this in mind, our result offers a simplified expression of the asymptotic variance of $\widehat{\gamma}_Y(k_N, k'_N)$, compared to the one provided in [54]. We also point out that, taking this asymptotic independence result into account, the asymptotic distribution we find in the case $c = 1$ essentially coincides with that of [54], although it should be pointed out that the bias term μ therein should read like a difference rather than a sum (this is revealed by inspecting Equation (3.10) therein) and their variance σ^2 should be divided by 2 (otherwise the Gaussian representation stated early in their Theorem 2.1 would contradict their asymptotic normality result). Finally, our result includes the possibility of taking proportional sequences k_N and k'_N, which is useful since one may obtain better finite-sample performance by selecting a value k'_N different from k_N if the marginal distributions of Y^* and T^* have very different second-order parameters ρ_Y^* and ρ_T^*. A related point about the estimation of a common tail index based on several samples of data is made in the Introduction of [94].

2.6.3.3 Joint convergence of marginal Conditional Tail Moments

Another consequence of our theoretical results can be formulated in terms of the joint convergence of high marginal Conditional Tail Moments (CTMs). As in Section 2.6.3.1, let $(\mathbf{X}_i)_{1\le i\le n}$, with $\mathbf{X}_i = (X_i^{(1)}, X_i^{(2)}, \ldots, X_i^{(d)})$, be a sample of independent copies of a random vector $\mathbf{X} = (X^{(1)}, X^{(2)}, \ldots, X^{(d)})$ such that each component $X^{(j)}$ has a continuous distribution function F_j and satisfies condition

$\mathcal{C}_2(\gamma_j, \rho_j, A_j)$. We let also, for $1 \le j \le d$, $k_j = k_j(n) \to \infty$ be such that $k_j/n \to 0$. The CTM of order a_j of the variable $X^{(j)}$ above its high quantile $U_j(n/k_j)$ is then

$$\mathrm{CTM}^{(j)}_{a_j}(U_j(n/k_j)) = \mathbb{E}([X^{(j)}]^{a_j} \,|\, X^{(j)} > U_j(n/k_j)).$$

With the language of Definition 2.6.1, this quantity is exactly the expected f_{a_j}−shortfall of $X^{(j)}$ above the quantile $U_j(n/k_j)$, where $f_{a_j}(x) = x^{a_j}$ for $x > 0$. Following the ideas of Section 2.6.2, its empirical counterpart is

$$\widehat{\mathrm{CTM}}^{(j)}_{a_j}(U_j(n/k_j)) = \frac{1}{k_j}\sum_{i=1}^{k_j}[X^{(j)}_{n-i+1,n}]^{a_j}.$$

This estimator is studied in [14, 114] (see also Section 2.2.3.3). The asymptotic results therein provide information on the joint convergence of estimators of several CTMs of a *single* variable X. Our next result below adopts the point of view of joint convergence of these estimators *across marginals.*

Theorem 2.6.10. *Work under the conditions of Theorem 2.6.7, with $\sqrt{k_j}A_j(n/k_j) \to \lambda_j \in \mathbb{R}$ replaced by the weaker assumption $\sqrt{k_j}A_j(n/k_j) = \mathrm{O}(1)$. Let also $a_1, a_2, \ldots, a_d > 0$ be such that $2a_j\gamma_j < 1$ for any j. Then*

$$\sqrt{k_1}\left(\frac{\widehat{\mathrm{CTM}}^{(j)}_{a_j}(U_j(n/k_j))}{\mathrm{CTM}^{(j)}_{a_j}(U_j(n/k_j))} - 1\right)_{1\le j\le d} \xrightarrow{d} \mathcal{N}_d\left(0,\, \boldsymbol{\Sigma}(\mathbf{a},\mathbf{c})\right)$$

with

$$\boldsymbol{\Sigma}_{j,l}(\mathbf{a},\mathbf{c}) = \begin{cases} c_j\dfrac{2a_j^2\gamma_j^2(1-a_j\gamma_j)}{1-2a_j\gamma_j} & \text{if } j = l, \\ a_j\gamma_j a_l\gamma_l(1-a_j\gamma_j)(1-a_l\gamma_l)\displaystyle\int_0^1\int_0^1 \frac{R_{j,l}\left(c_l u, c_j v\right)}{u^{a_j\gamma_j+1}v^{a_l\gamma_l+1}}du\,dv & \text{if } j < l. \end{cases}$$

A corollary of Theorem 2.6.10 is the following result on the joint convergence of several CTM estimators

$$\widehat{\mathrm{CTM}}_{a_j}(U(n/k)) = \frac{1}{k}\sum_{i=1}^{k} X^{a_j}_{n-i+1,n}$$

of a *single* heavy-tailed random variable X. This corollary immediately follows from the fact that the random pair (X, X) is asymptotically perfectly dependent, in the sense that

$$\lim_{t\to\infty} t\,\mathbb{P}\left(\overline{F}(X) \le \frac{x_1}{t},\ \overline{F}(X) \le \frac{x_2}{t}\right) = \min(x_1, x_2)$$

for any $(x_1, x_2) \in [0,\infty]^2 \setminus \{(\infty,\infty)\}$.

Corollary 2.6.11. *Suppose that X satisfies condition $\boldsymbol{\mathcal{C}_2(\gamma,\rho,A)}$. Let $k = k(n) \to \infty$ be such that $k/n \to 0$ and $\sqrt{k}A(n/k) = \mathrm{O}(1)$. Let also $a_1, a_2, \ldots, a_d > 0$ be such that $\gamma < 1/2a_j$ for any j. Then*

$$\sqrt{k}\left(\frac{\widehat{\mathrm{CTM}}_{a_j}(U(n/k))}{\mathrm{CTM}_{a_j}(U(n/k))} - 1\right)_{1\le j\le d} \xrightarrow{d} \mathcal{N}_d\left(0,\, \mathbf{M}(\mathbf{a})\right)$$

with

$$\mathbf{M}_{j,l}(\mathbf{a}) = \frac{a_j a_l\gamma^2(2-[a_j+a_l]\gamma)}{1-(a_j+a_l)\gamma}.$$

This result complements Theorem 2 in [14] (which in the present manuscript is Theorem 2.2.8), for distortion functions all equal to the identity function. A related result, in the presence of a finite-dimensional covariate, is Theorem 1 in [114].

While Theorem 2.6.10 is informative regarding the structure of the dependence between CTM estimators at high levels, its scope may however be limited from the practical point of view, as the focus in applied setups is usually on risk measures at out-of-sample levels. To put it differently, one would generally want to estimate a CTM of the form

$$\mathrm{CTM}_{a_j}^{(j)}(U_j(1/p_n)) = \mathbb{E}([X^{(j)}]^{a_j} \mid X^{(j)} > U_j(1/p_n))$$

where $p_n \to 0$ is such that $np_n = \mathrm{O}(1)$, a typical choice then being $p_n = 1/n$. In this context, following [14, 114], one may replace the estimator $\widehat{\mathrm{CTM}}_{a_j}^{(j)}$ by an extrapolated Weissman-type version:

$$\overline{\mathrm{CTM}}_{a_j}^{(j)}(U_j(1/p_n)) := \left(\frac{k_j}{np_n}\right)^{a_j\widehat{\gamma}_j(k_j)} \widehat{\mathrm{CTM}}_{a_j}^{(j)}(U_j(n/k_j)).$$

Here $\widehat{\gamma}_j(k_j)$ is the jth marginal Hill estimator introduced in Section 2.6.3.1. A combination of Theorems 2.6.7 and 2.6.10 makes it possible to obtain the following joint asymptotic normality result for these extrapolated estimators across marginals.

