On Extreme Value Approximation to Tails of Distribution Functions

ISBN 90 5170 805 X

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. **342** of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.
(end of this page)

On Extreme Value Approximation to Tails of Distribution Functions

Over extreme waarden benadering van staarten van verdelingsfunkties

PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Erasmus Universiteit Rotterdam

op gezag van de Rector Magnificus,

Prof.dr. S.W.J. Lamberts

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

donderdag 14 oktober 2004 om 16.00 uur

door

Deyuan Li

geboren te Sichuan, P.R.China.

Promotiecommissie

Promotor:	Prof.dr. C.G. de Vries
Overige leden:	Prof.dr. J.H.J. Einmahl Prof.dr. P. Groeneboom Prof.dr. L.F.M. de Haan

Acknowledgements

First I am very grateful to my supervisors, Prof. Laurens de Haan and Prof. Casper de Vries, for their encouragement, guidance, patience and continuous support during the last four years. It is my pleasure to have the opportunity of sharing their knowledge and wisdom. Without their kind help, I can not see the completion of this thesis on time.

I would like to express my gratitude to Prof. Shihong Cheng for his introducing me to Prof. Laurens de Haan.

I would like to thank my co-authors: Prof. Holger Drees, Prof. John Einmahl, Prof. Helena Pereira, Dr. Isabel Barao and Dr. Liang Peng for their cooperation. I am very happy to work with them.

I would like to thank Dr. Jaap Geluk, Dr. Alex Koing, Ms. Olga Gilissen, Ms. Tineke Kurtz, Dr. Tao Lin and Dr. Ana Ferreira for their help in various aspects. Also I would like to thank all the staff in Tinbergen Institute and Econometric Institute for their kindness.

I would like to thank my Chinese friends in Rotterdam. We spend the most colorful time in our lives together. It is a good memory.

Finally I am deeply indebted to my wife, Lei Wang, and my parents for their understanding, confidence and support.

Contents

Chapter 1

Introduction

This thesis is devoted to statistics in extreme value theory, where the estimation of quantities related to extreme events is of particular interest. For example in the design of dikes, a typical requirement is that the sea wall must be high enough so that the chance of a flood is no more than once in ten thousand years. We consider such a flood an extreme event. Or, in insurance mathematics it is of great interest to have statistical insight into the occurrence of large claims, due to natural catastrophes such as floods, hurricanes, high temperatures that give potential risk of fires of enormous proportions, etc. A common feature of this kind of events is that not many (if any at all) events of similar size have been observed in the past. Hence when making inferences related to extreme events, in particular one faces the problem of estimation where information from previous 'experiments' is scarce.

Let us illustrate our problem with an example. In Figure 1.1.a, in the real line are indicated 1877 observations from the sea level (in cm) at Delfzijl, which is located in the north coast of The Netherlands, measured during winterstorms in the years 1882-1991. The storm season lasts from October 1 until March 15. For more details on the data set see Dillingh et al. (1993). Now suppose that we are interested in estimating those sea levels (during winterstorms) that have probability .05 or .0001 of being exceeded. To start organising the information contained in the sample, we construct the empirical distribution function F_n, i.e. we put mass $1/n$ at every one of the observations (where n represents the sample size throughout). This is shown in Figure 1.1.b. Then, from this distribution the desired levels could be obtained by making the correspondence between the given probability and the level, as shown in Figure 1.1.b. But, it becomes clear that extra information is needed as the sea level to be estimated becomes larger, with the most extreme cases being when the given probability is smaller than $1/n$. Figures 1.2.a display the empirical distribution on a log-scale, i.e. the step function $-\log(1 - F_n)$, which is often an appropriate scale when one is mainly interested in the larger values of a sample.

Under rather general conditions, extreme value theory provides a class of

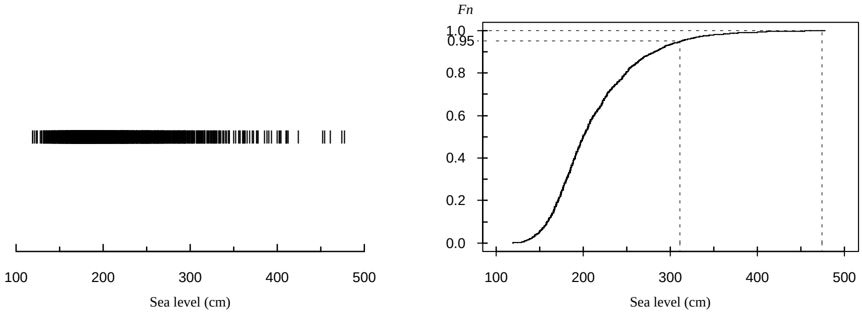

Figure 1.1: a) Left: sea level (cm) at Delfzijl (the data was trend corrected). b) Right: empirical distribution of the sea level sample.

functions to fit to the distribution of the largest observations. Figure 1.2.b shows some of these functions. A real parameter, γ say, comes into play which determines their shape. To fit the appropriate function to the tail of the distribution, one has then to decide on the shape, and moreover on the appropriate shift and scaling constants; for instance in case of deciding for a straight line ($\gamma = 0$) then one has to decide on the appropriate slope and origin of the line to fit to the tail.

The conditions that allow us to make such extrapolation are the extreme value conditions. If $X_1, X_2, ..., X_n$ are independent identically distributed (i.i.d.) random variables with common distribution function F, we require that there exist constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that the sequence of distribution functions

$$P\Big(\frac{\max_{1\le i\le n} X_i - b_n}{a_n} \le x\Big)$$

converges weakly to a non-degenerate distribution function. It then can be shown that with a judicious choice of a_n and b_n we have

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G_\gamma(x) := \exp\big(-(1+\gamma x)^{-1/\gamma}\big) \qquad (1.0.1)$$

for all x with $1 + \gamma x > 0$, where γ is a real parameter. This condition can be written in an equivalent form suitable for the application to quantiles. Let the function U be the generalized (left-continuous) inverse of the function $1/(1-F)$, i.e.

$$U(t) := \Big(\frac{1}{1-F}\Big)^{\leftarrow}(t) = F^{\leftarrow}(1-\frac{1}{t}), \quad t > 1.$$

Then (1.0.1) is equivalent to: for all $x > 0$

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}. \qquad (1.0.2)$$

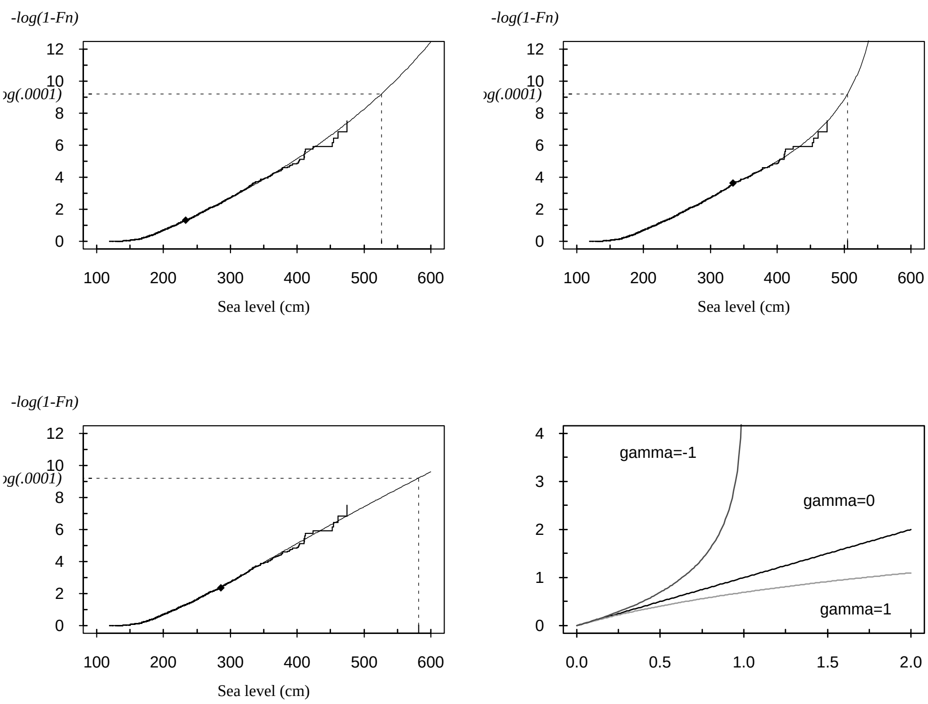

Figure 1.2: a) Top and bottom left: step function $-\log(1-F_n)$ of the sea level sample with estimated models. b) Bottom right: theoretical models.

This is relevant to high quantile estimation: it means that a high quantile $U(y)(= F^{\leftarrow}(1-1/y))$ for large y ($y = tx$ in (1.0.2)), is connected in a simple way asymptotically to a lower quantile $U(t)$ and $U(t)$ can be estimated using the empirical quantile function. Hence

$$U(y) \simeq U(t) + a(t)\frac{(y/t)^{\gamma} - 1}{\gamma}.$$

The left hand side is what we want to know and everything at the right hand side can be estimated using the empirical distribution function.

This means that the extreme value conditions are more or less unavoidable for estimation of high quantiles (outside the scope of the observations). Nevertheless these conditions are not always fulfilled. Hence it is useful to develop a check on the correctness of the asymptotic model (1.0.1) based on the observations at hand.

Such a check (or a test) is developed in Chapter 2 of this thesis for the one-dimensional case and in Chapter 4 for the two-dimensional case. In the course of the development of the test, some interesting results are obtained concerning tail empirical processes.

Chapter 3 is on large deviation for extremes. It is directly motivated by Proposition 2.3.2 in chapter 2. Under a second order condition we derive that

$$\lim_{n\to\infty} \frac{1 - F^n(a_n x_n + b_n)}{1 - G_\gamma(x)} = 1$$

for any sequence $x_n \uparrow 1/((-\gamma) \vee 0)$.

In chapter 5 we compare two estimators for the dependence function and the spectral measure by simulation. One estimator is based on maximum likelihood and the other estimator is based on non-parametric. We also present several methods to generate samples from distributions in the domain of attraction of a multi-dimensional extreme value distribution.

Chapter 6 deals with stable distribution. We derive alternative necessary and sufficient conditions for the domain of of attraction of a stable distribution in $\mathcal{R}^d$ which are phrased entirely in terms of (the joint distribution of) linear combinations of the marginals.

Chapter 2

Approximations to the Tail Empirical Distribution Function with Application to Testing Extreme Value Conditions

co-authors: Holger Drees and Laurens de Haan

Abstract. A weighted approximation to the tail empirical distribution function is derived which is suitable for applications in extreme value statistics. The approximation is used to develop an Anderson-Darling type test of the null hypothesis that the distribution function belongs to the domain of attraction of an extreme value distribution. A useful auxiliary result is a tail approximation to the distribution function.

2.1 Introduction

To assess the risk of extreme events that have not occurred yet, one needs to estimate the distribution function (d.f.) in the far tail. Extreme value theory provides a natural framework for an extrapolation of the distribution function beyond the range of available observations via the so-called Pareto approximation of the tail.

Assume that i.i.d. random variables (r.v.'s) X_i, $1 \le i \le n$, with d.f. F are observed such that

$$\lim_{n\to\infty} P\Big(a_n^{-1}(\max_{1\le i\le n} X_i - b_n) \le x\Big) = G_\gamma(x)$$

for all $x \in \mathbb{R}$, with some normalizing constants $a_n > 0$ and $b_n \in \mathbb{R}$; in short we write $F \in D(G_\gamma)$. Here

$$G_\gamma(x) := \exp\big(-(1+\gamma x)^{-1/\gamma}\big) \tag{2.1.1}$$

for all $x \in \mathbb{R}$ such that $1+\gamma x > 0$, and $\gamma \in \mathbb{R}$ is the so-called extreme value index. For $\gamma = 0$, the right-hand side of (2.1.1) is defined as $\exp(-\mathrm{e}^{-x})$.

This extreme value condition can be rephrased in the following way:

$$\lim_{t\to\infty} t\bar{F}(\tilde{a}(t)x + \tilde{b}(t)) = (1+\gamma x)^{-1/\gamma} \tag{2.1.2}$$

for all x with $1+\gamma x > 0$. Here $\bar{F} := 1 - F$, $\tilde{a}$ is some positive normalizing function and $\tilde{b}(t) := U(t)$ with

$$U(t) := \Big(\frac{1}{1-F}\Big)^{\leftarrow}(t) = F^{\leftarrow}\Big(1 - \frac{1}{t}\Big)$$

and $F^{\leftarrow}$ denoting the generalized inverse of F. In other words, if X is a r.v. with d.f. F, then

$$\lim_{t\to\infty} P\Big(\frac{X - \tilde{b}(t)}{\tilde{a}(t)} \le x \,\Big|\, X > \tilde{b}(t)\Big) = 1 - (1+\gamma x)^{-1/\gamma} =: V_\gamma(x)$$

for $x > 0$, where V_γ is a so-called generalized Pareto distribution. Thus, roughly speaking, we have for large t and $x > \tilde{b}(t)$

$$\bar{F}(x) = P\{X > x\} \approx t^{-1}\Big(1 + \gamma\frac{x - \tilde{b}(t)}{\tilde{a}(t)}\Big)^{-1/\gamma}, \tag{2.1.3}$$

that is, the tail of the d.f. can be approximated by a rescaled tail of a generalized Pareto distribution with suitable scale and location parameter and shape parameter γ. Since the latter can be easily extrapolated beyond the range of the observations, this framework offers an approach for estimating the d.f. F in the far tail.

Condition (2.1.2) holds for most standard distribution, but not for all distributions. Hence before applying approximation (2.1.3) one should check whether (2.1.2) is a reasonable assumption for the data set under consideration. To this end, we do not want to specify the exact parameters of the approximating generalized Pareto distribution beforehand.

A natural way to check the validity of (2.1.2) is to compare the tail of the empirical d.f. and a generalized Pareto distribution with estimated parameters by some goodness-of-fit test. Here we focus on tests of Anderson-Darling-type; however, using the empirical process approximations that will be established in the paper, similar results can be easily proved for other goodness-of-fit tests.

In the classical setting when a simple null hypothesis $F = F_0$ is to be tested, test statistics of Anderson-Darling type can be written in the form

$$\int_0^1 \Big(F_n(F_0^{-1}(x)) - x\Big)^2 \psi(x)\,dx$$

for a suitable weight function ψ which is unbounded near the boundary of the interval $[0,1]$; here F_n denotes the empirical d.f. defined by

$$F_n(x) := \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \le x\}}, \quad x \in \mathbb{R}.$$

If the null hypothesis is composite (but of parametric form), then F_0 is replaced with a d.f. with estimated parameters.

In the present framework two important differences must be taken into account. First, we do not assume that the left hand side and the right hand side of (2.1.3) are exactly equal, but the unknown d.f. F is only approximated by the "theoretical" generalized Pareto d.f. Second, this approximation is expected to hold only in the right tail, for $x > \tilde{b}(n/k)$ with $k \ll n$, say. In the asymptotic setting, we will assume that $k = k_n$ is an intermediate sequence, that is,

$$\lim_{n\to\infty} k_n = \infty, \quad \lim_{n\to\infty} k_n/n = 0.$$

The first condition is necessary to ensure consistency of the test, while the second condition reflects the restriction to the tail.

To be more specific, here we consider the test statistic

$$T_n := \int_0^1 \Big(\frac{n}{k_n}\bar{F}_n\big(\hat{a}(\frac{n}{k_n})\frac{x^{-\hat{\gamma}_n}-1}{\hat{\gamma}_n} + \hat{b}(\frac{n}{k_n})\big) - x\Big)^2 x^{\eta-2}\, dx \tag{2.1.4}$$

with $\bar{F}_n := 1 - F_n$. Here $\hat{\gamma}_n$, $\hat{a}(n/k_n)$ and $\hat{b}(n/k_n)$ are suitable estimators of the shape, scale and location parameter to be discussed later on, and η is an arbitrary positive constant. Since this test statistic measures a distance between the conditional distribution of the excesses above $\hat{b}(n/k_n)$ and an approximating generalized Pareto distribution (cf. (2.1.2)), a plot of this statistic as a function of $k = k_n$ may also be a useful tool for determining the point from which on approximation (2.1.3) is sufficiently accurate.

In the classical setting with simple null hypothesis, the asymptotic distribution of the Anderson-Darling test statistic under the null hypothesis is usually derived from a weighted approximation of the empirical distribution function. In analogy, in Theorem 2.2.1 we state a weighted approximation to the tail empirical process

$$E_n(x) := \sqrt{k_n}\bigg(\frac{n}{k_n}\bar{F}_n\Big(a\Big(\frac{n}{k_n}\Big)x + b\Big(\frac{n}{k_n}\Big)\Big) - (1+\gamma x)^{-1/\gamma}\bigg), \quad x \in \mathbb{R}. \tag{2.1.5}$$

For the uniform distribution such approximations are well known; see, e.g., Csörgő and Horváth (1993, Theorems 5.1.5 and 5.1.10). For more general d.f.'s $F \in D(G_\gamma)$, one must very carefully choose suitable modifications a and b of the normalizing functions to obtain accurate weighted approximations (cf. Lemma 2.2.1). Moreover, it turns out that for a certain class of d.f.'s with extreme value index $\gamma = 0$, a qualitatively different result holds. Proposition 2.2.1 gives an analogous approximation to the corresponding process with estimated parameters in the case $\gamma > -1/2$. The asymptotic normality of T_n then follows easily (Theorem 2.2.2).

As an auxiliary result to the approximation of E_n, we first need a weighted approximation to the (deterministic) tail d.f. $\bar{F}$ or, more precisely, to $t\bar{F}(a(t)x+b(t)) - (1+\gamma x)^{-1/\gamma}$, which is proved in Section 2.3 (see Proposition 2.3.2). This result is very useful in a wider context. For instance, Drees et al. (2003) have derived large deviation results in extreme value theory from this approximation. The Sections 2.4 and 2.5 contain the proofs of the main results, while in Section 2.6 asymptotic critical values are determined and the actual size of the Anderson-Darling type test with nominal size 5% is examined in a simulation study.

2.2 Main results

If i.i.d. uniformly distributed r.v.'s U_i are observed, then (2.1.2) holds with $\tilde{a}(t) = 1/t$ and $\gamma = -1$. For this particular case, Csörgő and Horváth (1993, Theorems 5.1.5 and 5.1.10) gave a weighted approximation to the normalized tail empirical process E_n defined in (2.1.5). Let

$$U_n(t) := \frac{1}{n}\sum_{i=1}^{n} I_{\{U_i \le t\}}, \quad t \in \mathbb{R},$$

denote the uniform tail empirical d.f. Then there exists a sequence of Brownian motions W_n such that

$$\sup_{t>0} t^{-1/2} \mathrm{e}^{-\varepsilon|\log t|} \left| \sqrt{k_n}\Big(\frac{n}{k_n} U_n\Big(\frac{k_n}{n}t\Big) - t\Big) - W_n(t) \right| \xrightarrow{P} 0 \tag{2.2.1}$$

as $n \to \infty$ for all intermediate sequences k_n, $n \in \mathbb{N}$ (see also Einmahl (1997, Corollary 3.3)).

By the well-known quantile transformation, $(F^{\leftarrow}(1-U_i))_{1\le i\le n}$ has the same distribution as $(X_i)_{1\le i\le n}$. Because $\bar{F}(t) \le t$ is equivalent to $F^{\leftarrow}(1-t) \le x$, it follows that $\bar{F}_n$ has the same distribution as

$$x \mapsto 1 - \frac{1}{n}\sum_{i=1}^{n} 1_{\{F^{\leftarrow}(1-U_i) \le x\}} = \frac{1}{n}\sum_{i=1}^{n} 1_{\{U_i < \bar{F}(x)\}} = U_n(\bar{F}(x) - 0),$$

that is the left hand limit of U_n at $\bar{F}(x)$. Hence, by the continuity of W_n, we obtain for suitable versions of $\bar{F}_n$ that

$$\begin{aligned} \sup_{\{x: z_n(x)>0\}} & (z_n(x))^{-1/2} \mathrm{e}^{-\varepsilon|\log z_n(x)|} \times \\ & \times \left| \sqrt{k_n}\Big[\frac{n}{k_n}\bar{F}_n\Big(\tilde{a}\big(\frac{n}{k_n}\big)x + \tilde{b}\big(\frac{n}{k_n}\big)\Big) - z_n(x)\Big] - W_n(z_n(x)) \right| \xrightarrow{P} 0 \end{aligned} \tag{2.2.2}$$

with

$$z_n(x) := \frac{n}{k_n}\bar{F}\Big(\tilde{a}(\frac{n}{k_n})x + \tilde{b}(\frac{n}{k_n})\Big).$$

In view of (2.1.2), one may conjecture that (2.2.2) still holds if $z_n(x)$ is replaced with $(1+\gamma x)^{-1/\gamma}$. However, for this to be justified, one must replace the normalizing functions $\tilde{a}$ and $\tilde{b}$ with suitable modifications such that (2.1.2) holds in a certain uniform sense. Moreover, we must bound the speed at which k_n tends to ∞.

In the sequel, we will focus on distributions which satisfy the following second order refinement of condition (2.1.2):

$$\lim_{t\to\infty} \frac{t\bar{F}\big(\tilde{a}(t)x + \tilde{b}(t)\big) - (1+\gamma x)^{-1/\gamma}}{\tilde{A}(t)} = (1+\gamma x)^{-1-1/\gamma} H_{\gamma,\rho}\big((1+\gamma x)^{-1/\gamma}\big) \tag{2.2.3}$$

for all x with $1+\gamma x > 0$ and some $\rho \le 0$ where

$$H_{\gamma,\rho}(x) := \frac{1}{\rho}\left(\frac{x^{\gamma+\rho}-1}{\gamma+\rho} - \frac{x^\gamma - 1}{\gamma}\right).$$

De Haan and Stadtmüller (1996) proved that (2.2.3) is equivalent to

$$\lim_{t\to\infty} \frac{\dfrac{U(tx)-\tilde{b}(t)}{\tilde{a}(t)} - \dfrac{x^\gamma-1}{\gamma}}{\tilde{A}(t)} = H_{\gamma,\rho}(x) \tag{2.2.4}$$

for all $x > 0$. Moreover, they showed that in (2.2.3) and (2.2.4) all possible non-trivial limits must essentially be of the given types, and that $|\tilde{A}|$ is necessarily ρ–varying.

Under this assumption, Drees (1998) and Cheng and Jiang (2001) determined suitable normalizing functions a and b such that convergence (2.2.4) holds uniformly in the following sense. In what follows, $f(t) \sim g(t)$ means $f(t)/g(t) \to 1$.

Lemma 2.2.1. *Suppose the second order condition (2.2.4) holds. Then there exist a function A, satisfying $A(t) \sim \tilde{A}(t)$ as $t \to \infty$, and for all $\varepsilon > 0$ a constant $t_\varepsilon > 0$ such that for all t and x with $\min(t, tx) \ge t_\varepsilon$*

$$x^{-(\gamma+\rho)} e^{-\varepsilon|\log x|} \left| \frac{\dfrac{U(tx)-b(t)}{a(t)} - \dfrac{x^\gamma-1}{\gamma}}{A(t)} - K_{\gamma,\rho}(x) \right| < \varepsilon. \tag{2.2.5}$$

Here

$$a(t) := \begin{cases} ct^\gamma & \text{if } \rho < 0, \\ \gamma U(t) & \text{if } \rho = 0,\ \gamma > 0, \\ -\gamma(U(\infty) - U(t)) & \text{if } \rho = 0,\ \gamma < 0, \\ U^{**}(t) + U^*(t) & \text{if } \rho = 0,\ \gamma = 0 \end{cases}$$

with $c := \lim_{t\to\infty} t^{-\gamma}\tilde{a}(t)$ *(which exists in that case),*

$$b(t) := \begin{cases} U(t) - a(t)A(t)/(\gamma+\rho) & \text{if } \gamma+\rho \neq 0,\ \rho < 0, \\ U(t) & \text{else,} \end{cases}$$

and

$$K_{\gamma,\rho}(x) := \begin{cases} \frac{1}{\gamma+\rho} x^{\gamma+\rho} & \text{if} \quad \rho < 0, \gamma+\rho \neq 0, \\ \log x & \text{if} \quad \rho < 0, \gamma+\rho = 0, \\ \frac{1}{\gamma} x^{\gamma} \log x & \text{if} \quad \rho = 0 \neq \gamma, \\ \frac{1}{2}\log^2 x & \text{if} \quad \rho = 0 = \gamma, \end{cases}$$

and for any integrable function g *the function* g^* *is defined by*

$$g^*(t) := g(t) - \frac{1}{t}\int_0^t g(u)dt.$$

In the sequel, we denote the right endpoint of the support of the generalized Pareto d.f. with extreme value index γ by

$$\frac{1}{(-\gamma)\vee 0} = \begin{cases} -1/\gamma & \text{if} \quad \gamma < 0, \\ \infty & \text{if} \quad \gamma \geq 0, \end{cases}$$

and its left endpoint by

$$-\frac{1}{\gamma \vee 0} = \begin{cases} -\infty & \text{if} \quad \gamma \leq 0, \\ -1/\gamma & \text{if} \quad \gamma > 0. \end{cases}$$

We have the following approximation to the tail empirical process E_n defined in (2.1.5):

Theorem 2.2.1. *Suppose that the second order condition (2.2.4) holds for some* $\gamma \in \mathbb{R}$ *and* $\rho \leq 0$*. Let* k_n *be an intermediate sequence such that* $\sqrt{k_n}A(n/k_n)$*,* $n \in \mathbb{N}$*, is bounded and choose* a*,* b *and* A *as in Lemma 2.2.1. Then there exist versions of* $\bar{F}_n$ *and a sequence of Brownian motions* W_n *such that for all* $x_0 > -1/(\gamma \vee 0)$
(i)

$$\sup_{x_0 \leq x < 1/((-\gamma)\vee 0)} \big((1+\gamma x)^{-1/\gamma}\big)^{-1/2+\varepsilon} \Big| E_n(x) - W_n\big((1+\gamma x)^{-1/\gamma}\big) - \sqrt{k_n}A\big(\frac{n}{k_n}\big)(1+\gamma x)^{-1/\gamma-1} K_{\gamma,\rho}\big((1+\gamma x)^{1/\gamma}\big)\Big| \xrightarrow{P} 0$$

if $\gamma \neq 0$ or $\rho < 0$, and

(ii)

$$\sup_{x_0 \le x < \infty} \Big(\max\Big(\mathrm{e}^{-x}, \frac{n}{k_n}\bar{F}\Big(a(\frac{n}{k_n})x + b(\frac{n}{k_n})\Big)\Big)\Big)^{-1/2+\varepsilon} \times$$

$$\times \Big| E_n(x) - W_n(\mathrm{e}^{-x}) - \sqrt{k_n}A(\frac{n}{k_n})\mathrm{e}^{-x}\frac{x^2}{2}\Big| \xrightarrow{P} 0$$

if $\gamma = \rho = 0$.

Remark 2.2.1. *If, in particular, $\sqrt{k_n}A(n/k_n)$ tends to 0, then the bias term $\sqrt{k_n}A(n/k_n)\ (1+\gamma x)^{-1/\gamma-1}K_{\gamma,\rho}((1+\gamma x)^{1/\gamma})$ is asymptotically negligible. In order for this statement to be true, it is sufficient to assume that the left-hand side of (2.2.3) remains bounded (rather than the present limit requirement) provided that k_n tends to infinity sufficiently slowly.*

The assertion in Theorem 2.2.1(ii) is wrong if the maximum of e^{-x} and $n/k_n\bar{F}(a(n/k_n)x+b(n/k_n))$ is replaced with just one of these two terms. Hence the asymptotic behavior of the tail empirical d.f. in the case $\gamma = \rho = 0$ is qualitatively different from the behavior in the case (i). This is due to the fact that in the case $\gamma \neq 0$ or $\rho < 0$ the tail behavior of F is essentially determined by the parameters γ and ρ, while in the case $\gamma = \rho = 0$ tail behaviors as diverse as $\bar{F}(x) \sim \exp(-\log^2 x)$, $\bar{F}(x) \sim \exp(-\sqrt{x})$ and $\bar{F}(x) \sim \exp(-x^2)$, say, are possible (cf. Example 2.3.1).

Nevertheless, also in the case $\gamma = \rho = 0$ results similar to the one in case (i) hold if $\max\big(\mathrm{e}^{-x},\ (n/k_n)\bar{F}(a(n/k_n)x + b(n/k_n))\big)$ is replaced with some weight function converging to ∞ much slower than e^{-x} as x tends to ∞:

Corollary 2.2.1. *Under the conditions of Theorem 2.2.1 with $\gamma = \rho = 0$ one has for all $\tau > 0$*

$$\sup_{x_0 \le x < \infty} \max(1, x^\tau) \Big| E_n(x) - W_n(\mathrm{e}^{-x}) - \sqrt{k_n}A(\frac{n}{k_n})\mathrm{e}^{-x}\frac{x^2}{2}\Big| \xrightarrow{P} 0.$$

The proofs of Theorem 2.2.1 and Corollary 2.2.1 are given in section 2.4.

According to these results, the standardized tail empirical d.f.

$$E_n((x^{-\gamma}-1)/\gamma) = \sqrt{k_n}\Big(\frac{n}{k_n}\bar{F}_n\Big(a(\frac{n}{k_n})\frac{x^{-\gamma}-1}{\gamma} + b(\frac{n}{k_n})\Big) - x\Big), \quad x \in (0,1]$$

converges to a Brownian motion plus a bias term if k_n tends to ∞ not too fast. This may be used to construct a test for $F \in D(G_\gamma)$. However, to this end, first the unknown parameters $\gamma, a(n/k_n)$ and $b(n/k_n)$ must be replaced with suitable estimators. The following result is an analog to Theorem 2.2.1(i) and Corollary 2.2.1 for the process with estimated parameters in the case $\gamma > -1/2$.

Proposition 2.2.1. *Suppose that the conditions of Theorem 2.2.1 are satisfied for some* $\gamma > -1/2$. *Let* $\hat{\gamma}_n$, $\hat{a}(n/k_n)$ *and* $\hat{b}(n/k_n)$ *be estimators such that*

$$\sqrt{k_n}\Big(\hat{\gamma}_n-\gamma,\ \frac{\hat{a}(n/k_n)}{a(n/k_n)}-1,\ \frac{\hat{b}(n/k_n)-b(n/k_n)}{a(n/k_n)}\Big)-\big(\Gamma(W_n),\ \alpha(W_n),\ \beta(W_n)\big) \xrightarrow{P} 0 \tag{2.2.6}$$

for some measurable real-valued functionals Γ, α *and* β *of the Brownian motions* W_n *used in Theorem 2.2.1. Then, for the versions of* $\bar{F}_n$ *used in Theorem 2.2.1 and every* $\varepsilon > 0$ *and* $\tau > 0$, *one has*

$$\begin{aligned}\sup_{0<x\le 1} h(x)\bigg|\sqrt{k_n}\,\Big(\frac{n}{k_n}\bar{F}_n\Big(\hat{a}\big(\frac{n}{k_n}\big)\frac{x^{-\hat{\gamma}_n}-1}{\hat{\gamma}_n}+\hat{b}\big(\frac{n}{k_n}\big)\Big)-x\Big)\\ -W_n(x)-L_n^{(\gamma)}(x)-\sqrt{k_n}A\big(\frac{n}{k_n}\big)x^{\gamma+1}K_{\gamma,\rho}\big(\frac{1}{x}\big)\bigg| \xrightarrow{P} 0\end{aligned} \tag{2.2.7}$$

with

$$L_n^{(\gamma)}(x) := \begin{cases} \frac{1}{\gamma}x\big(\frac{1}{\gamma}\Gamma(W_n)-\alpha(W_n)\big)+\frac{1}{\gamma}\Gamma(W_n)x\log x \\ \qquad -\frac{1}{\gamma}x^{1+\gamma}\big(\gamma\beta(W_n)+\frac{1}{\gamma}\Gamma(W_n)-\alpha(W_n)\big) & \text{if} \quad \gamma \neq 0, \\ x\big(-\beta(W_n)-\frac{1}{2}\Gamma(W_n)\log^2 x+\alpha(W_n)\log x\big) & \text{if} \quad \gamma = 0, \end{cases}$$

and

$$h(x) = \begin{cases} x^{-1/2+\varepsilon} & \text{if} \quad \gamma \neq 0 \text{ or } \rho < 0, \\ (1+|\log x|)^{\tau} & \text{if} \quad \gamma = \rho = 0. \end{cases}$$

Remark 2.2.2. *(i) If* $\gamma < -1/2$, *a rate of convergence of* $k_n^{-1/2}$ *for the estimators in (2.2.6) is not sufficient to ensure the approximation (2.2.7). To see this, note that in this case* $\hat{b}(n/k_n) - b(n/k_n)$ *is of larger order than* $k_n^{-1/2}(n/k_n)^{\gamma-\varepsilon}$ *and hence also of larger order than the difference between the* i_n*th largest order statistic and the right endpoint* $F^{\leftarrow}(1)$ *for some sequence* $i_n \to \infty$ *not too fast, leading, for small* $x > 0$, *to a non-negligible difference between* $\bar{F}_n(a(n/k_n)(x^{-\gamma}-1)/\gamma + b(n/k_n))$ *and the corresponding expression with estimated parameters.*

(ii) Typically the functionals Γ, α *and* β *depend on the underlying d.f.* F *only through* γ *if the estimators* $\hat{\gamma}_n$, $\hat{a}(n/k_n)$ *and* $\hat{b}(n/k_n)$ *use only the largest* $k_n + 1$ *order statistics and* $\sqrt{k_n}A(n/k_n) \to 0$. *This justifies the notation* $L_n^{(\gamma)}$ *for the limiting function occurring in (2.2.7) in that case. However, if* $\sqrt{k_n}A(n/k_n) \to c \neq 0$ *then* $L_n^{(\gamma)}$ *will also depend on c; for simplicity, we ignore this dependence in the notation.*

Example 2.2.1. In Proposition 2.2.1 one may use the so-called maximum likelihood estimator in a generalized Pareto model (cf. Smith (1987)). Denote the jth order statistic by $X_{j,n}$. Since the excesses $X_{n-i+1,n} - X_{n-k_n,n}$, $1 \le i \le k_n$ over the random threshold $X_{n-k_n,n}$ are approximately distributed according to a generalized Pareto distribution with shape parameter γ and scale parameter $\sigma_n := a(n/k_n)$ if $F \in D(G_\gamma)$ and k_n is not too big, γ and σ_n are estimated by the pertaining maximum likelihood estimators $\hat{\gamma}_n$ and $\hat{\sigma}_n$ in an exact generalized Pareto model for the excesses. They can be calculated as the solutions to the equations

$$\begin{aligned} \frac{1}{k}\sum_{i=1}^{k} \log\Big(1+\frac{\gamma}{\sigma}(X_{n-i+1,n}-X_{n-k,n})\Big) &= \gamma \\ \frac{1}{k}\sum_{i=1}^{k} \frac{1}{1+\frac{\gamma}{\sigma}(X_{n-i+1,n}-X_{n-k,n})} &= \frac{1}{\gamma+1}. \end{aligned}$$

In Theorem 2.1 of Drees et al. (2004) it is proved that $\hat{\gamma}_n$, $\hat{a}(n/k_n) := \hat{\sigma}_n$ and $\hat{b}(n/k_n) := X_{n-k_n,n}$ satisfy (2.2.6) with

$$\begin{aligned} \Gamma(W_n) &= -\frac{(\gamma+1)^2}{\gamma}\big((2\gamma+1)S_n - R_n\big) + (\gamma+1)W_n(1), \\ \alpha(W_n) &= -\frac{\gamma+1}{\gamma}\big(R_n - (\gamma+1)(2\gamma+1)S_n\big) - (\gamma+2)W_n(1), \\ \beta(W_n) &= W_n(1), \end{aligned}$$

where

$$\begin{aligned} R_n &:= \int_0^1 t^{-1} W_n(t)\,dt, \\ S_n &:= \int_0^1 t^{\gamma-1} W_n(t)\,dt, \end{aligned}$$

provided $\sqrt{k_n}A(n/k_n) \to 0$; if $\sqrt{k_n}A(n/k_n) \to c > 0$ then additional bias terms enter the formulas. As usual, for $\gamma = 0$, these expressions are to be interpreted as their limits as γ tends to 0, that is,

$$\begin{aligned} \Gamma(W_n) &= -\int_0^1 (2+\log t)t^{-1}W_n(t)\,dt + W_n(1), \\ \alpha(W_n) &= \int_0^1 (3+\log t)t^{-1}W_n(t)\,dt - 2W_n(1), \\ \beta(W_n) &= W_n(1). \end{aligned}$$

(Applying Vervaat's (1972) lemma to the approximation to the tail empirical distribution function given in Theorem 2.2.1, restricted to a compact interval bounded away from 0, and then using a Taylor expansion of $t \mapsto (t^{-\gamma}-1)/\gamma$

shows that the Brownian motions used by Drees et al. (2004) are indeed the Brownian motions used in Proposition 2.2.1 multiplied with -1.)

