DISSERTATION

HEAVY TAIL ANALYSIS FOR FUNCTIONAL AND INTERNET ANOMALY DATA

Submitted by

Mihyun Kim

Department of Statistics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2021

Doctoral Committee:

Advisor: Piotr Kokoszka

Daniel Cooley
Mary Meyer
Olivier Pinaud

ABSTRACT

HEAVY TAIL ANALYSIS FOR FUNCTIONAL AND INTERNET ANOMALY DATA

This dissertation is concerned with the asymptotic theory of statistical tools used in extreme value analysis of functional data and internet anomaly data. More specifically, we study four problems associated with analyzing the tail behavior of functional principal component scores in functional data and interarrival times of internet traffic anomalies, which are available only with a round-off error. The first problem we consider is the estimation of the tail index of scores in functional data. We employ the Hill estimator for the tail index estimation and derive conditions under which the Hill estimator computed from the sample scores is consistent for the tail index of the unobservable population scores. The second problem studies the dependence between extremal values of functional scores using the *extremal dependence measure* (EDM). After extending the EDM defined for positive bivariate observations to multivariate observations, we study conditions guaranteeing that a suitable estimator of the EDM based on these scores converges to the population EDM and is asymptotically normal. The third and last problems investigate the asymptotic and finite sample behavior of the Hill estimator applied to heavy–tailed data contaminated by errors. For the third one, we show that for time series models often used in practice, whose non–contaminated marginal distributions are regularly varying, the Hill estimator is consistent. For the last one, we formulate conditions on the errors under which the Hill and Harmonic Moment estimators applied to i.i.d. data continue to be asymptotically normal. The results of large and finite sample investigations are applied to internet anomaly data.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Professor Piotr Kokoszka, for all he has taught me about functional data analysis and extreme value theory, and also for all of his encouragement, support, and patience over the last five years. I have learned that "a little progress each day adds up to big results," from working with him, and making such progress was a full of joy. Also, he has talked to me about his attitude towards life, “live everyday as a vacation.” This quote has dramatically, positively changed my PhD life, and I still try to do things that give me a full of fun, inspiration, and fulfillment everyday.

I would also like to thank my committee members, Professor Dan Cooley, Professor Mary Meyer, and Professor Olivier Pinaud for their assistance, guidance, and professional discussions throughout my dissertation. To the faculty, staff, and students in the statistics department at Colorado State University, I am grateful for the opportunity to study in a wealthy research environment with great minds who are passionate about their work. Special thanks must go to Professor Jana Anderson, who gave me opportunities to run and teach a graduate level online course. I found my passion for teaching and gained confidence to pursue my career goal from those teaching opportunities. I would also like to thank Professor Julia Sharp for all she has taught me about collaborative scientific research. I am thankful for her willingness to work with me and for her patience over the past two years. Lastly, I would like to thank Professor Wen Zhou, who encouraged me to start a PhD program. He taught statistical machine learning, which was so fun and got me motivated and inspired.

I would like to thank my parents and my parents-in-law for their love and unconditional support. Most of all, I would like to thank my loving and supportive husband, Youngseok, for walking with me through this PhD journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Heavy-tailed phenomena occur in a variety of fields and have been studied for several decades. Benoit Mandelbrot, see e.g. [1], referred to them as the “Noah effect"; an observation can occur after surviving a flood, long after all other observations have perished. The origins of the field of extreme value theory are actually in mathematical research motivated by flood prevention engineering problems, see e.g. Chapter 1 of [2]. The currently used theory began to take shape in early 1970s, but many fundamental problems, motivated by mathematical curiosity, have been solved much earlier, with one of the most spectacular achievements being the solution to the problem of the extremal domains of attraction obtained by [3].

In this dissertation, we derive asymptotic properties of statistical tools for the analysis of heavy–tailed behavior observed in functional data and internet traffic anomaly data. In certain applications, most notably in finance, functional principal component scores in functional data exhibit heavy tails. Heavy–tailed characteristics are also found in anomalies arrival times. Quantifying the tail behavior of such data is needed to further apply methods of extreme value theory.

This dissertation consists of four main chapters. Chapters 2 and 3 make a contribution at the nexus of functional data analysis and heavy–tail analysis. In Chapters 4 and 5 we study the Hill estimator applied to observations contaminated by some errors. We now outline the main ideas of each chapter.

In Chapter 2, we study the tail behavior of functional principal component scores that are commonly used to reduce mathematically infinitely dimensional functional data to finite dimensional vectors. To estimate the tail index of the scores, we consider the Hill estimator that is the most commonly used tool for inference on the tail index. We derive conditions under which the Hill estimator computed from the sample scores is consistent for the tail index of the unobservable population scores.

In Chapter 3, we assess extremal dependence between functional principal component scores by means of the *extremal dependence measure* (EDM). Estimated scores form a triangular array of dependent random variables. We derive conditions under which an estimator of the EDM based on these scores is asymptotically normal. These conditions are completely different from those encountered in the second-order theory of functional data. They are formulated within the framework of functional regular variation. Large sample theory is complemented by an application to intraday return curves for certain stocks and by a simulation study.

Chapter 4 is concerned with the estimation of the tail index of heavy–tailed time series contaminated by measurement or other errors. We investigate asymptotic and finite sample properties of the Hill estimator applied time series observed with errors. We derive conditions under which the effect of the errors is asymptotically negligible. We show by means of a simulation study that the Hill estimator is asymptotically robust to relatively large errors.

In Chapter 5, we establish the asymptotic normality of the Hill estimator and of the harmonic moment estimator applied to heavy-tailed observations with measurement errors. The latter estimator is actually a class of estimators generalizing to the Hill estimator. Essentially, the only assumption on the errors needed to obtain the asymptotic normality is that they have lighter tails than the underlying unobservable process. The interarrival times of anomalies in a backbone internet network, computed with a roundoff error, are used in an application study in Chapters 4 and 5.

In the remainder of this chapter, we give a general introduction to the Hill estimator that we will see in Chapters 2, 4, and 5, and to the EDM that will appear in Chapter 3. We also discuss functional principal component analysis, to the extent needed to understand Chapters 2 and 3. It is not possible to explain all relevant concepts fully in a brief account, but we attempt to provide enough information to make this dissertation reasonably self-contained. As we introduce the required concepts, we give references to monographs that provide extensive, in-depth treatments.

1.1 Hill estimator and the EDM

In this section, we discuss two statistical tools used in heavy-tail analysis, the Hill estimator and the EDM, that this dissertation focuses on. We start by introducing the theory of regular variation, which provides a suitable mathematical framework. More detailed background is provided in Chapters 2 and 6 of [4].

A function $U : \mathbb{R}_+ \to \mathbb{R}_+$ is regularly varying with index $\alpha > 0$, $U \in RV_{-\alpha}$, if for any $x > 0$,

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

If the function U satisfies the above condition, we write $U \in RV_{-\alpha}$. In applications, we often consider tail probabilities $U(t) = P(X > t)$, where the random variable X has the same distribution as observations of an underlying random process. The tail index α characterizes the tail behavior of the process.

Suppose $X_1, \ldots, X_n$ are independent, nonnegative random variables with a common marginal distribution function F_X, which has regularly varying tail probabilities:

$$\bar{F}_X = 1 - F_X = P(X_i > \cdot) \in RV_{-\alpha}.$$

There is an extensive body of work on estimating the tail index α. One of the well–known estimators is the Hill estimator defined as

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

with the convention that $X_{(i)}$ is the i–th largest order statistic. It uses only the k largest observations, which intuitively makes sense because any inference on the tail index should be based on the extreme observations. The asymptotic properties of the Hill estimator have been studied as the number of upper order statistics, k, tends to infinity with the sample size n, in such a way that $k/n \to 0$.

The EDM quantifies the tendency of large values between two components to occur simultaneously. Its construction is based on the theory of heavy–tailed regularly varying random vectors. Thus, we first introduce multivariate regular variation and then discuss the EDM.

There are various equivalent formulations of multivariate regular variation, see Theorem 6.1 of [4]. We present here the definition with a polar coordinate representation. Fix a norm $\|\cdot\|$ in $\mathbb{R}^d$, and let $\mathbb{S}_+^d$ be the unit sphere in the nonnegative orthant in $\mathbb{R}^d$. A d–dimensional random vector $\mathbf{Z} = [Z_1, \dots, Z_d]^\top$ is regularly varying if and only if there exists a sequence $b_R(n) \to \infty$ and an angular probability measure Γ on $\mathbb{S}_+^d$ such that for $(R, \mathbf{\Theta}) = (\|\mathbf{Z}\|, \mathbf{Z}/\|\mathbf{Z}\|)$,

$$nP\left(\left(\frac{R}{b_R(n)}, \mathbf{\Theta}\right) \in \cdot\right) \xrightarrow{v} c\nu_\alpha \times \Gamma, \quad \text{in } M_+((0,\infty] \times \mathbb{S}_+^d),$$

where $\nu_\alpha(x,\infty] = x^{-\alpha}$ and $c = \nu\{\mathbf{x} : \|\mathbf{x}\| > 1\} > 0$. Basically, it can be interpreted that the radius R is involved with the tail index α and the angular measure Γ has all information on extremal dependence of the components in $\mathbf{Z}$.

Given a regularly varying nonnegative bivariate random vector $\mathbf{Z} = [Z_1, Z_2]^\top$, the EDM is defined by

$$\mathrm{EDM}(Z_1, Z_2) = \int_{\mathbb{S}_+^2} a_1 a_2 \Gamma(d\mathbf{a}),$$

see [5]. The EDM takes the minimal value of zero iff the coordinates of $\mathbf{Z}$ are asymptotically independent. Also, if the norm is symmetric, the EDM achieves its maximal value iff the coordinates of $\mathbf{Z}$ exhibit asymptotic full dependence.

1.2 Functional principal component scores

Functional data analysis (FDA) is the statistical analysis of samples of curves, and it has become an active field of statistics over the last three decades. Functional principal component analysis is one of the most fundamental tools of FDA. It leads to an efficient representation of infinitely dimensional objects, like curves, by means of multivariate vectors of scores, e.g., [6], [7], [8], [9].

A finite number of these scores encode the shape of the curves and are amenable to various statistical procedures.

We assume that all curves can be viewed as observations from a functional space $L^2 = L^2(\mathcal{T})$, where the measure space $\mathcal{T}$ is such that $L^2(\mathcal{T})$, with the usual inner product, is a *separable* Hilbert space. Suppose $X_1, \ldots, X_n$ are mean zero iid functions in L^2 with $\mathrm{E}\,\|X_i\|^2 < \infty$. Then, by Karhunen–Loéve expansion, see e.g., Chapter 11 of [10],

$$X_i(t) = \sum_{j=1}^{\infty} \xi_{ij} v_j(t), \quad \xi_{ij} = \langle X_i, v_j \rangle\,, \quad \mathrm{E}\xi_{ij}^2 = \lambda_j,$$

where the v_j are functional principal components (FPCs) and the ξ_{ij} are functional principal component scores.

The FPCs v_j and the eigenvalues λ_j are estimated by $\hat{v}_j$ and $\hat{\lambda}_j$, which are solutions to the equations

$$\widehat{C}(\hat{v}_j)(t) = \hat{\lambda}_j \hat{v}_j(t), \quad \text{for almost all } \ t \in \mathcal{T},$$

where $\widehat{C}$ is the sample covariance operator defined by

$$\widehat{C}(x)(t) = \frac{1}{n} \sum_{i=1}^{n} \langle X_i, x \rangle\, X_i, \quad x \in L^2.$$

The population scores ξ_{ij} are estimated by the sample scores $\hat{\xi}_{ij} = \langle X_i, \hat{v}_j \rangle$, which is the projection of X_i onto the estimated FPC $\hat{v}_j$. Thus, each $\hat{\xi}_{ij}$ quantifies the contribution of the curve $\hat{v}_j$ to the shape of the curve X_i. Furthermore, the vector of the sample scores, $[\hat{\xi}_{i1}, \ldots, \hat{\xi}_{ip}]^\top$, encodes the shape of X_i to a good approximation since each curve X_i can be approximated by a linear combination of a finite set of the estimated FPCs $\hat{v}_j$, i.e., $X_i(t) \approx \sum_{j=1}^{p} \hat{\xi}_{ij} \hat{v}_j(t)$.

Chapter 2

Hill estimator of projections of functional data on principal components

2.1 Introduction

A fundamental technique of functional data analysis is to replace infinite dimensional curves by coefficients of their projections onto suitable, fixed or data–driven, systems, e.g. [6], [7], [8], [9]. A finite number of these coefficients encode the shape of the curves and are amenable to various statistical procedures. The best systems are those that lead to low dimensional representations, and so provide the most efficient dimension reduction. Of these, the functional principal components (FPCs) have been most extensively used, with hundreds of papers dedicated to the various aspects of their theory and applications.

If $X_1, X_2, \ldots, X_n$ are mean zero iid functions in L^2 with $E\|X_i\|^2 < \infty$, then

$$X_i(t) = \sum_{j=1}^{\infty} \xi_{ij} v_j(t), \quad E\xi_{ij}^2 = \lambda_j, \tag{2.1}$$

where the v_j are the FPCs. The theory behind the Karhunen–Loéve expansion (2.1) is well–known, see e.g. Chapter 11 of [10], so we do not repeat the details.

The FPCs v_j and the eigenvalues λ_j are estimated by $\hat{v}_j$ and $\hat{\lambda}_j$ defined by

$$\int \hat{c}(t,s)\hat{v}_j(s)ds = \hat{\lambda}_j \hat{v}_j(t), \tag{2.2}$$

where $\hat{c}(t,s) = N^{-1}\sum_{n=1}^{N} X_n(t)X_n(s)$. In most inferential scenarios, replacing the v_j by the $\hat{v}_j$, and the λ_j by the $\hat{\lambda}_j$ is asymptotically negligible, see [11], [12], [13], [14], [15], among dozens of recent papers by other authors. Even though many different inferential problems have been considered, they are all related to some form of "second order inference", which utilizes estimators of means and covariance structures. In this paper, we study a totally different type of estimator, the Hill estimator, which is one of the most widely used tools of extreme value theory, see e.g. [16], [17] and [4]. Its definition is given in Section 2.2. We now describe a motivation for our study. We present an example based on financial data, but similar questions arise in the analysis of annual precipitation or other climate related curves.

Denote by $P_i(t)$ the price of an asset at time t of trading day i. For the assets we consider in our example, t is time in minutes between 9:30 and and 16:00 EST (NYSE opening times) rescaled to the unit interval $(0, 1)$. The intraday return curve on day i is defined by $X_i(t) = \log P_i(t) - \log P_i(0)$. In practice, $P_i(0)$ is the price after the first minute of trading. The curves R_i show how the return accumulates over the trading day, see e.g. Figure 1 in [18]; examples of are shown in Figure 2.1.

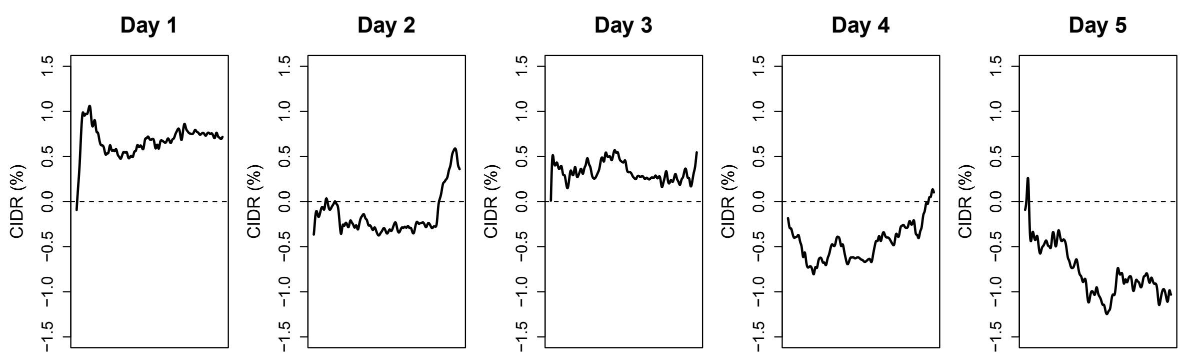

Figure 2.1: Five consecutive intraday return curves, Walmart stock.

The first three sample FPCs, $\hat{v}_1, \hat{v}_2, \hat{v}_3$, are shown in Figure 2.2. They are computed, using (2.2), from minute-by-minute Walmart returns form July 05, 2006 to Dec 30, 2011, $n = 1,378$ trading days. (This period is used for the other assets we consider.) The curves $\widehat{X}_i(t) = \sum_{j=1}^{3} \hat{\xi}_{ij}\hat{v}_j$, with the scores $\hat{\xi}_{ij} = \int X_i(t)\hat{v}_j(t)dt$, approximate the curves X_i well. Figure 2.3 shows the Hill plots of the sample score $\hat{\xi}_{ij}$ for two stocks and for $j = 1, 2, 3$. These plots are used to estimate the tail index α. Asymptotically, $\hat{\alpha}$ is obtained as the number of upper order statistics, k, tends to infinity with the sample size n, in such a way that $k/n \to 0$. In the plots, the values of k between 100 and

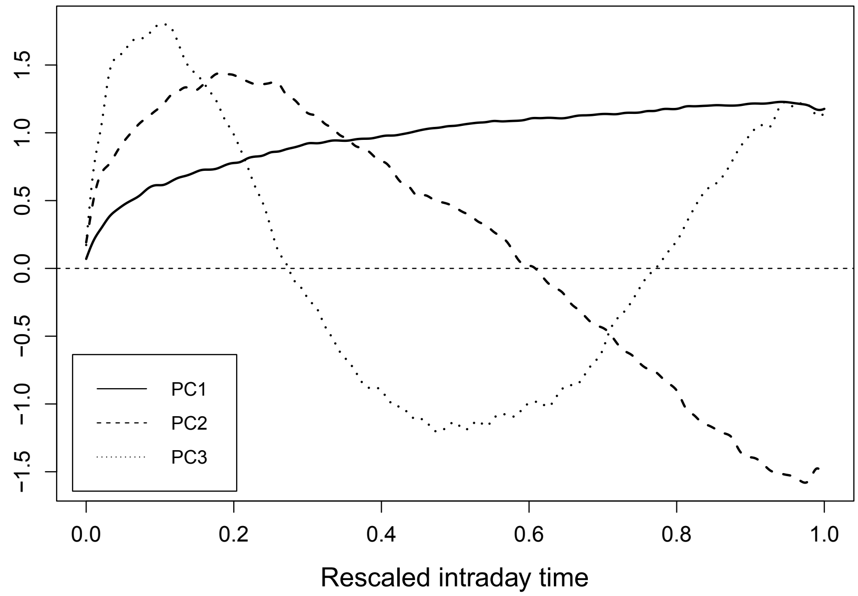

Figure 2.2: The first three sample FPCs of intraday returns on Walmart stock.

30 are used (recall, $n = 1,378$). These plots show that it is reasonable to assume that the scores have Pareto tails, with the tail index between 2 and 4.

It is important to emphasize that the Hill plots Figure 2.3 are computed using the samples score $\hat{\xi}_{ij} = \langle X_i, \hat{v}_j \rangle$, whereas the population parameter is the tail index α of the unobservable scores $\xi_{ij} = \langle X_i, v_j \rangle$. The question is if the Hill estimator based on the $\hat{\xi}_{ij}$ will be consistent for α, at least under some additional conditions, or if there is a systematic bias due to the effect of the estimation of the v_j by the $\hat{v}_j$. A problem of this type has not be studied. Consistency of the Hill estimator has been established in several settings, but always assuming that the regularly varying data are available.

Even for samples of iid positive random variables, the consistency of the Hill estimator is far from trivial. The first proof in the iid setting was developed by [19]. [20] introduced a general approach to establishing the consistency in case of dependent data, including both stationary times series and triangular arrays. Another extension was obtained by [21]. The sample scores do form a triangular array, but we were unable to adapt Hsing's method to accommodate the transition from

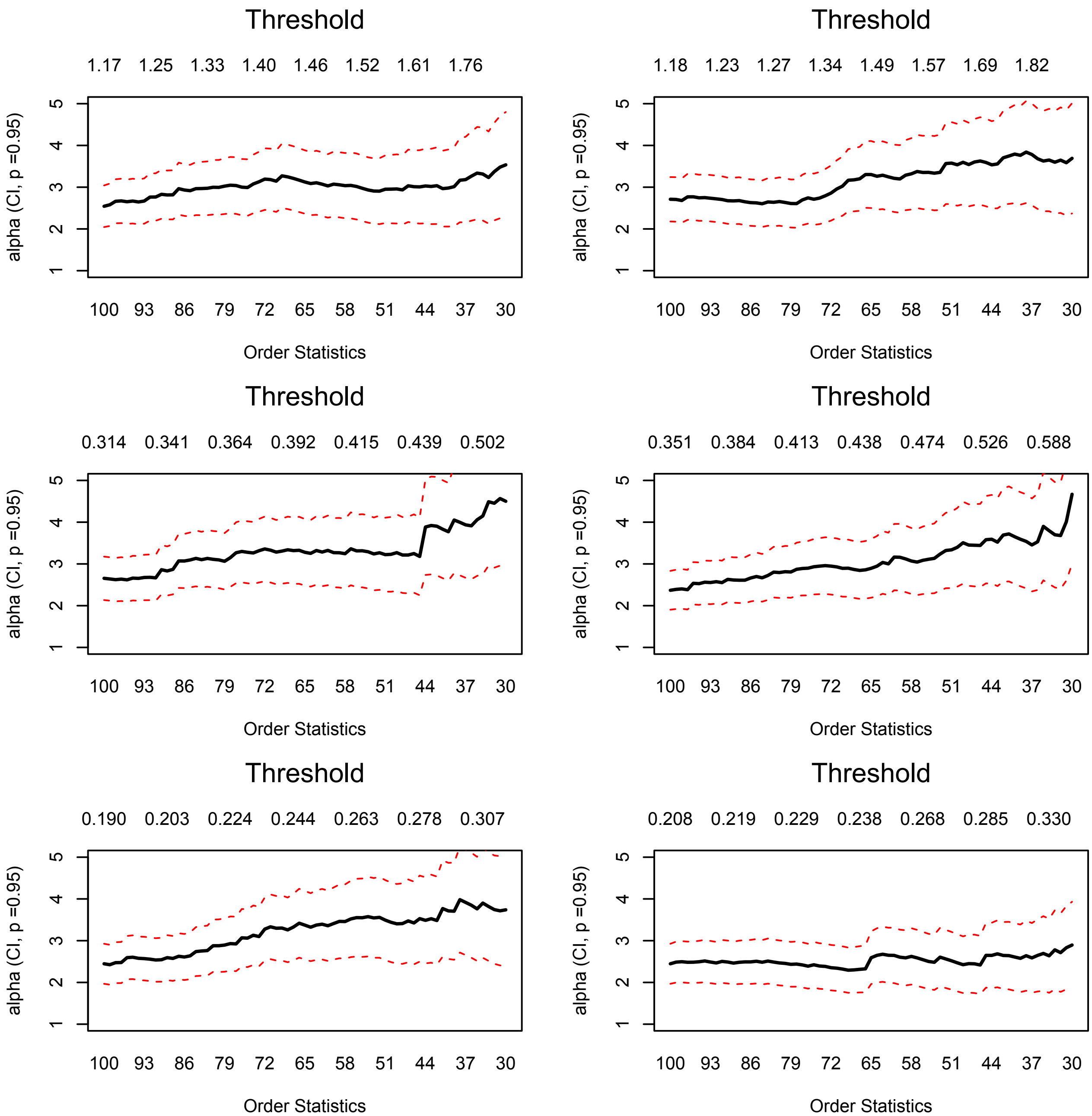

Figure 2.3: Hill plots for sample FPC scores for **Walmart** (left) and **IBM** (right). From top to bottom: levels $j = 1, 2, 3$.

the sample scores to the unobservable population scores. We developed an approach based on the vague convergence of radon measures, see [22], [4].

In Section 2.2 we introduce the framework and state our main result, Theorem 1, which is proven in Section 2.4, after some preparation in Section 2.3.

2.2 Assumptions and the main result

The most elegant, but in fact unnecessarily strong, assumption is that the function X whose copies X_i, $1 \leq i \leq n$, we observe is regularly varying in L^2. The space L^2 is infinitely dimensional and not locally compact, so we cannot define regular variation using the framework of [22], [4], but we can use a similar and more general framework of [23] who use M_0 convergence in place of the vague convergence in the Euclidean space with zero removed and compactified at infinity. Since we work with projections onto the real line, any definition of regular variation in L^2 which implies regular variation of these projections would work. According to [23] a function X in L^2 (or any Banach space) is regularly varying with index $\alpha > 0$ if

$$P(\|X\| > u) = u^{-\alpha} L(u) \tag{2.1}$$

and

$$\frac{P(u^{-1}X \in \cdot)}{P(\|X\| > u)} \xrightarrow{M_0} \mu(\cdot), \quad u \to \infty, \tag{2.2}$$

where μ is a non-null measure (exponent measure), and L is a slowly varying function. There are several equivalent definitions, see Chapter 2 of [24], which also contains all details.

Set

$$U(u) = P(|\langle X, v\rangle| > u), \quad \widehat{U}(u) = P(|\langle X, \hat{v}\rangle| > u),$$

where v is is one of the FPCs v_j in (2.1) and $\hat{v}$ its estimated defined by (2.2). The function U is regularly varying with index α, in the notation of [22], $U \in \mathrm{RV}_{-\alpha}$. To see this, consider the set

$\mathcal{A}_v = \{x : |\langle x, v\rangle| > 1\}$, and observe that $|\langle X, v\rangle| > u$ iff $u^{-1}X \in A_v$. By (2.1) and (2.2),

$$\frac{U(tu)}{U(u)} = \frac{P((tu)^{-1}X \in \mathcal{A}_v)}{P(\|X\| > tu)} \frac{P(\|X\| > tu)}{P(\|X\| > u)} \frac{P(\|X\| > u)}{P(u^{-1}X \in \mathcal{A}_v)} \to t^{-\alpha},$$

provided $\mu(\mathcal{A}_v) > 0$. It cannot be expected that $\widehat{U} \in \mathrm{RV}_{-\alpha}$; for a fixed n, $\hat{v}$ is a random function whose distribution will, in general, influence the distribution of $\langle X, \hat{v}\rangle$. Only some form of asymptotic regular variation can be expected because $\widehat{U}$ approaches U, in several ways, as $n \to \infty$.

The same argument shows that if $\mu(\{x : \langle x, v\rangle > 1\}) > 0$, then the function $U_+(u) = P(\langle x, v\rangle > u)$ is in $\mathrm{RV}_{-\alpha}$, and if $\mu(\{x : \langle x, v\rangle < -1\}) > 0$, then $U_-(u) = P(\langle x, v\rangle < -u)$ is in $\mathrm{RV}_{-\alpha}$. To avoid repetitions of almost identical statements, we focus in the following on the estimation of the tail index of the function U. We will work under the following assumption.

ASSUMPTION 1. The functions $X_1, X_2, \dots X_n$ are independent and have the same distribution as X. The function v is such that the function $U(u) = P(|\langle X, v\rangle| > u)$ is regularly varying with index $\alpha > 2$, $\alpha \neq 4$.

The assumption $\alpha > 2$ is needed because if $X \in \mathrm{RV}_{-\alpha}$ with $0 < \alpha < 2$, then, by (2.1), $E\|X\|^2 = \infty$, so the FPCs are not defined. If $\alpha = 2$, then either $E\|X\|^2 = \infty$ or $E\|X\|^2 < \infty$ are possible, and complex assumptions on the slowly varying functions L are needed to derive various rather technical results. We therefore assume $\alpha > 2$. Another phase transition occurs at $\alpha = 4$ separating, in a similar way, the cases with $E\|X\|^4 = \infty$ and $E\|X\|^4 < \infty$.

In our theory, the index α can depend on the direction v, but we do not emphasize it in our notation. We also note that even though the observed functions $X_1, X_2, \dots X_n$ are iid, the sample scores $\langle X_i, \hat{v}\rangle$ are no longer independent because $\hat{v}$ depends on all $X_1, X_2, \dots X_n$. They form a triangular array of dependent random variables, which are identically distributed for each fixed n. The Hill estimator must be based on the projections $\langle X_i, \hat{v}\rangle$. Before recalling its definition, we introduce the following random variables:

$$Y = |\langle X, v\rangle|, \quad \hat{Y} = |\langle X, \hat{v}\rangle|,$$

$$Y_i = |\langle X_i, v\rangle|, \quad \hat{Y}_i = |\langle X_i, \hat{v}\rangle|.$$

This allows us to define

$$H_{k,n} = \frac{1}{k}\sum_{i=1}^{k-1} \ln Y_{(i)} - \ln Y_{(k)}, \quad \widehat{H}_{k,n} = \frac{1}{k}\sum_{i=1}^{k-1} \ln \hat{Y}_{(i)} - \ln \hat{Y}_{(k)},$$

with the convention that $Y_{(n)}$ is the largest order statistic. In the functional data context, $H_{k,n}$ is an infeasible Hill estimator because the FPC v is not observed; $\widehat{H}_{k,n}$ is the Hill estimator that can be actually computed. We want to establish condition under which it converges in probability to α^{-1}, where α is the index of regular variation of Y.

We further define

$$1 - F(u) = P(Y > u) = U(u), \quad b(t) = F^{\leftarrow}\left(1 - \frac{1}{t}\right).$$

We will use the representation

$$b(t) = t^{1/\alpha} L_b(t), \tag{2.3}$$

where L_b is a slowly varying function.

The approach in Chapter 4 of [4] is based on vague convergence to the measure on the positive half–line, which is defined by

$$\nu_\alpha(x, \infty] = x^{-\alpha}, \;\; x > 0.$$

Our approach involves a sequence of "increasingly empirical" measures, with only the last one being observable. We set

$$\nu_n = \frac{1}{k}\sum_{i=1}^{n} I_{Y_i/b(\frac{n}{k})}, \quad \nu_n^\star = \frac{1}{k}\sum_{i=1}^{n} I_{Y_i/Y_{(k)}}, \quad \nu_n^\dagger = \frac{1}{k}\sum_{i=1}^{n} I_{\hat{Y}_i/b(\frac{n}{k})}, \quad \hat{\nu}_n = \frac{1}{k}\sum_{i=1}^{n} I_{\hat{Y}_i/\hat{Y}_{(k)}}.$$

Any argument must involve some bounds on a suitable distance between $\hat{Y}_i$ and Y_i. We now explain what can be assumed. If v is the jth eigenfunction of C, the population covariance operator, and $\hat{v}_j$ is the jth eigenfunction of $\widehat{C}$, then (see e.g. Lemma 2.3 in [8]),

$$\|\hat{v} - v\| \le \frac{2\sqrt{2}}{d_j} \|\widehat{C} - C\|_{\mathcal{L}}, \tag{2.4}$$

where $d_1 = \lambda_1 - \lambda_2, \quad d_j = \min\{\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1}\}$. Assuming that the eigenfunctions λ_j of C are such that for the j of interest $d_j > 0$, we obtain

$$\|\hat{v} - v\| \le A_v \|\widehat{C} - C\|_{\mathcal{L}}. \tag{2.5}$$

Since

$$|\hat{Y}_i - Y_i| \le |\langle X_i, \hat{v} - v\rangle| \le \|X_i\| \|\hat{v} - v\|,$$

we conclude from (2.5) that

$$|\hat{Y}_i - Y_i| \le A_v \|X_i\| \|\widehat{C} - C\|_{\mathcal{L}}. \tag{2.6}$$

If $\alpha > 4$, then, see e.g. Theorem 2.5 in [8],

$$E\|\widehat{C} - C\|_{\mathcal{L}}^2 = O(n^{-1}). \tag{2.7}$$

The case of regularly varying X with tail index $\alpha \in (2, 4)$ is studied in [25]. Under week conditions, relation (2.7) must be replaced by

$$E\|\widehat{C} - C\|_{\mathcal{L}}^{\beta} \le L_\beta(n) n^{-\beta(1-2/\alpha)}, \quad \forall\, \beta \in (0, \alpha/2), \tag{2.8}$$

where L_β is a slowly varying function. For a fixed α, the strongest bound is obtained as $\beta \nearrow \alpha/2$, in which case $\beta(1 - 2/\alpha) \nearrow \alpha/2 - 1$. As $\alpha \nearrow 4$ and $\beta \nearrow \alpha/2$, relation (2.8) thus approaches, in a heuristic sense, relation (2.7). It is enough to impose a slightly weaker, but more convenient, condition:

$$E\|\widehat{C} - C\|_{\mathcal{L}}^{\beta} = O\left(n^{-\kappa}\right), \quad \forall\, \beta \in \left(1, \frac{\alpha}{2}\right), \forall\, \kappa \in \left(0, \beta\left(1 - \frac{2}{\alpha}\right)\right). \tag{2.9}$$

The above discussion shows that the following Assumption 2 basically always holds as long as $d_j > 0$ in (2.4). We formulated it for ease of reference, and to emphasize that only certain properties of the sample covariance operator $\widehat{C}$ are used; $\widehat{C}$ could, in principle, be a different estimator of C, which has those properties.

ASSUMPTION 2. Relation (2.5) holds. The estimator $\widehat{C}$ satisfies (2.7) if $\alpha > 4$ and (2.9) if $\alpha \in (2, 4)$.

Since the Y_i are iid and in $\mathrm{RV}_{-\alpha}$, the only conditions needed to ensure that $H_{k,n} \stackrel{P}{\to} \alpha^{-1}$ are $k = k(n) \to \infty$ and $k/n \to 0$, as $n \to \infty$. In our setting, we want to estimate the tail index of unobservable random variables Y_i based on their observed approximations $\hat{Y}_i$. It can be expected that the rate of the approximation will impose additional conditions on the rate at which k tends to infinity with n. A sufficient condition is formulated in Assumption 3 below.

Define the function

$$\gamma(\alpha) = \begin{cases} \dfrac{\alpha - 2}{2\alpha - 2}, & \alpha \in (4, \infty), \\[2ex] \dfrac{1}{\alpha - 1}, & \alpha \in (3, 4], \\[2ex] 2 - \dfrac{\alpha}{2}, & \alpha \in (2, 3]. \end{cases} \tag{2.10}$$

Observe that $\gamma(\cdot)$ is continuous at $\alpha = 4$ with $\gamma(4) = 1/3$, and at $\alpha = 3$ with $\gamma(3) = 1/2$. It is increasing on $(4, \infty)$ with $\lim_{\alpha \nearrow \infty} \gamma(\alpha) = \frac{1}{2}$ and decreasing on $(2, 4)$ with $\lim_{\alpha \searrow 2} \gamma(\alpha) = 1$.

We will write $k >> n^{\gamma}$, for some $\gamma \in (0, 1)$, if $k/n^{\gamma} \to \infty$.

ASSUMPTION 3. We assume that $k >> n^{\gamma}$ for some $\gamma \in (\gamma(\alpha), 1)$, with $\gamma(\alpha)$ defined in (2.10).

According to Assumption 3, as $\alpha \searrow 2$, the order of k approaches n. One can say that as the value of α approaches the smallest possible value for which the functional principal components exit, only the very largest observations must be used to ensure the consistency of the Hill estimator.

THEOREM 1. *Suppose Assumptions 1, 2 and 3 hold. Then* $\widehat{H}_{k,n} \stackrel{P}{\to} \alpha^{-1}$.

While Theorem 1 is formulated in the specific setting of projections of functional data onto population and estimated FPCs, it is hoped that the approach we develop will be, in general outlines, applicable to other contexts where the tail indexed must be inferred from approximations to unobserved data. For example, only $Y_i + \varepsilon_i$ with correlated errors ε_i may be observed. It is also hoped that the theory developed for the most commonly used Hill estimator may be used to guide similar developments for other estimators of the tail index.

2.3 Preliminary results

We collect in this section several results, none of which is particularly profound or difficult to prove, but put together they play an important role in the proof of Theorem 1. By placing them in a preparatory section, we will also avoid repeatedly distracting from the main flow of the argument in Section 2.4.

Following [4], denote by $M_+ = M_+(0, \infty]$ the space of Radon measures on $(0, \infty]$.

LEMMA 1. *The function h on M_+ defined by $h(\mu) = \mu(z, \infty]$ is continuous at ν_α.*

Proof. Suppose $\mu_n \to \nu_\alpha$. This implies that for any relatively compact B with $\nu_\alpha(\partial B) = 0$, $\mu_n(B) \to \nu_\alpha(B)$. Taking $B = (z, \infty]$, we obtain $h(\mu_n) = \mu_n(B) \to \nu_\alpha(B) = h(\nu_\alpha)$.

□

LEMMA 2. *The function h on M_+ defined by*

$$h(\mu) = \int_z^M \mu(x, \infty] x^{-1} dx$$

is continuous at ν_α.

Proof. Suppose $\mu_n \to \nu_\alpha$. By Lemma 1, for every $x > 0$, $\mu_n(x, \infty]x^{-1} \to \nu_\alpha(x, \infty]x^{-1}$. The convergence

$$\int_z^M \mu_n(x, \infty] x^{-1} dx \to \int_z^M \nu_\alpha(x, \infty] x^{-1} dx$$

follows from the dominated convergence theorem because for $x > z$ and sufficiently large n,

$$\mu_n(x, \infty] \le \mu_n(z, \infty] \le 2\nu_\alpha(z, \infty] = 2z^{-\alpha}.$$

□

The measure ν_n is a random element of M_+, ν_α its deterministic (constant) element. The following lemma follows from Theorem 4.1 and relation (4.21) in [4].

LEMMA 3. *In the space* $M_+(0, \infty]$, $\nu_n \xrightarrow{P} \nu_\alpha$ *and* $\nu_n^\star \xrightarrow{P} \nu_\alpha$.

The next lemma follows from relation (4.17) in the proof of Theorem 4.2 in [4].

LEMMA 4. *If the* Y_i *are iid and in* $\mathrm{RV}_{-\alpha}$, *then* $Y_{(k)}/b(\frac{n}{k}) \xrightarrow{P} 1$.

LEMMA 5. *For any* $a, b \ge 0$, $|[a \wedge 1] - [b \wedge 1]| \le |a - b|$.

Proof. There are four cases:

1) $a > 1, b > 1,\ |1 - 1| = 0 \le |a - b|$;

2) $a > 1, b \le 1,\ |1 - b| = 1 - b < a - b = |a - b|$;

3) $a \le 1, b > 1,\ |a - 1| = 1 - a < b - a = |a - b|$;

4) $a \le 1, b \le 1,\ |a - b| \le |a - b|$.

□

The following statements are proven is Section 3.4 of [22]. The metric ρ which compactifies $(0, \infty]$ at ∞ is

$$\rho(u, v) = \left| \frac{1}{u} - \frac{1}{v} \right|.$$

The distance between measures $\mu_1, \mu_1 \in M_+(0, \infty]$ is defined by

$$d(\mu_1, \mu_2) = \sum_{m=1}^{\infty} 2^{-m} \left\{ |\mu_1(f_m) - \mu_2(f_m)| \wedge 1 \right\}. \tag{2.1}$$

The functions $f_m \in C_K(0,\infty]$ are of the form

$$f(x) = 1 - [c\rho(x,B) \wedge 1], \tag{2.2}$$

for some $c > 0$ and relatively compact $B \subset (0,\infty]$.

LEMMA 6. *For any metric ρ and any set B, $|\rho(a_1,B) - \rho(a_2,B)| \le \rho(a_1,a_2)$.*

Proof. Recall that $\rho(a,B) = \inf_{b\in B}\rho(a,b)$. For any $b \in B$, $\rho(a_1,b) \le \rho(a_1,a_2) + \rho(a_2,b)$. Taking the infimum of the left–hand side, we obtain $\rho(a_1,B) \le \rho(a_1,a_2) + \rho(a_2,b)$. Taking the infimum of the right–hand side, we obtain $\rho(a_1,B) \le \rho(a_1,a_2) + \rho(a_2,B)$. Consequently, $\rho(a_1,B) - \rho(a_2,B) \le \rho(a_1,a_2)$. Switching a_1 and a_2, we obtain the claim.

□

LEMMA 7. *Suppose random variables $H_m(n)$, $m,n \ge 1$, satisfy $0 \le H_m(n) \le 1$ and $\forall\ m \ge 1,\quad H_m(n) \xrightarrow{P} 0,\quad \text{as } n \to \infty$. Then, $\sum_{m=1}^{\infty} 2^{-m} H_m(n) \xrightarrow{P} 0,\quad \text{as } n \to \infty$.*

Proof. Define

$$\begin{aligned} S(n) &= \sum_{m=1}^{\infty} 2^{-m} H_m(n) \\ &= \sum_{m\le M} 2^{-m} H_m(n) + \sum_{m>M} 2^{-m} H_m(n) \\ &=: S_M(n) + S_M^{\star}(n). \end{aligned}$$

Fix $\varepsilon > 0$ and observe that $P(S(n) > \varepsilon) \le P\left(S_M(n) > \varepsilon/2\right) + P\left(S_M^{\star}(n) > \varepsilon/2\right)$. Since $S_M^{\star}(n) \le 2^{-M}$, we can choose M so large that $P\left(S_M^{\star}(n) > \varepsilon/2\right) = 0$. For such a (fixed) M, $P(S(n) > \varepsilon) \le P\left(S_M(n) > \varepsilon/2\right) \to 0$.

□

2.4 Proof of Theorem 1

The proof of Theorem 1 is constructed from a series of results, of which Proposition 1 is the most prominent. To facilitate the understanding of the proofs of Proposition 1 and Theorem 1, we note that

If $\alpha \in (3,4)$, then $\frac{1}{\alpha-1} > 2 - \frac{\alpha}{2}$,

If $\alpha \in (2,3)$, then $\frac{1}{\alpha-1} < 2 - \frac{\alpha}{2}$.

We may thus write

$$\gamma(\alpha) = \max\left\{\frac{1}{\alpha-1},\ 2-\frac{\alpha}{2}\right\}, \quad \alpha \in (2,4]. \tag{2.1}$$

PROPOSITION 1. *Under the assumptions of Theorem 1,* $d(\nu_n^\dagger, \nu_n) \xrightarrow{P} 0$.

Proof. Since each function f_m in (2.1) has compact support in $(0,\infty]$, $s_m := \inf\{\mathrm{supp}(f_m)\} > 0$. Therefore

$$\begin{aligned}
|\nu_n^\dagger(f_m) - \nu_n(f_m)| &= \left| \int f_m d\nu_m^\dagger - \int f_m d\nu_n \right| \\
&\le \frac{1}{k}\sum_{i=1}^{n} \left| f_m\left(\frac{\hat{Y}_i}{b(n/k)}\right) - f_m\left(\frac{Y_i}{b(n/k)}\right)\right| \\
&= \frac{1}{k}\sum_{i\in\mathcal{I}_m} \left| f_m\left(\frac{\hat{Y}_i}{b(n/k)}\right) - f_m\left(\frac{Y_i}{b(n/k)}\right)\right|,
\end{aligned}$$

where

$$\mathcal{I}_m = \left\{ i \ge 1 :\ \hat{Y}_i > s_m b(n/k) \text{ or } Y_i > s_m b(n/k) \right\}.$$

Since each f_m is of the form (2.2), by Lemmas 5 and 6,

$$\begin{aligned}
|\nu_n^\dagger(f_m) - \nu_n(f_m)| &\le \frac{c_m}{k}\sum_{i\in\mathcal{I}_m} \left| \rho\left(\frac{\hat{Y}_i}{b(n/k)}, B_m\right) - \rho\left(\frac{Y_i}{b(n/k)}, B_m\right)\right| \\
&\le \frac{c_m}{k}\sum_{i\in\mathcal{I}_m} \left| \rho\left(\frac{\hat{Y}_i}{b(n/k)}, \frac{Y_i}{b(n/k)}\right)\right|.
\end{aligned}$$

The claim will thus follow from the convergence

$$\sum_{m=1}^{\infty} 2^{-m} \left\{ \left[\frac{c_m}{k} \sum_{i \in \mathcal{I}_m} \left| \frac{b(n/k)}{\hat{Y}_i} - \frac{b(n/k)}{Y_i} \right| \right] \wedge 1 \right\} \xrightarrow{P} 0,$$

which, in turn, by Lemma 7, will follow from

$$\frac{1}{k} \sum_{i \in \mathcal{I}(n)} \left| \frac{b(n/k)}{\hat{Y}_i} - \frac{b(n/k)}{Y_i} \right| \xrightarrow{P} 0, \tag{2.2}$$

where, for some $s^\star > 0$,

$$\mathcal{I}(n) = \left\{ i \geq 1 : \hat{Y}_i > s^\star b(n/k) \text{ or } Y_i > s^\star b(n/k) \right\}.$$

Set

$$\mathcal{I}^{(1)}(n) = \{ i \geq 1 : Y_i > s^\star b(n/k) \}, \quad \mathcal{I}^{(2)}(n) = \left\{ i \geq 1 : \hat{Y}_i > s^\star b(n/k) \right\}.$$

Relation (2.2) will follow once we have shown that for $g = 1$ and $g = 2$,

$$\frac{b(n/k)}{k} \sum_{i \in \mathcal{I}^{(g)}(n)} \frac{|\hat{Y}_i - Y_i|}{\hat{Y}_i Y_i} \xrightarrow{P} 0. \tag{2.3}$$

We verify (2.3) for $g = 1$. The argument for $g = 2$ is basically the same; the roles of $\hat{Y}_i$ and Y_i must be interchanged.

Fix $\varepsilon > 0$. First observe that

$$P \left(\frac{b(n/k)}{k} \sum_{i \in \mathcal{I}^{(1)}(n)} \frac{|\hat{Y}_i - Y_i|}{\hat{Y}_i Y_i} > \varepsilon \right) \leq P \left(G(n) > \varepsilon \right),$$

where

$$G(n) = \frac{1}{s^\star k} \sum_{i \in \mathcal{I}^{(1)}(n)} \frac{|\hat{Y}_i - Y_i|}{\hat{Y}_i}.$$

Next, use the decomposition

$$P\left(G(n) > \varepsilon\right) = P_1(n) + P_2(n),$$

with

$$P_1(n) = P\left(G(n) > \varepsilon,\ \exists\, i \in \mathcal{I}^{(1)}(n):\ \hat{Y}_i \le \frac{1}{2}s^\star b(n/k)\right);$$

$$P_2(n) = P\left(G(n) > \varepsilon,\ \forall\, i \in \mathcal{I}^{(1)}(n) : \hat{Y}_i > \frac{1}{2}s^\star b(n/k)\right).$$

Observe that

$$\begin{aligned} P_1(n) &\le P\left(\exists\, i \in \mathcal{I}^{(1)}(n):\ \hat{Y}_i \le \frac{1}{2}s^\star b(n/k)\right) \qquad (2.4)\\ &\le P\left(\exists\, i \le n:\ Y_i > s^\star b(n/k) \text{ and } \hat{Y}_i \le \frac{1}{2}s^\star b(n/k)\right)\\ &\le P\left(\exists\, i \le n:\ |\hat{Y}_i - Y_i| > \frac{1}{2}s^\star b(n/k)\right)\\ &= P\left(\max_{1\le i\le n} |\hat{Y}_i - Y_i| > \frac{1}{2}s^\star b(n/k)\right). \end{aligned}$$

By (2.6),

$$\begin{aligned} P_1(n) &\le P\left(A_v\|\widehat{C} - C\|_{\mathcal{L}} \max_{1\le i\le n} \|X_i\| > \frac{1}{2}s^\star b(n/k)\right) \qquad (2.5)\\ &\le \frac{2A_v}{s^\star b(n/k)} E\left[\|\widehat{C} - C\|_{\mathcal{L}} \max_{1\le i\le n} \|X_i\|\right]. \end{aligned}$$

We first consider the case of $\alpha > 4$. By (2.5) and (2.7),

$$\begin{aligned} P_1(n) &\le \frac{2A_v}{s^\star b(n/k)} \left\{E\|\widehat{C} - C\|_{\mathcal{L}}^2\right\}^{1/2} \left\{E \max_{1\le i\le n} \|X_i\|^2\right\}^{1/2}\\ &= O\left(\frac{1}{b(n/k)} n^{-1/2} n^{1/2}\right) = O\left(\frac{1}{b(n/k)}\right) = o(1). \end{aligned}$$

By Markov's inequality,

$$P_2(n) \leq P\left(\frac{2}{s^{\star 2}kb(n/k)}\sum_{i=1}^{n}|\hat{Y}_i - Y_i| > \varepsilon\right) \tag{2.6}$$
$$\leq \frac{2}{\varepsilon s^{\star 2}kb(n/k)}\sum_{i=1}^{n}E|\hat{Y}_i - Y_i|.$$

By (2.6) and (2.7),

$$E|\hat{Y}_i - Y_i| \leq A_v\left\{E\|X_i\|^2\right\}^{1/2}\left\{E\|\widehat{C} - C\|_{\mathcal{L}}^2\right\}^{1/2} = O(n^{-1/2}).$$

Therefore,

$$P_2(n) = O\left(\frac{n^{1/2}}{kb(n/k)}\right) = o(1).$$

The last equality follows from the assumption $k >> n^{\gamma(\alpha)}$ and (2.3).

Now consider the case of $\alpha \in (2,4)$. We first show that $P_1(n) = o(1)$. By (2.4), (2.6), and Markov's inequality

$$P_1(n) = O\left(\frac{1}{\sqrt{b(n/k)}}\right)E\left[\|\widehat{C} - C\|_{\mathcal{L}}^{1/2}\max_{1\leq i\leq n}\|X_i\|^{1/2}\right].$$

We apply Hölder's inequality with $p = 2\beta$ and $q = 2\beta/(2\beta - 1)$ to get

$$P_1(n) = O\left(\frac{1}{\sqrt{b(n/k)}}\right)\left\{E\|\widehat{C} - C\|_{\mathcal{L}}^{\beta}\right\}^{\frac{1}{2\beta}}\left\{E\max_{1\leq i\leq n}\|X_i\|^{\frac{\beta}{2\beta-1}}\right\}^{\frac{2\beta-1}{2\beta}}.$$

For the above bound to be effective, we need $E\|X_i\|^{\frac{\beta}{2\beta-1}} < \infty$, which is implied by $\frac{\beta}{2\beta-1} < \alpha$. Since $2\beta - 1 > 1$ and $\beta < \alpha/2$, this condition always holds. It therefore follows from (2.9) that

$$P_1(n) = O\left(\frac{1}{\sqrt{b(n/k)}}n^{-\frac{\kappa}{\beta}}n^{\frac{2\beta-1}{2\beta}}\right)$$

We can thus conclude that $P_1(n) = o(1)$, if there are β and κ such that $-\kappa + 2\beta - 1 < 0$. This is possible if $2\beta - 1 < \beta\left(1 - \frac{2}{\alpha}\right)$. The above condition can be equivalently stated as $\beta\left(1 - \frac{2}{\alpha}\right) < 1$. Since $\beta < \frac{\alpha}{2}$, $\beta\left(1 - \frac{2}{\alpha}\right) < \frac{\alpha}{2} - 1 < 1$ because $\alpha < 4$. This completes the verification of $P_1(n) = o(1)$ for $\alpha \in (2, 4)$.

To show that $P_2(n) = o(1)$, observe that by (2.6), Markov's inequality with $0 < r \leq 1$, and (2.6),

$$\begin{aligned}
P_2(n) &\leq P\left(\frac{2}{s^{\star 2} k b(n/k)} \sum_{i=1}^{n} |\hat{Y}_i - Y_i| > \varepsilon\right) \\
&= O\left(\frac{1}{k^r b^r(n/k)}\right) E\left(\sum_{i=1}^{n} |\hat{Y}_i - Y_i|\right)^r \\
&= O\left(\frac{n^r}{k^r b^r(n/k)}\right) E\left[\|\widehat{C} - C\|_{\mathcal{L}} \frac{1}{n} \sum_{i=1}^{n} \|X_i\|\right]^r.
\end{aligned}$$

Applying Hölder's inequality with $p = \beta/r$ and $q = \beta/(\beta - r)$, we obtain

$$E\left[\|\widehat{C} - C\|_{\mathcal{L}}^r \left(\frac{1}{n} \sum_{i=1}^{n} \|X_i\|\right)^r\right] \leq \left\{E\|\widehat{C} - C\|_{\mathcal{L}}^{\beta}\right\}^{r/\beta} \left\{E\left(\frac{1}{n} \sum_{i=1}^{n} \|X_i\|\right)^{rq}\right\}^{1/q}.$$

For $E\|X_i\|^{rq}$ to be finite, we need

$$rq = \frac{r\beta}{\beta - r} < \alpha. \tag{2.7}$$

Choosing

$$r = \frac{\beta}{\beta + 1} \tag{2.8}$$

implies $rq = 1$. We thus obtain, with r specified in (2.8),

$$P_2(n) = O\left(\frac{n^r}{k^r b^r(n/k)}\right) \{E\|X\|\}^{1/q} n^{-\kappa r/\beta} = O\left(\frac{n^{r - r\kappa/\beta}}{k^r b^r(n/k)}\right).$$

By (2.3), the claim $P_2(n) = o(1)$ will thus follow if $k >> n^{\gamma}$, where

$$\gamma = \frac{r - \frac{\kappa r}{\beta} - \frac{r}{\alpha}}{r - \frac{r}{\alpha}} = \frac{1 - \frac{1}{\alpha} - \frac{\kappa}{\beta}}{1 - \frac{1}{\alpha}}.$$

The exponent is smaller than 1, and attains its smallest value as $\frac{\kappa}{\beta}$ approaches its largest possible value, i.e. $1 - 2/\alpha$. It remains to observe that

$$\frac{1 - \frac{1}{\alpha} - \frac{\kappa}{\beta}}{1 - \frac{1}{\alpha}} = \frac{1}{\alpha - 1}, \quad \text{if } \frac{\kappa}{\beta} = 1 - \frac{2}{\alpha}.$$

$\square$

REMARK 2.4.1. The proof of Proposition 1, in the case $\alpha \in (2,4)$, is valid in (2.1) is replaced by $\gamma(\alpha) = (\alpha - 1)^{-1}$. Only the latter bound was used. The bound $2 - \alpha/2$ is needed in the proof of Theorem 1.

Using Lemma 3, we obtain the following corollary.

COROLLARY 1. *Under the assumptions of Theorem 1,* $\nu_n^\dagger \xrightarrow{P} \nu_\alpha$.

The arguments used in the proofs of Propositions 2 and 3 are similar to those developed in Sections 4.3. and 4.4 of [4].

PROPOSITION 2. *Under the assumptions of Theorem 1,*

$$\frac{\hat{Y}_{(k)}}{b(n/k)} \xrightarrow{P} 1.$$

Proof. Fix $\varepsilon > 0$ and set

$$P_+(n) = P\left(\frac{\hat{Y}_{(k)}}{b(n/k)} > 1 + \varepsilon\right), \quad P_-(n) = P\left(\frac{\hat{Y}_{(k)}}{b(n/k)} < 1 - \varepsilon\right).$$

Observe that

$$
\begin{aligned}
P_+(n) &= P\left(I_{\frac{\hat{Y}_{(k)}}{b(n/k)}}(1+\varepsilon,\infty] = 1\right) \\
&\le P\left(\sum_{i=1}^{n} I_{\frac{\hat{Y}_{(i)}}{b(n/k)}}(1+\varepsilon,\infty] \ge k\right) \\
&= P\left(\frac{1}{k}\sum_{i=1}^{n} I_{\frac{\hat{Y}_{(i)}}{b(n/k)}}(1+\varepsilon,\infty] \ge 1\right) \\
&= P\left(\nu_n^{\dagger}(1+\varepsilon,\infty] \ge 1\right).
\end{aligned}
$$

A similar argument shows that $P_-(n) \le P\left(\nu_n^{\dagger}(1-\varepsilon,\infty] < 1\right)$. The claim follows because by Corollary 1 and Lemma 1,

$$
\begin{aligned}
\nu_n^{\dagger}(1+\varepsilon,\infty] &\xrightarrow{P} \nu_\alpha(1+\varepsilon,\infty] = (1+\varepsilon)^{-\alpha} < 1; \\
\nu_n^{\dagger}(1-\varepsilon,\infty] &\xrightarrow{P} \nu_\alpha(1-\varepsilon,\infty] = (1-\varepsilon)^{-\alpha} > 1.
\end{aligned}
$$

$\square$

PROPOSITION 3. *Under the assumptions of Theorem 1,* $\hat{\nu}_n \xrightarrow{P} \nu_\alpha$.

Proof. Consider the map $T : M_+ \times (0,\infty) \to M_+$ defined by

$$
T(\mu, x)(A) = \mu(xA), \quad \text{for Borel } A \subset (0,\infty].
$$

[4], pp. 83–84 shows that T is continuous. Observe that

$$
T\left(\nu_n^{\dagger}, \frac{\hat{Y}_{(k)}}{b(n/k)}\right) = \hat{\nu}_n, \quad T(\nu_\alpha, 1) = \nu_\alpha.
$$

The claim thus follows because by Corollary 1 and Proposition 2,

$$\left(\nu_n^\dagger,\ \frac{\hat{Y}_{(k)}}{b(n/k)}\right) \xrightarrow{P} (\nu_\alpha, 1) \quad \text{in } M_+ \times (0, \infty).$$

□

The following lemma may be of independent interest and more general utility.