Corollary 2.6.12. *Work under the conditions of Theorem 2.6.7. Let $a_1, a_2, \ldots, a_d > 0$ be such that $2a_j\gamma_j < 1$ for any j. Assume also that $\rho_j < 0$ for any j, and that $np_n \to C < \infty$ and $\sqrt{k_1}/\log(k_1/[np_n]) \to \infty$. Then*

$$\sqrt{k_1}\left(\frac{1}{a_j \log(k_j/[np_n])}\left[\frac{\overline{\mathrm{CTM}}_{a_j}^{(j)}(U_j(1/p_n))}{\mathrm{CTM}_{a_j}^{(j)}(U_j(1/p_n))} - 1\right]\right)_{1\le j\le d} \xrightarrow{d} \mathcal{N}_d\left(\mathbf{b}(\mathbf{c}),\, \mathbf{V}(\mathbf{c})\right)$$

with the notation of Theorem 2.6.7.

In the case $a_1 = a_2 = \cdots = a_d$, such a result may then be used to test the equality of CTMs across marginals. For instance, if $d = 2$ and $a_1 = a_2 = 1$, one may test whether the two variables $X^{(1)}$ and $X^{(2)}$ have the same Expected Shortfall at high levels. The possibility of assessing the equality of financial tail risk within a multivariate context using such results is explored in Sections 3 and 4 in [156]; a related result is Proposition 3 in [156], albeit for a different type of estimator of the Expected Shortfall, and in a time series context.

2.6.4 Perspectives for future research

Stationary and weakly dependent sequences As [156] argues, assessing tail homogeneity of the marginals of a random vector is of particular interest in certain financial applications, where it is important to develop asymptotic theory for dependent but stationary sequences. Since the cornerstone of the proofs of Theorems 2.6.1 and 2.6.3 is a weighted Gaussian approximation of the univariate tail empirical process, it is reasonable to expect that analogues of these two results can also be shown in this kind of framework: for instance, [104] proves such an approximation in the case of $\beta-$mixing sequences under certain regularity conditions and upper bounds on the size of clusters of exceedances. This would therefore result in analogues of Theorems 2.6.1 and 2.6.3 and ultimately of Theorems 2.6.7 and 2.6.10. We note though that the expression of the asymptotic covariance matrices in such analogues of Theorems 2.6.7 and 2.6.10 would certainly be more complicated than in the present section, which, among others, makes the derivation of a testing procedure of tail homogeneity a difficult task. This issue is avoided in [156] by resorting to a self-normalised test statistic in the spirit of [229], although the limiting distribution does not have a simple expression. An interesting open question is to, based on an analogue of Theorem 2.6.7 in a dependent and stationary setting, construct an asymptotically chi-squared test statistic of tail homogeneity and contrast its finite-sample performance with that of the test statistic of [156].

Applications to multivariate extreme risk analysis The potential of Theorems 2.6.1 and 2.6.3 for the analysis of random vectors with marginal heavy tails is briefly described in Section 2.6.3.3 through the statement of the joint convergence of empirical and extrapolated Conditional Tail Moments estimators. The implications of these theoretical results on joint convergence of risk measures at extreme levels are wider. Indeed, we already know that our Weissman-type estimators in Sections 2.2–2.5 have their convergence dictated by that of the tail index estimator used, but this statement can be strengthened in the form of an asymptotic representation of such extrapolated estimators involving this tail index estimator. To see this, let r_τ be a risk measure satisfying the extrapolation relationship

$$r_{\tau_n'} \approx \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\gamma} r_{\tau_n}$$

where $\tau_n \to 1$ is an intermediate sequence and $\tau_n' \to 1$ is an extreme sequence, i.e. $n(1-\tau_n) \to \infty$ and $n(1-\tau_n') \to c < \infty$. Assume that, at the intermediate level τ_n, there is a $\sqrt{n(1-\tau_n)}-$relatively consistent estimator $\widehat{r}_{\tau_n}$ of r_{τ_n}, and let $\widehat{\gamma}_n$ be a $\sqrt{n(1-\tau_n)}-$consistent tail index estimator. Define a Weissman-type estimator of $r_{\tau_n'}$ by

$$\widehat{r}^W_{\tau_n'} := \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}_n} \widehat{r}_{\tau_n}.$$

This construction is exactly the one we used for the estimation of extreme expectiles, L^p-quantiles, tail L^p-medians and extremiles. Write then

$$\log\left(\frac{\widehat{r}^W_{\tau_n'}}{r_{\tau_n'}}\right) = (\widehat{\gamma}_n - \gamma)\log\left(\frac{1-\tau_n}{1-\tau_n'}\right) + \log\left(\frac{\widehat{r}_{\tau_n}}{r_{\tau_n}}\right) - \log\left(\left[\frac{1-\tau_n'}{1-\tau_n}\right]^{\gamma} \frac{r_{\tau_n'}}{r_{\tau_n}}\right).$$

The second term on the right-hand side converges at the rate $\sqrt{n(1-\tau_n)}$, so under a bias condition allowing to control the third term, the first term dominates, that is:

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\log\left(\frac{\widehat{r}^W_{\tau_n'}}{r_{\tau_n'}}\right) = \sqrt{n(1-\tau_n)}(\widehat{\gamma}_n - \gamma) + \mathrm{o}_{\mathbb{P}}(1).$$

This means that the joint convergence of the extrapolated estimators of extreme risk measures across marginals is exactly that of the tail index estimators of those marginals. When the tail index estimator is chosen to be the Hill estimator, this convergence is given by Theorem 2.6.7. In other words, and at extreme levels, the results of [4] automatically give the joint convergence of Weissman-type estimators of any risk measure whose (empirical, say) estimator at intermediate levels converges at the standard $\sqrt{n(1-\tau_n)}$ rate. Of course, this does not give a complete picture of the joint inference for such large risk measures, as we know that the behaviour of the (asymptotically negligible) estimator at the intermediate step may nonetheless have a substantial influence upon finite-sample performance. An interesting and challenging topic is therefore to analyse the joint convergence of empirical risk measure estimators at intermediate levels to gather some more insight as to how asymptotic dependence drives the joint behaviour of extreme risk measure estimators in multivariate problems.

Publications & References

List of published or accepted papers post-PhD (most recent first)

[1] Daouia, A., Girard, S., Stupfler, G. (2020). Tail expectile process and risk assessment, *Bernoulli* **26**(1): 531–556.

[2] Falk, M., Stupfler, G. (2019). The min-characteristic function: characterizing distributions by their min-linear projections, to appear in *Sankhyā*.

[3] Gardes, L., Girard, S., Stupfler, G. (2019). Beyond tail median and conditional tail expectation: extreme risk estimation using tail L^p-optimisation, to appear in *Scandinavian Journal of Statistics*.

[4] Stupfler, G. (2019). On a relationship between randomly and non-randomly thresholded empirical average excesses for heavy tails, *Extremes* **22**(4): 749–769.

[5] Daouia, A., Gijbels, I., Stupfler, G. (2019). Extremiles: A new perspective on asymmetric least squares, *Journal of the American Statistical Association* **114**(527): 1366–1381.