Hence one may apply Proposition 2.2.1 to obtain the asymptotics of the tail empirical distribution function with estimated parameters. □

It is easy to devise tests for $F \in D(G_\gamma)$ with $\gamma > -1/2$ using approximation (2.2.7). For example, using the following limit theorem, the critical values of the Anderson-Darling type test can be calculated which rejects the null hypothesis if $k_n T_n$ (defined in (2.1.4)) is too large.

Theorem 2.2.2. *Under the conditions of Proposition 2.2.1 with* $\sqrt{k_n}A(n/k_n) \to 0$ *one has*

$$k_n T_n - \int_0^1 \Big(W_n(x) + L_n^{(\gamma)}(x)\Big)^2 x^{\eta-2}\, dx \xrightarrow{P} 0 \tag{2.2.8}$$

for all $\eta > 0$ *if* $\gamma \neq 0$ *or* $\rho < 0$*, and all* $\eta \geq 1$ *if* $\gamma = \rho = 0$.

Since the continuous distribution of $\int_0^1 (W_n(x) + L_n^{(\gamma)}(x))^2 x^{\eta-2} dx$ does not depend on n, for fixed $\gamma > -1/2$ its quantiles $Q_{p,\gamma}$ defined by $P\{\int_0^1 (W_n(x) + L_n^{(\gamma)}(x))^2 x^{\eta-2} dx \leq Q_{p,\gamma}\} = p$ can be obtained by simulations (see Section 6). Then the one-sided test rejecting $F \in D(G_\gamma)$ if $k_n T_n > Q_{1-\bar\alpha,\gamma}$ has asymptotic size $\bar\alpha \in (0,1)$.

If one wants to test $F \in D(G_\gamma)$ for an arbitrary unknown $\gamma > -1/2$, one may use the test rejecting the null hypothesis if $k_n T_n > Q_{1-\bar\alpha,\tilde\gamma_n}$ for some estimator $\tilde\gamma_n$ which is consistent for γ if $F \in D(G_\gamma)$. If the functionals Γ, α and β determining the limit distributions of $\hat\gamma_n$, $\hat a(n/k_n)$ and $\hat b(n/k_n)$ are continuous functions of γ (like the ones obtained in Example 2.2.1), then also $L_n^{(\gamma)}(x)$ and hence the quantiles $Q_{p,\gamma}$ are continuous functions of γ. Thus the test has asymptotic size $\bar\alpha$.

However, recall that, in fact, for (2.2.8) to hold we have not merely assumed that $F \in D(G_\gamma)$ but also that the second order condition (2.2.4) holds and, for the particular k_n used in the definition of the test statistic T_n, in addition we have assumed that $A(t) \to 0$ sufficiently fast such that $\sqrt{k_n}A(n/k_n) \to 0$. Hence, we actually test the subset of the null hypothesis $F \in D(G_\gamma)$ described by these additional assumptions. This, however, is exactly what is needed in statistical applications. For instance, note that typically the very same assumptions are made when confidence intervals for extreme quantiles or for exceedance probability over high thresholds are calculated. Therefore, for this purpose, one must not only check whether $F \in D(G_\gamma)$ but whether the Pareto approximation is sufficiently accurate for the number of order statistics used for estimation! Moreover, if one lets k vary, then the test statistic can also be used to find the largest k for which the Pareto approximation of the tail distribution beyond $X_{n-k,n}$ is justified.

A test for a similar hypothesis, but based on the tail empirical quantile function instead of the tail empirical distribution function, has been discussed by Dietrich et al. (2002). That test does not require $\gamma > -1/2$ but, on the other hand, $U(\infty) > 0$ and a slightly different second order condition were assumed.

The test based on the statistic $k_n T_n$ becomes particularly simple if Γ, α and β are the zero functional, that is, the estimators $\hat{\gamma}_n$, $\hat{a}(n/k_n)$ and $\hat{b}(n/k_n)$ converge at a faster rate than $k_n^{-1/2}$. This can be achieved by using suitable estimators based on m_n largest order statistics with $k_n = o(m_n)$ and $\sqrt{m_n}A(n/m_n) \to 0$. (For example, γ may be estimated by the estimator given in Example 2.2.1 with m_n instead of k_n, and $b(n/k_n)$ by a quantile estimator of the type described in de Haan and Rootzén (1993).) In that case the limit distribution $\int_0^1 W_n^2(x)x^{\eta-2}dx$ of the test statistic $k_n T_n$ does not depend on γ, so that no consistent estimator $\tilde{\gamma}_n$ for γ is needed. However, this approach has two disadvantages. Firstly, in practice it is often not an easy task to choose k_n such that the bias is negligible (i.e. $\sqrt{k_n}A(n/k_n) \to 0$). It is even more delicate to choose two numbers k_n and m_n such that k_n is much smaller than m_n but not too small and, at the same time, the bias of the estimators of the parameters is still not dominating when these are based on m_n order statistics. Secondly, while this approach may lead to a test whose actual size is closer to the nominal value $\bar{\alpha}$, the power of the test will probably higher if one choose a larger value for k_n, e.g. $k_n = m_n$, because the larger k_n the larger will typically be the test statistic $k_n T_n$ if the tail empirical d.f. is not well approximated by a generalized Pareto d.f. For these reasons, in the simulation study we will focus on the case where the tail empirical d.f. and the estimators $\hat{\gamma}_n$, $\hat{a}(n/k_n)$ and $\hat{b}(n/k_n)$ are based on the same number of largest order statistics.

2.3 Tail Approximation to the Distribution Function

A substantial part of the proof of Theorem 2.2.1 consists of proving an approximation to the tail of the (deterministic) distribution function.

For all $c, \delta > 0$ define sets

$$D_{t,\rho} := D_{t,\rho,\delta,c} := \begin{cases} \{x : t\bar{F}(a(t)x + b(t)) \le ct^{-\delta+1}\} & \text{if } \rho < 0, \\ \{x : t\bar{F}(a(t)x + b(t)) \le |A(t)|^{-c}\} & \text{if } \rho = 0. \end{cases}$$

Check that, in particular, eventually $[x_0, \infty) \subset D_{t,\rho}$ for all $x_0 > -1/(\gamma \vee 0)$.

Proposition 2.3.1. *Suppose that the second order relation (2.2.4) holds for some $\gamma \in \mathbb{R}$ and $\rho \le 0$. For $\varepsilon > 0$, define*

$$w_t(x) :=$$
$$\begin{cases} \big(t\bar{F}(a(t)x + b(t))\big)^{\rho-1} \mathrm{e}^{-\varepsilon|\log(t\bar{F}(a(t)x+b(t)))|}, & \gamma \ne 0 \ \text{or}\ \rho \ne 0, \\ \min\Big((t\bar{F}(a(t)x + b(t)))^{-1} \mathrm{e}^{-\varepsilon|\log(t\bar{F}(a(t)x+b(t)))|},\ \mathrm{e}^{x-\varepsilon|x|} \Big), & \gamma = \rho = 0. \end{cases}$$

Then, for all $\varepsilon, \delta, c > 0$,

$$\sup_{x\in D_{t,\rho}} w_t(x) \Bigg| \frac{t\bar{F}(a(t)x+b(t)) - (1+\gamma x)^{-1/\gamma}}{A(t)} - \big(t\bar{F}(a(t)x+b(t))\big)^{1+\gamma} K_{\gamma,\rho}\Big(\frac{1}{t\bar{F}(a(t)x+b(t))}\Big)\Bigg| \to 0$$

as $t \to \infty$.

Moreover, we establish an analogous result where $t\bar{F}(a(t)x+b(t))$ is replaced with $(1+\gamma x)^{-1/\gamma}$. To this end, let for $\delta, c > 0$

$$\tilde{D}_{t,\rho} := \tilde{D}_{t,\rho,\delta,c} := \begin{cases} \{x : (1+\gamma x)^{-1/\gamma} \le ct^{-\delta+1}\} & \text{if } \rho < 0, \\ \{x : (1+\gamma x)^{-1/\gamma} \le |A(t)|^{-c}\} & \text{if } \rho = 0, \end{cases}$$

and, for $\gamma \neq 0$ or $\rho < 0$,

$$\tilde{w}_t(x) := \big((1+\gamma x)^{-1/\gamma}\big)^{\rho-1} \exp\big(-\varepsilon|\log((1+\gamma x)^{-1/\gamma})|\big).$$

Proposition 2.3.2. *If the second order relation (2.2.4) holds for some $\gamma \in \mathbb{R}$ and $\rho \le 0$, then*

$$\sup_{x\in D_{t,\rho}} w_t(x)\Big|\frac{t\bar{F}(a(t)x+b(t)) - (1+\gamma x)^{-1/\gamma}}{A(t)} - (1+\gamma x)^{-1/\gamma-1} K_{\gamma,\rho}\big((1+\gamma x)^{1/\gamma}\big)\Big| \to 0$$

as $t \to \infty$. Moreover, if $\gamma \neq 0$ or $\rho < 0$, then

$$\sup_{x\in \tilde{D}_{t,\rho}} \tilde{w}_t(x)\Big|\frac{t\bar{F}(a(t)x+b(t)) - (1+\gamma x)^{-1/\gamma}}{A(t)} - (1+\gamma x)^{-1/\gamma-1} K_{\gamma,\rho}\big((1+\gamma x)^{1/\gamma}\big)\Big| \to 0,$$

and for $\gamma = \rho = 0$

$$\sup_{x\in \tilde{D}_{t,0}} w_t(x)\Big|\frac{t\bar{F}(a(t)x+b(t)) - \mathrm{e}^{-x}}{A(t)} - \mathrm{e}^{-x}\frac{x^2}{2}\Big| \to 0$$

for all $\delta, c > 0$ as $t \to \infty$.

At first glance, it is somewhat surprising that the results look differently in the case $\gamma = \rho = 0$ in that one needs a more complicated weight function, namely the minimum of a function of the standardized tail d.f. $t\bar{F}(a(t)x+b(t))$ and the corresponding function of the limiting exponential d.f. The following example shows that indeed the straightforward analog to the assertion in the case $\gamma \neq 0$ or $\rho < 0$ does not hold, because, in the case $\gamma = \rho = 0$, these two functions may behave quite differently for large x, despite the fact that for fixed x the former converges to the latter.

Example 2.3.1. Here we give an example of a d.f. satisfying (2.2.4) such that

$$\sup_{\{x:x>c\log|A(t)|\}} e^{x-\varepsilon|x|}\left|\frac{t\bar{F}\big(a(t)x+b(t)\big)-e^{-x}}{A(t)}-e^{-x}\frac{x^2}{2}\right| \tag{2.3.1}$$

does not tend to 0 for any $c,\varepsilon>0$.

Let $F(x):=1-e^{-\sqrt{x}}$, $x>0$, and $a(t):=2\log t$, $b(t):=\log^2 t$, $A(t):=1/\log t$. Then $U(x)=\log^2 x$ satisfies the second order condition (2.2.4):

$$\frac{1}{A(t)}\Big(\frac{U(tx)-U(t)}{a(t)}-\log x\Big)\to\frac{\log^2 x}{2}$$

as $t\to\infty$. Moreover

$$t\bar{F}\big(a(t)x+b(t)\big)=t\exp\Big(-\sqrt{2x\log t+\log^2 t}\Big)=\exp\Big(-\log t\big(\sqrt{1+2x/\log t}-1\big)\Big).$$

Hence, for $x=x(t)=\lambda(t)\log t/2$ with $\lambda(t)\to\infty$ as $t\to\infty$, one obtains

$$\begin{aligned}
&t\bar{F}\big(a(t)x+b(t)\big)=\exp\Big(-\log t\sqrt{\lambda(t)}(1+o(1))\Big),\\
&e^{-x}\frac{x^2}{2}=\frac{1}{8}\exp\Big(2(\log\log t+\log\lambda(t))-\frac{1}{2}\lambda(t)\log t\Big)=o(t\bar{F}(a(t)x+b(t))),\\
&e^{-x}=o\big(t\bar{F}(a(t)x+b(t))\big),
\end{aligned}$$

so that

$$\frac{t\bar{F}\big(a(t)x+b(t)\big)-e^{-x}}{A(t)}-e^{-x}\frac{x^2}{2}=\frac{t\bar{F}\big(a(t)x+b(t)\big)}{A(t)}(1+o(1)).$$

However, this contradicts the convergence of (2.3.1) to 0 as $t\to\infty$:

$$\begin{aligned}
&(e^{-x})^{-1+\varepsilon}\left|\frac{t\bar{F}(a(t)x+b(t))-e^{-x}}{A(t)}-e^{-x}\frac{x^2}{2}\right|\\
&=(e^{-x})^{-1+\varepsilon}\frac{t\bar{F}(a(t)x+b(t))}{A(t)}(1+o(1))\\
&=\exp\Big(\frac{1-\varepsilon}{2}\lambda(t)\log t-\sqrt{\lambda(t)}\log t(1+o(1))\Big)\cdot\frac{1+o(1)}{A(t)}\\
&\to\infty.
\end{aligned}$$

Likewise one can show that $F(x)=1-e^{-x^2}$, $x>0$, satisfies the second order condition (2.2.4) but that

$$\sup_{x\in D_{t,0}}\big(t\bar{F}(a(t)x+b(t))\big)^{-1}\exp\big(-\varepsilon|\log(t\bar{F}(a(t)x+b(t)))|\big)\times$$
$$\times\left|\frac{t\bar{F}\big(a(t)x+b(t)\big)-e^{-x}}{A(t)}-e^{-x}\frac{x^2}{2}\right|\to\infty.$$

□

Before proving the propositions, we need an auxiliary lemma. Let

$$q_t(x) := \frac{U(tx) - b(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma}.$$

Lemma 2.3.1. *For each $\varepsilon > 0$, there exists $\tilde{t}_\varepsilon > 0$ such that*

$$\sup_{x \geq \tilde{t}_\varepsilon / t} x^{-(\gamma+\rho)} \mathrm{e}^{-\varepsilon |\log x|} |q_t(x)| = O(A(t))$$

as $t \to \infty$.

Proof. We focus on the case $\gamma = \rho = 0$; the assertion can be proved by the same arguments in the other cases. From Lemma 2.2.1 we know that, for each $\delta > 0$, there exists t_δ such that for $t, tx \geq t_\delta$

$$\mathrm{e}^{-\varepsilon |\log x|} |q_t(x)| \leq \mathrm{e}^{-\varepsilon |\log x|} |A(t)| \Big(\frac{\log^2 x}{2} + \delta \mathrm{e}^{\delta |\log x|} \Big).$$

Choose $\delta < \varepsilon$ and $\tilde{t}_\varepsilon = t_\delta$ to obtain the assertion, since $\sup_{x>0} \mathrm{e}^{-\varepsilon |\log x|} \log^2 x < \infty$. □

Let

$$B_{t,\rho} := B_{t,\rho,\delta,c} := \begin{cases} [ct^{\delta-1}, \infty) & \text{if } \rho < 0, \\ \{y : |\log y| \leq c |\log |A(t)||\} & \text{if } \rho = 0, \end{cases}$$

with $\delta, c > 0$.

Corollary 2.3.1. *For all $c, \delta > 0$,*

$$\sup_{x \in B_{t,\rho}} x^{-\gamma} |q_t(x)| \to 0$$

as $t \to \infty$.

Proof. For $\rho < 0$, choose $\varepsilon \leq |\rho|$ in Lemma 2.3.1 to obtain

$$\sup_{x \geq 1} x^{-\gamma} |q_t(x)| \leq \sup_{x \geq 1} x^{-(\gamma+\rho)} \mathrm{e}^{-\varepsilon |\log x|} |q_t(x)| = O(A(t)) = o(1).$$

For all $c, \delta, \tilde{t}_\varepsilon > 0$, eventually $ct^{\delta-1}$ is greater than $\tilde{t}_\varepsilon / t$. Hence, by Lemma 2.3.1,

$$\sup_{ct^{\delta-1} \leq x < 1} x^{-\gamma} |q_t(x)| \leq O(A(t)) \cdot \sup_{ct^{\delta-1} \leq x < 1} x^{\rho - \varepsilon} = O\big(A(t) \cdot t^{(\delta-1)(\rho-\varepsilon)}\big) \to 0$$

if $(\delta - 1)(\rho - \varepsilon) < -\rho$ (which is satisfied for sufficient small $\varepsilon > 0$), since $A(t)$ is ρ-varying and hence $A(t) = o(t^{\eta+\rho})$ for all $\eta > 0$.

In the case $\rho = 0$, one has for all $\varepsilon \in (0, 1/c)$

$$\sup_{x \in B_{t,\rho}} x^{-\gamma} |q_t(x)| \leq O(A(t)) \cdot \sup_{x \in B_{t,\rho}} \mathrm{e}^{\varepsilon |\log x|} = O\big(A(t) \mathrm{e}^{\varepsilon c |\log |A(t)||}\big) \to 0,$$

since eventually $|A(t)|^c > \tilde{t}_\varepsilon/t$ and hence $x > \tilde{t}_\varepsilon/t$ for all $x \in B_{t,\rho}$. □

Proof of Proposition 2.3.1.

For simplicity assume that F is eventually strictly increasing. (For more general F, the assertion follows by standard extra arguments using the second order condition (2.2.3).) Let $g(x) := (1+\gamma x)^{-1/\gamma}$ and

$$y := \frac{1}{t\bar{F}(a(t)x + b(t))},$$

which implies that $x = (U(ty) - b(t))/a(t)$. Then $g'(x) = -(g(x))^{\gamma+1}$ and $g''(x) = (\gamma+1)(g(x))^{2\gamma+1}$, and so

$$\begin{aligned}
&t\bar{F}\big(a(t)x + b(t)\big) - (1+\gamma x)^{-1/\gamma} \\
&= -\Bigg(\Big(1+\gamma\frac{U(ty)-b(t)}{a(t)}\Big)^{-1/\gamma} - \Big(1+\gamma\frac{y^\gamma - 1}{\gamma}\Big)^{-1/\gamma}\Bigg) \\
&= -\Bigg(g\Big(\frac{U(ty)-b(t)}{a(t)}\Big) - g\Big(\frac{y^\gamma-1}{\gamma}\Big)\Bigg) \\
&= q_t(y)\Bigg(-g'\Big(\frac{y^\gamma-1}{\gamma}\Big)\Bigg) - \int_0^{q_t(y)}\int_0^s g''\Big(\frac{y^\gamma-1}{\gamma}+u\Big)\,du\,ds \\
&= q_t(y)y^{-\gamma-1} - \int_0^{q_t(y)}\int_0^s (1+\gamma)\Bigg(1+\gamma\Big(\frac{y^\gamma-1}{\gamma}+u\Big)\Bigg)^{-1/\gamma-2} du\,ds
\end{aligned}$$

with $(1+\gamma x)^{-1/\gamma-j} := \mathrm{e}^{-x}$ for $\gamma = 0$ and $j = 1,2$.

Since $(1+\gamma((y^\gamma-1)/\gamma+u))^{-1/\gamma-2}$ lies between $(1+\gamma(y^\gamma-1)/\gamma)^{-1/\gamma-2} = y^{-1-2\gamma}$ and $(1+\gamma((y^\gamma-1)/\gamma+q_t(y)))^{-1/\gamma-2} = y^{-1-2\gamma}(1+\gamma y^{-\gamma}q_t(y))^{-1/\gamma-2}$, Corollary 2.3.1 yields

$$\Big|t\bar{F}(a(t)x+b(t)) - (1+\gamma x)^{-1/\gamma} - q_t(y)y^{-1-\gamma}\Big| \le 2|1+\gamma|y^{-1-2\gamma}q_t^2(y) \quad (2.3.2)$$

for all $y \in B_{t,\rho}$ and sufficiently large t.

Since $ty \to \infty$ uniformly for $y \in B_{t,\rho}$, (2.2.5), Lemma 2.3.1 and Corollary 2.3.1 imply

$$\begin{aligned}
&\sup_{y\in B_{t,\rho}} w_t(x)\Bigg|\frac{t\bar{F}\big(a(t)x+b(t)\big) - (1+\gamma x)^{-1/\gamma}}{A(t)} \\
&\qquad - \big(t\bar{F}(a(t)x+b(t))\big)^{1+\gamma}K_{\gamma,\rho}\Big(\frac{1}{t\bar{F}(a(t)x+b(t))}\Big)\Bigg| \\
&\le \sup_{y\in B_{t,\rho}} y^{1-\rho}\mathrm{e}^{-\varepsilon|\log y|}\Bigg(\Bigg|\frac{q_t(y)y^{-(1+\gamma)}}{A(t)} - y^{-(1+\gamma)}K_{\gamma,\rho}(y)\Bigg| + 2|1+\gamma|\frac{y^{-(1+2\gamma)}}{|A(t)|}q_t^2(y)\Bigg) \\
&\le \sup_{y\in B_{t,\rho}} y^{-(\gamma+\rho)}\mathrm{e}^{-\varepsilon|\log y|}\Bigg|\frac{q_t(y)}{A(t)} - K_{\gamma,\rho}(y)\Bigg| \\
&\qquad + 2|1+\gamma|\sup_{y\in B_{t,\rho}} y^{-(\gamma+\rho)}\mathrm{e}^{-\varepsilon|\log y|}\frac{|q_t(y)|}{|A(t)|}\sup_{y\in B_{t,\rho}} y^{-\gamma}|q_t(y)| \\
&\to 0.
\end{aligned}$$

Because $x \in D_{t,\rho,\delta,c}$ is equivalent to $y \in B_{t,\rho,\delta,1/c}$ if $\rho < 0$, the assertion is proved in that case.

In the case $\rho = 0$, we have proved that for all $c > 0$

$$\sup_{\{x:|A(t)|^c \le 1/y \le |A(t)|^{-c}\}} w_t(x) \left| \frac{y^{-1} - (1+\gamma x)^{-1/\gamma}}{A(t)} - y^{-(1+\gamma)} K_{\gamma,0}(y) \right| \to 0.$$

Thus it suffices to prove that

$$\sup_{\{x:1/y<|A(t)|^c\}} w_t(x) \left| \frac{y^{-1} - (1+\gamma x)^{-1/\gamma}}{A(t)} - y^{-(1+\gamma)} K_{\gamma,0}(y) \right| \to 0,$$

where we may assume that $c > 0$ is sufficiently large. Note that

$$w_t(x) \left| y^{-(1+\gamma)} K_{\gamma,0}(y) \right| = O\left(\mathrm{e}^{-\varepsilon|\log y|} \log^2 y \right) = o(1)$$

uniformly for $1/y < |A(t)|^c$. Moreover, for $c > 1/\varepsilon$,

$$w_t(x) y^{-1} \le \mathrm{e}^{-\varepsilon|\log y|} \le |A(t)|^{c\varepsilon} = o(A(t))$$

for all x such that $1/y < |A(t)|^c$.

Therefore, it suffices to verify that

$$\sup_{\{x:\ t\bar{F}(a(t)x+b(t))<|A(t)|^c\}} w_t(x)(1+\gamma x)^{-1/\gamma} = o(A(t)). \tag{2.3.3}$$

To this end, we distinguish three cases.

First suppose $\gamma > 0$. Then $(1+\gamma x)U(t) = a(t)x + b(t) \to \infty$ uniformly for all x such that $1/y = t\bar{F}(a(t)x+b(t)) < |A(t)|^c \to 0$. By the Potter bounds (see Bingham et al. (1987), Theorem 1.5.6)

$$\begin{aligned} y^{-1} &= t\bar{F}\big(a(t)x + b(t)\big) \\ &= \frac{\bar{F}\big((1+\gamma x)U(t)\big)}{\bar{F}(U(t))} \ge \frac{1}{2}(1+\gamma x)^{-1/(\gamma(1-\varepsilon/2))} \end{aligned} \tag{2.3.4}$$

for sufficient large t. Hence the left-hand side of (2.3.3) is bounded by

$$\sup_{\{x:1/y<|A(t)|^c\}} y^{1-\varepsilon}\big(2y^{-1}\big)^{1-\varepsilon/2} \le 2|A(t)|^{\varepsilon c/2} = o(A(t))$$

when we choose $c > 2/\varepsilon$.

Likewise, for $\gamma < 0$, one has

$$\begin{aligned} y^{-1} &= t\bar{F}\big(a(t)x + b(t)\big) \\ &= \frac{\bar{F}\big(U(\infty) - (1+\gamma x)(U(\infty) - U(t))\big)}{\bar{F}\big(U(\infty) - (U(\infty) - U(t))\big)} \ge \frac{1}{2}(1+\gamma x)^{-1/(\gamma(1-\varepsilon/2))} \end{aligned} \tag{2.3.5}$$

and one can argue like in the case $\gamma > 0$.

Finally, if $\gamma = 0$ then the left-hand side of (2.3.3) is bounded by

$$\sup_{\{x:\ t\bar{F}(a(t)x_t+b(t))<|A(t)|^c\}} \mathrm{e}^{-\varepsilon x} = \mathrm{e}^{-\varepsilon x_t}$$

with $x_t = \inf\{x : t\bar{F}(a(t)x + b(t)) \le |A(t)|^c\}$. According to (2.3.2), Lemma 2.3.1, and Corollary 2.3.1, one has eventually

$$\begin{aligned}
\mathrm{e}^{-x_t} &\le |A(t)|^c + \big|q_t\big(|A(t)|^{-c}\big)\big||A(t)|^c + 2q_t^2\big(|A(t)|^{-c}\big)|A(t)|^c \\
&= |A(t)|^c\Big(1 + \big|q_t\big(|A(t)|^{-c}\big)\big| + 2q_t^2\big(|A(t)|^{-c}\big)\Big) \\
&= |A(t)|^c\Big(1 + O\big(|A(t)|^{1-\varepsilon c}\big)\Big) \\
&= O\big(|A(t)|^{c(1-\varepsilon)}\big)
\end{aligned}$$

which implies that $\mathrm{e}^{-\varepsilon x_t} = O(|A(t)|^{\varepsilon c(1-\varepsilon)}) = o(A(t))$ for $c > 2/\varepsilon$ and $\varepsilon < 1/2$.

The proof of Proposition 2.3.1 is complete. □

Proof of Proposition 2.3.2.

Recall the definition $y := 1/\big(t\bar{F}(a(t)x + b(t))\big)$. We consider three cases.

Case (i): $\rho < 0$.

Inequality (2.3.2) and Corollary 2.3.1 imply

$$\sup_{x\in D_{t,\rho}} \left|y(1+\gamma x)^{-1/\gamma} - 1\right| \le \sup_{y\ge ct^{\delta-1}} y^{-\gamma}|q_t(y)| + 2|1+\gamma| \sup_{y\ge ct^{\delta-1}} (y^{-\gamma}q_t(y))^2 \to 0. \tag{2.3.6}$$

Hence, for $\gamma + \rho \ne 0$, by the definition of $K_{\gamma,\rho}$

$$\begin{aligned}
&\sup_{x\in D_{t,\rho}} w_t(x)\left|(1+\gamma x)^{-(1+1/\gamma)}K_{\gamma,\rho}\big((1+\gamma x)^{1/\gamma}\big) - y^{-(1+\gamma)}K_{\gamma,\rho}(y)\right| \\
&= \sup_{x\in D_{t,\rho}} \mathrm{e}^{-\varepsilon|\log y|}\frac{1}{|\gamma+\rho|}\left|\big(y(1+\gamma x)^{-1/\gamma}\big)^{1-\rho} - 1\right| \\
&\to 0.
\end{aligned} \tag{2.3.7}$$

If $\gamma + \rho = 0$, then the left-hand side of (2.3.7) equals

$$\begin{aligned}
\sup_{x\in D_{t,\rho}} \mathrm{e}^{-\varepsilon|\log y|}\Big|&\big(y(1+\gamma x)^{-1/\gamma}\big)^{1+\gamma}\log\big(y(1+\gamma x)^{-1/\gamma}\big) \\
&+ \Big(\big(y(1+\gamma x)^{-1/\gamma}\big)^{1+\gamma} - 1\Big)(-\log y)\Big| \to 0.
\end{aligned}$$

Now the first assertion is immediate from Proposition 2.3.1. In view of (2.3.6), $\tilde{w}_t(x)/w_t(x)$ tends to 1 uniformly for $x \in D_{t,\rho}$. Moreover, $(1+\gamma x)^{-1/\gamma} \le ct^{-\delta+1}$ implies $t\bar{F}(a(t)x+b(t)) \le 2ct^{-\delta+1}$ for sufficient large t. Thus the second assertion follows immediate from the first.

Case (ii): $\rho = 0,\ \gamma \neq 0$.
Define

$$\begin{aligned} D^1_{t,0} &:= \big\{x\ :\ |A(t)|^c \le t\bar F(a(t)x+b(t)) \le |A(t)|^{-c}\big\} \\ &= \big\{x\ :\ \big|\log\big(t\bar F(a(t)x+b(t))\big)\big| \le c|\log|A(t)||\big\}, \\ D^2_{t,0} &:= \big\{x\ :\ t\bar F(a(t)x+b(t)) \le |A(t)|^c\big\}, \end{aligned}$$

so that $D_{t,0} = D^1_{t,0} \cup D^2_{t,0}$. As in case (i), (2.3.2) and Corollary 2.3.1 imply

$$\sup_{x\in D^1_{t,0}} \Big|y(1+\gamma x)^{-1/\gamma} - 1\Big| \le \sup_{y\in B_{t,0}} y^{-\gamma}|q_t(y)| + 2|1+\gamma| \sup_{y\in B_{t,0}} (y^{-\gamma}q_t(y))^2 \to 0. \tag{2.3.8}$$

Hence

$$\begin{aligned} &\sup_{x\in D^1_{t,0}} w_t(x)\Big|(1+\gamma x)^{-(1+1/\gamma)}K_{\gamma,0}\big((1+\gamma x)^{1/\gamma}\big) - y^{-(1+\gamma)}K_{\gamma,0}(y)\Big| \\ &= \frac{1}{|\gamma|}\sup_{x\in D^1_{t,0}} \mathrm{e}^{-\varepsilon|\log y|}\Big|y(1+\gamma x)^{-1/\gamma}\log\big(y(1+\gamma x)^{-1/\gamma}\big) \\ &\qquad\qquad + \big(y(1+\gamma x)^{-1/\gamma} - 1\big)(-\log y)\Big| \\ &\to 0. \end{aligned} \tag{2.3.9}$$

Note that

$$\sup_{x\in D^2_{t,0}} w_t(x)\Big|y^{-(1+\gamma)}K_{\gamma,0}(y)\Big| = \sup_{x\in D^2_{t,0}} \frac{1}{|\gamma|}\mathrm{e}^{-\varepsilon|\log y|}|\log y| \to 0. \tag{2.3.10}$$

Therefore, for the first assertion it remains to verify that

$$\begin{aligned} &\sup_{x\in D^2_{t,0}} w_t(x)\Big|(1+\gamma x)^{-1/\gamma-1}K_{\gamma,0}\big((1+\gamma x)^{1/\gamma}\big)\Big| \\ &= \frac{1}{|\gamma|}\sup_{x\in D^2_{t,0}} w_t(x)(1+\gamma x)^{-1/\gamma}\big|\log\big((1+\gamma x)^{-1/\gamma}\big)\big| \end{aligned} \tag{2.3.11}$$

tends to 0. For $\gamma > 0$, (2.3.4) shows that the right-hand side of (2.3.11) is bounded by

$$\begin{aligned} &\frac{1}{|\gamma|}\sup_{x\in D^2_{t,0}} w_t(x)\big(2y^{-1}\big)^{1-\varepsilon/2}\big|\log\big(2y^{-1}\big)\big| \\ &\le \frac{1}{|\gamma|}\sup_{y\ge|A(t)|^{-c}} y^{1-\varepsilon}\big(2y^{-1}\big)^{1-\varepsilon/2}\big|\log\big(2y^{-1}\big)\big| \to 0, \end{aligned}$$

and hence the convergence of (2.3.11) follows. In the case $\gamma < 0$, we can argue likewise. So the first assertion is immediate from (2.3.9)–(2.3.11) and Proposition 2.3.1.

For the second assertion, it suffices to prove that $\{x : (1+\gamma x)^{-1/\gamma} \le |A(t)|^{-c}\} \subset \{x : t\bar{F}(a(t)x+b(t)) \le |A(t)|^{-2c}\}$ eventually, and that $\sup_{x\in D_{t,0}} \tilde{w}_t(x)/w_t(x)$ is bounded when we replace ε with $\varepsilon/2$ in the definition of w_t.

From (2.3.8), we have $\sup_{x\in D^1_{t,0}} \tilde{w}_t(x)/w_t(x) \to 1$, so we must check whether $\sup_{x\in D^2_{t,0}} \tilde{w}_t(x)/w_t(x)$ is bounded.

We only discuss the case $\gamma > 0$, since the arguments are similar for $\gamma < 0$. By the Potter bounds, for all $\eta > 0$ and sufficiently large t,

$$2(1+\gamma x)^{-(1-\eta)/\gamma} \ge y^{-1} = t\bar{F}\big(a(t)x+b(t)\big) \ge \frac{1}{2}(1+\gamma x)^{-1/(\gamma(1-\eta))}$$

uniformly for all $x \in D^2_{t,0}$ (cf. (2.3.4)). Thus

$$\begin{aligned}
&\sup_{x\in D^2_{t,0}} \frac{\tilde{w}_t(x)}{w_t(x)} = \sup_{x\in D^2_{t,0}} \frac{y^{\varepsilon/2-1}}{(1+\gamma x)^{-(1-\varepsilon)/\gamma}}\\
&\le \sup_{x\in D^2_{t,0}} 2^{1-\varepsilon/2}\frac{(1+\gamma x)^{-1/\gamma}}{(1+\gamma x)^{-(1-\varepsilon)/\gamma}} \le 2 \sup_{x\in D^2_{t,0}} (1+\gamma x)^{-\varepsilon/\gamma}\\
&\le 2 \sup_{x\in D^2_{t,0}} (2y^{-1})^{\varepsilon(1-\varepsilon/2)} \to 0.
\end{aligned}$$

Thus $\sup_{x\in D_{t,0}} \tilde{w}_t(x)/w_t(x)$ is bounded for $\gamma > 0$.