LEMMA 1. *Suppose* $y \mapsto P(Y > y) \in \mathrm{RV}_{-\alpha}$ *for some* $\alpha > 0$. *Then,*

$$\lim_{z\to\infty} \limsup_{t\to\infty} \int_z^\infty tP(Y > xb(t))x^{-1}dx = 0.$$

Proof. The function $b(\cdot)$ is defined by $P(Y > b(t)) = t^{-1}$. We know that $b(\cdot) \in \mathrm{RV}_{1/\alpha}$ and

$$\lim_{t\to\infty} tP(Y > xb(t)) = x^{-\alpha}, \quad x > 0. \tag{2.9}$$

Set $f_t(x) = tP(Y > xb(t))x^{-1}$. We want to show

$$\lim_{z\to\infty} \limsup_{t\to\infty} \int_z^\infty f_t(x)dx = 0.$$

By (2.9), $\forall\, x > 0 \quad f_t(x) \to x^{-\alpha-1}$, as $t \to \infty$. To conclude that

$$\int_z^\infty f_t(x)dx \to \int_z^\infty x^{-\alpha-1}dx, \quad \text{as } t \to \infty,$$

we must find a function g such that for $t > t_0$,

$$f_t(x) \le g(x) \quad \text{and} \quad \int_z^\infty g(x)dx < \infty.$$

Set $U(y) = P(Y > y)$. Potter bounds state that $\forall\, \delta > 0$, $\exists\, u_0$, $\forall\, u \ge u_0$, $\forall\, y \ge 1$,

$$(1-\delta)y^{-\alpha-\delta} \le \frac{U(yu)}{U(u)} \le (1+\delta)y^{-\alpha+\delta}.$$

Since $b(t) \to \infty$ as $t \to \infty$, $\exists\, t_0, \forall\, t > t_0$, $U(xb(t)) \le (1+\delta)x^{-\alpha+\delta}U(b(t))$. Since $U(b(t)) = 1/t$, we obtain, for $t \ge t_0$, $f_t(x) = tU(xb(t))x^{-1} \le (1+\delta)x^{-\alpha+\delta-1} =: g(x)$. The function g is integrable if $\delta < \alpha$.

$\square$

PROOF OF THEOREM 1: Since

$$\widehat{H}_{k,n} = \int_1^\infty \hat{\nu}_n(x,\infty]x^{-1}dx,$$

we must show that

$$\int_1^\infty \hat{\nu}_n(x,\infty]x^{-1}dx \xrightarrow{P} \int_1^\infty \nu_\alpha(x,\infty]x^{-1}dx = \alpha^{-1}.$$

The verification is based on the commonly used truncation argument, Theorem 3.2 in [26], also stated as Theorem 3.5 in [4]. Set

$$\begin{aligned} V_n &= \int_1^\infty \hat{\nu}_n(x,\infty]x^{-1}dx, \quad V = \int_1^\infty \nu_\alpha(x,\infty]x^{-1}dx; \\ V_n^{(M)} &= \int_1^M \hat{\nu}_n(x,\infty]x^{-1}dx, \quad V^{(M)} = \int_1^M \nu_\alpha(x,\infty]x^{-1}dx. \end{aligned}$$

To establish the desired convergence $V_n \xrightarrow{P} V$, equivalently $V_n \xrightarrow{d} V$, we must verify that

$$\forall\, M > 1, \quad V_n^{(M)} \xrightarrow{d} V_n^{(M)}, \quad \text{as } n \to \infty; \tag{2.10}$$

$$V^{(M)} \xrightarrow{d} V, \quad \text{as } M \to \infty; \tag{2.11}$$

$$\forall\, \varepsilon > 0, \quad \lim_{M\to\infty} \limsup_{n\to\infty} P\left(|V_n^{(M)} - V_n| > \varepsilon\right) = 0. \tag{2.12}$$

Convergence (2.10) follows from Proposition 3 and Lemma 2. Convergence (2.11) is trivial because $\int_M^\infty \nu_\alpha(x,\infty]x^{-1}dx = \alpha^{-1}M^{-\alpha}$. Since $|V_n^{(M)} - V_n| = \int_M^\infty \hat{\nu}_n(x,\infty]x^{-1}dx$, (2.12) is equivalent to

$$\forall\, \varepsilon > 0, \quad \lim_{M\to\infty} \limsup_{n\to\infty} P\left(\int_M^\infty \hat{\nu}_n(x,\infty]x^{-1}dx > \varepsilon\right) = 0.$$

The steps of the verification of the above relation, up to (2.13), are pretty much the same as those developed on pp. 84-85 of [4]. We provide the details because we work with the measure $\nu_n^\dagger$ rather than with the measure ν_n, and the context for the remainder of the proof is helpful. Following (2.13), we use a different argument.

Fix $\varepsilon > 0$ and $\eta > 0$. Observe that

$$P\left(\int_M^\infty \hat{\nu}_n(x,\infty]x^{-1}dx > \varepsilon\right) \le Q_1(n) + Q_2(n),$$

where

$$Q_1(n) = P\left(\int_M^\infty \hat{\nu}_n(x,\infty]x^{-1}dx > \varepsilon,\ \left|\frac{\hat{Y}_{(k)}}{b(n/k)} - 1\right| < \eta\right),$$

$$Q_2(n) = P\left(\left|\frac{\hat{Y}_{(k)}}{b(n/k)} - 1\right| \ge \eta\right).$$

By Proposition 2, $\limsup_{n\to\infty} Q_2(n) = 0$, so we focus on $Q_1(n)$. We start with the bound

$$\begin{aligned} Q_1(n) &\le P\left(\int_M^\infty \hat{\nu}_n(x,\infty]x^{-1}dx > \varepsilon,\ \frac{\hat{Y}_{(k)}}{b(n/k)} > 1-\eta\right) \\ &= P\left(\int_M^\infty \frac{1}{k}\sum_{i=1}^n I_{\hat{Y}_i/\hat{Y}_{(k)}}(x,\infty]x^{-1}dx > \varepsilon,\ \frac{\hat{Y}_{(k)}}{b(n/k)} > 1-\eta\right). \end{aligned}$$

Conditions $\hat{Y}_i/\hat{Y}_{(k)} > x$ and $\hat{Y}_{(k)}/b(n/k) > 1-\eta$ imply $\hat{Y}_i/b(n/k) > x(1-\eta)$, so

$$\begin{aligned} Q_1(n) &\le P\left(\int_M^\infty \frac{1}{k}\sum_{i=1}^n I_{\hat{Y}_i/b(n/k)}(x(1-\eta),\infty]x^{-1}dx > \varepsilon\right) \\ &= P\left(\int_M^\infty \nu_n^\dagger(x(1-\eta),\infty]x^{-1}dx > \varepsilon\right) \\ &= P\left(\int_{M(1-\eta)}^\infty \nu_n^\dagger(x,\infty]x^{-1}dx > \varepsilon\right). \end{aligned}$$

Consequently, because the $\hat{Y}_i$, $1 \le i \le n$, have the same distribution,

$$Q_1(n) \le \frac{1}{\varepsilon} \int_{M(1-\eta)}^{\infty} E\left[\nu_n^{\dagger}(x,\infty]\right] x^{-1} dx = \frac{1}{\varepsilon} \int_{M(1-\eta)}^{\infty} \frac{n}{k} P\left(\frac{\hat{Y}}{b(n/k)} > x\right) x^{-1} dx.$$

It thus remains to show that

$$\lim_{z\to\infty} \limsup_{n\to\infty} \int_z^{\infty} \frac{n}{k} P\left(\hat{Y} > xb(n/k)\right) x^{-1} dx = 0. \tag{2.13}$$

We use the decomposition

$$\begin{aligned} P\left(\hat{Y} > xb(n/k)\right) &= P\left(\hat{Y} > xb(n/k),\ Y > \frac{1}{2}xb(n/k)\right) \\ &\quad + P\left(\hat{Y} > xb(n/k),\ Y \le \frac{1}{2}xb(n/k)\right) \\ &\le P\left(Y > \frac{1}{2}xb(n/k)\right) + P\left(|\hat{Y} - Y| > \frac{1}{2}xb(n/k)\right). \end{aligned}$$

By Lemma 1,

$$\lim_{z\to\infty} \limsup_{n\to\infty} \int_z^{\infty} \frac{n}{k} P\left(Y > \frac{1}{2}xb(n/k)\right) x^{-1} dx = 0.$$

If $\alpha > 4$, by (2.6) and (2.7)

$$\begin{aligned} \int_z^{\infty} \frac{n}{k} P\left(|\hat{Y} - Y| > \frac{1}{2}xb(n/k)\right) x^{-1} dx &\le \frac{n}{k} \int_z^{\infty} \frac{2}{xb(n/k)} E|\hat{Y} - Y| x^{-1} dx \\ &= O\left(\frac{n^{1/2}}{kb(n/k)}\right) \frac{1}{z}. \end{aligned}$$

By Assumption 3, for a slowly varying function L and $\gamma \in (\gamma(\alpha), 1)$,

$$\frac{n^{1/2}}{kb(n/k)} = \left\{\frac{n^{\gamma}}{k}\right\}^{1-\frac{1}{\alpha}} \left\{n^{\gamma(\alpha)-\gamma} L\left(\frac{n}{k}\right)\right\}^{1-\frac{1}{\alpha}} \to 0, \quad \text{as } n \to \infty.$$

If $\alpha \in (2,4)$, we use the bound ($r \in (0,1]$):

$$\int_z^\infty \frac{n}{k} P\left(|\hat{Y} - Y| > \frac{1}{2} x b(n/k)\right) x^{-1} dx \le \frac{n}{k} \int_z^\infty \left(\frac{2}{x b(n/k)}\right)^r E|\hat{Y} - Y|^r x^{-1} dx$$
$$= O\left(\frac{n}{k} \frac{E|\hat{Y} - Y|^r}{b^r(n/k)}\right) \frac{1}{z^r}.$$

The value of r will depend on α. Choosing it, and checking that it is available, requires some work. As in the proof of Proposition 1, $E|\hat{Y} - Y|^r = O\left(n^{-\kappa r/\beta}\right)$, provided (2.7) holds. Set

$$\gamma^* = \frac{1 - \frac{r}{\alpha} - \frac{\kappa r}{\beta}}{1 - \frac{r}{\alpha}}.$$

Then, for some slowly varying L,

$$\frac{n}{k} \frac{E|\hat{Y} - Y|^r}{b^r(n/k)} = O\left(\frac{n^{\gamma^*}}{k} L\left(\frac{n}{k}\right)\right)^{1 - \frac{r}{\alpha}}.$$

Clearly $\gamma^* < 1$. We must verify that there are β, κ and r, in permitted ranges, such that γ^* can be arbitrarily close to $\gamma(\alpha)$ given by (2.1). With α and r fixed, γ^* will approach its smallest possible value as κ/β approaches its largest possible value, i.e. $1 - 2/\alpha$. In this case, γ^* is greater than and approaches

$$\gamma_L(\alpha, r) := \frac{1 - \frac{r}{\alpha} - \left(1 - \frac{2}{\alpha}\right) r}{1 - \frac{r}{\alpha}} = \frac{\alpha - \alpha r + r}{\alpha - r}.$$

Condition (2.7) restricts the available values of r. A direct calculation shows that it is equivalent to

$$r < \frac{\beta\alpha}{\beta + \alpha}.$$

For a fixed α, the right–hand side is an increasing function of β and attains its upper limit if $\beta = \alpha/2$. This means that r must be less than, but can be arbitrarily close, to $\alpha/3$. Thus, γ^* can be arbitrarily close to

$$\gamma_L\left(\alpha, \frac{\alpha}{3}\right) = 2 - \frac{\alpha}{2}.$$

Combining it with Remark 2.4.1 concludes the proof.

Chapter 3

Extremal dependence measure for functional data

3.1 Introduction

We first concisely state main contributions of the paper with the caveat that detailed definitions and formulations will be provided in the following. Consider a sample of functions $X_i(t), t \in \mathcal{T}$, such that each of them has the same distribution as X. The Karhunen-Loéve expansion is $X(t) = \sum_{j=1}^{\infty} \xi_j v_j(t)$. The functions v_j are the functional principal components (FPCs) and the random variables ξ_j are their scores. We want to estimate extremal dependence of ξ_j and $\xi_{j'}$. We define a measure of such a dependence, which we denote by $D(\xi_j, \xi_{j'})$. We then define an estimator of $D(\xi_j, \xi_{j'})$ and formulate conditions under which it is consistent (Theorem 1) and asymptotically normal (Theorem 2). The main difficulty is that the population scores $\xi_{ij} = \langle X_i, v_j \rangle$ are not observable.

This paper thus makes a contribution at the nexus of functional data analysis (FDA) and extreme value theory (EVT). We assume that the reader is familiar with mathematical foundations of functional data analysis and central principles of extreme value theory. The FDA background given in Chapters 2 and 3 of [8] is sufficient. More detailed treatment is provided in [9]. Chapters 2 and 6 of [4] provide sufficient background in extreme value theory. Other references are cited when needed. We assume that all functions are elements of the space $L^2 = L^2(\mathcal{T})$, where the measure space $\mathcal{T}$ is such that $L^2(\mathcal{T})$, with the usual inner product, is a *separable* Hilbert space. This will be ensured if the measure on $\mathcal{T}$ is σ–finite and defined on a countably generated σ-algebra, see e.g. Proposition 3.4.5 in [27]. In particular, $\mathcal{T}$ can be taken to be a complete separable metric space (Polish space).

Suppose $X_1, \ldots, X_n$ are mean zero iid functions in L^2 with $\mathrm{E}\, \|X_i\|^2 < \infty$, and denote by X a generic random function with the same distribution as each X_i. A main dimension reduction tool of functional data analysis is to project the infinite dimensional functions X_i onto a finite dimensional subspace spanned by the FPCs. We now recall the required definitions. Consider the population covariance operator of X, defined by

$$C(x) := \mathrm{E}[\langle X, x\rangle\, X], \quad x \in L^2. \tag{3.1}$$

The eigenfunctions of C are the FPCs, denoted by $v_j, j \geq 1$, i.e., $C(v_j) = \lambda_j v_j$, where the λ_j are the eigenvalues of C. The FPCs lead to the commonly used Karhunen–Loéve expansion

$$X_i(t) = \sum_{j=1}^{\infty} \xi_{ij} v_j(t), \quad \xi_{ij} = \langle X_i, v_j\rangle\,, \quad \mathrm{E}\xi_{ij}^2 = \lambda_j. \tag{3.2}$$

The FPCs v_j and the eigenvalues λ_j are estimated by $\hat{v}_j$ and $\hat{\lambda}_j$, which are solutions to the equations

$$\widehat{C}(\hat{v}_j)(t) = \hat{\lambda}_j \hat{v}_j(t), \quad \text{for almost all } \; t \in \mathcal{T}, \tag{3.3}$$

where $\widehat{C}$ is the sample covariance operator defined by

$$\widehat{C}(x)(t) = \frac{1}{n} \sum_{i=1}^{n} \langle X_i, x\rangle\, X_i, \quad x \in L^2.$$

Each curve X_i can then be approximated by a linear combination of a finite set of the estimated FPCs $\hat{v}_j$, i.e., $X_i(t) \approx \sum_{j=1}^{p} \hat{\xi}_{ij} \hat{v}_j(t)$, where the $\hat{\xi}_{ij} = \langle X_i, \hat{v}_j\rangle$ are the sample scores. Each $\hat{\xi}_{ij}$ quantifies the contribution of the curve $\hat{v}_j$ to the shape of the curve X_i. Thus, the vector of the sample scores, $[\hat{\xi}_{i1}, \ldots, \hat{\xi}_{ip}]^\top$, encodes the shape of X_i to a good approximation. To illustrate, Fig. 3.1 displays the first three sample FPCs, $\hat{v}_1, \hat{v}_2, \hat{v}_3$, for intraday return curves $R_i, 1 \leq i \leq 1,378$, for Walmart stock from July 05, 2006 to Dec 30, 2011. These data are described in detail in Section II of the supplement. The curves R_i show how a return on an investment changes throughout a trading day as two examples are shown in Fig. 3.2. The curve $\hat{v}_1$ is a monotonic trend throughout the day. If the score corresponding to it is large, trading in this stock on a given day was dominated by a systematic increase (or decline if the score is negative) in the price of

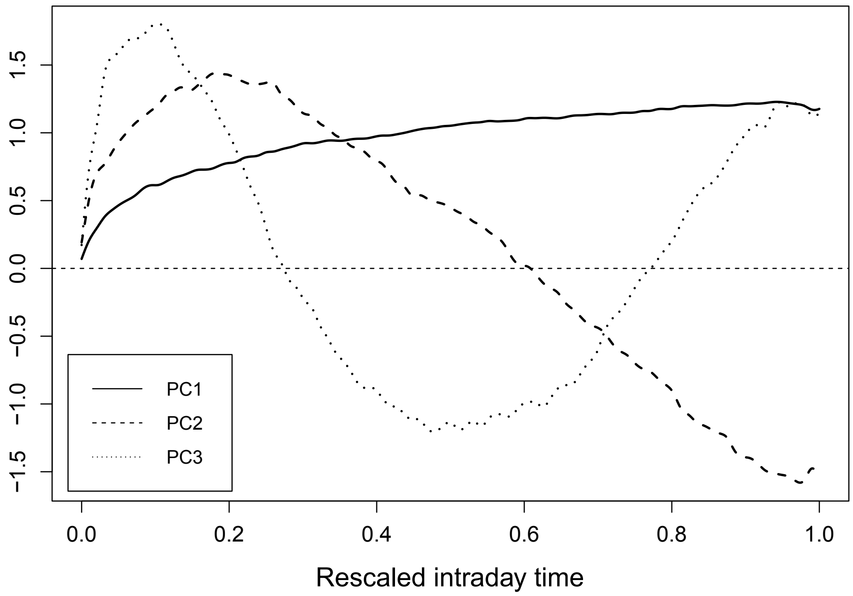

Figure 3.1: The first three sample FPCs of intraday returns on Walmart stock based on sample of 1,378 curves.

the stock. Notice the gradually decreasing slope of $\hat{v}_1$, which reflects the well-known fact that the most intense trading takes place after the opening of the trading floor. The second FPC, $\hat{v}_2$, has a large score, if there is a significant reversal in investor sentiment during a given trading day. These observations are illustrated in Fig. 3.2.

The main interest in this paper is the estimation of extremal dependence between the scores corresponding to different FPCs. Extremal dependence is a tendency of large values of one component to be coupled with large values of another component of a random vector. In the context of our Walmart stock example, extreme dependence between the first and second scores indicates that an extremely high monotonic trend and a pronounced reversion tend to occur simultaneously. Therefore, knowledge of the extreme value dependence of the scores may enhance the management of intraday risk.

We assess extremal dependence of the scores by means of the *extremal dependence measure* (EDM), which is constructed based on the theory of heavy–tailed regularly–varying random vectors. There has been considerable research on quantifying the tail dependence between extreme

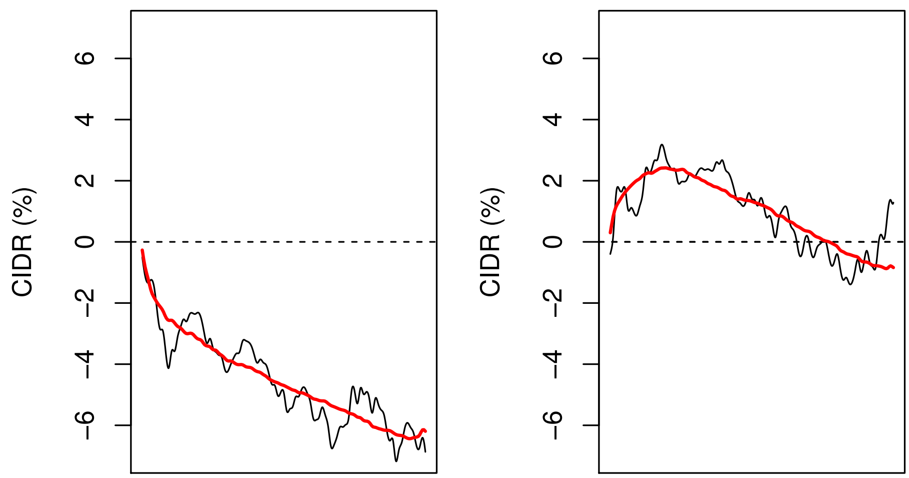

Figure 3.2: Walmart intraday cumulative return curves on two trading days and their approximations by $\sum_{i=1}^{3} \hat{\xi}_{ij}\hat{v}_j(t)$. In the left panel, $\hat{\xi}_1 = -4.7, \hat{\xi}_2 = 0.4, \hat{\xi}_3 = -0.1$, observed on October 7, 2008. In the right panel, $\hat{\xi}_1 = 0.8, \hat{\xi}_2 = 1.2, \hat{\xi}_3 = 0.1$, observed on November 18, 2008.

values in a heavy–tailed random vector. [28–30] defined the *coefficient of tail dependence*, which was later generalized to the *extremogram* by [31]. While these approaches are essentially based on the exponent measure of a random vector, the EDM is defined in terms of the spectral measure. The EDM was introduced by [32] and further investigated by [5]. Important related papers are [33] and [34].

In this paper, we quantify extremal dependence of scores using the EDM. To estimate the EDM of population scores, we consider an extension of the estimator proposed by [5]. It is important to emphasize that in our functional setting, the estimator can only be computed using the sample scores $\hat{\xi}_{ij} = \langle X_i, \hat{v}_j \rangle$, not the population scores $\xi_{ij} = \langle X_i, v_j \rangle$ because the ξ_{ij} are unobservable. Establishing large sample properties of any estimator based on sample scores requires taking the effect of the estimation of the scores into account. Since the $\hat{\xi}_{ij}$ depend on the whole sample $X_1, \ldots, X_n$, the vectors $[\hat{\xi}_{i1}, \ldots, \hat{\xi}_{ip}]^\top$ are no longer independent, even if $X_1, \ldots, X_n$ are i.i.d functions.

The remainder of the paper is organized as follows. In Section 3.2, we introduce preliminaries on multivariate regular variation and the EDM, and extend the concept of the EDM to multivariate data. Our main large sample results are presented in Section 3.3, which deals with the EDM for scores of functional observations. Section 3.4 presents a number of preliminary results. These results allow us to streamline the exposition of the proofs of the results of Section 3.3, which are presented in Section 3.5. Sections 3.6 and 3.7, present, respectively, an application to functional return data and a simulation study.

The paper is accompanied by supplementary material in Appendix A, which contains a couple of sections. Section A.1 explains how to normalize tail indexes of components of multivariate vectors. This is a well-researched topic in EVT, but may be less known in the FDA community, so a brief account needed to understand the application in Section 3.6 is provided. Section A.1 contains additional tables discussed in Section 3.7.

We hope that this work will be received with some interest by researchers working in two exciting and dynamic fields: functional data analysis and extreme value theory.

3.2 Multivariate regular variation and the EDM

We start by introducing multivariate regular variation for random vectors with positive components because the *extremal dependence measure* (EDM) was defined in such context. Following [4], we denote by $\mathbb{E}_d = [0, \infty]^d \setminus \{\mathbf{0}\}$ the nonnegative orthant compactified at infinity. We denote by $M_+(\mathbb{E}_d)$ the space of Radon measures on $\mathbb{E}_d$, and by $\xrightarrow{v}$ the vague convergence in $M_+(\mathbb{E}_d)$. An $\mathbb{E}_d$–valued random vector $\mathbf{Z} = [Z_1, \ldots, Z_d]^\top$ with distribution function F is regularly varying with index $-\alpha$, $\alpha > 0$, if there exists a sequence $b(n) \to \infty$ and a Radon measure ν on $\mathbb{E}_d$ such that

$$n \Pr\left(\frac{\mathbf{Z}}{b(n)} \in \cdot\right) \xrightarrow{v} \nu, \quad \text{in } M_+(\mathbb{E}_d). \tag{3.1}$$

Unless stated otherwise, all limits are taken as $n \to \infty$. The exponent measure ν has the property, $\nu(t\cdot) = t^{-\alpha}\nu(\cdot)$. We assume that one-dimensional marginal distributions of ν are nondegenerate. In (3.1), all components are normalized by the same sequence $\{b(n)\}$, which means that all marginal distributions are tail equivalent with the index $-\alpha$. A possible choice for $b(n)$ is the quantile function, defined by $\Pr(Z_1 > b(n)) = n^{-1}$. When $b(n) = n$, all marginal distributions are tail equivalent to the standard Pareto distribution with $\alpha = 1$, which is called the standard case.

There are various equivalent formulations of multivariate regular variation, see Theorem 6.1 of [4]. The formulation with a polar coordinate representation is commonly used due to its computational convenience and intuitive interpretation. Fix a norm $\|\cdot\|$ in $\mathbb{R}^d$, and set $\mathbb{S}_+^d = \{x \in \mathbb{R}^d : \|x\| = 1\} \cap \mathbb{E}_d$, the unit sphere in the nonnegative orthant. A d–dimensional random vector $\mathbf{Z} = [Z_1, \dots, Z_d]^\top$ is regularly varying if and only if there exists a sequence $b_R(n) \to \infty$ and an angular probability measure Γ on $\mathbb{S}_+^d$ such that for $(R, \boldsymbol{\Theta}) = (\|\mathbf{Z}\|, \mathbf{Z}/\|\mathbf{Z}\|)$,

$$n \Pr\left(\left(\frac{R}{b_R(n)}, \boldsymbol{\Theta}\right) \in \cdot\right) \xrightarrow{v} c\nu_\alpha \times \Gamma, \quad \text{in } M_+((0,\infty] \times \mathbb{S}_+^d), \tag{3.2}$$

where $\nu_\alpha(x,\infty] = x^{-\alpha}$ and $c = \nu\{\mathbf{x} : \|\mathbf{x}\| > 1\} > 0$. The sequence $\{b_R(n)\}$ in (3.2) is defined by $\Pr(R > b_R(n)) = n^{-1}$, so in this case $b_R(\cdot)$ depends on the choice of the norm $\|\cdot\|$. Definitions (3.1) and (3.2) can be extended directly to an $\mathbb{R}^d$-valued random vector with ν on $\mathbb{R}^d \setminus \{\mathbf{0}\}$ and Γ on $\mathbb{S}^d = \{x \in \mathbb{R}^d : \|x\| = 1\}$, see, e.g., Propositions 2.2.5 and 2.2.6 of [24]. In practice, the components of a random vector might not be tail equivalent. The case of different tail indexes of the coordinates, and transformations which make the coordinates tail equivalent are discussed in Section A.1.

We now turn to the EDM. Given a regularly varying nonnegative bivariate random vector $\mathbf{Z} = [Z_1, Z_2]^\top$, [5] define the EDM by

$$\mathrm{EDM}(Z_1, Z_2) = \int_{\mathbb{S}_+^2} a_1 a_2 \Gamma(d\mathbf{a}). \tag{3.3}$$

The EDM takes the minimal value of zero, $\mathrm{EDM}(Z_1, Z_2) = 0$, iff the coordinates of $\mathbf{Z}$ are asymptotically independent. This means that the angular measure Γ concentrates on $\{(1,0)/\|(1,0)\|, (0,1)/\|(0,1)\|\}$, or equivalently, the exponent measure ν concentrates on the axes. Also, if the norm is symmetric, $\mathrm{EDM}(Z_1, Z_2)$ achieves its maximal value iff the distribution of $\mathbf{Z}$ has asymptotic full dependence; i.e., Γ has mass on $\{(1,1)/\|(1,1)\|\}$, or equivalently, ν concentrates on the line $\{t(1,1), t > 0\}$.

[5] show that the EDM can be interpreted as the limit of cross moments between normalized Z_1 and Z_2 conditional on large values of $R = \|\mathbf{Z}\|$;

$$\mathrm{EDM}(Z_1, Z_2) = \lim_{r\to\infty} \mathrm{E}\left[\frac{Z_1}{R}\frac{Z_2}{R}\middle| R > r\right].$$

Based on this relation, they propose an estimator for $\mathrm{EDM}(Z_1, Z_2)$, defined by

$$\mathrm{D}_n(Z_1, Z_2) = \frac{1}{k}\sum_{i=1}^{n} \frac{Z_{i1}}{R_i}\frac{Z_{i2}}{R_i} I_{R_i \geq R_{(k)}}, \tag{3.4}$$

where $\mathbf{Z}_i = [Z_{i1}, Z_{i2}]^\top$, $1 \leq i \leq n$ are iid copies of $\mathbf{Z} = [Z_1, Z_2]^\top$, $R_i = \|\mathbf{Z}_i\|$, and $R_{(k)}$ is the kth largest order statistics with the convention $R_{(1)} = \max\{R_1, \dots, R_n\}$.

[5] consider non–negative bivariate vectors. To be able to work with the vectors of scores of functional data, we first have to extend their definitions to a setting of multivariate random vectors of an arbitrary dimension. Our first objective is to generalize (3.3) to a d–dimensional vector $\mathbf{Z} = [Z_1, \dots, Z_d]^\top$. We formulate the EDM between the components Z_1 and Z_2 for simplicity. We first assume that all components are positive. Given the angular measure Γ on $\mathbb{S}_+^d$ for $\mathbf{Z}$, we define the EDM for Z_1 and Z_2 as

$$\mathrm{D}(Z_1, Z_2) = \int_{\mathbb{S}_+^d} \frac{a_1 a_2}{\|(a_1, a_2, 0, \dots, 0)\|^2} \Gamma(d\mathbf{a}). \tag{3.5}$$

We set $a_1 a_2 / \|(a_1, a_2, 0, \dots, 0)\|^2 = 0$ when $a_1 = a_2 = 0$. Definition (3.5) is different from a simple extension of (3.3) given by

$$\mathrm{D}'(Z_1, Z_2) = \int_{\mathbb{S}_+^d} a_1 a_2 \Gamma(d\mathbf{a}). \tag{3.6}$$

We will now argue that for a d–dimensional vector $\mathbf{Z}$, with $d \geq 3$, D is a better measure for assessing extremal dependence between Z_1 and Z_2 than D'. Suppose that a random vector $\mathbf{Z} = [Z_1, Z_2, Z_3]^\top$ is regularly varying with an angular measure Γ on $\mathbb{S}_+^3$, and fix the Euclidean norm $\|\cdot\|$ in $\mathbb{R}_+^3$. Consider the following four cases.

1. The angular measure Γ_1 has unit mass on $(1,1,10)/\sqrt{102}$; the exponent measure ν_1 concentrates on $\{t(1,1,10), t > 0\}$.
2. The angular measure Γ_2 has unit mass on $(1,1,1)/\sqrt{3}$; the exponent measure ν_2 concentrates on $\{t(1,1,1), t > 0\}$.
3. The angular measure Γ_3 has unit mass on $(7,7,2)/\sqrt{102}$; the exponent measure ν_3 concentrates on $\{t(7,7,2), t > 0\}$.
4. The angular measure Γ_4 has mass $1/2$ on each $(1,1,10)/\sqrt{102}$ and $(7,7,2)/\sqrt{102}$; the exponent measure ν_4 concentrates on $\{t(1,1,10), t > 0\} \cup \{t(7,7,2), t > 0\}$.

Suppose Z has a Pareto distribution with index $\alpha > 0$. The following random vectors have extremal distribution corresponding to each of the above cases:

$$\mathbf{Z}^{(1)} = [Z,\ Z,\ 10Z], \quad \mathbf{Z}^{(2)} = [Z,\ Z,\ Z], \quad \mathbf{Z}^{(3)} = [7Z,\ 7Z,\ 2Z],$$

$$\mathbf{Z}^{(4)} = \xi[Z,\ Z,\ 10Z] + (1-\xi)[7Z,\ 7Z,\ 2Z],$$

where ξ is a Bernoulli random variable with probability of success $1/2$.

Set $\mathbb{P}_{12} = \{[t_1, t_2, 0], t_1, t_2 \in \mathbb{R}\}$. The projections of the random vectors $\mathbf{Z}^{(1)}$, $\mathbf{Z}^{(2)}$, $\mathbf{Z}^{(3)}$, and $\mathbf{Z}^{(4)}$ onto $\mathbb{P}_{12}$ are, respectively,

$$\widetilde{\mathbf{Z}}^{(1)} = [Z,\ Z],\ \ \widetilde{\mathbf{Z}}^{(2)} = [Z,\ Z],\ \ \widetilde{\mathbf{Z}}^{(3)} = [7Z,\ 7Z],\ \ \widetilde{\mathbf{Z}}^{(4)} = [\xi Z + 7(1-\xi)Z,\ \xi Z + 7(1-\xi)Z].$$

For all of the projected random vectors, the two components are equal, so a good measure of extremal dependence between them should attain its maximal value. Since we use the Euclidean norm and Γ is normalized to unity, the maximum value of both D and D' is $1/2$. Direct verification shows that we achieve the maximum value for all cases using the measure D. The measure D' however does not give the maximum value. For each case:

$$\mathrm{D}'(Z_1^{(1)}, Z_2^{(1)}) = \frac{1}{102}, \quad \mathrm{D}'(Z_1^{(2)}, Z_2^{(2)}) = \frac{34}{102},$$
$$\mathrm{D}'(Z_1^{(3)}, Z_2^{(3)}) = \frac{49}{102}, \quad \mathrm{D}'(Z_1^{(4)}, Z_2^{(4)}) = \frac{1}{102}\frac{1}{2} + \frac{49}{102}\frac{1}{2} = \frac{25}{102}.$$

It can be further shown that, for any norm $\|\cdot\|$ in $\mathbb{R}^d$, the measures D and D', defined for d–dimensional vector $\mathbf{Z}$ with $d \geq 3$, are not equivalent in the sense of Definition 1 on p.234 of [5], which we now recall. For a given $\mathbf{Z}$, let $\rho_i(\mathbf{Z}) = \int_{\mathbb{S}_+^d} k_i(\mathbf{a})\Gamma(d\mathbf{a})$ for some nonnegative map $k_i : \mathbb{S}_+^d \mapsto \mathbb{R}_+$. Then $\rho_1(\mathbf{Z})$ and $\rho_2(\mathbf{Z})$ are equivalent if and only if there are constants $0 < m \leq M < \infty$ such that

$$m\rho_1(\mathbf{Z}) \leq \rho_2(\mathbf{Z}) \leq M\rho_1(\mathbf{Z}).$$

It is obvious that the measures D and D' are equivalent for a bivariate vector $\mathbf{Z}$. We formalize the nonequivalence between the measures for a d–dimensional vector $\mathbf{Z}$ with $d \geq 3$ in the following proposition.

PROPOSITION 1. *Suppose that a $\mathbb{E}_d$–valued random vector $\mathbf{Z} = [Z_1, \ldots, Z_d]^\top$ is regularly varying with angular measure Γ on $\mathbb{S}_+^d$, with $d \geq 3$. Then $\mathrm{D}(Z_1, Z_2)$ and $\mathrm{D}'(Z_1, Z_2)$, defined in (3.5), (3.6), respectively, are not equivalent for any norm $\|\cdot\|$ in $\mathbb{R}^d$.*

Proof. Proposition 1 of [5] shows that $\rho_1(\mathbf{Z})$ and $\rho_2(\mathbf{Z})$ are equivalent if and only if there are constants $0 < m \leq M < \infty$ such that

$$mk_1(\mathbf{a}) \leq k_2(\mathbf{a}) \leq Mk_1(\mathbf{a}), \quad \forall\, \mathbf{a} \in \mathbb{S}_+^d. \tag{3.7}$$

Observe that the ratio of the integrand in $\mathrm{D}'(Z_1, Z_2)$ to the integrand in $\mathrm{D}(Z_1, Z_2)$ is $\|(a_1, a_2, 0, \ldots, 0)\|^2$. This ratio is clearly zero at $\mathbf{a} = \mathbf{0}$, violating (3.7), but $\mathbf{0} \notin \mathbb{S}_+^d$. We therefore consider a path in $\mathbb{S}_+^d$ defined by

$$\mathbf{a}(x) = (x, x, 1, 0, \ldots, 0)/\|(x, x, 1, 0, \ldots, 0)\|, \quad x \searrow 0.$$

Then,

$$\|(a_1(x), a_2(x), 0, 0, \dots, 0)\|^2 = \frac{\|(x, x, 0, 0, \dots, 0)\|^2}{\|(x, x, 1, 0, \dots, 0)\|^2} \to 0,$$

as $x \searrow 0$ because every norm in $\mathbb{R}^d$ is equivalent to the Euclidean norm.

□

Another question of interest is the relationship between $\mathrm{D}(Z_1, Z_2)$ in (3.5) and $\mathrm{EDM}(Z_1, Z_2)$ in (3.3). We clarify it in the following proposition.

PROPOSITION 2. *Suppose that the exponent measure and angular measure of a d–dimensional regularly–varying random vector $\mathbf{Z} = [Z_1, \dots, Z_d]^\top$ are, respectively, ν on $\mathbb{E}_d$ and Γ on $\mathbb{S}_+^d$. Denote the exponent measure and angular measure of the bivariate vector $[Z_1, Z_2]^\top$, respectively, by ν_2 on $\mathbb{E}_2$ and Γ_2 on $\mathbb{S}_+^2$. Then,*

$$\mathrm{D}(Z_1, Z_2) = \int_{\mathbb{S}_+^d} \frac{a_1 a_2}{\|(a_1, a_2, 0, \dots, 0)\|^2} \Gamma(d\mathbf{a}) = \int_{\mathbb{S}_+^2} b_1 b_2 \Gamma_2(d\mathbf{b}) = \mathrm{EDM}(Z_1, Z_2)$$

and, for any Borel set $G \subset \mathbb{E}_2$,

$$\nu_2(G) = \nu(G \times [0, \infty]^{d-2}).$$

Proof. We first clarify the connection between the measure ν on $\mathbb{E}_d$ and the measure ν_2 on $\mathbb{E}_2$. By (3.1), for any measurable rectangle $A \times B \subset \mathbb{E}_2$,

$$\frac{\nu(A \times B \times [0, \infty]^{d-2})}{\nu_2(A \times B)} = \lim_{n \to \infty} \frac{n \Pr\left(\mathbf{Z}/b(n) \in A \times B \times [0, \infty]^{d-2}\right)}{n \Pr\left(Z_1/b(n) \in A, Z_2/b(n) \in B\right)} = 1.$$

We conclude that the measure ν_2 is obtained by integrating the entire measure ν over all coordinates except for the first two.

According to formulas on page 239 of [5], $\mathrm{EDM}(Z_1, Z_2)$ can be expressed as

$$\int_{\mathbb{S}_+^2} b_1 b_2 \Gamma_2(d\mathbf{b}) = \frac{1}{\nu_2\left(\|(y_1, y_2)\| > 1\right)} \int_{\|(y_1, y_2)\| > 1} \frac{y_1 y_2}{\|(y_1, y_2)\|^2} \, \nu_2(dy_1 dy_2).$$

Therefore, using the relationship between ν_2 and ν,

$$\begin{aligned}
&\int_{\mathbb{S}_+^2} b_1 b_2 \Gamma_2(d\mathbf{b}) \\
&= \frac{1}{\nu\left(\{\mathbf{y} : \|(y_1, y_2, 0, \ldots, 0)\| > 1\}\right)} \int_{\{\mathbf{y}:\|(y_1,y_2,0,\ldots,0)\|>1\}} \frac{y_1 y_2}{\|(y_1, y_2, 0, \ldots, 0)\|^2}\, \nu(d\mathbf{y}).
\end{aligned}$$

Applying the polar transformation T defined by $T(\mathbf{y}) = (\|\mathbf{y}\|, \mathbf{y}/\|\mathbf{y}\|)$ for $\mathbf{y} \in \mathbb{E}_d$, we obtain

$$\begin{aligned}
&\int_{\mathbb{S}_+^2} b_1 b_2 \Gamma_2(d\mathbf{b}) \\
&= \frac{1}{\nu\left(\{\mathbf{y} : \|(y_1, y_2, 0, \ldots, 0)\| > 1\}\right)} \int_{T(\{\mathbf{y}:\|(y_1,y_2,0,\ldots,0)\|>1\})} f \circ T^{-1}(r, \mathbf{a})\, \nu \circ T^{-1}(dr \times d\mathbf{a}),
\end{aligned}$$

where $f(\mathbf{y}) = y_1 y_2 / \|(y_1, y_2, 0, \ldots, 0)\|^2$. First observe that

$$\begin{aligned}
T\left(\{\mathbf{y} : \|(y_1, y_2, 0, \ldots, 0)\| > 1\}\right) &= \{(r, (a_1, a_2, \ldots, a_d)) : \|(ra_1, ra_2, 0, \ldots, 0)\| > 1\} \\
&= \{(r, (a_1, a_2, \ldots, a_d)) : r > \|(a_1, a_2, 0, \ldots, 0)\|^{-1}\}.
\end{aligned}$$

Using the fact that $\nu \circ T^{-1} = c\nu_\alpha \times \Gamma$, where $c = \nu\left(\|\mathbf{y}\| > 1\right)$, we obtain

$$\begin{aligned}
&\nu\left(\{\mathbf{y} : \|(y_1, y_2, 0, \ldots, 0)\| > 1\}\right) \\
&\quad = \nu \circ T^{-1}\left(T\left(\{\mathbf{y} : \|(y_1, y_2, 0, \ldots, 0)\| > 1\}\right)\right) \\
&\quad = c\nu_\alpha \times \Gamma\left(\{(r, (a_1, a_2, \ldots, a_d)) : r > \|(a_1, a_2, 0, \ldots, 0)\|^{-1}\}\right) \\
&\quad = c\,\|(a_1, a_2, 0, \ldots, 0)\|^\alpha.
\end{aligned}$$

Therefore,

$$
\begin{aligned}
&\int_{\mathbb{S}_+^2} b_1 b_2 \Gamma_2(d\mathbf{b}) \\
&= \frac{1}{c\,\|(a_1,a_2,0,\ldots,0)\|^\alpha} \int_{\mathbb{S}_+^d} \int_{r>\|(a_1,a_2,0,\ldots,0)\|^{-1}} \frac{a_1 a_2}{\|(a_1,a_2,0,\ldots,0)\|^2}\, c\nu_\alpha(dr)\Gamma(d\mathbf{a}) \\
&= \int_{\mathbb{S}_+^d} \frac{a_1 a_2}{\|(a_1,a_2,0,\ldots,0)\|^2}\, \Gamma(d\mathbf{a}).
\end{aligned}
$$

$\square$

By Proposition 2 we can use the estimator (3.4), originally introduced for $\mathrm{EDM}(Z_1, Z_2)$, to estimate $\mathrm{D}(Z_1, Z_2)$ as well.

A further extension of the EDM (3.3) is that from the nonnegative quadrant to the four quadrants, as a vector of the scores takes on values in $\mathbb{R}^d$. [5] define the EDM for a nonnegative random vector, but (3.3) can be readily generalized to a random vector $\mathbf{Z} = [Z_1, Z_2]^\top$ with real components. Suppose that $\mathbf{Z} = [Z_1, Z_2]^\top$ in $\mathbb{R}^2$ is regularly varying with an angular measure Γ_2 on $\mathbb{S}^2$. Then, we define the EDM for $\mathbf{Z} = [Z_1, Z_2]^\top$ by

$$
\mathrm{EDM}(Z_1, Z_2) = \int_{\mathbb{S}^2} a_1 a_2 \Gamma_2(d\mathbf{a}). \tag{3.8}
$$

The above definition allows us to quantify the strength of the extremal dependence between Z_1 and Z_2 in $\mathbb{R}^2$. Unlike (3.3), (3.8) can take a negative value depending on which quadrants Γ_2 has its mass on, so careful interpretation is needed. To explore the dependence spectrum that (3.8) can measure, we fix the Euclidean norm $\|\cdot\|$ in $\mathbb{R}^2$. Then, (3.8) has a range from -1/2 to 1/2. The maximal value, $1/2$, indicates a perfect positive extremal dependence; here, "positive" means that Z_1 and Z_2 have the same signs, and "perfect" means that the magnitudes of Z_1 and Z_2 show asymptotic full dependence, i.e., Γ_2 concentrates on $\{(1,1)/\sqrt{2},\ (-1,-1)/\sqrt{2}\}$. Similarly, the minimum value, $-1/2$, indicates a perfect negative extremal dependence; "negative" means that Z_1 and Z_2 have the opposite signs, and in this case Γ has mass on $\{(-1,1)/\sqrt{2},\ (1,-1)/\sqrt{2}\}$.

Note that if $\mathbf{Z}$ exhibits asymptotic independence, i.e., its exponent measure concentrates on the standard axes, then (3.8) is 0, but the reverse does not necessarily hold true. For example, if Γ_2 concentrates equally on each element of

$$\{(1,1)/\sqrt{2},\ (-1,1)/\sqrt{2},\ (-1,-1)/\sqrt{2},\ (1,-1)/\sqrt{2}\},$$

then (3.8) is 0, but each quadrant shows the perfect dependence. To avoid this issue and take into account the extremal dependence in each quadrant, we suggest to complement (3.8) on the unit sphere $\mathbb{S}^2$ with its decomposition into the four quadrants. Let $\mathbb{S}^2_{(+,+)} = \mathbb{S}^2 \cap \{(x_1, x_2) \in \mathbb{R}^2 : x_1 \geq 0, x_2 \geq 0\}$. Similarly, let $\mathbb{S}^2_{(-,+)} = \mathbb{S}^2 \cap \{x_1 \leq 0, x_2 \geq 0\}$, $\mathbb{S}^2_{(-,-)} = \mathbb{S}^2 \cap \{x_1 \leq 0, x_2 \leq 0\}$, and $\mathbb{S}^2_{(+,-)} = \mathbb{S}^2 \cap \{x_1 \geq 0, x_2 \leq 0\}$. We define the supplementary measure for (3.8) by splitting the EDM into the four quadrant spheres,

$$\left[\int_{\mathbb{S}^2_{(+,+)}} a_1 a_2 \Gamma_2(d\mathbf{a}),\ \int_{\mathbb{S}^2_{(-,+)}} a_1 a_2 \Gamma_2(d\mathbf{a}),\ \int_{\mathbb{S}^2_{(-,-)}} a_1 a_2 \Gamma_2(d\mathbf{a}),\ \int_{\mathbb{S}^2_{(+,-)}} a_1 a_2 \Gamma_2(d\mathbf{a})\right]. \tag{3.9}$$

To estimate each of the components in (3.9), we slightly modify (3.4); for example, an estimator for $\int_{\mathbb{S}^2_{(+,+)}} a_1 a_2 \Gamma(d\mathbf{a})$ is

$$\mathrm{D}_n^{(+,+)}(Z_1, Z_2) = \frac{1}{k} \sum_{i=1}^{n} \frac{Z_{i1}}{R_i} \frac{Z_{i2}}{R_i} I_{R_i \geq R_{(k)}} I_{Z_{i1} \geq 0, Z_{i2} \geq 0}.$$

To elaborate, we first order the n bivariate vectors by norm and consider the top k vectors with large norm. We then use only those for which $Z_{i1} \geq 0$ and $Z_{i2} \geq 0$ from the k vectors. Estimators for the other components in (3.9) can be obtained in the same manner reflecting the different quadrants.

We conclude this section with an analog of Proposition 2. Given an $\mathbb{R}^d$–valued random vector $[Z_1, \ldots, Z_d]^\top$, we can measure extremal dependence between Z_1 and Z_2 using (3.5), but integrated over the whole sphere $\mathbb{S}^d$. Following the steps in the proof of Proposition 2, it is readily shown that $\mathrm{D}(Z_1, Z_2)$ for two components of an $\mathbb{R}^d$–valued vector is in fact the same as (3.8).

COROLLARY 1. *Suppose the angular measure of a $\mathbb{R}^d$–valued random vector $[Z_1, \ldots, Z_d]^\top$ is Γ on $\mathbb{S}^d$ and the angular measure of $[Z_1, Z_2]^\top$ is Γ_2 on $\mathbb{S}^2$. Then,*

$$\mathrm{D}(Z_1, Z_2) = \int_{\mathbb{S}^d} \frac{a_1 a_2}{\|(a_1, a_2, 0, \ldots, 0)\|^2} \Gamma(d\mathbf{a}) = \int_{\mathbb{S}^2} b_1 b_2 \Gamma_2(d\mathbf{b}) = \mathrm{EDM}(Z_1, Z_2).$$

3.3 The EDM for scores of functional data

In this section, we consider the estimation of the EDM of scores of functional data. Following the framework introduced in Section 3.1, recall that $X_1, \ldots, X_n$ are mean zero iid functions in L^2 with $\mathrm{E}\|X_i\|^2 < \infty$, and that each X_i admits the Karhunen–Loéve expansion (3.2). The unknown population scores $\xi_{ij} = \langle X_i, v_j \rangle$ in (3.2) are estimated by the sample scores $\hat{\xi}_{ij} = \langle X_i, \hat{v}_j \rangle$, where the $\hat{v}_j$ are estimators of the FPCs v_j. We introduce the following random variables:

$$\mathbf{Y}^d = [\xi_1, \ldots, \xi_d]^\top, \;\; \xi_j = \langle X, v_j \rangle \,, \quad \mathbf{Y}_i^d = [\xi_{i1}, \ldots, \xi_{id}]^\top, \;\; \xi_{ij} = \langle X_i, v_j \rangle \,,$$

$$\widehat{\mathbf{Y}}^d = [\hat{\xi}_1, \ldots, \hat{\xi}_d]^\top, \;\; \hat{\xi}_j = \langle X, \hat{v}_j \rangle \,, \quad \widehat{\mathbf{Y}}_i^d = [\hat{\xi}_{i1}, \ldots, \hat{\xi}_{id}]^\top, \;\; \hat{\xi}_{ij} = \langle X_i, \hat{v}_j \rangle \,.$$

To quantify the extremal dependence between components ξ_j and $\xi_{j'}$ in $\mathbf{Y}^d$, we consider the EDM, $\mathrm{D}(\xi_j, \xi_{j'})$, defined in (3.5). Then, by Corollary 1,

$$\mathrm{D}(\xi_j, \xi_{j'}) = \int_{\mathbb{S}^2} a_1 a_2 \Gamma_{jj'}(d\mathbf{a}), \tag{3.1}$$

where $\Gamma_{jj'}$ on $\mathbb{S}^2$ is the angular measure of the bivariate random vector $[\xi_j, \xi_{j'}]^\top$.

Set $\mathbf{Y}_i = [\xi_{ij}, \xi_{ij'}]^\top$, $\widehat{\mathbf{Y}}_i = [\hat{\xi}_{ij}, \hat{\xi}_{ij'}]^\top$, $1 \leq i \leq n$, where we suppress the dependence of the bivariate vectors on j and j'. In light of (3.4), we consider two random variables that approximate $\mathrm{D}(\xi_j, \xi_{j'})$:

$$\mathrm{D}_n(\xi_j, \xi_{j'}) := \frac{1}{k} \sum_{i=1}^{n} \frac{\xi_{ij}}{R_i} \frac{\xi_{ij'}}{R_i} I_{R_i \geq R_{(k)}},$$

$$\widehat{\mathrm{D}}_n(\xi_j, \xi_{j'}) := \frac{1}{\hat{k}} \sum_{i=1}^{n} \frac{\hat{\xi}_{ij}}{\widehat{R}_i} \frac{\hat{\xi}_{ij'}}{\widehat{R}_i} I_{\widehat{R}_i \geq \widehat{R}_{(\hat{k})}}, \tag{3.2}$$

where $R_i = \|\mathbf{Y}_i\|$, $\widehat{R}_i = ||\widehat{\mathbf{Y}}_i||$, and $R_{(k)}$ and $\widehat{R}_{(\hat{k})}$ are the respective largest order statistics. There is a fundamental difference between $\mathrm{D}_n(\xi_j, \xi_{j'})$ and $\widehat{\mathrm{D}}_n(\xi_j, \xi_{j'})$; $\mathrm{D}_n(\xi_j, \xi_{j'})$ is an infeasible estimator because the FPCs v_j are not observable, so the ξ_{ij} cannot be computed from the data. The estimator based on the sample scores, $\widehat{\mathrm{D}}_n(\xi_j, \xi_{j'})$, is what we can actually compute. Therefore, the consistency of $\widehat{\mathrm{D}}_n(\xi_j, \xi_{j'})$ for $\mathrm{D}(\xi_j, \xi_{j'})$ must be established. As noted in the Introduction, the sample scores $\hat{\xi}_{ij}$ are no longer independent in i (nor in j); they form a triangular array of dependent identically distributed vectors of dimension d. This new aspect of EDM estimation is specific to functional data. To handle it rigorously, we must introduce a suitable framework for regular variation of functional data. We follow [23] and [24].

[23] introduced a framework based on M_0 convergence, where M_0 is the space of measures on a complete separable metric space. [24] further investigated regular variation in Banach spaces using the notion of M_0 convergence. We define a regularly varying function in a separable Banach space $\mathbb{B}$ as follows.

DEFINITION 1. Denote the norm in $\mathbb{B}$ by $\|\cdot\|_{\mathbb{B}}$ and the unit sphere in $\mathbb{B}$ by $\mathbb{S} := \{x \in \mathbb{B} : \|x\|_{\mathbb{B}} = 1\}$. A random element X in $\mathbb{B}$ is regularly varying with index $-\alpha, \alpha > 0$ if any of the following conditions hold:

(i) There exists a measure ν and a regularly varying sequence $b(n) \to \infty$ with index $1/\alpha$ such that

$$n \Pr\left(\frac{X}{b(n)} \in \cdot\right) \xrightarrow{M_0} \nu(\cdot), \quad n \to \infty, \tag{3.3}$$

where ν is a non-null measure (exponent measure) on the Borel σ–field $\mathcal{B}(\mathbb{B}_0)$ of $\mathbb{B}_0 = \mathbb{B} \setminus \{\mathbf{0}\}$.

(ii) There exists a probability measure Γ on $\mathbb{S}$ and a regularly varying sequence $b_R(n) \to \infty$ such that, for any $y > 0$,

$$n \Pr\left(\|X\|_{\mathbb{B}} > y b_R(n), X/\|X\|_{\mathbb{B}} \in \cdot\right) \xrightarrow{w} c y^{-\alpha} \Gamma(\cdot), \quad n \to \infty, \tag{3.4}$$

for some $c > 0$.

There are several equivalent definitions, see Section 2.2 of [24], which also contains all details. The quantile function $b(t)$ in (3.3) admits the representation

$$b(t) = t^{1/\alpha} L(t), \quad t > 0, \tag{3.5}$$

where L is slowly varying as $t \to \infty$. An analogous representation holds for the function b_R. With the choice of $b_R(n)$, defined by $\Pr(\|X\|_{\mathbb{B}} > b_R(n)) = n^{-1}$, we get $c = 1$ in (3.4) since $\Gamma(\mathbb{S}) = 1$ for any $y > 0$.

We briefly review the theory of M_0 convergence. Let $B_\varepsilon := \{z \in \mathbb{B} : \|z\|_{\mathbb{B}} < \varepsilon\}$ be the open ball of radius $\varepsilon > 0$ centered at the origin. A Borel measure ν defined on $\mathbb{B}_0$ is said to be *boundedly finite* if $\nu(A) < \infty$, for all Borel sets that are bounded away from $\mathbf{0}$, i.e., $A \cap B_\varepsilon = \emptyset$, for some $\varepsilon > 0$. Let $\mathbb{M}_0$ be the collection of all such measures. For $\nu_n, \nu \in \mathbb{M}_0$, the ν_n converge to ν in the M_0 topology, if $\nu_n(A) \to \nu(A)$, for all bounded away from $\mathbf{0}$, ν–continuity Borel sets A, i.e., $\nu(\partial A) = 0$, where ∂A is the boundary of A. If $\mathbb{B}$ is an Euclidean space, Definition 1 is equivalent to regular variation as defined in Section 3.2.