[6] Falk, M., Stupfler, G. (2019). On a class of norms generated by nonnegative integrable distributions, *Dependence Modeling* **7**(1): 259–278.

[7] Church, O., Derclaye, E., Stupfler, G. (2019). An empirical analysis of the design case law of the EU Member States, *International Review of Intellectual Property and Competition Law* **50**(6): 685–719.

[8] Stupfler, G. (2019). On the study of extremes with dependent random right-censoring, *Extremes* **22**(1): 97–129.

[9] Gardes, L., Stupfler, G. (2019). An integrated functional Weissman estimator for conditional extreme quantiles, *REVSTAT: Statistical Journal* **17**(1): 109–144.

[10] Daouia, A., Girard, S., Stupfler, G. (2019). Extreme M-quantiles as risk measures: From L^1 to L^p optimization, *Bernoulli* **25**(1): 264–309.

[11] El Methni, J., Stupfler, G. (2018). Improved estimators of extreme Wang distortion risk measures for very heavy-tailed distributions, *Econometrics and Statistics* **6**: 129–148.

[12] Daouia, A., Girard, S., Stupfler, G. (2018). Estimation of tail risk based on extreme expectiles, *Journal of the Royal Statistical Society: Series B* **80**(2): 263–292.

[13] Stupfler, G., Yang, F. (2018). Analyzing and predicting CAT bond premiums: a Financial Loss premium principle and extreme value modeling, *ASTIN Bulletin* **48**(1): 375–411.

[14] El Methni, J., Stupfler, G. (2017). Extreme versions of Wang risk measures and their estimation for heavy-tailed distributions, *Statistica Sinica* **27**(2): 907–930.

[15] Girard, S., Stupfler, G. (2017). Intriguing properties of extreme geometric quantiles, *REVSTAT: Statistical Journal* **15**(1): 107–139.

[16] Falk, M., Stupfler, G. (2017). An offspring of multivariate extreme value theory: the max-characteristic function, *Journal of Multivariate Analysis* **154**: 85–95.

[17] Stupfler, G. (2016). On the weak convergence of the kernel density estimator in the uniform topology, *Electronic Communications in Probability* **21**(17): 1–13.

[18] Stupfler, G. (2016). Estimating the conditional extreme-value index under random right-censoring, *Journal of Multivariate Analysis* **144**: 1–24.

[19] Girard, S., Stupfler, G. (2015). Extreme geometric quantiles in a multivariate regular variation framework, *Extremes* **18**(4): 629–663.

[20] Meintanis, S.G., Stupfler, G. (2015). Transformations to symmetry based on the probability weighted characteristic function, *Kybernetika* **51**(4): 571–587.

[21] Goegebeur, Y., Guillou, A., Stupfler, G. (2015). Uniform asymptotic properties of a nonparametric regression estimator of conditional tails, *Annales de l'Institut Henri Poincaré (B): Probability and Statistics* **51**(3): 1190–1213.

[22] Gardes, L., Stupfler, G. (2015). Estimating extreme quantiles under random truncation, *TEST* **24**(2): 207–227.

[23] Stupfler, G. (2014). On the weak convergence of kernel density estimators in L^p spaces, *Journal of Nonparametric Statistics* **26**(4): 721–735.

[24] Gardes, L., Stupfler, G. (2014). Estimation of the conditional tail index using a smoothed local Hill estimator, *Extremes* **17**(1): 45–75.

[25] Stupfler, G. (2013). A moment estimator for the conditional extreme-value index, *Electronic Journal of Statistics* **7**: 2298–2343.

List of published papers linked to my PhD work (most recent first)

[26] Guillou, A., Loisel, S., Stupfler, G. (2015). Estimating the parameters of a seasonal Markov-modulated Poisson process, *Statistical Methodology* **26**: 103–123.

[27] Girard, S., Guillou, A., Stupfler, G. (2014). Uniform strong consistency of a frontier estimator using kernel regression on high order moments, *ESAIM: Probability and Statistics* **18**: 642–666.

[28] Guillou, A., Loisel, S., Stupfler, G. (2013). Estimation of the parameters of a Markov-modulated loss process in insurance, *Insurance: Mathematics and Economics* **53**(2): 388–404.

[29] Girard, S., Guillou, A., Stupfler, G. (2013). Frontier estimation with kernel regression on high order moments, *Journal of Multivariate Analysis* **116**: 172–189.

[30] Girard, S., Guillou, A., Stupfler, G. (2012). Estimating an endpoint with high order moments in the Weibull domain of attraction, *Statistics and Probability Letters* **82**(12): 2136–2144.

[31] Girard, S., Guillou, A., Stupfler, G. (2012). Estimating an endpoint with high-order moments, *TEST* **21**(4): 697–729.

Papers submitted or under revision

[32] Girard, S., Stupfler, G., Usseglio-Carleve, A. (2019). An L^p-quantile methodology for tail index estimation, submitted.

[33] Girard, S., Stupfler, G., Usseglio-Carleve, A. (2019). Nonparametric extreme conditional expectile estimation, submitted.

[34] Daouia, A., Gijbels, I., Stupfler, G. (2018). Extremile regression, submitted.

[35] Mitchell, E.G., Crout, N.M.J., Wilson, P., Wood, A.T.A., Stupfler, G. (2018). Operating at the extreme: estimating the upper yield boundary of winter wheat production in commercial practice, submitted.

[36] Daouia, A., Girard, S., Stupfler, G. (2018). ExpectHill estimation, extreme risk and heavy tails, submitted.

References

[37] Aban, I.B., Meerschaert, M.M., Panorska, A.K. (2006). Parameter estimation for the truncated Pareto distribution, *Journal of the American Statistical Association* **101**(473): 270–277.

[38] Abdous, B., Remillard, B. (1995). Relating quantiles and expectiles under weighted-symmetry, *Annals of the Institute of Statistical Mathematics* **47**(2): 371–384.

[39] Abramovitz, M., Stegun, I.A. (1972). *Handbook of Mathematical Functions* (10th printing), National Bureau of Standards Applied Mathematics Series, Washington D.C.

[40] Acerbi, C. (2002). Spectral measures of risk: A coherent representation of subjective risk aversion, *Journal of Banking & Finance* **26**(7): 1505–1518.

[41] Acerbi, C., Tasche, D. (2002). On the coherence of expected shortfall, *Journal of Banking & Finance* **26**(7): 1487–1503.

[42] Acharya, V.V., Pedersen, L.H., Philippon, T., Richardson, M. (2017). Measuring systemic risk, *The Review of Financial Studies* **30**(1): 2–47.

[43] Aloui, C., Mabrouk, S. (2010). Value-at-risk estimations of energy commodities via long-memory, asymmetry and fat-tailed GARCH models, *Energy Policy* **38**(5): 2326–2339.

[44] Artzner, P., Delbaen, F., Eber, J.-M., Heath, D. (1999). Coherent measures of risk, *Mathematical Finance* **9**(3): 203–228.

[45] Balslev, U., Monforte, A.D., Stergiou, G., Antunes, F., Mulcahy, F., Pehrson, P.O., Phillips, A., Pedersen, C., Lundgren, J.D. (1997). Influence of age on rates of new AIDS-defining diseases and survival in 6546 AIDS patients, *Scandinavian Journal of Infectious Diseases* **29**(4): 337–343.