Next we verify $\{x : (1+\gamma x)^{-1/\gamma} \le |A(t)|^{-c}\} \subset \{x : t\bar{F}(a(t)x+b(t)) \le |A(t)|^{-2c}\}$. To this end, define $x_t := \inf\{x : t\bar{F}(a(t)x+b(t)) \le |A(t)|^{-2c}\}$. Then by the analog to (2.3.8), $(1+\gamma x_t)^{-1/\gamma} \sim t\bar{F}(a(t)x_t+b(t)) = |A(t)|^{-2c}$. Hence for x satisfying $(1+\gamma x)^{-1/\gamma} \le |A(t)|^{-c}$, we have for sufficient large t, $(1+\gamma x)^{-1/\gamma} < (1+\gamma x_t)^{-1/\gamma}$, which implies $x > x_t$, and $t\bar{F}(a(t)x+b(t)) < t\bar{F}(a(t)x_t+b(t)) = |A(t)|^{-2c}$. Hence we obtain $\{x : (1+\gamma x)^{-1/\gamma} \le |A(t)|^{-c}\} \subset \{x : t\bar{F}(a(t)x+b(t)) \le |A(t)|^{-2c}\}$, and the proof of the second assertion is complete.

Case (iii): $\gamma = \rho = 0$.
In the very same way as for $\rho < 0$, we obtain for all $d > 0$

$$\sup_{\{x:|\log y|\le d|\log|A(t)||\}} |y\mathrm{e}^{-x} - 1| \to 0. \tag{2.3.12}$$

Thus

$$\begin{aligned}
&\sup_{x\in D^1_{t,0}} w_t(x)\Big|\mathrm{e}^{-x}x^2 - \frac{\log^2 y}{y}\Big|\\
&\le \sup_{x\in D^1_{t,0}} \mathrm{e}^{-\varepsilon|\log y|}\Big|y\mathrm{e}^{-x}\log\big(y\mathrm{e}^{-x}\big)\log(\mathrm{e}^{-x}/y) + (y\mathrm{e}^{-x}-1)\log^2 y\Big|\\
&\to 0.
\end{aligned}$$

Moreover, in view of (2.3.12) with $d = 2c$, eventually $-\log y < c \log |A(t)|$ implies $-x < c \log |A(t)|/2$. Hence

$$\begin{aligned}
&\sup_{x \in D_{t,0}^2} w_t(x) \Big| \mathrm{e}^{-x} x^2 - \frac{\log^2 y}{y} \Big| \\
&\le \sup_{x > -c \log |A(t)|/2} \mathrm{e}^{-\varepsilon |x|} x^2 + \sup_{x > -c \log |A(t)|/2} \mathrm{e}^{-\varepsilon |\log y|} \log^2 y \to 0.
\end{aligned}$$

Again the first assertion follows from Proposition 2.3.1.

Finally, in view of (2.3.12), $\mathrm{e}^{-x} < |A(t)|^{-c}$ implies $1/y < |A(t)|^{-2c}$ for sufficiently large t, so that the second assertion is obvious.

The proof of Proposition 2.3.2 is complete. □

2.4 Tail Approximation to the Empirical Distribution Function

For the proof of Theorem 2.2.1, we need two additional Lemmas.

Lemma 2.4.1. *Suppose* $x_0 > -1/(\gamma \vee 0)$.
(i) If $\rho < 0$, *then*

$$\sup_{x_0 \le x < 1/((-\gamma) \vee 0)} \Big| \frac{(1+\gamma x)^{-1/\gamma}}{t \bar{F}(a(t)x + b(t))} - 1 \Big| \ \to \ 0.$$

(ii) If $\rho = 0$ *and* $\gamma \ne 0$, *then for all* $\eta > 0$

$$\sup_{x_0 \le x < 1/((-\gamma) \vee 0)} \frac{t \bar{F}\big(a(t)x + b(t)\big) - (1 - \gamma x)^{-1/\gamma}}{\Big((1+\gamma x)^{-1/\gamma}\Big)^{1-\eta}} \ \to \ 0$$

as $t \to \infty$ *and thus*

$$\sup_{x_0 \le x < 1/((-\gamma) \vee 0)} \frac{t \bar{F}\big(a(t)x + b(t)\big)}{\Big((1+\gamma x)^{-1/\gamma}\Big)^{1-\eta}} \qquad \textit{is bounded.}$$

(iii) If $\gamma = \rho = 0$, *then for all* $\eta, c > 0$

$$\sup_{x_0 \le x < -c \log |A(t)|} \frac{t \bar{F}\big(a(t)x + b(t)\big) - \mathrm{e}^{-x}}{\mathrm{e}^{-(1-\eta)x}} \ \to \ 0$$

as $t \to \infty$ *and so*

$$\sup_{x_0 \le x < -c \log |A(t)|} \frac{t \bar{F}\big(a(t)x + b(t)\big)}{\mathrm{e}^{-(1-\eta)x}} \qquad \textit{is bounded.}$$

Proof. (i). By (2.3.6), one has for all $\delta \in (0,1)$ and $c>0$

$$\Big[x_0,\ \frac{1}{(-\gamma)\vee 0}\Big) \subset \Big\{x:\ (1+\gamma x)^{-1/\gamma} \le \frac{c}{2}t^{-\delta+1}\Big\} \subset D_{t,\rho}$$

for sufficiently large t. Hence, again by (2.3.6),

$$\sup_{x_0\le x<1/((-\gamma)\vee 0)} \Big|\frac{(1+\gamma x)^{-1/\gamma}}{t\bar F\big(a(t)x+b(t)\big)} - 1\Big| \to 0.$$

(ii). By similar arguments as in (i), one concludes $[x_0, 1/((-\gamma)\vee 0)) \subset D_{t,0}$. Hence Proposition 2.3.2 with $\varepsilon = \eta$ implies

$$\begin{aligned}
&\sup_{x_0\le x<1/((-\gamma)\vee 0))} \big((1+\gamma x)^{-1/\gamma}\big)^{\eta-1}\Big|\frac{t\bar F\big(a(t)x+b(t)\big)-(1+\gamma x)^{-1/\gamma}}{A(t)}\\
&\qquad\qquad - (1+\gamma x)^{-1/\gamma-1}K_{\gamma,0}\big((1+\gamma x)^{1/\gamma}\big)\Big|\\
&= \sup_{x_0\le x<1/((-\gamma)\vee 0))} \frac{1}{A(t)}\Big|\frac{t\bar F\big(a(t)x+b(t)\big)-(1+\gamma x)^{-1/\gamma}}{\big((1+\gamma x)^{-1/\gamma}\big)^{1-\eta}}\\
&\qquad\qquad - A(t)\big((1+\gamma x)^{-1/\gamma}\big)^{\gamma+\eta}K_{\gamma,0}\big((1+\gamma x)^{1/\gamma}\big)\Big|\\
&\to\ 0.
\end{aligned}$$

Because $A(t)\to 0$ and $(1+\gamma x)^{-1/\gamma}$ is bounded for $x \ge x_0$, the assertions are immediate from the definition of $K_{\gamma,0}$.

(iii). The proof is similar to the one of (ii). Note that (2.3.12) shows that $w_t(x)/\mathrm{e}^{(1-\varepsilon)x} \to 1$ uniformly for $x_0 \le x < -c\log|A(t)|$. □

Lemma 2.4.2. *Let W denote a Brownian motion.*
(i) If $\gamma \ne 0$ or $\rho < 0$, then

$$\sup_{x_0\le x<1/((-\gamma)\vee 0)} \big((1+\gamma x)^{-1/\gamma}\big)^{-1/2+\varepsilon}\Big|W\big(t\bar F\big(a(t)x+b(t)\big)\big)-W\big((1+\gamma x)^{-1/\gamma}\big)\Big| \to 0 \quad a.s.$$

as $t\to\infty$.
(ii) If $\rho=\gamma=0$, then

$$\sup_{x_0\le x}\Big(\max\big(\mathrm{e}^{-x}, t\bar F\big(a(t)x+b(t)\big)\big)\Big)^{-1/2+\varepsilon}\Big|W\Big(t\bar F\big(a(t)x+b(t)\big)\Big)-W(\mathrm{e}^{-x})\Big| \to 0 \quad a.s.$$

as $t\to\infty$.

Proof. (i). Let $s := (1+\gamma x)^{-1/\gamma}$ and $u(t,s) := t\bar F(a(t)x+b(t)) - s$. Then $x_0 \le x < 1/((-\gamma)\vee 0)$ implies $0 < s \le s_0$, where s_0 is a constant depending on x_0. So we only need to prove

$$\sup_{0<s\le s_0} s^{-1/2+\varepsilon}\big|W\big(u(t,s)+s\big)-W(s)\big| \to 0 \quad a.s. \tag{2.4.1}$$

as $t \to \infty$. From Lemma 2.4.1, one can easily conclude that $s^{\eta-1}u(t,s) \to 0$ uniformly for $s \in (0, s_0]$, and hence in the sequel we may assume $u(t,s) \le s^{1-\eta}$ with $\eta = \varepsilon/(1-\varepsilon)$ for sufficiently large t.

For all $0 < a < 1$

$$\begin{aligned}&\sup_{0<s\le a} s^{-1/2+\varepsilon}|W(u(t,s)+s) - W(s)|\\ &\le \sup_{0<s\le a} s^{-1/2+\varepsilon}\big(s+u(t,s)\big)^{(1-\varepsilon)/2} \cdot \sup_{0<s\le a}\left|\frac{W(u(t,s)+s)}{\big(s+u(t,s)\big)^{(1-\varepsilon)/2}}\right| + \sup_{0<s\le a}\left|\frac{W(s)}{s^{1/2-\varepsilon}}\right|.\end{aligned}$$

Since

$$\lim_{a\to 0}\sup_{0<s\le a} s^{-1/2+\varepsilon}\big(s+u(t,s)\big)^{(1-\varepsilon)/2} = \lim_{a\to 0}\sup_{0<s\le a}\left(s^{2\varepsilon/(1-\varepsilon)} + \frac{u(t,s)}{s^{1-\varepsilon/(1-\varepsilon)}}\right)^{(1-\varepsilon)/2} = 0,$$

the law of iterated logarithm yields

$$\lim_{a\to 0}\sup_{0<s\le a} s^{-1/2+\varepsilon}|W(u(t,s)+s) - W(s)| = 0 \quad a.s.$$

On the other hand, by the continuity of W, for all fixed $a > 0$

$$\lim_{n\to\infty}\sup_{a<s\le s_0} s^{-1/2+\varepsilon}|W(u(t,s)+s) - W(s)| = 0 \quad a.s.$$

as $t \to \infty$. Now assertion (2.4.1) is obvious.

(ii). We consider $x \in [x_0, -c\log|A(t)|)$ and $x \in [-c\log|A(t)|, \infty)$ separately.

As in the proof of (i), one may conclude from Lemma 2.4.1 that

$$\sup_{x_0\le x<-c\log|A(t)|} (\mathrm{e}^{-x})^{-1/2+\varepsilon}\left|W\Big(t\bar{F}\big(a(t)x+b(t)\big)\Big) - W(\mathrm{e}^{-x})\right| \to 0 \qquad a.s. \tag{2.4.2}$$

Since $\mathrm{e}^{-x} \to 0$ and $t\bar{F}(a(t)x+b(t)) \to 0$ uniformly for $x \ge -c\log|A(t)|$, we get

$$\begin{aligned}&\sup_{x\ge -c\log|A(t)|}\Big(\max\big(\mathrm{e}^{-x}, t\bar{F}\big(a(t)x+b(t)\big)\big)\Big)^{-1/2+\varepsilon}\left|W\Big(t\bar{F}\big(a(t)x+b(t)\big)\Big) - W(\mathrm{e}^{-x})\right|\\ &\le \sup_{x\ge -c\log|A(t)|}\Big(t\bar{F}\big(a(t)x+b(t)\big)\Big)^{-1/2+\varepsilon}\left|W\Big(t\bar{F}\big(a(t)x+b(t)\big)\Big)\right|\\ &\quad + \sup_{x\ge -c\log|A(t)|}(\mathrm{e}^{-x})^{-1/2+\varepsilon}|W(\mathrm{e}^{-x})|\\ &\to 0 \quad a.s.\end{aligned} \tag{2.4.3}$$

by the law of the iterated logarithm. A combination of (2.4.2) and (2.4.3) proves the assertion. □

Proof of Theorem 2.2.1.

We focus on the case $\gamma \neq 0$ or $\rho < 0$, because the other case can be treated similarly.

Define

$$\varepsilon^* = \begin{cases} \varepsilon & \text{if } \rho < 0, \\ \varepsilon/2 & \text{if } \rho = 0 \neq \gamma, \end{cases}$$

and

$$\begin{aligned}
I :=& \Big((1+\gamma x)^{-1/\gamma}\Big)^{-1/2+\varepsilon} \Big| E_n(x) - W_n\big((1+\gamma x)^{-1/\gamma}\big) \\
& - \sqrt{k_n} A\big(\frac{n}{k_n}\big)(1+\gamma x)^{-1/\gamma-1} K_{\gamma,\rho}\big((1+\gamma x)^{1/\gamma}\big)\Big| \\
\le & \frac{\big((1+\gamma x)^{-1/\gamma}\big)^{-1/2+\varepsilon}}{\big(\frac{n}{k_n}\bar{F}(a(\frac{n}{k_n})x + b(\frac{n}{k_n}))\big)^{-1/2+\varepsilon^*}} \Big(\frac{n}{k_n}\bar{F}\big(a(\frac{n}{k_n})x + b(\frac{n}{k_n})\big)\Big)^{-1/2+\varepsilon^*} \times \\
& \times \Big| \sqrt{k_n}\Big(\frac{n}{k_n}\bar{F}_n\big(a(\frac{n}{k_n})x + b(\frac{n}{k_n})\big) - \frac{n}{k_n}\bar{F}\big(a(\frac{n}{k_n})x + b(\frac{n}{k_n})\big)\Big) \\
& \quad - W_n\Big(\frac{n}{k_n}\bar{F}\big(a(\frac{n}{k_n})x + b(\frac{n}{k_n})\big)\Big)\Big| \\
& + \frac{\big((1+\gamma x)^{-1/\gamma}\big)^{-1/2+\varepsilon}}{\tilde{w}_t(x)} \tilde{w}_t(x)\sqrt{k_n}A\big(\frac{n}{k_n}\big) \times \\
& \times \left| \frac{\frac{n}{k_n}\bar{F}\big(a(\frac{n}{k_n})x + b(\frac{n}{k_n})\big) - (1+\gamma x)^{-1/\gamma}}{A(\frac{n}{k_n})} - (1+\gamma x)^{-1/\gamma-1}K_{\gamma,\rho}\big((1+\gamma x)^{1/\gamma}\big)\right| \\
& + \big((1+\gamma x)^{-1/\gamma}\big)^{-1/2+\varepsilon}\Big| W_n\Big(\frac{n}{k_n}\bar{F}\big(a(\frac{n}{k_n})x + b(\frac{n}{k_n})\big)\Big) - W_n\big((1+\gamma x)^{-1/\gamma}\big)\Big| \\
:=& I_1 + I_2 + I_3.
\end{aligned}$$

By (2.2.2) (with a and b instead of $\tilde{a}$ and $\tilde{b}$) and Lemma 2.4.1 one has $\sup_{x_0 \le x < 1/((-\gamma)\vee 0)} I_1 \xrightarrow{P} 0$. From Proposition 2.3.2 and the fact that $((1+\gamma x)^{-1/\gamma})^{-1/2+\varepsilon}/\tilde{w}_t(x)$ is bounded uniformly for $x_0 \le x < 1/((-\gamma)\vee 0)$, it follows that $\sup_{x_0 \le x < 1/((-\gamma)\vee 0)} I_2 \to 0$. Finally Lemma 2.4.2 shows that $\sup_{x_0 \le x < 1/((-\gamma)\vee 0)} I_3 \xrightarrow{P} 0$. □

Proof of Corollary 2.2.1.

Because of Theorem 2.2.1(ii) and $\max(1, x^\tau) = o(\mathrm{e}^{(1/2-\varepsilon)x})$ as $x \to \infty$ for all $\tau > 0$ and $\varepsilon \in (0, 1/2)$, it suffices to prove that

$$\sup_{x_0 \le x < \infty} \Big(\frac{n}{k_n}\bar{F}\Big(a\big(\frac{n}{k_n}\big)x + b\big(\frac{n}{k_n}\big)\Big)\Big)^{1/2-\varepsilon} \max(1, x^\tau) = O(1).$$

According to Lemma 2.2 of Resnick (1987), there exists a function $\bar{a}$ such that $a(t)/\bar{a}(t) \to 1$ as $t \to \infty$ and

$$F^t\big(\bar{a}(t)x + b(t)\big) \geq 1 - (1+\delta)^3(1+\delta x)^{-1/\delta}$$

for all $\delta > 0$, sufficiently large t and $x \geq x_0$. Thus, by the mean value theorem, there exists $\theta_{t,x} \in (0,1)$ such that

$$\begin{aligned} t\bar{F}\big(\bar{a}(t)x + b(t)\big) &\leq t\Big(1 - \Big(1 - (1+\delta)^3(1+\delta x)^{-1/\delta}\Big)^{1/t}\Big) \\ &= (1+\delta)^3(1+\delta x)^{-1/\delta}\Big(1 - \theta_{t,x}(1+\delta)^3(1+\delta x)^{-1/\delta}\Big)^{1/t-1} \\ &\leq 2(1+\delta x)^{-1/\delta} \end{aligned}$$

if $x \geq 0$ and $\delta > 0$ is sufficiently small. Since by the locally uniform convergence in (2.1.2)

$$\sup_{x_0 \leq x < 0} \Big(\frac{n}{k_n}\bar{F}\Big(a\big(\frac{n}{k_n}\big)x + b\big(\frac{n}{k_n}\big)\Big)\Big)^{1/2-\varepsilon} \max(1, x^\tau) = O(1),$$

it follows that

$$\begin{aligned} &\sup_{x_0 \leq x < \infty} \Big(\frac{n}{k_n}\bar{F}\Big(a\big(\frac{n}{k_n}\big)x + b\big(\frac{n}{k_n}\big)\Big)\Big)^{1/2-\varepsilon} \max(1, x^\tau) \\ &= O(1) + 2 \sup_{0 \leq x < \infty} \Big(1 + \frac{\delta}{2}x\Big)^{-1/\delta} \max(1, x^\tau) \\ &= O(1) \end{aligned}$$

if δ is chosen smaller than $1/\tau$. □

2.5 Tail Empirical Process With Estimated Parameters: Proofs

In this section we prove the approximation to the tail empirical process with estimated parameters stated in Proposition 2.2.1 and the limit theorem 2.2.2 for the test statistic T_n. To this end, we need a sequence of lemmas.

Define

$$\begin{aligned} A_{n,k_n} &:= \frac{\hat{a}(n/k_n)}{a(n/k_n)}, \\ B_{n,k_n} &:= \frac{\hat{b}(n/k_n) - b(n/k_n)}{a(n/k_n)}, \\ y_n(x) &:= \Big(1 + \gamma\big(B_{n,k_n} + A_{n,k_n}\frac{x^{-\hat{\gamma}_n} - 1}{\hat{\gamma}_n}\big)\Big)^{-1/\gamma}. \end{aligned}$$

Recall from (2.2.6) that

$$\begin{aligned} A_{n,k_n} &= 1 + k_n^{-1/2}\alpha(W_n) + o_P(k_n^{-1/2}), \\ B_{n,k_n} &= k_n^{-1/2}\beta(W_n) + o_P(k_n^{-1/2}), \\ \hat{\gamma}_n &= \gamma + k_n^{-1/2}\Gamma(W_n) + o_P(k_n^{-1/2}). \end{aligned} \tag{2.5.1}$$

Lemma 2.5.1. *Suppose (2.5.1) holds. Let $\lambda_n > 0$ be such that $\lambda_n \to 0$, and $k_n^{-1/2}\lambda_n^{\gamma} \to 0$ if $\gamma < 0$, or $k_n^{-1/2}\log^2\lambda_n \to 0$ if $\gamma = 0$.*
(i) If $\gamma > 0$ then, for all $\varepsilon > 0$, $x^{-1/2+\varepsilon}\big(\sqrt{k_n}(y_n(x) - x) - L_n^{(\gamma)}(x)\big) \xrightarrow{P} 0$, and $x^{\varepsilon-1}(y_n(x) - x) \xrightarrow{P} 0$ as $n \to \infty$ uniformly for $x \in (0,1]$.
(ii) If $-1/2 < \gamma \le 0$ then, for all $\varepsilon > 0$, $x^{-1/2+\varepsilon}\big(\sqrt{k_n}(y_n(x)-x)-L_n^{(\gamma)}(x)\big) \xrightarrow{P} 0$ and $(y_n(x) - x)/x \xrightarrow{P} 0$ as $n \to \infty$ uniformly for $x \in [\lambda_n, 1]$.

Proof. For $\gamma \ne 0$, define $\delta_n := 1 + \gamma B_{n,k_n} - A_{n,k_n}\gamma/\hat{\gamma}_n$, and $\Delta_n := \Delta_{n,x} := \delta_n\hat{\gamma}_n/(\gamma A_{n,k_n}x^{-\hat{\gamma}_n})$, so that $\delta_n = O_P(k_n^{-1/2})$.
(i). By the mean value theorem there exist $\theta_{n,x} \in (0,1)$ such that

$$\begin{aligned} y_n(x) &= \Big(1 + \gamma\big(B_{n,k_n} + A_{n,k_n}\frac{x^{-\hat{\gamma}_n} - 1}{\hat{\gamma}_n}\big)\Big)^{-1/\gamma} \\ &= \Big(\frac{\gamma}{\hat{\gamma}_n}A_{n,k_n}x^{-\hat{\gamma}_n} + \delta_n\Big)^{-1/\gamma} \\ &= \Big(\frac{\gamma}{\hat{\gamma}_n}A_{n,k_n}x^{-\hat{\gamma}_n}\Big)^{-1/\gamma}(1 + \Delta_n)^{-1/\gamma} \\ &= \big(\frac{\gamma}{\hat{\gamma}_n}A_{n,k_n}x^{-\hat{\gamma}_n}\big)^{-1/\gamma} - \frac{1}{\gamma}\Big(\frac{\gamma}{\hat{\gamma}_n}A_{n,k_n}x^{-\hat{\gamma}_n}(1 + \theta_{n,x}\Delta_n)\Big)^{-1/\gamma-1}\delta_n \\ &= \big(\frac{\gamma}{\hat{\gamma}_n}A_{n,k_n}x^{-\hat{\gamma}_n}\big)^{-1/\gamma} - \frac{1}{\gamma}x^{\hat{\gamma}_n(1/\gamma+1)}\delta_n(1 + o_P(1)) \end{aligned} \tag{2.5.2}$$

where the $o_P(1)$-term tends to 0 uniformly for $x \in (0,1]$. Hence again by the mean value theorem and (2.5.1), for some $\theta_{n,x} \in (0,1)$,

$$\begin{aligned} &y_n(x) - x \\ &= \Big(\big(\frac{\gamma}{\hat{\gamma}_n}A_{n,k_n}\big)^{-1/\gamma} - 1\Big)x^{\hat{\gamma}_n/\gamma} + (x^{\hat{\gamma}_n/\gamma} - x) - \frac{1}{\gamma}x^{\hat{\gamma}_n(1/\gamma+1)}\delta_n(1 + o_p(1)) \\ &= -\frac{1}{\gamma}(1 + o_p(1))\big(\frac{\gamma}{\hat{\gamma}_n}A_{n,k_n} - 1\big)x^{\hat{\gamma}_n/\gamma} + x^{1+\theta_{n,x}(\hat{\gamma}_n/\gamma-1)}\log x\big(\frac{\hat{\gamma}_n}{\gamma} - 1\big) \\ &\quad -\frac{1}{\gamma}x^{\hat{\gamma}_n(1/\gamma+1)}\delta_n(1 + o_p(1)) \\ &= \frac{1}{\gamma}(1 + o_p(1))x^{\hat{\gamma}_n/\gamma}\Big(\frac{\hat{\gamma}_n - \gamma}{\hat{\gamma}_n}A_{n,k_n} - (A_{n,k_n} - 1)\Big) + x^{1+\theta_{n,x}(\hat{\gamma}_n/\gamma-1)}\log x\frac{\hat{\gamma}_n - \gamma}{\gamma} \\ &\quad -\frac{1}{\gamma}x^{\hat{\gamma}_n(1/\gamma+1)}\Big(\gamma B_{n,k_n} + \frac{1}{\hat{\gamma}_n}(\hat{\gamma}_n - \gamma) - \frac{\gamma}{\hat{\gamma}_n}(A_{n,k_n} - 1)\Big)(1 + o_p(1)). \end{aligned} \tag{2.5.3}$$

Now the first assertion is a straightforward consequence of (2.5.1). For example,

$$\begin{aligned}
&x^{-1/2+\varepsilon}\sqrt{k_n}\frac{1}{\gamma}(1+o_P(1))x^{\hat\gamma_n/\gamma}\Big(\frac{\hat\gamma_n-\gamma}{\hat\gamma_n}A_{n,k_n}-(A_{n,k_n}-1)\Big)\\
&= \frac{1}{\gamma}x^{-1/2+\varepsilon}\exp\big((\hat\gamma_n/\gamma-1)\log x\big)\times\\
&\qquad\times x\Big(\sqrt{k_n}\frac{\hat\gamma_n-\gamma}{\hat\gamma_n}A_{n,k_n}-\sqrt{k_n}(A_{n,k_n}-1)\Big)(1+o_P(1))\\
&= \frac{1}{\gamma}x^{-1/2+\varepsilon}x\Big(\frac{\Gamma(W_n)}{\gamma}-\alpha(W_n)\Big)(1+o_P(1))
\end{aligned}$$

uniformly for $x\in(0,1]$.

Moreover, in view of (2.5.3),

$$\begin{aligned}
&x^{\varepsilon-1}(y_n(x)-x)\\
&=-\frac{1}{\gamma}(1+o_p(1))\big(\frac{\gamma}{\hat\gamma_n}A_{n,k_n}-1\big)x^{\hat\gamma_n/\gamma-1+\varepsilon}+x^{\varepsilon+\theta_{n,x}(\hat\gamma_n/\gamma-1)}\log x\big(\frac{\hat\gamma_n}{\gamma}-1\big)\\
&\qquad-\frac{1}{\gamma}x^{\hat\gamma_n-1+\varepsilon+\hat\gamma_n/\gamma}\delta_n(1+o_p(1))\\
&\xrightarrow{P}0
\end{aligned}$$

as $n\to\infty$ uniformly for $x\in(0,1]$.

(ii). First we consider the case $\gamma=0$. Then

$$\begin{aligned}
&y_n(x)-x\\
&=\exp\Big(-\big(B_{n,k_n}+A_{n,k_n}\frac{x^{-\hat\gamma_n}-1}{\hat\gamma_n}\big)\Big)-x\\
&=x\Big(\exp\Big(-\big(B_{n,k_n}+A_{n,k_n}\big(\frac{x^{-\hat\gamma_n}-1}{\hat\gamma_n}+\log x\big)-(A_{n,k_n}-1)\log x\big)\Big)-1\Big).
\end{aligned}\tag{2.5.4}$$

An application of the mean value theorem to $\gamma\mapsto x^{-\gamma}$ together with (2.5.1) yields

$$\frac{x^{-\hat\gamma_n}-1}{\hat\gamma_n}=-\log x+\frac{1}{2}\hat\gamma_n\log^2 x\exp\big(-\theta_{n,x}\hat\gamma_n\log x\big)\tag{2.5.5}$$

for some $\theta_{n,x}\in(0,1)$. It follows that

$$\frac{x^{-\hat\gamma_n}-1}{\hat\gamma_n}+\log x\xrightarrow{P}0$$

as $n\to\infty$ uniformly for $x\in[\lambda_n,1]$, since then $k_n^{-1/2}\log^2 x\le k_n^{-1/2}\log^2\lambda_n\to 0$ and likewise $\hat\gamma_n\log x\xrightarrow{P}0$. Hence

$$B_{n,k_n}+A_{n,k_n}\big(\frac{x^{-\hat\gamma_n}-1}{\hat\gamma_n}+\log x\big)-(A_{n,k_n}-1)\log x\xrightarrow{P}0,$$

and by (2.5.1) and (2.5.5)

$$\begin{aligned}
&x^{-1/2+\varepsilon}\sqrt{k_n}(y_n(x)-x)\\
&= -x^{-1/2+\varepsilon}x\Big(\sqrt{k_n}B_{n,k_n}+\sqrt{k_n}A_{n,k_n}\big(\frac{x^{-\hat{\gamma}_n}-1}{\hat{\gamma}_n}+\log x\big)\\
&\qquad -\sqrt{k_n}(A_{n,k_n}-1)\log x\Big)(1+o_P(1))\\
&= -x^{-1/2+\varepsilon}x\Big(\beta(W_n)+\frac{1}{2}\Gamma(W_n)\log^2 x-\alpha(W_n)\log x\Big)(1+o_P(1))
\end{aligned}$$

uniformly for $x\in[\lambda_n,1]$, that is, the first assertion.

Likewise one concludes from (2.5.4), (2.5.1) and (2.5.5) that $(y_n(x)-x)/x$ tends to 0 uniformly for $x\in[\lambda_n,1]$.

Next assume $-1/2<\gamma<0$. Because $\delta_n=O_P(k_n^{-1/2})$ and, by the definition of λ_n and (2.5.1),

$$k_n^{-1/2}x^{\hat{\gamma}_n}\le k_n^{-1/2}\lambda_n^{\hat{\gamma}_n}=o\Big(\exp\big(\log\lambda_n(\hat{\gamma}_n-\gamma)\big)\Big)=o_P(1),$$

$\Delta_n\to 0$ in probability uniformly for $x\in[\lambda_n,1]$. Therefore, the first assertion can be established as in the case $\gamma>0$.

Furthermore, according to (2.5.3),

$$\begin{aligned}
&\frac{y_n(x)-x}{x}\\
&=-\frac{1}{\gamma}(1+o_p(1))\big(\frac{\gamma}{\hat{\gamma}_n}A_{n,k_n}-1\big)x^{\hat{\gamma}_n/\gamma-1}+x^{\theta_{n,x}(\hat{\gamma}_n/\gamma-1)}\log x\big(\frac{\hat{\gamma}_n}{\gamma}-1\big)\\
&\quad-\frac{1}{\gamma}x^{\hat{\gamma}_n+\hat{\gamma}_n/\gamma-1}\delta_n(1+o_p(1))\\
&=-\frac{1}{\gamma}\big(\frac{\gamma}{\hat{\gamma}_n}A_{n,k_n}-1\big)\exp\Big(\frac{\sqrt{k_n}(\hat{\gamma}_n-\gamma)}{\gamma}\frac{\log x}{\sqrt{k_n}}\Big)(1+o_p(1))\\
&\quad+\frac{\log x}{\sqrt{k_n}}\frac{\sqrt{k_n}(\hat{\gamma}_n-\gamma)}{\gamma}\exp\Big(\frac{\theta_{n,x}\sqrt{k_n}(\hat{\gamma}_n-\gamma)}{\gamma}\frac{\log x}{\sqrt{k_n}}\Big)\\
&\quad-\frac{1}{\gamma}\exp\Big(\frac{\sqrt{k_n}(\hat{\gamma}_n-\gamma)}{\gamma}\frac{\log x}{\sqrt{k_n}}+\sqrt{k_n}(\hat{\gamma}_n-\gamma)\frac{\log x}{\sqrt{k_n}}\Big)\frac{x^\gamma}{\sqrt{k_n}}\sqrt{k_n}\delta_n(1+o_p(1))\\
&\xrightarrow{P}0
\end{aligned}$$

as $n\to\infty$ uniformly for $x\in[\lambda_n,1]$ by the choice of λ_n. □

Lemma 2.5.2. *Under the conditions of Lemma 2.5.1 one has for all* $\varepsilon>0$*:*
(i) If $\gamma>0$*, then* $x^{-1/2+\varepsilon}\big(W_n(y_n(x))-W_n(x)\big)\xrightarrow{P}0$ *as* $n\to\infty$ *uniformly for* $x\in(0,1]$.
(ii) If $-1/2<\gamma\le 0$*, then* $x^{-1/2+\varepsilon}\big(W_n(y_n(x))-W_n(x)\big)\xrightarrow{P}0$ *as* $n\to\infty$ *uniformly for* $x\in[\lambda_n,1]$.

Proof. Let $u_n > 0$, $n \in \mathbb{N}$, be an arbitrary sequence converging to 0. According to Lemma 2.5.1 and the law of iterated logarithm

$$x^{-1/2+\varepsilon}\big(W_n(y_n(x)) - W_n(x)\big) = \frac{(y_n(x))^{1/2-\varepsilon/2}}{x^{1/2-\varepsilon}} \frac{W_n(y_n(x))}{(y_n(x))^{1/2-\varepsilon/2}} - \frac{W_n(x)}{x^{1/2-\varepsilon}} \xrightarrow{P} 0$$

uniformly for $x \in (0, u_n]$ if $\gamma > 0$, and uniformly for $x \in [\lambda_n, u_n]$ if $-1/2 < \gamma \leq 0$. Since, due to the continuity of W_n and Lemma 2.5.1,

$$\sup_{u \leq x \leq 1} x^{-1/2+\varepsilon}\big|W_n(y_n(x)) - W_n(x)\big| \xrightarrow{P} 0$$

for all $u \in (0, 1]$, the assertion follows readily. □

Lemma 2.5.3. *Under the conditions of Lemma 2.5.1 one has for all $\varepsilon > 0$:*
(i) For $\gamma > 0$

$$x^{-1/2+\varepsilon}\Big((y_n(x))^{\gamma+1} K_{\gamma,\rho}\big(\frac{1}{y_n(x)}\big) - x^{\gamma+1} K_{\gamma,\rho}\big(\frac{1}{x}\big)\Big) \xrightarrow{P} 0$$

as $n \to \infty$ uniformly for $x \in (0, 1]$.
(ii) For $-1/2 < \gamma \leq 0$

$$x^{-1/2+\varepsilon}\Big((y_n(x))^{\gamma+1} K_{\gamma,\rho}\big(\frac{1}{y_n(x)}\big) - x^{\gamma+1} K_{\gamma,\rho}\big(\frac{1}{x}\big)\Big) \xrightarrow{P} 0$$

as $n \to \infty$ uniformly for $x \in [\lambda_n, 1]$.

Proof. (i). We only consider the case $\gamma > 0 = \rho$; the assertion can be proved similarly in the case $\gamma > 0 > \rho$. Equation (2.5.2) implies

$$\begin{aligned}
\log \frac{y_n(x)}{x} &= O\Big(\log\Big(\frac{\gamma}{\hat{\gamma}_n} A_{n,k_n}\Big)\Big) + O\Big(\Big(\frac{\hat{\gamma}_n}{\gamma} - 1\Big)\log x\Big) + O(\Delta_{n,x}) \\
&= O_P\big(k_n^{-1/2}(1 + |\log x|)\big)
\end{aligned}$$

uniformly for $x \in (0, 1]$. Hence, by the definition of $K_{\gamma,0}$ and Lemma 2.5.1(i),

$$\begin{aligned}
&x^{-1/2+\varepsilon}\Big((y_n(x))^{\gamma+1} K_{\gamma,0}\big(\frac{1}{y_n(x)}\big) - x^{\gamma+1} K_{\gamma,0}\big(\frac{1}{x}\big)\Big) \\
&= x^{-1/2+\varepsilon}\Big(-\frac{y_n(x)\log(y_n(x))}{\gamma} + \frac{x \log x}{\gamma}\Big) \\
&= -\frac{1}{\gamma}\Big(x^{-1/2+\varepsilon} y_n(x) \log \frac{y_n(x)}{x} + x^{-1/2+\varepsilon}(y_n(x) - x)\log x\Big) \\
&= -\frac{1}{\gamma}\Big(x^{\varepsilon-1} y_n(x) x^{1/2} O_P\big(k_n^{-1/2}(1 + |\log x|)\big) + x^{\varepsilon-1}(y_n(x) - x) x^{1/2} \log x\Big) \\
&\xrightarrow{P} 0
\end{aligned}$$

as $n \to \infty$ uniformly for $x \in (0, 1]$.
(ii). In the case $\gamma = 0 > \rho$, according to the definition of $K_{0,\rho}$, Lemma 2.5.1(ii) and the mean value theorem, there exists $\theta_{n,x} \in (0, 1)$ such that

$$\begin{aligned}
& x^{-1/2+\varepsilon}\Big(y_n(x)K_{0,\rho}(\frac{1}{y_n(x)}) - xK_{0,\rho}(\frac{1}{x})\Big) \\
= \;& x^{-1/2+\varepsilon}\Big(\frac{(y_n(x))^{1-\rho}}{\rho} - \frac{x^{1-\rho}}{\rho}\Big) \\
= \;& \frac{1-\rho}{\rho}x^{1/2+\varepsilon}\frac{y_n(x)-x}{x}\big(x + \theta_{n,x}(y_n(x)-x)\big)^{-\rho} \\
\xrightarrow{P} \;& 0
\end{aligned}$$

as $n \to \infty$ uniformly for $x \in [\lambda_n, 1]$.