We work in the Hilbert space L^2, so in the following we replace the general Banach space $\mathbb{B}$ with a separable Hilbert space $\mathbb{H}$. We define the finite–dimensional projection of $z \in \mathbb{H}$ on the subspace spanned by $f_1, \ldots, f_d \in \mathbb{H}$ by

$$\pi_{f_1,\ldots,f_d}(z) := [\langle z, f_1 \rangle, \ldots, \langle z, f_d \rangle]^\top.$$

We claim in the following proposition that regular variation in $\mathbb{H}$ implies regular variation of the finite–dimensional projections in $\mathbb{R}^d$. To lighten the notation, we suppress the subscript $f_1, \ldots, f_d$ so that $\pi(z) = \pi_{f_1,\ldots,f_d}(z)$. Let $\mathcal{B}(\mathbb{S}^d)$ be the Borel σ–field on $\mathbb{S}^d$. For any set S in $\mathcal{B}(\mathbb{S}^d)$, define a set of elements in $\mathbb{H}$ by

$$\mathcal{A}_\pi(S) := \{z \in \mathbb{H} : \|\pi(z)\| > 1, \ \pi(z)/\|\pi(z)\| \in S\}. \tag{3.6}$$

PROPOSITION 1. *If a random element X in $\mathbb{H}$ is regularly varying with index $-\alpha, \alpha > 0$, and $\nu(\mathcal{A}_\pi(\mathbb{S}^d)) > 0$, then $\pi(X)$ is regularly varying in $\mathbb{R}^d$ with index $-\alpha$.*

In our FDA context, the functions $f_1, \ldots, f_d$ of interest are the FPCs $v_1, \ldots, v_d$. We work under the following assumption.

ASSUMPTION 1. The functions $X_1, \ldots X_n$ are i.i.d copies of X, which is regularly varying in L^2 according to Definition 1 with $\alpha > 2, \alpha \neq 4$. The FPCs $v_1, \ldots, v_d$ satisfy $\nu(\mathcal{A}_{\pi_{v_1,\ldots,v_d}}(\mathbb{S}^d)) > 0$. (The set $\mathcal{A}_{\pi_{v_1,\ldots,v_d}}$ is defined according to (3.6).)

By Proposition 1, under Assumption 1, the projection $\mathbf{Y}^d = \pi_{v_1,\ldots,v_d}(X)$ is regularly varying in $\mathbb{R}^d$ with the same index as X. The assumption $\alpha > 2$ ensures that $\mathrm{E}\|X\|^2 < \infty$, so that the FPCs can be defined. If $\alpha = 2$, then either $\mathrm{E}\|X\|^2 = \infty$ or $\mathrm{E}\|X\|^2 < \infty$ are possible, and complex assumptions on the slowly varying function L would be needed to ensure that $\mathrm{E}\|X\|^2 < \infty$. Similarly, if $\alpha = 4$, then either $\mathrm{E}\|X\|^4 = \infty$ or $\mathrm{E}\|X\|^4 < \infty$ are possible. There is a phase transition at $\alpha = 4$ found in the functional context by [25]. The phase transitions at $\alpha = 2$ and $\alpha = 4$ in various context related to regular variation have been well-known since the 1980s, see, e.g., Theorem 3.5 in [35], earlier papers of [36–38], and [16] for a broad picture. We therefore exclude $\alpha = 2$ and $\alpha = 4$ from our analysis. In the context of research on regularly varying and heavy-tailed random elements, the chief restriction is $\alpha > 2$, needed to ensure that the FPC are readily defined. It is conceivable that in the context of functions whose projections are heavy-tailed, data-driven bases different from the FPC might be appropriate, but such bases have not been devised yet.

As noted earlier, the sample scores $\hat{\xi}_{ij} = \langle X_i, \hat{v}_j \rangle$ form a triangular array whose elements are dependent across i and j. We now review bounds on the distance $\hat{v}_j - v_j$. As noted in the Introduction, these bounds apply to $\mathrm{sign}\left(\langle \hat{v}_j, v_j \rangle\right) \hat{v}_j - v_j$, but the sign always cancels in final formulas, so we assume that $\mathrm{sign}\left(\langle \hat{v}_j, v_j \rangle\right) = 1$. Recall that v_j is the jth eigenfunction of the covariance operator C in (3.1) corresponding to the eigenvalue λ_j, and $\hat{v}_j$ is the jth eigenfunction of its estimator $\widehat{C}$ in (3.3). By Lemma 2.3 in [8],

$$\|\hat{v}_j - v_j\| \leq A_j \|\widehat{C} - C\|_{\mathcal{L}}, \tag{3.7}$$

provided $d_j > 0$, where $A_j = 2\sqrt{2}/d_j$, and

$$d_1 = \lambda_1 - \lambda_2, \quad d_j = \min\left\{\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1}\right\}, \ j \geq 2. \tag{3.8}$$

The asymptotic properties of the distance between $\widehat{C}$ and C are separated into two cases depending on the range of α. If $\alpha > 4$, then $\mathrm{E}\|X\|^4 < \infty$, so, by Theorem 2.5 in [8],

$$\mathrm{E}\|\widehat{C} - C\|^2 = O(n^{-1}). \tag{3.9}$$

Using (3.7), we have

$$\mathrm{E}\|\hat{v}_j - v_j\|^2 = O(n^{-1}). \tag{3.10}$$

The case of regularly varying X with tail index $\alpha \in (2, 4)$, which implies $\mathrm{E}\|X\|^2 < \infty$ and $\mathrm{E}\|X\|^4 = \infty$, is studied in [25]. Under week conditions, relation (3.9) must be replaced by

$$\mathrm{E}\|\widehat{C} - C\|_{\mathcal{L}}^{\beta} \leq L_\beta(n) n^{-\beta(1-2/\alpha)}, \quad \forall\, \beta \in (0, \alpha/2), \tag{3.11}$$

where L_β is a slowly varying function. For a fixed α, the strongest bound is obtained as $\beta \nearrow \alpha/2$, in which case $\beta(1 - 2/\alpha) \nearrow \alpha/2 - 1$. As $\alpha \nearrow 4$ and $\beta \nearrow \alpha/2$, relation (3.11) thus approaches, in a heuristic sense, relation (3.9). From (3.7) and (3.11), we get the condition

$$\mathrm{E}\|\hat{v}_j - v_j\|^{\beta} = o\left(n^{-\kappa}\right), \quad \forall\, \beta \in \left(1,\ \frac{\alpha}{2}\right), \forall\, \kappa \in \left(0,\ \beta\left(1 - \frac{2}{\alpha}\right)\right). \tag{3.12}$$

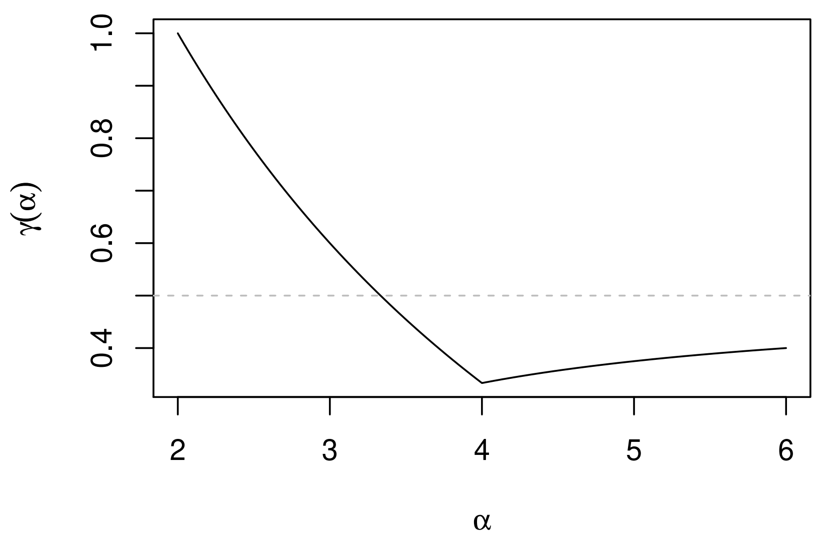

Figure 3.3: The graph of the function $\gamma(\alpha)$ for $\alpha \in (2, 6)$.

To see this, observe that

$$n^\kappa \mathrm{E}\|\hat{v}_j - v_j\|^\beta \leq A_j n^\kappa \mathrm{E}\|\widehat{C} - C\|_{\mathcal{L}}^\beta \leq A_j L_\beta(n) n^{-\beta(1-2/\alpha)+\kappa}.$$

Since $-\beta(1 - 2/\alpha) + \kappa < 0$, by Proposition 2.6 (i) of [4], we obtain (3.12).

The following Assumption 2 thus always holds as long as the eigenvalue separations d_j defined by (3.8) are positive, but this is a sufficient condition, so we state what is needed for our results to hold.

ASSUMPTION 2. The estimators $\hat{v}_j$ satisfy (3.10) if $\alpha > 4$ and (3.12) if $\alpha \in (2, 4)$.

Asymptotic properties in extreme value theory are typically derived as the number of upper order statistics, k, tends to infinity with the sample size n, in such a way that $k/n \to 0$. This condition remains to be sufficient for $\mathrm{D}_n(\xi_j, \xi_{j'}) \xrightarrow{P} \mathrm{D}(\xi_j, \xi_{j'})$, since the population scores $\mathbf{Y}_i$ are i.i.d. and regularly varying under Assumption 1. In our setting, however, we estimate the EDM based on $\widehat{\mathrm{D}}_n(\xi_j, \xi_{j'})$ calculated from the observed approximations $\widehat{\mathbf{Y}}_i$. It can be therefore expected that this additional approximation will, to some extent, restrict the rate at which k tends to infinity with n. We formulate a sufficient condition on the order of k in Assumption 3 below. We first define the function

$$\gamma(\alpha) = \begin{cases} \dfrac{6-\alpha}{\alpha+2}, & \alpha \in (2,4], \\ \\ \dfrac{\alpha-2}{2\alpha-2}, & \alpha \in (4,\infty). \end{cases} \tag{3.13}$$

Fig. 3.3 shows that $\gamma(\cdot)$ is continuous at the phase transition point $\alpha = 4$ with $\gamma(4) = 1/3$. It increases on $(4,\infty)$ with $\lim_{\alpha\nearrow\infty}\gamma(\alpha) = \frac{1}{2}$. For $\alpha \in (2,4)$, $\gamma(\alpha)$ decreases with $\lim_{\alpha\searrow 2}\gamma(\alpha) = 1$. For each value of $\alpha > 2$, the interval $(\gamma(\alpha), 1)$ is not empty. We write

$$k >> n^{\gamma}, \quad \text{for some } \gamma \in (0,1), \text{ if } k/n^{\gamma} \to \infty.$$

ASSUMPTION 3. We assume that $k >> n^{\gamma}$ for some $\gamma \in (\gamma(\alpha), 1)$, with $\gamma(\alpha)$ defined in (3.13).

Assumption 3 implies that $k > \sqrt{n}$ always works if $\alpha > 4$, but as $\alpha \searrow 2$, almost all observations must be used to ensure the consistency of the estimator.

With all assumptions formulated and explained, we are ready to state the first main result of this section.

THEOREM 1. *Recall the definitions of the EDM* $\mathrm{D}(\xi_j, \xi_{j'})$ *and its estimator* $\widehat{\mathrm{D}}_n(\xi_j, \xi_{j'})$ *given, respectively, in* (3.1) *and* (3.2)*. Under Assumptions 1, 2, and 3,*

$$\widehat{\mathrm{D}}_n(\xi_j, \xi_{j'}) \xrightarrow{P} \mathrm{D}(\xi_j, \xi_{j'}).$$

Recall that $\mathrm{D}(\xi_j, \xi_{j'})$ integrates extremal dependence over the whole sphere $\mathbb{S}^2$. As noted in Section 3.2, this might distort the true dependence, so we decompose $\mathrm{D}(\xi_j, \xi_{j'})$ into components measuring dependence over the four quadrants:

$$\begin{aligned}&\left[\mathrm{D}^{(+,+)}(\xi_j,\xi_{j'}),\mathrm{D}^{(-,+)}(\xi_j,\xi_{j'}),\mathrm{D}^{(-,-)}(\xi_j,\xi_{j'}),\mathrm{D}^{(+,-)}(\xi_j,\xi_{j'})\right]\\&:=\left[\int_{\mathbb{S}^2_{(+,+)}}a_1a_2\Gamma_{jj'}(d\mathbf{a}),\ \int_{\mathbb{S}^2_{(-,+)}}a_1a_2\Gamma_{jj'}(d\mathbf{a}),\ \int_{\mathbb{S}^2_{(-,-)}}a_1a_2\Gamma_{jj'}(d\mathbf{a}),\ \int_{\mathbb{S}^2_{(+,-)}}a_1a_2\Gamma_{jj'}(d\mathbf{a})\right].\end{aligned}$$

The corresponding estimators for the components are given by, respectively,

$$\begin{aligned}\widehat{\mathrm{D}}_n^{(+,+)}(\xi_j,\xi_{j'})&:=\frac{1}{k}\sum_{i=1}^n\frac{\hat{\xi}_{ij}}{\widehat{R}_i}\frac{\hat{\xi}_{ij'}}{\widehat{R}_i}I_{\widehat{R}_i\geq\widehat{R}_{(k)}}I_{\hat{\xi}_{ij}\geq0,\hat{\xi}_{ij'}\geq0},\\\widehat{\mathrm{D}}_n^{(-,+)}(\xi_j,\xi_{j'})&:=\frac{1}{k}\sum_{i=1}^n\frac{\hat{\xi}_{ij}}{\widehat{R}_i}\frac{\hat{\xi}_{ij'}}{\widehat{R}_i}I_{\widehat{R}_i\geq\widehat{R}_{(k)}}I_{\hat{\xi}_{ij}\leq0,\hat{\xi}_{ij'}\geq0},\\\widehat{\mathrm{D}}_n^{(-,-)}(\xi_j,\xi_{j'})&:=\frac{1}{k}\sum_{i=1}^n\frac{\hat{\xi}_{ij}}{\widehat{R}_i}\frac{\hat{\xi}_{ij'}}{\widehat{R}_i}I_{\widehat{R}_i\geq\widehat{R}_{(k)}}I_{\hat{\xi}_{ij}\leq0,\hat{\xi}_{ij'}\leq0},\\\widehat{\mathrm{D}}_n^{(+,-)}(\xi_j,\xi_{j'})&:=\frac{1}{k}\sum_{i=1}^n\frac{\hat{\xi}_{ij}}{\widehat{R}_i}\frac{\hat{\xi}_{ij'}}{\widehat{R}_i}I_{\widehat{R}_i\geq\widehat{R}_{(k)}}I_{\hat{\xi}_{ij}\geq0,\hat{\xi}_{ij'}\leq0}.\end{aligned}\tag{3.14}$$

Note that k in (3.14) is the same as in (3.2). In application, we first select k, the number of upper order statistics $\widehat{R}_{(i)}$ and then use it to compute (3.2) and (3.14). We will describe this with details in Section 3.6.

We establish the consistency of these estimators in the following corollary.

COROLLARY 1. *Under Assumptions 1, 2, and 3,*

$$\widehat{\mathrm{D}}_n^{(+,+)}(\xi_j,\xi_{j'})\xrightarrow{P}\mathrm{D}^{(+,+)}(\xi_j,\xi_{j'}),\quad\widehat{\mathrm{D}}_n^{(-,+)}(\xi_j,\xi_{j'})\xrightarrow{P}\mathrm{D}^{(-,+)}(\xi_j,\xi_{j'}),$$

$$\widehat{\mathrm{D}}_n^{(-,-)}(\xi_j,\xi_{j'})\xrightarrow{P}\mathrm{D}^{(-,-)}(\xi_j,\xi_{j'}),\quad\widehat{\mathrm{D}}_n^{(+,-)}(\xi_j,\xi_{j'})\xrightarrow{P}\mathrm{D}^{(+,-)}(\xi_j,\xi_{j'}).$$

Theorem 1 and Corollary 1 are proven in Section 3.5. Our approach to prove the consistency for the EDM is based on weak convergence of tail empirical measures. Set $\widehat{\mathbf{\Theta}}_i=\widehat{\mathbf{Y}}_i/||\widehat{\mathbf{Y}}_i||$. The estimator $\widehat{\mathrm{D}}_n(\xi_j, \xi_{j'})$ can then be written as an integral of a tail empirical measure, i.e.,

$$\widehat{\mathrm{D}}_n(\xi_j, \xi_{j'}) = \int_{\mathbb{S}^2} a_1 a_2 \widehat{\Gamma}_n(d\mathbf{a}), \quad \widehat{\Gamma}_n := \frac{1}{k} \sum_{i=1}^{n} I_{\widehat{\boldsymbol{\Theta}}_i} I_{\widehat{R}_i \geq \widehat{R}_{(k)}}.$$

The key argument to prove the consistency is therefore to show

$$\widehat{\Gamma}_n \Rightarrow \Gamma_{jj'} \quad \text{in } M_+\left(\mathbb{S}^2\right), \tag{3.15}$$

with $\Gamma_{jj'}$ in (3.1). Relation (3.15) is established by proving a series of weak convergence results.

We now turn to the asymptotic normality. The asymptotic normality of the estimator for the EDM is proven for i.i.d. bivariate observations in [5]. To show the asymptotic normality of an estimator based on heavy–tailed data, additional conditions are required even in fully observable i.i.d. settings. For example, for the Hill estimator, second–order regular variation with restrictions on the rate of k is assumed, see [39], [40], [41, 42]. The aforementioned condition is a univariate concept, which is not applicable to our context. Instead, we use a multivariate version of second–order regular variation, defined by [43]. With some constraint on k, i.e., $\sqrt{k}A(b(n/k)) \to 0$, where A is defined in formula (15) in [43], the multivariate second–order regular variation implies the following weaker condition, which is also assumed by [5].

ASSUMPTION 4. The R_i, $\boldsymbol{\Theta}_i$ satisfy

$$\sqrt{k}\left[\frac{n}{k}\Pr\left(\left(\frac{R_1}{b(n/k)}, \boldsymbol{\Theta}_1\right) \in \cdot\right) - c\nu_\alpha \times \Gamma_{jj'}\right] \xrightarrow{v} 0 \quad \text{in } M_+((0, \infty] \times \mathbb{S}^2).$$

Assumption 4 means that R_1 and $\boldsymbol{\Theta}_1$ are asymptotically independent. We emphasize that this assumption applies to population quantities, which are not observable in our setting. We now formulate the asymptotic normality of our estimator for the EDM, which is based on projections of functional data.

THEOREM 2. *Under Assumptions 2, 3 and 4*

$$\sqrt{k}\left(\widehat{\mathrm{D}}_n(\xi_j, \xi_{j'}) - \mathrm{D}(\xi_j, \xi_{j'})\right) \Rightarrow \mathcal{N}(0, \sigma^2),$$

where $\sigma^2 = \mathrm{Var}(\widetilde{\Theta}_1 \widetilde{\Theta}_2) > 0$*, with* $\widetilde{\Theta}_1$ *and* $\widetilde{\Theta}_2$ *being the components of a random vector with distribution* $\Gamma_{jj'}$.

3.4 Preliminary results

We put together several preliminary results in this section to avoid burdening the proofs in Section 3.5, so that readers can keep track of the main flow of the argument made in Section 3.5.

The first lemma follows from Lemma 3.7 of [44] and is needed to prove Lemma 2.

LEMMA 1. *Suppose random variables* $H_m(n)$*,* $m, n \geq 1$*, satisfy* $0 \leq H_m(n) \leq 1$ *and* $\forall\, m \geq 1, \quad H_m(n) \xrightarrow{P} 0, \quad \text{as } n \to \infty$*. Then,* $\sum_{m=1}^{\infty} 2^{-m} H_m(n) \xrightarrow{P} 0, \quad \text{as } n \to \infty$.

In the following lemma, we present a sufficient condition to guarantee the convergence between random measures defined on a nice space. We denote a locally compact topological space with countable base by $\mathbb{E}$. Following page 51 of [4], the vague metric $d(\cdot,\cdot)$ on $M_+(\mathbb{E})$ is defined by

$$d(\mu_1, \mu_2) = \sum_{i=1}^{\infty} \frac{|\mu_1(f_i) - \mu_2(f_i)| \wedge 1}{2^i}, \quad \mu_1, \mu_2 \in M_+(\mathbb{E}), \tag{3.1}$$

for some sequence of functions $f_i \in C_K^+(\mathbb{E})$ where $C_K^+(\mathbb{E})$ is the space of continuous functions with compact support on $\mathbb{E}$. By Lemma 1, the following is readily proven.

LEMMA 2. *Suppose that* μ_n*,* ν_n *are random measures in* $M_+(\mathbb{E})$*. If, for any* $f \in C_K^+(\mathbb{E})$*,* $|\mu_n(f) - \nu_n(f)| \xrightarrow{P} 0, \;\; n \to \infty$*, then* $d(\mu_n, \nu_n) \xrightarrow{P} 0$.

In the following lemma, we show that a continuous mapping with a compactness condition preserves convergence of random measures. Suppose that $\mathbb{E}_1$ and $\mathbb{E}_2$ are locally compact topological spaces with countable base. Denote by $\mathcal{K}(\mathbb{E})$ a set of all compact subsets of $\mathbb{E}$.

LEMMA 3. *Suppose that* $H : \mathbb{E}_1 \mapsto \mathbb{E}_2$ *is a continuous function such that*

$$H^{-1}(K_2) \in \mathcal{K}(\mathbb{E}_1), \quad \forall K_2 \in \mathcal{K}(\mathbb{E}_2). \tag{3.2}$$

If random measures μ_n, ν_n *in* $M_+(\mathbb{E}_1)$ *satisfy* $d(\mu_n, \nu_n) \xrightarrow{P} 0$, *as* $n \to \infty$, *then* $d(\mu_n \circ H^{-1}, \nu_n \circ H^{-1}) \xrightarrow{P} 0$, *in* $M_+(\mathbb{E}_2)$.

Proof. By Lemma 2, it suffices to show that, for any $f \in C_K^+(\mathbb{E}_2)$,

$$\mu_n \circ H^{-1}(f) - \nu_n \circ H^{-1}(f) \xrightarrow{P} 0. \tag{3.3}$$

Using the change of variables, we have $(\mu_n - \nu_n) \circ H^{-1}(f) = \int_{\mathbb{E}_2} f(e_2)(\mu_n - \nu_n) \circ H^{-1}(de_2) = \int_{\mathbb{E}_1} f(H(e_1))(\mu_n - \nu_n)(de_1)$. Thus, we have $(\mu_n - \nu_n) \circ H^{-1}(f) = (\mu_n - \nu_n)(f \circ H)$. Since f and H are both continuous, and with (3.2), we get $f \circ H \in C_K^+(\mathbb{E}_1)$, see page 142 of [4]. Then, since $d(\mu_n, \nu_n) \xrightarrow{P} 0$ by assumption, we get (3.3).

□

Consider the polar coordinate transform $T : [-\infty, \infty]^2 \setminus \{\mathbf{0}\} \mapsto (0, \infty] \times \mathbb{S}^2$ defined by, for $\mathbf{x} \in [-\infty, \infty]^2 \setminus \{\mathbf{0}\}$,

$$T(\mathbf{x}) = \left(\|\mathbf{x}\|, \frac{\mathbf{x}}{\|\mathbf{x}\|} \right). \tag{3.4}$$

Note that T is not bijective since its boundaries at infinity are included. Thus, Lemma 3 cannot be directly applied to T to show that it preserves convergence of random measures. Instead, we will show that by using, say, "restrict and then extend space" strategy, which is used in a different setting on page 176∼179 of [4]. We follow the technique in the proof of the next lemma.

LEMMA 4. *Suppose that random measures* μ_n, ν_n *satisfy*

$$d(\mu_n, \nu_n) \xrightarrow{P} 0, \quad \text{in } M_+([-\infty, \infty]^2 \setminus \{\mathbf{0}\}), \tag{3.5}$$

as $n \to \infty$. *Then,* $d(\mu_n \circ T^{-1}, \nu_n \circ T^{-1}) \xrightarrow{P} 0, \quad \text{in } M_+((0, \infty] \times \mathbb{S}^2)$.

Proof. Consider the transform $T^{'} : (-\infty,\infty)^2 \setminus \{\mathbf{0}\} \mapsto (0,\infty) \times \mathbb{S}^2$ defined by (3.4). Our first claim is that (3.5) implies

$$d(\mu_n, \nu_n) \xrightarrow{P} 0, \quad \text{in } M_+((-\infty,\infty)^2 \setminus \{\mathbf{0}\}). \tag{3.6}$$

Let $f_i \in C_K^+((-\infty,\infty)^2 \setminus \{\mathbf{0}\})$, and suppose that $K_i \in \mathcal{K}((-\infty,\infty)^2 \setminus \{\mathbf{0}\})$ is the compact support of f_i. Let $\tilde{f}_i := f_i(x) I_{x \in K_i}$, then $\tilde{f}_i \in C_K^+([-\infty,\infty]^2 \setminus \{\mathbf{0}\})$. Observe that $d(\mu_n,\nu_n) = \sum_{i=1}^{\infty} 2^{-i} |(\mu_n - \nu_n)(\tilde{f}_i)| = \sum_{i=1}^{\infty} 2^{-i} |(\mu_n - \nu_n)(f_i)| \xrightarrow{P} 0$, by (3.5), so we get (3.6).

Our second claim is that (3.6) implies

$$d(\mu_n \circ (T^{'})^{-1}, \nu_n \circ (T^{'})^{-1}) \xrightarrow{P} 0, \quad \text{in } M_+((0,\infty) \times \mathbb{S}^2). \tag{3.7}$$

This is readily proven by Lemma 3, since $T^{'}$ is continuous and satisfy (3.2).

The last step is now to extend $T^{'}$ to the bigger space, where ∞ is included. Let $f_i \in C_K^+((0,\infty] \times \mathbb{S}^2)$, and set $\|f_i\| = \sup f_i < \infty$. We define a smooth truncation function of r, for fixed M, δ, by

$$\phi(r; M, \delta) := I_{0<r\leq M} + \{-(r-M)/\delta + 1\} I_{M<r\leq M+\delta}.$$

Then, observe that

$$\begin{aligned}
& d(\mu_n \circ T^{-1}, \nu_n \circ T^{-1}) \\
& = \sum_{i=1}^{\infty} 2^{-i} |(\mu_n - \nu_n) \circ T^{-1}(f_i)| - \sum_{i=1}^{\infty} 2^{-i} |(\mu_n - \nu_n) \circ T^{-1}(f_i \phi)| \\
& + \sum_{i=1}^{\infty} 2^{-i} |(\mu_n - \nu_n) \circ T^{-1}(f_i \phi)| - \sum_{i=1}^{\infty} 2^{-i} |(\mu_n - \nu_n) \circ (T^{'})^{-1}(f_i \phi)| \\
& + \sum_{i=1}^{\infty} 2^{-i} |(\mu_n - \nu_n) \circ (T^{'})^{-1}(f_i \phi)| =: A + B + C.
\end{aligned}$$

Now, we will show that each of the components goes to 0. First, observe that

$$A \leq \sum_{i=1}^{\infty} 2^{-i} \left| \int_{(0,\infty]\times\mathbb{S}^2} f_i(r,\theta)(1-\phi(r))(\mu_n - \nu_n) \circ T^{-1}(dr, d\theta) \right|$$
$$\leq \sum_{i=1}^{\infty} 2^{-i} \|f_i\| \left| \int_{(M,\infty]\times\mathbb{S}^2} (\mu_n - \nu_n) \circ T^{-1}(dr, d\theta) \right|.$$

Taking a sufficiently large M, then A gets arbitrarily small. Next, for each M,

$$B \leq \sum_{i=1}^{\infty} 2^{-i} \|f_i\| \left| \int_{(0,M]\times\mathbb{S}^2} (\mu_n - \nu_n) \circ (T^{-1} - (T^{'})^{-1})(dr, d\theta) \right| = 0.$$

Since $f_i(r,\theta)\phi(r; M, \delta) \in C_K^+((0,\infty) \times \mathbb{S}^2)$, the last term C goes to 0 by (3.7).

$\square$

The next lemma shows that the distance between a population score and its corresponding approximation is asymptotically negligible.

LEMMA 5. *Under Assumptions 1, 2, for* $\alpha > 4$, $\mathrm{E}|\hat{\xi}_j - \xi_j| = O(n^{-1/2})$, *and for* $2 < \alpha < 4$, $\mathrm{E}|\hat{\xi}_j - \xi_j|^r = o(n^{-\kappa r/\beta})$, *for some* $r > 0$ *satisfying*

$$r < \frac{2\beta}{\beta + 2}, \tag{3.8}$$

where κ, β *are defined in* (3.12).

Proof. For $\alpha > 4$, by the Cauchy–Schwarz inequality, $|\hat{\xi}_j - \xi_j| \leq \|X\| \|\hat{v}_j - v_j\|$, so by Assumption 2,

$$\mathrm{E}|\hat{\xi}_j - \xi_j| \leq \{\mathrm{E}\|X\|^2\}^{1/2} \{\mathrm{E}\|\hat{v}_j - v_j\|^2\}^{1/2} = O(n^{-1/2}).$$

Now consider the case of $2 < \alpha < 4$. Since for any β, $\frac{2\beta}{\beta+2} < \beta$, condition (3.8), implies that $r < \beta$. Applying Hölder's inequality with $p = \beta/r > 1$ and $q = \beta/(\beta - r)$, we get $\mathrm{E}|\hat{\xi}_j - \xi_j|^r \leq \{\mathrm{E}\|\hat{v}_j - v_j\|^{\beta}\}^{\frac{r}{\beta}} \{\mathrm{E}\|X\|^{rq}\}^{1/q}$. Direct verification shows that condition (3.8) is equivalent to

$$\frac{2\beta^2}{(\beta+2)(\beta-r)} < 2,$$

which implies $rq < 2$. Hence, by Assumption 1, $\{\mathrm{E}\|X\|^{rq}\}^{1/q} < \infty$. Therefore, by (3.12), $\mathrm{E}|\hat{\xi}_j - \xi_j|^r = o(n^{-\kappa r/\beta})$.

□

In the following lemmas, we verify the continuity of functions that will be used in Section 3.5 with the continuous mapping theorem.

LEMMA 6. *Suppose that the map $H : M_+((0,\infty] \times \mathbb{S}^2) \times (0,\infty) \to M_+((0,\infty] \times \mathbb{S}^2)$, defined by for any measurable set $A \times B \subset (0,\infty] \times \mathbb{S}^2$,*

$$H(U,x)(A \times B) = U(xA \times B).$$

The map H is continuous at $(\nu_\alpha \times \Gamma_{jj'}, x)$.

Proof. Suppose $W_n \xrightarrow{v} \nu_\alpha \times \Gamma_{jj'}$ in $M_+((0,\infty] \times \mathbb{S}^2)$, and $x_n \to x$ in $(0,\infty)$. Then we must show that

$$H(W_n, x_n) = W_n((x_n \cdot) \times \cdot) \xrightarrow{v} H(\nu_\alpha \times \Gamma_{jj'}, x) = \nu_\alpha \times \Gamma_{jj'}((x \cdot) \times \cdot).$$

To verify this, it suffices to show that for any $f \in C_K^+((0,\infty] \times \mathbb{S}^2)$,

$$W_n((x_n \cdot) \times \cdot)(f) = \int_{(0,\infty] \times \mathbb{S}^2} f(t,\mathbf{a})\, W_n(x_n dt, d\mathbf{a}) = \int_{(0,\infty] \times \mathbb{S}^2} f(y/x_n, \mathbf{a})\, W_n(dy, d\mathbf{a})$$
$$\to \nu_\alpha \times \Gamma_{jj'}((x \cdot) \times \cdot)(f) = \int_{(0,\infty] \times \mathbb{S}^2} f(t,\mathbf{a})\, \nu_\alpha(x_n dt)\Gamma_{jj'}(d\mathbf{a}) = \int_{(0,\infty] \times \mathbb{S}^2} f(y/x,\mathbf{a})\, \nu_\alpha(dy)\Gamma_{jj'}(d\mathbf{a}).$$

The following verification is mostly based on pp. 83–84 of [4], whose test functions are univariate. Our test functions are however bivariate. We must employ a product metric to apply uniform continuity of the test functions.

First observe that

$$\begin{aligned}
&\left| \int_{(0,\infty]\times\mathbb{S}^2_+} f(y/x_n,\mathbf{a})\, W_n(dy,d\mathbf{a}) - \int_{(0,\infty]\times\mathbb{S}^2} f(y/x,\mathbf{a})\, \nu_\alpha(dy)\Gamma_{jj'}(d\mathbf{a}) \right| \\
&\qquad \le \left| \int_{(0,\infty]\times\mathbb{S}^2} f(y/x_n,\mathbf{a})\, W_n(dy,d\mathbf{a}) - \int_{(0,\infty]\times\mathbb{S}^2} f(y/x,\mathbf{a})\, W_n(dy,d\mathbf{a}) \right| \\
&\qquad\quad + \left| \int_{(0,\infty]\times\mathbb{S}^2} f(y/x,\mathbf{a})\, W_n(dy,d\mathbf{a}) - \int_{(0,\infty]\times\mathbb{S}^2} f(y/x,\mathbf{a})\, \nu_\alpha(dy)\Gamma_{jj'}(d\mathbf{a}) \right|.
\end{aligned}$$

Since $W_n \xrightarrow{v} \nu_\alpha \times \Gamma_{jj'}$ and $f(\frac{\cdot}{x},\cdot) \in C_K^+((0,\infty]\times\mathbb{S}^2)$, the second term of the right–hand side goes to zero. Now, we focus on the first term. Since f has compact support in $(0,\infty]\times\mathbb{S}^2$, we can take $\delta > 0$ such that the supports of $f(\frac{\cdot}{x},\cdot)$ and $f(\frac{\cdot}{x_n},\cdot)$, for large n, are contained in $[\delta,\infty]\times\mathbb{S}^2$. Then we get the bound

$$\begin{aligned}
&\left| \int_{(0,\infty]\times\mathbb{S}^2_+} f(y/x_n,\mathbf{a})\, W_n(dy,d\mathbf{a}) - \int_{(0,\infty]\times\mathbb{S}^2_+} f(y/x,\mathbf{a})\, W_n(dy,d\mathbf{a}) \right| \\
&\le \int_{[\delta,\infty]\times\mathbb{S}^2_+} |f(y/x_n,\mathbf{a}) - f(y/x,\mathbf{a})|\, W_n(dy,d\mathbf{a}) \\
&\le \sup_{y\ge\delta,\ \mathbf{a}\in\mathbb{S}^2} |f(y/x_n,\mathbf{a}) - f(y/x,\mathbf{a})|\, W_n([\delta,\infty]\times\mathbb{S}^2).
\end{aligned}$$

Since $W_n([\delta,\infty]\times\mathbb{S}^2)$ is bounded, it remains to show that as $x_n \to x$,

$$\sup_{y\ge\delta,\ \mathbf{a}\in\mathbb{S}^2} |f(y/x_n,\mathbf{a}) - f(y/x,\mathbf{a})| \to 0. \tag{3.9}$$

We use the fact that a continuous function with compact support is uniformly continuous. The metric on $(0,\infty]\times\mathbb{S}^2$ is given by $d_{\text{prod}}((u,\mathbf{a}),(v,\mathbf{b})) = d_{(0,\infty]}(u,v) + d_{\mathbb{S}^2}(\mathbf{a},\mathbf{b})$, see p.57 of [4]. Define the metric on $(0,\infty]$ by

$$d_{(0,\infty]}(u,v) = |u^{-1} - v^{-1}|,$$

for $u, v \in (0, \infty]$, which measures the distance between points in $(0, \infty]$ with one point compactification at ∞. Since $x_n \to x$ and $y \geq \delta_0$,

$$d_{\text{prod}}\left((y/x_n, \mathbf{a}), (y/x, \mathbf{a})\right) = \frac{|x_n - x|}{y} \leq \frac{|x_n - x|}{\delta_0} \to 0.$$

Therefore, by the uniform continuity of f, we get (3.9).

□

LEMMA 7. *The function g on $M_+\left((0, \infty] \times \mathbb{S}^2\right)$ defined by for any measurable sets $A \subset (0, \infty]$, $B \subset \mathbb{S}^2$, $g(U) = U\left(A \times B\right)$ is continuous at $\nu_\alpha \times \Gamma_{jj'}$.*

Proof. Suppose $W_n \xrightarrow{v} \nu_\alpha \times \Gamma_{jj'}$ in $M_+\left((0, \infty] \times \mathbb{S}^2\right)$. Since $A \times B$ is relatively compact in $(0, \infty] \times \mathbb{S}^2$, by Theorem 3.2 of [4] $g(W_n) = W_n\left(A \times B\right) \to g(\nu_\alpha \times \Gamma_{jj'}) = \nu_\alpha(A)\Gamma_{jj'}(B)$.

□

LEMMA 8. *The function h on $M_+\left(\mathbb{S}^2\right)$ defined by for $B \in \{\mathbb{S}^2, \mathbb{S}^2_{(+,+)}, \mathbb{S}^2_{(-,+)}, \mathbb{S}^2_{(-,-)}, \mathbb{S}^2_{(+,-)}\}$, $h(U) = \int_B \theta_1\theta_2 U(d\boldsymbol{\theta})$ is continuous at $\Gamma_{jj'}$.*

Proof. Suppose $W_n \xrightarrow{v} \Gamma_{jj'}$ in $M_+\left(\mathbb{S}^2\right)$. Consider a map $f : \mathbb{S}^2 \to \mathbb{R}$, defined by $f(\boldsymbol{\theta}) = \theta_1\theta_2 I_{\boldsymbol{\theta} \in B}$. Note that every continuous function on a compact space has compact support. Since f is continuous with compact support, by the definition of vague convergence,

$$h(W_n) = \int_B \theta_1\theta_2 W_n(d\boldsymbol{\theta}) \to h(\Gamma_{jj'}) = \int_B \theta_1\theta_2 \Gamma_{jj'}(d\boldsymbol{\theta}).$$

□

3.5 Proofs of the results of Section 3.3

Proof of Proposition 1: To prove the regular variation of $\pi(X)$ in $\mathbb{R}^d$, we will show that there exists a probability measure Γ on $\mathbb{S}^d$ and a regularly varying sequence $b(n)$ satisfying (3.4); for any $y > 0$,

$$n\Pr(\|\pi(X)\| > yb(n),\ \pi(X)/\|\pi(X)\| \in \cdot) \xrightarrow{w} cy^{-\alpha}\Gamma(\cdot) \ \text{ as } n \to \infty,$$

for some $c > 0$.

First, note that $\|\pi(X)\| > yb(n)$ and $\pi(X)/\|\pi(X)\| \in \cdot$ iff $(yb(n))^{-1}X \in \mathcal{A}_\pi(\cdot)$. Observe that, for any set S in $\mathcal{B}(\mathbb{S}^d)$,

$$n\Pr(\|\pi(X)\| > yb(n),\ \pi(X)/\|\pi(X)\| \in S) = n\Pr\left(\frac{X}{yb(n)} \in \mathcal{A}_\pi(S)\right).$$

To use (3.3) implied by the M_0 convergence, we must show that the $\mathcal{A}_\pi(S)$ are continuity sets of ν, i.e., $\nu(\partial\mathcal{A}_\pi(S)) = 0$. The verification uses the same idea described in the proof of Proposition 3.1 of [25], but the difference is that we work with the different projection $\pi(z)$ and its relevant set $\mathcal{A}_\pi(S)$.

By (3.6), we have

$$\partial\mathcal{A}_\pi(S) = \{z \in \mathbb{H} : \|\pi(z)\| = 1,\ \pi(z)/\|\pi(z)\| \in S\},$$

and

$$\partial(r\mathcal{A}_\pi(S)) = \{z \in \mathbb{H} : \|\pi(z)\| = r,\ \pi(z)/\|\pi(z)\| \in S\}.$$

Note that $\partial(r\mathcal{A}_\pi(S)) = r\partial\mathcal{A}_\pi(S)$, and the sets $\partial(r\mathcal{A}_\pi(S))$ are all disjoint in r. We assume $\nu(\partial\mathcal{A}_\pi(S)) > 0$ and get a contradiction. Since $\mathcal{A}_\pi(S) \supset \cup_{n\geq 1}\partial(n^{1/\alpha}\mathcal{A}_\pi(S))$, for all $\alpha > 0$, and ν is homogeneous,

$$\nu(\mathcal{A}_\pi(S)) \geq \sum_{n=1}^{\infty} \nu(n^{1/\alpha}\partial\mathcal{A}_\pi(S)) = \sum_{n=1}^{\infty} n^{-1}\nu(\partial\mathcal{A}_\pi(S)) = \infty.$$

This contradicts to the fact that ν is *boundedly finite*. Therefore, the $\mathcal{A}_\pi(S)$ are continuity sets of ν.

Now, by (3.3), we obtain

$$n\Pr(\|\pi(X)\| > yb(n),\ \pi(X)/\|\pi(X)\| \in S) \to \nu(y\mathcal{A}_\pi(S)) = y^{-\alpha}\nu(\mathcal{A}_\pi(S)).$$

Setting

$$\Gamma(\cdot) := \frac{\nu(\mathcal{A}_\pi(\cdot))}{c}, \qquad c = \nu(\mathcal{A}_\pi(\mathbb{S}^d)), \tag{3.1}$$

we get the claim.

Proof of Theorem 1:

Recall that

$$\mathbf{Y}_i = [\xi_{ij}, \xi_{ij'}]^\top,\ R_i = \|\mathbf{Y}_i\|,\ \mathbf{\Theta}_i = \mathbf{Y}_i/R_i,\ \widehat{\mathbf{Y}}_i = [\hat{\xi}_{ij}, \hat{\xi}_{ij'}]^\top,\ \widehat{R}_i = \|\widehat{\mathbf{Y}}_i\|,\ \widehat{\mathbf{\Theta}}_i = \widehat{\mathbf{Y}}_i/\widehat{R}_i.$$

Under Assumption 1, the $\mathbf{Y}_i$ are regularly varying with index $-\alpha$ by Proposition 1. More specifically, there exist a sequence $\{b(n)\}$ (the same as in (3.3)) and a probability angular measure $\Gamma_{jj'}$ defined as (3.1) satisfying

$$n\Pr\left(\left(\frac{R_i}{b(n)},\ \mathbf{\Theta}_i\right) \in \cdot\right) \xrightarrow{v} cv_\alpha \times \Gamma_{jj'} \quad \text{in } M_+((0,\infty] \times \mathbb{S}^2). \tag{3.2}$$

The constant c depends on the choice of $b(n)$. In the following, we assume $c = 1$ to keep the notation simple.

Our approach is to establish several weak convergences of tail empirical measures. We start with an empirical measure based on i.i.d. $\mathbf{Y}_i$:

$$U_n := \frac{1}{k}\sum_{i=1}^{n} I_{(R_i/b(n/k),\ \mathbf{\Theta}_i)} \Rightarrow \nu_\alpha \times \Gamma_{jj'} \quad \text{in } M_+\left((0,\infty] \times \mathbb{S}^2\right). \tag{3.3}$$

We then extend (3.3) to

$$\widehat{U}_n := \frac{1}{k}\sum_{i=1}^{n} I_{\left(\widehat{R}_i/b(n/k),\ \widehat{\mathbf{\Theta}}_i\right)} \Rightarrow \nu_\alpha \times \Gamma_{jj'} \quad \text{in } M_+\left((0,\infty] \times \mathbb{S}^2\right). \tag{3.4}$$

Since the $\widehat{\mathbf{Y}}_i$ are no longer independent, this requires techniques involving the Slutsky theorem. We further proceed to replace the unknown sequence $b(n/k)$ by its estimate $\widehat{R}_{(k)}$:

$$\widehat{U}_n^\star := \frac{1}{k}\sum_{i=1}^{n} I_{\left(\widehat{R}_i/\widehat{R}_{(k)},\, \widehat{\mathbf{\Theta}}_i\right)} \Rightarrow \nu_\alpha \times \Gamma_{jj'} \quad \text{in } M_+\left((0,\infty] \times \mathbb{S}^2\right). \tag{3.5}$$

Applying the continuous mapping theorem, we finally get (3.15), i.e.,

$$\widehat{\Gamma}_n = \frac{1}{k}\sum_{i=1}^{n} I_{\widehat{\mathbf{\Theta}}_i} I_{\widehat{R}_i \geq \widehat{R}_{(k)}} \Rightarrow \Gamma_{jj'} \quad \text{in } M_+\left(\mathbb{S}^2\right).$$

The consistency of $\widehat{\mathrm{D}}_n(\xi_j, \xi_{j'})$ for $\mathrm{D}(\xi_j, \xi_{j'})$ is then established because $\widehat{\mathrm{D}}_n(\xi_j, \xi_{j'}) = \int_{\mathbb{S}^2} a_1 a_2 \widehat{\Gamma}_n(d\mathbf{a})$.

We now present a series of the results mentioned above, of which Proposition 1 is the most essential and important step toward Theorem 1. The following lemma verifies (3.3), which is readily proven from (3.2) by Theorem 5.3 (ii) of [4].

LEMMA 1. *Under Assumption 1, relation* (3.3) *holds.*

The next result shows that the infeasible samples $\mathbf{Y}_i$ in (3.3) can be replaced by their approximations $\widehat{\mathbf{Y}}_i$.

PROPOSITION 1. *Under Assumptions 1, 2, and 3, relation* (3.4) *holds.*

Proof. By Lemma 1 and the Slutsky theorem, it suffices to prove that

$$d(\widehat{U}_n, U_n) = d\left(\frac{1}{k}\sum_{i=1}^{n} I_{\left(\widehat{R}_i/b(n/k),\, \widehat{\mathbf{\Theta}}_i\right)}, \frac{1}{k}\sum_{i=1}^{n} I_{(R_i/b(n/k),\, \mathbf{\Theta}_i)}\right) \xrightarrow{P} 0. \tag{3.6}$$

To show (3.6), we set $\widehat{V}_n := \frac{1}{k}\sum_{i=1}^{n} I_{\widehat{\mathbf{Y}}_i/b(n/k)}, \quad V_n := \frac{1}{k}\sum_{i=1}^{n} I_{\mathbf{Y}_i/b(n/k)}$, and prove

$$d(\widehat{V}_n, V_n) = d\left(\frac{1}{k}\sum_{i=1}^{n} I_{\widehat{\mathbf{Y}}_i/b(n/k)}, \frac{1}{k}\sum_{i=1}^{n} I_{\mathbf{Y}_i/b(n/k)}\right) \xrightarrow{P} 0. \tag{3.7}$$

Applying the polar transformation defined in (3.4), we get (3.6) from (3.7) by Lemma 4.

To prove (3.7), it suffices to show that, by Lemma 2, for any $f \in C_K^+([-\infty, \infty]^2 \setminus \{\mathbf{0}\})$, and any $\tau > 0$,

$$\Pr\left(\left|\frac{1}{k}\sum_{i=1}^{n} f\left(\frac{\widehat{\mathbf{Y}}_i}{b(n/k)}\right) - \frac{1}{k}\sum_{i=1}^{n} f\left(\frac{\mathbf{Y}_i}{b(n/k)}\right)\right| > \tau\right) \to 0. \tag{3.8}$$

Since f has compact support in $[-\infty, \infty]^2 \setminus \{\mathbf{0}\}$, set

$$a := \inf\{\|s\| : s \in \operatorname{supp}(f)\} > 0. \tag{3.9}$$

To prove (3.8), we consider a decomposition using the following sets. For $0 < \eta < a/2$, set

$$A_n(k) := \left\{1 \le i \le n : \left\|\frac{\widehat{\mathbf{Y}}_i}{b(n/k)} - \frac{\mathbf{Y}_i}{b(n/k)}\right\| \le \eta,\ \left\|\frac{\mathbf{Y}_i}{b(n/k)}\right\| \ge a - \eta\right\},$$

$$B_n(k) := \left\{1 \le i \le n : \left\|\frac{\widehat{\mathbf{Y}}_i}{b(n/k)} - \frac{\mathbf{Y}_i}{b(n/k)}\right\| \le \eta,\ \left\|\frac{\mathbf{Y}_i}{b(n/k)}\right\| < a - \eta\right\},$$

and

$$C_n(k) := \left\{1 \le i \le n : \left\|\frac{\widehat{\mathbf{Y}}_i}{b(n/k)} - \frac{\mathbf{Y}_i}{b(n/k)}\right\| > \eta\right\}.$$

Then, we have

$$\begin{aligned}
&\Pr\left(\left|\frac{1}{k}\sum_{i=1}^{n} f\left(\frac{\widehat{\mathbf{Y}}_i}{b(n/k)}\right) - \frac{1}{k}\sum_{i=1}^{n} f\left(\frac{\mathbf{Y}_i}{b(n/k)}\right)\right| > \tau\right) \\
&\le \Pr(S(A_n) > \tau/3) + \Pr(S(B_n) > \tau/3) + \Pr(S(C_n) > \tau/3),
\end{aligned}$$

where

$$S(A_n) = \frac{1}{k}\sum_{i \in A_n(k)} \left| f\left(\frac{\widehat{\mathbf{Y}}_i}{b(n/k)}\right) - f\left(\frac{\mathbf{Y}_i}{b(n/k)}\right)\right|,$$

and $S(B_n)$ and $S(C_n)$ are defined analogously with $\sum_{i \in B_n(k)}$ and $\sum_{i \in C_n(k)}$, respectively.

We will show that each of the three parts goes to 0. We first investigate $\Pr(S(A_n) > \tau/3)$. Since f is uniformly continuous,

$$w_\eta(f) := \sup_{\|\mathbf{x}-\mathbf{y}\|\leq\eta,\ \mathbf{x},\mathbf{y}\in[-\infty,\infty]^2\setminus\{\mathbf{0}\}} |f(\mathbf{x}) - f(\mathbf{y})| \to 0,\ \eta \to 0.$$

Observe that

$$S(A_n) \leq w_\eta(f)\frac{1}{k}\#\left\{1 \leq i \leq n : \left\|\frac{\mathbf{Y}_i}{b(n/k)}\right\| \geq a - \eta\right\} = w_\eta(f)U_n(E_{a-\eta}),$$

with the measure U_n defined in (3.3), and with the set $E_b \subset (0,\infty] \times \mathbb{S}^2$ defined by

$$E_b = \left\{(r,\theta) \in (0,\infty] \times \mathbb{S}^2 : r \geq b\right\},\ \ b > 0.$$

Now consider the function g on $M_+\left((0,\infty] \times \mathbb{S}^2\right)$, defined by, for any measurable set $A \subset (0,\infty]$, $g(U) = U\left(A \times \mathbb{S}^2\right)$. Then, by Lemma 7 and the continuous mapping theorem, for a fixed η, $U_n(E_{a-\eta}) \xrightarrow{P} \nu_\alpha(a-\eta,\infty] = (a-\eta)^{-\alpha}$. Therefore,

$$\limsup_{n\to\infty} \Pr(S(A_n) > \tau/3) \leq \Pr\left(w_\eta(f)(a-\eta)^{-\alpha} > \tau/3\right) \leq \Pr\left(w_\eta(f) > 2^{-\alpha}a^\alpha\tau/3\right).$$

By taking sufficiently small η, we can ensure that $\Pr\left(w_\eta(f) > 2^{-\alpha}a^\alpha\tau/3\right) = 0$, hence $\lim_{n\to\infty}\Pr(S(A_n) > \tau/3) = 0$.

Next, we consider the second probability in the decomposition. Observe that for each $i \in B_n(k)$,

$$\left\|\frac{\widehat{\mathbf{Y}}_i}{b(n/k)}\right\| \leq \left\|\frac{\widehat{\mathbf{Y}}_i}{b(n/k)} - \frac{\mathbf{Y}_i}{b(n/k)}\right\| + \left\|\frac{\mathbf{Y}_i}{b(n/k)}\right\| < a,\quad \left\|\frac{\mathbf{Y}_i}{b(n/k)}\right\| < a - \eta.$$

Thus, the two points $\widehat{\mathbf{Y}}_i/b(n/k), \mathbf{Y}_i/b(n/k)$ are outside of the support of f for all $i \in B_n(k)$, so $S(B_n) = 0$ by construction, and so $\Pr(S(B_n) > \tau/3) = 0$.

It remains to show that for any $\eta > 0$, $\lim_{n\to\infty} \Pr(S(C_n) > \tau/3) = 0$. Set

$$\|f\|_\infty = \sup_{\mathbf{x}\in[-\infty,\infty]^2\setminus\{\mathbf{0}\}} |f(\mathbf{x})|. \tag{3.10}$$

First, consider the case of $\alpha > 4$. By Markov's inequality,

$$\begin{aligned}
\Pr(S(C_n) > \tau/3) &\leq \frac{3}{\tau k}\mathrm{E}\left[\sum_{i\in C_n(k)} \left| f\left(\frac{\widehat{\mathbf{Y}}_i}{b(n/k)}\right) - f\left(\frac{\mathbf{Y}_i}{b(n/k)}\right)\right|\right] \\
&\leq \frac{6\|f\|_\infty}{\tau k}\mathrm{E}\left[\sum_{i=1}^{n} I_{\|\widehat{\mathbf{Y}}_i - \mathbf{Y}_i\| > \eta b(n/k)}\right] \\
&\leq \frac{6\|f\|_\infty}{\tau}\frac{n}{k}\Pr\left(\|\widehat{\mathbf{Y}}_i - \mathbf{Y}_i\| > \eta b(n/k)\right) \\
&\leq \frac{6\|f\|_\infty}{\tau\eta}\frac{n}{kb(n/k)}\mathrm{E}||\widehat{\mathbf{Y}}_i - \mathbf{Y}_i||.
\end{aligned}$$

Since all norms in $\mathbb{R}^2$ are equivalent, we get

$$||\widehat{\mathbf{Y}}_i - \mathbf{Y}_i|| \leq C\left(|\hat{\xi}_{ij} - \xi_{ij}| + |\hat{\xi}_{ij'} - \xi_{ij'}|\right), \tag{3.11}$$

for some $C > 0$. Since $\mathrm{E}||\widehat{\mathbf{Y}}_i - \mathbf{Y}_i|| \leq O(n^{-1/2})$ by Lemma 5, we have $\Pr(S(C_n) > \tau/3) = O\left(n^{1/2}/\{kb(n/k)\}\right)$. By Assumption 3 and (3.5), $\Pr(S(C_n) > \tau/3) = o(1)$.

Now consider the case of $\alpha \in (2,4)$. We will use Lemma 5, which refers to relation (3.12). Observe that since $\beta < \alpha/2 < 2$ in (3.12), it holds that $\frac{2\beta}{\beta+2} < 1$. This implies that r satisfying (3.8) also satisfies $r < 1$. Applying Markov's and Lyapunov's inequalities, we thus obtain

$$\begin{aligned}
\Pr(S(C_n) > \tau/3) &\leq \Pr\left(\frac{2\|f\|_\infty}{k}\sum_{i=1}^{n} I_{\|\widehat{\mathbf{Y}}_i - \mathbf{Y}_i\| > \eta b(n/k)} > \frac{\tau}{3}\right) \\
&\leq \frac{6^r\|f\|_\infty^r n^r}{\tau^r k^r}\mathrm{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} I_{\|\widehat{\mathbf{Y}}_i - \mathbf{Y}_i\| > \eta b(n/k)}\right)^r\right] \\
&\leq \frac{6^r\|f\|_\infty^r n^r}{\tau^r k^r}\left\{\mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} I_{\|\widehat{\mathbf{Y}}_i - \mathbf{Y}_i\| > \eta b(n/k)}\right]\right\}^r \\
&= \frac{6^r\|f\|_\infty^r n^r}{\tau^r k^r}\Pr\left(\|\widehat{\mathbf{Y}}_i - \mathbf{Y}_i\| > \eta b(n/k)\right)^r.
\end{aligned}$$

Applying Markov's inequality with the same r again and (3.11), we obtain

$$\Pr(S(C_n) > \tau/3) \leq c\frac{n^r}{k^r\{b(n/k)\}^{r^2}}\left\{\mathrm{E}\left[\max\left(|\hat{\xi}_{ij} - \xi_{ij}|, |\hat{\xi}_{ij'} - \xi_{ij'}|\right)^r\right]\right\}^r,$$

for some $c > 0$. Then by Lemma 5 and (3.5)

$$\Pr(S(C_n) > \tau/3) = o\left(\frac{n^{r-\kappa r^2/\beta}}{k^r\{b(n/k)\}^{r^2}}\right) = o\left(\frac{n^{r-\kappa r^2/\beta - r^2/\alpha}}{k^{r-r^2/\alpha}}\right).$$

Let

$$\gamma = \frac{r - \frac{\kappa r^2}{\beta} - \frac{r^2}{\alpha}}{r - \frac{r^2}{\alpha}} = \frac{1 - \frac{r}{\alpha} - \frac{\kappa r}{\beta}}{1 - \frac{r}{\alpha}}.$$

Then, γ is smaller than 1 for all $2 < \alpha < 4$, as κ/β gets close to 0, and it attains its smallest value as κ/β approaches its largest possible value, i.e., $1 - 2/\alpha$, see (3.12). We now set a lower bound of γ as a function of r for α fixed,

$$\gamma_L(r;\alpha) := \frac{1 - \frac{r}{\alpha} - (1 - \frac{2}{\alpha})r}{1 - \frac{r}{\alpha}} = \frac{\alpha - \alpha r + r}{\alpha - r}. \tag{3.12}$$

Since $2\beta/(\beta + 2)$ in (3.8) is an increasing function of β and attains its upper limit when $\beta = \alpha/2$, see (3.12), we obtain $r < 2\alpha/(\alpha + 4)$. Then, since $\gamma_L(r;\alpha)$ is an decreasing function of r, γ can be arbitrarily close to $\gamma_L(2\alpha/(\alpha+4);\alpha) = (6-\alpha)/(\alpha+2)$. Thus, by Assumption 3, $k >> n^\gamma$, and we get $\Pr(S(C_n) > \tau/3) = o(1)$.