[46] Bassett, G.W., Koenker, R., Kordas, G. (2004). Pessimistic portfolio allocation and Choquet expected utility, *Journal of Financial Econometrics* **2**(4): 477–492.

[47] Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*, John Wiley & Sons, Chichester.

[48] Beirlant, J., Guillou, A., Dierckx, G., Fils-Villetard, A. (2007). Estimation of the extreme value index and extreme quantiles under random censoring, *Extremes* **10**(3): 151–174.

[49] Beirlant, J., Guillou, A., Toulemonde, G. (2010). Peaks-Over-Threshold modeling under random censoring, *Communications in Statistics – Theory and Methods* **39**(7): 1158–1179.

[50] Beirlant, J., Vynckier, P., Teugels, J.L. (1996). Excess functions and estimation of the extreme value index, *Bernoulli* **2**(4): 293–318.

[51] Belles-Sampera, J., Guillén, M., Santolino, M. (2014). Beyond Value-at-Risk: GlueVaR distortion risk measures, *Risk Analysis* **34**(1): 121–134.

[52] Bellini, F., Di Bernardino, E. (2017). Risk management with expectiles, *The European Journal of Finance* **23**(6): 487–506.

[53] Bellini, F., Klar, B., Müller, A., Gianin, E.R. (2014). Generalized quantiles as risk measures, *Insurance: Mathematics and Economics* **54**: 41–48.

[54] Benchaira, S., Meraghni, D., Necir, A. (2015). On the asymptotic normality of the extreme value index for right-truncated data, *Statistics & Probability Letters* **107**: 378–384.

[55] Benchaira, S., Meraghni, D., Necir, A. (2016a). Tail product-limit process for truncated data with application to extreme value index estimation, *Extremes* **19**(2): 219–251.

[56] Benchaira, S., Meraghni, D., Necir, A. (2016b). Kernel estimation of the tail index of a right-truncated Pareto-type distribution, *Statistics & Probability Letters* **119**: 186–193.

[57] Bickel, P.J. (1982). On adaptive estimation, *Annals of Statistics* **10**(3): 647–671.

[58] Bickel, P.J., Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates, *Annals of Statistics* **1**(6): 1071–1095.

[59] Bingham, N.H., Goldie, C.M., Teugels, J.L. (1987). *Regular Variation*, Cambridge University Press.

[60] Borisov, I.S., Baklanov, E.A. (2001). Probability inequalities for generalized $L-$statistics, *Siberian Mathematical Journal* **42**(2): 217–231.

[61] Box, G.E.P., Cox, D.R. (1964). An analysis of transformations, *Journal of the Royal Statistical Society: Series B* **26**(2): 211–243.

[62] Bradley, R.C. (2005). Basic properties of strong mixing conditions. A survey and some open questions, *Probability Surveys* **2**: 107–144.

[63] Breckling, J., Chambers, R. (1988). M-quantiles, *Biometrika* **75**(4): 761–771.

[64] Brownlees, C.T., Engle, R. (2017). SRISK: A conditional capital shortfall measure of systemic risk, *The Review of Financial Studies* **30**(1): 48–79.

[65] Caeiro, F., Gomes, M.I., Pestana, D. (2005). Direct reduction of bias of the classical Hill estimator, *REVSTAT: Statistical Journal* **3**(2): 113–136.

[66] Caeiro, F., Gomes, M.I., Rodrigues, L.H. (2009). Reduced-bias tail index estimators under a third-order framework, *Communications in Statistics – Theory and Methods* **38**(7): 1019–1040.

[67] Cai, J.-J., Einmahl, J.H.J., de Haan, L. (2011). Estimation of extreme risk regions under multivariate regular variation, *Annals of Statistics* **39**(3): 1803–1826.

[68] Cai, J.J., Einmahl, J.H.J., de Haan, L., Zhou, C. (2015). Estimation of the marginal expected shortfall: the mean when a related variable is extreme, *Journal of the Royal Statistical Society: Series B* **77**(2): 417–442.

[69] Chaouch, M., Goga, C. (2010). Design-based estimation for geometric quantiles with application to outlier detection, *Computational Statistics & Data Analysis* **54**(10): 2214–2229.

[70] Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data, *Journal of the American Statistical Association* **91**(434): 862–872.

[71] Chavez-Demoulin, V., Davison, A.C. (2005). Generalized additive modelling of sample extremes, *Journal of the Royal Statistical Society: Series C* **54**(1): 207–222.

[72] Chavez-Demoulin, V., Embrechts, P., Sardy, S. (2014). Extreme-quantile tracking for financial time series, *Journal of Econometrics* **181**(1): 44–52.

[73] Chen, Z. (1996). Conditional L_p-quantiles and their application to testing of symmetry in non-parametric regression, *Statistics* & *Probability Letters* **29**(2): 107–115.

[74] Chernozhukov, V. (2005). Extremal quantile regression, *Annals of Statistics* **33**(2): 806–839.

[75] Cherny, A., Madan, D. (2009). New measures for performance evaluation, *The Review of Financial Studies* **22**(7): 2571–2606.

[76] Cirillo, P., Taleb, N.N. (2016). Expected shortfall estimation for apparently infinite-mean models of operational risk, *Quantitative Finance* **16**(10): 1485–1494.

[77] Cooley, D., Hunter, B.D., Smith, R.L. (2019). Univariate and multivariate extremes for the environmental sciences, in Gelfand, A.E., Fuentes, M., Hoeting, J.A., Smith, R.L. (eds.) *Handbook of Environmental and Ecological Statistics*: 153–180.

[78] Cotter, J., Dowd, K. (2006). Extreme spectral risk measures: An application to futures clearinghouse margin requirements, *Journal of Banking & Finance* **30**(12): 3469–3485.

[79] Cramér, H., Wold, H. (1936). Some theorems on distribution functions, *Journal of the London Mathematical Society* **s1-11**(4): 290–294.

[80] Csörgő, M., Csörgő, S., Horváth, L., Mason, D.M. (1986). Normal and stable convergence of integral functions of the empirical distribution function, *Annals of Probability* **14**(1): 86–118.

[81] Csörgő, M., Horváth, L. (1987). Approximation of intermediate quantile processes, *Journal of Multivariate Analysis* **21**(2): 250–262.

[82] Csörgő, M., Horváth, L. (1993). *Weighted Approximations in Probability and Statistics*, John Wiley & Sons, Chichester.

[83] Csörgő, S., Horváth, L., Mason, D.M. (1986). What portion of the sample makes a partial sum asymptotically stable or normal?, *Probability Theory and Related Fields* **72**(1): 1–16.

[84] Daouia, A., Gardes, L., Girard, S. (2013). On kernel smoothing for extremal quantile regression, *Bernoulli* **19**(5B): 2557–2589.

[85] Daouia, A., Gardes, L., Girard, S., Lekina, A. (2011). Kernel estimators of extreme level curves, *Test* **20**(2): 311–333.

[86] Daouia, A., Gijbels, I. (2011). Estimating frontier cost models using extremiles, in Van Keilegom, I., Wilson, P. (eds.) *Exploring Research Frontiers in Contemporary Statistics and Econometrics - Festschrift in honor of Leopold Simar*: 65–81.

[87] Darby, S.C., Ewart, D.W., Giangrande, P.L., Spooner, R.J., Rizza, C.R. (1996). Importance of age at infection with HIV-1 for survival and development of AIDS in UK haemophilia population, *The Lancet* **347**(9015): 1573–1579.