In the other cases the assertion can be proved likewise. □

Remark 2.5.1. *The part (ii) of Lemma 2.5.1 with weight function $x^{\varepsilon-1-\gamma}$ instead of $x^{-1/2+\varepsilon}$, and of the Lemmas 2.5.2 and 2.5.3 also hold true for $-1 < \gamma \le 0$.*

Lemma 2.5.4. *Suppose $p_n \to 0$, $np_n \to 0$, and $k_n^{-1/2}\log^2(np_n) \to 0$ as $n \to \infty$. Define*

$$\hat{x}_{p_n} := \frac{(\frac{k_n}{np_n})^{\hat{\gamma}_n} - 1}{\hat{\gamma}_n}\hat{a}(\frac{n}{k_n}) + \hat{b}(\frac{n}{k_n}).$$

Then, under the conditions of Proposition 2.2.1 for $-\frac{1}{2} < \gamma \le 0$, $P\{\hat{x}_{p_n} \le X_{n,n}\} \to 0$ as $n \to \infty$.

Proof. According to Theorem 1 of de Haan and Stadtmüller (1996), one has

$$\frac{\frac{a(tx)}{x^\gamma a(t)} - 1}{A(t)} \to \frac{x^\rho - 1}{\rho}$$

as $t \to \infty$. By similar arguments as used by Drees (1998) and Cheng and Jiang (2001) it follows that, for all $0 < \varepsilon < \frac{1}{2}$, there exists $t_\varepsilon > 0$ such that for all $t \ge t_\varepsilon$ and $x \ge 1$

$$\left|\frac{\frac{a(tx)}{x^\gamma a(t)} - 1}{A(t)} - \frac{x^\rho - 1}{\rho}\right| \le \varepsilon x^{\rho+\varepsilon}.$$

Hence

$$\frac{a(n)}{k_n^\gamma a(n/k_n)} = 1 + A\Big(\frac{n}{k_n}\Big)\frac{k_n^\rho - 1}{\rho} + o\Big(A\Big(\frac{n}{k_n}\Big)k_n^{\rho+\varepsilon}\Big) \to 1 \tag{2.5.6}$$

because $\rho \le 0$ and $\sqrt{k_n}A(n/k_n) = O(1)$.

Now, we distinguish two cases.

Case (i): $-1/2 < \gamma < 0$.
Then

$$\begin{aligned}
&\frac{\hat{x}_{p_n} - X_{n,n}}{a(n/k_n)} \\
= \quad & -\frac{1}{\gamma}\Big(\frac{\hat{a}(n/k_n)}{a(n/k_n)} - 1\Big) + \frac{1}{\hat{\gamma}_n}\frac{\hat{a}(n/k_n)}{a(n/k_n)}\Big(\frac{k_n}{np_n}\Big)^{\hat{\gamma}_n} + \frac{\hat{a}(n/k_n)}{a(n/k_n)}\Big(\frac{1}{\gamma} - \frac{1}{\hat{\gamma}_n}\Big) \\
& + \frac{\hat{b}(n/k_n) - b(n/k_n)}{a(n/k_n)} - \Big(\frac{b(n) - b(n/k_n)}{a(n/k_n)} + \frac{1}{\gamma}\Big) - \frac{X_{n,n} - b(n)}{a(n)} \cdot \frac{a(n)}{a(n/k_n)} \\
=: \quad & T_1 + T_2 + T_3 + T_4 - T_5 - T_6.
\end{aligned}$$

Assumption (2.5.1) implies $T_1 + T_3 + T_4 = O_P(k_n^{-1/2}) = o_p(k_n^{\gamma})$ and

$$T_2 = O_P\bigg(\Big(\frac{k_n}{np_n}\Big)^{\gamma} \exp\Big((\hat{\gamma}_n - \gamma)\log\frac{k_n}{np_n}\Big)\bigg) = O_P\bigg(\Big(\frac{k_n}{np_n}\Big)^{\gamma}\bigg) = o_P(k_n^{\gamma})$$

because $np_n \to 0$ and $k_n^{-1/2}\log(np_n) \to 0$.

Since, in view of (2.5.6) and the definition of $b(n)$,

$$\frac{U(n) - b(n)}{a(n/k_n)} = \frac{a(n)}{a(n/k_n)} \cdot \frac{A(n)}{\gamma + \rho} 1_{\{\rho < 0\}} = o(k_n^{\gamma}),$$

approximation (2.2.5) yields

$$T_5 = \frac{k_n^{\gamma} - 1}{\gamma} + o\Big(k_n^{\gamma+\rho+\varepsilon} A\big(\frac{n}{k_n}\big)\Big) + o(k_n^{\gamma}) + \frac{1}{\gamma} = \frac{k_n^{\gamma}}{\gamma} + o(k_n^{\gamma}).$$

Finally, $k_n^{-\gamma} T_6$ converges to G_γ in distribution because of $F \in D(G_\gamma)$ and (2.5.6).

Summing up, one obtains

$$\frac{\hat{x}_{p_n} - X_{n,n}}{k_n^{\gamma} a(n/k_n)} \xrightarrow{d} -\Big(M + \frac{1}{\gamma}\Big)$$

for a G_γ-distributed r.v. M. Now the assertion follows from the fact that $-(M + 1/\gamma) > 0$ a.s.

Case (ii): $\gamma = 0$.

By similar arguments as in the first case one obtains

$$
\begin{aligned}
&\frac{\hat{x}_{p_n} - X_{n,n}}{a(n/k_n)} \\
=\ & \Big(\frac{\big(\frac{k_n}{np_n}\big)^{\hat{\gamma}_n} - 1}{\hat{\gamma}_n} - \log\frac{k_n}{np_n}\Big)\frac{\hat{a}(n/k_n)}{a(n/k_n)} + \Big(\frac{\hat{a}(n/k_n)}{a(n/k_n)} - 1\Big)\log k_n + \frac{\hat{a}(n/k_n)}{a(n/k_n)}\log\frac{1}{np_n} \\
&+ \frac{\hat{b}(n/k_n) - b(n/k_n)}{a(n/k_n)} - \Big(\frac{b(n) - b(n/k_n)}{a(n/k_n)} - \log k_n\Big) - \frac{X_{n,n} - b(n)}{a(n)} \cdot \frac{a(n)}{a(n/k_n)} \\
=\ & o_P(1) + o_P(1) + \log\frac{1}{np_n}(1 + o_P(1)) + O_P(k_n^{-1/2}) + o(1) + O_P(1) \\
=\ & \log\frac{1}{np_n}(1 + o_P(1)) \\
\xrightarrow{P}\ & \infty
\end{aligned}
$$

from which the assertion is obvious. □

Proof of Proposition 2.2.1.

Recall the definition

$$
y_n(x) := \Bigg(1 + \gamma\Big(\frac{\hat{b}(\frac{n}{k_n}) - b(\frac{n}{k_n})}{a(\frac{n}{k_n})} + \frac{\hat{a}(\frac{n}{k_n})}{a(\frac{n}{k_n})}\frac{x^{-\hat{\gamma}_n} - 1}{\hat{\gamma}_n}\Big)\Bigg)^{-1/\gamma}.
$$

Observe that

$$
\begin{aligned}
I :=\ & x^{-1/2+\varepsilon}\Bigg(\sqrt{k_n}\Big[\frac{n}{k_n}\bar{F}_n\Big(\hat{a}\big(\frac{n}{k_n}\big)\frac{x^{-\hat{\gamma}_n} - 1}{\hat{\gamma}_n} + \hat{b}\big(\frac{n}{k_n}\big)\Big) - x\Big] \\
&\qquad - W_n(x) - L_n^{(\gamma)}(x) - \sqrt{k_n}A\big(\frac{n}{k_n}\big)x^{\gamma+1}K_{\gamma,\rho}\big(\frac{1}{x}\big)\Bigg) \\
=\ & \frac{x^{-1/2+\varepsilon}}{(y_n(x))^{-\frac{1+\varepsilon}{2}}}(y_n(x))^{-\frac{1+\varepsilon}{2}}\Bigg(\sqrt{k_n}\Big[\frac{n}{k_n}\bar{F}_n\Big(a\big(\frac{n}{k_n}\big)\frac{(y_n(x))^{-\gamma} - 1}{\gamma} + b\big(\frac{n}{k_n}\big)\Big) - y_n(x)\Big] \\
&\qquad - W_n(y_n(x)) - \sqrt{k_n}A\big(\frac{n}{k_n}\big)(y_n(x))^{\gamma+1}K_{\gamma,\rho}\big(\frac{1}{y_n(x)}\big)\Bigg) \\
&+ x^{-1/2+\varepsilon}\Big(\sqrt{k_n}(y_n(x) - x) - L_n^{(\gamma)}(x)\Big) \\
&+ x^{-1/2+\varepsilon}\Big(W_n(y_n(x)) - W_n(x)\Big) \\
&+ x^{-1/2+\varepsilon}\Big(\sqrt{k_n}A\big(\frac{n}{k_n}\big)(y_n(x))^{\gamma+1}K_{\gamma,\rho}\big(\frac{1}{y_n(x)}\big) - \sqrt{k_n}A\big(\frac{n}{k_n}\big)x^{\gamma+1}K_{\gamma,\rho}\big(\frac{1}{x}\big)\Big) \\
:=\ & I_1 + I_2 + I_3 + I_4
\end{aligned}
$$

Now we distinguish three cases.

Case (i): $\gamma > 0$.
By Lemma 2.5.1(i), $\sup_{x\in(0,1]} x^{-1/2+\varepsilon}/(y_n(x))^{-1/2+\varepsilon/2}$ is stochastically bounded.

Combining this with Theorem 2.2.1, we obtain $\sup_{x\in(0,1]}|I_1| \to 0$ in probability as $n \to \infty$. An application of Lemma 2.5.1(i), Lemma 2.5.2(i), and Lemma 2.5.3(i) gives

$$\sup_{x\in(0,1]}|I_2| \xrightarrow{d} 0, \qquad \sup_{x\in(0,1]}|I_3| \xrightarrow{P} 0, \qquad \sup_{x\in(0,1]}|I_4| \xrightarrow{P} 0,$$

respectively. Hence $\sup_{x\in(0,1]}|I| \to 0$ in probability as $n \to \infty$.

Case (ii): $-1/2 < \gamma < 0$, or $\gamma = 0$ and $\rho < 0$.
Let $\lambda_n := 1/(k_n \log k_n)$. Obviously $\lambda_n \to 0$, $k_n^{-1/2}\lambda_n^{\gamma} \to 0$ and $k_n^{-1/2}\log^2\lambda_n \to 0$ as $n \to \infty$, and hence the Lemmas 2.5.1, 2.5.2 and 2.5.3 apply. Like in case (i), we obtain $\sup_{x\in(\lambda_n,1]}|I| \to 0$ in probability as $n \to \infty$.

It remains to prove that $\sup_{x\in(0,\lambda_n]}|I| \to 0$ in probability. To this end, let $p_n := 1/(n\log k_n)$, and so $np_n \to 0$ and $k_n^{-1/2}\log^2(np_n) \to 0$ as $n \to \infty$. Thus, for $x \in (0, \lambda_n]$,

$$\begin{aligned} z_n(x) :&= \hat{a}(\frac{n}{k_n})\frac{x^{-\hat{\gamma}_n}-1}{\hat{\gamma}_n} + \hat{b}(\frac{n}{k_n}) \\ &\geq \hat{a}(\frac{n}{k_n})\frac{\lambda_n^{-\hat{\gamma}_n}-1}{\hat{\gamma}_n} + \hat{b}(\frac{n}{k_n}) = \hat{a}(\frac{n}{k_n})\frac{\left(\frac{k_n}{np_n}\right)^{\hat{\gamma}_n}-1}{\hat{\gamma}_n} + \hat{b}(\frac{n}{k_n}), \end{aligned}$$

and so by Lemma 2.5.4

$$P\Big\{z_n(x) < X_{n,n} \quad \text{for some } x \in (0,\lambda_n]\Big\} \to 0.$$

Let

$$\tau_n := \sup_{x\in(0,\lambda_n]} x^{-1/2+\varepsilon}\frac{n}{\sqrt{k_n}}\bar{F}_n\Big(\hat{a}(\frac{n}{k_n})\frac{x^{-\hat{\gamma}_n}-1}{\hat{\gamma}_n} + \hat{b}(\frac{n}{k_n})\Big).$$

By the definition of $\bar{F}_n$, $z_n(x) < X_{n,n}$ for some $x \in (0,\lambda_n]$ is equivalent to $\tau \neq 0$. Therefore,

$$P\{\tau_n \neq 0\} \to 0 \tag{2.5.7}$$

as $n \to \infty$.

Furthermore, it is easy to check that

$$\begin{aligned} &x^{-1/2+\varepsilon}\sqrt{k_n}x \to 0, \qquad x^{-1/2+\varepsilon}W_n(x) \xrightarrow{P} 0, \\ &x^{-1/2+\varepsilon}L_n^{(\gamma)}(x) \xrightarrow{P} 0, \quad x^{-1/2+\varepsilon}\sqrt{k_n}A(\frac{n}{k_n})x^{\gamma+1}K_{\gamma,\rho}(\frac{1}{x}) \to 0 \end{aligned} \tag{2.5.8}$$

uniformly for $x \in (0,\lambda_n]$ as $n \to \infty$. For example, the second convergence is an immediate consequence of the law of the iterated logarithm, and in the case

$-1/2 < \gamma < 0$

$$\begin{aligned}
&\sup_{x\in(0,\lambda_n]} x^{-1/2+\varepsilon}|L_n^{(\gamma)}(x)| \\
&\le \sup_{x\in(0,\lambda_n]} \frac{1}{|\gamma|}x^{1/2+\varepsilon}\Big|\frac{1}{\gamma}\Gamma(W_n)-\alpha(W_n)\Big| + \sup_{x\in(0,\lambda_n]} \frac{1}{|\gamma|}|\Gamma(W_n)|x^{1/2+\varepsilon}\log x \\
&\quad + \sup_{x\in(0,\lambda_n]} \frac{1}{|\gamma|}x^{1/2+\gamma+\varepsilon}\Big|\gamma\beta(W_n)+\frac{1}{\gamma}\Gamma(W_n)-\alpha(W_n)\Big| \\
&\xrightarrow{P} 0.
\end{aligned}$$

In view of (2.5.7) and (2.5.8), the assertion $\sup_{x\in(0,\lambda_n]}|I| \to 0$ in probability is immediate.

Case (iii): $\gamma = \rho = 0$.
According to Lemma 2.5.1, $y_n(x)/x \to 1$ in probability uniformly for $x \in [\lambda_n, 1]$ with $\lambda_n := 1/(k_n \log k_n)$, and hence

$$\frac{(1+|\log x|)^\tau}{(1+|\log y_n(x)|)^\tau} = \Big(\frac{1+|\log x|}{1+|\log x|+o_P(1)}\Big)^\tau = O_P(1)$$

uniformly for $x \in [\lambda_n, 1]$. Therefore, one can argue as in case (ii) (using Corollary 2.2.1 instead of Theorem 2.2.1) to establish the assertion. □

Proof of Theorem 2.2.2.
By Proposition 2.2.1 one has

$$\begin{aligned}
&\Big(\sqrt{k_n}\Big[\frac{n}{k_n}\bar F_n\Big(\hat a\big(\frac{n}{k_n}\big)\frac{x^{-\hat\gamma_n}-1}{\hat\gamma_n}+\hat b\big(\frac{n}{k_n}\big)\Big)-x\Big]\Big)^2 \\
&= \Big(W_n(x)+L_n^{(\gamma)}(x)+\sqrt{k_n}A\big(\frac{n}{k_n}\big)x^{\gamma+1}K_{\gamma,\rho}\big(\frac{1}{x}\big)+\frac{o_p(1)}{h(x)}\Big)^2.
\end{aligned} \tag{2.5.9}$$

Using the law of iterated logarithm, it is readily checked that

$$\begin{aligned}
\int_0^1 \big(W_n(x)+L_n^{(\gamma)}\big)^2 x^{\eta-2}\,dx &= O_P(1) \\
\int_0^1 \Big(x^{\gamma+1}K_{\gamma,\rho}\big(\frac{1}{x}\big)\Big)^2 x^{\eta-2}\,dx &< \infty \\
\int_0^1 \frac{x^{\eta-2}}{h^2(x)}\,dx &< \infty
\end{aligned}$$

for $\eta > 0$, and $\eta \ge 1$ if $\gamma = \rho = 0$. Hence the assertion is an immediate consequence of (2.5.9) and $\sqrt{k_n}A(n/k_n) \to 0$. □

$\gamma =$ / p	2	1.5	1	0.5	0.25	0	−0.25	−0.375	−0.499
0.995	0.545	0.513	0.507	0.525	0.553	0.621	0.672	0.739	0.909
0.99	0.477	0.462	0.459	0.474	0.494	0.554	0.604	0.667	0.795
0.975	0.408	0.389	0.383	0.390	0.409	0.459	0.510	0.558	0.657
0.95	0.349	0.337	0.330	0.337	0.355	0.390	0.431	0.468	0.552
0.9	0.289	0.281	0.278	0.285	0.295	0.318	0.355	0.381	0.444
0.8	0.231	0.227	0.224	0.229	0.239	0.254	0.280	0.299	0.343
0.7	0.197	0.193	0.191	0.195	0.201	0.213	0.235	0.253	0.286
0.6	0.171	0.168	0.166	0.169	0.175	0.185	0.204	0.217	0.243
0.5	0.151	0.148	0.147	0.149	0.154	0.162	0.178	0.189	0.211
0.4	0.132	0.131	0.130	0.132	0.136	0.144	0.157	0.164	0.183
0.3	0.116	0.114	0.114	0.116	0.120	0.126	0.135	0.144	0.158
0.2	0.100	0.099	0.098	0.100	0.103	0.108	0.116	0.122	0.134
0.1	0.083	0.082	0.081	0.082	0.085	0.089	0.095	0.099	0.106
0.05	0.071	0.070	0.070	0.071	0.073	0.078	0.080	0.083	0.090
0.025	0.062	0.062	0.062	0.063	0.064	0.068	0.071	0.073	0.078
0.01	0.053	0.054	0.054	0.055	0.056	0.059	0.060	0.062	0.067
0.005	0.048	0.049	0.049	0.050	0.051	0.052	0.054	0.055	0.060

Table 2.1: Quantiles $Q_{p,\gamma}$ of the limit distribution of k_nT_n.

2.6 Simulations

First we want to calculate the limiting distribution of the test statistic k_nT_n defined by (2.1.4), where we use the maximum likelihood estimator $\hat\gamma_n$, $\hat a(n/k_n)$ and $\hat b(n/k_n)$ described in Example 2.2.1. Here we have chosen $\eta = 1$, thus giving maximal weight to deviations in the extreme tail region that is possible in the framework of Theorem 2.2.2 for all values of $\gamma > -1/2$.

To simulate $\int_0^1 (W_n(x) + L_n^{(\gamma)}(x))^2 x^{-1}\,dx$, the Brownian motion W_n on the unit interval is simulated on a grid with 50 000 points. Then the integral is approximated by a Riemann sum for the extreme value indices $\gamma = 2$, 1.5, 1, 0.5, 0.25, 0, −0.25, −0.375 and −0.499. Note that for $\gamma < -1/2$ the term $L_n^{(\gamma)}$ is not defined since the integral $S_n = \int_0^1 t^{\gamma-1} W_n(t)\,dt$ defined in Example 2.2.1 may not exist. The empirical quantiles of the integral statistic obtained in 20 000 runs are reported in Table 2.1. It is not surprising that the extreme upper quantiles increase rapidly as $\gamma < 0$ decreases, since $|S_n| \to \infty$ in probability as $\gamma \downarrow -1/2$, and thus the limit distribution of k_nT_n converges weakly to ∞, too.

Next we investigate the finite sample behavior of the test described in section 2.2, that rejects the hypothesis that $F \in D(G_\gamma)$ for some $\gamma > -1/2$ if k_nT_n exceeds $\hat Q_{1-\bar\alpha,\tilde\gamma_n}$. Here we use the maximum likelihood estimator for γ also as the pilot estimator, that is, $\tilde\gamma_n = \hat\gamma_n$; the estimates are calculated using Grimshaw's (1993) algorithm. Since we have approximately determined the quantiles $Q_{p,\gamma}$ only for 9 different values of γ, we use linear interpolation to approximate the quantiles for intermediate values of γ, that is, for $\tilde{\gamma}_n \in [\gamma_1, \gamma_2]$ we define

$$\hat{Q}_{p,\tilde{\gamma}_n} = Q_{p,\gamma_1} + \frac{\tilde{\gamma}_n - \gamma_1}{\gamma_2 - \gamma_1}(Q_{p,\gamma_2} - Q_{p,\gamma_1})$$

where Q_{p,γ_i} denote the quantiles given in Table 2.1. Moreover, we define $\hat{Q}_{p,\tilde{\gamma}_n} := Q_{p,2}$ if $\tilde{\gamma}_n > 2$.

As usually in extreme value theory, the choice of the number k_n of order statistics used for the inference is a crucial point. Here we consider $k_n = 25, 50, \ldots, 150$ for sample size $n = 200$, and $k_n = 25, 50, \ldots, 400$ for sample size $n = 1000$.

We have drawn 1000 samples from each of the following distribution functions belonging to the domain of attraction of G_γ for some $\gamma > -1/2$:

- Cauchy distribution $\quad (\gamma = 1,\ \rho = -2)$:

  $$F(x) = \frac{1}{2} + \frac{1}{\pi}\arctan x, \quad x \in \mathbb{R}.$$

- Burr(β, τ, λ) distribution $\quad (\gamma = 1/(\tau\lambda),\ \rho = -1/\lambda)$:

  $$F(x) = 1 - \left(\frac{\beta}{\beta + x^\tau}\right)^\lambda, \quad x > 0,$$

  with $(\beta, \tau, \lambda) = (1, 2, 2)$.

- Extreme value distribution $EV(\gamma) \quad (\gamma \in \mathbb{R}, \rho = -1)$:

  $$F(x) = \exp\big(- (1 + \gamma x)^{-1/\gamma}\big), \quad 1 + \gamma x > 0,$$

  with $\gamma = 0.25$ and $\gamma = 0$.

- Weibull(λ, τ) distribution $\quad (\gamma = 0, \rho = 0)$:

  $$F(x) = 1 - \exp(-\lambda x^\tau), \quad x > 0,$$

  with $(\lambda, \tau) = (1, 0.5)$.

- Reversed Burr(β, τ, λ) distribution $\quad (\gamma = -1/(\tau\lambda),\ \rho = -1/\lambda)$:

  $$F(x) = 1 - \left(\frac{\beta}{\beta + (x^* - x)^{-\tau}}\right)^\lambda, \quad x < x^*,$$

  with $(\beta, \tau, \lambda) = (1, 4, 1)$ and $x^* = 1$.

In some simulations either there exists no solution to the likelihood equations, or the maximum likelihood estimate of γ is less than $-1/2$, so that the test cannot be applied. The relative frequency of simulations in which this happened are given in the Tables 2.2–2.5; for all other values of k_n not mentioned in these tables, the test could be performed in at least 99.6% of the simulations, except for the reversed Burr distribution and sample size n=1000 where in up to 2% of the simulations the estimate for the extreme value index was less than $-1/2$ in the cases not mentioned in Table 2.5. For this distribution, one gets estimates of γ less than $-1/2$ in at least 1% of the simulations for all values of k_n and in more than 30% of all simulations with $n = 200$, while for all other distributions this happened only if a small proportion of the data is used for the inference. It is obvious that the problem of pilot estimates of γ being smaller than $-1/2$ becomes more and more acute as the true extreme value index approaches $-1/2$; this is particularly true for small sample sizes.

In the Tables 2.6 and 2.7 the empirical size of the test with nominal size $\bar{\alpha} = 0.05$ is reported, that is, the relative frequency of simulations in which the hypothesis is rejected. These frequencies are based only on those simulations in which the test could actually be applied. In addition, in 10 000 simulations for each d.f. and each k_n, we determine the number of order statistics for which the maximum likelihood estimator of γ has minimal mean squared error. The corresponding empirical sizes are given in bold face.

With the exception of the Weibull distribution for $n = 1000$, the number of order statistics which is optimal for the maximum likelihood estimator of γ is approximately equal to the value of k_n where the empirical size of the test starts to grow rapidly, while for smaller values of k_n the empirical size is quite close to the nominal value. This indicates that the test statistic T_n can indeed be used to choose the sample fraction on which extreme value estimators are based.

This conclusion is also supported by Figure 2.1 that displays both the empirical size of the test and the mean squared error of $\hat{\gamma}_n$ versus k for the Cauchy distribution and sample size $n = 1000$. The mean squared error is minimal for about $k = 250$ which is also the point where the empirical size increases sharply.

At first glance, it might be surprising that, unlike estimators of γ, the test behaves equally well for small and large values of $|\rho|$. However, recall that for the actual size to be close to the nominal value it is not important how accurate the estimators are but only how precise the Gaussian approximation for the tail empirical distribution function with estimated parameters is. While the rate of convergence of estimators of the extreme value index deteriorates as ρ tends to 0, this is not necessarily true for the accuracy of the normal approximation.

Acknowledgment: Part of the work of Holger Drees and Laurens de Haan was done while visiting the Stochastics Center at Chalmers University Gothenburg. Grateful acknowledgement is made for hospitality particularly to Holger Rootzén. During that time Holger Drees was supported by the Euro-

k_n	Cauchy	Burr(1,2,2)	EV(0.25)	EV(0)	Weibull(1,0.5)	Rev. Burr(1,4,1)
25	0.000	0.0003	0.002	0.018	0.004	0.090

Table 2.2: Relative frequency of simulations in which no maximum likelihood estimate was found for sample size $n = 200$.

k_n	Cauchy	Burr(1,2,2)	EV(0.25)	EV(0)	Weibull(1,0.5)	Rev. Burr(1,4,1)
25	0.000	0.013	0.035	0.097	0.014	0.343
50	0.000	0.000	0.000	0.025	0.001	0.317

k_n	75	100	125	150
Rev. Burr(1,4,1)	0.351	0.523	0.796	0.973

Table 2.3: Relative frequency of simulations in which $\hat{\gamma}_n < -0.5$ for sample size $n = 200$.

k_n	Cauchy	Burr(1,2,2)	EV(0.25)	EV(0)	Weibull(1,0.5)	Rev. Burr(1,4,1)
25	0.000	0.001	0.006	0.026	0.005	0.073

Table 2.4: Relative frequency of simulations in which no maximum likelihood estimate was found for sample size $n = 1000$.

k_n	Cauchy	Burr(1,2,2)	EV(0.25)	EV(0)	Weibull(1,0.5)	Rev. Burr(1,4,1)
25	0.001	0.004	0.020	0.091	0.025	0.274
50	0.000	0.000	0.002	0.012	0.000	0.157
75	0.000	0.000	0.000	0.004	0.000	0.077

k_n	75	100	125	150
Rev. Burr(1,4,1)	0.077	0.043	0.031	0.020

Table 2.5: Relative frequency of simulations in which $\hat{\gamma}_n < -0.5$ for sample size $n = 1000$.

k_n	Cauchy	Burr(1,2,2)	EV(0.25)	EV(0)	Weibull(1,0.5)	Rev. Burr(1,4,1)
25	0.044	0.034	0.036	0.024	**0.042**	0.018
50	0.055	0.044	0.046	0.028	0.056	**0.035**
75	**0.083**	**0.061**	0.054	0.034	0.087	0.065
100	0.253	0.065	**0.074**	**0.061**	0.141	0.177
125	0.721	0.099	0.107	0.116	0.264	0.505
150	0.990	0.180	0.183	0.259	0.517	0.778

Table 2.6: Empirical size the one-sided test with nominal size $\bar{\alpha} = 0.05$ for sample size $n = 200$.

k_n	Cauchy	Burr(1,2,2)	EV(0.25)	EV(0)	Weibull(1,0.5)	Rev. Burr(1,4,1)
25	0.045	0.033	0.031	0.032	0.035	0.023
50	0.047	0.036	0.040	0.044	**0.046**	0.021
75	0.051	0.038	0.028	0.043	0.041	0.030
100	0.058	0.039	0.051	0.049	0.047	0.031
125	0.047	0.042	0.049	0.044	0.059	0.043
150	0.053	0.047	0.054	0.048	0.082	0.042
175	0.056	0.050	0.061	0.042	0.086	**0.040**
200	0.055	0.047	0.058	0.049	0.098	0.062
225	0.059	**0.052**	0.060	0.054	0.110	0.066
250	**0.073**	0.050	0.060	0.052	0.120	0.098
275	0.100	0.064	0.061	**0.052**	0.130	0.119
300	0.115	0.068	0.058	0.063	0.158	0.147
325	0.146	0.068	0.073	0.063	0.204	0.189
350	0.203	0.065	**0.068**	0.081	0.235	0.273
375	0.280	0.068	0.071	0.088	0.275	0.350
400	0.345	0.093	0.083	0.113	0.319	0.445

Table 2.7: Empirical size of the one-sided test with nominal size $\bar{\alpha} = 0.05$ for sample size $n = 1000$.

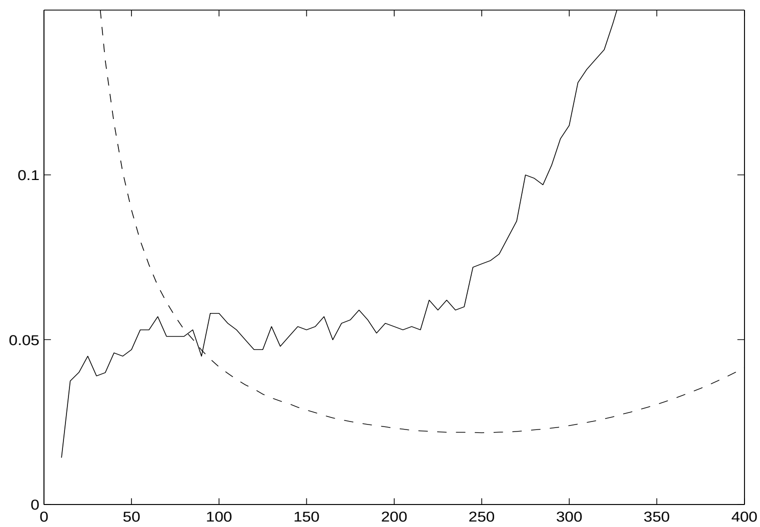

Figure 2.1: Empirical size of the test with nominal size $\bar{\alpha} = 0.05$ (solid line) and the mean squared error of $\hat{\gamma}_n$ (dashed line) as a function of k_n for Cauchy samples of size $n = 1000$.

pean Union TMR grant ERB-FMRX-CT960095. His work was also partly supported by the Netherlands Organization for Scientific Research through the Netherlands Mathematical Research Foundation and by the Heisenberg program of the DFG. Laurens de Haan's research was partially supported by CEAUL/FCT/POCTI/FEDER.

Chapter 3

On Large Deviation for Extremes

co-authors: Holger Drees and Laurens de Haan

Abstract. Recently a weighted approximation for the tail empirical distribution function has been developed (chapter 2). We show that the same result can also be used to improve a known uniform approximation of the distribution of the maximum of a random sample. From this a general result about large deviations of this maximum is derived. In addition, the relationship between two second order conditions used in extreme value theory is clarified.

3.1 Introduction

Let $\{X_n, n \geq 1\}$ be independent identically distributed random variables with common distribution function $F(x)$. Suppose F is in the domain of attraction of the extreme value distribution with index $\gamma \in \mathbb{R}$

$$G_\gamma(x) := \exp\left(-(1+\gamma x)^{-1/\gamma}\right), \quad 1+\gamma x > 0,$$

that is, there exist normalizing constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$P(M_n \leq a_n x + b_n) \to G_\gamma(x), \qquad x \in \mathbb{R}, \tag{3.1.1}$$

as $n \to \infty$, where $M_n := \max(X_1, X_2, ..., X_n)$.

Since the limit function G_γ is continuous, we have

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F^n(a_n x + b_n) - G_\gamma(x)| = 0.$$

Cheng and Jiang (2001) proved that under the second order strengthening (3.1.4) of condition (3.1.1) one can find a sequence $A(n)$ satisfying $A(n) \to 0$, as

$n \to \infty$ and A is regularly varying with index $\rho \le 0$ and normalizing constants $\tilde{a}_n > 0$ and $\tilde{b}_n \in \mathbb{R}$ such that

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \left| \frac{F^n(\tilde{a}_n x + \tilde{b}_n) - G_\gamma(x)}{A(n)} + (1+\gamma x)^{-1/\gamma - 1} G_\gamma(x) \bar{H}_{\gamma,\rho}((1+\gamma x)^{1/\gamma}) \right| = 0, \tag{3.1.2}$$

where the function $\bar{H}_{\gamma,\rho}$ is defined in (3.3.1) below, and where the last item is defined by continuity when $1+\gamma x \to 0$. De Haan and Resnick (1996) established a similar approximation under a somewhat stronger second order condition. We will show that, under the same condition but with slightly different normalizing constants, a weighted version of this result holds, that is more accurate for values of x close to the right endpoint $1/(-\gamma) \vee 0)$ of G_γ, that is, ∞ if $\gamma \ge 0$ and $1/(-\gamma)$ if $\gamma < 0$.

From this result it is easily deduced that

$$\lim_{n\to\infty} \frac{1 - F^n(a_n x_n + b_n)}{1 - G_\gamma(x_n)} = 1 \tag{3.1.3}$$

for all sequences $x_n \uparrow 1/((-\gamma) \vee 0)$. Convergence (3.1.3), which can be considered a result about large deviations of the maximum M_n from its 'typical' behavior, was also studied in Section 2.3 of the monograph by Resnick (1987). There, for different normalizing constants a_n and b_n, quite complicated necessary and sufficient conditions on the maximal rate at which x_n may tend to $1/((-\gamma) \vee 0)$ were given such that (3.1.3) holds. In contrast, for our choice of the normalizing constants, the large deviations result (3.1.3) holds for *all* sequences $x_n \uparrow 1/((-\gamma) \vee 0)$, provided the second order condition (3.1.4) is met with $\rho < 0$.