$\square$

The following proposition is used to prove the asymptotic normality in Theorem 2. We put it in this section to help readers follow its proof easily since it uses several elements of the proof of Proposition 1. The claim is similar to (3.6), but $1/k$ is replaced by a suitably chosen power of k, so a more delicate argument is needed.

PROPOSITION 2. *Suppose that Assumptions 2 3 and 4 hold. Then,*

$$d\left(\frac{1}{\sqrt{k}}\sum_{i=1}^{n} I_{\left(\widehat{R}_i/b(n/k),\ \widehat{\mathbf{\Theta}}_i\right)}, \frac{1}{\sqrt{k}}\sum_{i=1}^{n} I_{(R_i/b(n/k),\ \mathbf{\Theta}_i)}\right) \xrightarrow{P} 0.$$

Proof. We follow the approach used in the proof of Proposition 1, so we skip fully analogous parts and focus on the new aspects. To get the claim, it suffices to show that

$$\Pr\left(\frac{1}{\sqrt{k}}\sum_{i=1}^{n}\left|f\left(\frac{\widehat{\mathbf{Y}}_i}{b(n/k)}\right) - f\left(\frac{\mathbf{Y}_i}{b(n/k)}\right)\right| > \tau\right) \to 0,$$

for every $f \in C_K^+([-\infty,\infty]^2 \setminus \{\mathbf{0}\})$. For $0 < \eta < a/2$, with a defined in (3.9), set

$$A_n(k) := \left\{1 \le i \le n : \left\|\frac{\widehat{\mathbf{Y}}_i}{k^p b(n/k)} - \frac{\mathbf{Y}_i}{k^p b(n/k)}\right\| \le \eta,\ \left\|\frac{\mathbf{Y}_i}{k^p b(n/k)}\right\| \ge a-\eta\right\},$$
$$B_n(k) := \left\{1 \le i \le n : \left\|\frac{\widehat{\mathbf{Y}}_i}{k^p b(n/k)} - \frac{\mathbf{Y}_i}{k^p b(n/k)}\right\| \le \eta,\ \left\|\frac{\mathbf{Y}_i}{k^p b(n/k)}\right\| < a-\eta\right\},$$

and

$$C_n(k) := \left\{1 \le i \le n : \left\|\frac{\widehat{\mathbf{Y}}_i}{k^p b(n/k)} - \frac{\mathbf{Y}_i}{k^p b(n/k)}\right\| > \eta\right\},$$

where p is a positive constant such that $p\min\{r,1\} = 1/2$ for some r satisfying (3.8). Except for the factor k^p, these sets of indexes are analogous to those used in the proof of Proposition 1. Then, we have

$$\Pr\left(\frac{1}{\sqrt{k}}\sum_{i=1}^{n}\left|f\left(\frac{\widehat{\mathbf{Y}}_i}{b(n/k)}\right)-f\left(\frac{\mathbf{Y}_i}{b(n/k)}\right)\right|>\tau\right)$$
$$\leq \Pr(S(A_n)>\tau/3)+\Pr(S(B_n)>\tau/3)+\Pr(S(C_n)>\tau/3),$$

where

$$S(A_n)=\frac{1}{\sqrt{k}}\sum_{i\in A_n(k)}\left|f\left(\frac{\widehat{\mathbf{Y}}_i}{b(n/k)}\right)-f\left(\frac{\mathbf{Y}_i}{b(n/k)}\right)\right|,$$

and $S(B_n)$ and $S(C_n)$ are defined analogously with $\sum_{i\in B_n(k)}$ and $\sum_{i\in C_n(k)}$, respectively. Our claim is that each of the three terms converges to 0. Before we proceed, we note some results about p in k^p to facilitate the understanding of the proofs;

$$pr=\frac{1}{2}\ \text{ for } 2<\alpha<4, \quad p\geq\frac{1}{2}\ \text{ for } \alpha>2. \tag{3.13}$$

To see this, observe that $\beta<\alpha/2$ in (3.12) and $\frac{2\beta}{\beta+2}$ in (3.8) is increasing of β. It thus holds that $r<\frac{2\beta}{\beta+2}<\frac{2\alpha}{\alpha+4}$. This implies that $0<r<1$ for $2<\alpha<4$, and $0<r<2$ for $\alpha>2$.

First, observe that

$$\begin{aligned} S(A_n) &\leq 2\|f\|_\infty\sqrt{k}\frac{1}{k}\sum_{i=1}^{n} I_{\|\mathbf{Y}_i/k^p b(n/k)\|\geq a-\eta} \\ &= c\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{n} I_{R_i/b(n/k)\geq k^p(a-\eta)}-\nu_\alpha(k^p(a-\eta),\infty]\right)+ck^{1/2-p\alpha}(a-\eta)^{-\alpha}, \end{aligned}$$

where $\|f\|_\infty$ is defined in (3.10) and c is a positive constant. The last term goes to 0 since $p\alpha>p\geq 1/2$ for $\alpha>2$. Now, we focus on the first term. Assumption 4 implies

$$\mu_n:=\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{n} I_{R_i/b(n/k)}-\nu_\alpha\right)\xrightarrow{P}0. \tag{3.14}$$

Consider the map g_M on $M_+(0,\infty]$, defined by $g_M(U)=U([M,\infty])$. We must show that $g_{k^p(a-\eta)}(\mu_n)\xrightarrow{P}0$. This follows from the following more general argument. We have a sequence of signed measures on $(0,\infty]$, such that $\mu_n \xrightarrow{P} 0$. Since we can decompose μ_n into positive and negative parts, we can assume that the μ_n are positive. For $a_n \to \infty$ (in our case $a_n = k^p(a-\eta)$), we claim that $\mu_n([a_n,\infty]) \xrightarrow{P} 0$. By Lemma 7, the map g_M is continuous, so for each fixed M, $\mu_n([M,\infty]) \xrightarrow{P} 0$. For sufficiently large n, $a_n > 1$, $\mu_n([a_n,\infty]) \le \mu_n([1,\infty])$, and the claim follows.

Next, we obtain $\Pr(S(B_n) > \tau/3) = 0$ in the same manner in Proposition 1.

For $S(C_n)$, we first consider the case of $\alpha > 4$. 0bserve that by Markov's inequality,

$$\Pr(S(C_n) > \tau/3) \le \frac{1}{\sqrt{k}} \frac{6\|f\|_\infty}{\tau\eta} \frac{n}{k^p b(n/k)} \mathrm{E}||\widehat{\mathbf{Y}}_i - \mathbf{Y}_i||.$$

By Lemma 5, $\mathrm{E}||\widehat{\mathbf{Y}}_i - \mathbf{Y}_i|| = O(n^{-1/2})$, so $\Pr(S(C_n) > \tau/3) = O\left(n^{1/2-1/\alpha}/k^{1/2+p-1/\alpha}\right)$. We must thus verify that $n^{1/2-1/\alpha}/k^{1/2+p-1/\alpha} \to 0$. We know that $n^\gamma/k \to 0$ if $\gamma > \gamma(\alpha)$. We use the factorization

$$\frac{n^{1/2-1/\alpha}}{k^{1/2+p-1/\alpha}} = \left(\frac{n^\gamma}{k}\right)^{\frac{1}{2\gamma}-\frac{1}{\alpha\gamma}} \left(\frac{1}{k}\right)^{\frac{1}{2}+p-\frac{1}{\alpha}-\frac{1}{2\gamma}+\frac{1}{\alpha\gamma}}.$$

Since $\alpha > 2$, $\frac{1}{2\gamma} - \frac{1}{\alpha\gamma} > 0$, so we must be able to claim that $\frac{1}{2}+p-\frac{1}{\alpha}-\frac{1}{2\gamma}+\frac{1}{\alpha\gamma} > 0$. Since $p \ge \frac{1}{2}$, this will follow from $1 - \frac{1}{\alpha} - \frac{1}{\gamma}\left(\frac{1}{2} - \frac{1}{\alpha}\right) > 0$. A few algebraic manipulations show that the above inequality is equivalent to $\gamma > \frac{\alpha-2}{2\alpha-2} = \gamma(\alpha)$. For the case of $\alpha \in (2,4)$, we apply Markov's and Lyapunov's inequalities, just as we did in Proposition 1. Then, by Lemma 5 and (3.13) we obtain

$$\Pr(S(C_n) > \tau/3) = o\left(\frac{n^{r-\kappa r^2/\beta - r^2/\alpha}}{k^{r/2+pr^2-r^2/\alpha}}\right) = o\left(\frac{n^{r-\kappa r^2/\beta-r^2/\alpha}}{k^{r-r^2/\alpha}}\right).$$

It is verified at the end of the proof of Proposition 1 that the last quantity tends to zero under Assumption 3.

□

The next lemma will be used in Proposition 3 to replace $b(n/k)$ in (3.4) with $\widehat{R}_{(k)}$.

LEMMA 2. *Under Assumptions 1, 2, and 3,* $\widehat{R}_{(k)}/b(n/k) \xrightarrow{P} 1$.

Proof. Fix $\varepsilon > 0$ and set

$$P_+(n) = \Pr\left(\frac{\widehat{R}_{(k)}}{b(n/k)} > 1+\varepsilon\right), \quad P_-(n) = \Pr\left(\frac{\widehat{R}_{(k)}}{b(n/k)} < 1-\varepsilon\right).$$

Observe that

$$\begin{aligned} P_+(n) &= \Pr\left(I_{\widehat{R}_{(k)}/b(n/k)}(1+\varepsilon,\infty] = 1\right) \\ &\leq \Pr\left(\frac{1}{k}\sum_{i=1}^{n} I_{\widehat{R}_i/b(n/k)}(1+\varepsilon,\infty] \geq 1\right) \\ &= \Pr\left(\widehat{U}_n\left((1+\varepsilon,\infty]\times\mathbb{S}^2\right) \geq 1\right). \end{aligned}$$

A similar argument shows that $P_-(n) \leq \Pr\left(\widehat{U}_n\left((1-\varepsilon,\infty]\times\mathbb{S}^2\right) < 1\right)$. The claim follows because by Lemma 7 and the continuous mapping theorem, we obtain $\widehat{U}_n((1+\varepsilon,\infty]\times\mathbb{S}^2) \xrightarrow{P} \nu_\alpha(1+\varepsilon,\infty] = (1+\varepsilon)^{-\alpha} < 1$; $\widehat{U}_n((1-\varepsilon,\infty]\times\mathbb{S}^2) \xrightarrow{P} \nu_\alpha(1-\varepsilon,\infty] = (1-\varepsilon)^{-\alpha} > 1$.

□

PROPOSITION 3. *Under Assumptions 1, 2, and 3, relation* (3.5) *holds.*

Proof. By Proposition 1 and Lemma 2, we obtain joint weak convergence $\left(\widehat{U}_n, \frac{\widehat{R}_{(k)}}{b(n/k)}\right) \Rightarrow (\nu_\alpha \times \Gamma_{jj'}, 1)$ in $M_+\left((0,\infty]\times\mathbb{S}^2\right)\times(0,\infty)$. Consider the operator $H : M_+\left((0,\infty]\times\mathbb{S}^2\right)\times(0,\infty) \to M_+((0,\infty]\times\mathbb{S}^2)$, defined by for any measurable set $A\times B \subset (0,\infty]\times\mathbb{S}^2$, $H(U,x)(A\times B) = U(xA\times B)$. Since $H\left(\widehat{U}_n, \widehat{R}_{(k)}/b(n/k)\right) = \frac{1}{k}\sum_{i=1}^{n} I_{\left(\widehat{R}_i/\widehat{R}_{(k)},\ \widehat{\boldsymbol{\Theta}}_i\right)}$, $H\left(\nu_\alpha\times\Gamma_{jj'}, 1\right) = \nu_\alpha\times\Gamma_{jj'}$, we get (3.5) by Lemma 6 and the continuous mapping theorem.

□

Proof of Theorem 1: Consider the map $g : M_+\left((0,\infty]\times\mathbb{S}^2\right) \to M_+\left(\mathbb{S}^2\right)$, defined by for any measurable set $A \subset \mathbb{S}^2$, $g(U) = U\left([1,\infty]\times A\right)$. Then, by Lemma 7 and the continuous mapping theorem, we obtain (3.15) from (3.5). Now we consider the map h on $M_+(\mathbb{S}^2)$ defined by $h(U) = \int_{\mathbb{S}^2}\theta_1\theta_2 U(d\boldsymbol{\theta})$. By Lemma 8 and the continuous mapping theorem, we obtain, from (3.15),

$\int_{\mathbb{S}^2} \theta_1\theta_2\widehat{\Gamma}_n(d\boldsymbol{\theta}) \Rightarrow \int_{\mathbb{S}^2} \theta_1\theta_2\Gamma_{jj'}(d\boldsymbol{\theta})$. Since

$$\int_{\mathbb{S}^2} \theta_1\theta_2\widehat{\Gamma}_n(d\boldsymbol{\theta}) = \frac{1}{k}\sum_{i=1}^{n} I_{R_i \geq R_{(k)}} \int_{\mathbb{S}^2} \theta_1\theta_2 I_{\boldsymbol{\Theta}_i \in d\boldsymbol{\theta}} = \widehat{\mathrm{D}}_n(\hat{\xi}_{ij}, \hat{\xi}_{ij'}),$$

we get the claim.

Proof of Corollary 1: Consider the map h on $M_+(\mathbb{S}^2)$ defined by

$$h(S) = \int_{\mathbb{S}^2_{(+,+)}} \theta_1\theta_2 S(d\boldsymbol{\theta}).$$

Applying the map to (3.15), we obtain the consistency of $\widehat{\mathrm{D}}_n^{(+,+)}(\xi_j, \xi_{j'})$ for $\mathrm{D}^{(+,+)}(\xi_j, \xi_{j'})$, by Lemma 8 and the continuous mapping theorem. The consistency of the remaining estimators can be proven in the same way, just using different quadrant domains in the map h.

Proof of Theorem 2: Define the empirical process based on the sample scores by

$$W_n(t) = \frac{1}{\sigma\sqrt{k}}\sum_{i=1}^{n}\left(\widehat{\Theta}_{i1}\widehat{\Theta}_{i2} - \mathrm{E}\left[\widetilde{\Theta}_1\widetilde{\Theta}_2\right]\right) I_{\widehat{R}_i/b(n/k) \geq t^{-1/\alpha}}, \quad t \geq 0.$$

The main argument to prove the asymptotic normality is the weak convergence of W_n to the standard Brownian motion W;

$$W_n \Rightarrow W, \quad \text{in } D[0,\infty), \tag{3.15}$$

where $D[0,\infty)$ is the usual Skorokhod space. Once we verify (3.15), then by Lemma 2 we obtain the joint convergence

$$\left(W_n(\cdot), \left(\frac{\widehat{R}_{(k)}}{b(n/k)}\right)^{-\alpha}\right) \Rightarrow (W(\cdot), 1), \quad \text{in } D[0,\infty) \times [0,\infty).$$

Applying the composition map $(x(\cdot), c) \mapsto x(c)$, we conclude that

$$\sqrt{k}\left(\widehat{\mathrm{D}}_n(\xi_j, \xi_{j'}) - \mathrm{E}\left[\widetilde{\Theta}_1\widetilde{\Theta}_2\right]\right) = \sigma W_n\left(\left(\frac{\widehat{R}_{(k)}}{b(n/k)}\right)^{-\alpha}\right) \Rightarrow \sigma W(1).$$

The general strategy is thus similar to the one employed to prove Theorem 1 in [5]. However, in our setting, new arguments are needed to establish relations (3.17) and (3.18). These terms are zero in the proof of [5].

Now, to show (3.15), consider the following decomposition

$$\begin{aligned} W_n(t) &= \frac{1}{\sigma\sqrt{k}} \sum_{i=1}^{n} \left(\Theta_{i1}\Theta_{i2} - \mathrm{E}\left[\widetilde{\Theta}_1\widetilde{\Theta}_2\right] \right) I_{R_i/b(n/k) \geq t^{-1/\alpha}} \\ &+ \frac{1}{\sigma\sqrt{k}} \sum_{i=1}^{n} \left(\widehat{\Theta}_{i1}\widehat{\Theta}_{i2} I_{\widehat{R}_i/b(n/k) \geq t^{-1/\alpha}} - \Theta_{i1}\Theta_{i2} I_{R_i/b(n/k) \geq t^{-1/\alpha}} \right) \\ &+ \frac{1}{\sigma\sqrt{k}} \sum_{i=1}^{n} \mathrm{E}\left[\widetilde{\Theta}_1\widetilde{\Theta}_2\right] \left(I_{R_i/b(n/k) \geq t^{-1/\alpha}} - I_{\widehat{R}_i/b(n/k) \geq t^{-1/\alpha}} \right). \end{aligned}$$

We will verify that

$$\frac{1}{\sigma\sqrt{k}} \sum_{i=1}^{n} \left(\Theta_{i1}\Theta_{i2} - \mathrm{E}\left[\widetilde{\Theta}_1\widetilde{\Theta}_2\right] \right) I_{R_i/b(n/k) \geq (\cdot)^{-1/\alpha}} \Rightarrow W, \quad \text{in } D[0,\infty), \tag{3.16}$$

and for any $s \geq 0$,

$$\sup_{0 \leq t \leq s} \left| \frac{1}{\sigma\sqrt{k}} \sum_{i=1}^{n} \left(\widehat{\Theta}_{i1}\widehat{\Theta}_{i2} I_{\widehat{R}_i/b(n/k) \geq t^{-1/\alpha}} - \Theta_{i1}\Theta_{i2} I_{R_i/b(n/k) \geq t^{-1/\alpha}} \right) \right| \xrightarrow{P} 0; \tag{3.17}$$

$$\mathrm{E}\left[\widetilde{\Theta}_1\widetilde{\Theta}_2\right] \sup_{0 \leq t \leq s} \left| \frac{1}{\sigma\sqrt{k}} \sum_{i=1}^{n} \left(I_{R_i/b(n/k) \geq t^{-1/\alpha}} - I_{\widehat{R}_i/b(n/k) \geq t^{-1/\alpha}} \right) \right| \xrightarrow{P} 0. \tag{3.18}$$

We begin with (3.16). Since the empirical process in (3.16) is based on i.i.d. population scores, if we verify

$$\sqrt{k}\left[\frac{n}{k} \Pr\left(\left(\frac{R_1}{b(n/k)}, \mathbf{\Theta}_1 \right) \in \cdot \right) - \frac{n}{k} \Pr\left(\frac{R_1}{b(n/k)} \in \cdot \right) \times \Gamma_{jj'} \right] \xrightarrow{v} 0, \quad \text{in } M_+((0,\infty] \times \mathbb{S}^2), \tag{3.19}$$

then (3.16) readily holds by Theorem 1 of [5]. Their theorem is proven for nonnegative random vectors, but the proof also works for random vectors in $\mathbb{R}^d$, with a small modification.

To prove (3.19), we use the equivalent conditions for vague convergence presented in Theorem 3.2 of [4]. Take any relatively compact set $B \in (0, \infty]$. Then, $B \times \mathbb{S}^2$ is also relatively compact in $(0, \infty] \times \mathbb{S}^2$, so we obtain from Assumption 4,

$$\sqrt{k}\left[\frac{n}{k}\Pr\left(\frac{R_1}{b(n/k)} \in B\right) - \nu_\alpha(B)\right] \to 0. \tag{3.20}$$

The constant c in Assumption 4 depends on the choice of $b(n)$, so we set $c = 1$ for simplicity. Now, take any relatively compact set $A \times S \in (0, \infty] \times \mathbb{S}^2$, and observe that

$$\begin{aligned}
&\sqrt{k}\left[\frac{n}{k}\Pr\left(\left(\frac{R_1}{b(n/k)}, \boldsymbol{\Theta}_1\right) \in A \times S\right) - \frac{n}{k}\Pr\left(\frac{R_1}{b(n/k)} \in A\right) \times \Gamma_{jj'}(S)\right] \\
&= \sqrt{k}\left[\frac{n}{k}\Pr\left(\left(\frac{R_1}{b(n/k)}, \boldsymbol{\Theta}_1\right) \in A \times S\right) - \nu_\alpha(A)\Gamma_{jj'}(S)\right] \\
&+ \sqrt{k}\left[\nu_\alpha(A) - \frac{n}{k}\Pr\left(\frac{R_1}{b(n/k)} \in A\right)\right]\Gamma_{jj'}(S) \to 0.
\end{aligned}$$

The first term vanishes by Assumption 4. Also, since A is relatively compact in $(0, \infty]$ and $0 \leq \Gamma_{jj'}(S) \leq 1$, the second term goes to 0 by (3.20).

For (3.17) and (3.18), we Proposition 2, i.e.,

$$\frac{1}{\sqrt{k}}\sum_{i=1}^{n} I_{\left(\widehat{R}_i/b(n/k),\ \widehat{\boldsymbol{\Theta}}_i\right)} - \frac{1}{\sqrt{k}}\sum_{i=1}^{n} I_{(R_i/b(n/k),\ \boldsymbol{\Theta}_i)} \xrightarrow{P} 0. \tag{3.21}$$

Consider the map $h : M_+\left((0, \infty] \times \mathbb{S}^2\right) \to M_+(0, \infty]$, defined by $h(U) = \int_{\mathbb{S}^2} \theta_1\theta_2 U(dr, d\boldsymbol{\theta})$. Applying h to (3.21), by Lemma 8 and the continuous mapping theorem we obtain

$$\phi_n := \frac{1}{\sqrt{k}}\sum_{i=1}^{n}\left(\widehat{\Theta}_{i1}\widehat{\Theta}_{i2} I_{\widehat{R}_i/b(n/k)} - \Theta_{i1}\Theta_{i2} I_{R_i/b(n/k)}\right) \xrightarrow{P} 0. \tag{3.22}$$

We thus have a sequence of signed measures on $(0, \infty]$, such that $\phi_n \xrightarrow{P} 0$. Since a signed measure can be decomposed into positive and negative parts, we can assume that the ϕ_n are positive. Now, consider the map g_M on $M_+(0, \infty]$, defined by $g_M(U) = U([M, \infty])$. By Lemma 7, the map g_M is continuous, so for each fixed M, $\phi_n([M, \infty]) \xrightarrow{P} 0$. Therefore, for any $s \geq 0$, taking M such that

Table 3.1: Time periods related to the subprime mortgage crisis.

Designation	Time span	Sample size n (days)
Before	07/05/2006 - 09/28/2007	313
During	10/01/2007 - 02/27/2009	351
After 1	03/02/2009 - 07/30/2010	356
After 2	08/02/2010 - 12/30/2011	358

$M > s$, we obtain

$$\sup_{0\le t\le s}\left|\frac{1}{\sigma\sqrt{k}}\sum_{i=1}^{n}\left(\widehat{\Theta}_{i1}\widehat{\Theta}_{i2}I_{\widehat{R}_i/b(n/k)\ge t^{-1/\alpha}} - \Theta_{i1}\Theta_{i2}I_{R_i/b(n/k)\ge t^{-1/\alpha}}\right)\right| \le \phi_n([M,\infty]) \stackrel{P}{\to} 0.$$

Similarly, considering the map $\ell : M_+\left((0,\infty]\times\mathbb{S}^2\right) \to M_+(0,\infty]$, defined by $\ell(U) = \int_{\mathbb{S}^2} U(dr, d\boldsymbol{\theta})$, we conclude (3.18).

3.6 Application to intraday returns

In this section, we quantify the extremal dependence between scores of cumulative intraday returns (CIDRs) for Walmart and IBM stocks, taken from July 05, 2006 to Dec 30, 2011. We define CIDRs as follows: Denote by $P_i(t)$ the price of an asset on trading day i at time t. For the assets in our example, t is time in minutes between 9:30 and 16:00 EST (NYSE opening times) rescaled to the unit interval $(0,1)$. We define the CIDR on day i as the curve

$$R_i(t) = \ln P_i(t) - \ln P_i(0).$$

In practice, $P_i(0)$ is the price after the first minute of trading. The curves R_i show how the return accumulates over the trading day, see, e.g., Figure 3.2.

[45] sought to identify the curves R_i that are in some sense extreme. They did so by looking for curves for which scores $\hat{\xi}_{ij}$ are extreme for some $j = 1, 2$ or 3, or by looking at the norm $\|X_i\| = \{\sum_{j=1}^{p}\xi_{ij}^2\}^{1/2}$.

A question we seek to investigate is if the financial crisis of 2008 affected the extremal dependence between the scores of the CIDRs. Over the last decade, the 2008 crisis has been extensively studied in finance and economics literature, see, e.g., [46], [47] and [48] who cite many references. We consider four time intervals, "before", "during", "after1", and "after2", defined in Table 3.1. For each interval, we compute estimates of the EDM for the three pairs of the first three scores, ξ_1, ξ_2, ξ_3, by which the shapes of the observed CIDR curves are encoded: $R_i(t) \approx \sum_{j=1}^{3} \hat{\xi}_{ij} \hat{v}_j(t)$. We fix the Euclidean norm in this application.

Hill plots of the sample scores $\hat{\xi}_{ij}$ for $j = 1, 2, 3$ are shown in Figure 2.3 in Section 2.1. The Hill plot is a tool commonly used to detect the presence of heavy tails approximately following a Pareto distribution for large values, see, for example, page 80 of [4]. Figure 2.3 indicates that it is reasonable to assume that the scores follow Pareto distributions with the tail index between 2 and 4 since stable horizontal lines lie between these values. Since we cannot guarantee that the first three scores are always tail equivalent, we first transform the scores so that they all have the same tail index α. We use a power transformation approach, similar to that described in Section A.1, but here we transform the scores to have $\alpha = 3$ since our theory requires $\alpha > 2$, see Assumption 3. To compute (3.14), we must choose k, the number of upper order statistics $\widehat{R}_{(i)}$. We use a data-driven method proposed by [49]. It is based on the scaling property of the exponent measure: $\nu(t\cdot) = t^{-\alpha}\nu(\cdot)$. More specifically, using the weak convergence result (3.5) proven by Proposition 3 in Section 3.5, we obtain

$$\frac{u^{\alpha} \widehat{U}_n^{\star}\left([u, \infty) \times \mathbb{S}^2\right)}{\widehat{U}_n^{\star}\left([1, \infty) \times \mathbb{S}^2\right)} \approx 1,$$

where u is in a neighborhood of 1. We then graph, for each fixed k,

$$\left\{ \left(u, \frac{u^{\alpha} \widehat{U}_n^{\star}\left([u, \infty) \times \mathbb{S}^2\right)}{\widehat{U}_n^{\star}\left([1, \infty) \times \mathbb{S}^2\right)} \right),\ 0.1 \leq u \leq 5 \right\},$$

and choose k that makes the ratio to hover around 1 for most of the values of u.

Tables 3.2 reports estimates for the three pairs of the first three scores, ξ_1, ξ_2, ξ_3, for the Walmart stock. Table A.1 reports analogous information for the IBM stock. We note that $\mathrm{D}^{*p}(\xi_j, \xi_{j'})$, for

Table 3.2: Estimates of EDM for Walmart stock. Standard errors in parentheses are computed using Theorem 2.

	$\mathrm{D}^{(+,+)}(\xi_1, \xi_2)$	$\mathrm{D}^{(-,+)}(\xi_1, \xi_2)$	$\mathrm{D}^{(-,-)}(\xi_1, \xi_2)$	$\mathrm{D}^{(+,-)}(\xi_1, \xi_2)$
Before	0.07 (0.02)	-0.06 (0.02)	0.07 (0.02)	-0.09 (0.02)
During	0.11 (0.02)	-0.06 (0.01)	0.07 (0.02)	-0.06 (0.01)
After 1	0.13 (0.03)	-0.07 (0.02)	0.07 (0.03)	-0.04 (0.02)
After 2	0.09 (0.03)	-0.05 (0.02)	0.03 (0.02)	-0.10 (0.03)
	$\mathrm{D}^{(+,+)}(\xi_1, \xi_3)$	$\mathrm{D}^{(-,+)}(\xi_1, \xi_3)$	$\mathrm{D}^{(-,-)}(\xi_1, \xi_3)$	$\mathrm{D}^{(+,-)}(\xi_1, \xi_3)$
Before	0.09 (0.03)	-0.05 (0.02)	0.04 (0.02)	-0.04 (0.02)
During	0.10 (0.02)	-0.07 (0.02)	0.05 (0.02)	-0.07 (0.02)
After 1	0.06 (0.02)	-0.11 (0.03)	0.07 (0.03)	-0.07 (0.03)
After 2	0.07 (0.02)	-0.08 (0.02)	0.05 (0.02)	-0.05 (0.02)
	$\mathrm{D}^{(+,+)}(\xi_2, \xi_3)$	$\mathrm{D}^{(-,+)}(\xi_2, \xi_3)$	$\mathrm{D}^{(-,-)}(\xi_2, \xi_3)$	$\mathrm{D}^{(+,-)}(\xi_2, \xi_3)$
Before	0.08 (0.03)	-0.09 (0.03)	0.05 (0.02)	-0.05 (0.02)
During	0.10 (0.02)	-0.08 (0.02)	0.07 (0.02)	-0.07 (0.02)
After 1	0.10 (0.02)	-0.08 (0.02)	0.10 (0.02)	-0.10 (0.02)
After 2	0.07 (0.02)	-0.08 (0.02)	0.07 (0.02)	-0.05 (0.02)

$*p \in \{(+,+),(-,-)\}$, has positive values because ξ_j and $\xi_{j'}$ have the same signs, and $\mathrm{D}^{*n}(\xi_j, \xi_{j'})$, for $*n \in \{(-,+),(+,-)\}$, has negative values because ξ_j and $\xi_{j'}$ have the opposite signs. Tables 3.2 and A.1 also report estimated standard errors of the estimates for EDM. Using Theorem 2, the standard error was computed by $\hat{\sigma}/\sqrt{k}$ where

$$\hat{\sigma}^2 = \frac{1}{k}\sum_{i=1}^{n}\left(\frac{\xi_{ij}}{R_i}\frac{\xi_{ij'}}{R_i}\right)^2 I_{R_i \geq R_{(k)}} - \left(\frac{1}{k}\sum_{i=1}^{n}\frac{\xi_{ij}}{R_i}\frac{\xi_{ij'}}{R_i} I_{R_i \geq R_{(k)}}\right)^2. \tag{3.1}$$

First, we see that there are apparent differences in estimates over the four quadrants for each time period and each pair. This indicates that extremal dependence could provide different information depending on quadrants, so it is more useful to obtain the EDM for each quadrant rather than the EDM integrated over all quadrants when measurements take on values in $\mathbb{R}^d$. We observe that the differences between the four periods are not significant, and they are within two estimated standard errors. As documented in Section A.2, formula (3.1) tends to produce slightly underestimated standard errors. Thus the available data do not provide evidence for differences between the periods. This is probably due to small sample sizes. Estimators of any form of extreme behavior require large sample sizes because only the most extremal observations matter. With these caveats, it is nevertheless interesting to examine the patterns whose statistical significance cannot be claimed, but which provide some exploratory insights.

We then look at the results for Walmart. We see that estimates for "during" and "after 1" in general have higher absolute values than those for "before" and "after 2" for all the three pairs. This means that the crisis increased the level of extremal dependence, and its impact continued for about a year after the crisis. For IBM stock, such a tendency is observed only for pair (ξ_2, ξ_3). For the other pairs, there is no distinguishable differences or trends over the four periods. To investigate the impact of the crisis more precisely, we interpret the EDM for each pair of the first three scores since each score quantifies a different characteristic of the shape of a curve. First, we examine the EDM between ξ_1 and ξ_2, both of which together have a considerable contribution to the shape of CIDR curves. For Walmart, $\mathrm{D}^{(+,+)}(\xi_1, \xi_2)$ at "during", "after 1", and for IMB, $\mathrm{D}^{(-,+)}(\xi_1, \xi_2)$ at "before" are relatively strong. Such a dependence means that an extremely high monotonic trend and a strong reversion are closely associated. Next, for ξ_1 and ξ_3, each of which quantifies a monotonic trend and a pronounced swing, respectively, the extremal dependence for Walmart are affected by the crisis: the estimates for the four quadrants increased for "during" and "after 1". For ξ_2 and ξ_3, the crisis increases the extremal dependence for Walmart again, implying that the chance of a pronounced inflection is highly related with a strong swing during that time.

The scores ξ_j have different variances and looking at the dependence measure between them may not take into account the effect of the different variances and their estimation. The normalized scores $Z_j = \xi_j / \sqrt{\lambda_j}$ have variance 1. We repeated the application to Walmart and IBM returns using $\hat{Z}_{ij} = \hat{\xi}_{ij} / \hat{\lambda}_j^{1/2}$ in place of the $\hat{\xi}_{ij}$. We got basically the same results, reported in Tables A.7 and A.8. An intuitive explanation might be that the EDM describes dependence between large values in the orthogonal directions j and j'. If these values are rescaled, the dependence should not change.

One may consider various extensions of the analysis presented above for which suitable theory would need to be developed. For example extending the seminal work of [50], who study dependence between pairs of stocks, might be of particular interest.

An important point to note is that our theory is valid under the assumption that the curves R_i are independent. By construction and by the results of [15], it is reasonable to assume that they form a stationary functional time series. The results of [51] indicate that they are in some sense “uncorrelated", just like point-to-point returns. They could however be dependent in some nonlinear, GARCH-type, way. Even in such a case, the analysis of extreme values focuses on very few observations generally separated by long time intervals, so the extremal scores are nearly independent, a common assumption in extreme value theory. It might nevertheless be of interest to investigate under what temporal dependence assumptions our results would remain valid.

3.7 A simulation study

In this section, we investigate finite sample performances of the estimator of the EDM computed from sample scores by means of a simulation study.

The design of our study is as follows. We generate a sample of functions of the form

$$X_i(t) = \sum_{j=1}^{3} \xi_{ij} v_j(t), \quad 0 < t \leq 1, \quad 1 \leq i \leq n,$$

where the v_j are the FPCs of the Wiener Process, i.e.,

$$v_j(t) = \frac{\sqrt{2}}{\left(j - \frac{1}{2}\right)\pi} \sin\left(\left(j - \frac{1}{2}\right)\pi t\right), \quad j = 1, 2, 3.$$

This choice is motivated by the observation that the estimated FPCs of the data we consider in Section 3.6 are similar to the above v_j, see Figure 3.1. This reflects the well-known fact that, to a rough approximation, stock prices follow a random walk.

For the scores $\mathbf{Y}_i = [\xi_{i1}, \xi_{i2}, \xi_{i3}]^\top$, $1 \leq i \leq n$, we generate random vectors that are regularly varying with the true tail index $\alpha = 3$. We must generate these vectors in such a way that the

Table 3.3: Theoretical values of EDM for each case and each pair of the first three population scores, ξ_1, ξ_2, ξ_3.

n	$\mathrm{D}(\xi_1,\xi_2)$	$\mathrm{D}^{(+,+)}(\xi_1,\xi_2)$	$\mathrm{D}^{(-,+)}(\xi_1,\xi_2)$	$\mathrm{D}^{(-,-)}(\xi_1,\xi_2)$	$\mathrm{D}^{(+,-)}(\xi_1,\xi_2)$
Case1	0.000	0.000	0.000	0.000	0.000
Case2	0.000	0.100	-0.100	0.100	-0.100
Case3	0.000	0.125	-0.125	0.125	-0.125
n	$\mathrm{D}(\xi_1,\xi_3)$	$\mathrm{D}^{(+,+)}(\xi_1,\xi_3)$	$\mathrm{D}^{(-,+)}(\xi_1,\xi_3)$	$\mathrm{D}^{(-,-)}(\xi_1,\xi_3)$	$\mathrm{D}^{(+,-)}(\xi_1,\xi_3)$
Case1	0.000	0.000	0.000	0.000	0.000
Case2	0.000	0.075	-0.075	0.075	-0.075
Case3	0.000	0.125	-0.125	0.125	-0.125
n	$\mathrm{D}(\xi_2,\xi_3)$	$\mathrm{D}^{(+,+)}(\xi_2,\xi_3)$	$\mathrm{D}^{(-,+)}(\xi_2,\xi_3)$	$\mathrm{D}^{(-,-)}(\xi_2,\xi_3)$	$\mathrm{D}^{(+,-)}(\xi_2,\xi_3)$
Case1	0.000	0.000	0.000	0.000	0.000
Case2	0.000	0.115	-0.115	0.115	-0.115
Case3	0.000	0.125	-0.125	0.125	-0.125

the theoretical value of the EDM can be computed analytically, so that we can see how close the estimated EDM is to the true value. To construct vectors of scores with a known population EDM, we start with Z_1, Z_2, Z_3 that are i.i.d. random variables following a generalized Pareto distribution, $\Pr(Z > z) = (1 + \xi(z-\mu)/\sigma)^{-1/\xi}$, with location $\mu = 0$, shape $\xi = 1/3$, and scale $\sigma = 1$. Next, suppose that U_1, U_2, U_3 are i.i.d. random variables that take values -1 or 1, each with probability $1/2$. We consider the following three cases for the $\mathbf{Y}_i$.

Case 1 [Independence] The $\mathbf{Y}_i$ are i.i.d. random variables generated from

$$[U_1Z_1,\ U_2Z_2,\ U_3Z_3]^\top.$$

Case 2 The $\mathbf{Y}_i$ are i.i.d. random variables generated from

$$[U_1Z_1,\ 2U_2Z_1,\ 3U_3Z_1]^\top.$$

Case 3 [full dependence] The $\mathbf{Y}_i$ are i.i.d. random variables generated from

$$[U_1Z_1,\ U_2Z_1,\ U_3Z_1]^\top.$$

Table 3.4: Empirical biases (standard errors) of the estimator of the EDM for **Case 1** [Independence]

n	$\mathrm{D}(\xi_1,\xi_2)$	$\mathrm{D}^{(+,+)}(\xi_1,\xi_2)$	$\mathrm{D}^{(-,+)}(\xi_1,\xi_2)$	$\mathrm{D}^{(-,-)}(\xi_1,\xi_2)$	$\mathrm{D}^{(+,-)}(\xi_1,\xi_2)$
200	0.000 (0.04)	0.035 (0.02)	-0.033 (0.01)	0.034 (0.02)	-0.033 (0.01)
600	0.000 (0.02)	0.027 (0.01)	-0.027 (0.01)	0.027 (0.01)	-0.026 (0.01)
1000	0.000 (0.02)	0.024 (0.01)	-0.025 (0.01)	0.024 (0.01)	-0.025 (0.01)
n	$\mathrm{D}(\xi_1,\xi_3)$	$\mathrm{D}^{(+,+)}(\xi_1,\xi_3)$	$\mathrm{D}^{(-,+)}(\xi_1,\xi_3)$	$\mathrm{D}^{(-,-)}(\xi_1,\xi_3)$	$\mathrm{D}^{(+,-)}(\xi_1,\xi_3)$
200	0.000 (0.04)	0.023 (0.02)	-0.024 (0.01)	0.024 (0.02)	-0.023 (0.01)
600	0.000 (0.01)	0.017 (0.01)	-0.017 (0.01)	0.017 (0.01)	-0.017 (0.03)
1000	0.000 (0.01)	0.015 (0.01)	-0.015 (0.01)	0.016 (0.01)	-0.015 (0.01)
n	$\mathrm{D}(\xi_2,\xi_3)$	$\mathrm{D}^{(+,+)}(\xi_2,\xi_3)$	$\mathrm{D}^{(-,+)}(\xi_2,\xi_3)$	$\mathrm{D}^{(-,-)}(\xi_2,\xi_3)$	$\mathrm{D}^{(+,-)}(\xi_2,\xi_3)$
200	0.002 (0.04)	0.051 (0.02)	-0.047 (0.02)	0.048 (0.02)	-0.049 (0.02)
600	0.000 (0.03)	0.040 (0.02)	-0.040 (0.01)	0.041 (0.02)	-0.041 (0.01)
1000	0.000 (0.01)	0.037 (0.01)	-0.036 (0.01)	0.037 (0.01)	-0.037 (0.01)

Table 3.3 reports theoretical values of EDM for each case and each pair of the first three population scores in $\mathbf{Y}_i$. Note that for Case 1, we obtain $\mathrm{D}^*(\xi_1,\xi_2) = \mathrm{D}^*(\xi_2,\xi_3) = \mathrm{D}^*(\xi_1,\xi_3)$ = 0, where $* \in \{(+,+),(-,+),(-,-),(+,-)\}$, since all pairs are independent. For Case 2, each pair of the scores has a different value of extremal dependence; $\mathrm{D}^{*p}(\xi_1,\xi_2) = 1/10$, $\mathrm{D}^{*p}(\xi_2,\xi_3) = 3/26$, $\mathrm{D}^{*p}(\xi_1,\xi_3) = 3/40$, and $\mathrm{D}^{*n}(\xi_1,\xi_2) = -1/10$, $\mathrm{D}^{*n}(\xi_2,\xi_3) = -3/26$, $\mathrm{D}^{*n}(\xi_1,\xi_3) = -3/40$ where $*p \in \{(+,+),(-,-)\}$ and $*n \in \{(-,+),(-,+)\}$. For Case 3, since all pairs have asymptotic full dependence, each quadrant has a perfect extremal dependence; we obtain $\mathrm{D}^{*p}(\xi_1,\xi_2) = \mathrm{D}^{*p}(\xi_2,\xi_3) = \mathrm{D}^{*p}(\xi_1,\xi_3) = 1/8$, which is the maximum value that each quadrant can get. $\mathrm{D}^{*n}(\xi_1,\xi_2) = \mathrm{D}^{*n}(\xi_2,\xi_3) = \mathrm{D}^{*n}(\xi_1,\xi_3) = -1/8$, which is the minimum value. For all of the three cases, the EDM integrated over the four quadrants is 0, i.e., $\mathrm{D}(\xi_1,\xi_2) = \mathrm{D}(\xi_2,\xi_3) = \mathrm{D}(\xi_1,\xi_3) = 0$, since extremal dependence is symmetric over the four quadrants. For each case, we generate a sample of functions X_i using the FPCs v_1, v_2, v_3 and obtain sample scores. Then, based on the sample scores, we compute an estimate for the EDM using the estimator (3.14), and get the average and the estimated standard error of estimates over 1000 replications. We consider sample sizes $n = 200, 600, 1000$.

Tables 3.4, A.2, A.3 report empirical biases (average minus theoretical value) and standard errors computed as sample standard deviations of the 1000 replications. The results show that the bias tends to 0 as the sample size increases. The standard errors also decrease with increasing sample size, but slower, as expected. These results confirm the desirable performance of the estimator.

It is also of interest to compare the standard errors in Tables 3.4, A.2, A.3 with those obtained by the application of Theorem 2 and formula (3.1). Such comparisons are presented in Tables A.4, A.5, A.6 of Section A.2, which report the average of estimated standard errors computed using (3.1). This formula generally leads to standard errors which are smaller than the empirical standard errors, but the differences are small, especially for sample sizes used in Section 3.6.

There are some findings that are less expected. First, for the same sample size, the empirical bias depends on the structure of extremal dependence of population score $\mathbf{Y}_i$. We see from Tables A.2 and A.3 that Case 2 has relatively small biases, but Case 3 of full dependence has relatively large biases. This might be due to the fact that the theoretical EDM values in case 3 are larger in absolute value or due to some bias introduced by estimating the population scores. Within the same dependence structure (case), the bias seems to depend on pairs of scores. The biases for $\mathrm{D}^*(\xi_1, \xi_2)$, $\mathrm{D}^*(\xi_1, \xi_3)$ are larger in Case 3, but in Case 1 and Case 2, $\mathrm{D}^*(\xi_2, \xi_3)$ is larger. This again might be attributable to the discrepancy between the population scores and their approximations for each pair. Table 3.4 gives some idea of biases and standard errors that can be expected in the case of independence. We see that for the sample sizes considered in Section 3.6, the standard errors are somewhere in the range 0.02 to 0.04. For Case 3 of full dependence, the estimator computed from the sample scores underestimates the population EDMs.

Chapter 4

Consistency of the Hill estimator for time series observed with measurement errors

4.1 Introduction

Our objective is to establish the consistency of the Hill estimator applied to heavy-tailed time series observed with measurement errors, and to explore the impact of the errors in finite samples. Heavy–tailed time series commonly occur in fields such as finance, insurance, hydrology, and computer network traffic. The theory of regular variation provides a suitable mathematical framework. Suppose $X_1, \ldots, X_n$ is a realization of a strictly stationary time series with one-dimensional distribution function F_X, which has a regularly varying tail with index $\alpha > 0$, i.e. $P(X_i > x)$ behaves roughly like $x^{-\alpha}$, for large x. An estimate of α is essential for further inference related to extreme behavior of the time series. Risk measures, like the VaR or the expected shortfall, require an estimate $\hat{\alpha}$. The joint dependence structure is usually estimated by normalizing the data to the standard Fréchet distribution with $\alpha = 1$, which requires some estimate $\hat{\alpha}$. Many more applications are discussed in the monographs cited in the next paragraph. A well known and commonly used estimator of the index α is the Hill estimator, whose definition is recalled in Section 4.2. It is often used after an examination of the Hill plot, which is also a tool for detecting the presence of heavy tails. This paper studies the Hill estimator in situations in which the data are contaminated by measurement errors.

The Hill estimator is studied in practically all monographs on extreme value theory, see e.g. [16], [17], [4] and [2]. Its consistency for samples of i.i.d. random variables was first proven by [19]. Consistency of the Hill estimator has been established beyond the i.i.d. setting. [20] derives a general approach to establishing its consistency for stationary time series satisfying a certain mixing condition. [52] and [53] also consider mixing conditions. Other extensions are obtained by [21], [54], and [55], who show the consistency of the Hill estimator for time series using a tail empirical random measure proposed by [56]. Recently, [57] have proven the consistency of the Hill estimator for network data in a linear preferential attachment model.

In many applications, we do not observe $X_1, \ldots, X_n$ directly. Instead, the data are measured with noise, measurement or roundoff error. In other words, we observe $Y_i = X_i + \varepsilon_i,\ i = 1, 2, \ldots, n$, where $\{\varepsilon_i\}$ is an error process. The question is whether the Hill estimator computed from the Y_i will be still consistent for α under suitable assumptions on the errors ε_i. The research presented in this paper has been partially motivated by our work on modeling the stochastic behavior of internet traffic anomalies, whose arrival times are available only with a roundoff error. The database we have reports these times in 5 min. resolution.

Putting together known results, it is fairly straightforward to establish the consistency if the X_i are i.i.d., but a more in-depth investigation is needed when they follow a stochastic process model with a complex dependence structure. We investigate this question in the context of models considered by [55]. These include infinite moving averages with heavy-tailed innovations, bilinear processes driven by heavy-tailed noise variables, solutions of stochastic difference equations, the ARCH process of [58], and interarrival times of heavy-tailed hidden Markov chains. The models considered by [55] thus cover practically all known stochastic processes whose marginal distributions are regularly varying. Finite sample properties are investigated by means of a simulation study based on these models and by an application to the interarrival times of internet traffic anomalies. The main general conclusions of our research are as follows. 1) Asymptotically, the Hill estimator is robust to relatively large errors. 2) This robustness is confirmed in finite samples. 3) Five minute resolution is sufficient to estimate the tail index of the interarrival times of the anomalies we study.

Consistency of the Hill estimator based on data observed with measurement error has not been studied, but there has been considerable interest in a related problem, estimation of the end-point of a distribution function in the presence of additive observation errors, see [59], [60], [61], and [62]. They all consider Gaussian measurement errors. We, however, do not place this restriction on the errors. We assume a broader class of error distributions. Intuitively, we can relax the assumptions on the measurement errors because heavy-tailed X_i are “much larger" random variables than those with a finite end-point.

In Section 4.2, we introduce notation and assumptions. Our framework and main results are presented in Section 4.3. Finite sample performance of the Hill estimator in the presence of errors is investigated in Section 4.4. In Section 4.5, we present an application to the interarrival times of internet traffic anomalies. The proofs are developed in Section 4.7, preceded by some preparation in Section 4.6.

4.2 Notation and assumptions

We start by introducing some notation, generally following [22]. Recall that $X_1, \ldots, X_n$ are nonnegative random variables with common distribution F_X, which has regularly varying tail probabilities:

$$\bar{F}_X = 1 - F_X = P(X > \cdot) \in RV_{-\alpha},\ \alpha > 0. \tag{4.1}$$

We denote by X a generic random variable with the same distribution as each X_i. A function $U : \mathbb{R}_+ \to \mathbb{R}_+$ is regularly varying with index $\alpha > 0$, $U \in RV_{-\alpha}$, if for any $x > 0$,

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

For two functions $U, V : \mathbb{R}_+ \to \mathbb{R}_+$, we write $U(x) \sim V(x)$ if $U(x)/V(x) \to 1$, as $x \to \infty$, and $U(x) = o(V(x))$ if $U(x)/V(x) \to 0$, as $x \to \infty$.

The Hill estimator for the X_i is defined as

$$H_{k,n} := \frac{1}{k} \sum_{i=1}^{k-1} \log \frac{X_{(i)}}{X_{(k)}}, \tag{4.2}$$

with the convention that $X_{(1)}$ is the largest order statistic. We use definition (4.2) rather than the commonly used, asymptotically equivalent, definition with the $1/k$ replaced by $1/(k-1)$ because it leads to visually shorter formulas in the proofs. The consistency of the Hill estimator has been studied as the number of upper order statistics, k, tends to infinity with the sample size n, in such a way that $k/n \to 0$, i.e.

$$n \to \infty,\ k \to \infty,\ \frac{k}{n} \to 0. \tag{4.3}$$

We assume throughout the paper that condition (4.3) holds.

We consider the Hill estimator based on observations contaminated by measurement or other errors whose source is difficult to quantify. We thus assume that we observe $Y_i = X_i + \varepsilon_i,\ 1 \leq i \leq n$, where $\{\varepsilon_i\}$ are i.i.d. random errors following F_ε, and independent of the $\{X_i\}$. Then, the Hill estimator for the observations Y_i is defined as

$$\widehat{H}_{k,n} := \frac{1}{k} \sum_{i=1}^{k-1} \log \frac{Y_{(i)}}{Y_{(k)}}.$$

In our context, $\widehat{H}_{k,n}$ is the Hill estimator that can be actually used since what we observe are the Y_i. [55] show that the Hill estimator based on the X_i, $H_{k,n}$, is consistent for the tail index of $\bar{F}_X$, when it is applied to certain classes of heavy-tailed stationary processes. In our context, the X_i are unobservable. We want to establish conditions on F_ε under which $\widehat{H}_{k,n}$ is consistent for the tail index of $\bar{F}_X$. We solve this problem for all classes of the X_i considered by [55].

The approach of [55] is based on the weak convergence to the measure ν on $(0, \infty]$, satisfying $\int_1^\infty \log(u)\nu(du) < \infty$. One example of the measure ν is ν_α, defined by $\nu_\alpha(x, \infty] = x^{-\alpha},\ x > 0$. Our approach involves tail empirical random measures on $(0, \infty]$, based on the X_i, Y_i, and their weak convergence to the measure ν in $M_+(0, \infty]$, the space of Radon measures on $(0, \infty]$. We study the limit relations

$$\frac{1}{k} \sum_{i=1}^{n} I_{X_i/b(n/k)} \Rightarrow \nu, \quad \frac{1}{k} \sum_{i=1}^{n} I_{Y_i/b(n/k)} \Rightarrow \nu, \tag{4.4}$$

where $b(\cdot)$ is the quantile function, defined by

$$P(X_i > b(t)) = t^{-1}.$$

We investigate when the first convergence in (4.4) implies the second one. We use $\Rightarrow$ to denote weak convergence of random measures and $\stackrel{v}{\to}$ to denote vague convergence in $M_+(0,\infty]$, see [4].

We now state assumptions on the unobservable random variables X_i. We consider several conditions. We first assume that the unobservable variables are independent and have a common, regularly varying tail distribution. We then relax this assumption by considering three classes introduced by [55]. We first assume that the X_i follow a heavy-tailed stationary process which can be approximated by sequences of m-dependent random variables, and the m-dependent sequences carry enough information on the tail behavior of the original process. Then, we consider random coefficient autoregressive model. The final class consists of heavy-tailed hidden semi-Markov models.

ASSUMPTION 1. The X_i are nonnegative, independent random variables with common one-dimensional distribution F_X, which has regularly varying tail probabilities, i.e. (4.1) holds.

ASSUMPTION 2. The X_i form a stationary sequence, which can be approximated by stationary m-dependent sequences $\{X_i^{(m)}\}$ as follows. There exist Radon measures $\nu^{(m)}$, ν on $(0,\infty]$ with $\int_1^\infty \log(u)\nu(du) < \infty$ and $\nu^{(m)} \stackrel{v}{\to} \nu$, as $m \to \infty$. The X_i, $X_i^{(m)}$, and the ν, $\nu^{(m)}$ satisfy the following relations.

(a) For any fixed $m \geq 1$ (under (4.3)),

$$\frac{n}{k} P\left(\frac{X_i^{(m)}}{b(n/k)} \in \cdot \right) \stackrel{v}{\to} \nu^{(m)}.$$

(b) For any $\tau > 0$,

$$\lim_{m\to\infty} \limsup_{n\to\infty} \frac{n}{k} P\left(\left| \frac{X_i^{(m)}}{b(n/k)} - \frac{X_i}{b(n/k)} \right| > \tau \right) = 0.$$

(c) For each $m \geq 1$, the function $y \mapsto \nu^{(m)}(y, \infty]$ is right–continuous.

Condition (c) is not assumed by [55]. We need it to deal with the impact of the measurement errors. This condition is however not restrictive in practice because in all examples considered by [55], the functions $y \mapsto \nu^{(m)}(y, \infty]$ are continuous.

ASSUMPTION 3. The X_i form a stationary sequence, which satisfies the stochastic autoregressive equation

$$X_i = A_i X_{i-1} + B_i, \ -\infty < i < \infty,$$

where $\{(A_i, B_i), \ -\infty < i < \infty\}$ are i.i.d. $\mathbb{R}_+^2$-valued random pairs satisfying the following conditions. There exists $\alpha > 0$ with

$$EA_0^\alpha = 1, \ \ EA_0^\alpha \log^+ A_0 < \infty, \ \ 0 < EB_0^\alpha < \infty,$$

where $\log^+ x = \log x \vee 0$, $B_0/(1 - A_0)$ is nondegenerate, and the conditional distribution of $\log A_0$ given $A_0 \neq 0$ is nonlattice.

The conditions imposed on (A_i, B_i) ensure that the X_i are regularly varying, see Lemma 5 (i).

The final class we consider consists of hidden semi-Markov Models. These models generalize the commonly used hidden Markov models, and have recently found application in biology, computer science, operations research and meteorology, see e.g. [63] and [64]. The heavy-tailed hidden Markov has one or more states following heavy–tailed distributions. We first state its building blocks and then state the assumption. Let $\{J_n, n \geq 0\}$ be an ergodic, m-state Markov chain on the state space $\{1, 2, \ldots, m\}$ with the stationary distribution $\tilde{\pi} = (\pi_1, \ldots, \pi_m)$, and $P = \{p_{ij}, 1 \leq i, j \leq m\}$ be the transition probability matrix of the chain. Suppose $\{D_n^{(r)}, n \geq 0\}$, $r = 1, 2, \ldots, m$, are i.i.d. holding time random variables with common distributions $\{q_n^{(r)}, n \geq 0\}$, for each r. Define $\{V_i, i \geq 0\}$ by

$$V_i = \sum_{n=0}^{\infty} J_n 1_{[\sum_{l=0}^{n-1} D_l^{(j_l)} \leq i \leq \sum_{l=0}^{n} D_l^{(j_l)}]},$$

and define for $i \geq 0$,

$$X_i = F_{V_i}^{\leftarrow}(U_i), \tag{4.5}$$

where the U_i are i.i.d. uniform random variables with support $[0,1]$, and $F_1, \ldots, F_m$ are distributions on $\mathbb{R}_+$. The $\{U_i, i \geq 0\}, \{J_n, n \geq 0\}, \{D_n^{(r)}, n \geq 0, 1 \leq r \leq m\}$ are all independent. The X_i can be thought of as interarrivals which are generated from distribution F_r when $V_i = r$.

ASSUMPTION 4. The X_i form a sequence satisfying (4.5) with

$$ED_n^{(r)} < \infty, \;\; r = 1, \ldots, m,$$

and

$$\bar{F}_1(\cdot) \in RV_{-\alpha} \text{ and } \lim_{x \longrightarrow \infty} \frac{\bar{F}_j(x)}{\bar{F}_1(x)} = 0, \;\; j = 2, \ldots, m. \tag{4.6}$$

Under Assumption 4, we define $b(\cdot)$ by $\bar{F}_1(b(t)) = t^{-1}$.