[88] Davison, A.C., Ramesh, N.I. (2000). Local likelihood smoothing of sample extremes, *Journal of the Royal Statistical Society: Series B* **62**(1): 191–208.

[89] Davison, A.C., Smith, R.L. (1990). Models for exceedances over high thresholds, *Journal of the Royal Statistical Society: Series B* **52**(3): 393–442.

[90] Degiannakis, S. (2004). Volatility forecasting: evidence from a fractional integrated asymmetric power ARCH skewed-t model, *Applied Financial Economics* **14**(18): 1333–1342.

[91] Deheuvels, P. (1984). Probabilistic aspects of multivariate extremes, in Tiago de Oliveira, J. (ed.) Statistical extremes and applications, *NATO ASI Series (Series C: Mathematical and Physical Sciences)* **131**: 117–130.

[92] Dekkers, A.L.M., Einmahl, J.H.J., de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution, *Annals of Statistics* **17**(4): 1833–1855.

[93] Delbaen, F. (2002). *Coherent risk measures – Lectures given at the Cattedra Galileiana at the Scuola Normale Superiore di Pisa, March 2000*, Publications of the Scuola Normale Superiore, Pisa.

[94] Dematteo, A., Clémençon, S. (2016). On tail index estimation based on multivariate data, *Journal of Nonparametric Statistics* **28**(1): 152–176.

[95] Deme, E.H., Girard, S., Guillou, A. (2013). Reduced-bias estimator of the Proportional Hazard Premium for heavy-tailed distributions, *Insurance: Mathematics and Economics* **52**(3): 550–559.

[96] Deme, E.H., Girard, S., Guillou, A. (2015). Reduced-bias estimators of the Conditional Tail Expectation of heavy-tailed distributions, in Hallin, M., Mason, D.M., Pfeifer, D., Steinebach, J.G. (eds.) *Mathematical Statistics and Limit Theorems*: 105–123.

[97] Deprins, D., Simar, L., Tulkens, H. (1984). Measuring labor efficiency in post offices, in Marchand, M., Pestieau, P., Tulkens, H. (eds.) *The Performance of Public Enterprises: Concepts and Measurement*: 243–267.

[98] Dietrich, D., de Haan, L., Hüsler, J. (2002). Testing extreme value conditions, *Extremes* **5**(1): 71–85.

[99] Dony, J., Einmahl, U. (2006). Weighted uniform consistency of kernel density estimators with general bandwidth sequences, *Electronic Journal of Probability* **11**: 844–859.

[100] Doukhan, P. (1994). *Mixing: Properties and Examples*, Springer-Verlag, New York.

[101] Drees, H. (1998). On smooth statistical tail functionals, *Scandinavian Journal of Statistics* **25**(1): 187–210.

[102] Drees, H. (2000). Weighted approximations of tail processes for β-mixing random variables, *Annals of Applied Probability* **10**(4): 1274–1301.

[103] Drees, H. (2002). Tail empirical processes under mixing conditions, in Dehling, H.G., Mikosch, T., Sørensen, M. (eds.) *Empirical Processes Techniques for Dependent Data*: 325–342.

[104] Drees, H. (2003). Extreme quantile estimation for dependent data, with applications to finance, *Bernoulli* **9**(4): 617–657.

[105] Drees, H., de Haan, L., Li, D. (2006). Approximations to the tail empirical distribution function with application to testing extreme value conditions, *Journal of Statistical Planning and Inference* **136**(10): 3498–3538.

[106] Drees, H., de Haan, L., Resnick, S. (2000). How to make a Hill plot, *Annals of Statistics* **28**(1): 254–274.

[107] Drees, H., Huang, X. (1998). Best attainable rates of convergence for estimators of the stable tail dependence function, *Journal of Multivariate Analysis* **64**(1): 25–47.

[108] Ebrahimi, N., Molefe, D. (2003). Survival function estimation when lifetime and censoring time are dependent, *Journal of Multivariate Analysis* **87**(1): 101–132.

[109] Ehm, W., Gneiting, T., Jordan, A., Krüger, F. (2016). Of quantiles and expectiles: consistent scoring functions, Choquet representations and forecast rankings (with discussion), *Journal of the Royal Statistical Society: Series B* **78**(3): 505–562.

[110] Einmahl, J.H.J., Fils-Villetard, A., Guillou, A. (2008). Statistics of extremes under random censoring, *Bernoulli* **14**(1): 207–227.

[111] Einmahl, J.H.J., Mason, D.M. (1988). Strong limit theorems for weighted quantile processes, *Annals of Probability* **16**(4): 1623–1643.

[112] Einmahl, U., Mason, D.M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators, *Journal of Theoretical Probability* **13**(1): 1–37.

[113] Einmahl, U., Mason, D.M. (2005). Uniform in bandwidth consistency of kernel-type function estimators, *Annals of Statistics* **33**(3): 1380–1403.

[114] El Methni, J., Gardes, L., Girard, S. (2014). Non-parametric estimation of extreme risk measures from conditional heavy-tailed distributions, *Scandinavian Journal of Statistics* **41**(4): 988–1012.

[115] Embrechts, P., Klüppelberg, C., Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, Springer-Verlag, Berlin.

[116] Engle, R.F., Jondeau, E., Rockinger, M. (2015). Systemic risk in Europe, *Review of Finance* **19**(1): 145–190.

[117] Falk, M. (2019). *Multivariate Extreme Value Theory and D-Norms*, Springer Nature Switzerland.

[118] Falk, M., Hüsler, J., Reiss, R.-D. (2011). *Laws of Small Numbers: Extremes and Rare Events*, Springer, Basel.

[119] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications – Volume II* (second edition), Wiley, New York.

[120] Fissler, T., Ziegel, J.F. (2016). Higher order elicitability and Osband's principle, *Annals of Statistics* **44**(4): 1680–1707.

[121] Föllmer, H., Knispel, T. (2013). Convex risk measures: Basic facts, law-invariance and beyond, asymptotics for large portfolios, in *Handbook of the Fundamentals of Financial Decision Making: Part II*: 507–554.

[122] Fraga Alves, M.I., Gomes, M.I., de Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter, *Portugaliae Mathematica* **60**(2): 193–213.

[123] Frahm, G., Junker, M., Schmidt, R. (2005). Estimating the tail-dependence coefficient: properties and pitfalls, *Insurance: Mathematics and Economics* **37**(1): 80–100.

[124] Friederichs, P., Wahl, S., Buschow, S. (2018). Postprocessing for extreme events, in Vannitsem, S., Wilks, D.S., Messner, J.W. (eds.) *Statistical Postprocessing of Ensemble Forecasts*: 127–154.

[125] Furman, E., Landsman, Z. (2006). On some risk-adjusted tail-based premium calculation principles, *Journal of Actuarial Practice* **13**: 175–190.

[126] Galambos, J. (1978). *The asymptotic theory of extreme order statistics*, John Wiley & Sons, New York-Chichester-Brisbane.

[127] Gardes, L. (2018). Tail dimension reduction for extreme quantile estimation, *Extremes* **21**(1): 57–95.

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

[129] Gardes, L., Girard, S. (2010). Conditional extremes from heavy-tailed distributions: an application to the estimation of extreme rainfall return levels, *Extremes* **13**(2): 177–204.