Condition (3.1.1) is equivalent to the existence of a positive function a^* such that

$$\lim_{t\to\infty} \frac{V(tx) - V(t)}{a^*(t)} = \frac{x^\gamma - 1}{\gamma}$$

for all $x > 0$, where the function V is defined as a generalized inverse:

$$V(t) := \left(\frac{1}{-\log F}\right)^{\leftarrow}(t) = F^{\leftarrow}(\mathrm{e}^{-1/t}).$$

Cheng and Jiang (2001) proved that the following second order condition is necessary for a uniform approximation of type (3.1.2): there exists a (positive or negative) function A^* and a parameter $\rho \le 0$ such that for all $x > 0$

$$\lim_{t\to\infty} \frac{\frac{V(tx)-V(t)}{a^*(t)} - \frac{x^\gamma - 1}{\gamma}}{A^*(t)} = H_{\gamma,\rho}(x) := \begin{cases} \frac{x^{\gamma+\rho}-1}{\gamma+\rho}, & \rho < 0,\ \gamma+\rho \ne 0, \\ \log x, & \rho < 0,\ \gamma + \rho = 0, \\ \frac{1}{\gamma} x^\gamma \log x, & \rho = 0 \ne \gamma, \\ \frac{1}{2}\log^2 x, & \rho = 0 = \gamma. \end{cases} \tag{3.1.4}$$

Then there exist functions a and A satisfying $a(t) \sim a^*(t)$ and $A(t) \sim A^*(t)$ as $t \to \infty$, such for all $\epsilon > 0$ there exists a constant $t_\epsilon > 0$ such that for all t, $tx \geq t_\epsilon$

$$x^{-(\gamma+\rho)} e^{-\epsilon|\log x|} \left| \frac{\frac{V(tx)-V(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma}}{A(t)} - H_{\gamma,\rho}(x) \right| < \epsilon \tag{3.1.5}$$

(see Drees (1998)). Cheng and Jiang (2001) gave explicit representations of the functions a and A in terms of F. Under this second order condition, following the lines of chapter 2, one may prove a weighted approximation to the tail of the empirical distribution function which will be central for the proof of our main result.

While here we work with a second order condition for the function V, usually the analogous condition for $U := (1/1-F)^{\leftarrow}$ is considered. The relationship between these two conditions is clarified in the Appendix.

3.2 Main results

Our main result is a weighted approximation to the normalized distribution function $F^n(a_n x + b_n)$ of the maximum M_n where the additive constant b_n is chosen equal to $V(n)$. Before stating the main result we make the following conventions:

$$(1+\gamma x)^{1/\gamma} = \begin{cases} \exp(x) & \text{if } \gamma = 0, \\ 0 & \text{if } \gamma > 0 \text{ and } 1+\gamma x \leq 0, \\ \infty & \text{if } \gamma < 0 \text{ and } 1+\gamma x \leq 0. \end{cases}$$

By this convention, we have

$$G_\gamma(x) = \begin{cases} 0 & \text{if } \gamma > 0 \text{ and } 1+\gamma x \leq 0, \\ 1 & \text{if } \gamma < 0 \text{ and } 1+\gamma x \leq 0. \end{cases}$$

Theorem 3.2.1. *Assume that V satisfies (3.1.4) with $\gamma \neq 0$ or $\rho < 0$. Define for $n \in \mathbb{N}$*

$$\begin{aligned} a_n &:= \begin{cases} a(n)\Big(1 + \frac{\gamma}{\gamma+\rho} A(n)\Big) & \textit{if } \rho < 0, \gamma+\rho \neq 0, \\ a(n) & \textit{otherwise,} \end{cases} \\ b_n &:= V(n) \end{aligned}$$

and

$$\tilde{H}_{\gamma,\rho}(x) := \begin{cases} \frac{x^{\gamma+\rho} - x^\gamma}{\gamma+\rho} & \textit{if } \rho < 0, \gamma+\rho \neq 0, \\ H_{\gamma,\rho}(x) & \textit{otherwise.} \end{cases}$$

Then for each $\epsilon > 0$

$$\sup_{x<\frac{1}{(-\gamma)\vee 0}} \max\left(1, \left((1+\gamma x)^{1/\gamma}\right)^{1-\epsilon I_{\{\rho=0\}}}\right)\times$$
$$\times \left| \frac{F^n(a_n x+b_n)-G_\gamma(x)}{A(n)} + G_\gamma(x)(1+\gamma x)^{-1/\gamma-1}\tilde{H}_{\gamma,\rho}((1+\gamma x)^{1/\gamma})\right| = o(1)$$

as $n \to \infty$.

The proof of Theorem 3.2.1 is based on a similar approximation where a less natural additive constant $\bar{b}_n$ and the scaling constant $\bar{a}_n = a(n)$ are used (see Proposition 3.3.1). At first glance, seemingly one has to pay for the natural choice $b_n = V(n)$ by a more complicated scaling constant a_n. However, (3.1.5) also holds when $a(t)$ is replaced with $a(t)(1+\gamma/(\gamma+\rho)A(t))$ in the case $\rho < 0, \gamma+\rho \neq 0$ and $H_{\gamma,\rho}$ is replaced with $\tilde{H}_{\gamma,\rho}$. Hence there is nothing special about the normalizing function a (and hence also about $\bar{a}_n = a(n)$), but its particular form is only due to the quite arbitrary choice of the limiting function $H_{\gamma,\rho}$ often considered in the literature.

From the weighted approximation established in Theorem 3.2.1, results on the relative error of the extreme value approximation of F^n and on large deviations follow readily:

Corollary 3.2.1. *Under the conditions of Theorem 3.2.1 with $\rho < 0$ one has*

$$\sup_{-\frac{1}{\gamma\vee 0}<x<\frac{1}{(-\gamma)\vee 0}} \left| \frac{\frac{1-F^n(a_n x+b_n)}{1-G_\gamma(x)}-1}{A(n)} - \frac{G_\gamma(x)}{1-G_\gamma(x)}(1+\gamma x)^{-1/\gamma-1}\tilde{H}_{\gamma,\rho}((1+\gamma x)^{1/\gamma})\right| = o(1)$$

as $n \to \infty$. In particular,

$$\lim_{n\to\infty} \frac{1-F^n(a_n x_n+b_n)}{1-G_\gamma(x_n)} = 1$$

for any sequence $x_n \uparrow 1/((-\gamma)\vee 0)$.

In fact the restriction $x < 1/((-\gamma)\vee 0)$ in Theorem 3.2.1 is not essential:

Remark 3.2.1. *If $\gamma < 0$, then*

$$\sup_{x\geq -1/\gamma} \left| \frac{F^n(a_n x+b_n)-G_\gamma(x)}{A(n)} + G_\gamma(x)(1+\gamma x)^{-1/\gamma-1}\tilde{H}_{\gamma,\rho}((1+\gamma x)^{1/\gamma})\right|$$

is zero for sufficiently large n.

3.3 Proofs

The main ingredient of the proof of Theorem 3.2.1 is an approximation similar to the one asserted in Theorem 3.2.1 but using different normalizing constants and, as a consequence, the following modification of the limiting function:

$$\bar{H}_{\gamma,\rho}(x) := \begin{cases} \frac{x^{\gamma+\rho}}{\gamma+\rho} & \text{if } \rho < 0, \gamma+\rho \neq 0, \\ H_{\gamma,\rho}(x) & \text{otherwise.} \end{cases} \tag{3.3.1}$$

Proposition 3.3.1. *Suppose that not $\gamma = \rho = 0$. Let*

$$\begin{aligned} \bar{a}_n &:= a(n), \\ \bar{b}_n &:= \begin{cases} V(n) - \frac{1}{\gamma+\rho} a(n) A(n) & \textit{if } \rho < 0, \gamma+\rho \neq 0, \\ V(n) & \textit{otherwise.} \end{cases} \end{aligned}$$

Then, under the conditions of Theorem 3.2.1,

$$\sup_{(1+\gamma x)^{-1/\gamma} \leq -\log A^2(n)} \max\left(1, \left((1+\gamma x)^{1/\gamma}\right)^{1-\rho-\varepsilon}\right) \times$$
$$\times \left| \frac{F^n(\bar{a}_n x + \bar{b}_n) - G_\gamma(x)}{A(n)} + G_\gamma(x)(1+\gamma x)^{-1/\gamma-1} \bar{H}_{\gamma,\rho}((1+\gamma x)^{1/\gamma}) \right| = o(1)$$

as $n \to \infty$.

Proof. By the very same arguments as used in the proof of the Propositions 2.3.1 and 2.3.2 (see chapter 2), one obtains

$$\sup_{x \in \tilde{D}_{n,\rho}} w(x) \left| \frac{n(-\log F(\bar{a}_n x + \bar{b}_n)) - (1+\gamma x)^{-1/\gamma}}{A(n)} \right.$$
$$\left. - (1+\gamma x)^{-1/\gamma-1} \bar{H}_{\gamma,\rho}((1+\gamma x)^{1/\gamma}) \right| = o(1)$$

with $w(x) := \left((1+\gamma x)^{1/\gamma}\right)^{1-\rho-\varepsilon}$ and

$$\tilde{D}_{n,\rho} := \begin{cases} \{x : (1+\gamma x)^{-1/\gamma} \leq c n^{-\delta+1}\} & \rho < 0, \\ \{x : (1+\gamma x)^{-1/\gamma} \leq |A(n)|^{-c}\} & \rho = 0. \end{cases}$$

In particular,

$$\sup_{(1+\gamma x)^{-1/\gamma} \leq -\log A^2(n)} w(x) \left| \frac{n(-\log F(\bar{a}_n x + \bar{b}_n)) - (1+\gamma x)^{-1/\gamma}}{A(n)} \right.$$
$$\left. - (1+\gamma x)^{-1/\gamma-1} \bar{H}_{\gamma,\rho}((1+\gamma x)^{1/\gamma}) \right| = o(1).$$

This implies

$$\begin{aligned} F^n(\bar{a}_n x + \bar{b}_n) &= \exp\left(n \log F(\bar{a}_n x + \bar{b}_n)\right) \\ &= \exp\left(-(1+\gamma x)^{-1/\gamma} - A(n)\Phi(x) - o(1)\frac{A(n)}{w(x)}\right) \\ &= G_\gamma(x) \exp\left(-A(n)\Phi(x) - o(1)\frac{A(n)}{w(x)}\right) \end{aligned} \tag{3.3.2}$$

as $n \to \infty$, where the $o(1)$-term is uniform in x and

$$\Phi(x) := (1+\gamma x)^{-1/\gamma-1}\bar{H}_{\gamma,\rho}((1+\gamma x)^{1/\gamma})$$
$$= \begin{cases} \frac{(1+\gamma x)^{(\rho-1)/\gamma}}{\gamma+\rho} & \rho<0, \gamma+\rho\neq 0 \\ (1+\gamma x)^{-1/\gamma-1}\log((1+\gamma x)^{1/\gamma}) & \rho<0, \gamma+\rho=0 \\ \frac{1}{\gamma}(1+\gamma x)^{-1/\gamma}\log((1+\gamma x)^{1/\gamma}) & \rho=0\neq\gamma. \end{cases}$$

Now let's look at the item

$$A(n)\Phi(x)+o(1)\frac{A(n)}{w(x)}.$$

In case of $\rho<0, \gamma+\rho\neq 0$,

$$\begin{aligned} |A(n)\Phi(x)| &= |A(n)|\cdot|\frac{(1+\gamma x)^{(\rho-1)/\gamma}}{\gamma+\rho}| \\ &\leq \frac{1}{|\gamma+\rho|}|A(n)|\left(-\log A^2(n)\right)^{1-\rho} \\ &\to 0 \end{aligned}$$

as $n \to \infty$ uniformly in x. Further for $\varepsilon < 1-\rho$

$$\begin{aligned} |\frac{A(n)}{w(x)}| &= |A(n)|\cdot\left((1+\gamma x)^{-1/\gamma}\right)^{1-\rho-\varepsilon} \\ &\leq |A(n)|\left(-\log A^2(n)\right)^{1-\rho-\varepsilon} \\ &\to 0 \end{aligned}$$

as $n \to \infty$ uniformly in x. Hence

$$A(n)\Phi(x)+o(1)\frac{A(n)}{w(x)} \to 0 \tag{3.3.3}$$

as $n \to \infty$ uniformly in x. In other cases, the proofs of (3.3.3) are similar.

Because of

$$1-x \leq e^{-x} \leq 1-x+x^2$$

for $-1 \leq x \leq 1$, we have eventually

$$\begin{aligned} &1-A(n)\Phi(x)-o(1)\frac{A(n)}{w(x)} \\ &\leq \exp\left(-A(n)\Phi(n)-o(1)\frac{A(n)}{w(x)}\right) \\ &\leq 1-A(n)\Phi(x)-o(1)\frac{A(n)}{w(x)}+\left(A(n)\Phi(x)+o(1)\frac{A(n)}{w(x)}\right)^2. \end{aligned}$$

It's easy to check that $(A(n)\Phi(x))^2 = o(1)A(n)/w(x)$ uniformly in x. Hence by (3.3.3), in view of (3.3.2),

$$F^n(\bar{a}_n x + \bar{b}_n) = G_\gamma(x)\left(1 - A(n)\Phi(x) - o(1)\frac{A(n)}{w(x)}\right)$$

as $n \to \infty$, uniformly for $(1+\gamma x)^{-1/\gamma} \le -\log A^2(t)$. Hence

$$\max\left(1,\left((1+\gamma x)^{1/\gamma}\right)^{1-\rho-\varepsilon}\right)\left|\frac{F^n(\bar{a}_n x + \bar{b}_n) - G_\gamma(x)}{A(n)} + G_\gamma(x)\Phi(x)\right|$$
$$= \max\left(1,\, w(x)\right)\frac{G_\gamma(x)}{w(x)}|o(1)|$$

uniformly in x. Since $\max\left(1,\, w(x)\right)G_\gamma(x)/w(x)$ is bounded uniformly in x, the statement follows. □

For the proof of Theorem 3.2.1 we need two additional lemmas. Define

$$\tilde{x} := x + \Delta_x A(n) \quad \text{with} \quad \Delta_x := \frac{1+\gamma x}{\gamma+\rho} I_{\{\rho<0,\gamma+\rho\neq 0\}}. \tag{3.3.4}$$

Lemma 3.3.1.

$$\sup_{-\frac{1}{\gamma\vee 0}<x<\frac{1}{(-\gamma)\vee 0}}\left|\frac{(1+\gamma\tilde{x})^{-1/\gamma}/(1+\gamma x)^{-1/\gamma} - 1}{A(n)} + \frac{1}{\gamma+\rho}I_{\{\rho<0,\gamma+\rho\neq 0\}}\right| = o(1).$$

Proof. Suppose $\gamma \neq 0$. Since $\Delta_x/(1+\gamma x)$ is constant, a Taylor expansion yields

$$\begin{aligned}(1+\gamma\tilde{x})^{-1/\gamma} &= (1+\gamma x + \gamma\Delta_x A(n))^{-1/\gamma}\\ &= (1+\gamma x)^{-1/\gamma}\left(1+\gamma\frac{\Delta_x}{1+\gamma x}A(n)\right)^{-1/\gamma}\\ &= (1+\gamma x)^{-1/\gamma}\left(1 - \frac{\Delta_x A(n)}{1+\gamma x} + o(A(n))\right)\end{aligned}$$

and hence the assertion. The proof is similar in the case $\gamma = 0$. □

Next define for the ease of writing

$$y := (1+\gamma x)^{1/\gamma} \quad \text{and} \quad \tilde{y} := (1+\gamma\tilde{x})^{1/\gamma}.$$

Then Lemma 3.3.1 can be reformulated as follows:

$$\lim_{n\to\infty}\sup_{y>0}\left|\frac{y/\tilde{y} - 1}{A(n)} + \frac{1}{\gamma+\rho}I_{\{\rho<0,\gamma+\rho\neq 0\}}\right| = 0. \tag{3.3.5}$$

Lemma 3.3.2.

$$\lim_{n\to\infty}\sup_{y>0}\, ye^{1/(3y)}\left|\frac{\mathrm{e}^{-1/\tilde{y}}-\mathrm{e}^{-1/y}}{A(n)}-\frac{\mathrm{e}^{-1/y}}{y}\cdot\frac{1}{\gamma+\rho}I_{\{\rho<0,\gamma+\rho\neq 0\}}\right|=0. \quad (3.3.6)$$

Proof. We prove the statement by two steps.
(i) For $y\geq(-\log A^2(n))^{-1}$, we have by (3.3.5) and $|A(n)|/y\leq-\log A^2(n)|A(n)|\to 0$ that $y^{-1}(\tilde{y}/y-1)\to 0$ uniformly. Hence

$$\begin{aligned}e^{-1/\tilde{y}}-e^{-1/y}&=e^{-1/y}\left(\exp\left(-\frac{y/\tilde{y}-1}{y}\right)-1\right)\\&=e^{-1/y}\left(\frac{A(n)}{y(\gamma+\rho)}I_{\{\rho<0,\gamma+\rho\neq 0\}}(1+o(1))\right)\end{aligned}$$

where the o(1)-term is uniformly in y. Hence

$$\begin{aligned}&\lim_{n\to\infty}\sup_{y\geq(-\log A^2(n))^{-1}} ye^{1/(3y)}\left|\frac{e^{-1/\tilde{y}}-e^{-1/y}}{A(n)}-\frac{e^{-1/y}}{y}\cdot\frac{1}{\gamma+\rho}I_{\{\rho<0,\gamma+\rho\neq 0\}}\right|\\&=\lim_{n\to\infty}\sup_{y\geq(-\log A^2(n))^{-1}}|o(1)|e^{-2/(3y)}=0.\end{aligned}$$

(ii) Now let's consider $0<y\leq(-\log A^2(n))^{-1}$. We consider the various terms of (3.3.6) separately. Note that

$$\begin{aligned}ye^{1/(3y)}\cdot\frac{e^{-1/\tilde{y}}}{|A(n)|}&=ye^{1/(3y)}\cdot\frac{e^{-(1+O(A(n)))/y}}{|A(n)|}\\&=ye^{-1/(2y)}e^{-(1/6+O(A(n))/y}/|A(n)|\\&\leq(-\log A^2(n))^{-1}e^{\frac{1}{2}\log A^2(n))}c/|A(n)|\\&=c(-\log A^2(n))^{-1}\to 0\end{aligned}$$

as $n\to\infty$ uniformly in y. In a similar, we check that

$$ye^{1/(3y)}\cdot\frac{e^{-1/y}}{A(n)}\to 0,\quad ye^{1/(3y)}\cdot\frac{e^{-1/y}}{y}\to 0$$

as $n\to\infty$ uniformly in y. Hence

$$\lim_{n\to\infty}\sup_{0<y\leq(-\log A(n))^{-1}} ye^{1/(3y)}\left|\frac{e^{-1/\tilde{y}}-e^{-1/y}}{A(n)}-\frac{e^{-1/y}}{y}\cdot\frac{1}{\gamma+\rho}I_{\{\rho<0,\gamma+\rho\neq 0\}}\right|=0.$$

Combining (i) and (ii), we get (3.3.6). □

Proof of Theorem 3.2.1. We prove the statement by three steps.

(i) We'll prove that

$$\begin{aligned} &\sup_{y\geq(-\log A^2(n))^{-1}} \max\left(1,\, y^{1-\varepsilon I_{\{\rho=0\}}}\right)\times \\ &\quad\times\left|\frac{F^n(a_nx+b_n)-e^{-1/y}}{A(n)}+e^{-1/y}y^{-\gamma-1}\tilde{H}_{\gamma,\rho}(y)\right|=o(1). \end{aligned} \tag{3.3.7}$$

In view of Lemma 3.3.1, Proposition 3.3.1 implies that

$$\begin{aligned} &\sup_{(1+\gamma\tilde{x})^{-1/\gamma}\leq-\log A^2(n)} \max\left(1,\, \left((1+\gamma\tilde{x})^{1/\gamma}\right)^{1-\rho-\varepsilon}\right)\times \\ &\times\left|\frac{F^n(\bar{a}_n\tilde{x}+\bar{b}_n)-G_\gamma(\tilde{x})}{A(n)}+G_\gamma(\tilde{x})(1+\gamma\tilde{x})^{-1/\gamma-1}\bar{H}_{\gamma,\rho}((1+\gamma\tilde{x})^{1/\gamma})\right|=o(1). \end{aligned}$$

Note that $\bar{a}_n\tilde{x}+\bar{b}_n=a_nx+b_n$, so that the last approximation can be rewritten as

$$\sup_{\tilde{y}\geq(-\log A^2(n))^{-1}} \max\left(1,\, \tilde{y}^{1-\rho-\varepsilon}\right)\left|\frac{F^n(a_nx+b_n)-\mathrm{e}^{-1/\tilde{y}}}{A(n)}+\mathrm{e}^{-1/\tilde{y}}\tilde{y}^{-\gamma-1}\bar{H}_{\gamma,\rho}(\tilde{y})\right|=o(1).$$

Now by (3.3.5) we have

$$\sup_{y\geq(-\log A^2(n))^{-1}} \frac{\max\left(1,\, y^{1-\varepsilon I_{\{\rho=0\}}}\right)}{\max\left(1,\, \tilde{y}^{1-\rho-\varepsilon}\right)}<\infty, \tag{3.3.8}$$

and by Lemma 3.3.2, we have

$$\begin{aligned} &\max\left(1,\, y^{1-\varepsilon I_{\{\rho=0\}}}\right)\left|\frac{e^{-1/\tilde{y}}-e^{-1/y}}{A(n)}-\frac{e^{-1/y}}{y}\cdot\frac{1}{\gamma+\rho}I_{\{\rho<0,\gamma+\rho\neq0\}}\right| \\ &=\max\left(1,\, y^{1-\varepsilon I_{\{\rho=0\}}}\right)\}\frac{e^{-1/(3y)}}{y}|o(1)|\to 0 \end{aligned} \tag{3.3.9}$$

uniformly in y. Combining (3.3.7), (3.3.8) and (3.3.9) we see that it suffices to prove that

$$\begin{aligned} &\sup_{y\geq(-\log A^2(n))^{-1}} \max\left(1,\, y^{1-\varepsilon I_{\{\rho=0\}}}\right)\Big|\mathrm{e}^{-1/\tilde{y}}\tilde{y}^{-\gamma-1}\bar{H}_{\gamma,\rho}(\tilde{y}) \\ &\quad -\mathrm{e}^{-1/y}y^{-\gamma-1}\tilde{H}_{\gamma,\rho}(y)-\frac{e^{-1/y}}{y}\cdot\frac{1}{\gamma+\rho}I_{\{\rho<0,\gamma+\rho\neq0\}}\Big|\to 0. \end{aligned} \tag{3.3.10}$$

Check that the term the absolute value of which is considered can be represented as $g(\tilde{y})-g(y)$ with

$$g(t)=\begin{cases} \frac{1}{\gamma+\rho}t^{\rho-1}\mathrm{e}^{-1/t}, & \rho<0,\ \gamma+\rho\neq0, \\ t^{-\gamma-1}\log t\,\mathrm{e}^{-1/t}, & \rho<0,\ \gamma+\rho=0, \\ \frac{1}{\gamma}t^{-1}\log t\,\mathrm{e}^{-1/t}, & \rho=0\neq\gamma. \end{cases}$$

By the mean value theorem, the left-hand side of (3.3.10) equals

$$\sup_{y \geq (-\log A^2(n))^{-1}} \max\left(y,\, y^{2-\varepsilon I_{\{\rho=0\}}}\right)\left|g'(\bar{y})\left(\frac{\tilde{y}}{y}-1\right)\right|$$

for some $\bar{y}$ between y and $\tilde{y}$. Now (3.3.5) implies that $y/\tilde{y} \to 1$ uniformly. Further using (3.3.5), we get

$$\begin{aligned} &\sup_{y \geq (-\log A^2(n))^{-1}} \max\left(y,\, y^{2-\varepsilon I_{\{\rho=0\}}}\right)|g'(\bar{y})| \\ = &\sup_{\bar{y} \geq (-\log A^2(n))^{-1}} \max\left(\bar{y},\, \bar{y}^{2-\varepsilon I_{\{\rho=0\}}}\right)|g'(\bar{y})|(1+o(1)). \end{aligned}$$

It is checked easily that this is bounded. Hence (3.3.7) holds.

(ii) We'll prove that

$$\begin{aligned} \sup_{(1+\gamma x)^{-1/\gamma} \geq -\log A^2(n)} &\left| \frac{F^n(a_n x + b_n) - G_\gamma(x)}{A(n)} \right. \\ &\left. + G_\gamma(x)(1+\gamma x)^{-1/\gamma-1}\tilde{H}_{\gamma,\rho}((1+\gamma x)^{1/\gamma}) \right| = o(1). \end{aligned} \tag{3.3.11}$$

We consider the various terms of (3.3.11) separately. Define x_n such that $(1+\gamma x_n)^{-1/\gamma} = -\log A^2(n)$. By (3.3.2)

$$\begin{aligned} &\sup_{(1+\gamma x)^{-1/\gamma} \geq -\log A^2(n)} \frac{F^n(a_n x + b_n)}{|A(n)|} \\ \leq &\sup_{(1+\gamma \tilde{x})^{-1/\gamma} \geq -\log A^2(n)} \frac{F^n(\bar{a}_n \tilde{x} + \bar{b}_n)}{|A(n)|} \\ \leq &\frac{F^n(\bar{a}_n x_n + \bar{b}_n)}{|A(n)|} \\ = &G_\gamma(x_n)\exp\left(-A(n)\Phi(x_n) - o(1)\frac{A(n)}{w(x_n)}\right)\frac{1}{|A(n)|} \\ = &\exp(\log A^2(n))\exp\left(-A(n)\Phi(x_n) - o(1)\frac{A(n)}{w(x_n)}\right)\frac{1}{|A(n)|}. \end{aligned}$$

Since

$$\frac{\exp(\log A^2(n))}{|A(n)|} \to 0 \quad \text{and} \quad -A(n)\Phi(x_n) - o(1)\frac{A(n)}{w(x_n)} \to 0$$

as $n \to \infty$ (cf. (3.3.2)), we get $F^n(a_n x + b_n)/|A(n)| \to 0$ as $n \to \infty$ uniformly in x. In a similar, we can also prove that

$$\frac{G_\gamma(x)}{A(n)} \to 0 \quad \text{and} \quad G_\gamma(x)(1+\gamma x)^{-1/\gamma-1}\tilde{H}_{\gamma,\rho}((1+\gamma x)^{1/\gamma}) \to 0$$

as $n \to \infty$ uniformly in x. Hence (3.3.11) holds.

Combining (i) and (ii), we have proved that

$$\sup_{-\frac{1}{\gamma\vee 0}<x<\frac{1}{(-\gamma)\vee 0}} \max\left(1,\ \left((1+\gamma x)^{1/\gamma}\right)^{1-\varepsilon I_{\{\rho=0\}}}\right)\times$$
$$\times\left|\frac{F^n(a_nx+b_n)-G_\gamma(x)}{A(n)}+G_\gamma(x)(1+\gamma x)^{-1/\gamma-1}\tilde{H}_{\gamma,\rho}((1+\gamma x)^{1/\gamma})\right|=o(1)$$

as $n \to \infty$. For $\gamma \le 0$, It is same as the statement of Theorem 3.2.1. For $\gamma > 0$, we need to prove that

$$\sup_{x<-1/\gamma}|F^n(a_nx+b_n)-G_\gamma(x)|/|A(n)|=o(1) \tag{3.3.12}$$

as $n \to \infty$ since $G_\gamma(x)(1+\gamma x)^{-1/\gamma-1}\tilde{H}_{\gamma,\rho}((1+\gamma x)^{1/\gamma})=0$ for all $x<-1/\gamma$.

(iii) $\gamma > 0,\ x < -1/\gamma$. Let $\delta_n = (-\log A^2(n))^{-\gamma}/\gamma > 0$ and $x_n = -1/\gamma + \delta_n$, then $(1+\gamma x_n)^{-1/\gamma} = -\log A^2(n)$. By (i) and (ii)

$$\begin{aligned}&\sup_{x<-1/\gamma}|F^n(a_nx+b_n)-G_\gamma(x)|/|A(n)|\\&=\sup_{x<-1/\gamma}F^n(a_nx+b_n)/|A(n)|\\&\le F^n(a_nx_n+b_n)/|A(n)|\\&\le G_\gamma(x_n)/|A(n)|+|G_\gamma(x_n)(1+\gamma x_n)^{-1/\gamma-1}\tilde{H}_{\gamma,\rho}((1+\gamma x_n)^{1/\gamma})|+|o(1)|.\end{aligned}$$

It's easy to check that

$$\frac{G_\gamma(x_n)}{A(n)}\to 0 \quad\text{and}\quad G_\gamma(x_n)(1+\gamma x_n)^{-1/\gamma-1}\tilde{H}_{\gamma,\rho}((1+\gamma x_n)^{1/\gamma})\to 0$$

as $n \to \infty$. Thus (3.3.12) holds, and assertion of Theorem 3.2.1. □

Proof of Corollary 3.2.1. The first assertion follows from Theorem 3.2.1 and the boundness of $\max(1,\ (1+\gamma x)^{1/\gamma})(1-G_\gamma(x))$ uniformly for $-\frac{1}{\gamma\vee 0} < x < \frac{1}{(-\gamma)\vee 0}$. The second assertion is now obvious since

$$\frac{G_\gamma(x)}{1-G_\gamma(x)}(1+\gamma x)^{-1-1/\gamma}\tilde{H}_{\gamma,\rho}((1+\gamma x)^{1/\gamma})$$

is bounded uniformly for $-\frac{1}{\gamma\vee 0} < x < \frac{1}{(-\gamma)\vee 0}$. □

Proof of Remark 3.2.1. Note the fact: for $x \ge -1/\gamma$

$$G_\gamma(x)(1+\gamma x)^{-1/\gamma-1}\tilde{H}_{\gamma,\rho}((1+\gamma x)^{1/\gamma})=0.$$

For any positive sequence M_n, let $\delta_n = -\exp(\gamma \max(n, M_n))/\gamma > 0$ and $x_n := -1/\gamma - \delta_n$, then $(1+\gamma x_n)^{-1/\gamma} = \exp(-\max(n, M_n))$. For $x \geq -1/\gamma$, by Theorem 3.2.1

$$\begin{aligned}
&|F^n(a_n x + b_n) - G_\gamma(x)|/|A(n)| \\
&\leq (1 - F^n(a_n x_n + b_n))/|A(n)| \\
&= (1 - G_\gamma(x_n))/|A(n)| + G_\gamma(x_n)(1+\gamma x_n)^{-1/\gamma-1}\tilde{H}_{\gamma,\rho}((1+\gamma x_n)^{1/\gamma}) \\
&\quad + o(1)(1+\gamma x_n)^{-(1-\varepsilon)/\gamma}.
\end{aligned}$$

Note that

$$\begin{aligned}
M_n(1 - G_\gamma(x_n))/A(n) &= M_n(1 - \exp(-\exp(-\max(n, M_n))))/A(n) \\
&= M_n \exp(-\max(n, M_n))(1 + o(1))/A(n) \to 0
\end{aligned}$$

and

$$M_n(1+\gamma x_n)^{-(1-\varepsilon)/\gamma} o(1) = M_n \exp(-(1-\varepsilon)\max(n, M_n)) o(1) \to 0$$

and also

$$M_n G_\gamma(x_n)(1+\gamma x_n)^{-1/\gamma-1}\tilde{H}_{\gamma,\rho}((1+\gamma x_n)^{1/\gamma}) \to 0$$

as $n \to \infty$. Thus

$$M_n \cdot \sup_{x \geq -1/\gamma} \left| \frac{F^n(a_n x + b_n) - G_\gamma(x)}{A(n)} \right| = o(1) \tag{3.3.13}$$

as $n \to \infty$. The assertion now follows by contradiction. If Remark 3.2.1 is not true, relation (3.3.13) does not hold for

$$M_n := n / \sup_{x \geq -1/\gamma} \left| \frac{F^n(a_n x + b_n) - G_\gamma(x)}{A(n)} \right| .$$

□

3.4 Appendix

In the present paper we use the second order condition (3.1.4) on $V = (1/(-\log F))^{\leftarrow}(t)$, while the analogous condition on $U = (1/(1-F))^{\leftarrow}(t)$ is more common in the literature. In this appendix, we will discuss the relationship between these two conditions. To this end, we first examine the effect of certain transformations on the so-called second order extended regular variation, that is, condition (3.1.4) in a slightly more abstract framework.

Proposition 3.4.1. *Suppose $g \in ERV^{(2)}(\gamma_1, \rho_1)$ with $\gamma_1 \in \mathbb{R}$, $\rho_1 \le 0$, i.e.*

$$\frac{\frac{g(tx)-g(t)}{a_1(t)} - \frac{x^{\gamma_1}-1}{\gamma_1}}{A_1(t)} \to H_{\gamma_1,\rho_1}(x) \tag{3.4.1}$$

and f satisfies

$$\frac{\frac{f(tx)}{f(t)} - x^{\gamma_2}}{A_2(t)} \to x^{\gamma_2}\frac{x^{\rho_2}-1}{\rho_2} \tag{3.4.2}$$

with $\gamma_2 > 0$ and $|A_2| \in RV(\rho_2)$ for some $\rho_2 \le 0$ and $A(t) \to 0$ as $t \to \infty$. If

$$\frac{A_1(f(t))}{A_2(t)} \to c \tag{3.4.3}$$

as $t \to \infty$ for some $c \in [-\infty, +\infty]$, then

$$\begin{aligned} &\frac{\frac{g(f(tx))-g(f(t))}{\gamma_2 a_1(f(t))} - \frac{x^{\gamma_1\gamma_2}-1}{\gamma_1\gamma_2}}{|A_1(f(t))| + |A_2(t)|} \\ &\to \operatorname{sgn}(A_2)\left(\frac{1}{1+|c|}\cdot\frac{x^{\gamma_1\gamma_2}}{\gamma_2}\frac{x^{\rho_2}-1}{\rho_2} + \frac{c}{1+|c|}\cdot\frac{1}{\gamma_2}H_{\gamma_1,\rho_1}(x^{\gamma_2})\right) \end{aligned} \tag{3.4.4}$$

with $\operatorname{sgn}(A_2)$ denoting the eventually constant sign of $A_2(t)$ and $c/(1+|c|)$ defined as ±1 for $c \pm \infty$.