We next state an assumption on the tail distribution $\bar{F}_\varepsilon$, which says that the measurement error ε has a lighter tail than X. This assumption is reasonable as measurement errors are thought to be small relative to the quantity being measured.

ASSUMPTION 5. The ε_i are i.i.d. random errors with a common tail distribution $\bar{F}_\varepsilon$, which has an asymptotic tail property

$$P(|\varepsilon| > x) = o(P(X > x)), \text{ as } x \to \infty.$$

The sequence $\{\varepsilon_i\}$ is independent of the sequence $\{X_i\}$ and of the approximating sequences $\{X_i^{(m)}\}$ in Assumption 2.

The order statistics used to compute the Hill estimator must be positive. In the following, all statements are tacitly assumed to hold conditional on the event $\{Y_{(k)} > 0\}$.

4.3 Main results

The underlying idea of our argument is that to get the consistency of the Hill estimator computed from error contaminated observations, it is enough to show that

$$\frac{n}{k} P\Big(\frac{Y_i}{b(n/k)} \in \cdot\Big) \xrightarrow{v} \nu \quad \text{and} \quad \frac{1}{k}\sum_{i=1}^{n} 1_{Y_i/b(n/k)} \Rightarrow \nu$$

in $M_+(0,\infty]$, where ν is the measure to which $\frac{1}{k}\sum_{i=1}^{n} 1_{X_i/b(n/k)}$ weakly converges. One can then obtain

$$\widehat{H}_{k,n} \xrightarrow{P} \int_1^\infty \log(u)\nu(du), \tag{4.1}$$

by Proposition 2.4 of [21]. If $\nu = \nu_\alpha$, defined in Section 4.2, (4.1) leads to

$$\widehat{H}_{k,n} \xrightarrow{P} \frac{1}{\alpha}. \tag{4.2}$$

We start with the i.i.d. case. We show that $Y = X + \varepsilon$ has regularly varying tail probabilities with the same index as $\bar{F}_X$, i.e. $\bar{F}_Y \in RV_{-\alpha}$. This approach allows us to conclude consistency for *any* estimator of α, provided it is consistent based on the X_i. For the *Hill* estimator, regular variation of the underlying tail distribution $\bar{F}_X$ is actually equivalent to the consistency of the estimator based on the Y_i. These results are presented respectively in parts (a) and (b) of Theorem 1, for which Proposition 1 is a preparation.

PROPOSITION 1. *Denote $Y = X + \varepsilon$, and let $\bar{F}_Y$ be the tail distribution of Y. Suppose that $P(X > \cdot) \in RV_{-\alpha}$, $P(|\varepsilon| > x) = o(P(X > x))$, and ε is independent of X. Then,*

$$\bar{F}_Y \in RV_{-\alpha}.$$

THEOREM 1. *(a) Under Assumptions 1 and 5, any estimator of α computed from the Y_i is consistent, if its counterpart computed from the unobservable X_i is consistent. (b) For the Hill estimator, under Assumption 5, the X_i satisfy Assumption 1 if and only if* (4.2) *holds.*

We now turn to dependent X_i that follow one of the assumptions specified in Section 4.2. The contaminated variables $X_i + \varepsilon_i$ need not satisfy these assumptions, and so a careful investigation is required.

We first consider the stationary process $\{X_i\}$ and its approximating m-dependent processes $\{X_i^{(m)}\}$ satisfying Assumption 2. Set

$$Y_i = X_i + \varepsilon_i, \quad Y_i^{(m)} = X_i^{(m)} + \varepsilon_i. \tag{4.3}$$

THEOREM 2. *If the unobservable sequences $\{X_i\}$ and $\{X_i^{(m)}\}$ satisfy Assumption 2 and if Assumption 5 holds, then the sequences $\{Y_i\}$ and $\{Y_i^{(m)}\}$ defined by* (4.3) *satisfy Assumption 2 as well.*

[55] provide three examples of processes satisfying Assumption 2.

(a) Infinite-order moving averages of heavy-tail innovations defined by

$$X_i = \sum_{j=0}^{\infty} c_j Z_{i-j}, \ -\infty < i < \infty,$$

where the Z_i are i.i.d. nonnegative random errors with a regularly varying tail distribution,

$$P(Z_i > \cdot) \in RV_{-\alpha}, \ \alpha > 0, \tag{4.4}$$

and the c_j contain at least one positive number, and satisfy $\sum_{j=0}^{\infty} |c_j|^{\delta} < \infty$, for some $0 < \delta < \alpha \wedge 1$. This model was recently studied by [65].

(b) A simple bilinear model driven by heavy-tail innovations defined by

$$X_i = cX_{i-1}Z_{i-1} + Z_i, \ \ -\infty < i < \infty,$$

where $c > 0$ and the Z_i are i.i.d. nonnegative random errors satisfying (4.4) and $c^{\alpha/2} E Z_1^{\alpha/2} < 1$.

(c) Solutions of stochastic equations of the form

$$X_i = A_i X_{i-1} + Z_i, \ \ -\infty < i < \infty,$$

where the Z_i are i.i.d. nonnegative random errors satisfying (4.4) and $\{(A_i, Z_i) \in \mathbb{R}_+^2, \ -\infty < i < \infty\}$ are i.i.d. random pairs with $EA_0^\alpha < 1$, $EA_0^\beta < \infty$, for some $0 < \alpha < \beta$.

By Corollary 3.1 of [55], processes (a), (b) and (c) satisfy Assumption 2, and for process (b), α is replaced by $\alpha/2$. We thus obtain the following corollary to Theorem 2.

COROLLARY 1. *Convergence* (4.1) *holds under Assumptions 2 and 5, i.e.* (4.2) *holds for the processes (a) or (c), and for the process (b),* $\widehat{H}_{k,n} \xrightarrow{P} 2/\alpha$.

We next assume that the unobservable stationary process $\{X_i\}$ satisfies Assumption 3. In this case, it cannot be claimed that the contaminated process also satisfies Assumption 3. For example, if the X_i follow an ARCH model, then $X_i + \varepsilon_i$ will not follow this model.

THEOREM 3. *Relations* (4.1) *and* (4.2) *hold under Assumptions 3 and 5.*

The ARCH process introduced by [58] is defined by

$$X_i = N_i(\beta + \lambda X_{i-1}^2)^{1/2}, \ -\infty < i < \infty, \tag{4.5}$$

where the N_i are i.i.d. $N(0,1)$ random variables, $\beta > 0$, and $\lambda > 0$. We assume $0 < \lambda < 2e^E$, where $E = 0.5772...$ is Euler's constant, to guarantee the existence of α stated in Assumption 3, see Lemma 8.4.6 of [16]. The process $\{X_i^2\}$ therefore satisfies Assumption 3 with $A_i = \lambda N_i^2$ and $B_i = \beta N_i^2$. We then obtain the following corollary to Theorem 3.

COROLLARY 2. *Relation* (4.2) *holds for the ARCH(1) process, under Assumption 5, provided* $\beta > 0$ *and* $0 < \lambda < 2e^E$.

We finally study the consistency of the Hill estimator for interarrival times generated by a heavy-tailed hidden Markov model, and which are observed with measurement errors. We assume that the process $\{X_i\}$ satisfies Assumption 4, under which [55] show that $H_{k,n} \xrightarrow{P} 1/\alpha$. We consider $\widehat{H}_{k,n}$ based on $Y_i = X_i + \varepsilon_i$.

THEOREM 4. *Convergence* (4.2) *holds under Assumptions 4 and 5.*

4.4 Impact of measurement errors in finite samples

In this section, we report the results of simulation studies of the Hill estimator applied to various processes contaminated by additive errors. We investigate the impact of these errors, especially how large they can be compared to be tolerated in practice.

We generate observations $Y_i = X_i + \varepsilon_i,\ i = 1, 2, \ldots, n$, where $\{X_i\}$ and $\{\varepsilon_i\}$ are independent sets of random variables. We use four models for the X_i, those considered in Section 4.2.

Model 1 The X_i are **i.i.d.** random variables, which follow a Pareto distribution with $\alpha = 2$, $P(X_i > x) = x^{-2},\ x > 1$.

Model 2 The X_i form the **AR(2)** process $X_i = 1.3X_{i-1} - 0.7X_{i-2} + Z_i$, where the Z_i follow a Pareto distribution with $\alpha = 2$, $P(Z_i > z) = z^{-2},\ z > 1$.

Model 3 The X_i form the simple **bilinear** model driven by heavy-tail innovations defined by $X_i = 0.7X_{i-1}Z_{i-1} + Z_i$, where the Z_i follow a Pareto distribution with $\alpha = 4$, $P(Z_i > z) = z^{-4},\ z > 1$.

Model 4 The X_i follow the **ARCH** process $X_i = N_i(1 + 0.5773X_{i-1}^2)^{1/2}$, where the N_i are i.i.d. $N(0, 1)$ random variables.

Model 2 is causal and thus has an infinite moving average representation, which satisfies Assumption 2. Each X_i therefore has tail index $\alpha = 2$. Model 3 also satisfies Assumption 2, and each X_i has tail index $\alpha/2 = 2$. Model 4 with $\beta = 1$, $\lambda = 0.5773$ satisfies Assumption 3, and X_i^2 has tail index α which satisfies $E(0.5773N_0^2)^\alpha = 1$. We get a numerical solution for the equation, $\alpha \approx 2$, since the equation cannot be solved explicitly. Thus, in all four models, the true value of the tail index of the X_i (X_i^2 for Model 4) is 2.

The ε_i are drawn from a normal distribution with mean 0 and standard deviation σ, a scaled t-distribution with 4 degrees of freedom (scaled t_4), and a generalized Pareto distribution (GPD), $P(|\varepsilon| > z) = (1 + \xi(z - \mu)/\sigma)^{-1/\xi}$, with location $\mu = 0$, shape $\xi = 1/4$, and scale σ_{GPD}. The

Table 4.1: Empirical bias and standard error of $\hat{\alpha}$ of the Hill estimator applied to various models with additive errors following $N(0, \sigma^2)$, t_4, or GPD, with **fixed error SD**.

		No error	$N(0,\sigma^2)$					t_4	GPD
Error SD		0	0.1	0.2	0.3	0.4	0.5	1.41	0.47
Model 1	bias	0.02	0.05	0.05	0.07	0.08	0.11	0.31	0.07
SD = 2.88	(SE)	(0.14)	(0.11)	(0.11)	(0.11)	(0.11)	(0.12)	(0.17)	(0.24)
Model 2	bias	0.42	0.42	0.42	0.43	0.43	0.43	0.49	0.43
SD = 6.24	(SE)	(0.33)	(0.33)	(0.33)	(0.33)	(0.33)	(0.33)	(0.34)	(0.34)
Model 3	bias	0.23	0.23	0.23	0.22	0.23	0.23	0.23	0.23
SD = 31.5	(SE)	(0.54)	(0.54)	(0.54)	(0.53)	(0.54)	(0.53)	(0.53)	(0.54)
Model 4	bias	-0.23	-0.22	-0.22	-0.21	-0.21	-0.21	-0.18	-0.21
SD = 7.00	(SE)	(0.21)	(0.21)	(0.21)	(0.21)	(0.22)	(0.22)	(0.22)	(0.22)

scale parameters for each error distribution vary. They can be fixed or determined by the ratio of the standard deviation of error distribution (error SD) to the standard deviation of underlying process (model SD). For example, if we consider the ratio of 0.1 for Model 1 whose standard deviation is 2.88, then the corresponding scale parameter is 0.288 for the normal distribution, 0.204 for the scaled t_4, and 0.125 for the GPD. All distributions of the measurement error have a lighter tail than the X_i (X_i^2 for Model 4), so the tail distributions satisfy Assumption 5.

We estimate the tail index using the Hill estimator with a data–driven cut-off k, the number of upper order statistics used to compute it. We use the threshold selection method introduced by [66], which employs a bootstrap procedure to choose k that minimizes the asymptotic mean square error (AMSE). This procedure is implemented by the function `hall` of the `R` package `tea`. For each model/error pair, we compute the average of the estimates over 1000 replications, and the estimated standard error based on these replications. The sample size is $n = 5,000$.

Table 4.1 reports the results for fixed error SDs, Table 4.2 for fixed ratios of error SD to model SD. The error SD of 0, or the ratio 0 means that there are no errors. Model SD is calculated from the generated X_i. Table 4.3 provides information on the effects of errors on the selection of optimal k.

Tables 4.1 and 4.2 show that additive errors lead to estimates which indicate lighter tails than those indicated by the estimates computed from uncontaminated data. This can be intuitively

Table 4.2: Empirical bias and standard error of $\hat{\alpha}$ for the **fixed ratio** of the error SD to model SD.

Model	Error Type	Error SD/Model SD Ratio							
		0	0.005	0.01	0.02	0.04	0.06	0.1	0.2
Model 1 SD = 2.88	Normal	0.02 (0.14)	0.04 (0.13)	0.05 (0.11)	0.05 (0.11)	0.05 (0.11)	0.05 (0.11)	0.06 (0.11)	0.12 (0.12)
	scaled t_4	0.02 (0.14)	0.04 (0.12)	0.05 (0.11)	0.05 (0.11)	0.05 (0.11)	0.04 (0.12)	0.05 (0.12)	0.09 (0.13)
	GPD	0.02 (0.14)	0.05 (0.12)	0.05 (0.11)	0.05 (0.11)	0.04 (0.12)	0.04 (0.12)	0.04 (0.13)	0.07 (0.14)
Model 2 SD = 6.24	Normal	0.42 (0.33)	0.42 (0.33)	0.42 (0.33)	0.42 (0.33)	0.42 (0.33)	0.43 (0.33)	0.43 (0.33)	0.47 (0.33)
	scaled t_4	0.42 (0.33)	0.42 (0.33)	0.42 (0.33)	0.42 (0.33)	0.43 (0.33)	0.43 (0.33)	0.43 (0.33)	0.47 (0.33)
	GPD	0.42 (0.33)	0.42 (0.33)	0.42 (0.33)	0.42 (0.33)	0.43 (0.34)	0.42 (0.33)	0.44 (0.34)	0.47 (0.35)
Model 3 SD = 31.5	Normal	0.23 (0.54)	0.23 (0.54)	0.23 (0.54)	0.23 (0.54)	0.22 (0.53)	0.23 (0.53)	0.24 (0.53)	0.26 (0.53)
	scaled t_4	0.23 (0.54)	0.23 (0.54)	0.22 (0.54)	0.23 (0.54)	0.23 (0.54)	0.23 (0.53)	0.23 (0.53)	0.26 (0.54)
	GPD	0.23 (0.54)	0.23 (0.54)	0.23 (0.54)	0.23 (0.54)	0.23 (0.54)	0.23 (0.53)	0.23 (0.53)	0.27 (0.55)
Model 4 SD = 7.00	Normal	−0.23 (0.21)	−0.22 (0.21)	−0.22 (0.21)	−0.22 (0.21)	−0.21 (0.21)	−0.21 (0.22)	−0.20 (0.22)	−0.18 (0.23)
	scaled t_4	−0.23 (0.21)	−0.22 (0.21)	−0.22 (0.21)	−0.22 (0.21)	−0.22 (0.21)	−0.21 (0.22)	−0.20 (0.22)	−0.18 (0.23)
	GPD	−0.23 (0.21)	−0.22 (0.21)	−0.22 (0.21)	−0.22 (0.21)	−0.22 (0.21)	−0.21 (0.22)	−0.21 (0.22)	−0.17 (0.23)

expected because the errors in a sense “dilute" the true heavy tails. For Models 1–3, the biases increase with the error SD. For Model 4, the bias becomes smaller in absolute value, but this cannot be interpreted that the error helps the bias to be small. Instead, this behavior is in agreement with the previous observation; the bias for uncontaminated process is negative, and it becomes less negative (lighter tail) as the error SD increases.

A rather unexpected finding is that the bias is not affected a lot even by large errors. We see from Table 4.1 that for iid X_i, Model 1, even error SD equal to half the model SD causes bias of 0.31, which is not large given the uncertainty about the selection of k. Such a level of contamination could however indicate that the data have finite variance, whereas in fact they may have infinite variance. Even more remarkable is that for dependent data with heavy–tailed marginal distributions, the errors have almost no impact on the bias and the SE of the estimator. Table 4.2 is designed to take a closer look at this finding by controlling the ratio of error SD to model SD. We first observe that the bias increases with this ratio. Second, this increase is relatively flat. Only for iid X_i, the ratio of 20 percent causes a bump in bias. For dependent X_i, such a ratio does not change the bias much compared to uncontaminated data; Models 2, 3 and 4 are surprisingly insensitive to the errors. Finally, the bias caused by the errors does not depend on the type of the error distribution. Standard errors of the estimates are basically unaffected by the errors. In some cases, the errors lead to smaller or larger estimated standard errors, but these estimates are so close that the differences are probably not statistically significant.

Table 4.3 reports on average optimal ks and their standard errors for all combinations of the underlying models and error distributions. We do not observe any clear positive or negative relationships between the average optimal k and the ratio, but the dependent data still show a considerably weaker dependence on the ratio; again, only a large increase of the ratio has a relatively large impact on the average optimal k.

Another question of interest is how the errors affect the shape of the Hill plot. The Hill estimator is *location variant*, see Section 4.2.2 of [4]. The lack of location invariance makes it sensitive to a shift in location; but it does not theoretically affect the tail index estimate. Since adding measurement errors could be thought of as a location shift, there could be sensitivity to additive errors over some range of k, which will not show up when examining the averages. The Hill plot is a useful tool to examine this property by describing estimates as a function of the minimal order statistic k used to compute the estimates.

Figure 4.1 shows the Hill plots for observations generated by Model 1 with the ratio of 0.1, along with a vertical line showing the optimal k. The true tail index α is 2 for all the plots. The impact due to the additive errors, for the ratio of 0.1, turns out to be surprisingly weak as all the plots in Figure 4.1 look stable. We also consider the Hill plots for Model 1 observed with the same types of errors, but with a relatively large error SD corresponding to the ratio 0.2. These plots are

Table 4.3: Average optimal k (standard error) for the Hill estimator as the function of the ratio of the error SD to model SD. The sample size is 5000, and the number of replications is 1000.

Model	Error Type	Error SD/Model SD Ratio							
		0	0.005	0.01	0.02	0.04	0.06	0.1	0.2
Model 1 SD = 2.88	Normal	2114 (1622)	2933 (1727)	3066 (1625)	2973 (1552)	2809 (1378)	2690 (1201)	2514 (903)	2141 (537)
	scaled t_4	2114 (1622)	2990 (1702)	3061 (1633)	2959 (1558)	2777 (1429)	2635 (1291)	2354 (1080)	1951 (735)
	GPD	2114 (1622)	3022 (1690)	3053 (1634)	2928 (1585)	2751 (1483)	2587 (1380)	2295 (1208)	1700 (892)
Model 2 SD = 6.24	Normal	1249 (827)	1250 (827)	1251 (824)	1261 (815)	1264 (811)	1280 (789)	1307 (697)	1293 (588)
	scaled t_4	1249 (827)	1250 (826)	1254 (822)	1250 (820)	1255 (812)	1268 (787)	1225 (711)	1140 (632)
	GPD	1249 (827)	1252 (826)	1252 (826)	1248 (815)	1261 (808)	1246 (794)	1161 (725)	1041 (649)
Model 3 SD = 31.5	Normal	638 (907)	639 (907)	639 (906)	651 (908)	654 (906)	668 (901)	749 (881)	821 (827)
	scaled t_4	638 (907)	644 (907)	642 (907)	649 (903)	652 (904)	667 (892)	703 (852)	709 (789)
	GPD	638 (907)	639 (906)	641 (906)	646 (906)	647 (902)	657 (886)	654 (831)	621 (756)
Model 4 SD = 7.00	Normal	246 (134)	242 (136)	243 (140)	241 (141)	243 (146)	248 (156)	314 (249)	510 (376)
	scaled t_4	246 (134)	243 (136)	242 (137)	243 (141)	244 (147)	251 (157)	300 (211)	377 (267)
	GPD	246 (134)	243 (136)	244 (139)	244 (141)	246 (145)	253 (157)	281 (184)	311 (204)

shown in Figure 4.2. The shape of Hill plot is more affected by the larger error SD; this sensitivity is especially noticeable for errors with the normal and scaled t_4 distributions. The Hill plots for Models 2, 3, and 4, even without errors, do not look as stable as for iid observations, there is an upward trend. The presence of the errors changes their shape a little bit for the ratio of 0.2, but one would not say that these errors make the plots any worse.

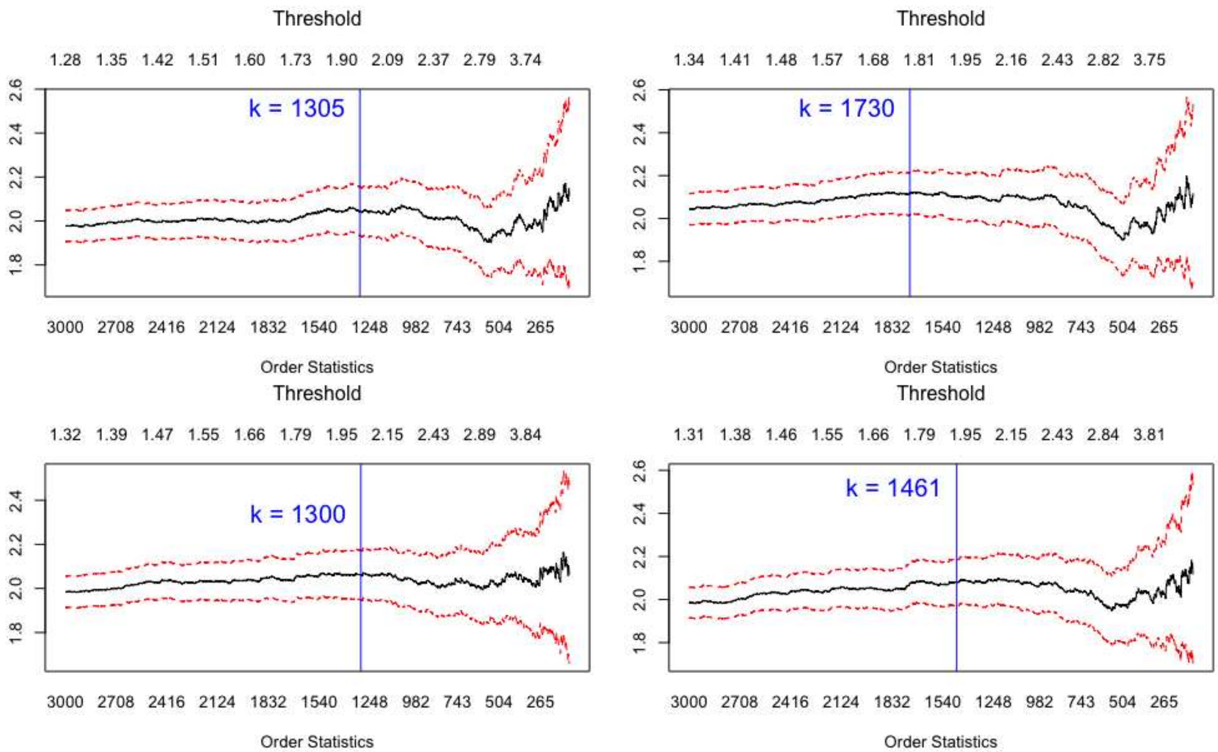

Figure 4.1: Hill plots with a vertical line showing the optimal k for Model 1 (a single realization) observed with no measurement errors (top left), with errors following the normal (top right), scaled t_4 (bottom left), and GPD (bottom right) with the **ratio of 0.1**.

4.5 Application to internet traffic anomalies

In this section, we illustrate the relevance of studying the Hill estimator for error contaminated data by an application to the interarrival times of internet traffic anomalies. We first provide some background, limited in scope to conserve space, and focus on the aspects relevant to this paper. More detailed network background is presented in [67], a paper which to some extent motivates the present research. We hope that that the analysis presented below may guide other applications where the tail index must be estimated from error contaminated data.

Figure 4.3 shows the backbone internet network in the United States known as Internet2. A traffic disruption in any of the links can slow down service in the whole country. For this reason, anomalies in the internet traffic have been extensively studied. An anomaly is a time and space confined traffic whose volume is much higher than typical. An anomaly can result from a malfunction of network resources, like routers, or from malicious activity, like denial of service attacks. [67]

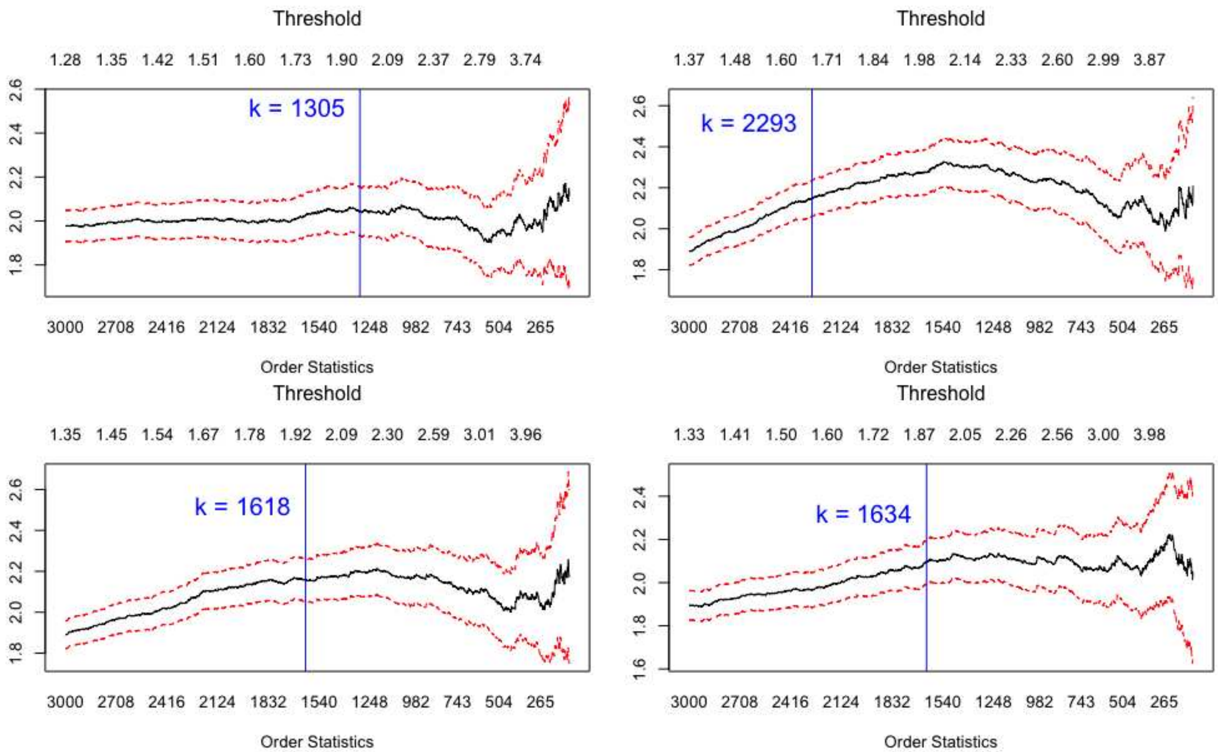

Figure 4.2: Hill plots with a vertical line showing the optimal k for Model 1 (a single realization) observed with no measurement errors (top left), with errors following the normal (top right), scaled t_4 (bottom left), and GPD (bottom right) with the **ratio of 0.2**.

developed a simple algorithm, based on the Fourier transform, which, among other characteristics, allowed them to identify the arrival time of an anomaly in any unidirectional link. They created a database of anomalies and their characteristics for 28 unidirectional links, corresponding to the 14 two–directional links shown in Figure 4.3, for the time period of 50 weeks, starting October 16, 2005. Due to a huge amount of data to be processed, the algorithm computes an anomaly arrival time only with the precision of five minutes. There is therefore uncertainty as to when the anomaly actually arrived, a rounding error. A key element in the analysis of anomalous traffic is to understand the distribution of the interarrival times, the time separation between the arrivals of two consecutive anomalies. This may be helpful in provisioning network resources. [67] perform a preliminary fitting, based on the exponential distribution. We take a closer look at this problem and place it in the context of this paper.

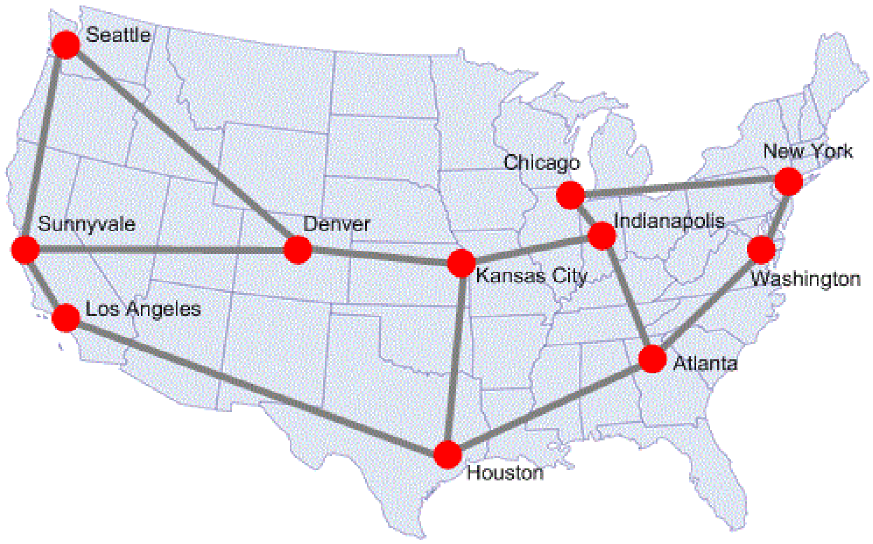

Figure 4.3: A map showing 14 two-directional links of the Internet2 network.

We index the unidirectional links by integers from 1 to 28, it is not important for our analysis to which nodes they correspond. The count of anomalies detected by the algorithm of [67] varies from link to link, as shown in Table 4.4. We have examined the Hill plots and performed other diagnostic tests, and determined that it is reasonable to assume that for each link the distribution of the interarrival times is regularly varying with the tail index between 1 and 3. The values computed using the Hill estimator with the optimal k introduced in Section 4.4 are shown in Table 4.4, in the rows α_{obs}. Exploratory data analysis in [68] strongly suggests that the interarrival times form an iid sequence, so the setting of Theorem 1 holds.

We illustrate our analysis using the interarrival times in link 5, corresponding to anomalies traveling from Chicago to New York. In Figure 4.4 we display the Hill plot and the QQ plot of the log transformed data matched against exponential quantiles beyond the exceedance threshold corresponding to the optimal k. We should get approximately a line whose slope is $1/\alpha_{obs}$ if our data had a Pareto tail with index α_{obs}, see Section 4.6.4 of [4]. The QQ plot looks linear with the fit of a straight line whose slope is $1/1.53$, which tells us that it is reasonable to assume a Pareto tail with index 1.53. The same conclusion can be drawn for other links. The smallest value of α_{obs} is 1.27. It corresponds to anomalies traveling from Los Angeles to Sun Valley. The largest is 2.22, from the Indianapolis to Atlanta link.

Table 4.4: Results of a simulation study based on anomalous internet traffic. The tail index α_{obs} is computed from the interarrival times produced by the algorithm. The average $\bar{\alpha}$ is computed from 1,000 replications of the interarrival times with errors, σ_a is the standard deviation of the 1,000 estimates.

Link	1	2	3	4	5	6	7
sample size	405	247	362	454	347	345	603
α_{obs}	1.69	1.50	1.62	1.62	1.53	1.59	1.68
$\bar{\alpha}$	1.72	1.49	1.63	1.63	1.58	1.68	1.65
σ_a	0.03	0.05	0.02	0.02	0.05	0.11	0.02
Link	8	9	10	11	12	13	14
sample size	300	387	345	382	304	476	507
α_{obs}	1.56	1.47	1.44	1.79	2.22	2.11	1.93
$\bar{\alpha}$	1.51	1.50	1.50	1.80	2.24	2.12	1.90
σ_a	0.03	0.03	0.05	0.03	0.03	0.03	0.05
Link	15	16	17	18	19	20	21
sample size	478	319	402	388	433	493	340
α_{obs}	2.07	1.48	1.91	1.35	1.27	1.97	1.97
$\bar{\alpha}$	2.00	1.45	1.91	1.36	1.29	1.96	2.00
σ_a	0.05	0.04	0.02	0.01	0.02	0.04	0.03
Link	22	23	24	25	26	27	28
sample size	417	597	296	258	340	348	264
α_{obs}	1.46	1.65	1.43	1.83	1.43	1.95	1.69
$\bar{\alpha}$	1.46	1.65	1.51	1.80	1.43	1.90	1.58
σ_a	0.01	0.03	0.06	0.03	0.04	0.05	0.07

In the context of this paper, each interarrival time Y_i, computed by the algorithm, is treated as a “true" interarrival time X_i measured with a roundoff error, i.e. $Y_i = X_i + \varepsilon_i$. The unobserved X_i is not rigorously defined, but we can think of it as the time separation based on a more precise algorithm, or just a different algorithm. In the latter case, the analysis that follows provides information about the uncertainty in the estimation of α caused by the choice of a specific algorithm. Since the smallest value of Y_i in physical units is 5 min., we use 5 minutes as a unit lag. We therefore assume that the errors ε_i are uniformly distributed on $[-1, 1]$. We experimented with other beta distributions on $[-1, 1]$, the results were basically unaffected.

We perform the following numerical experiment. For each link, we generate $R = 1,000$ samples of unobservable interarrival times $X_i^{(r)}$ from the observations Y_i, i.e. $X_i^{(r)} = Y_i - \varepsilon_i^{(r)}$, where the $\varepsilon_i^{(r)}$ are drawn from the uniform distribution on $[-1, 1]$, for $r = 1, \ldots, R$. We get estimates $\hat{\alpha}_r$ for each sample and then compute the average of the estimates, $\bar{\alpha}$, and their estimated

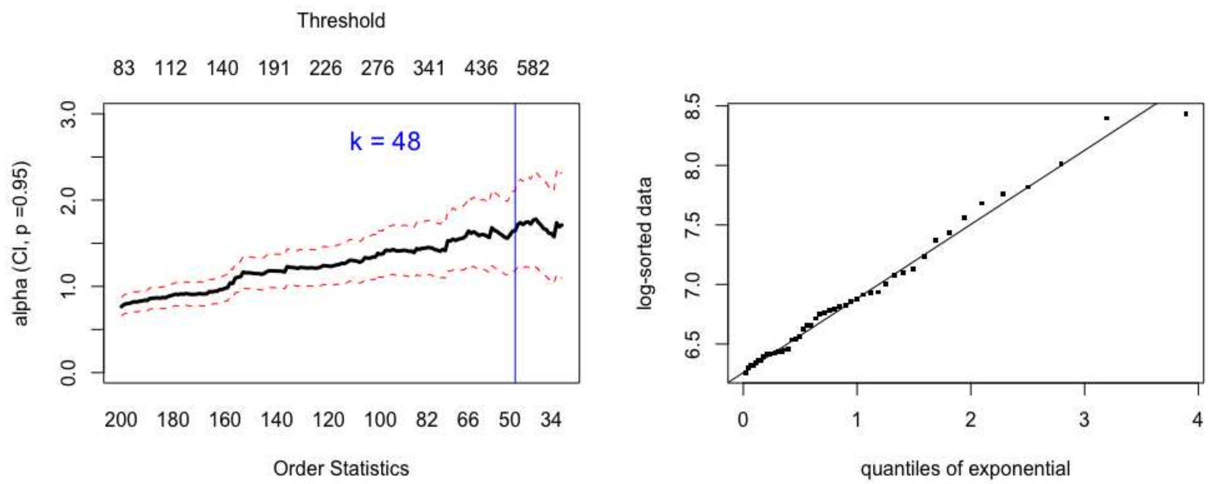

Figure 4.4: Hill plot (left) and QQ plot (right) for link 5.

standard error, σ_a, i.e.,

$$\bar{\alpha} = \frac{1}{R}\sum_{r=1}^{R}\hat{\alpha}_r, \quad \sigma_a = \left\{\frac{1}{R}\sum_{r=1}^{R}(\hat{\alpha}_r - \bar{\alpha})^2\right\}^{1/2}.$$

The results in Table 4.4 show that $\bar{\alpha}$ is close to α_{obs} for most links. For each link, the ratio of the error SD to the observations SD is less than 0.001, so one might expect such an outcome based on the simulations in Section 4.4, but the sample sizes for these data are much smaller than 5,000, so the result was not clear a priori. We find a couple links, 6 and 28, which have a relatively large discrepancy between α_{obs} and $\bar{\alpha}$, with a high value of σ_a. All discrepancies are however within $2\sigma_a$, so these differences are not significant. Overall, our numerical experiment shows that for the purpose of the estimation of the tail index of the anomalies interarrival times, an algorithm that identifies arrivals of anomalies with 5 min. resolution is sufficient.

4.6 Preliminary results

We list in this section several lemmas which are used in Section 4.7. The first lemma follows from the definition of regular variation. The second lemma states three equivalent conditions for regularly varying functions. It follows from Theorem 3.6 in [4].

LEMMA 1. *Suppose that $U \in RV_{-\alpha}$ and $V(x) \sim cU(x)$ for some two functions $U, V : \mathbb{R}_+ \to \mathbb{R}_+$, and $0 < c < \infty$. Then, $V \in RV_{-\alpha}$.*

LEMMA 2. *Suppose X is a nonnegative random variable with its complementary distribution function $\bar{F}$. The following are equivalent:*

(i) $P(X > \cdot) \in RV_{-\alpha}$, $\alpha > 0$.

(ii) There exists a sequence $a(n)$ with $a(n) \to \infty$ such that $\lim_{n\to\infty} nP(X > a(n)x) = x^{-\alpha},\ x > 0$.

(iii) There exists a sequence $a(n)$ with $a(n) \to \infty$ such that $nP(X/a(n) \in \cdot) \xrightarrow{v} \nu_\alpha(\cdot)$,

in $M_+(0, \infty]$, where $\nu_\alpha(x, \infty] = x^{-\alpha}$.

The sequence $a(n)$ is the same in (ii) and (iii).

We next summarize what has been established for sums of regularly varying functions.

LEMMA 3. *(i) Let X and Y be two independent non-negative random variables with their corresponding complementary distributions $P(X > \cdot) \in RV_{-\alpha_1}$ and $P(Y > \cdot) \in RV_{-\alpha_2}$, for some $\alpha_1, \alpha_2 > 0$. Then $P(X > \cdot) + P(Y > \cdot) \in RV_{\max(-\alpha_1, -\alpha_2)}$.*

(ii) Under the assumptions of part (i), $P(X + Y > x) \sim P(X > x) + P(Y > x)$, and $P(X + Y > \cdot) \in RV_{\max(-\alpha_1, -\alpha_2)}$.

(iii) Let X and Y be two independent non-negative random variables, $P(X > \cdot) \in RV_{-\alpha}$ for $\alpha > 0$ and $P(Y > x) = o(P(X > x))$. Then $P(X + Y > x) \sim P(X > x)$, and $P(X + Y > \cdot) \in RV_{-\alpha}$.

Proof. Statement (i) is proven as Proposition 1.5.7 (iii) of [69]. Statement (ii) follows from Lemma 3 (i), Lemma 1, and calculations on p. 278 of [70], which establish a convolution property of finite sums of regularly varying variables. Its proof is also found in Theorem 1.1 of [71]. Statement (iii) is proven as Theorem 2.1 of [71]. □

The following lemma follows from Proposition 2.2 of [55].

LEMMA 4. *Under Assumption 2,*

$$\frac{1}{k}\sum_{i=1}^{k-1}\log\frac{X_{(i)}}{X_{(k)}} \xrightarrow{P} \int_1^\infty \log(u)\nu(du).$$

LEMMA 5. *Under Assumption 3,*

(i) $P(X_i > \cdot) \in RV_{-\alpha}$.

(ii) There exist γ *and* c_0 *such that* $0 < \gamma < \alpha$, $0 < c_0 < 1$, $EA_0^\gamma = c_0$.

Proof. Statement (i) is shown in [72], [73] and [74]. Statement (ii) is shown on p. 220 of [75]. □

The following lemma follows from Propositions 2.3 of [55].

LEMMA 6. *Suppose a stationary sequence* X_i *satisfies the following conditions:*

$$\frac{n}{k}P\left(\frac{X_i}{(n/k)^{1/\alpha}} \in \cdot\right) \xrightarrow{v} \nu_\alpha, \tag{4.1}$$

where $\nu_\alpha(x,\infty] = x^{-\alpha}$ *in* $M_+(0,\infty]$. *For any* $x > 0, y > 0$,

$$\lim_{n\to\infty}\frac{n}{k^2}\sum_{j=2}^{k}P\left(\frac{X_1}{(n/k)^{1/\alpha}} > x, \frac{X_j}{(n/k)^{1/\alpha}} > y\right) = 0. \tag{4.2}$$

For any sequence l_n *such that* $l_n \to \infty$, $l_n/k \to 0$, $n/k = o(l_n)$, *and intervals*

$$I_1 = [1, k-l_n], I_2 = [k+1, 2k-l_n], \ldots, I_{[n/k]} = [([n/k]-1)k+1, [n/k]k - l_n],$$

$$\lim_{n\to\infty}\left|E\prod_{j=1}^{[n/k]}\left(1-\frac{1}{k}\sum_{i\in I_j}f\left(\frac{X_i}{(n/k)^{1/\alpha}}\right)\right) - \prod_{j=1}^{[n/k]}E\left(1-\frac{1}{k}\sum_{i\in I_j}f\left(\frac{X_i}{(n/k)^{1/\alpha}}\right)\right)\right| = 0, \tag{4.3}$$

where f *is any function of the form* $f = \sum_{h=1}^{s}\beta_h 1_{(x_h,\infty]}$, *for* $\beta_h > 0$, $h = 1,\ldots,s$, *and* $x_h > 0$, $h = 1,\ldots,s$. *Then,*

$$H_{k,n} = \frac{1}{k}\sum_{i=1}^{k-1}\log\frac{X_{(i)}}{X_{(k)}} \xrightarrow{P} \frac{1}{\alpha}. \tag{4.4}$$

4.7 Proofs of the results of Section 4.3

Proof of Proposition 1 : The result almost follows from Lemma 3 (iii). We must take care of the absolute value.

For $y > 0$, $\{X + \varepsilon > y\} \subset \{X + |\varepsilon| > y\}$. Therefore, $P(X + \varepsilon > y) \leq P(X + |\varepsilon| > y)$. By Lemma 3 (iii),

$$\limsup_{y \to \infty} \frac{P(X + \varepsilon > y)}{P(X > y)} = \limsup_{y \to \infty} \frac{P(X + \varepsilon > y)}{P(X + |\varepsilon| > y)} \leq 1. \tag{4.1}$$

Now, taking any $\delta > 0$, we obtain $\{X + \varepsilon > y\} \supset \{X > (1 + \delta)y, \varepsilon > -\delta y\}$. Thus,

$$P(X + \varepsilon > y) \geq P(X > (1 + \delta)y)P(\varepsilon > -\delta y),$$

by the independence of X and ε. Since $P(X > \cdot) \in RV_{-\alpha}$ and $\delta > 0$ is arbitrary,

$$\liminf_{y \to \infty} \frac{P(X + \varepsilon > y)}{P(X > y)} \geq 1. \tag{4.2}$$

Combining (4.1) and (4.2), we obtain $\lim_{y \to \infty} P(X + \varepsilon > y)/P(X > y) = 1$, and $P(X+\varepsilon > \cdot) \in RV_{-\alpha}$ by Lemma 1.

Proof of Theorem 1 : Since the Y_i are i.i.d random variables with a common tail distribution $\bar{F}_Y \in RV_{-\alpha}$, by Proposition 1, the consistency follows, e.g., from Theorems 4.1 and 4.2 in [4]. By Theorem 2 of [19] the consistency of the Hill estimator computed from the Y_i implies $\bar{F}_Y \in RV_{-\alpha}$, and by Proposition 1 $\bar{F}_X \in RV_{-\alpha}$.

Proof of Theorem 2 : First, $\{Y_i\}$ and $\{Y_i^{(m)}\}$ are stationary because the sequence $\{\varepsilon_i\}$ is independent of the sequence $\{X_i\}$ and of the approximating sequences $\{X_i^{(m)}\}$. Also, for any $\tau > 0$,

$$\lim_{m \to \infty} \limsup_{n \to \infty} \frac{n}{k} P\left(\left|\frac{Y_i^{(m)}}{b(n/k)} - \frac{Y_i}{b(n/k)}\right| > \tau\right) = \lim_{m \to \infty} \limsup_{n \to \infty} \frac{n}{k} P\left(\left|\frac{X_i^{(m)}}{b(n/k)} - \frac{X_i}{b(n/k)}\right| > \tau\right) = 0.$$

So it remains to show that for each $m \geq 1$,

$$\frac{n}{k} P\Big(\frac{Y_i^{(m)}}{b(n/k)} \in \cdot\Big) \xrightarrow{v} \nu^{(m)}, \text{ as } (4.3). \tag{4.3}$$

In fact, it suffices to find M such that (4.3) holds for each $m \geq M$, since $m \to \infty$ in Proposition 2.2 of [55].

We first show that the following is true. There exists M such that for each $m \geq M$,

$$\frac{n}{k}P\Big(\frac{Y_i^{(m)}}{b(n/k)} > y\Big) = \frac{n}{k}P\Big(\frac{X_i^{(m)} + \varepsilon_i}{b(n/k)} > y\Big) \to \nu^{(m)}(y, \infty], \text{ as (4.3).}$$

Take any $\delta > 0$. For $y > 0$, $\{X_i^{(m)} + \varepsilon_i > yb(n/k)\} \supset \{X_i^{(m)} > (1+\delta)yb(n/k),\ \varepsilon_i > -\delta yb(n/k)\}$. Therefore, $\frac{n}{k}P(X_i^{(m)} + \varepsilon_i > yb(n/k)) \geq \frac{n}{k}P(X_i^{(m)} > (1+\delta)yb(n/k))P(\varepsilon_i > -\delta yb(n/k))$, by the independence of X_i and ε_i. By Assumption 2(a), since $b(n/k) \to \infty$,

$$\liminf_{n \to \infty} \frac{n}{k}P(X_i^{(m)} + \varepsilon_i > yb(n/k)) \geq \nu^{(m)}((1+\delta)y, \infty].$$

Therefore, by the right–continuity of $y \mapsto \nu^{(m)}(y, \infty]$,

$$\liminf_{n \to \infty} \frac{n}{k}P(X_i^{(m)} + \varepsilon_i > yb(n/k)) \geq \nu^{(m)}(y, \infty]. \tag{4.4}$$

Next observe that

$$\begin{aligned}\{X_i^{(m)} + \varepsilon_i > yb(n/k)\} &\subset \{X_i^{(m)} > (1-\delta)yb(n/k)\} \\ &\cup \{|\varepsilon_i| > (1-\delta)yb(n/k)\} \cup \{X_i^{(m)} > \delta yb(n/k),\ |\varepsilon_i| > \delta yb(n/k)\}.\end{aligned}$$

By the independence of X_i and ε_i,

$$\begin{aligned}\frac{n}{k}P(X_i^{(m)} + \varepsilon_i > yb(n/k)) &\leq \frac{n}{k}P(X_i^{(m)} > (1-\delta)yb(n/k)) \\ &+ \frac{n}{k}P(|\varepsilon_i| > (1-\delta)yb(n/k)) \\ &+ \frac{n}{k}P(X_i^{(m)} > \delta yb(n/k))P(|\varepsilon_i| > \delta yb(n/k)) \\ &=: Q_1(n,m) + Q_2(n) + Q_3(n,m).\end{aligned}$$

Then, by Assumption 2(a), for each $m \geq 1$

$$\limsup_{n\to\infty} Q_1(n,m) = \nu^{(m)}((1-\delta)y,\infty], \ \ \limsup_{n\to\infty} Q_3(n,m) = \nu^{(m)}(\delta y,\infty] \times 0 = 0.$$

Observe that

$$\limsup_{n\to\infty} Q_2(n) = \limsup_{n\to\infty} \frac{n}{k} P(X_i > (1-\delta)yb(n/k)) \frac{P(|\varepsilon_i| > (1-\delta)yb(n/k))}{P(X_i > (1-\delta)yb(n/k))}.$$

Since by Assumption 5,

$$\limsup_{n\to\infty} \frac{P(|\varepsilon_i| > (1-\delta)yb(n/k))}{P(X_i > (1-\delta)yb(n/k))} = 0,$$

we must verify that

$$\limsup_{n\to\infty} \frac{n}{k} P(X_i > (1-\delta)yb(n/k)) < \infty. \tag{4.5}$$

Since for $0 < \eta < (1-y)\delta$,

$$\{X_i > (1-\delta)yb(n/k)\} \subset \{X_i - X_i^{(m)} > \eta b(n/k)\} \cup \{X_i^{(m)} > \{(1-\delta)y - \eta\} b(n/k)\},$$

we have

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{n}{k} P(X_i > (1-\delta)yb(n/k)) \\
&\leq \limsup_{n\to\infty} \frac{n}{k} P(X_i - X_i^{(m)} > \eta b(n/k)) + \limsup_{n\to\infty} \frac{n}{k} P(X_i^{(m)} > \{(1-\delta)y-\eta\} b(n/k)).
\end{aligned}$$

By Assumption 2 (b), for any $0 < \gamma < 1$, there exists M such that for $m \geq M$,

$$\limsup_{n\to\infty} \frac{n}{k} P(X_i - X_i^{(m)} > \eta b(n/k)) < \gamma.$$

By Assumption 2 (a),

$$\limsup_{n\to\infty} \frac{n}{k} P(X_i^{(m)} > \{(1-\delta)y-\eta\} b(n/k)) = \nu^{(m)}((1-\delta)y - \eta, \infty].$$

We therefore conclude that for $m \geq M$,

$$\limsup_{n\to\infty} \frac{n}{k} P(X_i > (1-\delta)yb(n/k)) \leq \gamma + \nu^{(m)}((1-\delta)y - \eta, \infty] < \infty,$$

concluding the verification of (4.5) and leading to

$$\limsup_{n\to\infty} \frac{n}{k} P(X_i^{(m)} + \varepsilon_i > yb(n/k)) \leq \nu^{(m)}(y, \infty]. \tag{4.6}$$

Combining (4.4) and (4.6), we obtain

$$\lim_{n\to\infty} \frac{n}{k} P(X_i^{(m)} + \varepsilon_i > yb(n/k)) = \nu^{(m)}(y, \infty],$$

and thus, by Lemma 2 (ii), (iii), we conclude (4.3).

Proof of Corollary 1 : Processes (a) and (c) satisfy Assumption 2, with $\nu = \nu_\alpha$, and process (b) with $\nu(x, \infty] = x^{-\alpha/2}$. The claim follows from Theorem 2 and Lemma 4.

We now describe a decomposition and state two lemmas, which will be used in the proof of Theorem 3. By iterating $X_i = A_i X_{i-1} + B_i$ $i - j$ times, for $j < i$, X_i can be decomposed into two components

$$X_i := X_i^{j,i} + \Pi_{j+1}^i X_j \tag{4.7}$$

where

$$X_i^{j,i} = B_i + A_i B_{i-1} + A_i A_{i-1} B_{i-2} + \cdots + A_i A_{i-1} \cdots A_{j+2} B_{j+1},$$

and

$$\Pi_{j+1}^i = A_i A_{i-1} \cdots A_{j+1}.$$

Note that X_j is independent of $X_i^{j,i}$, Π_{j+1}^i. We will work with the decomposition

$$Y_i = X_i + \varepsilon_i = X_i^{j,i} + \Pi_{j+1}^i X_j + \varepsilon_i. \tag{4.8}$$

LEMMA 1. *Under Assumptions 3 and 5, there exists* $C < \infty$ *such that for* $0 < \eta < x$, $0 < \delta < y$, $\tau > 0$, *and* $s < t$,

$$
\begin{aligned}
&P(Y_s > x, Y_t > y) \\
&\quad \leq P(X_0 > x - \eta) P(X_0 > y - \delta - \tau) + C \frac{k}{n} c_0^{t-1} + P(\varepsilon_0 > \delta) + P(\varepsilon_0 > \eta).
\end{aligned}
$$

Proof.

$$
\begin{aligned}
P(Y_s > x, Y_t > y) &\leq P(X_s + \varepsilon_s > x, Y_t > y, \varepsilon_s \leq \eta) + P(\varepsilon_s > \eta) \\
&\leq P(X_0 > x - \eta) P(X_0 > y - \delta - \tau) \\
&\quad + C \frac{k}{n} c_0^{t-1} + P(\varepsilon_0 > \delta) + P(\varepsilon_0 > \eta).
\end{aligned}
$$

The last inequality holds by Lemma $4.1(c)$ of [55].

□

LEMMA 2. *Suppose* $i_1 < i_2 < \cdots < i_s$, $y_i > 0$ *for* $i = 1, \ldots, s$, *and* $\tau > 0$. *Under Assumptions 3 and 5,*

(i)

$$
\begin{aligned}
&\left| P\Big(\frac{Y_{i_1}}{(n/k)^{1/\alpha}} > y_1, \ldots, \frac{Y_{i_s}}{(n/k)^{1/\alpha}} > y_s \Big) - P\Big(\frac{Y_{i_1}}{(n/k)^{1/\alpha}} > y_1 \Big) \cdots P\Big(\frac{Y_{i_s}}{(n/k)^{1/\alpha}} > y_s \Big) \right| \\
&\leq \sum_{q=1}^{s-1} \Big(\prod_{j=1}^{s-q} P(Y_0 > (n/k)^{1/\alpha} y_j) P(Y_0 \in ((n/k)^{1/\alpha}(y_{s-q+1} - \tau), (n/k)^{1/\alpha}(y_{s-q+1} + \tau)]) \\
&\times \prod_{j=s-q+2}^{s} P(Y_0 > (n/k)^{1/\alpha}(y_j - \tau)) \Big) + \sum_{j=2}^{s} P(\Pi_1^{i_j - i_{j-1}} X_0 > (n/k)^{1/\alpha} \tau)
\end{aligned}
$$

(ii) Moreover, there exists $M = M(y_1, \ldots, y_s)$ and $K = K(y_1, \ldots, y_s)$ such that for n large enough,

$$\left| P\Big(\frac{Y_{i_1}}{(n/k)^{1/\alpha}} > y_1, \ldots, \frac{Y_{i_s}}{(n/k)^{1/\alpha}} > y_s\Big) - P\Big(\frac{Y_{i_1}}{(n/k)^{1/\alpha}} > y_1\Big) \cdots P\Big(\frac{Y_{i_s}}{(n/k)^{1/\alpha}} > y_s\Big) \right|$$
$$\leq K\tau(s-1)M^{s-1}\Big(\frac{k}{n}\Big)^s + \tau^{-\gamma} E Y_0^{\gamma} \Big(\frac{k}{n}\Big)^{\gamma/\alpha} \sum_{j=2}^{s} c_0^{i_j - i_{j-1}}.$$

Proof. The verification of (i) uses a similar idea to that developed in the proof of Lemma 4.1 of [55], which uses induction. We however work with the observations Y_i, which include the measurement errors ε_i. We start with proving that (i) holds for $s = 2$ and then show that (i) also holds for $s = 3$. We use the decomposition described as (4.7) and (4.8).