[130] Gardes, L., Girard, S. (2012). Functional kernel estimators of large conditional quantiles, *Electronic Journal of Statistics* **6**: 1715–1744.

[131] Ghosh, S., Resnick, S. (2010). A discussion on mean excess plots, *Stochastic Processes and their Applications* **120**(8): 1492–1517.

[132] Giné, E., Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators, *Annales de l'Institut Henri Poincaré (B): Probability and Statistics* **38**(6): 907–921.

[133] Giné, E., Koltchinskii, V., Zinn, J. (2004). Weighted uniform consistency of kernel density estimators, *Annals of Probability* **32**(3B): 2570–2605.

[134] Gneiting, T. (2011). Making and evaluating point forecasts, *Journal of the American Statistical Association* **106**(494): 746–762.

[135] Goegebeur, Y., Beirlant, J., de Wet, T. (2010). Kernel estimators for the second order parameter in extreme value statistics, *Journal of Statistical Planning and Inference* **140**(9): 2632–2652.

[136] Goldie, C.M., Smith, R.L. (1987). Slow variation with remainder: theory and applications, *The Quarterly Journal of Mathematics* **38**(1): 45–71.

[137] Gomes, M.I., Brilhante, M.F., Pestana, D. (2016). New reduced-bias estimators of a positive extreme value index, *Communications in Statistics – Simulation and Computation* **45**(3): 833–862.

[138] Gomes, M.I., Guillou, A. (2015). Extreme value theory and statistics of univariate extremes: a review, *International Statistical Review* **83**(2): 263–292.

[139] Gomes, M.I., Neves, M.M. (2011). Estimation of the extreme value index for randomly censored data, *Biometrical Letters* **48**(1): 1–22.

[140] Gomes, M.I., Pestana, D., Caeiro, F. (2009). A note on the asymptotic variance at optimal levels of a bias-corrected Hill estimator, *Statistics & Probability Letters* **79**(3): 295–303.

[141] Gong, J., Li, Y., Peng, L., Yao, Q. (2015). Estimation of extreme quantiles for functions of dependent random variables, *Journal of the Royal Statistical Society: Series B* **77**(5): 1001–1024.

[142] González-Rivera, G., Drost, F.C. (1999). Efficiency comparisons of maximum-likelihood-based estimators in GARCH models, *Journal of Econometrics* **93**(1): 93–111.

[143] Greenwood, J.A., Landwehr, J.M., Matalas, N.C., Wallis, J.R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form, *Water Resources Research* **15**(5): 1049–1054.

[144] Gudendorf, G., Segers, J. (2010). Extreme-value copulas, in Jaworski, P., Durante, F., Härdle, W.K., Rychlik, T. (eds.) Copula Theory and Its Applications, *Lecture Notes in Statistics* **198**: 127–145.

[145] de Haan, L., Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, Springer, New York.

[146] de Haan, L., Mercadier, C., Zhou, C. (2016). Adapting extreme value statistics to financial time series: dealing with bias and serial dependence, *Finance and Stochastics* **20**(2): 321–354.

[147] Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators, *Journal of Multivariate Analysis* **14**(1): 1–16.

[148] Hall, P., Tajvidi, N. (2000). Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data, *Statistical Science* **15**(2): 153–167.

[149] Haouas, N., Necir, A., Meraghni, D., Brahimi, B. (2017). A Lynden-Bell integral estimator for the tail index of right-truncated data with a random threshold, *Afrika Statistika* **12**(1): 1159–1170.

[150] Härdle, W., Marron, J.S. (1985). Optimal bandwidth selection in nonparametric regression function estimation, *Annals of Statistics* **13**(4): 1465–1481.

[151] He, F., Cheng, Y., Tong, T. (2016). Estimation of high conditional quantiles using the Hill estimator of the tail index, *Journal of Statistical Planning and Inference* **176**: 64–77.

[152] Herbst, T. (1999). An application of randomly truncated data models in reserving IBNR claims, *Insurance: Mathematics and Economics* **25**(2): 123–131.

[153] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *Annals of Statistics* **3**(5): 1163–1174.

[154] Hill, J.B. (2013). Least tail-trimmed squares for infinite variance autoregressions, *Journal of Time Series Analysis* **34**(2): 168–186.

[155] Hill, J.B. (2015). Expected shortfall estimation and Gaussian inference for infinite variance time series, *Journal of Financial Econometrics* **13**(1): 1–44.

[156] Hoga, Y. (2018). Detecting tail risk differences in multivariate time series, *Journal of Time Series Analysis* **39**(5): 665–689.

[157] Holzmann, H., Klar, B. (2016). Expectile asymptotics, *Electronic Journal of Statistics* **10**(2): 2355–2371.

[158] Horváth, L., Yandell, B.S. (1988). Asymptotics of conditional empirical processes, *Journal of Multivariate Analysis* **26**(2): 184–206.

[159] Hosking, J.R.M., Wallis, J.R., Wood, E.F. (1985). Estimation of the generalized extreme-value distribution by the method of probability weighted moments, *Technometrics* **27**(3): 251–261.

[160] Hsing, T. (1991). On tail index estimation using dependent data, *Annals of Statistics* **19**(3): 1547–1569.

[161] Hu, X.J., Lawless, J.F. (1996a). Estimation from truncated life time data with supplementary information on covariates and censoring times, *Biometrika* **83**(4): 747–761.

[162] Hu, X.J., Lawless, J.F. (1996b). Estimation of rate and mean functions from truncated recurrent event data, *Journal of the American Statistical Association* **91**(433): 300–310.

[163] Hua, L., Joe, H. (2011). Second order regular variation and conditional tail expectation of multiple risks, *Insurance: Mathematics and Economics* **49**(3): 537–546.

[164] Huang, X., Zhang, N. (2008). Regression survival analysis with an assumed copula for dependent censoring: a sensitivity analysis approach, *Biometrics* **64**(4): 1090–1099.

[165] Hüsler, J., Li, D. (2006). On testing extreme value conditions, *Extremes* **9**(1): 69–86.

[166] Jackson, J.C. (1974). The analysis of quasar samples, *Monthly Notices of the Royal Astronomical Society* **166**(2): 281–295.

[167] Jackson, D., White, I.R., Seaman, S., Evans, H., Baisley, K., Carpenter, J. (2014). Relaxing the independent censoring assumption in the Cox proportional hazards model using multiple imputation, *Statistics in Medicine* **33**(27): 4681–4694.

[168] Jiang, Y., Sun, H., Zhang, Y., Long, H. (2017). Non-parametric tests for the tail equivalence via empirical likelihood, *Communications in Statistics – Theory and Methods* **46**(21): 10640–10656.

[169] Jones, B.L., Zitikis, R. (2003). Empirical estimation of risk measures and related quantities, *North American Actuarial Journal* **7**(4): 44–54.

[170] Jouini, E., Schachermayer, W., Touzi, N. (2006). Law invariant risk measures have the Fatou property, in Kusuoka, S., Yamazaki, A. (eds.) *Advances in Mathematical Economics* **9**: 49–71.

[171] Kalbfleisch, J.D., Lawless, J.F. (1989). Inference based on retrospective ascertainment: an analysis of the data on transfusion-related AIDS, *Journal of the American Statistical Association* **84**(406): 360–372.