Corollary 3.4.1. *(i) Suppose $U \in ERV^{(2)}(\gamma, \rho)$ with $\gamma \in \mathbb{R}$, $\rho \le 0$ and auxiliary functions a and A. If $2tA(t) \to c \in [-\infty, +\infty] \setminus \{1-\gamma\}$, then*

$$\frac{\frac{V(tx)-V(t)}{a^*(t)} - \frac{x^{\gamma}-1}{\gamma}}{A^*(t)} \to H_{\gamma,\rho^*}(x) \tag{3.4.5}$$

as $t \to \infty$ for all $x \in \mathbb{R}$ with

$$\begin{aligned} \rho^* &= \max(\rho, -1), \\ a^*(t) &= \Big(1 - \frac{\gamma}{1+|c|}A_0(t)\Big)a\Big(\frac{1}{1-\mathrm{e}^{-1/t}}\Big), \\ A^*(t) &= \frac{\gamma-1+c}{1+|c|}A_0(t), \\ A_0(t) &= \Big|A\Big(\frac{1}{1-\mathrm{e}^{-1/t}}\Big)\Big| + \frac{1}{2t}. \end{aligned}$$

(ii) Conversely, suppose $V \in ERV^{(2)}(\gamma, \rho)$ with $\gamma \in \mathbb{R}$, $\rho \le 0$ and auxiliary functions a and A. If $2tA(t) \to c \in [-\infty, +\infty] \setminus \{\gamma - 1\}$, then

$$\frac{\frac{U(tx)-U(t)}{a^*(t)} - \frac{x^{\gamma}-1}{\gamma}}{A^*(t)} \to H_{\gamma,\rho^*}(x)$$

as $t \to \infty$ for all $x \in \mathbb{R}$ with

$$\rho^* = \max(\rho, -1),$$
$$a^*(t) = \Big(1 + \frac{\gamma}{1+|c|}A_0(t)\Big)a\Big(\frac{1}{1-\mathrm{e}^{-1/t}}\Big),$$
$$A^*(t) = \frac{1-\gamma+c}{1+|c|}A_0(t).$$

Remark 3.4.1. *In case of $c = \gamma - 1$, V may or may not belong to $ERV^{(2)}(\gamma, \rho^*)$ for some $\rho^* \le 0$ if $U \in ERV^{(2)}(\gamma, \rho)$.*

Proof of Proposition 3.4.1. Because (3.4.2) with $\gamma_2 > 0$ implies $f(t) \to \infty$ as $t \to \infty$ and convergence (3.4.1) holds locally uniformly, one has for fixed $x > 0$

$$\begin{aligned}
&\frac{g(f(tx)) - g(f(t))}{a_1(f(t))} \\
&= \frac{g(\frac{f(tx)}{f(t)} \cdot f(t)) - g(f(t))}{a_1(f(t))} \\
&= \frac{(\frac{f(tx)}{f(t)})^{\gamma_1} - 1}{\gamma_1} + A_1(f(t))H_{\gamma_1,\rho_1}\left(\frac{f(tx)}{f(t)}\right) + o(A_1(f(t))) \\
&= x^{\gamma_1\gamma_2} \cdot \frac{(\frac{f(tx)}{x^{\gamma_2}f(t)})^{\gamma_1} - 1}{\gamma_1} + \frac{x^{\gamma_1\gamma_2} - 1}{\gamma_1} + A_1(f(t))H_{\gamma_1,\rho_1}\left(\frac{f(tx)}{f(t)}\right) + o(A_1(f(t))).
\end{aligned} \tag{3.4.6}$$

By (3.4.2), one has for fixed $x > 0$

$$\frac{f(tx)}{x^{\gamma_2}f(t)} = 1 + A_2(t)\frac{x^{\rho_2} - 1}{\rho_2} + o(A_2(t)).$$

Hence

$$\frac{(\frac{f(tx)}{x^{\gamma_2}f(t)})^{\gamma_1} - 1}{\gamma_1} = A_2(t)\frac{x^{\rho_2} - 1}{\rho_2} + o(A_2(t)). \tag{3.4.7}$$

From (3.4.6) and (3.4.7), one may conclude

$$\begin{aligned}
&\frac{g(f(tx)) - g(f(t))}{\gamma_2 a_1(f(t))} - \frac{x^{\gamma_1\gamma_2} - 1}{\gamma_1\gamma_2} \\
&= \frac{x^{\gamma_1\gamma_2}}{\gamma_2}\left(A_2(t)\frac{x^{\rho_2} - 1}{\rho_2} + o(A_2(t))\right) + \frac{1}{\gamma_2}A_1(f(t))H_{\gamma_1,\rho_1}(x^{\gamma_2}) + o(A_1(f(t))).
\end{aligned} \tag{3.4.8}$$

By (3.4.3), we can easily get (3.4.4). □

Proof of Corollary 3.4.1.
(i) The function f defined by $f(t) = 1/(1 - e^{-1/t})$ satisfies (3.4.2) with $\gamma_2 = 1$, $\rho_2 = -1$, and $A_2(t) = -1/2t$. Since $V(t) = U(f(t))$ and $A(f(t))/A_2(t) = -2tA(t)(1 + o(1))$, Proposition 3.4.1 yields

$$\frac{\frac{V(tx)-V(t)}{a(f(t))} - \frac{x^\gamma - 1}{\gamma}}{A_0(t)} \to \frac{1}{1+|c|}(x^{\gamma-1} - x^\gamma) + \frac{c}{1+|c|}H_{\gamma,\rho}(x).$$

Hence, because of $1/(1+y) = 1 - y + o(y)$ as $y \to 0$,

$$\begin{aligned} &\frac{V(tx) - V(t)}{a^*(t)} \\ &= \frac{x^\gamma - 1}{\gamma}\Big(1 + \frac{\gamma}{1+|c|}A_0(t) + o(A_0(t))\Big) \\ &\quad + \Big(\frac{1}{1+|c|}(x^{\gamma-1} - x^\gamma) + \frac{c}{1+|c|}H_{\gamma,\rho}(x)\Big)A_0(t) + o(A_0(t)) \\ &= \frac{x^\gamma - 1}{\gamma} + \Big(\frac{1}{1+|c|}(x^{\gamma-1} - 1) + \frac{c}{1+|c|}H_{\gamma,\rho}(x)\Big)A_0(t) + o(A_0(t)). \end{aligned} \tag{3.4.9}$$

If $\gamma = 1$ then $x^{\gamma-1} - 1$ vanishes, and the assertion is obvious, because c is assumed unequal to $1 - \gamma = 0$.

If $\gamma \neq 1$, then $x^{\gamma-1} - 1 = (\gamma - 1)H_{\gamma,-1}(x)$. So if $|c| = \infty$ (which implies $\rho \geq -1$) or $c = 0$ (and hence $\rho \leq -1$), then (3.4.5) is immediate from (3.4.9). Finally, if $c \in \mathbb{R} \setminus \{0, 1-\gamma\}$, then necessarily $\rho = -1$ and $\gamma - 1 + c \neq 0$, so that again the assertion follows from (3.4.9).

(ii) The proof is very similar to the one of (i). Here we use $f(t) = 1/(-\log(1 - 1/t))$, satisfying (3.4.2) with $\gamma_2 = 1$, $\rho_2 = -1$, and $A_2(t) = 1/2t$. □

Note. Chapter 3 is based on the paper Drees, de Haan and Li (2003). Here we present a little stronger results and add more remarks.

Chapter 4

Weighted Approximations of Tail Copula Processes with Application to Testing the Multivariate Extreme Value Condition

co-authors: John Einmahl and Laurens de Haan

Abstract. Consider n i.i.d. random vectors on $\mathbb{R}^2$, with unknown, common distribution function F. Under a sharpening of the extreme value condition on F, we derive a weighted approximation of the corresponding tail copula process. Then we construct a test to check whether the extreme value condition holds by comparing two estimators of the limiting extreme value distribution, one obtained from the tail copula process and the other obtained by first estimating the spectral measure which is then used as a building block for the limiting extreme value distribution. We derive the limiting distribution of the test statistic from the aforementioned weighted approximation. This limiting distribution contains unknown functional parameters. Therefore we show that a version with estimated parameters converges weakly to the true limiting distribution. Based on this result, the finite sample properties of our testing procedure are investigated through a simulation study. A real data application is also presented.

4.1 Introduction

Let (X, Y), $(X_1, Y_1), ..., (X_n, Y_n)$ be i.i.d. random vectors with continuous distribution function (d.f.) F. Suppose that there exist norming constants $a_n, c_n > 0$ and $b_n, d_n \in \mathbb{R}$ such that the sequence of d.f.'s

$$P\left(\frac{\max_{1\le i\le n} X_i - b_n}{a_n} \le x,\ \frac{\max_{1\le i\le n} Y_i - d_n}{c_n} \le y\right)$$

converges to a limit d.f., say $G(x, y)$, with non-degenerate marginal d.f., that is,

$$\lim_{n\to\infty} F^n(a_n x + b_n,\, c_n y + d_n) = G(x, y) \tag{4.1.1}$$

for all but countably many x and y. Then, for a suitable choice of a_n, b_n, c_n and d_n, there exist $\gamma_1,\, \gamma_2 \in \mathbb{R}$ such that

$$G(x,\infty) = \exp\left(-(1+\gamma_1 x)^{-1/\gamma_1}\right), \quad G(\infty, y) = \exp\left(-(1+\gamma_2 y)^{-1/\gamma_2}\right).$$

The d.f. G is called an extreme value d.f. and $\gamma_1,\, \gamma_2$ are called the (marginal) extreme value indices.

Any extreme value d.f. G can be represented as

$$G\left(\frac{x^{-\gamma_1}-1}{\gamma_1}, \frac{y^{-\gamma_2}-1}{\gamma_2}\right) = \exp\left(-\int_0^{\pi/2} (x(1\wedge\tan\theta)) \vee (y(1\wedge\cot\theta))\, \Phi(d\theta)\right), \tag{4.1.2}$$

with Φ the d.f. of the so-called spectral measure. There is a one-to-one correspondence between extreme value d.f.'s G and finite measures with d.f. Φ that satisfy

$$\int_0^{\pi/2} (1\wedge\tan\theta)\, \Phi(d\theta) = \int_0^{\pi/2} (1\wedge\cot\theta)\, \Phi(d\theta) = 1,$$

via (4.1.2).

Alternatively one can characterize the extreme value d.f.'s G by: there is a measure Λ on $[0,\infty]^2 \setminus \{(\infty,\infty)\}$ such that, with

$$l(x,y) := -\log G\left(\frac{x^{-\gamma_1}-1}{\gamma_1}, \frac{y^{-\gamma_2}-1}{\gamma_2}\right), \tag{4.1.3}$$

we have

$$\begin{aligned} &1.\;\; l(x,y) = \Lambda\left(\{(u,v)\in[0,\infty]^2 :\, u\le x \text{ or } v\le y\}\right), \\ &2.\;\; l(tx,ty) = t\,l(x,y) \quad \text{for } t,x,y>0. \end{aligned} \tag{4.1.4}$$

Combining the two characterizations we find

$$l(x,y) = \int_0^{\pi/2} (x(1\wedge\tan\theta)) \,\vee\, (y(1\wedge\cot\theta))\; \Phi(d\theta). \tag{4.1.5}$$

Relation (4.1.1) implies (cf. Einmahl, de Haan and Piterbarg (2001))

$$\lim_{t\downarrow 0} t^{-1} P\left((1-F_1(X))\wedge(1-F_2(Y)) \le t,\, 1-F_2(Y) \le (1-F_1(X))\tan\theta\right) = \Phi(\theta)$$

$$(4.1.6)$$

for continuity points $\theta \in (0, \pi/2]$ of Φ, where $F_1(x) := F(x, \infty)$ and $F_2(y) := F(\infty, y)$. Also

$$\lim_{t \downarrow 0} t^{-1} P\left(1 - F_1(X) \leq tx \text{ or } 1 - F_2(Y) \leq ty\right) = l(x, y) \tag{4.1.7}$$

for $(x, y) \in [0, \infty)^2$. More generally

$$\lim_{t \downarrow 0} t^{-1} P\left((1 - F_1(X),\, 1 - F_2(Y)) \in tA\right) = \Lambda(A) \tag{4.1.8}$$

for any Borel set A in $[0, \infty]^2 \setminus \{(\infty, \infty)\}$ (with $tA := \{(tx, ty) : (x, y) \in A\}$) provided $\Lambda(\partial A) = 0$.

A non-parametric estimator for Φ, suggested by the limit relation (4.1.6) is (Einmahl *et al.* (2001))

$$\hat{\Phi}(\theta) := \frac{1}{k} \sum_{i=1}^{n} I_{\{R_i^X \vee R_i^Y \geq n+1-k,\ n+1-R_i^Y \leq (n+1-R_i^X) \tan\theta\}} \tag{4.1.9}$$

where R_i^X is the rank of X_i among $X_1, X_2, ..., X_n$, R_i^Y is the rank of Y_i among $Y_1, Y_2, ..., Y_n$. Similarly a non-parametric estimator for l, suggested by the limit relation (4.1.7) is (Huang (1992), see also Drees and Huang (1998))

$$\begin{aligned} \hat{l}_2(x, y) &:= \frac{1}{k} \sum_{i=1}^{n} I_{\{X_i > X_{n+1-\lceil kx \rceil : n} \text{ or } Y_i > Y_{n+1-\lceil ky \rceil : n}\}} \\ &= \frac{1}{k} \sum_{i=1}^{n} I_{\{R_i^X > n+1-kx \text{ or } R_i^Y > n+1-ky\}}, \end{aligned} \tag{4.1.10}$$

where $X_{1:n} \leq \cdots \leq X_{n:n}$ are the order statistics of the X_i, $i = 1, 2, \ldots, n$ (similarly for the Y_i), with $\lceil z \rceil$ the smallest integer $\geq z$.

The mentioned papers give asymptotic normality results for $\hat{\Phi}$ and $\hat{l}_2$ under certain conditions and with sequences $k = k(n)$ satisfying $k(n) \to \infty$, $k(n)/n \to 0$, as $n \to \infty$. Another way of estimating l is via (4.1.5) and (4.1.9):

$$\hat{l}_1(x, y) := \int_0^{\pi/2} (x(1 \wedge \tan\theta)) \vee (y(1 \wedge \cot\theta))\, \hat{\Phi}(d\theta). \tag{4.1.11}$$

The multivariate extreme value framework that we sketched is the appropriate one when one, e.g., wants to estimate the probability of *extreme sets* i.e., sets outside the range of the observations.; see de Haan and Sinha (1999). Condition (4.1.1) is fulfilled for many standard distributions but not for all distributions. Hence before using this framework to estimate probabilities of extreme sets, it is important to check whether (4.1.1) is a reasonable assumption for the data set at hand. And one wants to do this beforehand, without specifying the exact structure of the limiting distribution.

A promising approach to this testing problem seems to be to see if the two estimators $\hat{l}_1$ and $\hat{l}_2$ for l, that have a different background, are not too different. The estimator $\hat{l}_2$ is a natural one mimicking more or less the tail of the distribution itself. But this estimator does not necessarily satisfy condition 2 of (4.1.4). On the other hand $\hat{l}_1$ does satisfy condition 2 of (4.1.4) but the estimator itself is of a somewhat more complicated nature. So one can maintain that such a test would check whether condition 2 of (4.1.4) holds.

The proposed test statistic is of Anderson-Darling type:

$$L_n := \iint_{0<x,y\leq 1} \left(\hat{l}_1(x,y) - \hat{l}_2(x,y)\right)^2 (x \vee y)^{-\beta}\, dxdy \tag{4.1.12}$$

for certain $\beta \geq 0$. The test statistic is similar to those used for testing a parametric null hypothesis (like testing for normality), where the empirical distribution function is compared with the true distribution function with estimated parameters. Here, however, the estimated parameter Φ is a function (and we only deal with the tail of the distribution). Also note that our methods allow us to deal with other test statistics than L_n as well.

Note that this test checks whether the dependence structure is of the right type. It is only based on the relative positions (ranks) of the data and completely independent of the marginal distributions of F for which tests have been developed already in Drees, de Haan and Li (2004) and Dietrich, de Haan and Hüsler (2002).

We shall establish the asymptotic distribution of kL_n as $n \to \infty$ under (4.1.1) and some extra conditions stemming from Huang (1992) and Einmahl *et al.* (2001), thus providing a basis for applying a test.

Note that the test statistic L_n is based on observations for which at least one component exceeds a certain threshold. Since the estimators depend on this threshold, one can plot L_n as a function of k. This plot can be used as an exploratory tool for determining from which threshold on the two estimators $\hat{l}_1$ and $\hat{l}_2$ are close to each other suggesting that the approximations (4.1.6) and (4.1.7) can be trusted, and hence yields a heuristic procedure for determining k. So this a second use of the test statistic L_n.

The weak convergence of kL_n is stated in Theorem 4.2.3. For the proof of this theorem the known asymptotic normality result for $\hat{\Phi}$ (Einmahl *et al.* (2001)) is sufficient but not the *known* one for $\hat{l}_2$ (Huang (1992)). Hence as a preliminary but important result, we first develop a Gaussian approximation for the weighted tail copula process on $(0,1]^2$

$$\sqrt{k}\left(\hat{l}_2(x,y) - l(x,y)\right)/(x \vee y)^{\eta}, \quad 0 \leq \eta < 1/2,$$

thus extending significantly the result of Huang (1992) where $\eta = 0$. This result, which seems to be useful in other contexts as well, is stated in Theorem 4.2.2. The proofs are given in section 4.3.

The limiting random variable in Theorem 4.2.3 is determined as an integral of a combination of Gaussian processes. They are parametrized by functions which can be estimated consistently. In section 4.4 it is proved that the probability distribution of the limiting random variable with these functions estimated converges to the distribution of the limiting random variable with these functions equal to the actual ones, which makes the procedure applicable in practice. In section 4.5 simulation results and an application to real data are reported.

4.2 Main results

Before stating the main results, we introduce some notation. Define W_Λ to be a Wiener process indexed by the Borel sets in $[0,\infty]^2 \setminus \{(\infty,\infty)\}$, depending on the parameter Λ from (4.1.4), which is a measure and we assume it has a density λ, in the following way: W_Λ is a centered Gaussian process and for Borel sets C and $\tilde{C}$: $EW_\Lambda(C)W_\Lambda(\tilde{C}) = \Lambda(C \cap \tilde{C})$. Define the sets C_θ by

$$C_\theta = \{(x,y) \in [0,\infty]^2 : \ x \wedge y \le 1,\ y \le x \tan\theta\}, \qquad \theta \in [0, \frac{\pi}{2}],$$

and the process Z by

$$\begin{aligned} Z(\theta) = &\int_0^{1\vee\frac{1}{\tan\theta}} \lambda(x, x\tan\theta)(W_1(x)\tan\theta - W_2(x\tan\theta))\,dx \\ &- W_2(1)\int_{1\vee\frac{1}{\tan\theta}}^{\infty} \lambda(x,1)\,dx - I_{(\frac{\pi}{4},\frac{\pi}{2}]}(\theta)W_1(1)\int_1^{\tan\theta} \lambda(1,y)\,dy, \quad \theta \in [0,\frac{\pi}{2}), \\ Z\left(\frac{\pi}{2}\right) = &-W_2(1)\int_1^\infty \lambda(x,1)\,dx \ - \ W_1(1)\int_1^\infty \lambda(1,y)\,dy, \end{aligned} \tag{4.2.1}$$

where λ is the density of Λ, with $W_1(x) = W_\Lambda([0,x]\times[0,\infty])$ and $W_2(y) = W_\Lambda([0,\infty]\times[0,y])$.

Define for $x, y > 0$

$$W_R(x,y) = W_\Lambda([0,x]\times[0,y]), \quad R(x,y) = \Lambda([0,x]\times[0,y]) \tag{4.2.2}$$

and

$$R_1(x,y) = \partial R(x,y)/\partial x, \quad R_2(x,y) = \partial R(x,y)/\partial y. \tag{4.2.3}$$

Theorem 4.2.1. *Assume that condition (4.1.8) and Conditions 1 and 2 of Einmahl et al. (2001) hold, and that Λ has a continuous density λ on $[0,\infty)^2 \setminus \{(0,0)\}$. Then for a special construction*

$$\sup_{0<x,y\le 1} \frac{\left|\sqrt{k}(\hat{l}_1(x,y) - l(x,y)) - A(x,y)\right|}{x \vee y} \xrightarrow{P} 0$$

as $n \to \infty$, where

$$A(x,y) := \begin{cases} x(W_\Lambda(C_{\frac{\pi}{2}}) + Z(\frac{\pi}{2})) + y \int_{\pi/4}^{\arctan \frac{y}{x}} \frac{1}{\sin^2\theta}(W_\Lambda(C_\theta) + Z(\theta))d\theta, & \text{if } y \geq x, \\ x(W_\Lambda(C_{\frac{\pi}{2}}) + Z(\frac{\pi}{2})) - x \int_{\arctan \frac{y}{x}}^{\pi/4} \frac{1}{\cos^2\theta}(W_\Lambda(C_\theta) + Z(\theta))d\theta, & \text{if } y < x. \end{cases}$$

Let

$$U_i = 1 - F_1(X_i), \quad V_i = 1 - F_2(Y_i), \quad i = 1, 2, ..., n. \tag{4.2.4}$$

Let $C(x,y)$ is the distribution function of (U_i, V_i). By (4.1.8) and (4.2.2) we have $R(x,y) = \lim_{t\downarrow 0} t^{-1}C(tx,ty)$. We assume, as in Huang (1992), that for some $\alpha > 0$

$$t^{-1}C(tx,ty) - R(x,y) = O(t^\alpha) \quad \text{as } t \downarrow 0, \tag{4.2.5}$$

uniformly for $x \vee y \leq 1$, $x, y \geq 0$.

Theorem 4.2.2. *Assume that conditions (4.1.8) and (4.2.5) hold and that $k = o\left(n^{\frac{2\alpha}{1+2\alpha}}\right)$. If R_1 and R_2 are continuous, then we have for $0 \leq \eta < 1/2$ and for a special construction*

$$\sup_{0<x,y\leq 1} \frac{\left|\sqrt{k}(\hat{l}_2(x,y) - l(x,y)) + B(x,y)\right|}{(x \vee y)^\eta} \xrightarrow{P} 0$$

as $n \to \infty$, where

$$B(x,y) := W_R(x,y) - R_1(x,y)W_1(x) - R_2(x,y)W_2(y).$$

Theorem 4.2.3. *Assume the conditions of Theorems 4.2.1 and 4.2.2 hold. Then for each $0 \leq \beta < 3$*

$$\iint_{0<x,y\leq 1} \frac{k\left(\hat{l}_1(x,y) - \hat{l}_2(x,y)\right)^2}{(x \vee y)^\beta} dxdy \xrightarrow{d} \iint_{0<x,y\leq 1} \frac{(A(x,y) + B(x,y))^2}{(x \vee y)^\beta} dxdy \tag{4.2.6}$$

as $n \to \infty$, and the limit is finite almost surely.

Remark 4.2.1. *The case $\beta = 0$ is similar to the Cramér-von Mises test. Note that for $\beta < 2$, Theorem 4.2.3 easily follows from an unweighted approximation in Theorems 4.2.1 and 4.2.2. Therefore the case $\beta = 2(!)$ is similar to the Anderson-Darling test.*

Remark 4.2.2. *Note that we do not merely test the multivariate extreme value condition but also the refined conditions of Theorem 4.2.3. Hence we actually test a smaller null hypothesis. But such a smaller hypothesis is needed for statistical applications, since these refined conditions are the ones that yield that the normalized tail of F is sufficiently close to G.*

Remark 4.2.3. *The random variable on the right in Theorem 4.2.3 has a continuous distribution function. This follows from a property of Gaussian measures on Banach spaces: the measure of a closed ball is a continuous function of its radius, see, e.g., Paulauskas and Račkauskas (1989), Chapter 4, Theorem 1.2.*

Remark 4.2.4. *Since $x \vee y \leq l(x,y) \leq x+y \leq 2(x \vee y)$, (4.2.6) remains true with $x \vee y$ replaced with $l(x,y)$ or $x+y$, but when choosing $l(x,y)$, the left-hand-side of (4.2.6) is not a statistic and l has to be estimated.*

4.3 Proofs

Before proving Theorem 4.2.1, we first present two lemmas and a proposition.

Lemma 4.3.1.

$$l(x,y) = \begin{cases} x\Phi(\frac{\pi}{2}) + y\int_{\pi/4}^{\arctan\frac{y}{x}} \frac{1}{\sin^2\theta}\Phi(\theta)d\theta, & \text{if } y \geq x, \\ x\Phi(\frac{\pi}{2}) - x\int_{\arctan\frac{y}{x}}^{\pi/4} \frac{1}{\cos^2\theta}\Phi(\theta)d\theta, & \text{if } y < x. \end{cases}$$

Proof. Since

$$\begin{aligned} l(x,y) &= \int_0^{\pi/2} (x(1\wedge\tan\theta)) \vee (y(1\wedge\cot\theta))\ \Phi(d\theta) \\ &= \int_0^{\pi/4} (x\tan\theta) \vee y\ \Phi(d\theta) + \int_{\pi/4}^{\pi/2} x \vee (y\cot\theta)\ \Phi(d\theta) \end{aligned}$$

and

$$x\tan\theta > y \Leftrightarrow x > y\cot\theta \Leftrightarrow \theta > \arctan\frac{y}{x},$$

then

$$\begin{aligned} l(x,y) &= \int_0^{\frac{\pi}{4}\wedge\arctan\frac{y}{x}} y\,\Phi(d\theta) + \int_{\frac{\pi}{4}\wedge\arctan\frac{y}{x}}^{\frac{\pi}{4}} x\tan\theta\,\Phi(d\theta) \\ &\quad + \int_{\pi/4}^{\frac{\pi}{4}\vee\arctan\frac{y}{x}} y\cot\theta\,\Phi(d\theta) + \int_{\frac{\pi}{4}\vee\arctan\frac{y}{x}}^{\pi/2} x\,\Phi(d\theta) \\ &= \begin{cases} \int_0^{\pi/4} y\,\Phi(d\theta) + \int_{\pi/4}^{\arctan\frac{y}{x}} y\cot\theta\,\Phi(d\theta) + \int_{\arctan\frac{y}{x}}^{\pi/2} x\,\Phi(d\theta), & \text{if } y \geq x, \\ \int_0^{\arctan\frac{y}{x}} y\,\Phi(d\theta) + \int_{\arctan\frac{y}{x}}^{\pi/4} x\tan\theta\,\Phi(d\theta) + \int_{\pi/4}^{\pi/2} x\,\Phi(d\theta), & \text{if } y < x. \end{cases} \end{aligned}$$

In case of $y \geq x$, via integration by parts, one has

$$\begin{aligned} l(x,y) &= y\Phi(\frac{\pi}{4}) - y\Phi(0) + y\cot(\arctan\frac{y}{x})\Phi(\arctan\frac{y}{x}) - y\cot\frac{\pi}{4}\Phi(\frac{\pi}{4}) \\ &\quad - y\int_{\pi/4}^{\arctan\frac{y}{x}} \Phi(\theta)(-\frac{1}{\sin^2\theta})d\theta + x\Phi(\frac{\pi}{2}) - x\Phi(\arctan\frac{y}{x}) \\ &= x\Phi(\frac{\pi}{2}) + y\int_{\pi/4}^{\arctan\frac{y}{x}} \frac{1}{\sin^2\theta}\Phi(\theta)d\theta. \end{aligned}$$

In case of $y < x$, via integration by parts again, one has

$$\begin{aligned} l(x,y) &= y\Phi(\arctan\frac{y}{x}) - y\Phi(0) + x\tan\frac{\pi}{4}\Phi(\frac{\pi}{4}) - x\tan(\arctan\frac{y}{x})\Phi(\arctan\frac{y}{x}) \\ &\quad - x\int_{\arctan\frac{y}{x}}^{\pi/4} \Phi(\theta)\frac{1}{\cos^2\theta}d\theta + x\Phi(\frac{\pi}{2}) - x\Phi(\frac{\pi}{4}) \\ &= x\Phi(\frac{\pi}{2}) - x\int_{\arctan\frac{y}{x}}^{\pi/4} \frac{1}{\cos^2\theta}\Phi(\theta)d\theta. \end{aligned}$$

□

Write

$$R_n(x,y) = \frac{n}{k}C\left(\frac{kx}{n}, \frac{ky}{n}\right), \quad T_n(x,y) = \frac{1}{k}\sum_{i=1}^{n} I_{\{U_i < \frac{kx}{n}, V_i < \frac{ky}{n}\}} \tag{4.3.1}$$

$$v_n(x,y) = \sqrt{k}(T_n(x,y) - R_n(x,y)), \quad v_{n,\eta}(x,y) = \frac{v_n(x,y)}{(x \vee y)^\eta} \tag{4.3.2}$$

and

$$v_{n,\eta,1}(x) = \frac{v_n(x,\infty)}{x^\eta}, \quad v_{n,\eta,2}(y) = \frac{v_n(\infty,y)}{y^\eta}, \quad v_{n,j} = v_{n,0,j},\ j = 1,2. \tag{4.3.3}$$

Proposition 4.3.1. *Let $T > 0$. For $0 \leq \eta < 1/2$*

$$(v_{n,\eta}(x,y),\, x,y \in (0,T], \quad v_{n,\eta,1}(x),\, x \in (0,T], \quad v_{n,\eta,2}(y),\ y \in (0,T])$$

converges in distribution to

$$\left(\frac{W_R(x,y)}{(x \vee y)^\eta},\, x,y \in (0,T], \quad \frac{W_1(x)}{x^\eta},\ x \in (0,T], \quad \frac{W_2(y)}{y^\eta},\ y \in (0,T]\right)$$

as $n \to \infty$.

Proof. Define

$$Z_{n,i} = \frac{1}{\sqrt{k}}\delta_{(\frac{n}{k}U_i,\, \frac{n}{k}V_i)}$$

and for all $0 < x, y \le T$ define the functions

$$f_{x,y} = I_{[0,x)\times[0,y)}/(x \vee y)^\eta, \quad f_x^{(1)} = I_{[0,x)\times[0,\infty]}/x^\eta, \quad f_y^{(2)} = I_{[0,\infty]\times[0,y)}/y^\eta.$$

All these f's form the class $\mathcal{F}$. We equip $\mathcal{F}$ with the semi-metric d defined by

$$d(f_{x,y},\, f_{u,v}) = \sqrt{E\left(\frac{W_R(x,y)}{(x\vee y)^\eta} - \frac{W_R(u,v)}{(u\vee v)^\eta}\right)^2},$$

$$d(f_{x,y},\, f_u^{(1)}) = \sqrt{E\left(\frac{W_R(x,y)}{(x\vee y)^\eta} - \frac{W_1(u)}{u^\eta}\right)^2},$$

etc.

For any $\varepsilon > 0$, the bracketing number $N_{[\,]}(\varepsilon, \mathcal{F}, L_2^n)$ is the minimal number of sets N_ε in a partition $\mathcal{F} = \bigcup_{j=1}^{N_\varepsilon} \mathcal{F}_{\varepsilon j}$ of the index set into sets $\mathcal{F}_{\varepsilon j}$ such that, for every partitioning set $\mathcal{F}_{\varepsilon j}$

$$\sum_{i=1}^n E^* \sup_{f,g\in\mathcal{F}_{\varepsilon j}} |\mathcal{Z}_{n,i}(f) - \mathcal{Z}_{n,i}(g)|^2 \le \varepsilon^2. \tag{4.3.4}$$

We will use Theorem 2.11.9 in van der Vaart and Wellner (1996): For each n, let $\mathcal{Z}_{n,1}, \mathcal{Z}_{n,2}, \ldots, \mathcal{Z}_{n,n}$ be independent stochastic processes with finite second moments indexed by a totally bounded semimetric space $(\mathcal{F}, d)$. Suppose

$$\sum_{i=1}^n E^*\|\mathcal{Z}_{n,i}\|_{\mathcal{F}} 1_{\{\|\mathcal{Z}_{n,i}\|_{\mathcal{F}} > \lambda\}} \to 0, \text{ for every } \lambda > 0,$$

where $\|\mathcal{Z}_{n,i}\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |\mathcal{Z}_{n,i}(f)|$, and

$$\int_0^{\delta_n} \sqrt{\log N_{[\,]}(\varepsilon, \mathcal{F}, L_2^n)}\, d\varepsilon \to 0, \text{ for every } \delta_n \downarrow 0.$$

Then the sequence $\sum_{i=1}^n (\mathcal{Z}_{n,i} - E\mathcal{Z}_{n,i})$ is asymptotically tight in $\ell^\infty(\mathcal{F})$ and converges weakly, provided the finite-dimensional distributions converge weakly.

We briefly sketch the total boundedness of $(\mathcal{F}, d)$. We only consider the subclass $\mathcal{F}_2$ of $\mathcal{F}$ consisting of the bivariate $f_{x,y}$'s; moreover we restrict ourselves to the case $x \ge y$, $u \ge v$ and $x \ge u$, $y \ge v$. For any $\delta > 0$, assuming $|x - u| \le \delta$ and $|y - v| \le \delta$, one has

$$\begin{aligned} d^2(f_{x,y},\, f_{u,v}) &= E\left(\frac{W_R(x,y)}{(x\vee y)^\eta} - \frac{W_R(u,v)}{(u\vee v)^\eta}\right)^2 \\ &= E\left(\frac{u^\eta W_R(x,y) - x^\eta W_R(u,v)}{(xu)^\eta}\right)^2 \\ &= \frac{u^{2\eta} R(x,y) - 2x^\eta u^\eta R(u,v) + x^{2\eta} R(u,v)}{(xu)^{2\eta}}. \end{aligned}$$

If $u \leq \delta$, then

$$\begin{aligned} d^2(f_{x,y}, f_{u,v}) &\leq \frac{R(x,y)}{x^{2\eta}} + \frac{2R(u,v)}{u^{2\eta}} + \frac{R(u,v)}{u^{2\eta}} \\ &\leq x^{1-2\eta} + 3u^{1-2\eta} \\ &\leq (2\delta)^{1-2\eta} + 3\delta^{1-2\eta} \leq 5\delta^{1-2\eta}. \end{aligned}$$

If $u > \delta$, then, since

$$\begin{aligned} R(x,y) &\leq R(u,v) + \Lambda([u,x] \times [0,\infty]) + \Lambda([0,\infty] \times [v,y]) \\ &\leq R(u,v) + 2\delta, \end{aligned}$$

we have

$$\begin{aligned} d^2(f_{x,y}, f_{u,v}) &\leq \frac{R(u,v)(u^\eta - x^\eta)^2}{(xu)^{2\eta}} + \frac{2\delta u^{2\eta}}{(xu)^{2\eta}} \\ &\leq u^{1-4\eta}(u^\eta - x^\eta)^2 + 2\delta^{1-2\eta} \\ &\leq u^{1-4\eta} x^{2\eta-2}(x-u)^2 + 2\delta^{1-2\eta} \\ &\leq u^{-1-2\eta}(x-u)^2 + 2\delta^{1-2\eta} \leq 3\delta^{1-2\eta}. \end{aligned}$$

So, since $1 - 2\eta > 0$, we see that for every $\varepsilon > 0$ we can find a $\delta > 0$ such that for $|x - u| \leq \delta$ and $|y - v| \leq \delta$, $d^2(f_{x,y},\ f_{u,v}) < \varepsilon$. Hence, since $[0,T]^2$ is totally bounded with respect to the Euclidean metric, we obtain the total boundedness of $(\mathcal{F}, d)$.

Observe that

$$Z_{n,i}(f_{x,y}) = \frac{1}{\sqrt{k}} I_{\{U_i < \frac{k}{n}x,\, V_i < \frac{k}{n}y\}} / (x \vee y)^\eta,$$

$$\sum_{i=1}^{n} (Z_{n,i} - EZ_{n,i})(f_{x,y}) = v_{n,\eta}(x,y)$$

and similarly for the marginal processes. First we have to show that for every $\lambda > 0$

$$\sum_{i=1}^{n} E||Z_{n,i}||_{\mathcal{F}} I_{\{||Z_{n,i}||_{\mathcal{F}} > \lambda\}} \to 0 \tag{4.3.5}$$

as $n \to \infty$. Again we will restrict ourselves to the subclass $\mathcal{F}_2$. For the univariate $f_x^{(1)}$'s and $f_y^{(2)}$'s, it can be shown in a similar but easier way.