For $x > 0, 0 < \tau < y$, and $s < t$

$$\begin{aligned}
&P(Y_s > x, Y_t > y) \\
&\leq P(Y_s > x, X_t^{s,t} + \Pi_{s+1}^t X_s + \varepsilon_t > y,\ \Pi_{s+1}^t X_s \leq \tau) + P(\Pi_{s+1}^t X_s > \tau) \\
&\leq P(Y_s > x) P(X_t^{s,t} + \varepsilon_t > y - \tau) + P(\Pi_{s+1}^t X_s > \tau),
\end{aligned}$$

by the independence of Y_s and $X_t^{s,t} + \varepsilon_t$. Then,

$$P(Y_s > x, Y_t > y) \leq P(Y_s > x)P(Y_t > y) + P(Y_0 > x)P(Y_0 \in (y - \tau, y]) + P(\Pi_1^{t-s} X_0 > \tau).$$

Also, observe that

$$\begin{aligned}
&P(Y_s > x)P(Y_t > y) \\
&\leq P(Y_s > x)P(Y_t \in (y, y+\tau]) + P(Y_s > x)P(Y_t > y + \tau) \\
&\leq P(Y_0 > x)P(Y_0 \in (y, y+\tau]) + P(Y_s > x, Y_t > y) + P(\Pi_1^{t-s} X_0 > \tau).
\end{aligned}$$

Therefore,

$$|P(Y_s > x, Y_t > y) - P(Y_s > x)P(Y_t > y)| \leq P(Y_0 > x)P(Y_0 \in (y-\tau, y+\tau]) + P(\Pi_1^{t-s} X_0 > \tau).$$

We now prove the inequality for $s = 3$ using the result for $s = 2$. For $s < t < u$, and $x, y, z > 0$

$$\begin{aligned}
&P(Y_s > x, Y_t > y, Y_u > z) \\
&\leq P(Y_s > x, Y_t > y, X_u^{t,u} + \Pi_{t+1}^u X_t + \varepsilon_u > z,\ \Pi_{t+1}^u X_t \leq \tau) + P(\Pi_{t+1}^u X_u > \tau) \\
&\leq P(Y_s > x)P(Y_t > y)P(Y_u > z) + P(Y_0 > x)P(Y_0 > y)P(Y_0 \in (z-\tau, z]) \\
&+ P(Y_0 > x)P(Y_0 \in (y-\tau, y])P(Y_0 > z-\tau) \\
&+ P(\Pi_1^{t-s} X_0 > \tau) + P(\Pi_1^{u-t} X_0 > \tau).
\end{aligned}$$

Also, observe that

$$\begin{aligned}
&P(Y_s > x)P(Y_t > y)P(Y_u > z) \\
&\leq P(Y_s > x)P(Y_t > y)P(Y_u \in (z, z+\tau]) + P(Y_s > x)P(Y_t > y)P(Y_u > z+\tau) \\
&\leq P(Y_0 > x)P(Y_0 > y)P(Y_0 \in (z, z+\tau]) \\
&+ P(Y_s > x, Y_t > y, Y_u > z) + P(Y_0 > x)P(Y_0 \in (y, y+\tau])P(Y_0 > z-\tau) \\
&+ P(\Pi_1^{t-s} X_0 > \tau) + P(\Pi_1^{u-t} X_0 > \tau).
\end{aligned}$$

Therefore,

$$\begin{aligned}
|P(Y_s > x, Y_t > y, Y_u > z) &- P(Y_s > x)P(Y_t > y)P(Y_u > z)| \\
&\leq P(Y_0 > x)P(Y_0 > y)P(Y_0 \in (z-\tau, z+\tau]) \\
&+ P(Y_0 > x)P(Y_0 \in (y-\tau, y+\tau])P(Y_0 > z-\tau) \\
&+ P(\Pi_1^{t-s} X_0 > \tau) + P(\Pi_1^{u-t} X_0 > \tau).
\end{aligned}$$

Replacing Y_i by $Y_i/(n/k)^{1/\alpha}$, we conclude (i).

For (ii), we can use Lemma $4.1(b)$ of [55], since the X_i satisfy Assumption 3 and $P(Y_i > \cdot) \in RV_{-\alpha}$ by Lemma 5 and Proposition 1.

$\square$

Proof of Theorem 3 : First, Y_i satisfies (4.1) since $P(Y_i > \cdot) \in RV_{-\alpha}$ by Lemma 5 and Proposition 1. Thus, we conclude (4.1) by Lemma 2. Next, by Lemma 1, for any $x > 0, y > 0$,

$$
\begin{aligned}
&\frac{n}{k^2}\sum_{j=2}^{k} P(Y_1 > (n/k)^{1/\alpha}x, Y_j > (n/k)^{1/\alpha}y) \\
&\le \frac{n}{k^2}\sum_{j=2}^{k}\big\{P(X_0 > (n/k)^{1/\alpha}(x-\eta))P(X_0 > (n/k)^{1/\alpha}(y-\delta-\tau)) \\
&+ C\frac{k}{n}c_0^{t-1} + P(\varepsilon_0 > (n/k)^{1/\alpha}\delta) + P(\varepsilon_0 > (n/k)^{1/\alpha}\eta)\big\} \\
&\le \frac{n(k-1)}{k^2}P(X_0 > (n/k)^{1/\alpha}(x-\eta))P(X_0 > (n/k)^{1/\alpha}(y-\delta-\tau)) \\
&+ C\frac{1}{k}\sum_{j=2}^{k} c_0^{j-1} + \frac{n(k-1)}{k^2}P(\varepsilon_0 > (n/k)^{1/\alpha}\delta) + \frac{n(k-1)}{k^2}P(\varepsilon_0 > (n/k)^{1/\alpha}\eta).
\end{aligned}
$$

By Lemma 5 (i) and Lemma 2,

$$
\begin{aligned}
\frac{n}{k}P(X_0 > (n/k)^{1/\alpha}(x-\eta)) &\to \nu_\alpha(x-\eta,\infty] = (x-\eta)^{-\alpha} < \infty, \\
\frac{n}{k}P(X_0 > (n/k)^{1/\alpha}(y-\delta-\tau)) &\to \nu_\alpha(y-\delta-\tau,\infty] = (y-\delta-\tau)^{-\alpha} < \infty.
\end{aligned}
$$

Also, by Assumption 5,

$$
\frac{n}{k}P(\varepsilon_0 > (n/k)^{1/\alpha}\delta) = \frac{P(\varepsilon_0 > (n/k)^{1/\alpha}\delta)}{P(X_0 > (n/k)^{1/\alpha}\delta)}\frac{P(X_0 > (n/k)^{1/\alpha}\delta)}{P(X_0 > (n/k)^{1/\alpha})} \to 0 \times \delta^{-\alpha} = 0,
$$
$$
\frac{n}{k}P(\varepsilon_0 > (n/k)^{1/\alpha}\eta) \to 0.
$$

Since $\sum_{j=1}^{k} c_0^{j-1} < \infty$, we conclude (4.2). We also conclude that Y_i satisfies (4.3) by Lemma 2 (ii) and Lemma 4.2 of [55]. Therefore, we conclude the claim by Lemma 6.

Proof of Theorem 4 : We will show that

$$\frac{1}{k}\sum_{i=1}^{n} 1_{Y_i/b(n/k)} \Rightarrow \nu, \tag{4.9}$$

where $\nu(x,\infty] = \theta_1 x^{-\alpha}$ and

$$\theta_r = \frac{ED_1^{(r)}\pi_r}{E\sum_{j=1}^{m} D_1^{(j)}\pi_j}, \quad r = 1,\cdots,m.$$

To verify (4.9), we follow, up to (4.10), the argument developed in the proof of Proposition 5.1 of [55]. Following (4.10), we handle the difference caused by the additive error and verify that the difference is negligible under Assumption 5.

To show (4.9), it suffices to prove the convergence of Laplace transforms, that is, for $f \in C_K^+(0,\infty]$,

$$Ee^{-1/k\sum_{i=1}^{n} f(Y_i/b(n/k))} \to e^{-\nu(f)}.$$

Define for $n \geq 0$,

$$N_n^{(r)} = \sum_{l=0}^{n} 1_{[V_l=r]}, \quad r = 1,\ldots,m.$$

By the conditional independence of the $\{X_i\}$ given $\{V_i\}$ and the independence between $\{X_i\}$ and $\{\varepsilon_i\}$,

$$\begin{aligned}
Ee^{-1/k\sum_{i=1}^{n} 1_{Y_i/b(n/k)}(f)} &= Ee^{-1/k\sum_{i=1}^{n} f(Y_i/b(n/k))} \\
&= E(E(e^{-1/k\sum_{i=1}^{n} f(Y_i/b(n/k))}|V_1,\ldots,V_n)) \\
&= E\Big(\prod_{i=1}^{n} E(e^{-1/kf(Y_i/b(n/k))}|V_1,\ldots,V_n)\Big) \\
&= E\prod_{r=1}^{m}\Big(\int_0^\infty e^{-f(y)/k}P(Y/b(n/k)\in dy|V=r)\Big)^{N_n^{(r)}}.
\end{aligned}$$

Note that $N_n^{(r)}/n \xrightarrow{P} \theta_r$, shown as (5.6) in the proof of Proposition 5.1 of [55].

Since $1 - e^{-t} \leq t$, for all $t \in \mathbb{R}$,

$$\left(\int_0^\infty e^{-f(y)/k} P(Y/b(n/k) \in dy | V = r) \right)^{N_n^{(r)}}$$
$$= \left(1 - \frac{\int_0^\infty (1 - e^{-f(y)/k}) n P(Y/b(n/k) \in dy | V = r)}{n} \right)^{N_n^{(r)}}$$
$$\geq \left(1 - \frac{\int_0^\infty f(y) \frac{n}{k} P(Y/b(n/k) \in dy | V = r)}{n} \right)^{N_n^{(r)}}.$$

Also, since $1 - e^{-f(y)/k} = \frac{f(y)}{k} + \frac{1}{2} e^{-2c} (\frac{f(y)}{k})^2$, for some c, which satisfies $|c| \leq \frac{f(y)}{k}$,

$$\left(\int_0^\infty e^{-f(y)/k} P(Y/b(n/k) \in dy | V = r) \right)^{N_n^{(r)}}$$
$$\leq \left(1 - \frac{\int_0^\infty f(y) \frac{n}{k} P(Y/b(n/k) \in dy | V = r)}{n} \right)^{N_n^{(r)}}.$$

For $r = 1$,

$$\frac{n}{k} P(Y/b(n/k) \in \cdot | V = 1) \xrightarrow{v} \nu_\alpha,$$

since $\bar{F}_1 \in RV_{-\alpha}$. Therefore,

$$\left(\int_0^\infty e^{-f(y)/k} P(Y/b(n/k) \in dy | V = r) \right)^{N_n^{(1)}} \xrightarrow{P} e^{-\theta_1 \nu_\alpha (f)}. \tag{4.10}$$

For $2 \leq r \leq m$, $y > 0$,

$$\begin{aligned}
&\frac{n}{k} P(Y > b(n/k) y | V = r) \\
&\quad = \frac{P(Y > b(n/k) y | V = r)}{\bar{F}_1(b(n/k))} \\
&\quad \leq \frac{P(X > (1-\delta) b(n/k) y | V = r) + P(\varepsilon > \delta b(n/k) y | V = r)}{\bar{F}_1(b(n/k))} \\
&\quad = \frac{\bar{F}_r((1-\delta) b(n/k) y) + \bar{F}_\varepsilon(\varepsilon > \delta b(n/k) y)}{\bar{F}_1(b(n/k))} \\
&\quad = \frac{\bar{F}_r((1-\delta) b(n/k) y)}{\bar{F}_1((1-\delta) b(n/k) y)} \frac{\bar{F}_1((1-\delta) b(n/k) y)}{\bar{F}_1(b(n/k))} + \frac{\bar{F}_\varepsilon(\varepsilon > \delta b(n/k) y)}{\bar{F}_1(\delta b(n/k) y)} \frac{\bar{F}_1(\delta b(n/k) y)}{\bar{F}_1(b(n/k))} \to 0,
\end{aligned}$$

by Assumptions 4, 5. Therefore,

$$\int_0^\infty f(y)\frac{n}{k}P(Y/b(n/k) \in dy|V = r) \leq \|f\| \int_c^\infty \frac{n}{k}P(Y/b(n/k) \in dy|V = r)$$
$$= \|f\|\frac{n}{k}P(Y > b(n/k)c|V = r) \to 0,$$

where $c := \inf\{\text{supp}(f)\} > 0$ and $\|f\| := \sup_{(0,\infty]} f$, concluding

$$\left(\int_0^\infty e^{-f(y)/k}P(Y/b(n/k) \in dy|V = r)\right)^{N_n^{(r)}} \xrightarrow{P} (e^{-0})^{\theta_r} = 1.$$

Since $\prod_{r=1}^m \left(\int_0^\infty e^{-f(y)/k}P(Y/b(n/k) \in dy|V = r)\right)^{N_n^{(r)}}$ is bounded, it is uniformly integrable, which leads to (4.9). The weak convergence (4.9) implies $Y_{(k)}/b(n/k) \xrightarrow{P} {\theta_1}^{1/\alpha}$, whose proof is similar to that described in Proposition 2.1 of [21]. Following the same steps developed in Propositions 2.2, 2.3, and 2.4 of [21], we conclude (4.2).

Chapter 5

Hill-type estimators applied to error contaminated data: large sample normality and confidence intervals

5.1 Introduction

Heavy-tailed phenomena have been found in a variety of fields, including finance, insurance, computer network traffic and geophysics. The theory of regular variation provides a mathematical framework for their analysis. Hundreds of papers have been written on the subject, and it is difficult to present an unbiased selection of the most important contributions, so we merely cite here the book of [4], and discuss the most closely related references, as the presentation progresses. This work is concerned with the estimation of the tail index, α, of a heavy–tailed distribution from observations contaminated by measurement or other errors. We investigate asymptotic and finite sample properties of the Hill estimator, which is the most commonly used tool for inference on α, and of the harmonic moment estimator (HME), which is a class of estimators related to and generalizing the Hill estimator.

Suppose $\{X_i, i \geq 1\}$ is a sequence of independent, nonnegative random variables with common one dimensional marginal distribution function F, which has regularly varying tail probabilities, i.e.

$$\bar{F}(x) = 1 - F(x) = \mathbb{P}(X_i > x) = x^{-\alpha} L(x), \quad \alpha > 0, \tag{5.1}$$

where L is a slowly varying function. The class of distributions with tail behavior (5.1) coincides with the maximum domain of attraction of the Fréchet distribution, one of the three basic types of extreme value distributions. The Hill estimator is defined as

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

with the convention that $X_{(i)}$ is the i–th largest order statistic. Throughout the paper, we assume that

$$n \to \infty,\ k \to \infty,\ \frac{k}{n} \to 0. \tag{5.2}$$

The Hill estimator is often used after an examination of the Hill plot, which is also a tool for detecting the presence of heavy tails. The Hill plot and the Hill estimator have been extensively studied, and are introduced in all monographs on extreme value theory, see e.g. [16], [76], [2], [4] and [77]. Considerable research has been done to establish conditions for the asymptotic normality of the Hill estimator. If only the regular variation (5.1) is assumed, asymptotic normality holds with random centering. Several authors formulated conditions on F, which permit replacing the random centering by a deterministic one. The first result of this type was established by [78] for slowly varying functions, L, which converge to a constant at a polynomial rate. [79] showed that the estimator is asymptotically normal for any regularly varying function satisfying the von Mises condition, their centering, however, depends on the sample size n. To show that the Hill estimator centered by the exponent α^{-1} is asymptotically normal, second–order regular variation, a refinement of the concept of regular variation, is assumed, see [39], [40], [41], and [42]. The approach in Section 9.1 of [4], which is based on tail empirical processes, also requires the second–order regular variation. [80] also use the tail empirical process to study asymptotic normality of the Hill estimator for long memory stochastic volatility models assuming a second order condition.

The HME was introduced by [81] to provide a broad class of estimators, which, in a sense, extend the Hill estimator and have desirable robustness properties against large outliers. Consistency and asymptotic normality of the HME was established by [81] for the Pareto distribution and by [82] under a second–order regular variation condition. The HME was also studied, under a different name, by [83], [84] and [85]. The HME is defined in [82] by

$$H_{k,n}^{(\beta)} := \frac{1}{\beta - 1}\left\{\left[\frac{1}{k}\sum_{i=1}^{k}\left(\frac{X_{(k)}}{X_{(i)}}\right)^{\beta-1}\right]^{-1} - 1\right\},$$

where $\beta > 0$, $\beta \neq 1$, is a tuning parameter. For $\beta = 1$, the HME is defined by $H_{k,n}^{(1)} := \lim_{\beta\to 1} H_{k,n}^{(\beta)}$. We therefore obtain the Hill estimator as the limit of the HME as $\beta \to 1$.

We study the Hill estimator and the HME based on observations contaminated by measurement errors, or other errors whose origin is either difficult to understand and model or to quantify precisely. We thus assume that we observe

$$Y_i = X_i + \varepsilon_i,\ 1 \le i \le n,$$

where the ε_i are i.i.d. random errors and independent of the X_i. The Hill estimator computed from the observations Y_i is then

$$\widehat{H}_{k,n} := \frac{1}{k}\sum_{i=1}^{k-1}\log\frac{Y_{(i)}}{Y_{(k)}},$$

and the HME based on the Y_i is

$$\widehat{H}_{k,n}^{(\beta)} := \frac{1}{\beta - 1}\left\{\left[\frac{1}{k}\sum_{i=1}^{k}\left(\frac{Y_{(k)}}{Y_{(i)}}\right)^{\beta-1}\right]^{-1} - 1\right\}.$$

In our context, $\widehat{H}_{k,n}$, $\widehat{H}_{k,n}^{(\beta)}$ are the estimators that can be actually used since what we observe are the Y_i, not the X_i satisfying (5.1). The consistency of the Hill estimator $\widehat{H}_{k,n}$ has been established for a class of error distributions whose tail is lighter than the tail of the X_i, in [86]. In this paper, we want to find conditions on the errors under which the asymptotic normality of $\widehat{H}_{k,n}$, $\widehat{H}_{k,n}^{(\beta)}$ continues to hold. We start with conditions on the X_i satisfying (5.1), which implies the asymptotic normality with random centering. We then consider the second–order regular variation condition and the exact Pareto distribution to derive the asymptotic normality with a constant centering. Some specific questions we seek to answer are as follows. What must we assume about the errors ε_i to obtain asymptotic normality with random centering? What additional assumptions are needed for the deterministic centering? In either case, are any additional assumptions on the rate of k, beyond (5.2), needed? Which characteristics of the distribution of the ε_i enter into these assumptions? In finite samples, how “large", and in what sense, can the ε_i be for the asymptotic confidence intervals to remain useful? It is hoped that the research we present answers such questions in a useful and informative way.

The problem of estimation in the presence of errors has received considerable attention. For example, [59], [60], [61], and [62] study estimation of the end-point of data observed with additive measurement errors. While they all show asymptotic normality in the presence of Gaussian measurement errors, in our case we assume a broader class of error distributions, which includes the normal distribution. This is due to the fact that the heavy-tailed X_i are "much larger" random variables than those with a finite end-point. Most closely related is the work of [87], in which the asymptotic normality of the Hill estimator for round–off data is established. [87] assume that the observations have the form $Y_i = 10^{-l}[10^l U_i^{-1/\alpha}]$, where U_i is uniform on $[0, 1]$ and $[\cdot]$ denotes the integer part. Such data can be written in the form of $Y_i = X_i + \varepsilon_i$, where $X_i = U_i^{-1/\alpha}$ has the exact Pareto distribution and $\varepsilon_i = 10^{-l}[10^l U_i^{-1/\alpha}] - U_i^{-1/\alpha}$ is a bounded error of a specific form. We consider broader classes for both the X_i and the ε_i under the assumption that ε_i is independent of X_i, reflecting our treatment of the ε_i as a measurement error. We use a different asymptotic approach. We establish weak convergence of suitable empirical tail processes for observations contaminated by general errors. Asymptotic normality then follows easily from these general results, which are also of independent interest.

The paper is organized as follows. Assumptions and main theoretical results are stated in Section 5.2. In Section 5.3, we present simulation studies examining finite sample properties of confidence intervals based on the asymptotic normal distribution, focusing on the impact of errors. This numerical investigation is followed in Section 5.4 by an application to the interarrival times of internet traffic anomalies. The proofs are presented in Section 5.6 after some preparation in Section 5.5.

5.2 Assumptions and main asymptotic results

Recall that the observations are $Y_i = X_i + \varepsilon_i$, $1 \le i \le n$. We first state the assumptions on the unobservable random variables X_i. Recall that a function $U : \mathbb{R}_+ \to \mathbb{R}_+$ is regularly varying with index $-\alpha$, $\alpha > 0$, denoted $U \in RV_{-\alpha}$, if

$$\lim_{t \to \infty} \frac{U(tx)}{U(t)} = x^{-\alpha}, \quad \text{for any } x > 0.$$

ASSUMPTION 1. [Regular variation] The X_i are nonnegative, independent random variables with common distribution function F_X such that $\bar{F}_X = P(X_i > \cdot) \in RV_{-\alpha}$.

As noted in the Introduction, even without measurement errors, the assumption of regular variation implies asymptotic normality of the Hill estimator only with random centering. To conduct inference, in particular to obtain useful confidence intervals, one needs a result with centering by $1/\alpha$. For this, second–order regular variation is typically assumed.

ASSUMPTION 2. [Second–order regular variation (2RV)] The X_i are nonnegative, independent random variables with common distribution function F_X, which is second–order $(-\alpha, \rho)$ regularly varying (written $\bar{F}_X \in 2RV(-\alpha, \rho)$), i.e. there exists a positive function $g \in RV_\rho$ such that $g(t) \to 0$, as $t \to \infty$, and for $\alpha > 0$, $\rho \le 0$, $K \ne 0$.

$$\lim_{t \to \infty} \frac{1}{g(t)} \left(\frac{\bar{F}_X(tx)}{\bar{F}_X(t)} - x^{-\alpha} \right) = H(x) := K x^{-\alpha} \frac{x^\rho - 1}{\rho}, \quad x > 0. \tag{5.1}$$

Note that Assumption 2 implies Assumption 1. Observe, however, that condition (5.1) does not hold if the X_i have the exact Pareto distribution, i.e. $P(X_i > x) = x^{-\alpha}$. In this case one would need to allow $K = 0$, and would thus lose any information contained in the function g. The case of exact Pareto tails should however be included in any reasonable theory for heavy–tailed observations. We do so by introducing a parallel set of assumptions.

ASSUMPTION 3. [Pareto] The X_i are nonnegative, independent random variables with a common distribution function F_X such that $\bar{F}_X(x) = P(X_i > x) = x^{-\alpha},\ x \ge 1,\ \alpha > 0$.

The function g in (5.1) can be interpreted as the convergence rate of $\bar{F}_X(tx)/\bar{F}_X(t)$ to $x^{-\alpha}$. It has been used to restrict the sequence $k = k(n)$. [39], [40], [41], and [42] assume that

$$\sqrt{k} g(b(n/k)) \to 0, \tag{5.2}$$

along with the second–order regular variation for $\rho \leq 0$. In (5.2), and throughout the paper, $b(\cdot)$ is the quantile function, defined by $P(X_i > b(t)) = t^{-1}$. It has the representation

$$b(t) = t^{1/\alpha} L_b(t), \tag{5.3}$$

where L_b is a slowly varying function. The condition (5.2) is sufficient in our setting under the additional assumption $\rho > -1$. To cover the 2RV case with $\rho \leq -1$ and the pure Pareto case, we consider the following condition:

$$\frac{\sqrt{k}}{b(n/k)} \to 0. \tag{5.4}$$

Using (5.3), it is easy to verify that (5.2) implies $k = o(n^{-2\rho/(\alpha-2\rho)})$, and (5.4) implies $k = o(n^{2/(\alpha+2)})$. These two rates agree at the phase transition point $\rho = -1$. We use Assumption 4 in the 2RV case and Assumption 5 in the Pareto case.

ASSUMPTION 4. [2RV] The sequence $k = k(n)$ satisfies (5.2) if $\rho > -1$ and (5.4) if $\rho \leq -1$.

ASSUMPTION 5. [Pareto] The sequence $k = k(n)$ satisfies (5.4).

We now turn to the assumptions on the measurement errors ε_i. To get the consistency of the Hill estimator, the only assumption on the errors is that they have lighter tails than X, as was assumed in [86]. We will see that this assumption is also sufficient to establish the asymptotic normality of the Hill estimator with random centering.

ASSUMPTION 6. The ε_i are i.i.d. with tails satisfying

$$P(|\varepsilon| > x) = o(P(X > x)), \text{ as } x \to \infty.$$

The sequence $\{\varepsilon_i\}$ is independent of the sequence $\{X_i\}$.

To obtain the asymptotic normality of the Hill estimator and the HME with a constant centering, a stronger but still broadly applicable assumption on the errors is needed; the errors must have lighter tails than a power function. Assumption 7 is needed when we assume the second–order regular variation, and Assumption 8 is suitable for the Pareto distribution.

ASSUMPTION 7. [2RV] The ε_i satisfy Assumption 6 and

$$P(|\varepsilon| > x) = o(x^{-\kappa}), \text{ as } x \to \infty, \tag{5.5}$$

for some $\kappa > \alpha + \max(-\rho, 1)$.

ASSUMPTION 8. [Pareto] The ε_i satisfy Assumption 6 and (5.5) for some $\kappa > \alpha + 1$.

We now proceed to define the function spaces in which our functional convergence results hold. We work in $D[0,\infty)$, the Skorokhod space of real–valued, right–continuous functions on $[0,\infty)$ with finite left limits existing on $(0,\infty)$. For any $s > 0$, the Skorokhod metric in $D[0,s]$ is defined by

$$d_s(x,y) = \inf_{\lambda \in \Lambda_s} ||\lambda - e||_s \vee ||x - y \circ \lambda||_s, \quad x, y \in D[0,s],$$

where $\Lambda_s = \{\lambda : [0,s] \mapsto [0,s],\ \lambda(0) = 0,\ \lambda(s) = s,\ \lambda(\cdot)$ is continuous, strictly increasing$\}$, and $\|\cdot\|_s$ is the supremum norm on $[0,s]$. The Skorokhod metric on $D[0,\infty)$ is then defined by

$$d_\infty(x,y) = \int_0^\infty e^{-s}(d_s(r_s x, r_s y) \wedge 1) ds, \quad x, y \in D[0,\infty),$$

where $r_s x, r_s y$ are the restrictions of $x, y \in D[0,\infty)$ to the interval $[0,s]$. Given a sequence of random processes, $X_n, n \geq 0$, in $D[0,\infty)$, we denote weak convergence of X_n to X_0 by $X_n \Rightarrow X_0$. We also use $\Rightarrow$ to denote weak convergence of random variables.

We define two "increasingly empirical" measures, with only the last one being observable. We set

$$\nu_n := \frac{1}{k}\sum_{i=1}^{n} I_{Y_i/b(n/k)}, \quad \hat{\nu}_n := \frac{1}{k}\sum_{i=1}^{n} I_{Y_i/Y_{(k)}},$$

with $b(\cdot)$ defined in (5.3). The random measures ν_n, $\hat{\nu}_n$, and all other Radon measures of this type, are defined on $(0, \infty]$ compactified at ∞. Thus, for $s \geq 0$, we can define the random processes

$$W_n(s) = \sqrt{k}(\nu_n(s^{-1/\alpha}, \infty] - E\nu_n(s^{-1/\alpha}, \infty]),$$
$$\widehat{W}_n(s) = \sqrt{k}(\hat{\nu}_n(s^{-1/\alpha}, \infty] - E\hat{\nu}_n(s^{-1/\alpha}, \infty]).$$

We first investigate the asymptotic normality of the tail empirical processes W_n, $\widehat{W}_n$, then study when it implies the asymptotic normality of the Hill estimator $\widehat{H}_{k,n}$ and the HME $\widehat{H}_{k,n}^{(\beta)}$. Theorem 1 shows that even very general errors specified in Assumption 6 do not impact the asymptotic behavior of the tail empirical processes W_n nor $\widehat{W}_n$: the limit distributions of these statistics based on the Y_i are the same as those of the corresponding statistics based on the unobservable X_i.

THEOREM 1. *Under Assumptions 1 and 6,*

$$W_n \Rightarrow W \quad \text{in } D[0, \infty), \tag{5.6}$$

and

$$\widehat{W}_n \Rightarrow W \quad \text{in } D[0, \infty), \tag{5.7}$$

where W is the standard Brownian motion on $[0, \infty)$.

The Hill estimator can be written as an integral of the tail empirical measure $\hat{\nu}_n$, i.e.

$$\widehat{H}_{k,n} = \int_1^\infty \frac{1}{k}\sum_{i=1}^{n} I_{Y_i/Y_{(k)}}(s, \infty] s^{-1} ds = \int_1^\infty \hat{\nu}_n(s, \infty] s^{-1} ds.$$

Similarly, the HME can be expressed as a transformed integral of the tail empirical measure $\hat{\nu}_n$, i.e.

$$\widehat{H}_{k,n}^{(\beta)} = \frac{1}{\beta-1}\left\{\left[(1-\beta)\widehat{M}_{k,n}+1\right]^{-1}-1\right\}, \quad \beta \neq 1,$$

where

$$\widehat{M}_{k,n}^{(\beta)} := \int_1^\infty \hat{\nu}_n(s,\infty]s^{-\beta}ds = \frac{1}{1-\beta}\left[\frac{1}{k}\sum_{i=1}^{k}\left(\frac{Y_{(k)}}{Y_{(i)}}\right)^{\beta-1}-1\right].$$

The order statistics used to compute the Hill estimator and the HME must be positive. In the following, all statements are tacitly assumed to hold conditional on the event $\{Y_{(k)} > 0\}$, where k is the count of the largest order statistics in the definition of $\widehat{H}_{k,n}$, $\widehat{H}_{k,n}^{(\beta)}$.

THEOREM 2. *Suppose that Assumptions 1 and 6 hold. If $\alpha > 0$ and $\beta > 1 - \alpha/2$,*

$$\sqrt{k}\left(\int_1^\infty \hat{\nu}_n(s,\infty]s^{-\beta}ds - \int_{Y_{(k)}}^\infty \frac{n}{k}\bar{F}_Y(s)s^{-\beta}ds\right) \Rightarrow \frac{1}{\alpha}\int_0^1 W(s)s^{\frac{\beta-1}{\alpha}-1}ds.$$

By putting $\beta = 1$ in Theorem 2 we obtain the asymptotic normality of the Hill estimator with random centering, which is stated as Corollary 1 (a). Similarly, the asymptotic behavior of $\widehat{M}_{k,n}^{(\beta)}$ follows directly from Theorem 2, which is presented in Corollary 1 (b).

COROLLARY 1. *Under the Assumptions of Theorem 2,*

(a)

$$\sqrt{k}\left(\widehat{H}_{k,n} - \int_{Y_{(k)}}^\infty \frac{n}{k}\bar{F}_Y(s)\frac{ds}{s}\right) \Rightarrow \frac{1}{\alpha}\int_0^1 W(s)\frac{ds}{s},$$

(b) if $\beta \neq 1$, then

$$\sqrt{k}\left(\widehat{M}_{k,n}^{(\beta)} - \int_{Y_{(k)}}^\infty \frac{n}{k}\bar{F}_Y(s)\frac{ds}{s^\beta}\right) \Rightarrow \frac{1}{\alpha}\int_0^1 W(s)s^{\frac{\beta-1}{\alpha}-1}ds.$$

We emphasize that Theorem 1, Theorem 2, and Corollary 1 hold either under Assumption 2 or Assumption 3, since both imply Assumption 1.

The convergence in Theorem 2 requires random centering with $\int_{Y_{(k)}}^\infty n/k\bar{F}_Y(s)s^{-\beta}ds$, which makes Corollary 1 of limited practical use, but it provides a starting point for improvements. To replace it with a constant centering, we need the assumption of second–order regular variation (or of exact Pareto tails) and the stronger assumptions on the errors. In the following theorem, we establish the asymptotic normality of the integral of the tail empirical measure, $\int_1^\infty \hat{\nu}_n(s,\infty]s^{-\beta}ds$, with a constant centering.

THEOREM 3. *Suppose either Assumptions 2, 4, and 7 (2RV case), or Assumptions 3, 5, and 8 (Pareto case) hold. If $\alpha > 0$ and $\beta > 1 - \alpha/2$, then*

$$\sqrt{k}\left(\int_1^\infty \hat{\nu}_n(s,\infty]s^{-\beta}ds - \frac{1}{\alpha+\beta-1}\right) \Rightarrow N\left(0, \frac{\alpha}{(\alpha+\beta-1)^2(\alpha+2\beta-2)}\right).$$

The asymptotic normality of the Hill estimator $\widehat{H}_{k,n}$ follows easily from Theorem 3. To obtain the asymptotic normality of the HME $\widehat{H}_{k,n}^{(\beta)}$, we must apply Theorem 3 and the delta method. The corresponding results are stated in the following corollary.

COROLLARY 2. *Under the assumptions of Theorem 3,*

(a)

$$\sqrt{k}\left(\widehat{H}_{k,n} - \frac{1}{\alpha}\right) \Rightarrow N(0, 1/\alpha^2),$$

(b) if $\beta \neq 1$, then

$$\sqrt{k}\left(\left[(1-\beta)\widehat{M}_{k,n}^{(\beta)} + 1\right] - \frac{\alpha}{\alpha+\beta-1}\right) \Rightarrow N\left(0, \frac{\alpha(1-\beta)^2}{(\alpha+\beta-1)^2(\alpha+2\beta-2)}\right)$$

and

$$\sqrt{k}\left(\widehat{H}_{k,n}^{(\beta)} - \frac{1}{\alpha}\right) \Rightarrow N\left(0, \frac{(\alpha+\beta-1)^2}{\alpha^3(\alpha+2\beta-2)}\right).$$

We note that the results in Corollary 2 are the same as for observations without measurement errors; see Theorem 3.2.5 of [2] and Theorem 2 of [82]. The effect of relatively small errors ε_i is thus asymptotically negligible. We also remark that Corollary 2 (a) cannot be easily proven by verifying the conditions in Theorem 3.2.5 of [2]. If the X_i are exactly Pareto or second–order regularly varying, it is not clear if the Y_i are in any of these classes. Proposition 1 is a related result which plays an important role in the proof of Theorem 3, which is a general new result of independent interest.

In the next two sections, we explore how small the errors must be in finite samples to have a practically negligible effect on confidence interval inference. Then, we present preliminary results in Section 5.5, followed by the proof of the main results in Section 5.6. Supplementary tables are provided in Appendix B.

5.3 Impact of errors on confidence intervals

We investigate the effect of error contaminations on confidence intervals constructed using the more commonly used Hill estimator. The effect of Pareto errors on the harmonic moment estimator (HME) is studied in Section 5 of [82], in a more limited, but informative, simulation study.

The asymptotic level $1-p$ confidence interval for α^{-1} implied by Corollary 2 (a) is

$$\left(\frac{1}{\hat{\alpha}} - z_{p/2}\frac{1}{\hat{\alpha}\sqrt{k}},\ \frac{1}{\hat{\alpha}} + z_{p/2}\frac{1}{\hat{\alpha}\sqrt{k}}\right), \tag{5.1}$$

where $\hat{\alpha}^{-1} = \widehat{H}_{k,n}$, and z_q is the upper quantile of the standard normal distribution defined by $\Phi(z_q) = 1-q$. The above interval is implemented by the function `hill` of the `R` package `evir`, with the default asymptotic coverage $1-p=0.95$. According to Corollary 2 (a), it is asymptotically valid even if the observations are contaminated by fairly general errors. In this section, we investigate the impact of these errors on the empirical coverage probability of the interval (5.1). To obtain interval (5.1), the number of upper order statistics, k, has to be chosen. We consider a range of values of k for a given sample size n. We also employ a few methods of selecting k, which have been proposed.

The design of our simulation study is as follows. We generate observations $Y_i = X_i + \varepsilon_i,\ i = 1, 2, \ldots, n$, where $\{X_i\}$ and $\{\varepsilon_i\}$ are independent sets of random variables. For each model/error pair, we compute 1000 confidence intervals and report the fraction of the intervals that contain the reciprocal of the true tail index. We consider sample size $n = 500$, which is representative of the sample sizes occurring in the application presented in Section 5.4.

We use two models for the X_i, both satisfying the condition of Corollary 2 (a) and having the true tail index $\alpha = 2$. The first is the standard Pareto distribution, which is not second order regularly varying, and the second is a distribution in the Hall/Weiss class. The Hall/Weiss class provides examples of the second–order regular variation, see p. 142 of [88]. Model 2 satisfies Assumption 2 with $g(t) = t^{-5}$.

Model 1 [Pareto] The X_i are i.i.d. random variables, which follow a Pareto distribution with $\alpha = 2$, $P(X_i > x) = x^{-2}, \ \ x \geq 1$.

Model 2 [2RV] The X_i are i.i.d. random variables, which follow the Hall/Weiss class with $\alpha = 2$ and $\rho = -5$, $P(X_i > x) = x^{-2}(1 + x^{-5})/2, \ \ x \geq 1$.

We consider four different distributions for the errors ε_i. They all satisfy Assumptions 7 and 8 (with $\alpha = 2$), since for each of them $P(|\varepsilon| > x) = o(x^{-\kappa})$, for some $7 < \kappa < 8$.

Error 1 [Normal] The ε_i are i.i.d. random variables, drawn from a normal distribution with mean 0 and standard deviation σ_{Normal}.

Error 2 [scaled t_8] The ε_i are i.i.d. random variables, drawn from a scaled t-distribution with 8 degrees of freedom.

Error 3 [GPD] The ε_i are i.i.d. random variables, drawn from a generalized Pareto distribution, $P(|\varepsilon| > z) = (1 + \xi(z - \mu)/\sigma)^{-1/\xi}$, with location $\mu = 0$, shape $\xi = 1/8$, and scale σ_{GPD}.

Error 4 [Uniform] The ε_i are i.i.d. random variables, drawn from the uniform distribution on the interval $[-a, a]$, $a > 0$.

The scale parameters for each error distribution vary. They are determined by the ratio of the standard deviation of error distribution (error SD) to the standard deviation of underlying process (model SD). For example, if the ratio is 0.1 for Model 1 whose standard deviation is 2.44, then $\sigma_{Normal} = 0.244$ for Error 1, the corresponding scale for Error 2 is 0.183, $\sigma_{GPD} = 0.131$ for Error 3, and a for Error 4 is 0.423. We consider several values of the ratio and then obtain the confidence interval (5.1) for each of them.

We first examine the robustness of coverage probabilities to the errors in finite samples, considering a wide range of k for a given sample size n. Tables B.1 and B.2 in Section B.1 report coverage probabilities of the approximate 95% confidence intervals for the Pareto model, with $n = 500$ and $n = 2000$, respectively. We first observe that the coverage probabilities for samples generated from the Pareto distribution without the errors are close to the target coverage, 95%, for large k's. This is found in the row with the ratio 0 in each table. This result is in agreement with the typical behavior of the Hill plot showing stable, unbiased estimates for large k when its underlying distribution is exactly a Pareto distribution. Second, the coverage overall decreases with the ratio, but this decrease is relatively flat over a range of the ratio from 0.01 to 0.1, for all the error types. In particular, for $n = 2000$, the coverage is surprisingly acceptable for a wide range of values of k; in many cases it is close to the target of 95%. On the other hand, the coverage seems sensitive to relatively large errors with the ratio more than 10 percent. An interesting observation is that, in the presence of errors, the coverage gets worse as k gets larger. This result is consistent with Corollary 2 (a), which implies that the Hill estimator obtains the asymptotic normality if k satisfies Assumption 5; k goes to infinity with n, but not too fast. The reduction in the coverage probability caused by large k is not observed for data contaminated by relatively small errors. Finally, the impact on the coverage probability overall does not depend on the type of the error distribution. In particular, for the small ratios, the difference that the error type makes looks negligible.

Tables B.3 and B.4 in Section B.1 report coverage probabilities of the asymptotic 95% confidence intervals for the 2RV model, with $n = 500$ and $n = 2000$, respectively. Unlike the Pareto case, the 2RV model does not achieve the target coverage, 95%, even if there are no errors. This may be due to n not being sufficiently large. The errors with small ratio, however, have only a small impact on the coverage. It can be also seen that the impact on the coverage probability for small ratio does not depend on the error type. Finally, we see that k cannot increase too fast, indirectly confirming the need for Assumption 4.

We have found so far that the coverage can achieve the target probability for some properly chosen k or cannot achieve it for any k, given a finite sample. Even if we can identify some range of k for which the coverage approaches the target, the question still remains of how to select an optimal k in practice. There are various methods for choosing it. A commonly used approach is based on the minimization of the asymptotic mean squared error (AMSE), see e.g. [89], [66], [90], and [91]. These methods are however based on asymptotic arguments, which brings up a question of how well they perform in finite samples. [92] suggest a data driven method minimizing a penalty function of the distance between empirical quantiles and theoretical quantiles to improve the performance in finite samples. There are also heuristic methods, mainly trying to find the region where the Hill plot, a plot of estimates of the tail index against k, becomes more stable, see [42].

To provide practically useful information on choosing a data–driven cut–off k, we examined four methods based on different underlying ideas of selecting the optimal k. The first threshold selection method, introduced by [66], uses a bootstrap procedure to find the k which minimizes the AMSE. This value is computed by the function `hall` of the `R` package `tea`. (We also considered a few related methods based on the minimization of the AMSE argument, but they all gave disappointing results. The coverage that the Hall method produced was always among the best of these methods.) The second method, proposed by [92], is based on minimizing a penalty function of the distance between the observed quantile and the fitted Pareto type tail. This distance is in the quantile dimension, not in the probability dimension like the Kolmogorov–Smirnov distance. This method is suggested to remedy the behavior that a small change in probabilities makes a large difference in quantiles. We use two different penalty functions: the supremum of the absolute distance (KS), and the mean absolute distance (MAD). Both are implemented by the function `mindist` of the `R` package `tea`. The final method is an Eye–Ball technique whose automatic algorithm is developed by [92] and is carried out by the function `eye` of the `R` package `tea`. This heuristic method attempts to find a stable portion of the Hill plot and obtain the k at which a considerable drop in the variance occurs, as k increases.

Tables 5.1 and 5.2 report coverage probabilities and the average optimal k selected using the four different methods. For the Pareto model, the coverage decreases with the ratio for all the

Table 5.1: Proportion (in percent) of the approximate 95% confidence intervals including $1/\alpha$ and the average optimal k in parentheses, for $n = \mathbf{500}$ and the **Pareto** model. The Hall, MAD, KS, and Eye–Ball methods are used to choose the optimal k. The target coverage is 95 percent.

Method	Error Type	Error SD/Model SD Ratio						
		0	0.01	0.02	0.05	0.1	0.2	0.3
Hall	Normal	88.9 (283)	87.6 (311)	88.4 (330)	88.9 (329)	83.8 (301)	77.5 (256)	71.2 (222)
	scaled t_8	88.7 (283)	88.0 (321)	88.6 (337)	88.9 (322)	83.2 (289)	77.6 (242)	68.9 (201)
	GPD	89.4 (285)	88.9 (322)	88.8 (340)	88.7 (320)	83.7 (283)	76.9 (220)	72.7 (169)
	Uniform	89.1 (284)	88.2 (308)	88.3 (329)	87.9 (329)	80.1 (301)	73.1 (265)	61.3 (238)
MAD	Normal	97.0 (218)	97.4 (214)	96.8 (214)	97.6 (198)	96.8 (147)	97.4 (92)	96.2 (68)
	scaled t_8	97.1 (219)	97.2 (219)	97.4 (214)	97.8 (200)	97.2 (156)	97.2 (107)	97.2 (79)
	GPD	97.1 (219)	97.2 (216)	98.0 (214)	98.2 (191)	97.8 (145)	98.2 (95)	98.0 (77)
	Uniform	97.0 (218)	97.4 (219)	96.4 (220)	96.2 (195)	97.0 (151)	94.8 (99)	93.6 (70)
KS	Normal	83.4 (68)	82.2 (67)	84.0 (68)	81.2 (77)	77.2 (93)	75.0 (83)	67.6 (79)
	scaled t_8	83.6 (68)	83.6 (67)	83.5 (69)	84.2 (71)	81.7 (90)	77.4 (85)	71.9 (82)
	GPD	83.6 (68)	84.4 (68)	83.8 (66)	83.8 (72)	82.4 (83)	77.9 (74)	75.1 (62)
	Uniform	83.4 (68)	84.0 (70)	82.0 (69)	82.0 (80)	78.6 (92)	69.4 (103)	63.6 (101)
Eye	Normal	95.3 (51)	95.1 (51)	94.8 (51)	95.2 (51)	94.8 (51)	93.2 (51)	90.5 (50)
	scaled t_8	95.3 (51)	95.4 (51)	95.5 (51)	95.3 (51)	93.5 (51)	92.7 (50)	88.2 (50)
	GPD	95.3 (51)	95.2 (51)	94.9 (51)	95.2 (51)	93.5 (50)	91.7 (50)	86.0 (49)
	Uniform	95.3 (51)	95.1 (51)	95.0 (51)	95.6 (51)	94.3 (51)	93.6 (51)	92.0 (51)

Table 5.2: Proportion (in percent) of the approximate 95% confidence intervals including $1/\alpha$ and the average optimal k in parentheses, for $\mathbf{n = 500}$ and the **2RV** model. The Hall, MAD, KS, and Eye–Ball methods are used to choose the optimal k. The target coverage is 95 percent.

Method	Error Type	Error SD/Model SD Ratio 0	0.01	0.02	0.05	0.1	0.2	0.3
Hall	Normal	75.3 (118)	75.6 (119)	75.0 (142)	12.9 (395)	8.8 (416)	37.2 (326)	34.7 (250)
	scaled t_8	75.8 (118)	75.3 (119)	74.4 (130)	29.0 (339)	0.8 (429)	29.1 (366)	37.4 (293)
	GPD	75.8 (118)	76.2 (119)	72.5 (150)	28.3 (340)	1.9 (422)	17.6 (368)	38.3 (284)
	Uniform	75.6 (118)	75.2 (119)	74.0 (129)	33.8 (316)	26.4 (410)	35.0 (314)	30.3 (256)
MAD	Normal	18.7 (222)	18.2 (221)	18.5 (221)	16.3 (228)	7.5 (304)	36.6 (150)	70.8 (90)
	scaled t_8	18.7 (222)	18.9 (221)	18.8 (222)	17.0 (223)	8.5 (311)	21.0 (200)	51.7 (121)
	GPD	18.7 (222)	18.3 (221)	18.8 (221)	16.3 (223)	12.1 (270)	22.2 (203)	66.2 (127)
	Uniform	18.7 (222)	18.4 (222)	18.1 (221)	16.1 (240)	11.1 (282)	40.6 (134)	60.2 (82)
KS	Normal	66.6 (104)	66.6 (102)	67.0 (104)	66.4 (104)	56.7 (152)	52.5 (160)	53.0 (117)
	scaled t_8	66.6 (104)	66.9 (103)	66.8 (105)	67.0 (100)	66.3 (117)	53.9 (175)	53.0 (136)
	GPD	66.6 (104)	66.4 (103)	67.5 (104)	67.0 (101)	65.6 (109)	56.8 (155)	58.8 (115)
	Uniform	66.6 (104)	67.1 (102)	66.8 (102)	66.3 (107)	53.9 (169)	51.0 (166)	48.4 (140)
Eye	Normal	93.6 (51)	93.9 (51)	93.4 (51)	93.7 (51)	92.6 (51)	88.7 (51)	77.8 (50)
	scaled t_8	93.6 (51)	93.9 (51)	94.6 (51)	93.9 (51)	91.9 (51)	91.1 (50)	83.3 (50)
	GPD	93.6 (51)	93.8 (51)	93.8 (51)	93.9 (51)	92.5 (51)	88.1 (50)	82.3 (50)
	Uniform	93.6 (51)	93.7 (51)	93.8 (51)	94.0 (51)	92.7 (51)	90.4 (51)	82.9 (51)

selection methods as shown in Table 5.1; again, a small ratio has a relatively small impact on the coverage. The MAD and Eye–Ball methods achieve the target coverage, 95%, when the underlying process is not contaminated by the errors. These methods also are less sensitive to the ratio increase. For the Pareto model, the MAD approach generally leads to coverage probabilities which are higher than 95%. However, as shown in Table 5.2, it gives very low coverage for the 2RV model. It has an unexpected, difficult to explain, property of the coverage increasing with the ratio. The Hall method also shows some fluctuation over the ratio, but this fluctuation is not found when the ratio is 0.01 and 0.02. The other methods also exhibit this insensitivity for small ratios. The Eye–Ball method seems to work well for the Pareto and 2RV models since it gives relatively high values of coverage. Its average optimal k also falls into the optimal range which gives high values of coverage in Tables B.1 and B.3 in Section B.1.

The main conclusions of the above detailed discussion are as follows.

1. The Eye–Ball method of selecting k is recommended for both the Pareto and 2RV models.

2. For the heavy–tailed X_i with the tail index $\alpha = 2$, the coverage probability of the approximate 95% confidence interval containing the true index is robust to errors whose SD does not exceed 2 percent of model SD.

3. There is no clear evidence that the coverage probability depends on the error distribution. Instead, the coverage is mainly affected by how large the ε_i are compared to the X_i, regardless of the threshold selection methods.

We conclude this section with a discussion of the confidence interval for α obtained via an application of the delta method. Corollary 2 (a) and the delta method imply that

$$\sqrt{k}(\widehat{H}_{k,n}^{-1} - \alpha) \Longrightarrow N(0, \alpha^2).$$

Thus, setting $\tilde{\alpha} = \widehat{H}_{k,n}^{-1}$, we get the approximate level $1-p$ confidence interval for α of the form

$$\left(\tilde{\alpha} - z_{p/2}\frac{\tilde{\alpha}}{\sqrt{k}},\ \tilde{\alpha} + z_{p/2}\frac{\tilde{\alpha}}{\sqrt{k}}\right). \tag{5.2}$$

One might want to use the interval (5.2) rather than (5.1) to make inference on α, but care is needed in finite samples. Since the delta method is based on an additional asymptotic approximation, confidence intervals derived from it could provide a poor approximation for small sample sizes. We have performed a simulation study for the interval (5.2), similar to the one described earlier in this section. We have found that it almost always gives coverage probability worse than the interval (5.1). Therefore, when working with sample sizes similar to $n = 500$, we recommend using the reciprocals of the bounds of the interval (5.1).

5.4 Application to Internet2 anomalous traffic

In this section, we present an application to interarrival times of anomalies in a backbone internet network, Internet2. These times are available only with round–off errors. We provide only minimal background; more details are presented in [67], a paper which to some extent motivates the present research. We describe results of confidence interval inference for the tail index of these interarrival times. We restrict ourselves to confidence intervals based on the Hill estimator, the results for the HME are similar. We then examine the robustness of the Hill estimator to the round-off errors by a numerical experiment.

The Internet2 network consists of 14 two–directional links connecting major cities in the United States, as shown in Figure 4.3. A traffic disruption in any of these links can slow down service in the whole country. For this reason, anomalies in the internet traffic have been extensively studied. An anomaly is a time and space confined traffic whose volume is much higher than typical. [67] developed an anomaly extraction algorithm. [86] and [93] argue that the interarrival times have heavy tails. The anomaly extraction algorithm can identify the arrival time of an anomaly in any unidirectional link only in a resolution of five minutes. While network measurement devices operate at much higher frequencies, such a rough resolution is due to the limitation of the anomaly extraction algorithm. It is based on the Fourier transform, which eliminates noise by retaining only low frequency harmonics. [67] created a database for the time period of 50 weeks, starting October 16, 2005. A question we seek to answer in this section is if the round–off error has a negligible or a non–negligible impact on the confidence intervals for the tail index of the interarrival times. Additionally, we would like to see if the various data–driven methods of selecting k, discussed in Section 5.3 lead to overlapping confidence intervals, or if they suggest different ranges of α. These conclusions could potentially be different for each of the 28 unidirectional links. We index these links by integers from 1 to 28 since it is not important for the purpose of our investigation to which nodes they correspond.

Tables 5.3 and 5.4 report tail index estimates and 95% confidence intervals for each link, obtained using the four methods of selecting k discussed in Section 5.3. We first observe that all methods, except for the KS method, generally produce similar point estimates for each link. The interval estimates from the KS method are generally wider. In particular, some links have the infinity as the upper end. This is manually put to deal with a negative lower end of the of the interval (5.1). We now check whether intervals from the four methods overlap. We find 20 links with a nonempty intersection of the four intervals and 8 links with an empty intersection. The intersection does not have any interpretation in the usual frequentist sense of [94], but it provides, so to say, the safest region in an engineering sense, for the 20 links for which it is nonempty. For the links with the empty intersection, or even for all links, we recommend using the confidence interval produced from the Eye–Ball method, which can be considered the most reliable estimate based on the simulation result of Section 5.3.

In the context of this paper, each interarrival time Y_i, computed by the algorithm, is treated as a "true" interarrival time X_i measured with a round–off error, i.e. $Y_i = X_i + \varepsilon_i$. The unobserved X_i is not rigorously defined, but we can think of it as the time separation based on a more precise algorithm, or just a different algorithm. In the latter case, the analysis that follows provides information about the uncertainty in the estimation of α caused by the choice of a specific

Table 5.3: Point estimates and 95% confidence intervals for the tail index of the anomalies interarrival times. The link index along with the sample size are displayed. The estimates are obtained using the Hall, MAD, KS, and Eye–Ball methods. The intersection of the four intervals is shown if it is nonempty, an empty intersection is indicated by $\emptyset$.

Link	1 ($n = 405$)		2 ($n = 247$)		3 ($n = 362$)		4 ($n = 454$)	
Hall	1.70	$(1.3, 2.4)$	1.50	$(1.1, 2.5)$	1.63	$(1.3, 2.1)$	1.64	$(1.3, 2.1)$
MAD	1.43	$(1.2, 1.8)$	1.21	$(1.0, 1.6)$	1.51	$(1.1, 2.3)$	1.28	$(0.9, 2.6)$
KS	3.19	$(1.3, \infty)$	3.06	$(1.6, 24.8)$	4.66	$(2.0, \infty)$	2.08	$(1.2, 8.0)$
Eye	1.79	$(1.4, 2.4)$	1.23	$(1.0, 1.8)$	1.60	$(1.3, 2.2)$	1.59	$(1.3, 2.1)$
Overlap		$(1.4, 1.8)$		$(1.6, 1.6)$		$(2.0, 2.1)$		$(1.3, 2.1)$
	5 ($n = 347$)		6 ($n = 345$)		7 ($n = 603$)		8 ($n = 300$)	
Hall	1.54	$(1.2, 2.1)$	1.59	$(1.2, 2.2)$	1.64	$(1.3, 2.2)$	1.45	$(1.1, 2.1)$
MAD	1.43	$(1.2, 1.8)$	1.49	$(1.0, 2.9)$	1.31	$(0.9, 2.4)$	1.27	$(1.0, 1.7)$
KS	1.88	$(1.3, 3.2)$	3.35	$(1.9, 12.9)$	5.34	$(3.2, 17.4)$	3.43	$(2.1, 9.9)$
Eye	1.53	$(1.2, 2.1)$	1.52	$(1.2, 2.1)$	1.38	$(1.1, 1.7)$	1.38	$(1.1, 1.9)$
Overlap		$(1.3, 1.8)$		$(1.9, 2.1)$		$\emptyset$		$\emptyset$
	9 ($n = 387$)		10 ($n = 345$)		11 ($n = 382$)		12 ($n = 304$)	
Hall	1.48	$(1.2, 2.1)$	1.44	$(1.1, 2.1)$	1.83	$(1.4, 2.6)$	2.27	$(1.6, 3.7)$
MAD	1.31	$(1.1, 1.7)$	1.24	$(1.0, 1.6)$	1.36	$(1.1, 1.8)$	1.50	$(1.2, 2.0)$
KS	3.98	$(2.3, 15.4)$	2.85	$(1.7, 9.3)$	3.63	$(1.7, \infty)$	2.51	$(1.7, 4.9)$
Eye	1.52	$(1.2, 2.0)$	1.39	$(1.1, 1.9)$	1.72	$(1.4, 2.3)$	1.60	$(1.3, 2.2)$
Overlap		$\emptyset$		$\emptyset$		$(1.7, 1.8)$		$(1.7, 2.0)$
	13 ($n = 476$)		14 ($n = 507$)		15 ($n = 478$)		16 ($n = 319$)	
Hall	2.16	$(1.7, 3.0)$	1.96	$(1.5, 3.0)$	2.07	$(1.6, 3.0)$	1.44	$(1.1, 2.0)$
MAD	1.58	$(1.0, 3.9)$	1.44	$(0.9, 3.5)$	1.46	$(0.9, 3.4)$	1.36	$(1.0, 2.2)$
KS	2.06	$(1.6, 2.9)$	3.85	$(1.3, \infty)$	2.05	$(1.6, 2.9)$	3.32	$(1.9, 16.6)$
Eye	2.02	$(1.6, 2.7)$	1.60	$(1.3, 2.1)$	1.80	$(1.5, 2.3)$	1.47	$(1.2, 2.0)$
Overlap		$(1.7, 2.7)$		$(1.5, 2.1)$		$(1.6, 2.3)$		$(1.9, 2.0)$

algorithm. Since the smallest value of Y_i in physical units is 5 min., we use 5 minutes as a unit lag. We therefore assume that the errors ε_i are uniformly distributed on $[-1, 1]$. We performed the following numerical experiment. For each link, we treat the value of α estimated from the real interarrival times as the true value. We then generate $R = 1,000$ replications of error contaminated data $Y_i^{(r)} = X_i + \varepsilon_i^{(r)}$, $1 \leq r \leq 1000$. With some abuse of notation, the X_i are now the observed interarrival times. For each of these replications we compute the interval (5.1) with $p = 10\%$ and $p = 5\%$. To choose k, we use the Hall, MAD, KS, and Eye–Ball methods described in Section 5.3. For each link, we determine the percentage of these intervals that cover the value of α estimated

Table 5.4: Continuation of Table 5.3.