[172] Kalbfleisch, J.D., Lawless, J.F. (1991). Regression models for right truncated data with applications to AIDS incubation times and reporting lags, *Statistica Sinica* **1**(1): 19–32.

[173] Kaminsky, K.S. (1987). Prediction of IBNR claim counts by modelling the distribution of report lags, *Insurance: Mathematics and Economics* **6**(2): 151–159.

[174] Kaplan, E.L., Meier, P. (1958). Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association* **53**(282): 457–481.

[175] Karmarkar, M., Paul, S. (2019). Intraday portfolio risk management using VaR and CVaR: A CGARCH-EVT-Copula approach, *International Journal of Forecasting* **35**(2): 699–709.

[176] Kato, K. (2009). Asymptotics for argmin processes: Convexity arguments, *Journal of Multivariate Analysis* **100**(8): 1816–1829.

[177] Kinsvater, P., Fried, R., Lilienthal, J. (2016). Regional extreme value index estimation and a test of tail homogeneity, *Environmetrics* **27**(2): 103–115.

[178] Klein, J.P., Moeschberger, M.L. (1987). Independent or dependent competing risks: does it make a difference? *Communications in Statistics – Simulation and Computation* **16**(2): 507–533.

[179] Knight, K. (1998). Limiting distributions for L_1 regression estimators under general conditions, *Annals of Statistics* **26**(2): 755–770.

[180] Knight, K. (1999). Epi-convergence in distribution and stochastic equi-semicontinuity, technical report, University of Toronto.

[181] Koenker, R. (2005). *Quantile Regression*, Cambridge University Press.

[182] Koenker, R., Bassett, G. (1978). Regression quantiles, *Econometrica* **46**(1): 33–50.

[183] Kokic, P., Chambers, R., Breckling, J., Beare, S. (1997). A measure of production performance, *Journal of Business & Economic Statistics* **15**(4): 445–451.

[184] Koltchinskii, V.I. (1997). M-estimation, convexity and quantiles, *Annals of Statistics* **25**(2): 435–477.

[185] Kou, S., Peng, X. (2016). On the measurement of economic tail risk, *Operations Research* **64**(5): 1056–1072.

[186] Kou, S., Peng, X., Heyde, C.C. (2013). External risk measures and Basel accords, *Mathematics of Operations Research* **38**(3): 393–417.

[187] Koutsoyiannis, D. (2004). Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records, *Hydrological Sciences Journal* **49**(4): 591–610.

[188] Krätschmer, V., Zähle, H. (2017). Statistical inference for expectile-based risk measures, *Scandinavian Journal of Statistics* **44**(2): 425–454.

[189] Kratz, M., Resnick, S. (1996). The QQ-estimator and heavy tails, *Communications in Statistics – Stochastic Models* **12**(4): 699–724.

[190] Kuan, C.-M., Yeh, J.-H., Hsu, Y.-C. (2009). Assessing value at risk with CARE, the Conditional Autoregressive Expectile models, *Journal of Econometrics* **150**(2): 261–270.

[191] Kusuoka, S. (2001). On law invariant coherent risk measures, in Kusuoka, S., Maruyama, T. (eds.) *Advances in Mathematical Economics* **3**: 83–95.

[192] Lagakos, S.W., Barraj, L.M., De Gruttola, V. (1988). Nonparametric analysis of truncated survival data, with application to AIDS, *Biometrika* **75**(3): 515–523.

[193] Landwehr, J.M., Matalas, N.C., Wallis, J.R. (1979). Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles, *Water Resources Research* **15**(5): 1055–1064.

[194] Lawless, J.F. (1994). Adjustments for reporting delays and the prediction of occurred but not reported events, *Canadian Journal of Statistics* **22**(1): 15–31.

[195] Ling, S. (2005). Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *Journal of the Royal Statistical Society: Series B* **67**(3): 381–393.

[196] Luo, K., Law, M., Kaldor, J.M., McDonald, A.M., Cooper, D.A. (1995). The role of initial AIDS-defining illness in survival following AIDS, *AIDS* **9**(1): 57–63.

[197] Lynden-Bell, D. (1971). A method of allowing for known observational selection in small samples applied to 3CR quasars, *Monthly Notices of the Royal Astronomical Society* **155**(1): 95–118.

[198] Mao, T., Ng, K.W., Hu, T. (2015). Asymptotic expansions of generalized quantiles and expectiles for extreme risks, *Probability in the Engineering and Informational Sciences* **29**(3): 309–327.

[199] Mao, T., Yang, F. (2015). Risk concentration based on expectiles for extreme risks under FGM copula, *Insurance: Mathematics and Economics* **64**: 429–439.

[200] McNeil, A.J., Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach, *Journal of Empirical Finance* **7**(3-4): 271–300.

[201] McNeil, A.J., Frey, R., Embrechts, P. (2005). *Quantitative Risk Management - Concepts, Techniques, and Tools*, Princeton University Press, New Jersey.

[202] Meeker, W.Q., Escobar, L.A. (1998). *Statistical methods for reliability data*, John Wiley & Sons, New York.

[203] Meintanis, S.G., Swanepoel, J., Allison, J. (2014). The probability weighted characteristic function and goodness-of-fit testing, *Journal of Statistical Planning and Inference* **146**: 122–132.

[204] Modarres, M., Kaminskiy, M.P., Krivtsov, V. (2009). *Reliability engineering and risk analysis: a practical guide* (second edition), CRC Press.

[205] Ndao, P., Diop, A., Dupuy, J.-F. (2014). Nonparametric estimation of the conditional tail index and extreme quantiles under random censoring, *Computational Statistics & Data Analysis* **79**: 63–79.

[206] Necir, A., Meraghni, D. (2009). Empirical estimation of the proportional hazard premium for heavy-tailed claim amounts, *Insurance: Mathematics and Economics* **45**(1): 49–58.

[207] Nešlehová, J., Embrechts, P., Chavez-Demoulin, V. (2006). Infinite mean models and the LDA for operational risk, *Journal of Operational Risk* **1**(1): 3–25.

[208] Newey, W.K. (1988). Adaptive estimation of regression models via moment restrictions, *Journal of Econometrics* **38**(3): 301–339.

[209] Newey, W.K., Powell, J.L. (1987). Asymmetric least squares estimation and testing, *Econometrica* **55**(4): 819–847.

[210] Newey, W.K., Steigerwald, D.G. (1997). Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models, *Econometrica* **65**(3): 587–599.

[211] Nishiyama, Y. (2011). Impossibility of weak convergence of kernel density estimators to a nondegenerate law in $L_2(\mathbb{R}^d)$, *Journal of Nonparametric Statistics* **23**(1): 129–135.

[212] Peng, L. (1998). Asymptotically unbiased estimators for the extreme-value index, *Statistics & Probability Letters* **38**(2): 107–115.

[213] Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer-Verlag, New York.

[214] Resnick, S.I. (1987). *Extreme Values, Regular Variation, and Point Processes*, Springer, New York.

[215] Resnick, S.I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer, New York.

[216] Resnick, S., Stărică, C. (1997). Smoothing the Hill estimator, *Advances in Applied Probability* **29**(1): 271–293.

[217] Reynkens, T., Verbelen, R., Beirlant, J., Antonio, K. (2017). Modelling censored losses using splicing: a global fit strategy with mixed Erlang and extreme value distributions, *Insurance: Mathematics and Economics* **77**: 65–77.