Note that

$$\sup_{f_{x,y} \in \mathcal{F}_2} \frac{1}{\sqrt{k}} I_{\{U_i < \frac{k}{n}x,\, V_i < \frac{k}{n}y\}} / (x \vee y)^\eta \leq \frac{1}{\sqrt{k}} \frac{1}{(\frac{n}{k}(U_i \vee V_i))^\eta},$$

so for each $\lambda > 0$

$$\begin{aligned}
&\sum_{i=1}^{n} E||Z_{n,i}||_{\mathcal{F}_2} I_{\{||Z_{n,i}||_{\mathcal{F}_2}>\lambda\}} \\
&\le \frac{n}{\sqrt{k}} E \frac{1}{(\frac{n}{k}(U_i \vee V_i))^{\eta}} I_{\{\frac{n}{k}(U_i \vee V_i)<(\sqrt{k}\lambda)^{-1/\eta}\}} \\
&= \frac{n}{\sqrt{k}} \int_0^{(\sqrt{k}\lambda)^{-1/\eta}} x^{-\eta} dC(\frac{k}{n}x, \frac{k}{n}x) \\
&= \frac{n}{\sqrt{k}} \Big(\sqrt{k}\lambda C(\frac{k}{n}(\sqrt{k}\lambda)^{-1/\eta}, \frac{k}{n}(\sqrt{k}\lambda)^{-1/\eta}) + \eta \int_0^{(\sqrt{k}\lambda)^{-1/\eta}} C(\frac{k}{n}x, \frac{k}{n}x) x^{-\eta-1} dx\Big) \\
&\le \frac{n}{\sqrt{k}} \Big(\sqrt{k}\lambda \frac{k}{n}(\sqrt{k}\lambda)^{-1/\eta} + \eta \int_0^{(\sqrt{k}\lambda)^{-1/\eta}} \frac{k}{n} x^{-\eta} dx\Big) \\
&= \lambda^{1-1/\eta} k^{1-1/(2\eta)} + \sqrt{k}\frac{\eta}{1-\eta}(\sqrt{k}\lambda)^{1-1/\eta} \\
&= \frac{1}{1-\eta} \lambda^{1-1/\eta} k^{1-1/(2\eta)} \to 0, \quad (\eta < 1/2).
\end{aligned}$$

Next we want to show

$$\int_0^{\delta_n} \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, L_2^n)}\, d\varepsilon \to 0 \tag{4.3.6}$$

for every $\delta_n \downarrow 0$. We present the proof for $T = 1$ for notational convenience; for general $T > 0$ the proof is similar. Let $\varepsilon > 0$ be small, define $a = \varepsilon^{3/(1-2\eta)}$ and $\theta = 1 - \varepsilon^3$. We again consider only $\mathcal{F}_2$; the univariate f's are easier to handle. Define

$$\begin{aligned}
\mathcal{F}(a) &= \{f_{x,y} \in \mathcal{F}_2 : \; x \wedge y \le a\}, \\
\mathcal{F}(l, m) &= \{f_{x,y} \in \mathcal{F}_2 : \; \theta^{l+1} \le x \le \theta^l, \; \theta^{m+1} \le y \le \theta^m\}.
\end{aligned}$$

Then

$$\mathcal{F}_2 = \mathcal{F}(a) \bigcup \left(\bigcup_{m=0}^{[\frac{\log a}{\log \theta}]} \bigcup_{l=0}^{[\frac{\log a}{\log \theta}]} \mathcal{F}(l, m) \right)$$

First check (4.3.4) for $\mathcal{F}(a)$:

$$\begin{aligned}
&\sum_{i=1}^{n} E \sup_{f,g \in \mathcal{F}(a)} (Z_{n,i}(f) - Z_{n,i}(g))^2 = nE \sup_{f,g \in \mathcal{F}(a)} (Z_{n,i}(f) - Z_{n,i}(g))^2 \\
\le\ & 4nE \sup_{f \in \mathcal{F}(a)} Z_{n,i}^2(f) = \frac{4n}{k} E \sup_{\substack{x,y>0 \\ x \wedge y \le a}} I_{\{U_i<\frac{kx}{n}, V_i<\frac{ky}{n}\}}/(x \vee y)^{2\eta} \\
\le\ & \frac{4n}{k} E\,(\frac{n}{k}U_i)^{-2\eta} I_{\{\frac{n}{k}U_i<a\}} = \frac{4n}{k}\int_0^{ak/n} (\frac{n}{k}x)^{-2\eta}\, dx = \frac{4}{1-2\eta} a^{1-2\eta} \le \varepsilon^2.
\end{aligned}$$

Now we consider (4.3.4) for the $\mathcal{F}(l,m)$; w.l.o.g. we take $l \leq m$:

$$\begin{aligned}
&\sum_{i=1}^{n} E \sup_{f,g\in\mathcal{F}(l,m)} (Z_{n,i}(f) - Z_{n,i}(g))^2 \\
&\leq nE\left(\sup_{f\in\mathcal{F}(l,m)} Z_{n,i}(f) - \inf_{f\in\mathcal{F}(l,m)} Z_{n,i}(f)\right)^2 \\
&\leq \frac{n}{k}E\left(I_{\{U_i<\frac{k}{n}\theta^l,\, V_i<\frac{k}{n}\theta^m\}}/(\theta^{l+1}\vee\theta^{m+1})^\eta - I_{\{U_i<\frac{k}{n}\theta^{l+1},\, V_i<\frac{k}{n}\theta^{m+1}\}}/(\theta^l\vee\theta^m)^\eta\right)^2 \\
&= \frac{n}{k}E\Bigg(I_{\{U_i<\frac{k}{n}\theta^l,\, V_i<\frac{k}{n}\theta^m\}}(\frac{1}{\theta^{\eta(l+1)}} - \frac{1}{\theta^{\eta l}}) \\
&\qquad\qquad +(I_{\{U_i<\frac{k}{n}\theta^l,\, V_i<\frac{k}{n}\theta^m\}} - I_{\{U_i<\frac{k}{n}\theta^{l+1},\, V_i<\frac{k}{n}\theta^{m+1}\}})\frac{1}{\theta^{\eta l}}\Bigg)^2 \\
&\leq \frac{2n}{k}\left(C(\frac{k}{n}\theta^l, \frac{k}{n}\theta^m)\frac{1}{\theta^{2\eta l}}(\frac{1}{\theta^\eta}-1)^2 + \left[C(\frac{k}{n}\theta^l, \frac{k}{n}\theta^m) - C(\frac{k}{n}\theta^{l+1}, \frac{k}{n}\theta^{m+1})\right]\frac{1}{\theta^{2\eta l}}\right) \\
&\leq \frac{2n}{k}\left(\frac{k}{n}\frac{\theta^l}{\theta^{2\eta l}}(\frac{1}{\theta^\eta}-1)^2 + \frac{2k}{n}\frac{\theta^l}{\theta^{2\eta l}}(1-\theta)\right) \\
&\leq 2\left(\frac{1}{\theta^{1/2}}-1\right)^2 + 4(1-\theta) \leq \varepsilon^6 + 4\varepsilon^3 \leq \varepsilon^2.
\end{aligned}$$

It is easy to see that the number of elements of the "partition" of $\mathcal{F}_2$ is bounded by ε^{-7}, which yields (4.3.6). Hence we proved the asymptotic tightness condition.

It remains to prove that the finite-dimensional distributions of our process converge weakly. This follows from the fact that multivariate weak convergence follows from weak convergence of linear combinations of the components and the univariate Lindeberg-Feller central limit theorem. It is easily seen that the Lindeberg condition is satisfied for these linear combinations since the elements of $\mathcal{F}$ are weighted indicators and hence bounded. □

Lemma 4.3.2. *For $0 \leq \eta < 1/2$*

$$P\left(\sup_{\substack{x\vee y\leq\varepsilon\\ x,y>0}} \frac{|W_R(x,y)|}{(x\vee y)^\eta} \geq \lambda\right) \leq 16\sum_{m=0}^{\infty}\exp\left(-\frac{\lambda^2}{2}\frac{2^{m(1-2\eta)}}{\varepsilon^{1-2\eta}}\right).$$

Proof. For $m = 0, 1, 2, \ldots$ define

$$\mathcal{A}_m = \{(x,y) : \frac{\varepsilon}{2^{m+1}} \leq x \leq \frac{\varepsilon}{2^m},\ \frac{\varepsilon}{2^{m+1}} \leq y \leq \varepsilon\}.$$

Then, with Z a standard normal random variable,

$$
\begin{aligned}
& P\left(\sup_{\substack{x\vee y\le\varepsilon\\ 0<x\le y}}\frac{|W_R(x,y)|}{(x\vee y)^\eta}\ge\lambda\right) = P\left(\sup_{\substack{x\vee y\le\varepsilon\\ 0<x\le y}}\frac{|W_R(x,y)|}{y^\eta}\ge\lambda\right)\\
\le\ & P\left(\sup_{m\in\{0,1,2,\ldots\}}\ \sup_{(x,y)\in\mathcal{A}_m}\frac{|W_R(x,y)|}{y^\eta}\ge\lambda\right)\\
\le\ & \sum_{m=0}^{\infty} P\left(\sup_{(x,y)\in\mathcal{A}_m}|W_R(x,y)|\ge\lambda\left(\frac{\varepsilon}{2^{m+1}}\right)^\eta\right)\\
\le\ & 4\sum_{m=0}^{\infty} P\left(|W_R(\frac{\varepsilon}{2^m},\varepsilon)|\ge\lambda\left(\frac{\varepsilon}{2^{m+1}}\right)^\eta\right)\le 4\sum_{m=0}^{\infty} P\left(|Z|\ge\frac{\lambda}{2^\eta}\left(\frac{2^m}{\varepsilon}\right)^{1/2-\eta}\right)\\
\le\ & 8\sum_{m=0}^{\infty}\exp\left(-\frac{\lambda^2}{2}\frac{2^{m(1-2\eta)}}{\varepsilon^{1-2\eta}}\right),
\end{aligned}
$$

where the third inequality follows for instance from an adaptation of Lemma 1.2 in Orey and Pruitt (1973) and the last inequality from Mill's ratio. A symmetry argument completes the proof. □

By Theorem 2 in Einmahl *et al.* (2001) and Proposition 4.3.1 (and their proofs) it follows that

$$
\begin{aligned}
&\left(\sqrt{k}(\hat\Phi(\theta)-\Phi(\theta)), v_{n,\eta}(x,y), v_{n,\eta,1}(u), v_{n,\eta,2}(v)\right)\\
&\xrightarrow{d}\left(W_\Lambda(C_\theta)+Z(\theta)), \frac{W_R(x,y)}{(x\vee y)^\eta}, \frac{W_1(u)}{u^\eta}, \frac{W_2(v)}{v^\eta}\right),
\end{aligned}
$$

on $D[0,\,\pi/2]\times D[0,\,T]^2\times D[0,T]\times D[0,T]$. By the Skorohod construction, there exists now a probability space carrying $\hat\Phi^*$, v_n^*, $v_{n,1}^*$, $v_{n,2}^*$, $W_\Lambda^*(C.)$, Z^*, W_R^*, W_1^* and W_2^* such that

$$
\begin{aligned}
\left(\hat\Phi^*, v_n^*, v_{n,1}^*, v_{n,2}^*\right) &\stackrel{d}{=}\left(\hat\Phi, v_n, v_{n,1}, v_{n,2}\right),\\
\left(W_\Lambda^*(C.), Z^*, W_R^*, W_1^*, W_2^*\right) &\stackrel{d}{=}\left(W_\Lambda(C.), Z, W_R, W_1, W_2\right)
\end{aligned}
$$

and for $0\le\eta<1/2$

$$
D_n := \sup_{0\le\theta\le\pi/2}\left|\sqrt{k}(\hat\Phi^*(\theta)-\Phi(\theta)) - (W_\Lambda^*(C_\theta)+Z^*(\theta))\right| = o_P(1), \quad (4.3.7)
$$

$$
\sup_{0<x,y\le T}\frac{|v_n^*(x,y)-W_R^*(x,y)|}{(x\vee y)^\eta} = o_P(1), \quad (4.3.8)
$$

$$
\sup_{0<x\le T}\frac{|v_{n,1}^*(x)-W_1^*(x)|}{x^\eta} = o_P(1), \quad (4.3.9)
$$

$$\sup_{0<x,y\le T}\frac{|v^*_{n,2}(y)-W^*_2(y)|}{y^\eta}=o_P(1), \tag{4.3.10}$$

as $n\to\infty$. Henceforth we will work on this probability space, but drop the * from the notation.

Proof of Theorem 4.2.1. By Lemma 4.3.1

$$\begin{aligned}&\sqrt{k}(\hat{l}_1(x,y)-l(x,y))\\&=\begin{cases}x\sqrt{k}(\hat{\Phi}(\frac{\pi}{2})-\Phi(\frac{\pi}{2}))+y\int_{\pi/4}^{\arctan\frac{y}{x}}\frac{1}{\sin^2\theta}\sqrt{k}(\hat{\Phi}(\theta)-\Phi(\theta))d\theta, & \text{if } y\ge x,\\ x\sqrt{k}(\hat{\Phi}(\frac{\pi}{2})-\Phi(\frac{\pi}{2}))-x\int_{\arctan\frac{y}{x}}^{\pi/4}\frac{1}{\cos^2\theta}\sqrt{k}(\hat{\Phi}(\theta)-\Phi(\theta))d\theta, & \text{if } y<x.\end{cases}\end{aligned}$$

Now, let's first consider the case $y\ge x$.

$$\begin{aligned}&\sup_{0<x\le y\le 1}\left|\frac{\sqrt{k}(\hat{l}_1(x,y)-l(x,y))-A(x,y)}{x\vee y}\right|\\&=\frac{1}{x\vee y}\Big|x\left(\sqrt{k}(\hat{\Phi}^*(\frac{\pi}{2})-\Phi(\frac{\pi}{2}))-(W^*_\Lambda(C_{\frac{\pi}{2}})-Z^*(\frac{\pi}{2})\right)\\&\qquad+y\int_{\pi/4}^{\arctan\frac{y}{x}}\frac{1}{\sin^2\theta}\left(\sqrt{k}(\hat{\Phi}^*(\theta)-\Phi(\theta))-(W^*_\Lambda(C_\theta)-Z^*(\theta))\right)d\theta\Big|+o_P(1)\\&\le\frac{xD_n}{x\vee y}+\frac{yD_n}{x\vee y}\int_{\pi/4}^{\pi/2}\frac{1}{\sin^2\theta}d\theta+o_P(1)\to 0,\end{aligned}$$

in probability as $n\to\infty$. In case of $y<x$, the proof is similar. □

Let Q_{1n} and Q_{2n} be the empirical quantile functions of the $\{U_i\}_{i=1}^n$ and $\{V_i\}_{i=1}^n$, respectively. Define

$$\hat{R}(x,y)=\frac{1}{k}\sum_{i=1}^n I_{\{U_i<Q_{1n}(kx/n),\,V_i<Q_{2n}(ky/n)\}}.$$

Note that by (4.1.10)

$$\hat{l}_2(x,y)=\frac{1}{k}\sum_{i=1}^n I_{\{U_i<Q_{1n}(kx/n)\ or\ V_i<Q_{2n}(ky/n)\}}.$$

Proof of Theorem 4.2.2. It is easily seen that $\hat{l}_2(x,y)+\hat{R}(x,y)=(\lceil kx\rceil+$

$\lceil ky \rceil - 2)/k \leq ([kx] + [ky])/k$, for each $x, y \in (0, 1]$, almost surely. So we have

$$
\begin{aligned}
&\sup_{\substack{0<x,y\leq 1\\ x\vee y\geq 1/k}} \frac{|\sqrt{k}(\hat{l}_2(x,y) - l(x,y)) + \sqrt{k}(\hat{R}(x,y) - R(x,y))|}{(x\vee y)^\eta}\\
&\overset{a.s.}{=} \sup_{\substack{0<x,y\leq 1\\ x\vee y\geq 1/k}} \frac{\left|\sqrt{k}(\frac{1}{k}(\lceil kx\rceil + \lceil ky\rceil - 2) - (x+y))\right|}{(x\vee y)^\eta}\\
&\leq k^{-\eta} \sup_{0<x,y\leq 1} \sqrt{k}(x + y - ([kx] + [ky])/k)\\
&\leq 2\sqrt{k}\cdot k^{\eta-1} = 2k^{\eta-1/2} \to 0.
\end{aligned}
$$

Write $S_{jn}(x) = \frac{n}{k} Q_{jn}(\frac{k}{n}x)$, $j = 1, 2$. Then we have

$$
\begin{aligned}
&\sup_{\substack{0<x,y\leq 1\\ x\vee y\geq 1/k}} \frac{|\sqrt{k}(\hat{l}_2(x,y) - l(x,y)) + W_R(x,y) - R_1(x,y)W_1(x) - R_2(x,y)W_2(y)|}{(x\vee y)^\eta}\\
&\overset{a.s.}{=} \sup_{\substack{0<x,y\leq 1\\ x\vee y\geq 1/k}} \frac{|\sqrt{k}(\hat{R}(x,y) - R(x,y)) - W_R(x,y) + R_1(x,y)W_1(x) + R_2(x,y)W_2(y)|}{(x\vee y)^\eta}\\
&\quad + o(1)\\
&= \sup_{\substack{0<x,y\leq 1\\ x\vee y\geq 1/k}} \frac{|\sqrt{k}(\hat{R}(x,y) - R_n(S_{1n}(x), S_{2n}(y))) - W_R(x,y)|}{(x\vee y)^\eta}\\
&\quad + \sup_{\substack{0<x,y\leq 1\\ x\vee y\geq 1/k}} \frac{|\sqrt{k}(R_n(S_{1n}(x), S_{2n}(y))) - R(S_{1n}(x), S_{2n}(y))|}{(x\vee y)^\eta}\\
&\quad + \sup_{\substack{0<x,y\leq 1\\ x\vee y\geq 1/k}} \frac{|\sqrt{k}(R(S_{1n}(x), S_{2n}(y)) - R(x,y)) + R_1(x,y)W_1(x,y) + R_2(x,y)W_2(y)|}{(x\vee y)^\eta}\\
&\quad + o(1)\\
&=: D_1 + D_2 + D_3 + o(1).
\end{aligned}
$$

We will show that $D_j \to 0$ in probability, $j = 1, 2, 3$. We have

$$
\begin{aligned}
D_1 &= \sup_{\substack{0<x,y\leq 1\\ x\vee y\geq 1/k}} \frac{|\sqrt{k}(T_n(S_{1n}(x), S_{2n}(y)) - R_n(S_{1n}(x), S_{2n}(y))) - W_R(x,y)|}{(x\vee y)^\eta}\\
&\leq \sup_{\substack{0<x,y\leq 1\\ x\vee y\geq 1/k}} \frac{|\sqrt{k}(T_n(S_{1n}(x), S_{2n}(y)) - R_n(S_{1n}(x), S_{2n}(y))) - W_R(S_{1n}(x), S_{2n}(y))|}{(S_{1n}(x)\vee S_{2n}(y))^\eta}\\
&\quad \cdot \left(\frac{S_{1n}(x)\vee S_{2n}(y)}{x\vee y}\right)^\eta + \sup_{\substack{0<x,y\leq 1\\ x\vee y\geq 1/k}} \frac{|W_R(S_{1n}(x), S_{2n}(y)) - W_R(x,y)|}{(x\vee y)^\eta}
\end{aligned}
$$

$$\leq \sup_{0<s,t\leq 2} \frac{|v_n(s,t) - W_R(s,t)|}{(s \vee t)^\eta} \cdot \sup_{\substack{0<s,t\leq k/n \\ s\vee t\geq 1/n}} \left(\frac{Q_{1n}(s) \vee Q_{2n}(t)}{s \vee t} \right)^\eta$$
$$+ \sup_{\substack{0<x,y\leq 1 \\ x\vee y\geq 1/k}} \frac{|W_R(S_{1n}(x), S_{2n}(y)) - W_R(x,y)|}{(x \vee y)^\eta}$$
$$=: D_{11} \cdot D_{12} + D_{13},$$

where the last inequality holds with arbitrarily high probability. Then $D_{11} \to 0$ in probability because of (4.3.8) with $T = 2$. It is well known that

$$\sup_{s\geq 1/n} \frac{Q_{jn}(s)}{s} = O_P(1), \quad j = 1, 2 \tag{4.3.11}$$

(see Shorack and Wellner (1986), p. 419). Hence $D_{11} \cdot D_{12} \to 0$, in probability. Now consider for each $\varepsilon > 1/k$

$$D_{13} \leq \sup_{\substack{0<x,y\leq 1 \\ x\vee y\geq \varepsilon}} \frac{|W_R(S_{1n}(x), S_{2n}(y)) - W_R(x,y)|}{\varepsilon^\eta}$$
$$+ \sup_{\substack{0<x,y\leq 1 \\ 1/k\leq x\vee y\leq \varepsilon}} \frac{|W_R(S_{1n}(x), S_{2n}(y))|}{(S_{1n}(x) \vee S_{2n}(y))^\eta} \cdot \sup_{s,t\geq 1/n} \left(\frac{Q_{1n}(s) \vee Q_{2n}(y)}{s \vee t} \right)^\eta$$
$$+ \sup_{\substack{0<x,y\leq 1 \\ 1/k\leq x\vee y\leq \varepsilon}} \frac{|W_R(x,y)|}{(x \vee y)^\eta}$$
$$=: D_{14} + D_{15} + D_{16}.$$

By the (uniform) continuity of W_R and the fact that

$$\sup_{0<t\leq k/n} \frac{n}{k} |Q_{jn}(t) - t| \to 0, \quad a.s., \quad j = 1, 2, \tag{4.3.12}$$

$D_{14} \to 0$ in probability a.s. for any $\varepsilon > 0$. Let $\delta > 0$, by (4.3.11) and Lemma 4.3.2 we see that for large n, $P(D_{15} \geq \delta) \leq \delta$ for $\varepsilon > 0$ small enough. Again from Lemma 4.3.2 we have $P(D_{16} \geq \delta) \leq \delta$. Hence $D_{13} \to 0$ in probability and consequently $D_1 \to 0$, in probability.

Consider D_2. Take (a, b) with $a \vee b = u$. Then according to (4.2.5)

$$\frac{1}{t} C(ta, tb) = \frac{u}{ut} C(tu\frac{a}{u}, tu\frac{b}{u})$$
$$= uR(\frac{a}{u}, \frac{b}{u}) + u^{1+\alpha} O(t^\alpha)$$
$$= R(a,b) + (a \vee b)^{1+\alpha} O(t^\alpha).$$

Now with arbitrarily high probability

$$D_2 \leq \sup_{0<x,y\leq 2} \frac{|\sqrt{k}(R_n(x,y) - R(x,y))|}{(x \vee y)^\eta} \cdot \sup_{s\vee t\geq 1/n} \left(\frac{Q_{1n}(s) \vee Q_{2n}(t)}{s \vee t} \right)^\eta .$$

We have seen before that second term of this product is $O_P(1)$. So it is suffices to show that the first term is $o(1)$:

$$\sup_{0<x,y\le 2} \frac{|\sqrt{k}(R_n(x,y)-R(x,y))|}{(x\vee y)^\eta} = \left(\sup_{0<x,y\le 2} \frac{\sqrt{k}(x\vee y)^{1+\alpha}}{(x\vee y)^\eta}\right) O\left(\left(\frac{k}{n}\right)^\alpha\right)$$
$$= O\left(\frac{k^{\alpha+1/2}}{n^\alpha}\right) = o(1),$$

by assumption. Hence $D_2 \to 0$ in probability.

It remains to show that $D_3 \to 0$ in probability. By two applications of the mean-value theorem we obtain

$$\begin{aligned} &R(S_{1n}(x), S_{2n}(y)) - R(x,y) \\ &= R(S_{1n}(x), S_{2n}(y)) - R(x, S_{2n}(y)) + R(x, S_{2n}(y)) - R(x,y) \\ &= R_1(\theta_{1n}, S_{2n}(y))(S_{1n}(x) - x) + R_2(x, \theta_{2n})(S_{2n}(y) - y) \end{aligned}$$

with θ_{1n} between x and $S_{1n}(x)$ and θ_{2n} between y and $S_{2n}(y)$. So

$$\begin{aligned} D_3 \le \sup_{\substack{0<x,y\le 1\\ x\vee y\ge 1/k}} &\frac{|R_1(\theta_{1n}, S_{2n}(y))\sqrt{k}(S_{1n}(x)-x) + R_1(x,y)W_1(x)|}{(x\vee y)^\eta} \\ + \sup_{\substack{0<x,y\le 1\\ x\vee y\ge 1/k}} &\frac{|R_2(x,\theta_{2n})\sqrt{k}(S_{2n}(y)-y) + R_2(x,y)W_2(y)|}{(x\vee y)^\eta}. \end{aligned}$$

We consider only the first term in the right hand side of this expression; the second one can be dealt with similarly. Write $z_n(x) = \sqrt{k}(S_{1n}(x) - x)$. From (4.3.9) with $\eta = 0$ it follows that

$$\sup_{0<x\le 1} |z_n(x) + W_1(x)| \to 0$$

in probability. From this it can be shown that for $0 \le \eta < 1/2$

$$\sup_{1/k\le x\le 1} \frac{|z_n(x) + W_1(x)|}{x^\eta} \to 0 \tag{4.3.13}$$

in probability (see, e.g., Einmahl (1992)). Now

$$\begin{aligned} &\sup_{\substack{0<x,y\le 1\\ x\vee y\ge 1/k}} \frac{|R_1(\theta_{1n}, S_{2n}(y))z_n(x) + R_1(x,y)W_1(x)|}{(x\vee y)^\eta} \\ &\le \sup_{0<x,y\le 1} R_1(\theta_{1n}, S_{2n}(y)) \cdot \sup_{1/k\le x\le 1} \frac{|z_n(x) + W_1(x)|}{x^\eta} \\ &\quad + \sup_{0<x,y\le 1} |R_1(x,y) - R_1(\theta_{1n}, S_{2n}(y))| \cdot \sup_{0<x\le 1} \frac{|W_1(x)|}{x^\eta} \\ &=: D_{31} + D_{32}. \end{aligned}$$

Since R_1 is continuous on $[0,2]^2$ it is uniformly continuous and bounded. This together with (4.3.13) yields $D_{31} \to 0$ in probability. The uniform continuity of R_1 together with (4.3.12) and the fact that

$$\sup_{0<x\leq 1} \frac{|W_1(x)|}{x^\eta} < \infty \quad a.s.,$$

yields $D_{32} \to 0$ in probability and consequently $D_3 \to 0$ in probability.

Finally we show that

$$\sup_{0<x,y<1/k} \frac{|\sqrt{k}(\hat{l}_2(x,y) - l(x,y)) + B(x,y)|}{(x \vee y)^\eta} = o_P(1).$$

Observing that $\sup_{0<x,y<1/k} \hat{l}_2(x,y) = 0$ a.s., this follows easily. □

Proof of Theorem 4.2.3. For each $0 \leq \beta < 3$, there exist $\alpha \in [0,2)$ and $\eta \in [0,1/2)$ such that $\beta = \alpha + 2\eta$. By Theorem 4.2.1 and Theorem 4.2.2, and

$$\int_0^1 \int_0^1 \frac{1}{(x \vee y)^\alpha} \, dxdy < \infty,$$

it follows that as $n \to \infty$

$$\begin{aligned}
&\iint_{0<x,y\leq 1} \frac{k\left(\hat{l}_1(x,y) - \hat{l}_2(x,y)\right)^2}{(x \vee y)^\beta} \, dxdy \\
&= o_P(1) \iint_{0<x,y\leq 1} \frac{1}{(x \vee y)^\alpha} \, dxdy + \iint_{0<x,y\leq 1} \frac{(A(x,y) + B(x,y))^2}{(x \vee y)^\beta} \, dxdy \\
&\xrightarrow{d} \iint_{0<x,y\leq 1} \frac{(A(x,y) + B(x,y))^2}{(x \vee y)^\beta} \, dxdy \, .
\end{aligned}$$

□

4.4 Approximating the limit

For testing purposes, we have to find the probability distribution of the limiting random variable in Theorem 4.2.3. This can be done by simulating the processes A and B, but unfortunately their distributions depend on the unknown measure Λ. Therefore, we generate approximations A_n and B_n, respectively, of the processes A and B, not with parameter Λ but with approximated parameter Λ_n. In this section, we consider the convergence of the sequence of these approximated limiting random variables. Until further notice, we take $\{\Lambda_n\}_{n\geq 1}$ to be a sequence of *deterministic* measures.

Define

$$R_{1n}(x,y) := \frac{1}{2}k^{1/5}\Lambda_n([x-k^{-1/5}, x+k^{-1/5}]\times[0,y)),$$
$$R_{2n}(x,y) := \frac{1}{2}k^{1/5}\Lambda_n([0,x)\times[y-k^{-1/5}, y+k^{-1/5}]),$$
$$W_{1n}(x) := W_{\Lambda_n}([0,x]\times[0,\infty]), \quad W_{2n}(y) := W_{\Lambda_n}([0,\infty]\times[0,y]),$$
$$W_{R_n}(x,y) := W_{\Lambda_n}([0,x]\times[0,y]),$$

and the process B_n by

$$B_n(x,y) := W_{R_n}(x,y) - R_{1n}(x,y)W_{1n}(x) - R_{2n}(x,y)W_{2n}(y).$$

Based on the definition of Z in (4.2.1) and the homogeneity property of λ (i.e., $\lambda(tx,ty) = \frac{1}{t}\lambda(x,y)$), we define the approximating process Z_n by

$$Z_n(\theta) = \begin{cases} \lambda_n(1,\tan\theta)\tan\theta\displaystyle\int_0^{\frac{1}{\tan\theta}}\frac{W_{1n}(x)}{x}dx - \lambda_n(1,\tan\theta)\int_0^1\frac{W_{2n}(x)}{x}dx \\ \qquad - W_{2n}(1)\displaystyle\int_{\frac{1}{\tan\theta}}^{\infty}\lambda_n(x,1)dx, & \theta\in[0,\frac{\pi}{4}] \\ \lambda_n(\frac{1}{\tan\theta},1)\displaystyle\int_0^1\frac{W_{1n}(x)}{x}dx - \lambda_n(\frac{1}{\tan\theta},1)\frac{1}{\tan\theta}\int_0^{\tan\theta}\frac{W_{2n}(x)}{x}dx \\ \qquad - W_{2n}(1)\displaystyle\int_1^{\infty}\lambda_n(x,1)dx - W_{1n}(1)\int_1^{\tan\theta}\lambda_n(1,y)dy, & \theta\in(\frac{\pi}{4},\frac{\pi}{2}) \\ -W_{2n}(1)\displaystyle\int_1^{\infty}\lambda_n(x,1)dx - W_{1n}(1)\int_1^{\infty}\lambda_n(1,y)dy, & \theta=\frac{\pi}{2} \end{cases} \tag{4.4.1}$$

where λ_n is the approximation of λ defined by

$$\lambda_n(1,y) := \frac{1}{4}k^{1/3}\Lambda_n([1-k^{-1/6}, 1+k^{-1/6}]\times[y-k^{-1/6}, y+k^{-1/6}]), \quad y>0,$$
$$\lambda_n(x,1) := \frac{1}{4}k^{1/3}\Lambda_n([x-k^{-1/6}, x+k^{-1/6}]\times[1-k^{-1/6}, 1+k^{-1/6}]), \quad x>0.$$

Finally define the process A_n by

$$A_n(x,y) := \begin{cases} x(W_{\Lambda_n}(C_{\frac{\pi}{2}}) + Z_n(\frac{\pi}{2})) + y\int_{\pi/4}^{\arctan\frac{y}{x}}\frac{1}{\sin^2\theta}(W_{\Lambda_n}(C_\theta)+Z_n(\theta))d\theta & \text{if } y\ge x, \\ x(W_{\Lambda_n}(C_{\frac{\pi}{2}}) + Z_n(\frac{\pi}{2})) - x\int_{\arctan\frac{y}{x}}^{\pi/4}\frac{1}{\cos^2\theta}(W_{\Lambda_n}(C_\theta)+Z_n(\theta))d\theta & \text{if } y< x. \end{cases}$$

First we consider the weak convergence of the weighted approximating processes. We write $D_2 := D([0,1]^2)$ for the generalization of $D[0,1]$ to dimension 2, and $\mathcal{L}_d$ for the Borel σ-algebra on (D_2, d), where d is the metric on D_2 defined in Neuhaus (1971).

Proposition 4.4.1. *Let Λ be as in Theorem 4.2.3. Suppose that $\{\Lambda_n\}_{n\geq 1}$ is a sequence of measures on $[0,\infty]^2 \setminus \{(\infty,\infty)\}$ satisfying that for each $x, y \geq 0$*

$$\Lambda_n([0,x]\times[0,\infty]) = [kx]/k, \quad \Lambda_n([0,\infty]\times[0,y]) = [ky]/k \tag{4.4.2}$$

and

$$\sup_{0<x,y\leq 1} |\Lambda_n([0,x]\times[0,y]) - \Lambda([0,x]\times[0,y])| \to 0 \tag{4.4.3}$$

as $n\to\infty$. Further suppose that

$$\sup_{0<x\leq 1} |\lambda_n(x,1) - \lambda(x,1)| \to 0, \quad \sup_{0<y\leq 1} |\lambda_n(1,y) - \lambda(1,y)| \to 0, \tag{4.4.4}$$

$$\sup_{0<x,y\leq 1} |R_{jn}(x,y) - R_j(x,y)| \to 0, \quad j = 1,2, \tag{4.4.5}$$

as $n\to\infty$. Then for each $0\leq \eta < 1/2$

$$\left\{\frac{A_n(x,y)+B_n(x,y)}{(x\vee y)^\eta}, (x,y)\in[0,1]^2\right\} \to \left\{\frac{A(x,y)+B(x,y)}{(x\vee y)^\eta}, (x,y)\in[0,1]^2\right\},$$

weakly in D_2.

Before proving this proposition, we present three corollaries. The last one is the main result of this section.

Corollary 4.4.1. *Under the conditions of Proposition 4.4.1 for each $0\leq\beta<3$*

$$\iint_{0<x,y\leq 1} \frac{(A_n(x,y)+B_n(x,y))^2}{(x\vee y)^\beta}\,dxdy \xrightarrow{d} \iint_{0<x,y\leq 1} \frac{(A(x,y)+B(x,y))^2}{(x\vee y)^\beta}\,dxdy \tag{4.4.6}$$

as $n\to\infty$.

Let Q_{Λ_n} be the quantile function of the random variable on the left hand side of (4.4.6) and Q_Λ the quantile function of the random variable on the right hand side of (4.4.6).

Corollary 4.4.2. *Under the conditions of Proposition 4.4.1, for each $0\leq\beta<3$ and for each continuity point $1-\alpha$ $(0<\alpha<1)$ of Q_Λ,*

$$\lim_{n\to\infty} Q_{\Lambda_n}(1-\alpha) = Q_\Lambda(1-\alpha).$$

Next, with abuse of notation, we estimate Λ_n from the data, so it becomes random. In Einmahl *et al.* (2001), Λ_n is defined as

$$\begin{aligned}\Lambda_n(A) \quad &:= \quad \frac{1}{k}\sum_{i=1}^{n} I_{\frac{k}{n}A}\left(\frac{1}{n}\sum_{j=1}^{n} I_{(-\infty,U_i]}(U_j),\ \frac{1}{n}\sum_{j=1}^{n} I_{(-\infty,V_i]}(V_j)\right) \qquad (4.4.7)\\ &= \quad \frac{1}{k}\sum_{i=1}^{n} I_{kA}\left(n+1-R_i^X, n+1-R_i^Y\right)\end{aligned}$$

where $U_i := 1-F_1(X_i)$, $V_i := 1-F_2(Y_i)$ for $i = 1, 2, ..., n$. Note that for $x, y > 0$

$$\Lambda_n([0,x)\times[0,y)) = \frac{1}{k}\sum_{i=1}^{n} I_{\{U_i<Q_{1n}(kx/n),\, V_i<Q_{2n}(ky/n)\}}.$$

So $\Lambda_n([0,x)\times[0,\infty]) = (\lceil kx\rceil - 1)/k \le [kx]/k = \Lambda_n([0,x]\times[0,\infty])$ a.s. and $\Lambda_n([0,\infty]\times[0,y)) = (\lceil ky\rceil - 1)/k \le [ky]/k = \Lambda_n([0,\infty]\times[0,y])$ a.s.

The final and main corollary deals with the *random* measures Λ_n, where the functions derived from Λ_n, like λ_n, are defined as before. In particular, we define Q_{Λ_n}, as the quantile function of the random variable on the left hand side of (4.4.6), *conditional* on Λ_n, so it is also random.