Link	17 ($n = 402$)		18 ($n = 388$)		19 ($n = 433$)		20 ($n = 493$)	
Hall	1.91	$(1.5, 2.5)$	1.36	$(1.1, 1.9)$	1.27	$(1.1, 1.6)$	1.90	$(1.5, 2.6)$
MAD	1.51	$(1.0, 3.7)$	1.22	$(1.0, 1.6)$	1.27	$(1.0, 1.9)$	1.45	$(0.9, 3.3)$
KS	1.96	$(1.5, 2.8)$	3.22	$(1.7, 26.1)$	2.63	$(1.2, \infty)$	1.97	$(1.6, 2.6)$
Eye	1.86	$(1.5, 2.5)$	1.31	$(1.1, 1.8)$	1.51	$(1.2, 2.0)$	1.83	$(1.5, 2.4)$
Overlap		$(1.5, 2.5)$		$\emptyset$		$(1.2, 1.6)$		$(1.6, 2.4)$
	21 ($n = 340$)		22 ($n = 417$)		23 ($n = 597$)		24 ($n = 296$)	
Hall	1.97	$(1.5, 2.9)$	1.46	$(1.2, 1.9)$	1.67	$(1.3, 2.2)$	1.56	$(1.2, 2.2)$
MAD	1.51	$(1.0, 3.3)$	1.38	$(1.1, 2.0)$	1.26	$(0.8, 2.9)$	1.28	$(1.0, 1.7)$
KS	2.01	$(1.5, 3.0)$	3.61	$(1.5, \infty)$	3.67	$(2.1, 14.2)$	3.44	$(2.0, 13.3)$
Eye	1.87	$(1.5, 2.6)$	1.54	$(1.2, 2.1)$	1.50	$(1.2, 1.9)$	1.43	$(1.1, 2.0)$
Overlap		$(1.5, 2.6)$		$(1.5, 1.9)$		$\emptyset$		$\emptyset$
	25 ($n = 258$)		26 ($n = 340$)		27 ($n = 348$)		28 ($n = 264$)	
Hall	1.78	$(1.3, 2.9)$	1.48	$(1.1, 2.3)$	1.95	$(1.5, 2.9)$	1.58	$(1.2, 2.3)$
MAD	1.35	$(1.0, 1.9)$	1.20	$(0.9, 1.7)$	1.64	$(1.0, 4.7)$	1.38	$(1.0, 2.4)$
KS	4.11	$(1.7, \infty)$	3.57	$(2.2, 10.3)$	2.80	$(1.6, 14.0)$	2.70	$(1.3, \infty)$
Eye	1.38	$(1.1, 2.0)$	1.25	$(1.0, 1.7)$	1.71	$(1.4, 2.4)$	1.60	$(1.2, 2.3)$
Overlap		$(1.7, 1.9)$		$\emptyset$		$(1.6, 2.4)$		$(1.3, 2.3)$

from real data. If the interarrival times were measured perfectly, i.e. $\varepsilon_i \equiv 0$, then 100% of these intervals would cover the “true value", so our target in this experiment is 100% rather than 95% or 90% as in Section 5.3. If the actual coverage is $100(1-q)\%$, then we interpret q as the probability of getting a wrong interval estimate due to the round-off error. It turned out that for all links we achieved the target percentage, 100%, for both 95% and 90% confidence levels, regardless of the threshold selection methods. In light of the results of Section 5.3, the 100% coverage could be expected since the ratio of the Error SD to the observation SD is less than 0.001 for each link. We have seen from Tables 5.1 and 5.2 that the errors with the ratio of 0.01 had almost no impact on the coverage probability. Based on this 100% coverage, we conclude that the impact of the round–off error on the confidence interval estimate from the real data is practically negligible. This allows us to use the available rough interarrival times to make an inference on the tail index.

The conclusions of the research described in this section are as follows.

1. For the purpose of confidence interval inference on the tail index of the anomalies interarrival times, the five minute resolution is acceptable.

2. For most links the confidence intervals obtained using the four data–driven methods of selecting k have a nonempty intersection.

3. Based on the Eye–Ball method, one can be confident that for all links the true value of α is between 1.0 and 2.7. The most typical range for α is $(1.2, 2.3)$; each interval for half of the links falls into the range.

5.5 Preliminary results

We collect in this section a number of lemmas to avoid burdening the proofs in Section 5.6 with additional explanations. Many of these lemmas are results established earlier, in such cases we list their sources. Lemmas for which direct sources could not be found are proven. We denote by Y, X, ε, the random variables with the same distribution as, respectively, each Y_i, X_i, ε_i.

Lemma 1 states useful properties of the Skorokhod metric. It follows from facts presented on pp. 47, 48 of [4]. For a sequence of deterministic functions, $x_n, n \geq 0$, in $D[0,\infty)$, we denote convergence of x_n to x_0 by $x_n \to x_0$. The uniform metric on $D[0,s]$ is defined by $||x-y||_s := \sup_{0\leq t\leq s}|x(t)-y(t)|$.

LEMMA 1. *Suppose* $x, x_n, y \in D[0,s]$, $z_n \in D[0,\infty)$, *for* $n \geq 0$.

(i) The Skorokhod metric on $D[0,s]$ *is bounded above by the uniform metric on* $D[0,s]$, *i.e.* $d_s(x,y) \leq ||x-y||_s$.

(ii) If $d_s(x_n,x_0) \to 0$, *then for all* $0 \leq t \leq s$ *satisfying* $t \in C(x_0)$, *the set of continuity points of* x_0, $x_n(t) \to x_0(t)$.

(iii) If $d_s(x_n,x_0) \to 0$ *and* $x_0 \in C[0,s]$, *the space of continuous functions on* $[0,s]$, *then we have the uniform convergence,* $||x_n - x_0||_s \to 0$.

(iv) $d_\infty(z_n,z_0) \to 0$ *if and only if for any* $s \in C(z_0)$, $d_s(r_s z_n, r_s z_0) \to 0$, *where* $r_s z_n, r_s z_0$ *are the restrictions of* z_n, z_0 *to the interval* $[0,s]$.

We get the following lemma by combining the results of Lemma 1.

LEMMA 2. *Suppose $x_n, x \in D[0,\infty)$, for $n \geq 0$. Then, $x_n \to x$ in $D[0,\infty)$ and x is continuous if and only if for any $s \geq 0$,*

$$||x_n - x||_s = \sup_{0 \leq t \leq s} |x_n(t) - x(t)| \to 0.$$

LEMMA 3. *The functions $h_1, h_2 : D[0,\infty) \to \mathbb{R}$ defined, for any fixed M, by*

$$h_1(x) = \int_0^M e^{-s} x(s) ds, \;\; h_2(x) = \int_1^M x(s) s^{-\beta} ds, \;\; \text{for } \beta > 1,$$

are continuous at any function in $C[0,\infty)$.

Proof. Suppose $x_n \to x_0$ in $D[0,\infty)$, where x_0 is continuous. Then

$$\begin{aligned}\left| \int_0^M e^{-s} x_n(s) ds - \int_0^M e^{-s} x_0(s) ds \right| &\leq \int_0^M e^{-s} |x_n(s) - x_0(s)| ds \\ &\leq \sup_{0 \leq s \leq M} |x_n(s) - x_0(s)| \int_0^M e^{-s} ds \\ &= ||x_n - x_0||_M \int_0^M e^{-s} \to 0.\end{aligned}$$

The last term goes to 0 by Lemma 2. The same argument is used for the proof of the continuity of h_2.

□

LEMMA 4. *Suppose random processes D_n in $D[0,\infty)$, $n \geq 1$, satisfy $0 \leq D_n \leq 1$ and*

$$\forall\, s \geq 0, \;\; D_n(s) \xrightarrow{P} 0, \;\; \text{as } n \to \infty. \tag{5.1}$$

Then,

$$\int_0^\infty e^{-s} D_n(s) ds \xrightarrow{P} 0, \quad \text{as } n \to \infty.$$

Proof. Define

$$
\begin{aligned}
I(n) &= \int_0^\infty e^{-s} D_n(s) ds \\
&= \int_0^M e^{-s} D_n(s) ds + \int_M^\infty e^{-s} D_n(s) ds \\
&=: I_M(n) + I_M^\star(n).
\end{aligned}
$$

Fix $\varepsilon > 0$ and observe that

$$P(I(n) > \varepsilon) \le P\left(I_M(n) > \frac{\varepsilon}{2}\right) + P\left(I_M^\star(n) > \frac{\varepsilon}{2}\right).$$

Since $I_M^\star(n) \le e^{-M}$, we can choose M so large that $P\left(I_M^\star(n) > \varepsilon/2\right) = 0$. For such a (fixed) M, applying lemma 3 and the continuous mapping theorem, the assumption (5.1) implies $\int_0^M e^{-s} D_n(s) ds \xrightarrow{P} 0$. Therefore,

$$P(I(n) > \varepsilon) \le P\left(I_M(n) > \frac{\varepsilon}{2}\right) \to 0.$$

$\square$

LEMMA 5. *(i) Suppose $x, x_n, y_n, n \ge 1$, are deterministic functions in $D[0,\infty)$. If $x_n \to x$ in $D[0,\infty)$, and for any $s \ge 0$,*

$$\|y_n - x_n\|_s = \sup_{0 \le t \le s} |y_n(t) - x_n(t)| \to 0,$$

then $y_n \to x$ in $D[0,\infty)$.

(ii) Suppose $X, X_n, Y_n, n \ge 1$, are random processes in $D[0,\infty)$. If $X_n \Rightarrow X$, and for any $s \ge 0$,

$$\|Y_n - X_n\|_s = \sup_{0 \le t \le s} |Y_n(t) - X_n(t)| \xrightarrow{P} 0,$$

then $Y_n \Rightarrow X$ in $D[0,\infty)$.

Proof. For (i), by Lemma 2, $d_\infty(y_n, x_n) \to 0$. Then, we conclude the claim by the triangle inequality. For (ii), by Lemma 1 (i), $d_s(r_s Y_n, r_s X_n) \xrightarrow{P} 0$, for any $s \geq 0$. Then, by Lemma 4, $d_\infty(Y_n, X_n) \xrightarrow{P} 0$. We conclude the claim by the Slutsky theorem.

□

LEMMA 6. *The function* $h : D[0,\infty) \times D[0,\infty) \to D[0,\infty)$*, defined by*

$$h(x,y) = x \circ y = x(y(\cdot)), \text{ for } x, y \in D[0,\infty),\ y \geq 0,$$

is continuous at (x_0, y_0)*, for any* $x_0, y_0 \in C(0,\infty]$.

Proof. The metric on $D[0,\infty) \times D[0,\infty)$ is given by

$$d_{\text{prod}}((x_n, y_n), (x_0, y_0)) = d_\infty(x_n, x_0) + d_\infty(y_n, y_0),$$

see p.57 of [4]. Therefore, if $(x_n, y_n) \to (x_0, y_0)$, then $d_\infty(x_n, x_0) \to 0$, and $d_\infty(y_n, y_0) \to 0$. Now observe that

$$d_\infty(x_n(y_n), x_0(y_0)) \leq d_\infty(x_n(y_n), x_0(y_n)) + d_\infty(x_0(y_n), x_0(y_0)).$$

Since x_0 is continuous, $d_\infty(x_0(y_n), x_0(y_0)) \to 0$ as $d_\infty(y_n, y_0) \to 0$. Next, to show the convergence of $d_\infty(x_n(y_n), x_0(y_n))$, it suffices to show that, for any $s \geq 0$, $||x_n(y_n) - x_0(y_n)||_s = \sup_{0\leq t \leq s} |x_n(y_n(t)) - x_0(y_n(t))| \to 0$. Since $d_\infty(x_n, x_0) \to 0$ and x_0 is continuous on $[0,\infty)$, we can take s' such that $s' \geq \sup_{0 \leq t \leq s} y_n(t)$ and $||x_n - x_0||_{s'} \to 0$, by Lemma 2. Since, for any $s \geq 0$, $||x_n(y_n) - x_0(y_n)||_s \leq ||x_n - x_0||_{s'} \to 0$, we can conclude that $d_\infty(x_n(y_n), x_0(y_n)) \to 0$, by Lemma 2.

□

The next lemma follows from Theorem 4.1 in [4].

LEMMA 7. *Suppose* $\{X_i, i \geq 1\}$ *is a sequence of independent nonnegative random variables with common distribution function* F *such that* $\bar{F}(\cdot) \in RV_{-\alpha}$. *Then*

$$\frac{1}{k}\sum_{i=1}^{n} I_{X_i/b(n/k)} \Rightarrow \nu_\alpha$$

in $M_+(0,\infty]$, *the space of Radon measures on* $(0,\infty]$, *where* $\nu_\alpha(x,\infty] = x^{-\alpha}, x > 0$.

Recall that the function b is defined by $P(X > b(t)) = t^{-1}$. Part (i) of Lemma 8 follows from the definition of regular variation and $\lim_{t\to\infty} b(t) = \infty$; part (ii) was proven as Lemma 4.1 in [44]. It follows from an application of Potter bounds and dominated convergence.

LEMMA 8. *Suppose* $u \mapsto P(X > u) \in RV_{-\alpha}$, *for some* $\alpha > 0$. *Then*

(i) for each $u > 0$, $\lim_{t\to\infty} tP(X > ub(t)) = u^{-\alpha}$,

(ii)

$$\lim_{z\to\infty}\limsup_{t\to\infty}\int_z^\infty tP(X > xb(t))x^{-1}dx = 0.$$

The next lemma extends the results of Lemma 8 to $Y = X + \varepsilon$.

LEMMA 9. *Suppose* $Y = X + \varepsilon$, *and let* $\bar{F}_Y$ *be the tail distribution of* Y. *Suppose that* $u \mapsto P(X > u) \in RV_{-\alpha}$ *for some* $\alpha > 0$, $P(|\varepsilon| > x) = o(P(X > x))$, *and* ε *is independent of* X. *Then*

(i) $u \mapsto P(Y > u) \in RV_{-\alpha}$.

(ii) $\lim_{t\to\infty} \bar{F}_Y(t)/\bar{F}_X(t) = 1$.

(iii) for each $u > 0$, $\lim_{t\to\infty} tP(Y > ub(t)) = u^{-\alpha}$.

(iv) for each $\beta > 1 - \alpha/2$,

$$\lim_{z\to\infty}\limsup_{t\to\infty}\int_z^\infty tP(Y > yb(t))\frac{dy}{y^\beta} = 0.$$

Proof. Statement (i), (ii) are stated and proven in Proposition 3.1 of [86]. For statement (iii), observe

$$\lim_{t\to\infty} tP(Y > ub(t)) = \lim_{t\to\infty} \frac{P(Y > ub(t))}{P(X > b(t))} = \lim_{t\to\infty} \frac{P(Y > ub(t))}{P(X > ub(t))}\frac{P(X > ub(t))}{P(X > b(t))}.$$

By (ii) and $\bar{F}_X \in RV_{-\alpha}$ from the assumption, we get the conclusion. For statement (iv), set $f_t(y) = tP(Y > yb(t))y^{-\beta}$. We want to show

$$\lim_{z\to\infty} \limsup_{t\to\infty} \int_z^\infty f_t(y)dy = 0.$$

By (iii),

$$\forall\, y > 0 \quad f_t(y) \to y^{-\alpha-\beta}, \quad \text{as } t \to \infty.$$

To conclude that

$$\int_z^\infty f_t(y)dy \to \int_z^\infty y^{-\alpha-\beta}dy, \quad \text{as } t \to \infty,$$

we must find a function g such that for $t > t_0$,

$$f_t(y) \le g(y) \quad \text{and} \quad \int_z^\infty g(y)dy < \infty.$$

By the assumption $\bar{F}_X \in RV_{-\alpha}$, we obtain Potter bounds and combine this with (ii), then we get bounds such that $\forall\, \delta, c > 0, \exists\, t_0, \forall\, t \ge t_0, \forall\, y \ge 1$,

$$(1-c)(1-\delta)y^{-\alpha-\delta-\beta} \le f_t(y) = \frac{P(Y > yb(t))}{P(X > yb(t))}\frac{P(X > yb(t))}{P(X > b(t))}y^{-\beta} \le (1+c)(1+\delta)y^{-\alpha+\delta-\beta}.$$

Then $g := (1+c)(1+\delta)y^{-\alpha+\delta-\beta}$ is integrable if $\delta < \alpha/2$.

□

The next lemma follows from Remark B.3.8 in [2].

LEMMA 10. *Suppose that F satisfies Assumption 2. Then under Assumption 4,*

$$\lim_{n\longrightarrow\infty}\sqrt{k}\left(\frac{n}{k}\bar{F}(b(n/k)x)-x^{-\alpha}\right)=0,$$

locally uniformly for $x>0$.

The next result is elementary. It follows for the convergence of characteristic functions.

LEMMA 11. *Suppose that random vectors X_n, Y_n in $\mathbb{R}^d$ are independent for each n and that $X_n \Rightarrow X$ and $Y_n \Rightarrow Y$, where X and Y are independent. Then $X_n + Y_n \Rightarrow X + Y$.*

The next result is known as Vervaat's lemma. It is stated e.g. in Proposition 3.3 in [4]. Recall that for a nondecreasing function H, the left–continuous inverse is defined by $H^{\leftarrow}(u) = \inf\{s : H(s) \geq u\}$.

LEMMA 12. *Suppose $X_n, n \geq 0$, is a sequence of random processes in $D[0,\infty)$ such that each X_n has nondecreasing paths and X_0 has continuous paths. Define the identity function by $e(t) = t$. Then, $c_n(X_n - e) \Rightarrow X_0$ implies that $c_n(X_n^{\leftarrow} - e) \Rightarrow -X_0$.*

The last lemma is well–known, see e.g. Lemma 9.1 in [4].

LEMMA 13. *Suppose that $\alpha > 0$ and $\beta > 1 - \alpha/2$. If W is the standard Wiener process, then for some $a, b \in \mathbb{R}$,*

$$a\int_0^1 W(s)s^{\frac{\beta-1}{\alpha}-1}ds - bW(1) \stackrel{d}{=} N\left(0, \frac{2a^2\alpha^2}{(\alpha+\beta-1)(\alpha+2\beta-2)} - \frac{2ab\alpha}{\alpha+\beta-1} + b^2\right). \quad (5.2)$$

Proof. Since a linear combination of Gaussian random variables is still Gaussian, the left side of (5.2) is Gaussian with mean 0. To get its variance, we observe that

$$E\left(a\int_0^1 W(s)s^{\frac{\beta-1}{\alpha}-1}ds - bW(1)\right)^2$$
$$= E\left(a^2\int_0^1 W(s)s^{\frac{\beta-1}{\alpha}-1}ds\int_0^1 W(u)u^{\frac{\beta-1}{\alpha}-1}du - 2ab\int_0^1 W(s)W(1)s^{\frac{\beta-1}{\alpha}-1}ds + b^2W(1)^2\right).$$

Using $E\left(W(s)W(u)\right) = s \wedge u$, we get

$$
\begin{aligned}
&E\left(a\int_0^1 W(s)s^{\frac{\beta-1}{\alpha}-1}ds - bW(1)\right)^2 \\
&= 2a^2\int_0^1\int_0^u s\wedge u\, s^{\frac{\beta-1}{\alpha}-1}dsu^{\frac{\beta-1}{\alpha}-1}du - 2ab\int_0^1 s\wedge 1\, s^{\frac{\beta-1}{\alpha}-1}ds + b^2 \\
&= \frac{2a^2\alpha^2}{(\alpha+\beta-1)(\alpha+2\beta-2)} - \frac{2ab\alpha}{\alpha+\beta-1} + b^2.
\end{aligned}
$$

□

5.6 Proofs of the main results

5.6.1 Proof of Theorem 1

We begin by proving a couple of lemmas. The following fundamental lemma is known if there are no errors, i.e. $F_Y = F_X$. Using the results of Section 5.5, we can prove it in our setting.

LEMMA 1. *Let $\{E_i, i \geq 1\}$ be i.i.d. exponential random variables with mean 1. Set $S_n = \sum_{i=1}^n E_i$ and define*

$$\hat{\phi}_n(s) = \frac{n}{k}\bar{F}_Y(b(n/k)s^{-1/\alpha})\frac{S_{n+1}}{n}, \quad s > 0. \tag{5.1}$$

Then under Assumptions 1, 6,

$$\hat{\phi}_n(\cdot) \xrightarrow{P} e \quad \text{in } D[0,\infty). \tag{5.2}$$

where e is the identity function, $e(x) = x$.

Proof. By Lemma 9 (iii) and Lemma 2, $\frac{n}{k}\bar{F}_Y(b(n/k)(\cdot)^{-1/\alpha}) \to e \quad \text{in } D[0,\infty)$. Observe that for any $s > 0$,

$$\sup_{0\leq t\leq s}\left|\hat{\phi}_n(t) - \frac{n}{k}\bar{F}_Y(b(n/k)t^{-1/\alpha})\right| \leq \sup_{0\leq t\leq s}\frac{n}{k}\bar{F}_Y(b(n/k)t^{-1/\alpha})\left|\frac{S_{n+1}}{n} - 1\right|.$$

Since $n/k\bar{F}_Y(b(n/k)t^{-1/\alpha}) \to t$, for all $0 \le t \le s$, by Lemma 9 (iii), and $S_{n+1}/n \xrightarrow{P} 1$ by the weak law of large numbers, we obtain

$$\sup_{0\le t\le s}\left|\hat{\phi}_n(t) - \frac{n}{k}\bar{F}_Y(b(n/k)t^{-1/\alpha})\right| \xrightarrow{P} 0.$$

By Lemma 5 (ii), we get the conclusion.

$\square$

Denote by ν_α the measure in $M_+(0,\infty]$, the space of Radon measures on $(0,\infty]$, defined by $\nu_\alpha(x,\infty] = x^{-\alpha},\ x > 0$.

LEMMA 2. *Under Assumptions 1 and 6,* $\nu_n = \frac{1}{k}\sum_{i=1}^n I_{Y_i/b(n/k)} \Rightarrow \nu_\alpha$ *in* $M_+(0,\infty]$.

Proof. By Lemma 7, it suffices to verify that

$$d\left(\frac{1}{k}\sum_{i=1}^n I_{Y_i/b(n/k)}, \frac{1}{k}\sum_{i=1}^n I_{X_i/b(n/k)}\right) \xrightarrow{P} 0.$$

The vague metric $d(\cdot,\cdot)$ on $M_+(0,\infty]$ is defined by

$$d(\mu_1,\mu_2) = \sum_{i=1}^\infty \frac{|\mu_1(f_i) - \mu_2(f_i)| \wedge 1}{2^i}, \quad \mu_1,\mu_2 \in M_+(0,\infty],$$

for some $f_i \in C_K^+(0,\infty]$ where $C_K^+(0,\infty]$ is the space of continuous function with compact support. By the definition of the vague metric d on $M_+(0,\infty]$ and Lemma 3.7 of [44], it suffices to show that, for any $f \in C_K^+((0,\infty])$, $\tau > 0$,

$$P\left(\left|\frac{1}{k}\sum_{i=1}^n f\Big(\frac{Y_i}{b(n/k)}\Big) - \frac{1}{k}\sum_{i=1}^n f\Big(\frac{X_i}{b(n/k)}\Big)\right| > \tau\right) \xrightarrow{P} 0.$$

Since f has compact support in $(0,\infty]$, set $s := \inf\{\text{supp}(f)\} > 0$.

Also, since f is uniformly continuous,

$$w_\eta(f) := \sup_{|x-y|\le\eta,\ x,y\in(0,\infty]} |f(x)-f(y)| \to 0,\ \eta\to 0.$$

Fix $\tau > 0$. Let

$$G(n) = \frac{1}{k}\sum_{i=1}^{n} \left| f\Big(\frac{Y_i}{b(n/k)}\Big) - f\Big(\frac{X_i}{b(n/k)}\Big) \right|.$$

Observe that

$$\begin{aligned}
&P\left(\left| \frac{1}{k}\sum_{i=1}^{n} f\Big(\frac{Y_i}{b(n/k)}\Big) - \frac{1}{k}\sum_{i=1}^{n} f\Big(\frac{X_i}{b(n/k)}\Big) \right| > \tau \right) \\
&\le P(G(n) > \tau) \\
&\le P(G(n)I_{A_n} > \tau/3) + P(G(n)I_{B_n} > \tau/3) + P(G(n)I_{C_n} > \tau/3),
\end{aligned}$$

where

$$\begin{aligned}
A_n &= \Big\{ 1 \le i \le n : \left| \frac{Y_i}{b(n/k)} - \frac{X_i}{b(n/k)} \right| \le \eta,\ \frac{X_i}{b(n/k)} \ge s-\eta \Big\}, \\
B_n &= \Big\{ 1 \le i \le n : \left| \frac{Y_i}{b(n/k)} - \frac{X_i}{b(n/k)} \right| \le \eta,\ \frac{X_i}{b(n/k)} < s-\eta \Big\},
\end{aligned}$$

and

$$C_n = \Big\{ 1 \le i \le n : \left| \frac{Y_i}{b(n/k)} - \frac{X_i}{b(n/k)} \right| > \eta \Big\},$$

for $0 < \eta < s/2$. We start with the bound

$$P(G(n)I_{A_n} > \tau/3) \le P\Big(w_\eta(f) \frac{1}{k}\sum_{i=1}^{n} I_{X_i/b(n/k)}[s-\eta,\infty) > \tau/3 \Big).$$

By Lemma 7, and taking sufficiently small η,

$$P(G(n)I_{A_n} > \tau/3) \le P\left(w_\eta(f)(s-\eta)^{-\alpha} > \tau/3 \right) = 0.$$

Next,

$$P(G(n)I_{B_n} > \tau/3) = 0,$$

since

$$\frac{Y_i}{b(n/k)}, \ \frac{X_i}{b(n/k)} < s, \text{ for all } i \in B_n.$$

By Markov's inequality,

$$\begin{aligned}
P(G(n)I_{C_n} > \tau/3) &\leq \frac{3n}{k\tau} E\left[\left|f\Big(\frac{Y_i}{b(n/k)}\Big) - f\Big(\frac{X_i}{b(n/k)}\Big)\right| I_{C_n}\right] \\
&\leq \frac{6\sup_{x\in(s,\infty]}|f(x)|}{\tau} \frac{P\Big(\Big|\frac{Y_i}{b(n/k)} - \frac{X_i}{b(n/k)}\Big| > \eta\Big)}{P(X_1 > b(n/k))} \\
&\leq \frac{6\sup_{x\in(s,\infty]}|f(x)|}{\tau} \frac{P(|\varepsilon_i| > \eta b(n/k))}{P(X_1 > \eta b(n/k))} \frac{P(X_1 > \eta b(n/k))}{P(X_1 > b(n/k))}.
\end{aligned}$$

Since $\sup_{x\in(s,\infty]}|f(x)| < \infty$, $P(|\varepsilon_i| > \eta b(n/k))/P(X_1 > \eta b(n/k)) \to 0$ by Assumption 6, and $P(X_1 > \eta b(n/k))/P(X_1 > b(n/k)) \to \eta^{-\alpha}$, we obtain $P(G(n)I_{C_n} > \tau/3) \to 0$.

□

The proof of the following lemma is basically the same as the proof in Proposition 2.4.2 of [44], so we skip it.

LEMMA 3. *Under Assumptions 1 and 6,* $Y_{(k)}/b(n/k) \xrightarrow{P} 1$.

Proof of Theorem 1: For (5.6), we use the technique developed in the proof of Theorem 9.1 of [4]. We work with the Y_i which are observed with errors, whereas Theorem 9.1 of [4] applies to the unobservable X_i. We must show that the effect of the errors ε_i is negligible in every step of the proof.

Suppose $\{E_i, i \geq 1\}$ are i.i.d. exponential random variables with mean 1. Set $S_n = \sum_{i=1}^n E_i$, then

$$\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^n I_{S_i \leq kx} - x\right) \Rightarrow W(x), \quad \text{in } D[0,\infty), \tag{5.3}$$

which is established in the proof of Theorem 9.1 of [4]. Consider the functions $\hat{\phi}_n$ defined by (5.1). Combining (5.3) and (5.2), we obtain the joint convergence in $D[0,\infty) \times D[0,\infty)$,

$$\left(\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{n} I_{S_i \le k\cdot} - e\right), \hat{\phi}_n(\cdot)\right) \Rightarrow (W(\cdot), e).$$

By Lemma 6 and the continuous mapping theorem, we have

$$\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{n} I_{S_i \le k\hat{\phi}_n(\cdot)} - \hat{\phi}_n(\cdot)\right) \Rightarrow W(\cdot) \quad \text{in } D[0,\infty). \tag{5.4}$$

Recall that if $Z \sim \text{Gamma}(\alpha = a, \beta = 1)$, $V \sim \text{Gamma}(\alpha = b, \beta = 1)$, and Z and V are independent, then $Z/(Z+V) \sim \text{Beta}(a,b)$. Therefore,

$$\left(\frac{S_1}{S_{n+1}}, \cdots, \frac{S_n}{S_{n+1}}\right) \stackrel{d}{=} \left(1 - \frac{S_n}{S_{n+1}}, \cdots, 1 - \frac{S_1}{S_{n+1}}\right) \stackrel{d}{=} (U_{(n)}, \ldots, U_{(1)}),$$

where $U_{(n)} \le \ldots \le U_{(1)}$ are the order statistics of i.i.d. $U(0,1)$ random variables $U_1, \ldots, U_n$. Using the fact that $F_Y^{\leftarrow}(U_i) = Y_i$, one can easily verify that

$$\frac{1}{k}\sum_{i=1}^{n} I_{S_i \le k\hat{\phi}_n(s)} \stackrel{d}{=} \nu_n[s^{-1/\alpha}, \infty).$$

The verification is the same as shown in the proof of Theorem 9.1 of [4], which uses the fact that $F_X^{\leftarrow}(U_i) = X_i$, so we skip it.

To complete (5.6), it remains to show that

$$\sup_{0 \le t \le s} \sqrt{k}\left|\hat{\phi}_n(t) - E\nu_n[t^{-1/\alpha}, \infty)\right| \xrightarrow{P} 0. \tag{5.5}$$

Observe that for any $s > 0$

$$\sup_{0 \le t \le s} \sqrt{k}\left|\hat{\phi}_n(t) - E\nu_n[t^{-1/\alpha}, \infty)\right| = \sup_{0 \le t \le s} \frac{n}{k}\bar{F}_Y(b(n/k)t^{-1/\alpha})\sqrt{\frac{k}{n}}\left|\frac{S_{n+1} - n}{\sqrt{n}}\right|.$$

Since $n/k\bar{F}_Y(b(n/k)t^{-1/\alpha}) \to t$, for all $0 \le t \le s$, by Lemma 9 (iii), and using $|(S_{n+1}-n)/\sqrt{n}| = O_p(1)$, which follows from the central limit theorem, we obtain (5.5).

For (5.7), by (5.6) and Lemma 3, we have the following joint weak convergence:

$$\left(\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{n} I_{Y_i/b(n/k)}(y,\infty] - \frac{n}{k}\bar{F}_Y(b(n/k)y)\right), \frac{Y_{(k)}}{b(n/k)}\right) \Rightarrow (W(y^{-\alpha}), 1).$$

By Lemma 3 and the continuous mapping theorem, we get (5.7).

5.6.2 Proof of Theorem 2

Proof of Theorem 2: From (5.7), we obtain,

$$\sqrt{k}(\hat{\nu}_n(s,\infty] - E\hat{\nu}_n(s,\infty]) \Rightarrow W(s^{-1/\alpha}) \quad \text{in } D(0,\infty]. \tag{5.6}$$

If we apply the map

$$x \mapsto \int_1^\infty x(s)\frac{ds}{s^\beta} \tag{5.7}$$

to (5.6), we can conclude the claim. The steps of the justification of the use of the map (5.7) are similar to those developed on pp. 298-299 of [4]. We provide the details since we work with the observations Y_i, which include the measurement errors ε_i.

The verification is based on triangular array convergence argument, see Theorem 3.2 in [26]. Set

$$U_n = \int_1^\infty \sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{n} I_{Y_i/Y_{(k)}}(s,\infty] - \frac{n}{k}\bar{F}_Y(Y_{(k)}s)\right)s^{-\beta}ds, \quad U = \int_1^\infty W(s^{-\alpha})s^{-\beta}ds;$$
$$U_n^{(M)} = \int_1^M \sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{n} I_{Y_i/Y_{(k)}}(s,\infty] - \frac{n}{k}\bar{F}_Y(Y_{(k)}s)\right)s^{-\beta}ds, \quad U^{(M)} = \int_1^M W(s^{-\alpha})s^{-\beta}ds.$$

To establish the desired convergence $U_n \Rightarrow U$, we must verify that

$$\forall\, M > 1, \quad U_n^{(M)} \Rightarrow U^{(M)}, \quad \text{as (5.2)}; \tag{5.8}$$

$$U^{(M)} \Rightarrow U, \quad \text{as } M \to \infty; \tag{5.9}$$

$$\forall\, \varepsilon > 0, \quad \lim_{M\to\infty} \limsup_{n\to\infty} P\left(|U_n^{(M)} - U_n| > \varepsilon\right) = 0. \tag{5.10}$$

Convergence (5.8) follows from Theorem 1, Lemma 3 and the continuous mapping theorem. Convergence (5.9) holds since

$$P\left(\int_M^\infty W(s^{-\alpha}) s^{-\beta} ds > \delta\right) \le \frac{1}{\delta^2 \alpha^2} E\left[\left(\int_0^{M^{-\alpha}} W(s) s^{\frac{\beta-1}{\alpha}-1} ds\right)^2\right] \to 0.$$

Convergence (5.10) is equivalent to

$$\forall\, \varepsilon > 0, \quad \lim_{M\to\infty} \limsup_{n\to\infty} P\left(\sqrt{k}\int_M^\infty \left|\frac{1}{k}\sum_{i=1}^n I_{Y_i/Y_{(k)}}(s,\infty] - \frac{n}{k}\bar{F}_Y(Y_{(k)}s)\right| s^{-\beta} ds > \varepsilon\right) = 0.$$

Fix $\varepsilon > 0$ and $\eta > 0$. Observe that

$$P\left(\sqrt{k}\int_M^\infty \left|\frac{1}{k}\sum_{i=1}^n I_{Y_i/Y_{(k)}}(s,\infty] - \frac{n}{k}\bar{F}_Y(Y_{(k)}s)\right| s^{-\beta} ds > \varepsilon\right) \le Q_1(n) + Q_2(n),$$

where

$$Q_1(n) = P\left(\sqrt{k}\int_M^\infty \left|\frac{1}{k}\sum_{i=1}^n I_{Y_i/Y_{(k)}}(s,\infty] - \frac{n}{k}\bar{F}_Y(Y_{(k)}s)\right| s^{-\beta} ds > \varepsilon,\ \left|\frac{Y_{(k)}}{b(n/k)} - 1\right| < \eta\right),$$

$$Q_2(n) = P\left(\left|\frac{Y_{(k)}}{b(n/k)} - 1\right| \ge \eta\right).$$

By Lemma 3, $\limsup_{n\to\infty} Q_2(n) = 0$, so we focus on $Q_1(n)$. Since $Y_{(k)}/b(n/k) > 1-\eta$, with the change of variable $s = ub(n/k)/Y_{(k)}$,

$$Q_1(n) \le P\left(\sqrt{k}\int_{M(1-\eta)}^\infty \left|\frac{1}{k}\sum_{i=1}^n I_{Y_i/b(n/k)}(u,\infty] - \frac{n}{k}\bar{F}_Y(b(n/k)u)\right| u^{-\beta} du > \varepsilon\right).$$

By Chebychev's inequality,

$$\begin{aligned}
Q_1(n) &\le \frac{k}{\varepsilon^2} E\left[\left(\int_{M(1-\eta)}^{\infty}\left|\frac{1}{k}\sum_{i=1}^{n} I_{Y_i/b(n/k)}(u,\infty] - \frac{n}{k}\bar{F}_Y(b(n/k)u)\right| u^{-\beta}du\right)^2\right] \\
&\le \frac{k}{\varepsilon^2}\int_{M(1-\eta)}^{\infty} E\left[\left(\frac{1}{k}\sum_{i=1}^{n} I_{Y_i/b(n/k)}(u,\infty] - \frac{n}{k}\bar{F}_Y(b(n/k)u)\right)^2\right] u^{-\beta}du \\
&\le \frac{1}{\varepsilon^2}\int_{M(1-\eta)}^{\infty} \frac{n}{k}\bar{F}_Y(b(n/k)u)u^{-\beta}du \to 0,
\end{aligned}$$

as $M \to \infty$, by Lemma 9 (iv).

5.6.3 Proof of Theorem 3

Proposition 1 below is a key argument needed to prove Theorem 3 under second order regular variation. Lemmas 1, 2, and 3 are preparations for its proof.

Lemma 1 states, in a modified form, some results established in the proof of Theorem 3.2 of [88]. Its proof basically follows the arguments on pp. 150, 151 of [88], so we do not present it.

LEMMA 1. *Under Assumptions 2 and 7, we have*

$$\lim_{t\to\infty}\int_0^{t/2} t\{(1-u/t)^{-\alpha}-1\}dF_\varepsilon(u) = \lim_{t\to\infty}\int_0^{\infty} t\{(1-u/t)^{-\alpha}-1\}dF_\varepsilon(u) = \alpha E[\varepsilon I_{\varepsilon\ge 0}],$$

$$\lim_{t\to\infty}\int_{-t/2}^{0} t\{(1-u/t)^{-\alpha}-1\}dF_\varepsilon(u) = \lim_{t\to\infty}\int_{-\infty}^{0} t\{(1-u/t)^{-\alpha}-1\}dF_\varepsilon(u) = \alpha E[\varepsilon I_{\varepsilon\le 0}],$$

$$\lim_{t\to\infty}\int_0^{t/2}\frac{\bar{F}_X(t(1-u/t)) - (1-u/t)^{-\alpha}\bar{F}_X(t)}{\bar{F}_X(t)g(t)}dF_\varepsilon(u) = 0,$$

and

$$\lim_{t\to\infty}\int_{-t/2}^{0}\frac{\bar{F}_X(t(1-u/t)) - (1-u/t)^{-\alpha}\bar{F}_X(t)}{\bar{F}_X(t)g(t)}dF_\varepsilon(u) = 0.$$

LEMMA 2. *Under Assumptions 2 and 7,*

$$\lim_{t\to\infty}\frac{P(|\varepsilon|>t)}{\bar{F}_X(t)g(t)} = 0.$$

Proof. Since $\bar{F}_X(\cdot)g(\cdot) \in RV_{-\alpha+\rho}$, $\bar{F}_X(t)g(t) = t^{-\alpha+\rho}L(t)$, where L is a slowly varying function. Observe that

$$\frac{P(|\varepsilon| > t)}{\bar{F}_X(t)g(t)} = \frac{P(|\varepsilon| > t)}{t^{-\kappa}} \frac{t^{-\kappa}}{\bar{F}_X(t)g(t)} = \frac{P(|\varepsilon| > t)}{t^{-\kappa}} \frac{1}{t^{-\alpha+\rho+\kappa}L(t)}.$$

By Assumption 7, $\lim_{t\to\infty} P(|\varepsilon| > t)/t^{-\kappa} = 0$. Also, by Proposition 2.6 (i) of [4], $t^{-\alpha+\rho+\kappa}L(t) \in RV_{-\alpha+\rho+\kappa}$ goes to ∞, as $t \to \infty$.

□

LEMMA 3. *Under Assumptions 2 and 7,*

$$\lim_{t\to\infty} \int_0^{t/2} \frac{\bar{F}_\varepsilon(t-u) - \bar{F}_\varepsilon(t)}{\bar{F}_X(t)g(t)} dF_X(u) = 0.$$

Proof. Set $q_t(u) = \bar{F}_\varepsilon(t(1-u/t))/(\bar{F}_X(t)g(t))$. We want to show

$$\lim_{t\to\infty} \int_0^{t/2} q_t(u)dF_X(u) = 0 \tag{5.11}$$

so that we conclude the claim. By Assumption 2, 7,

$$\forall\, \frac{1}{2} \leq x \leq 1, \quad \frac{\bar{F}_\varepsilon(tx)}{\bar{F}_X(t)g(t)} = \frac{\bar{F}_\varepsilon(tx)}{\bar{F}_X(tx)g(tx)} \frac{\bar{F}_X(tx)g(tx)}{\bar{F}_X(t)g(t)} \to 0, \quad \text{as } t \to \infty.$$

By Proposition B.1.9 of [2], we get an upper bound such that $\forall\, \eta > 0$, $\exists\, t_0, c > 0$, $\forall\, t \geq t_0$, $\forall\, \frac{1}{2} \leq x \leq 1$,

$$\frac{\bar{F}_\varepsilon(tx)}{\bar{F}_X(tx)g(tx)} \frac{\bar{F}_X(tx)g(tx)}{\bar{F}_X(t)g(t)} \leq cx^{-\alpha+\rho-\eta}.$$

Since $cx^{-\alpha+\rho-\eta} \leq c2^{\alpha-\rho+\eta}$, we get (5.11).

□

In the following proposition we investigate the asymptotic behavior of $\bar{F}_Y$. Its proof is motivated by Theorem 3.2 of [88], which describes how two i.i.d. second–order regularly varying variables behave under convolution. We study in Proposition 1 the convolution of a second–order regularly varying variable and an error. The behavior of the convolution depends on whether or not $\lim_{t\to\infty} tg(t)$ is finite. By Proposition 2.6 (i) in [4], if $\rho > -1$, then $\lim_{t\to\infty} tg(t) = \infty$, and if $\rho < -1$, then $\lim_{t\to\infty} tg(t) = 0$. For $\rho = -1$, $\lim_{t\to\infty} tg(t)$ can be finite or infinity.

PROPOSITION 1. *Under Assumptions 2 and 7, there exist two functions $\tilde{g}$ and $\tilde{H}$ such that*

$$\lim_{t\to\infty} \frac{1}{\tilde{g}(t)} \left(\frac{\bar{F}_Y(tx)}{\bar{F}_X(t)} - x^{-\alpha} \right) = \tilde{H}(x),$$

for $x > 0$.

(i) if $\rho \geq -1$ and $\lim_{t\to\infty} tg(t) = \infty$, then $\tilde{g}(t) = g(t)$ and $\tilde{H}(x) = H(x)$.

(ii) if $\rho \leq -1$ and $l = \lim_{t\to\infty} tg(t) < \infty$, then $\tilde{g}(t) = t^{-1}$ and $\tilde{H}(x) = lH(x) + \alpha x^{-\alpha-1}E[\varepsilon]$.

Proof. Use the decomposition

$$\begin{aligned} P(X+\varepsilon > t) &= P(X+\varepsilon > t, X \vee \varepsilon > t) + P(X+\varepsilon > t, X \vee \varepsilon \leq t) \\ &=: P_1(t) + P_2(t). \end{aligned}$$

First, considering $P_1(t)$, we obtain

$$P_1(t) = P(X > t) + P(\varepsilon > t) - P(X > t, \varepsilon > t) - \int_{-\infty}^{0} (\bar{F}_X(t-u) - \bar{F}_X(t)) dF_\varepsilon(u).$$

Observe that

$$\begin{aligned} \frac{P_1(tx) - x^{-\alpha}\bar{F}_X(t)}{\bar{F}_X(t)\tilde{g}(t)} &= \frac{g(t)}{\tilde{g}(t)} \Bigg[\frac{P(X > tx) - x^{-\alpha}\bar{F}_X(t)}{\bar{F}_X(t)g(t)} + \frac{g(tx)}{g(t)} \frac{\bar{F}_X(tx)}{\bar{F}_X(t)} \frac{P(\varepsilon > tx)}{\bar{F}_X(tx)g(tx)} \\ &\quad - \frac{g(tx)}{g(t)} \frac{P(X > tx)}{\bar{F}_X(t)} \frac{P(\varepsilon > tx)}{g(tx)} \\ &\quad + \frac{\bar{F}_X(tx)}{\bar{F}_X(t)} \frac{g(tx)}{g(t)} \int_{-\infty}^{0} \frac{\bar{F}_X(tx(1-u/(tx))) - (1-u/(tx))^{-\alpha}\bar{F}_X(tx)}{\bar{F}_X(tx)g(tx)} dF_\varepsilon(u) \Bigg] \\ &\quad + \frac{1}{t\tilde{g}(t)} \left[\frac{\bar{F}_X(tx)}{x\bar{F}_X(t)} \int_{-\infty}^{0} tx\{(1-u/(tx))^{-\alpha} - 1\} dF_\varepsilon(u) \right] \\ &=: \frac{g(t)}{\tilde{g}(t)} P_{11}(tx) + \frac{1}{t\tilde{g}(t)} P_{12}(tx). \end{aligned}$$

Now, we use another decomposition for $P_2(t)$ such that

$$
\begin{aligned}
P_2(t) =& P(X+\varepsilon > t, X \vee \varepsilon \le t, X \wedge \varepsilon \le t/2) + P(X+\varepsilon > t, X \vee \varepsilon \le t, X \wedge \varepsilon \ge t/2) \\
=& \int_0^{t/2} (\bar{F}_X(t-u) - \bar{F}_X(t)) dF_\varepsilon(u) + \int_0^{t/2} (\bar{F}_\varepsilon(t-u) - \bar{F}_\varepsilon(t)) dF_X(u) \\
& + (\bar{F}_X(t/2) - \bar{F}_X(t))(\bar{F}_\varepsilon(t/2) - \bar{F}_\varepsilon(t)).
\end{aligned}
$$

Then,

$$
\begin{aligned}
\frac{P_2(tx)}{\bar{F}_X(t)\tilde{g}(t)} =& \frac{g(t)}{\tilde{g}(t)} \left[\frac{\bar{F}_X(tx)}{\bar{F}_X(t)} \frac{g(tx)}{g(t)} \int_0^{tx/2} \frac{\bar{F}_X(tx(1-u/(tx))) - (1-u/(tx))^{-\alpha}\bar{F}_X(tx)}{\bar{F}_X(tx)g(tx)} dF_\varepsilon(u) \right. \\
& \left. + \frac{\bar{F}_X(tx)}{\bar{F}_X(t)} \frac{g(tx)}{g(t)} \int_0^{tx/2} \frac{\bar{F}_\varepsilon(tx-u) - \bar{F}_\varepsilon(tx)}{\bar{F}_X(tx)g(tx)} dF_X(u) \right] \\
& + \frac{1}{t\tilde{g}(t)} \left[\frac{\bar{F}_X(tx)}{x\bar{F}_X(t)} \int_0^{tx/2} tx\{(1-u/(tx))^{-\alpha} - 1\} dF_\varepsilon(u) \right] \\
& + \frac{\bar{F}_X(t)}{\tilde{g}(t)} \left[\frac{\bar{F}_X(tx/2) - \bar{F}_X(tx)}{\bar{F}_X(t)} \frac{\bar{F}_\varepsilon(tx/2) - \bar{F}_\varepsilon(tx)}{\bar{F}_X(t)} \right] \\
=:& \frac{g(t)}{\tilde{g}(t)} P_{21}(tx) + \frac{1}{t\tilde{g}(t)} P_{22}(tx) + \frac{\bar{F}_X(t)}{\tilde{g}(t)} P_{23}(tx).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
& \frac{P(X+\varepsilon > tx) - x^{-\alpha}\bar{F}_X(t)}{\bar{F}_X(t)\tilde{g}(t)} \\
&= \frac{P_1(tx) - x^{-\alpha}\bar{F}_X(t)}{\bar{F}_X(t)\tilde{g}(t)} + \frac{P_2(tx)}{\bar{F}_X(t)\tilde{g}(t)} \\
&= \frac{g(t)}{\tilde{g}(t)} P_{11}(tx) + \frac{1}{t\tilde{g}(t)} P_{12}(tx) + \frac{g(t)}{\tilde{g}(t)} P_{21}(tx) + \frac{1}{t\tilde{g}(t)} P_{22}(tx) + \frac{\bar{F}_X(t)}{\tilde{g}(t)} P_{23}(tx).
\end{aligned}
$$

By Assumption 2, and Lemmas 1, 2, 3, $\lim_{t\to\infty} P_{11}(tx) = H(x)$, $\lim_{t\to\infty} P_{21}(tx) = \lim_{t\to\infty} P_{23}(tx) = 0$, $\lim_{t\to\infty} P_{12}(tx) = \alpha x^{-\alpha-1} E[\varepsilon I_{\varepsilon\le 0}]$, and $\lim_{t\to\infty} P_{22}(tx) = \alpha x^{-\alpha-1} E[\varepsilon I_{\varepsilon\ge 0}]$. Thus, we conclude the proof with the choices of $\tilde{H}(x)$ and $\tilde{g}(x)$.

□

LEMMA 4. *Under Assumptions 2, 4, and 7,*

$$\lim_{n \to \infty} \sqrt{k} \left(\frac{n}{k} \bar{F}_Y(b(n/k)y) - y^{-\alpha} \right) = 0, \tag{5.12}$$

locally uniformly in $(0, \infty)$ *and for any* $\beta > 1 - \alpha/2$,

$$\lim_{n \to \infty} \sqrt{k} \int_1^\infty \left(\frac{n}{k} \bar{F}_Y(b(n/k)s) - s^{-\alpha} \right) \frac{ds}{s^\beta} = 0. \tag{5.13}$$

Proof. Observe that

$$\begin{aligned}
&\sqrt{k} \left(\frac{n}{k} \bar{F}_Y(b(n/k)y) - y^{-\alpha} \right) \\
&= \sqrt{k} \left(\frac{n}{k} \bar{F}_X(b(n/k)y) - y^{-\alpha} \right) + \sqrt{k} \left(\frac{n}{k} \bar{F}_Y(b(n/k)y) - \frac{n}{k} \bar{F}_X(b(n/k)y) \right).
\end{aligned}$$

Since the local uniform convergence of the first part holds by Lemma 10, it suffices to show that for any $0 < s_1 \leq s_2$,

$$\sup_{s_1 \leq t \leq s_2} \sqrt{k} \left| \frac{n}{k} \bar{F}_Y(b(n/k)t) - \frac{n}{k} \bar{F}_X(b(n/k)t) \right| \to 0. \tag{5.14}$$

Then, we get (5.12) by Lemma 5 (i). To show (5.14), first observe that for $y > 0$,

$$\begin{aligned}
|\bar{F}_Y(y) - \bar{F}_X(y)| &= |P(X + \varepsilon > y) - P(X > y)| \\
&\leq P(\varepsilon > y/2) + \int_0^{y/2} \bar{F}_X(y-u) - \bar{F}_X(y) dF_\varepsilon(u).
\end{aligned}$$

Then we obtain

$$\begin{aligned}
&\sup_{s_1 \le t \le s_2} \sqrt{k} \left| \frac{n}{k} \bar{F}_Y(b(n/k)t) - \frac{n}{k} \bar{F}_X(b(n/k)t) \right| \\
&\qquad \le \sup_{s_1 \le t \le s_2} \sqrt{k} g(b(n/k)) \left\{ \frac{P(\varepsilon > b(n/k)t/2)}{\bar{F}_X(b(n/k)) g(b(n/k))} \right. \\
&\qquad \left. + \int_0^{b(n/k)t/2} \frac{\bar{F}_X(b(n/k)t - u) - \bar{F}_X(b(n/k)t)}{\bar{F}_X(b(n/k)) g(b(n/k))} dF_\varepsilon(u) \right\} \to 0,
\end{aligned}$$

by Assumptions 2, 4, 7, and Lemma 1.

Next, for (5.13), by Lemma 1,

$$\frac{n}{k} \bar{F}_Y(b(n/k)y) - y^{-\alpha} \sim \tilde{g}(b(n/k)) \tilde{H}(y)$$

as (5.2), for $y \ge 1$. To conclude that

$$\lim_{n \to \infty} \sqrt{k} \int_1^\infty \left(\frac{n}{k} \bar{F}_Y(b(n/k)s) - s^{-\alpha} \right) \frac{ds}{s^\beta} = \lim_{n \to \infty} \sqrt{k} \tilde{g}(b(n/k)) \int_1^\infty \tilde{H}(s) \frac{ds}{s^\beta} = 0,$$

we must find a function q such that for $t > t_0$,

$$\frac{\bar{F}_Y(b(t)y)}{\bar{F}_X(b(t))} y^{-\beta} - y^{-\alpha-\beta} \le q(y) \quad \text{and} \quad \int_1^\infty q(y) dy < \infty.$$

By Assumption 2, Lemma 9 (ii), and Potter bounds, $\forall\, \delta, c > 0$, $\exists\, t_0$, $\forall\, t \ge t_0$, $\forall\, y \ge 1$,

$$\frac{\bar{F}_Y(b(t)y)}{\bar{F}_X(b(t)y)} \frac{\bar{F}_X(b(t)y)}{\bar{F}_X(b(t))} y^{-\beta} - y^{-\alpha-\beta} \le (1+c)(1+\delta) y^{-\alpha+\delta-\beta} - y^{-\alpha-\beta}.$$

Then $q := (1+c)(1+\delta) y^{-\alpha+\delta-\beta} - y^{-\alpha-\beta}$ is integrable if $\delta < \alpha/2$.