[218] Rio, E. (2017). *Asymptotic Theory of Weakly Dependent Random Processes*, Springer-Verlag, Berlin-Heidelberg.

[219] Ripley, B.D., Solomon, P.J. (1994). A note on Australian AIDS survival, University of Adelaide Department of Statistics Research Report 94/3.

[220] Rootzén, H., Leadbetter, M.R., de Haan, L. (1998). On the distribution of tail array sums for strongly mixing stationary sequences, *Annals of Applied Probability* **8**(3): 868–885.

[221] Rootzén, H., Tajvidi, N. (1997). Extreme value statistics and wind storm losses: a case study, *Scandinavian Actuarial Journal* **1**: 70–94.

[222] Ruymgaart, F.H. (1998). A note on weak convergence of density estimators in Hilbert spaces, *Statistics* **30**(4): 331–343.

[223] Sayah, A., Yahia, D., Brahimi, B. (2014). On robust tail index estimation under random censorship, *Afrika Statistika* **9**(1): 671–683.

[224] Schmidt, R., Stadtmüller, U. (2006). Non-parametric estimation of tail dependence, *Scandinavian Journal of Statistics* **33**(2): 307–335.

[225] Schultze, J., Steinebach, J. (1996). On least squares estimates of an exponential tail coefficient, *Statistics and Decisions* **14**(4): 353–372.

[226] Schulze Waltrup, L., Sobotka, F., Kneib, T., Kauermann, G. (2015). Expectile and quantile regression - David and Goliath?, *Statistical Modelling* **15**(5): 433–456.

[227] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

[228] Serfling, R. (2002). Quantile functions for multivariate analysis: approaches and applications, *Statistica Neerlandica* **56**(2): 214–232.

[229] Shao, X. (2010). A self-normalized approach to confidence interval construction in time series, *Journal of the Royal Statistical Society: Series B* **72**(3): 343–366.

[230] Shorack, G.R. (1972). Functions of order statistics, *Annals of Mathematical Statistics* **43**(2): 412–427.

[231] Shorack, G.R. (2000). *Probability for Statisticians*, Springer, New York.

[232] Silverman, B.V. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives, *Annals of Statistics* **6**(1): 177–184.

[233] Small, C.G. (1990). A survey of multidimensional medians, *International Statistical Review* **58**(3): 263–277.

[234] Smith, R. L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone (with discussion), *Statistical Science* **4**(4): 367–393.

[235] Sobotka, F., Kneib, T. (2012). Geoadditive expectile regression, *Computational Statistics & Data Analysis* **56**(4): 755–767.

[236] Stute, W. (1982). A law of the logarithm for kernel density estimators, *Annals of Probability* **10**(2): 414–422.

[237] Stute, W. (1984). The oscillation behavior of empirical processes: the multivariate case, *Annals of Probability* **12**(2): 361–379.

[238] Stute, W. (1993). Almost sure representations of the product-limit estimator for truncated data, *Annals of Statistics* **21**(1): 146–156.

[239] Takahashi, R. (1987). Some properties of multivariate extreme value distributions and multivariate tail equivalence, *Annals of the Institute of Statistical Mathematics* **39**(3): 637–647.

[240] Takahashi, R. (1988). Characterizations of a multivariate extreme value distribution, *Advances in Applied Probability* **20**(1): 235–236.

[241] Tang, T.-L., Shieh, S.-J. (2006). Long memory in stock index futures markets: A value-at-risk approach, *Physica A: Statistical Mechanics and its Applications* **366**: 437–448.

[242] Taylor, J.W. (2008). Estimating value at risk and expected shortfall using expectiles, *Journal of Financial Econometrics* **6**(2): 231–252.

[243] Tsiatis, A. (1975). A non-identifiability aspect of the problem of competing risks, *Proceedings of the National Academy of Sciences* **72**(1): 20–22.

[244] Tsukahara, H. (2009). Some properties of distortion risk measures, in: Kusuoka, S., Maruyama, T. (eds.) *Advances in Mathematical Economics* **12**: 153–166.

[245] Utev, S.A. (1990). On the central limit theorem for $\varphi-$mixing arrays of random variables, *Theory of Probability & Its Applications* **35**(1): 131–139.

[246] Vanem, E. (2015). Uncertainties in extreme value modelling of wave data in a climate change perspective, *Journal of Ocean Engineering and Marine Energy* **1**(4): 339–359.

[247] van der Vaart, A.W. (1998). *Asymptotic Statistics*, Cambridge University Press, Cambridge.

[248] van der Vaart, A.W., Wellner, J.A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*, Springer–Verlag, New York.

[249] Vandewalle, B., Beirlant, J. (2006). On univariate extreme value statistics and the estimation of reinsurance premiums, *Insurance: Mathematics and Economics* **38**(3): 441–459.

[250] Wang, S. (1996). Premium calculation by transforming the layer premium density, *ASTIN Bulletin* **26**(1): 71–92.

[251] Wang, S.S. (2000). A class of distortion operators for pricing financial and insurance risks, *Journal of Risk and Insurance* **67**(1): 15–36.

[252] Wang, H.J., Li, D., He, X. (2012). Estimation of high conditional quantiles for heavy-tailed distributions, *Journal of the American Statistical Association* **107**(500): 1453–1464.

[253] Wang, H.J., Li, D. (2013). Estimation of extreme conditional quantiles through power transformation, *Journal of the American Statistical Association* **108**(503): 1062–1074.

[254] Wang, H., Tsai, C.L. (2009). Tail index regression, *Journal of the American Statistical Association* **104**(487): 1233–1240.

[255] Wang, R., Ziegel, J.F. (2015). Elicitable distortion risk measures: A concise proof, *Statistics & Probability Letters* **100**: 172–175.

[256] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations, *Journal of the American Statistical Association* **73**(364): 812–815.

[257] Wirch, J.L., Hardy, M.R. (1999). A synthesis of risk measures for capital adequacy, *Insurance: Mathematics and Economics* **25**(3): 337–347.

[258] Wirch, J.L., Hardy, M. R. (2002). Distortion risk measures: Coherence and stochastic dominance, in *International Congress on Insurance: Mathematics and Economics*, Lisbon.

[259] Woodroofe, M. (1985). Estimating a distribution function with truncated data, *Annals of Statistics* **13**(1): 163–177.

[260] Worms, J., Worms, R. (2014). New estimators of the extreme value index under random right censoring, for heavy-tailed distributions, *Extremes* **17**(2): 337–358.

[261] Worms, J., Worms, R. (2016). A Lynden-Bell integral estimator for extremes of randomly truncated data, *Statistics & Probability Letters* **109**: 106–117.

[262] Yeo, I.-K., Johnson, R.A. (2000). A new family of power transformations to improve normality or symmetry, *Biometrika* **87**(4): 954–959.

[263] Yeo, I.-K., Johnson, R.A. (2014). An empirical characteristic function approach to selecting a transformation to symmetry, in Lahiri, S., Schick, A., SenGupta, A., Sriram, T. (eds.) Contemporary Developments in Statistical Theory, *Springer Proceedings in Mathematics & Statistics* **68**: 191–202.

[264] Youssef, M., Belkacem, L., Mokni, K. (2015). Value-at-Risk estimation of energy commodities: A long-memory GARCH-EVT approach, *Energy Economics* **51**: 99–110.

[265] Ziegel, J.F. (2016). Coherence and elicitability, *Mathematical Finance* **26**(4): 901–918.