$\square$

LEMMA 5. *Under Assumptions 3, 5, and 8,* (5.12) *holds locally uniformly in* $[1, \infty)$ *and* (5.13) *holds.*

Proof. To prove (5.12), it suffices to show that for any $s \geq 1$,

$$\sup_{1\leq z\leq s} \sqrt{k}\left|\frac{n}{k}\bar{F}_Y(b(n/k)z) - z^{-\alpha}\right| \to 0,$$

by Lemma 2. Observe that for $y \geq 1$, $t \geq 1$,

$$\begin{aligned}
&\frac{P(X+\varepsilon > ty)}{P(X > t)} - y^{-\alpha} \\
&= \frac{\int_{-\infty}^{\infty}\int_{ty-u}^{\infty} dF_X(x)dF_\varepsilon(u)}{P(X > t)} - y^{-\alpha} \\
&= \frac{\int_{-\infty}^{ty-1}\int_{ty-u}^{\infty} dF_X(x)dF_\varepsilon(u) + \int_{ty-1}^{\infty}\int_{1}^{\infty} dF_X(x)dF_\varepsilon(u)}{P(X > t)} - y^{-\alpha} \\
&= \frac{\int_{-\infty}^{ty-1}(ty-u)^{-\alpha}dF_\varepsilon(u) + P(\varepsilon > ty-1)}{P(X > t)} - y^{-\alpha} \\
&= \int_{-\infty}^{0}\left(y-\frac{u}{t}\right)^{-\alpha} - y^{-\alpha}dF_\varepsilon(u) + \int_{0}^{ty-1}\left(y-\frac{u}{t}\right)^{-\alpha} - y^{-\alpha}dF_\varepsilon(u) + \frac{P(\varepsilon > ty-1)}{P(X > t)}.
\end{aligned}$$

By the mean value theorem, there exists $c_-(y,u,t) \in [y, y-u/t]$ and $c_+(y,u,t) \in [y-u/t, y]$ such that

$$\begin{aligned}
&\left|\frac{P(X+\varepsilon > ty)}{P(X > t)} - y^{-\alpha}\right| \\
&= \left|\int_{-\infty}^{0} \alpha c_-(y,u,t)^{-\alpha-1}\frac{u}{t}dF_\varepsilon(u) + \int_{0}^{ty-1} -\alpha c_+(y,u,t)^{-\alpha-1}\frac{u}{t}dF_\varepsilon(u) + \frac{P(\varepsilon > ty-1)}{P(X > t)}\right| \\
&\leq \frac{\alpha y^{-\alpha-1}}{t}\int_{-\infty}^{ty-1} |u|dF_\varepsilon(u) + \frac{P(\varepsilon > ty-1)}{P(X > t)}.
\end{aligned}$$

Observing

$$\begin{aligned}
\frac{P(\varepsilon > ty-1)}{P(X > t)} &= \frac{1}{t}\frac{P(\varepsilon > ty-1)}{P(X > ty-1)(ty-1)^{-1}}\frac{(y-1/t)^{-\alpha}(ty-1)^{-1}}{t^{-1}} \\
&= \frac{1}{t}(y-1/t)^{-\alpha-1}\frac{P(\varepsilon > ty-1)}{(ty-1)^{-\alpha-1}},
\end{aligned}$$

by Assumption 8, there exists t_0 such that $\forall\, t \geq t_0, \forall\, y \geq 1$,

$$(y-1/t)^{-\alpha-1}\frac{P(\varepsilon > ty-1)}{(ty-1)^{-\alpha-1}} \leq 2^{\alpha+1}y^{-\alpha-1}M, \tag{5.15}$$

where $M \in [0,\infty)$ is an upper bound satisfying $P(\varepsilon > ty-1)/(ty-1)^{-\alpha-1} \leq M, \forall\, t \geq t_0$. Therefore, for large t,

$$\left|\frac{P(X+\varepsilon > ty)}{P(X>t)} - y^{-\alpha}\right| \leq \frac{\alpha y^{-\alpha-1}}{t}E|\varepsilon| + \frac{1}{t}2^{\alpha+1}y^{-\alpha-1}M,$$

and then by Assumptions 5, we obtain

$$\sup_{1\leq z\leq s}\sqrt{k}\left|\frac{n}{k}\bar{F}_Y(b(n/k)z) - z^{-\alpha}\right| \leq \frac{\sqrt{k}}{b(n/k)}\sup_{1\leq z\leq s}\{\alpha z^{-\alpha-1}E|\varepsilon| + 2^{\alpha+1}z^{-\alpha-1}M\} \to 0.$$

To verify (5.13), set $f_t(y) = (y-1/t)^{-\alpha-1}y^{-\beta}P(\varepsilon > ty-1)/(ty-1)^{-\alpha-1}$. From (5.15), there exists t_0 such that $\forall\, t \geq t_0, \forall\, y \geq 1$,

$$f_t(y) \leq 2^{\alpha+1}y^{-\alpha-1-\beta}M.$$

Let $c := \int_1^\infty 2^{\alpha+1}y^{-\alpha-1-\beta}Mdy < \infty$, then by the dominated convergence theorem,

$$\limsup_{t\to\infty}\int_1^\infty f_t(y)dy \leq c.$$

Therefore, by Assumption 5,

$$\sqrt{k}\int_1^\infty\left|\frac{n}{k}\bar{F}_Y(b(n/k)y) - y^{-\alpha}\right|\frac{dy}{y^\beta} \leq \frac{\sqrt{k}}{b(n/k)}\int_1^\infty \alpha y^{-\alpha-1-\beta}E|\varepsilon| + f_{b(n/k)}(y)dy \to 0.$$

$\square$

LEMMA 6. *Under either Assumptions 2, 4, and 7 (2RV case) or Assumptions 3, 5, and 8 (Pareto case), for any $\tau \in \mathbb{R}$,*

$$\sqrt{k}\left(\left(\frac{Y_{(k)}}{b(n/k)}\right)^{\tau} - 1\right) \Rightarrow \frac{\tau}{\alpha} W(1).$$

Proof. Set $\gamma = 1/\alpha$ and observe

$$\begin{aligned}
&\sqrt{k}(\nu_n(y^{-\gamma}, \infty] - y) \\
&= \sqrt{k}\left(\nu_n(y^{-\gamma}, \infty] - \frac{n}{k}\bar{F}_Y(b(n/k)y^{-\gamma})\right) + \sqrt{k}\left(\frac{n}{k}\bar{F}_Y(b(n/k)y^{-\gamma}) - y\right).
\end{aligned}$$

By Lemma 5 (ii), Theorem 1, and (5.12), we obtain

$$\sqrt{k}(\nu_n(y^{-\gamma}, \infty] - y) \Rightarrow W(y^{-\gamma}),$$

for $y > 0$ under Assumptions 2, 4, and 7 (2RV case), and for $y \geq 1$ under Assumptions 3, 5, and 8 (Pareto case). Then, by Lemma 12,

$$\sqrt{k}\left(\nu_n((\cdot)^{-\gamma}, \infty]^{\leftarrow}(t) - t\right) \Rightarrow -W(t^{-\gamma}).$$

Since $\nu_n((\cdot))^{-\gamma}, \infty])^{\leftarrow}(t) = \inf\{s : \nu_n(s^{-\gamma}, \infty]) \geq t\} = \left(Y_{([kt])}/b(n/k)\right)^{-\alpha}$, we have

$$\sqrt{k}\left(\left(\frac{Y_{([kt])}}{b(n/k)}\right)^{-\alpha} - t\right) \Rightarrow -W(t^{-\gamma}).$$

By the delta method,

$$\sqrt{k}\left(\left(\frac{Y_{(k)}}{b(n/k)}\right)^{\tau} - 1\right) \Rightarrow \frac{\tau}{\alpha} W(1).$$

□

Proof of Theorem 3 : We start with the decomposition

$$
\begin{aligned}
&\sqrt{k}\left(\int_1^\infty \sum_{i=1}^n I_{Y_i/Y_{(k)}}(s,\infty]s^{-\beta}ds - \frac{1}{\alpha+\beta-1}\right)\\
&= \sqrt{k}\left(\int_1^\infty \sum_{i=1}^n I_{Y_i/Y_{(k)}}(s,\infty]s^{-\beta}ds - \int_1^\infty \frac{n}{k}\bar{F}_Y(Y_{(k)}s)s^{-\beta}ds\right)\\
&+ \sqrt{k}\left(\int_1^\infty \frac{n}{k}\bar{F}_Y(Y_{(k)}s)s^{-\beta}ds - \int_1^\infty \frac{n}{k}\bar{F}_Y(b(n/k)s)s^{-\beta}ds\right)\\
&+ \sqrt{k}\left(\int_1^\infty \frac{n}{k}\bar{F}_Y(b(n/k)s)s^{-\beta}ds - \int_1^\infty s^{-\alpha-\beta}ds\right).
\end{aligned}
$$

To establish the asymptotic normality, we verify that

$$
\sqrt{k}\left(\sum_{i=1}^n I_{Y_i/Y_{(k)}}(s,\infty]s^{-\beta}ds - \int_1^\infty \frac{n}{k}\bar{F}_Y(Y_{(k)}s)s^{-\beta}ds\right) \Rightarrow \frac{1}{\alpha}\int_0^1 W(s)s^{\frac{\beta-1}{\alpha}-1}ds; \tag{5.16}
$$

$$
\sqrt{k}\left(\int_1^\infty \frac{n}{k}\bar{F}_Y(Y_{(k)}s)s^{-\beta}ds - \int_1^\infty \frac{n}{k}\bar{F}_Y(b(n/k)s)s^{-\beta}ds\right) \Rightarrow -\frac{1}{\alpha+\beta-1}W(1); \tag{5.17}
$$

$$
\sqrt{k}\left(\int_1^\infty \frac{n}{k}\bar{F}_Y(b(n/k)s)s^{-\beta}ds - \int_1^\infty s^{-\alpha-\beta}ds\right) \to 0; \tag{5.18}
$$

$$
\int_1^\infty \frac{1}{k}\sum_{i=1}^n I_{Y_i/Y_{(k)}}(s,\infty]s^{-\beta}ds - \int_1^\infty \frac{n}{k}\bar{F}_Y(Y_{(k)}s)s^{-\beta}ds \text{ is independent of } Y_{(k)}. \tag{5.19}
$$

Convergence (5.16) follows from Theorem 2, and convergence (5.18) follows from (5.13). Since

$$
E\left[\int_1^\infty \frac{1}{k}\sum_{i=1}^n I_{Y_i/Y_{(k)}}(s,\infty]s^{-\beta}ds - \int_1^\infty \frac{n}{k}\bar{F}(Y_{(k)}s)s^{-\beta}ds \,\middle|\, Y_{(k)}\right] = 0,
$$

we get (5.19). Once we show (5.17), we can get the conclusion by Lemma 11.

To verify (5.17), first consider when $\beta = 1$ (for the Hill estimator). Observe that

$$
\sqrt{k}\int_{Y_{(k)}}^{b(n/k)} \frac{n}{k}\bar{F}_Y(s)s^{-1}ds = \sqrt{k}\int_{Y_{(k)}/b(n/k)}^1 \frac{n}{k}\bar{F}_Y(b(n/k)s)s^{-1}ds.
$$

By the mean value theorem, there exists $s(n) \in [Y_{(k)}/b(n/k), 1]$ such that

$$\begin{aligned}\sqrt{k}\int_{Y_{(k)}}^{b(n/k)} \frac{n}{k}\bar{F}_Y(s)s^{-1}ds &= \sqrt{k}\frac{n}{k}\bar{F}_Y(b(n/k)s(n))\int_{Y_{(k)}/b(n/k)}^{1} s^{-1}ds \\ &= \sqrt{k}\frac{n}{k}\bar{F}_Y(b(n/k)s(n))\left(-\log\frac{Y_{(k)}}{b(n/k)}\right).\end{aligned}$$

Since $s(n) \xrightarrow{P} 1$, $n/k\bar{F}_Y(b(n/k)s(n)) \xrightarrow{P} 1$ by Lemma 9 (iii). Consider the decomposition

$$\sqrt{k}\left(-\log\frac{Y_{(k)}}{b(n/k)}\right) = \sqrt{k}\left(1-\frac{Y_{(k)}}{b(n/k)}\right) - \sqrt{k}\left\{\log\frac{Y_{(k)}}{b(n/k)} - \left(\frac{Y_{(k)}}{b(n/k)}-1\right)\right\}.$$

From Lemma 6,

$$\sqrt{k}\left(1-\frac{Y_{(k)}}{b(n/k)}\right) \Rightarrow -\frac{1}{\alpha}W(1).$$

Observe that

$$\sqrt{k}\log\frac{Y_{(k)}}{b(n/k)} = \sqrt{k}\log\left(1+\frac{Y_{(k)}}{b(n/k)}-1\right),$$

and set $\varepsilon_n = Y_{(k)}/b(n/k) - 1$. Then, for some $c_n \in (1, 1+\varepsilon_n)$, $\log(1+\varepsilon_n) = \varepsilon_n - \frac{1}{2c_n^2}\varepsilon_n^2$. Therefore,

$$\sqrt{k}\left|\log\frac{Y_{(k)}}{b(n/k)} - \left(\frac{Y_{(k)}}{b(n/k)}-1\right)\right| = \sqrt{k}|\log(1+\varepsilon_n)-\varepsilon_n| \leq \sqrt{k}\varepsilon_n^2.$$

Since $\sqrt{k}\varepsilon_n \Rightarrow 1/\alpha W(1)$ from Lemma 6,

$$\sqrt{k}\left|\log\frac{Y_{(k)}}{b(n/k)} - \left(\frac{Y_{(k)}}{b(n/k)}-1\right)\right| \leq \frac{1}{\sqrt{k}}(\sqrt{k}\varepsilon_n)^2 \xrightarrow{P} 0.$$

We thus conclude that

$$\sqrt{k}\left(-\log\frac{Y_{(k)}}{b(n/k)}\right) \Rightarrow -\frac{1}{\alpha}W(1),$$

and prove that (5.17) holds when $\beta = 1$.

For $\beta \neq 1$ (the HME), observe that

$$\begin{aligned}
&\sqrt{k}\left(\int_1^\infty \frac{n}{k}\bar{F}_Y(Y_{(k)}s)s^{-\beta}ds - \int_1^\infty \frac{n}{k}\bar{F}_Y(b(n/k)s)s^{-\beta}ds\right) \\
&\quad = \sqrt{k}\left(\frac{Y_{(k)}}{b(n/k)}\right)^{\beta-1}\int_{Y_{(k)}/b(n/k)}^{1} \frac{n}{k}\bar{F}_Y(b(n/k)s)s^{-\beta}ds \\
&\quad + \sqrt{k}\left(\left(\frac{Y_{(k)}}{b(n/k)}\right)^{\beta-1} - 1\right)\int_1^\infty \frac{n}{k}\bar{F}_Y(b(n/k)s)s^{-\beta}ds.
\end{aligned}$$

By the mean value theorem, there exists $s(n) \in [Y_{(k)}/b(n/k), 1]$ such that

$$\begin{aligned}
&\sqrt{k}\left(\int_1^\infty \frac{n}{k}\bar{F}_Y(Y_{(k)}s)s^{-\beta}ds - \int_1^\infty \frac{n}{k}\bar{F}_Y(b(n/k)s)s^{-\beta}ds\right) \\
&\quad = \left(\frac{Y_{(k)}}{b(n/k)}\right)^{\beta-1}\frac{n}{k}\bar{F}_Y(b(n/k)s(n))\frac{1}{-\beta+1}\sqrt{k}\left(1 - \left(\frac{Y_{(k)}}{b(n/k)}\right)^{-\beta+1}\right) \\
&\quad + \sqrt{k}\left(\left(\frac{Y_{(k)}}{b(n/k)}\right)^{\beta-1} - 1\right)\int_1^\infty \frac{n}{k}\bar{F}_Y(b(n/k)s)s^{-\beta}ds.
\end{aligned}$$

Then, by Lemmas 9 (iii), 3, 6, and (5.13),

$$\begin{aligned}
&\sqrt{k}\left(\int_1^\infty \frac{n}{k}\bar{F}_Y(Y_{(k)}s)s^{-\beta}ds - \int_1^\infty \frac{n}{k}\bar{F}_Y(b(n/k)s)s^{-\beta}ds\right) \\
&\quad = \frac{1}{-\beta+1}\left(-\frac{-\beta+1}{\alpha}W(1)\right) + \frac{\beta-1}{\alpha}W(1)\frac{1}{\alpha+\beta-1} = -\frac{1}{\alpha+\beta-1}W(1).
\end{aligned}$$

Bibliography

[1] B .B. Mandelbrot and J. R. Wallis. Noah, Joseph, and operational hydrology. *Water Resour. Res.*, 4:909–918, 1968.

[2] L. de Haan and A. Ferreira. *Extreme Value Theory: an Introduction*. Springer, 2006.

[3] B. V. Gnedenko. Sur la distribution limite du terme maximum d'une série aléatoire. *The Annals of Mathematics*, 44:423–453, 1943.

[4] S. I. Resnick. *Heavy–Tail Phenomena*. Springer, 2007.

[5] M. Larsson and S. I. Resnick. Extremal dependence measure and extremogram: the regularly varying case. *Extremes*, 15:231–256, 2012.

[6] D. Bosq. *Linear Processes in Function Spaces*. Springer, 2000.

[7] J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer, 2005.

[8] L. Horváth and P. Kokoszka. *Inference for Functional Data with Applications*. Springer, 2012.

[9] T. Hsing and R. Eubank. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley, 2015.

[10] P. Kokoszka and M. Reimherr. *Introduction to Functional Data Analysis*. CRC Press, 2017.

[11] F. Yao, H-G. Müller, and J-L. Wang. Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, 33:2873–2903, 2005.

[12] R. Gabrys and P. Kokoszka. Portmanteau test of independence for functional observations. *Journal of the American Statistical Association*, 102:1338–1348, 2007.

[13] I. Berkes, R. Gabrys, L. Horváth, and P. Kokoszka. Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society (B)*, 71:927–946, 2009.

[14] L. Horváth, P. Kokoszka, and R. Reeder. Estimation of the mean of functional time series and a two sample problem. *Journal of the Royal Statistical Society (B)*, 75:103–122, 2013.

[15] L. Horváth, P. Kokoszka, and G. Rice. Testing stationarity of functional time series. *Journal of Econometrics*, 179:66–82, 2014.

[16] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin, 1997.

[17] J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of Extremes: Theory and Applications*. John Wiley & Sons, 2006.

[18] D. O. Lucca and E. Moench. The pre-FOMC announcement drift. *The Journal of Finance*, 70:329–371, 2015.

[19] D. M. Mason. Laws of large numbers for sums of extreme values. *The Annals of Probability*, 10:754–764, 1982.

[20] T. Hsing. On tail index estimation using dependent data. *The Annals of Statistics*, 19:1547–1569, 1991.

[21] S. I. Resnick and C. Stărică. Consistency of Hill's estimator for dependent data. *Journal of Applied Probabability*, 32:139–167, 1995.

[22] S. I. Resnick. *Extreme Values, Regular Variation, and Point Processes*. Springer, 1987.

[23] H. Hult and F. Lindskog. Regular variation for measures on metric spaces. *Publications de l'Institute Mathématique. Nouvelle Série*, 80(94):121–140, 2006.

[24] T. Meiguet. *Heavy Tailed Functional Time Series*. PhD thesis, Universite catholigue de Louvain, 2010.

[25] P. Kokoszka, S. Stoev, and Q. Xiong. Principal components analysis of regularly varying functions. *Bernoulli*, 25:3864–3882, 2019.

[26] P. Billingsley. *Convergence of Probability Measures; Second Edition*. Wiley, New York, 1999.

[27] D. L. Cohn. *Measure Theory*. Birkhäuser, second edition, 2013.

[28] A. W. Ledford and J. A. Tawn. Statistics for near independence in multivariate extreme values. *Biometrika*, 83:169–187, 1996.

[29] A. W. Ledford and J. A. Tawn. Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59:475–499, 1997.

[30] A. W. Ledford and J. A. Tawn. Diagnostics for dependence within time series extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65:521–543, 2003.

[31] R. A. Davis and T. Mikosch. The extremogram: A correlogram for extreme events. *Bernoulli*, 15:977–1009, 2009.

[32] S. I. Resnick. The extremal dependence measure and asymptotic independence. *Stochastic models*, 20:205–227, 2004.

[33] M. G. Genton, S. A. Padoan, and H. Sang. Multivariate max-stable spatial processes. *Biometrika*, 102:215–230, 2015.

[34] D. Cooley and E. Thibaud. Decompositions of dependence for high-dimensional extremes. *Biometrika*, 106:587–604, 2019.

[35] R. A. Davis and T. Mikosch. The sample autocorrelations of heavy-tailed processes with applications to ARCH. *The Annals of Statistics*, 26:2049–2080, 1998.

[36] R. A. Davis and S. I. Resnick. Limit theory for moving averages of random variables with regularly varying tail probabilities. *The Annals of Probability*, 13:179–195, 1985.

[37] R. A. Davis and S. I. Resnick. More limit theory for the sample correlation function of moving averages. *Stochastic Processes and the Applications*, 20:257–279, 1985.

[38] R. A. Davis and S. I. Resnick. Limit theory for the sample covariance and correlation functions of moving averages. *The Annals of Statistics*, 14:533–558, 1986.

[39] E. Haeusler and J. L. Teugels. On asymptotic normality of Hill's estimator for the exponent of regular variation. *The Annals of Statistics*, 13:743–756, 1985.

[40] S. Csörgő, P. Deheuvels, and D. Mason. Kernel estimates of the tail index of a distribution. *The Annals of Statistics*, 13:1050–1077, 1985.

[41] S. I. Resnick and C. Stărică. Asymptotic behavior of Hill's estimator for autoregressive data. *Communications in Statistics. Stochastic models*, 13:703–721, 1997.

[42] S. I. Resnick and C. Stărică. Smoothing the Hill estimator. *Advances in Applied Probability*, 29:271–293, 1997.

[43] S. I. Resnick. Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, 5:303–336, 2002.

[44] M. Kim and P. Kokoszka. Hill estimator of projections of functional data on principal components. *Statistics*, 53:699–720, 2019.

[45] P. Kokoszka, H. Miao, S. Stoev, and B. Zheng. Risk analysis of cumulative intraday return curves. *Journal of Time Series Econometrics*, 11:Issue 2 paper 3, 2018.

[46] M. Speigel. The academic analysis of the 2008 financial crisis: Round 1. *Review of Financial Studies*, 24:1773–1781, 2011.

[47] S. Claessens and N. Van Horen. The impact of the global financial crisis on banking globalization. *IMF Economic Review*, 63:868–918, 2015.

[48] P. R. Lane and G. M. Milesi-Ferretti. The external wealth of nations revisited: international financial integration in the aftermath of the global financial crisis. *IMF Economic Review*, 66:189–222, 2018.

[49] C. Stărică. Multivariate extremes for models with constant conditional correlations. *Journal of Empirical Finance*, 6:515–553, 1999.

[50] S.H. Poon, M. Rockinger, and J. Tawn. Extreme value dependence in financial markets: Diagnostics, models, and financial implications. *The Review of Financial Studies*, 17:581–610, 2004.

[51] P. Kokoszka and M. Reimherr. Predictability of shapes of intraday price curves. *The Econometrics Journal*, 16:285–308, 2013.

[52] H. Rootzén, M. Leadbetter, and L. De Haan. Tail and quantile estimation for strongly mixing stationary sequences. Technical report, Center for Stochastic Processes, Dep. Statistics, Univ. North Carolina, 1990.

[53] H. Rootzén. The tail empirical process for stationary sequences. Technical report, Department of Mathematics, Chalmers University, Sweden, 1995.

[54] R. A. Davis and S. I. Resnick. Limit theory for bilinear processes with heavy-tailed noise. *The Annals of Applied Probability*, 6(4):1191–1210, 1996.

[55] S. I. Resnick and C. Stărică. Tail index estimation for dependent data. *Annals of Applied Probability*, 8:1156–1183, 1998.

[56] D. M. Mason. A strong invariance theorem for the tail empirical process. *Ann. Inst. Henri Poincaré*, 24:491–506, 1988.

[57] T. Wang and S. I. Resnick. Consistency of Hill estimators in a linear preferential attachment model. *Extremes*, 22:1–28, 2019.

[58] R. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50:987–1007, 1982.

[59] P. Hall and L. Simar. Estimating a changepoint, boundary, or frontier in the presence of observation error. *Journal of the American Statistical Association*, 97:523–534, 2002.

[60] A. Goldenshluger and A. Tsybakov. Estimating the endpoint of a distribution in the presence of additive observation errors. *Statistics & Probability Letters*, 68:39–49, 2004.

[61] A. Kneip, L. Simar, and I. V. Keilegom. Frontier estimation in the presence of measurement error with unknown variance. *Journal of Econometrics*, 184:379–393, 2015.

[62] X. Leng, L. Peng, C. Zhou, and X. Wang. Endpoint estimation for observations with normal measurement errors. *Extremes*, 21:1–26, 2018.

[63] V. Barbu and N. Limnios. *Semi-Markov Chains and Hidden Semi-Markov Models toward Applications*. Lecture Notes in Statistics. Springer, 2008.

[64] S. Z. Yu. Hidden semi-markov models. *Artificial Intelligence*, 174:215–243, 2010.

[65] A. Bartlett and W. P. McCormick. Estimation for non-negative time series with heavy-tail innovations. *Journal of Time Series Analysis*, 34:96–115, 2013.

[66] P. Hall. Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate Analysis*, 32:177–203, 1990.

[67] V. W. Bandara, A. Pezeshki, and A. P. Jayasumana. A spatiotemporal model for internet traffic anomalies. *IET Networks*, 3:41–53, 2014.

[68] P. Kokoszka, H. Nguyen, H. Wang, and L. Yang. Statistical and probabilistic analysis of inter-arrival and waiting times of internet2 anomalies. Technical report, Colorado State University, 2019. Available at www.researchgate.net/publication/330913718.

[69] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation*. Cambridge University Press, 1987.

[70] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 2. Wiley, New York, 2nd edition, 1971.

[71] N. H. Bingham, C. M. Goldie, and E. Omey. Regularly varying probability densities. *Publications De L'Institut Mathematique*, 80:47–57, 2006.

[72] H. Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Math*, 131:207–248, 1973.

[73] W. Vervaat. On a stochastic difference equation and a representation of non-negative infinitely divisible random variables. *Adv. in Appl. Probab.*, 11:750–783, 1979.

[74] C. M. Goldie. Implicit renewal theory and tails of solutions of random equations. *The Annals of Applied Probability*, 1:126–166, 1991.

[75] L. de Haan, S. I. Resnick, H. Rootzén, and C. de Vries. Extremal behavior of solutions to a stochastic difference equation, with applications to ARCH processes. *Stochastic Processes and their Applications*, 32:213–224, 1989.

[76] J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of Extremes: Theory and Applications*. John Wiley & Sons, 2004.

[77] N. Markovich. *Nonparametric Analysis of Univariate Heavy–Tailed Data: Research and Practice*. John Wiley & Sons, 2008.

[78] P. Hall. On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44:37–42, 1982.

[79] R. A. Davis and S. I. Resnick. Tail estimates motivated by extreme value theory. *The Annals of Statistics*, 12:1467–1487, 1984.

[80] R. Kulik and P. Soulier. The tail empirical process for long memory stochastic volatility sequences. *Stochastic Processes and their Applications*, 121:109–134, 2011.

[81] III J.B. Henry. A harmonic moment tail index estimator. *Journal of Statistical Theory and Applications*, 8:141–162, 2009.

[82] J. Beran, D. Schell, and M. Stehlík. The harmonic moment tail index estimator: asymptotic distribution and robustness. *Annals of the Institute of Statistical Mathematics*, 66:193–220, 2014.

[83] M. Brilhante, M. Gomes, and D. Pestana. A simple generalisation of the Hill estimator. *Computational Statistics & Data Analysis*, 57:518–535, 2013.

[84] V. Paulauskas and M. Vaičiulis. On an improvement of Hill and some other estimators. *Lithuanian Mathematical Journal*, 53:336–355, 2013.

[85] F. Caeiro, M. Gomes, J. Beirlant, and T. de Wet. Mean–of–order p reduced–bias extreme value index estimation under a third–order framework. *Extremes*, 19:561–589, 2016.

[86] M. Kim and P. Kokoszka. Consistency of the Hill estimator for time series observed with measurement errors. *Journal of Time series Analysis*, 41:421–453, 2020.

[87] M. Matsui, T. Mikosch, and L. Tafakori. Estimation of the tail index for lattice-valued sequences. *Extremes*, 16:429–455, 2013.

[88] J. Geluk, L. de Haan, S. I. Resnick, and C. Stărică. Second-order regular variation, convolution and the central limit theorem. *Stochastic Processes and their Applications*, 69:139–159, 1997.

[89] P. Hall and A. H. Welsh. Adaptive estimates of parameters of regular variation. *The Annals of Statistics*, 13:331–341, 1985.

[90] H. Drees and E. Kaufmann. Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes and their Applications*, 75:149–172, 1998.

[91] J. Danielsson, L. de Haan, L. Peng, and C.G. de Vries. Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis*, 76, 2001.

[92] J. Danielsson, L. M. Ergun, L. de Haan, and C. G. de Vries. Tail index estimation: Quantile driven threshold selection, 2016.

[93] P. Kokoszka, H. Nguyen, H. Wang, and L. Yang. Statistical and probabilistic analysis of interarrival and waiting times of Internet2 anomalies. *Statistical Methods & Applications*, pages 1–18, 2020.

[94] J. Neyman. Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society of London*, 236:333–380, 1937.

[95] L. de Haan and J. de Ronde. Sea and wind: multivariate extremes at work. *Extremes*, 1:7–46, 1998.

[96] J. Einmahl, L. de Haan, and V. Piterbarg. Nonparametric estimation of the spectral measure of an extreme value distribution. *The Annals of Statistics*, 29:1401–1423, 2001.

Appendix A

Supplementary material of Chapter 3

A.1 General definition of regular variation and transformations to equivalent tails

Let $\mathbf{Z} = [Z^{(1)}, \ldots, Z^{(d)}]^\top$ be a $\mathbb{R}_+^d$–valued random vector. Suppose that there exist a Radon measure ν on $\mathbb{E}_d$ and sequences $\{b^{(j)}(n), n \geq 1\}$ with $\lim_{n\to\infty} b^{(j)}(n) = \infty$, for $j = 1, \ldots, d$, such that

$$n \Pr\left(\left(\frac{Z^{(j)}}{b^{(j)}(n)},\ j = 1, \ldots, d\right) \in \cdot\right) \xrightarrow{v} \nu, \quad \text{in } M_+(\mathbb{E}_d) \tag{A.1}$$

and for each $j = 1, \ldots, d$

$$n \Pr\left(\frac{Z^{(j)}}{b^{(j)}(n)} \in \cdot\right) \xrightarrow{v} \nu_{\alpha_j}, \quad \alpha_j > 0, \quad \text{in } M_+(0, \infty]. \tag{A.2}$$

The sequences $\{b^{(j)}(n)\}$, $j = 1, \ldots, d$ are defined by $\Pr(Z^{(j)} > b^{(j)}(n)) = n^{-1}$. If conditions (A.1) and (A.2) hold, we say that $\mathbf{Z}$ is regularly varying with marginal indexes $\alpha_1, \ldots, \alpha_d$.

Two methods are recommended to transform non–standard cases to the standard case with the tail index $\alpha = 1$. The first method is to assume that for $j = 1, \ldots, d$ the jth marginal tail asymptotically behaves like a Pareto tail with index α_j and then use a power transformation to make each transformed component regularly varying with $\alpha = 1$. This method is mathematically simple, but has the drawback of requiring the estimation of the marginal α_js. The uncertainty due to the estimation of the α_j can be avoided by using the ranks method, see [95], [96]. Let $\{\mathbf{Z}_i = [Z_i^{(1)}, \ldots, Z_i^{(d)}],\ 1 \leq i \leq n\}$ be a random sample of $\mathbb{R}_+^d$ random vectors with common distribution satisfying the global and marginal regular variation conditions (A.1) and (A.2). For each fixed j, define the the complementary rank of $Z_i^{(j)}$ by

$$r_i^{(j)} = \sum_{l=1}^{n} I_{Z_l^{(j)} \geq Z_i^{(j)}}, \quad j = 1, \ldots, d,$$

which is the number of jth components at least as large as $Z_i^{(j)}$. According to Proposition 9.4 of [4], a simple scaling argument for the tail empirical measure shows that

$$\frac{1}{k} \sum_{i=1}^{n} I_{\left(k/r_i^{(j)};\ j=1,\ldots,d\right)} \Rightarrow \nu_*, \quad \text{in } M_+(\mathbb{E}_d),$$

where ν_* is the limit measure for the standard case satisfying $\nu_*([\mathbf{0}, \mathbf{x}]^c) = \nu([\mathbf{0}, \mathbf{x}^{\mathbf{1}/\boldsymbol{\alpha}}]^c)$ for $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_d]$ and $\nu_*(t\cdot) = t^{-1}\nu_*(\cdot)$. Using this transformation, we can achieve the standard case, which allows estimation of the standard limit measure ν_* and the angular measure Γ associated with ν_*. The disadvantage of this method is that we cannot guarantee the sample $\{r_i = [r_{i1}, \ldots, r_{id}], 1 \leq i \leq n\}$ to be i.i.d.

A.2 Supplementary graphs and Tables

We conclude this section by presenting the results of the estimation of the EDM between the rescaled scores $\xi_j/\sqrt{\lambda_j}$. Tables A.7 and A.8 report estimates for the three pairs of the first three normalized scores, $\xi_1/\sqrt{\lambda_1}$, $\xi_2/\sqrt{\lambda_2}$, $\xi_3/\sqrt{\lambda_3}$, for Walmart and IBM, respectively. It is seen that there is not much difference in estimates compared to those for non–normalized scores in Tables 3.2 and A.1.

Table A.1: Estimates of EDM for IBM stock. Standard errors in parentheses are computed using Theorem 2.

	$\mathrm{D}^{(+,+)}(\xi_1,\xi_2)$	$\mathrm{D}^{(-,+)}(\xi_1,\xi_2)$	$\mathrm{D}^{(-,-)}(\xi_1,\xi_2)$	$\mathrm{D}^{(+,-)}(\xi_1,\xi_2)$
Before	0.07 (0.02)	-0.13 (0.03)	0.05 (0.02)	-0.08 (0.02)
During	0.07 (0.02)	-0.09 (0.02)	0.08 (0.02)	-0.08 (0.02)
After 1	0.07 (0.02)	-0.06 (0.01)	0.07 (0.02)	-0.10 (0.02)
After 2	0.07 (0.02)	-0.10 (0.02)	0.10 (0.02)	-0.06 (0.02)
	$\mathrm{D}^{(+,+)}(\xi_1,\xi_3)$	$\mathrm{D}^{(-,+)}(\xi_1,\xi_3)$	$\mathrm{D}^{(-,-)}(\xi_1,\xi_3)$	$\mathrm{D}^{(+,-)}(\xi_1,\xi_3)$
Before	0.10 (0.02)	-0.08 (0.02)	0.09 (0.02)	-0.10 (0.02)
During	0.08 (0.03)	-0.10 (0.03)	0.10 (0.03)	-0.04 (0.02)
After 1	0.06 (0.02)	-0.10 (0.02)	0.07 (0.02)	-0.07 (0.02)
After 2	0.06 (0.02)	-0.10 (0.02)	0.06 (0.02)	-0.06 (0.02)
	$\mathrm{D}^{(+,+)}(\xi_2,\xi_3)$	$\mathrm{D}^{(-,+)}(\xi_2,\xi_3)$	$\mathrm{D}^{(-,-)}(\xi_2,\xi_3)$	$\mathrm{D}^{(+,-)}(\xi_2,\xi_3)$
Before	0.09 (0.03)	-0.09 (0.03)	0.04 (0.02)	-0.07 (0.03)
During	0.13 (0.03)	-0.08 (0.02)	0.09 (0.02)	-0.08 (0.02)
After 1	0.08 (0.02)	-0.06 (0.02)	0.08 (0.02)	-0.09 (0.02)
After 2	0.06 (0.02)	-0.08 (0.02)	0.07 (0.02)	-0.09 (0.02)

Table A.2: Empirical biases (standard errors) of the estimator of the EDM for **Case 2**

n	$\mathrm{D}(\xi_1,\xi_2)$	$\mathrm{D}^{(+,+)}(\xi_1,\xi_2)$	$\mathrm{D}^{(-,+)}(\xi_1,\xi_2)$	$\mathrm{D}^{(-,-)}(\xi_1,\xi_2)$	$\mathrm{D}^{(+,-)}(\xi_1,\xi_2)$
200	-0.014 (0.10)	-0.011 (0.03)	0.006 (0.04)	-0.018 (0.03)	0.008 (0.04)
600	-0.007 (0.09)	-0.006 (0.02)	0.005 (0.04)	-0.007 (0.02)	0.001 (0.03)
1000	-0.009 (0.09)	-0.006 (0.02)	0.003 (0.03)	-0.007 (0.02)	0.002 (0.03)
n	$\mathrm{D}(\xi_1,\xi_3)$	$\mathrm{D}^{(+,+)}(\xi_1,\xi_3)$	$\mathrm{D}^{(-,+)}(\xi_1,\xi_3)$	$\mathrm{D}^{(-,-)}(\xi_1,\xi_3)$	$\mathrm{D}^{(+,-)}(\xi_1,\xi_3)$
200	-0.008 (0.05)	-0.019 (0.03)	0.014 (0.03)	-0.017 (0.03)	0.014 (0.03)
600	-0.002 (0.04)	-0.011 (0.03)	0.012 (0.03)	-0.015 (0.03)	0.012 (0.03)
1000	-0.002 (0.04)	-0.011 (0.03)	0.013 (0.03)	-0.014 (0.03)	0.010 (0.03)
n	$\mathrm{D}(\xi_2,\xi_3)$	$\mathrm{D}^{(+,+)}(\xi_2,\xi_3)$	$\mathrm{D}^{(-,+)}(\xi_2,\xi_3)$	$\mathrm{D}^{(-,-)}(\xi_2,\xi_3)$	$\mathrm{D}^{(+,-)}(\xi_2,\xi_3)$
200	-0.003 (0.07)	-0.039 (0.03)	0.040 (0.03)	-0.042 (0.03)	0.038 (0.05)
600	0.002 (0.06)	-0.031 (0.02)	0.032 (0.02)	-0.033 (0.03)	0.034 (0.04)
1000	0.001 (0.06)	-0.030 (0.02)	0.034 (0.02)	-0.031 (0.02)	0.028 (0.04)

Table A.3: Empirical biases (standard errors) of the estimator of the EDM for **Case 3** [full dependence]

n	$\mathrm{D}(\xi_1, \xi_2)$	$\mathrm{D}^{(+,+)}(\xi_1, \xi_2)$	$\mathrm{D}^{(-,+)}(\xi_1, \xi_2)$	$\mathrm{D}^{(-,-)}(\xi_1, \xi_2)$	$\mathrm{D}^{(+,-)}(\xi_1, \xi_2)$
200	0.005 (0.09)	-0.053 (0.03)	0.056 (0.03)	-0.055 (0.03)	0.057 (0.03)
600	0.005 (0.08)	-0.051 (0.02)	0.054 (0.02)	-0.053 (0.02)	0.054 (0.02)
1000	0.002 (0.07)	-0.052 (0.02)	0.053 (0.02)	-0.053 (0.02)	0.054 (0.02)
n	$\mathrm{D}(\xi_1, \xi_3)$	$\mathrm{D}^{(+,+)}(\xi_1, \xi_3)$	$\mathrm{D}^{(-,+)}(\xi_1, \xi_3)$	$\mathrm{D}^{(-,-)}(\xi_1, \xi_3)$	$\mathrm{D}^{(+,-)}(\xi_1, \xi_3)$
200	-0.001 (0.05)	-0.081 (0.02)	0.081 (0.02)	-0.081 (0.02)	0.081 (0.07)
600	0.001 (0.04)	-0.079 (0.01)	0.079 (0.01)	-0.079 (0.01)	0.080 (0.06)
1000	0.000 (0.03)	-0.079 (0.01)	0.079 (0.01)	-0.079 (0.01)	0.079 (0.06)
n	$\mathrm{D}(\xi_2, \xi_3)$	$\mathrm{D}^{(+,+)}(\xi_2, \xi_3)$	$\mathrm{D}^{(-,+)}(\xi_2, \xi_3)$	$\mathrm{D}^{(-,-)}(\xi_2, \xi_3)$	$\mathrm{D}^{(+,-)}(\xi_2, \xi_3)$
200	-0.002 (0.07)	-0.037 (0.03)	0.037 (0.03)	-0.037 (0.03)	0.035 (0.03)
600	0.002 (0.06)	-0.029 (0.03)	0.030 (0.03)	-0.030 (0.03)	0.030 (0.02)
1000	-0.001 (0.06)	-0.028 (0.02)	0.028 (0.02)	-0.029 (0.02)	0.028 (0.02)

Table A.4: Estimated standard errors, computed by (3.1), of the estimator of the EDM for **Case 1** [Independence]

n	$\mathrm{D}^{(+,+)}(\xi_1, \xi_2)$	$\mathrm{D}^{(-,+)}(\xi_1, \xi_2)$	$\mathrm{D}^{(-,-)}(\xi_1, \xi_2)$	$\mathrm{D}^{(+,-)}(\xi_1, \xi_2)$
200	0.01	0.01	0.01	0.01
600	0.01	0.01	0.01	0.01
1000	0.01	0.01	0.01	0.01
n	$\mathrm{D}^{(+,+)}(\xi_1, \xi_3)$	$\mathrm{D}^{(-,+)}(\xi_1, \xi_3)$	$\mathrm{D}^{(-,-)}(\xi_1, \xi_3)$	$\mathrm{D}^{(+,-)}(\xi_1, \xi_3)$
200	0.01	0.01	0.01	0.01
600	0.01	0.01	0.01	0.01
1000	0.01	0.01	0.01	0.01
n	$\mathrm{D}^{(+,+)}(\xi_2, \xi_3)$	$\mathrm{D}^{(-,+)}(\xi_2, \xi_3)$	$\mathrm{D}^{(-,-)}(\xi_2, \xi_3)$	$\mathrm{D}^{(+,-)}(\xi_2, \xi_3)$
200	0.01	0.01	0.01	0.01
600	0.01	0.01	0.01	0.01
1000	0.01	0.01	0.01	0.01

Table A.5: Estimated standard errors, computed by (3.1), of the estimator of the EDM for **Case 2**

n	$\mathrm{D}^{(+,+)}(\xi_1,\xi_2)$	$\mathrm{D}^{(-,+)}(\xi_1,\xi_2)$	$\mathrm{D}^{(-,-)}(\xi_1,\xi_2)$	$\mathrm{D}^{(+,-)}(\xi_1,\xi_2)$
200	0.03	0.02	0.02	0.02
600	0.02	0.02	0.02	0.02
1000	0.02	0.01	0.02	0.01
n	$\mathrm{D}^{(+,+)}(\xi_1,\xi_3)$	$\mathrm{D}^{(-,+)}(\xi_1,\xi_3)$	$\mathrm{D}^{(-,-)}(\xi_1,\xi_3)$	$\mathrm{D}^{(+,-)}(\xi_1,\xi_3)$
200	0.02	0.02	0.02	0.02
600	0.02	0.02	0.02	0.02
1000	0.01	0.01	0.01	0.01
n	$\mathrm{D}^{(+,+)}(\xi_2,\xi_3)$	$\mathrm{D}^{(-,+)}(\xi_2,\xi_3)$	$\mathrm{D}^{(-,-)}(\xi_2,\xi_3)$	$\mathrm{D}^{(+,-)}(\xi_2,\xi_3)$
200	0.03	0.02	0.03	0.02
600	0.02	0.02	0.02	0.02
1000	0.02	0.01	0.02	0.01

Table A.6: Estimated standard errors, computed by (3.1), of the EDM for **Case 3** [full dependence]

n	$\mathrm{D}^{(+,+)}(\xi_1,\xi_2)$	$\mathrm{D}^{(-,+)}(\xi_1,\xi_2)$	$\mathrm{D}^{(-,-)}(\xi_1,\xi_2)$	$\mathrm{D}^{(+,-)}(\xi_1,\xi_2)$
200	0.03	0.03	0.03	0.03
600	0.02	0.02	0.02	0.02
1000	0.01	0.01	0.01	0.01
n	$\mathrm{D}^{(+,+)}(\xi_1,\xi_3)$	$\mathrm{D}^{(-,+)}(\xi_1,\xi_3)$	$\mathrm{D}^{(-,-)}(\xi_1,\xi_3)$	$\mathrm{D}^{(+,-)}(\xi_1,\xi_3)$
200	0.02	0.02	0.02	0.02
600	0.01	0.01	0.01	0.01
1000	0.01	0.01	0.01	0.01
n	$\mathrm{D}^{(+,+)}(\xi_2,\xi_3)$	$\mathrm{D}^{(-,+)}(\xi_2,\xi_3)$	$\mathrm{D}^{(-,-)}(\xi_2,\xi_3)$	$\mathrm{D}^{(+,-)}(\xi_2,\xi_3)$
200	0.03	0.03	0.03	0.03
600	0.03	0.03	0.03	0.03
1000	0.02	0.02	0.02	0.02

Table A.7: Estimates of EDM for Walmart stock. Standard errors in parentheses are computed using Theorem 2.

	$\mathrm{D}^{(+,+)}(\xi_1,\xi_2)$	$\mathrm{D}^{(-,+)}(\xi_1,\xi_2)$	$\mathrm{D}^{(-,-)}(\xi_1,\xi_2)$	$\mathrm{D}^{(+,-)}(\xi_1,\xi_2)$
Before	0.06 (0.02)	-0.06 (0.02)	0.07 (0.02)	-0.10 (0.02)
During	0.11 (0.02)	-0.06 (0.01)	0.08 (0.02)	-0.06 (0.01)
After 1	0.11 (0.03)	-0.08 (0.02)	0.07 (0.02)	-0.06 (0.02)
After 2	0.11 (0.04)	-0.07 (0.03)	0.04 (0.02)	-0.11 (0.04)
	$\mathrm{D}^{(+,+)}(\xi_1,\xi_3)$	$\mathrm{D}^{(-,+)}(\xi_1,\xi_3)$	$\mathrm{D}^{(-,-)}(\xi_1,\xi_3)$	$\mathrm{D}^{(+,-)}(\xi_1,\xi_3)$
Before	0.08 (0.02)	-0.06 (0.02)	0.08 (0.02)	-0.06 (0.02)
During	0.08 (0.02)	-0.07 (0.02)	0.05 (0.02)	-0.08 (0.02)
After 1	0.05 (0.02)	-0.11 (0.03)	0.08 (0.03)	-0.07 (0.03)
After 2	0.08 (0.03)	-0.07 (0.03)	0.07 (0.03)	-0.05 (0.02)
	$\mathrm{D}^{(+,+)}(\xi_2,\xi_3)$	$\mathrm{D}^{(-,+)}(\xi_2,\xi_3)$	$\mathrm{D}^{(-,-)}(\xi_2,\xi_3)$	$\mathrm{D}^{(+,-)}(\xi_2,\xi_3)$
Before	0.08 (0.03)	-0.09 (0.03)	0.05 (0.02)	-0.05 (0.02)
During	0.09 (0.02)	-0.08 (0.02)	0.07 (0.02)	-0.07 (0.02)
After 1	0.10 (0.02)	-0.08 (0.02)	0.09 (0.02)	-0.09 (0.02)
After 2	0.07 (0.02)	-0.08 (0.02)	0.08 (0.02)	-0.06 (0.02)

Table A.8: Estimates of EDM for IBM stock. Standard errors in parentheses are computed using Theorem 2.

	$\mathrm{D}^{(+,+)}(\xi_1,\xi_2)$	$\mathrm{D}^{(-,+)}(\xi_1,\xi_2)$	$\mathrm{D}^{(-,-)}(\xi_1,\xi_2)$	$\mathrm{D}^{(+,-)}(\xi_1,\xi_2)$
Before	0.06 (0.02)	-0.13 (0.02)	0.06 (0.02)	-0.07 (0.02)
During	0.06 (0.02)	-0.07 (0.03)	0.10 (0.02)	-0.07 (0.02)
After 1	0.05 (0.02)	-0.05 (0.02)	0.08 (0.03)	-0.09 (0.03)
After 2	0.06 (0.02)	-0.10 (0.02)	0.10 (0.02)	-0.06 (0.02)
	$\mathrm{D}^{(+,+)}(\xi_1,\xi_3)$	$\mathrm{D}^{(-,+)}(\xi_1,\xi_3)$	$\mathrm{D}^{(-,-)}(\xi_1,\xi_3)$	$\mathrm{D}^{(+,-)}(\xi_1,\xi_3)$
Before	0.10 (0.02)	-0.08 (0.01)	0.08 (0.02)	-0.10 (0.02)
During	0.07 (0.03)	-0.10 (0.03)	0.10 (0.03)	-0.05 (0.02)
After 1	0.08 (0.02)	-0.09 (0.02)	0.06 (0.01)	-0.06 (0.01)
After 2	0.05 (0.02)	-0.12 (0.03)	0.10 (0.03)	-0.07 (0.03)
	$\mathrm{D}^{(+,+)}(\xi_2,\xi_3)$	$\mathrm{D}^{(-,+)}(\xi_2,\xi_3)$	$\mathrm{D}^{(-,-)}(\xi_2,\xi_3)$	$\mathrm{D}^{(+,-)}(\xi_2,\xi_3)$
Before	0.11 (0.03)	-0.08 (0.02)	0.05 (0.02)	-0.09 (0.02)
During	0.14 (0.03)	-0.09 (0.03)	0.09 (0.03)	-0.08 (0.03)
After 1	0.07 (0.02)	-0.07 (0.02)	0.07 (0.02)	-0.08 (0.02)
After 2	0.06 (0.02)	-0.10 (0.02)	0.06 (0.02)	-0.09 (0.02)

Appendix B

Supplementary material of Chapter 5

B.1 Tables with coverage probabilities for fixed k values

Table B.1: Proportion (in percent) of the asymptotic confidence intervals including $1/\alpha$ as a function of the number of upper order statistics, k, for $n = \mathbf{500}$ and the **Pareto** model. In the "Ratio" column, the ratio of error SD to model SD is displayed. Ratio equal to zero indicates no errors. The target coverage is 95 percent.

Ratio	Error Type	k								
		30	50	100	150	200	250	300	350	400
0		92.5	93.3	94.3	93.9	93.9	94.4	94.5	95.1	95.5
0.01	Normal	92.3	93.3	94.2	94.5	94.2	94.1	94.1	94.9	95.1
	scaled t_8	92.5	93.1	94.7	94.1	94.4	94.3	94.5	95.1	94.8
	GPD	92.4	93.5	94.5	94.1	94.0	94.3	94.1	94.9	95.4
	Uniform	92.7	93.6	94.6	94.3	94.5	93.4	94.8	95.3	95.1
0.05	Normal	91.7	93.6	93.9	94.4	92.8	93.1	93.1	94.1	96.0
	scaled t_8	91.2	94.0	95.2	94.0	93.5	93.3	92.9	94.5	96.2
	GPD	91.6	93.6	94.8	94.6	93.3	94.2	93.3	94.4	95.7
	Uniform	91.9	93.0	93.7	94.1	93.2	93.4	94.2	94.1	93.8
0.1	Normal	91.0	93.1	92.5	91.4	87.9	87.6	91.0	94.4	67.8
	scaled t_8	90.0	90.6	86.6	83.2	81.3	89.0	96.4	73.9	2.7
	GPD	90.5	92.8	92.8	91.8	91.1	92.4	94.8	95.2	88.0
	Uniform	90.4	93.1	93.7	93.3	92.1	89.8	87.5	90.7	89.3
0.2	Normal	90.3	90.2	80.9	71.4	72.9	87.9	93.4	30.5	0.0
	scaled t_8	88.5	89.3	84.1	84.2	87.0	92.4	93.9	53.1	0.6
	GPD	88.6	88.0	87.1	87.1	88.6	94.3	93.9	70.4	9.7
	Uniform	89.9	91.9	90.3	85.0	76.1	69.1	87.6	78.5	1.0
0.3	Normal	85.0	81.0	55.5	49.4	71.0	96.3	45.7	0.0	0.0
	scaled t_8	87.3	84.7	74.1	70.7	78.8	93.4	83.5	5.9	0.0
	GPD	84.6	84.0	78.8	84.1	91.5	95.2	72.4	12.2	0.0
	Uniform	88.9	88.1	80.4	59.7	44.1	70.9	90.4	5.7	0.0

Table B.2: Proportion (in percent) of the asymptotic confidence intervals including $1/\alpha$ as a function of the number of upper order statistics, k, for $n = \mathbf{2000}$ and the **Pareto** model. In the "Ratio" column, the ratio of error SD to model SD is displayed. Ratio equal to zero indicates no errors. The target coverage is 95 percent.

Ratio	Error Type	k									
		50	100	200	300	400	500	600	700	800	1500
0		90.1	92.5	94.3	94.8	95.5	94.9	95.4	95	95.4	95.9
0.01	Normal	89.9	92.0	94.0	94.6	95.3	94.9	95.5	94.9	95.4	96.0
	scaled t_8	90.0	92.5	94.4	94.3	95.4	95.1	95.0	95.0	95.7	95.6
	GPD	89.9	92.0	94.8	94.5	95.8	94.7	95.8	95.1	95.1	95.1
	Uniform	90.0	92.3	94.2	94.8	95.4	94.8	95.8	94.8	94.9	95.0
0.05	Normal	89.9	92.4	94.1	93.3	94.8	94.5	94.1	94.0	94.1	94.3
	scaled t_8	89.5	92.8	94.3	94.1	94.8	94.5	94.3	94.4	93.4	94.2
	GPD	89.7	91.8	93.9	94.7	94.8	94.4	94.7	93.3	94.1	95.6
	Uniform	90.2	91.8	94.5	94.0	94.7	93.9	95.2	94.4	94.0	90.7
0.1	Normal	90.1	91.9	94.6	93.5	93.2	92.6	91.7	90.6	87.4	90.9
	scaled t_8	90.3	91.6	93.4	93.3	92.2	91.7	88.6	86.5	83.0	73.7
	GPD	89.6	92.5	93.1	91.6	90.4	88.2	87.2	85.6	86.3	64.7
	Uniform	89.8	90.5	93.5	92.2	89.5	86.2	81.9	74.0	66.5	92.1
0.2	Normal	89.8	91.5	90.4	86.8	76.9	66.1	56.5	47.0	45.2	0.0
	scaled t_8	90.6	90.9	86.8	80.2	68.0	55.0	48.1	46.5	50.7	0.0
	GPD	88.4	88.1	84.6	75.3	68.9	65.7	68.2	74.6	84.1	0.0
	Uniform	88.5	89.2	85.1	71.4	49.2	25.7	9.0	2.6	1.0	0.0
0.3	Normal	89.1	88.6	82.9	67.8	56.9	50.7	49.4	54.1	64.6	0.0
	scaled t_8	88.3	86.1	70.1	43.7	26.9	19.0	18.3	27.0	46.8	0.0
	GPD	85.9	79.9	63.6	51.3	45.8	52.1	63.7	79.4	93.5	0.0
	Uniform	88.5	85.0	61.4	22.7	3.3	0.1	0.0	0.0	0.5	0.0

Table B.3: Proportion (in percent) of the asymptotic confidence intervals including $1/\alpha$ as a function of the number of upper order statistics, k, for $n = \mathbf{500}$ and the **2RV** model. In the "Ratio" column, the ratio of error SD to model SD is displayed. Ratio equal to zero indicates no errors. The target coverage is 95 percent.

Ratio	Error Type	k								
		30	50	100	150	200	250	300	350	400
0		92.2	92.0	85.8	59.1	23.0	4.9	0.1	0.1	0.0
0.01	Normal	92.0	92.3	85.4	59.1	21.7	4.6	0.2	0.1	0.0
	scaled t_8	92.2	92.4	86.0	59.3	22.6	4.9	0.1	0.1	0.0
	GPD	92.3	92.4	85.9	58.4	22.6	4.7	0.2	0.0	0.0
	Uniform	92.1	92.1	85.9	58.7	22.3	4.9	0.2	0.0	0.0
0.05	Normal	90.8	91.2	84.6	53.2	17.8	2.5	0.2	0.0	0.0
	scaled t_8	90.9	93.2	84.4	55.4	19.6	3.2	0.1	0.0	0.0
	GPD	91.0	92.3	85.2	53.9	18.7	3.6	0.2	0.1	0.0
	Uniform	91.6	92.5	84.4	53.2	17.6	2.4	0.1	0.1	0.0
0.1	Normal	89.9	91.0	79.0	38.1	10.0	1.4	0.1	0.2	0.4
	scaled t_8	91.0	91.6	81.5	44.8	13.3	1.9	0.1	0.0	0.0
	GPD	90.2	91.5	78.8	45.2	14.2	3.2	0.1	0.1	0.0
	Uniform	91.5	91.4	79.8	39.4	8.1	0.6	0.2	0.4	2.5
0.2	Normal	89.0	86.1	49.8	14.0	4.2	2.8	7.3	42.3	91.0
	scaled t_8	89.4	88.1	63.4	23.9	7.0	2.1	1.5	4.9	49.2
	GPD	88.2	85.0	64.2	33.5	14.6	6.8	3.7	3.9	19.6
	Uniform	90.2	87.3	55.2	9.3	0.8	0.6	9.0	79.1	38.9
0.3	Normal	85.0	74.3	26.3	6.9	5.8	17.8	74.4	75.5	0.0
	scaled t_8	86.5	82.1	43.5	15.9	7.4	7.7	22.4	80.4	51.3
	GPD	83.3	79.0	53.3	32.0	21.3	20.4	29.3	58.5	92.9
	Uniform	84.7	79.5	21.9	1.1	0.2	9.9	83.8	20.5	0.0

Table B.4: Proportion (in percent) of the asymptotic confidence intervals including $1/\alpha$ as a function of the number of upper order statistics, k, for $n = \mathbf{2000}$ and the **2RV** model. In the "Ratio" column, the ratio of error SD to model SD is displayed. Ratio equal to zero indicates no errors. The target coverage is 95 percent.

Ratio	Error Type	k									
		50	100	200	300	400	500	600	700	800	1500
0		90.1	92.4	92.6	88.1	73.0	44.4	16.2	2.1	0.1	0.0
0.01	Normal	89.9	91.7	92.7	88.2	73.6	44.2	16.5	2.2	0.1	0.0
	scaled t_8	90.0	92.5	93.1	88.0	73.2	43.8	16.3	2.3	0.1	0.0
	GPD	89.9	91.9	92.5	87.8	72.9	44.0	16.0	2.0	0.1	0.0
	Uniform	90.0	92.2	92.3	88.5	72.7	44.2	16.0	1.8	0.1	0.0
0.05	Normal	90.2	91.8	92.3	86.5	68.2	35.4	9.1	0.8	0.0	0.0
	scaled t_8	90.0	92.2	92.7	86.2	67.7	34.7	8.8	0.6	0.0	0.0
	GPD	89.9	91.9	91.6	86.0	68.3	35.9	9.7	0.8	0.0	0.0
	Uniform	90.7	91.1	92.0	86.5	67.3	34.5	9.4	0.5	0.0	0.0
0.1	Normal	90.4	91.2	91.3	78.5	45.3	11.6	0.7	0.0	0.0	0.0
	scaled t_8	90.4	91.5	91.5	81.9	56.0	20.5	2.1	0.1	0.0	0.0
	GPD	89.6	92.3	90.0	78.8	49.9	18.3	3.0	0.2	0.0	0.0
	Uniform	89.2	90.6	91.9	78.7	50.1	13.6	0.7	0.0	0.0	0.0
0.2	Normal	90.1	88.9	74.2	27.0	2.7	0.0	0.0	0.0	0.0	74.2
	scaled t_8	90.5	90.3	81.1	48.4	13.4	1.8	0.0	0.0	0.0	8.5
	GPD	88.2	87.3	72.9	44.0	16.7	3.8	0.7	0.1	0.0	0.1
	Uniform	89.0	88.4	77.8	39.5	3.3	0.0	0.0	0.0	0.0	0.4
0.3	Normal	88.2	82.3	30.6	1.6	0.1	0.0	0.0	0.0	0.0	0.0
	scaled t_8	88.2	85.2	54.0	13.8	0.7	0.0	0.0	0.0	0.0	3.9
	GPD	85.8	78.3	47.9	19.0	6.0	2.9	1.0	0.4	0.3	87.4
	Uniform	87.1	83.2	47.9	3.5	0.0	0.0	0.0	0.0	0.0	0.0