Corollary 4.4.3. *Let Λ_n be as in (4.4.7). Under the conditions of Theorem 4.2.3, we have for each $0 \le \beta < 3$ and each continuity point $1-\alpha$ ($0 < \alpha < 1$) of Q_Λ, that*

$$Q_{\Lambda_n}(1-\alpha) \xrightarrow{P} Q_\Lambda(1-\alpha), \quad \text{as} \ \ n\to\infty.$$

For testing purposes, Corollary 4.4.3 shows that simulation of the limiting random variable in Theorem 4.2.3 with Λ replaced with the estimated Λ_n is asymptotically correct.

Now we turn to the proofs. In order to prove Proposition 4.4.1, by Prohorov's theorem it is necessary and sufficient to prove that

(i) The finite-dimensional distributions of $\{(A_n(x,y)+B_n(x,y))/(x\vee y)^\eta,\ (x,y)\in[0,1]^2\}_{n\ge1}$ converge to those of $\{(A(x,y)+B(x,y))/(x\vee y)^\eta,\ (x,y)\in[0,1]^2\}$,

(ii) $\{(A_n(x,y)+B_n(x,y))/(x\vee y)^\eta,\ (x,y)\in[0,1]^2\}_{n\ge1}$ is relatively compact.

For the relative compactness, we need several lemmas. First we present in Lemma 4.4.1 sufficient conditions for relative compactness ; the proof is similar to that of Theorem 15.5 in Billingsley (1968), see also Neuhaus (1971).

Lemma 4.4.1. *Let P_n be probability measures on $(D_2, \mathcal{L}_d)$. Suppose that, for each positive η, there exists an $M > 0$ such that*

$$P_n(x\in D_2 : |x(0,0)| > M) \le \eta, \quad n\ge 1.$$

Suppose further that, for each positive ε *and* η*, there exist a* δ*,* $0<\delta<1$*, and an integer* n_0 *such that*

$$P_n(x \in D_2 : \sup_{|u_1-u_2|\le\delta, |v_1-v_2|\le\delta} |x(u_1,v_1)-x(u_2,v_2)| > \varepsilon) \le \eta, \quad n \ge n_0.$$

Then $\{P_n\}_{n\ge1}$ *is relatively compact.*

Lemma 4.4.2. *Under the conditions of Proposition 4.4.1, for each* c*,* $a>0$

(i) $\int_0^c \frac{W_{jn}(t)}{t} dt \sim N(0,\sigma_n^2)$*, with* $\sigma_n^2 \le 2c$*,* $\quad j=1,2$*,*

(ii) $P(\sup_{t\ge c}|\frac{W_{jn}(t)}{t}| \ge a) \le 2P(|W(2/c)| \ge a)$*,* $j=1,2$*, where* W *is a standard Wiener process.*

Proof. (i) This follows from Proposition 1, page 42, in Shorack and Wellner (1986).

(ii) Let W be a standard Wiener process. Since $\{W(t)/t, t\ge c\} =^d \{W(1/t), t \ge c\}$, then

$$P(\sup_{t\ge c}|W(t)/t| \ge a) = P(\sup_{0<s\le 1/c}|W(s)| \ge a) \le 2P(|W(1/c)| \ge a).$$

Write $\Lambda_{1n}(t)$ for $\Lambda_n([0,t]\times[0,\infty])$. Since $\{W_{1n}(t), t>0\} \stackrel{d}{=} \{W(\Lambda_{1n}(t)), t>0\}$, then

$$P(\sup_{t\ge c}|W_{1n}(t)/t| \ge a) = P(\sup_{t\ge c}\left|\frac{W(\Lambda_{1n}(t))\cdot\Lambda_{1n}(t)}{\Lambda_{1n}(t)\cdot t}\right| \ge a)$$
$$\le P(\sup_{\Lambda_{1n}(t)\ge c/2}|W(\Lambda_{1n}(t))/\Lambda_{1n}(t)| \ge a) \le 2P(|W(2/c)| \ge a),$$

eventually (since $t-1/k \le \Lambda_{1n}(t) \le t$). For $j=2$ the proof is the same. □

Lemma 4.4.3. *Define*

$$H_n := \sup_{\theta\in[0,\pi/2]} |W_{\Lambda_n}(C_\theta) + Z_n(\theta)|.$$

Then under the conditions of Proposition 4.4.1, there exists an $n_0 \in \mathbb{N}$ *such that*

$$\sup_{n\ge n_0} P(H_n \ge a) = O(e^{-a}) \text{ as } a\to\infty.$$

Proof. Define $H_{1n} := \sup_{\theta\in[0,\pi/4]}|W_{\Lambda_n}(C_\theta)+Z_n(\theta)|$, $H_{2n} := \sup_{\theta\in(\pi/4,\pi/2)}|W_{\Lambda_n}(C_\theta)+Z_n(\theta)|$, and $H_{3n} := |W_{\Lambda_n}(C_{\pi/2})+Z_n(\pi/2)|$. It suffices to verify that there exists an $n_0\in\mathbb{N}$ such that

$$\sup_{n\ge n_0} P(H_{jn} \ge a) = O(e^{-a}), \quad j=1,2,3$$

as $a\to\infty$. Here we only check it in case of $j=1$. For the other two cases, the proofs are similar.

Since for all $n \geq 1$

$$\{W_{\Lambda_n}(C_\theta), \theta \in [0, \pi/2]\} \stackrel{d}{=} \{W(\Lambda_n(C_\theta)), \theta \in [0, \pi/2]\},$$

with W a standard Wiener process, we have

$$\begin{aligned} P(H_{1n} \geq a) &\leq P(\sup_{\theta \in [0,\pi/4]} |W(\Lambda_n(C_\theta))| \geq a/2) + P(\sup_{\theta \in [0,\pi/4]} |Z_n(\theta)| \geq a/2) \\ &\leq 2P(|W(\Lambda_n(C_{\pi/4}))| \geq a/2) + P(\sup_{\theta \in [0,\pi/4]} |Z_n(\theta)| \geq a/2). \end{aligned}$$

Clearly $\Lambda_n(C_{\pi/4}) \leq 1$ for all $n \geq 1$, and hence $\sup_{n \geq 1} P(|W(\Lambda_n(C_{\pi/4}))| \geq a/2) = O(e^{-a})$, as $a \to \infty$.

From Einmahl *et al.* (2001), one has $\sup_{x>0} \lambda(x, 1) < \infty$ and $\sup_{y>0} \lambda(1, y) < \infty$. Then by (4.4.4) there exists a constant $\lambda_0 > 0$ such that $\sup_{0<x\leq 1} \lambda_n(x, 1) < \lambda_0$ and $\sup_{0<y\leq 1} \lambda_n(1, y) < \lambda_0$ for large n. Using (4.4.2) and the fact that Λ_n is a step function, one can prove with some effort that $\int_1^\infty \lambda_n(x, 1)dx \leq 2$ and $\int_1^\infty \lambda_n(1, y)dy \leq 2$ for sufficiently large n, hence by the definition of $Z_n(\theta)$, one has

$$\begin{aligned} &\sup_{\theta \in [0,\pi/4]} |Z_n(\theta)| \\ &\leq \lambda_0 \left| \int_0^1 \frac{W_{1n}(x)}{x} dx \right| + \lambda_0 \sup_{\theta \in [0,\pi/4]} \left| \tan\theta \int_1^{1/\tan\theta} \frac{W_{1n}(x)}{x} dx \right| \\ &\quad + \lambda_0 \left| \int_0^1 \frac{W_{2n}(x)}{x} dx \right| + 2 |W_{2n}(1)| \\ &\leq \lambda_0 \left| \int_0^1 \frac{W_{1n}(x)}{x} dx \right| + \lambda_0 \sup_{x \geq 1} \left| \frac{W_{1n}(x)}{x} \right| + \lambda_0 \left| \int_0^1 \frac{W_{2n}(x)}{x} dx \right| + 2 |W_{2n}(1)|, \end{aligned}$$

for sufficiently large n. By Lemma 4.4.2(i), $\int_0^1 \frac{W_{1n}(x)}{x} dx$ and $\int_0^1 \frac{W_{2n}(x)}{x} dx$ have centered normal distributions with uniformly bounded variances for all $n \geq 1$. By Lemma 4.4.2(ii) there exist an $n_0 \in \mathbb{N}$ such that

$$\sup_{n \geq n_0} P(\lambda_0 \sup_{x \geq 1} |W_{1n}(x)|/x \geq a/8) \leq 2P(W(2) \geq a/(8\lambda_0)) = O(e^{-a})$$

as $a \to \infty$. Hence

$$\sup_{n \geq n_0} P(\sup_{\theta \in [0,\pi/4]} |Z_n(\theta)| \geq a/2) = O(e^{-a})$$

as $a \to \infty$. So $\sup_{n \geq n_0} P(H_{1n} \geq a) = O(e^{-a})$ as $a \to \infty$. □

Lemma 4.4.4. *Under the conditions of Proposition 4.4.1, for each* $0 \leq \eta < 1/2$

$$\left\{ \frac{B_n(x, y)}{(x \vee y)^\eta}, (x, y) \in [0, 1]^2 \right\}_{n \geq 1}$$

is relatively compact.

Proof. By the definition of R_{1n} and R_{2n}, one has

$$\begin{aligned} R_{1n}(x,y) &= \frac{1}{2}k^{1/5}\Lambda_n([x-k^{-1/5}, x+k^{-1/5}]\times[0,\infty]) \\ &= \frac{1}{2}k^{1/5}\left(\frac{[k(x+k^{-1/5})]}{k} - \frac{[k(x-k^{-1/5})]}{k}\right) \\ &\le 1 + 1/k^{4/5} \le 2 \quad \text{if } k \ge 1. \end{aligned}$$

Also $R_{2n}(x,y) \le 2$ for $k \ge 1$. Hence it is sufficient to prove

$$\begin{aligned} &\{W_{R_n}(x,y)/(x\vee y)^\eta, x,y\in[0,1]\}_{n\ge 1}, \\ &\{W_{1n}(x)/x^\eta, x\in[0,1]\}_{n\ge 1}, \\ &\{W_{2n}(y)/y^\eta, y\in[0,1]\}_{n\ge 1} \end{aligned}$$

are relatively compact. Here we only show the proof of the first one. The proofs of the others are similar.

Setting $0/0 = 0$, by Lemma 4.4.1 it suffices to prove that for each positive ε, there exist a $\delta\,(0<\delta<1)$ and $n_0 \in \mathbb{N}$ (n_0 may depend on δ) such that

$$P\left(\sup_{\substack{x,y,u,v\in[0,1]\\ |x-u|\le\delta, |y-v|\le\delta}} \left|\frac{W_{\Lambda_n}([0,x]\times[0,y])}{(x\vee y)^\eta} - \frac{W_{\Lambda_n}([0,u]\times[0,v])}{(u\vee v)^\eta}\right| > \varepsilon\right) \le \varepsilon, \quad n \ge n_0. \tag{4.4.8}$$

We partition the square $[0,1]\times[0,1]$ into m^2 ($m\in\mathbb{N}$) small squares, say $[0,1]\times[0,1] = \bigcup_{i=1}^m \bigcup_{j=1}^m \Delta_{ij}$, with $\Delta_{ij} := \{(x,y) : i\delta \le x \le (i{+}1)\delta, j\delta \le y \le (j{+}1)\delta\}$, $\delta := 1/m$ and $i,j = 0,1,...,m-1$. In order to prove (4.4.8), it suffices to prove that for each positive ε, there exist a $\delta\,(0<\delta<1)$ and $n_0 = n_0(\delta)\in\mathbb{N}$ such that

$$\sum_{i=0}^{m-1}\sum_{j=0}^{m-1} P\left(\sup_{\Delta_{ij}}\left|\frac{W_{\Lambda_n}([0,x]\times[0,y])}{(x\vee y)^\eta} - \frac{W_{\Lambda_n}([0,i\delta]\times[0,j\delta])}{\delta^\eta(i\vee j)^\eta}\right| > \varepsilon\right) \le \varepsilon, \quad n\ge n_0. \tag{4.4.9}$$

We consider the case $i\vee j \ge 1$ and the case $i=j=0$ separately. Let's first look at the case $i \vee j \ge 1$. Assume $i > j$. Let $S(x,y) := [0,x]\times[0,y]$. Note that for $(x,y)\in\Delta_{ij}$

$$\begin{aligned} &\left|\frac{W_{\Lambda_n}([0,x]\times[0,y])}{(x\vee y)^\eta} - \frac{W_{\Lambda_n}([0,i\delta]\times[0,j\delta])}{\delta^\eta(i\vee j)^\eta}\right| \\ &= \left|\frac{W_{\Lambda_n}(S(x,y))}{x^\eta} - \frac{W_{\Lambda_n}(S(i\delta,j\delta))}{(i\delta)^\eta}\right| \\ &= \frac{|(i\delta)^\eta W_{\Lambda_n}(S(i\delta,j\delta)) + (i\delta)^\eta W_{\Lambda_n}(S(x,y)\backslash S(i\delta,j\delta)) - x^\eta W_{\Lambda_n}(S(i\delta,j\delta))|}{x^\eta(i\delta)^\eta} \\ &\le \frac{|(i\delta)^\eta W_{\Lambda_n}(S(x,y)\backslash S(i\delta,j\delta)) - (x^\eta-(i\delta)^\eta)W_{\Lambda_n}(S(i\delta,j\delta))|}{(i\delta)^{2\eta}} \end{aligned}$$

(since $x \geq i\delta \geq y$). Hence

$$P\left(\sup_{\Delta_{ij}}\left|\frac{W_{\Lambda_n}([0,x]\times[0,y])}{(x\vee y)^\eta}-\frac{W_{\Lambda_n}([0,i\delta]\times[0,j\delta])}{\delta^\eta(i\vee j)^\eta}\right|>\varepsilon\right)$$
$$\leq P\Big(\sup_{\Delta_{ij}}|\frac{W_{\Lambda_n}(S(x,y)\backslash S(i\delta,j\delta))}{(i\delta)^\eta}|>\frac{\varepsilon}{2}\Big)+P\Big(\sup_{\Delta_{ij}}|\frac{x^\eta-(i\delta)^\eta}{(i\delta)^{2\eta}}W_{\Lambda_n}(S(i\delta,j\delta))|>\frac{\varepsilon}{2}\Big)$$
$$\leq 4P\Big(|\frac{W_{\Lambda_n}(S((i+1)\delta,(j+1)\delta)\backslash S(i\delta,j\delta))}{(i\delta)^\eta}|>\frac{\varepsilon}{4}\Big)$$
$$+P\Big(|\frac{(1+1/i)^\eta-1}{(i\delta)^\eta}W_{\Lambda_n}(S(i\delta,j\delta))|>\frac{\varepsilon}{2}\Big).$$

Since $\Lambda_n(S((i+1)\delta,(j+1)\delta)\backslash S(i\delta,j\delta))\leq 2\delta+4/k$ for all $i\vee j\geq 1$, there exist $n_*=n_*(\delta)\in\mathbb{N}$ such that $k_*=k(n_*)\geq 1/\delta$ and hence

$$\Lambda_n(S((i+1)\delta,(j+1)\delta)\backslash S(i\delta,j\delta))\leq 6\delta,\quad n\geq n_*.$$

uniformly in $i\vee j\geq 1$. It follows that $(i\delta)^{-\eta}W_{\Lambda_n}(S((i+1)\delta,(j+1)\delta)\backslash S(i\delta,j\delta))$ has a normal distribution with mean zero and variance $\sigma_n^2(i,j)$ satisfying $\sigma_n^2(i,j)\leq 6\delta^{1-2\eta}$ for all $i>j, i\geq 1$, and $n\geq n_*$. Hence for all $\varepsilon>0$

$$\sup_{n\geq n_*}\sup_{i>j,i\geq 1}P\big(|(i\delta)^{-\eta}W_{\Lambda_n}(S((i+1)\delta,(j+1)\delta)\backslash S(i\delta,j\delta))|>\varepsilon/4\big)=O(e^{-\delta^{\eta-1/2}})$$

as $\delta\to 0$. On the other hand, note that $\frac{(1+1/i)^\eta-1}{(i\delta)^\eta}W_{\Lambda_n}(S(i\delta,j\delta))$ has a normal distribution with mean zero and variance $\tilde\sigma_n^2(i,j)$ satisfying $\tilde\sigma_n^2(i,j)\leq(i\delta)^{1-2\eta}((1+1/i)^\eta-1)^2\leq 4\delta^{1-2\eta}$. So

$$\sup_{n\geq n_*}\sup_{i>j,i\geq 1}P\Big(|\frac{(1+1/i)^\eta-1}{(i\delta)^\eta}W_{\Lambda_n}(S(i\delta,j\delta))|>\varepsilon/2\Big)=O(e^{-\delta^{\eta-1/2}})$$

as $\delta\to 0$.

In case of $j>i, j\geq 1$ and case of $i=j\geq 1$, we can get similar results as above. Hence

$$\sup_{n\geq n_*}\sum_{i\vee j\geq 1}^{m-1}P\left(\sup_{\Delta_{ij}}\left|\frac{W_{\Lambda_n}([0,x]\times[0,y])}{(x\vee y)^\eta}-\frac{W_{\Lambda_n}([0,i\delta]\times[0,j\delta])}{\delta^\eta(i\vee j)^\eta}\right|>\varepsilon\right)$$
$$=O(\delta^{-2}e^{-\delta^{\eta-1/2}}) \tag{4.4.10}$$

as $\delta\to 0$.

Now let us look at the case $i=j=0$. By Lemma 4.3.2 (in fact we can replace R by Λ_n in that lemma), one has

$$\sup_{n\geq 1}P\left(\sup_{x\vee y\leq\delta}\left|\frac{W_{\Lambda_n}([0,x]\times[0,y])}{(x\vee y)^\eta}\right|>\varepsilon\right)=O(e^{-\delta^{\eta-1/2}}) \tag{4.4.11}$$

as $\delta\to 0$.

Since (4.4.10) and (4.4.11) imply (4.4.9), the result follows. □

Lemma 4.4.5. *Under the conditions of Proposition 4.4.1, for each* $0 \le \eta < 1$

$$\left\{ \frac{A_n(x,y)}{(x \vee y)^\eta}, (x,y) \in [0,1]^2 \right\}_{n \ge 1}$$

is relatively compact.

Proof. The proof is similar to that of Lemma 4.4.4. We use the same notation for Δ_{ij} and S. We only need to check that for each positive ε, there exist a δ $(0 < \delta < 1)$ and $n_0 = n_0(\delta) \in \mathbb{N}$ such that

$$\sum_{i=0}^{m-1} \sum_{j=0}^{m-1} P\left(\sup_{\Delta_{ij}} \left| \frac{A_n(x,y)}{(x \vee y)^\eta} - \frac{A_n(i\delta, j\delta)}{\delta^\eta (i \vee j)^\eta} \right| > \varepsilon \right) \le \varepsilon, \quad n \ge n_0. \tag{4.4.12}$$

We consider the case $i \vee j \ge 1$ and the case $i = j = 0$ separately. Let us first look at the case $i \vee j \ge 1$. In case of $i > j$, $i \ge 1$, note that for $(x,y) \in \Delta_{ij}$

$$\begin{aligned}
&|A_n(x,y)/(x \vee y)^\eta - A_n(i\delta, j\delta)/((i\delta) \vee (j\delta))^\eta| \\
&= \Big|(x^{1-\eta} - (i\delta)^{1-\eta})(W_{\Lambda_n}(C_{\pi/2}) - Z_n(\pi/2)) \\
&\quad - (x^{1-\eta} - (i\delta)^{1-\eta}) \int_{\arctan y/x}^{\pi/4} \frac{1}{\cos^2\theta} (W_{\Lambda_n}(C_\theta) + Z_n(\theta))\, d\theta \\
&\quad + (i\delta)^{1-\eta} \int_{\arctan j/i}^{\arctan y/x} \frac{1}{\cos^2\theta} (W_{\Lambda_n}(C_\theta) + Z_n(\theta))\, d\theta \Big| \\
&\le (i\delta)^{1-\eta}((1+1/i)^\eta - 1)(1+\pi/2)H_n + (i\delta)^{1-\eta} \left(\arctan \frac{j+1}{i} - \arctan \frac{j}{i} \right) 2\, H_n
\end{aligned}$$

where H_n is defined in Lemma 4.4.3. Since $(i\delta)^{1-\eta}((1+1/i)^\eta - 1) = O(\delta^{1-\eta})$ and $(i\delta)^{1-\eta}(\arctan \frac{j+1}{i} - \arctan \frac{j}{i}) = O(\delta^{1-\eta})$ as $\delta \to 0$ and uniformly in i, j $(i > j, i \ge 1)$, then by Lemma 4.4.3 there exists $n_* = n_*(\delta) \in \mathbb{N}$ such that

$$\begin{aligned}
&\sup_{n \ge n_*} \sup_{i>j,\, i \ge 1} P\big(|A_n(x,y)/(x \vee y)^\eta - A_n(i\delta, j\delta)/((i\delta) \vee (j\delta))^\eta| > \varepsilon/2\big) \\
&= O(e^{-\delta^{(\eta-1)/2}})
\end{aligned} \tag{4.4.13}$$

as $\delta \to 0$.

In case of $j > i, j \ge 1$ and case of $i = j \ge 1$ we can get a similar result as (4.4.13). Hence there exists $n_{01} = n_{01}(\delta) \in \mathbb{N}$ such that

$$\begin{aligned}
&\sup_{n \ge n_{01}} \sum_{i \vee j \ge 1}^{m} P\big(|A_n(x,y)/(x \vee y)^\eta - A_n(i\delta, j\delta)/((i\delta) \vee (j\delta))^\eta| > \varepsilon\big) \\
&= O(\delta^{-2} e^{-\delta^{(\eta-1)/2}})
\end{aligned} \tag{4.4.14}$$

as $\delta \to 0$.

Now let's consider the case $i = j = 0$ and w.l.o.g. assume $y \geq x$. Then for $0 \leq x \leq y \leq \delta$

$$\begin{aligned} &|A_n(x,y)/(x \vee y)^\eta| \\ &= \left| xy^{-\eta} W_{\Lambda_n}(C_{\pi/2}) + Z_n(\pi/2)) + y^{1-\eta} \int_{\pi/4}^{\arctan y/x} \sin^{-2}\theta (W_{\Lambda_n}(C_\theta) + Z(\theta)) d\theta \right| \\ &\leq \delta^{1-\eta}(1+\pi/2) H_n. \end{aligned}$$

Hence there exists $n_{02} = n_{02}(\delta) \in \mathbb{N}$ such that

$$\sup_{n \geq n_{02}} P\big(\sup_{x \vee y \leq \delta} |A_n(x,y)/(x \vee y)^\eta| > \varepsilon \big) = O(e^{-\delta^{(\eta-1)/2}}) \tag{4.4.15}$$

as $\delta \to 0$.

Now (4.4.14) and (4.4.15) imply (4.4.12). □

Proof of Proposition 4.4.1. By Lemmas 4.4.4 and 4.4.5,

$$\left\{ \frac{A_n(x,y) + B_n(x,y)}{(x \vee y)^\eta},\ (x,y) \in [0,1]^2 \right\}_{n \geq 1} \tag{4.4.16}$$

is relatively compact. It is easy to check that the finite-dimensional distributions of our estimated processes in (4.4.16) converge to those of the limiting process, which completes the proof. □

Proof of Corollary 4.4.1. After applying a Skorohod construction to the weak convergence statement of Proposition 4.4.1, the proof is similar to that of Theorem 4.2.3. □

Proof of Corollary 4.4.2. Proposition 4.4.1 implies the weak convergence of the distribution function of the left hand side of (4.4.6) to the distribution function of the right hand side of (4.4.6). This property carries over to the inverse functions Q_{Λ_n} and Q_Λ. □

Proof of Corollary 4.4.3. From another Skorohod construction we obtain an a.s. version of the statement of Theorem 4.2.2; without changing the notation we now work with this construction. Since for $0 < x, y \leq 1$

$$\begin{aligned} \Lambda([0,x] \times [0,y]) &= x + y - l(x,y), \\ \Lambda_n([0,x] \times [0,y]) &= \lceil kx \rceil / k + \lceil ky \rceil / k - \hat{l}_2(x,y) - \delta_n(x,y)/k \end{aligned}$$

($\delta_n(x,y)$ takes values in $\{0,1,2\}$), it follows that for each $\varepsilon > 0$

$$\sup_{0 < x,y \leq 1} k^{1/2-\varepsilon} \Big| \Lambda_n([0,x] \times [0,y]) - \Lambda([0,x] \times [0,y]) \Big| \to 0 \quad \text{a.s.} \tag{4.4.17}$$

as $n \to \infty$.

We now show that (4.4.2), (4.4.3), (4.4.4), (4.4.5) hold a.s. We already saw, below (4.4.7), that (4.4.2) holds a.s. and the a.s. version of (4.4.3) follows immediately from (4.4.17).

By (4.4.17) and (4.4.2), it is easily follows that

$$\sup_{E\in\mathcal{E}} k^{1/2-\varepsilon}\Big|\Lambda_n(E)-\Lambda(E)\Big| \to 0 \quad \text{a.s.} \tag{4.4.18}$$

as $n\to\infty$, where $\mathcal{E} := \{E|\ E=[x_1,x_2]\times[y_1,y_2], 0<x_1\le x_2\le 2,\ 0<y_1\le y_2\le 2\}$. Let $E_n(x)=[x-k^{-1/6},x+k^{-1/6}]\times[1-k^{-1/6},1+k^{-1/6}]$. Then

$$\begin{aligned}
&\sup_{0<x\le 1}|\lambda_n(x,1)-\lambda(x,1)|\\
&=\sup_{0<x\le 1}|\frac{1}{4}k^{1/3}\Lambda_n(E_n(x))-\frac{1}{4}k^{1/3}\Lambda(E_n(x))+\frac{1}{4}k^{1/3}\Lambda(E_n(x))-\lambda(x,1)|\\
&\le \sup_{0<x\le 1}\frac{1}{4}k^{1/3}|\Lambda_n(E_n(x))-\Lambda(E_n(x))|+\sup_{0<x\le 1}|\frac{1}{4}k^{1/3}\Lambda(E_n(x))-\lambda(x,1)|\\
&\to 0 \quad \text{a.s. as } n\to\infty,
\end{aligned}$$

as $n\to\infty$, by (4.4.18) and $\lambda(0,1)=0$. The proofs of $\sup_{0<y\le 1}|\lambda_n(1,y)-\lambda(1,y)|\to 0$ a.s. and $\sup_{0<x,y\le 1}|R_{jn}(x,y)-R_j(x,y)|\to 0$, $j=1,2$, a.s. are similar. Hence (4.4.4) and (4.4.5) hold a.s.

According to Corollary 4.4.2 we have

$$Q_{\Lambda_n}(1-\alpha)\to Q_\Lambda(1-\alpha) \quad \text{a.s.}$$

as $n\to\infty$, hence also in probability. □

4.5 Simulation study and real data application

In this section we present a small simulation study, making use of the results of section 4.4. We will consider one distribution satisfying the domain of attraction condition and one that fails to satisfy it. At the end of the section, we will apply our procedure to financial data. *Throughout* we take $\beta=2$ in the test statistic of (4.1.12).

Consider the bivariate Cauchy distribution restricted to the first quadrant, with density

$$f(x,y)=\frac{2}{\pi(1+x^2+y^2)^{\frac{3}{2}}},\quad x,y>0.$$

It readily follows that

$$\Lambda([0,x]\times[0,y])=x+y-\sqrt{x^2+y^2},\quad \lambda(x,y)=\frac{xy}{(x^2+y^2)^{3/2}},\quad x,y>0.$$

This distribution satisfies the conditions of Theorem 4.2.3; in particular (4.2.5) holds with $\alpha = 2$ (see Einmahl *et al.* (2001), pp. 1409-1410). First we present in Table 4.1 the quantiles of the limiting random variable

$$\iint_{0<x,y\le 1} \frac{(A(x,y)+B(x,y))^2}{(x\vee y)^2} dxdy,$$

using the approximation of section 4.4. We used 100,000 replications. With high probability these quantiles are accurate up to 0.01.

p	0.25	0.50	0.75	0.90	0.95
$Q(p)$	0.10	0.14	0.22	0.34	0.44

Table 4.1: Quantiles of the limiting r.v. for $\beta = 2$ for the Cauchy distribution.

Now for sample size $n = 2000$, we simulate 1000 times the test statistic

$$k\iint_{0<x,y\le 1} \frac{(\hat{l}_1(x,y)-\hat{l}_2(x,y))^2}{(x\vee y)^2} dxdy,$$

for various values of k. Using the 0.95-th quantile above, we find the simulated type-I error probabilities; see Table 4.2. In the ideal situation the number of rejections is a binomial r.v. with parameters 1000 and 0.05. So the numbers in the table are remarkably close to 0.05. Only for $k = 400$, the bias seems to set in. In addition, in Figure 4.1 we see, for various k, on the left for one sample of size $n = 2000$ the values of the test statistic and on the right the median and 0.95-th quantile for the test statistic based on 800 samples. Note that the behavior of the test statistic fluctuates with k, but that for all k in the figure the value is far below 0.44, the 0.95-th quantile of the limiting random variable.

k	20	40	60	80	100	125	150	175	200	250	300	400
$\hat{\alpha}$	.049	.048	.055	.039	.038	.049	.046	.055	.049	.060	.055	.082

Table 4.2: Simulated type-I error for the Cauchy distribution: $n = 2000$ and $\alpha = 0.05$.

Next we consider a distribution with uniform-$(0,1)$ marginals (a copula), which does not satisfy the bivariate domain of attraction condition. Since both marginals are uniform, they are in the univariate domain of attraction of the reverse Weibull law. So it is the dependence structure that causes the failure. The distribution is an adaptation of a distribution in Schlather (2001): take a density of 3/2 on the following rectangles: $[2^{-(2m+1)}, 2^{-(2m)}] \times [2^{-(2r+1)}, 2^{-(2r)}]$, for $m = 0,1,2,\dots$ and $r = 0,1,2,\dots$; in this way a probability mass of 2/3 is assigned. The remaining 1/3 is assigned by taking the uniform distribution on the line segments from $(2^{-(2m+2)}, 2^{-(2m+2)})$ to $(2^{-(2m+1)}, 2^{-(2m+1)})$, $m =$

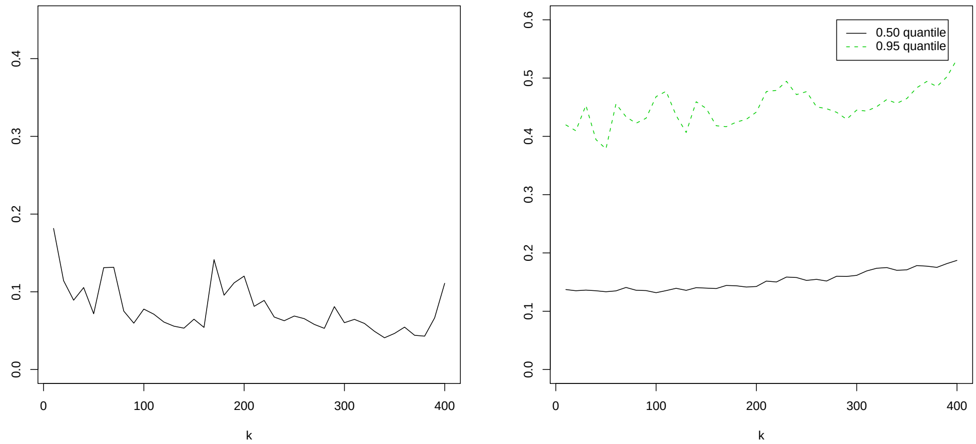

Figure 4.1: Cauchy distribution: test statistic (left) and quantiles of the test statistic (right).

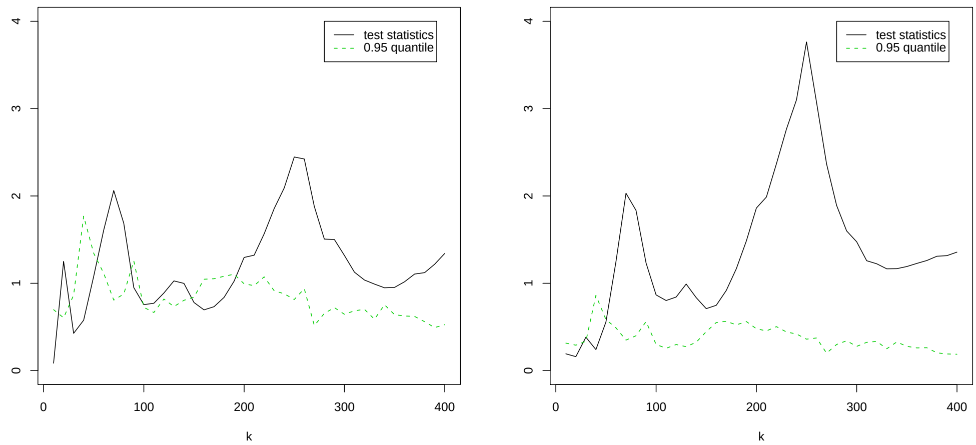

Figure 4.2: Alternative distribution: test statistics and 0.95-th quantiles of 2 samples.

$0, 1, 2, \ldots$, such that the mass of the m-th segment is equal to $2^{-(2m+2)}$. In Figure 4.2, we see for varying k the test statistics and simulated 0.95-th quantiles of two samples of size $n = 2000$ from this distribution. Again the test statistics fluctuate with k, but from a certain k on (and for most values of k), the null hypothesis is clearly rejected.

Finally, we apply the test to real data, similarly as we just did for the simulated data sets in Figure 4.2. The data are 3283 daily logarithmic equity returns over the period 1991-2003 for two Dutch banks, ING and ABN AMRO bank. The bivariate, heavy-tailed data are shown in Figure 4.3 on the left; on the right we see again the test statistic and 0.95-th quantile. Since the

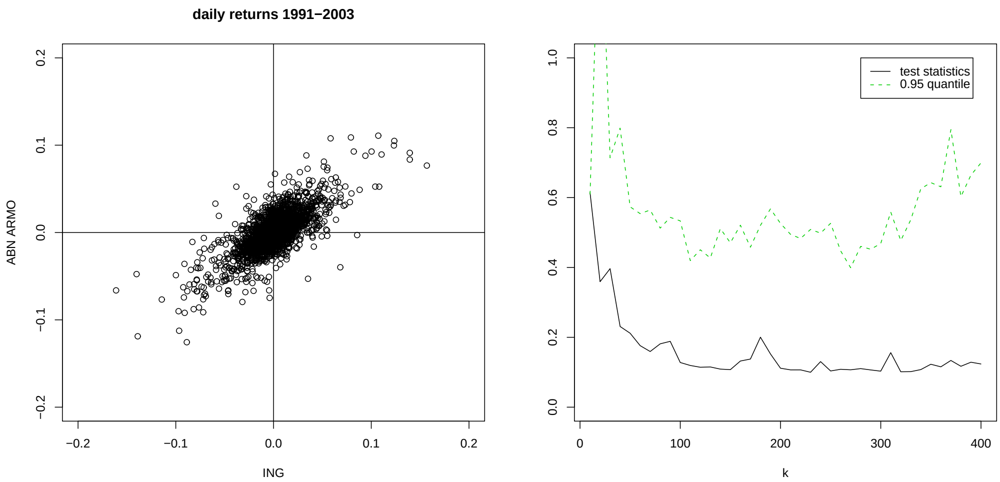

Figure 4.3: Daily equity returns of two Dutch banks (left) and test statistics and 0.95-th quantiles (right).

test statistic is everywhere clearly below the quantile, we cannot reject the null hypothesis. This is a satisfactory result, because it allows us to analyze these data further, using statistical theory of extremes.

Chapter 5

Comparison of Estimators in Multivariate EVT

co-authors: M. Isabel Barão and Laurens de Haan

Abstract. Several methods to generate samples from distributions in the domain of attraction of multivariate extreme value distributions are proposed and used for comparison of estimators by simulation.

5.1 Introduction

One of the main purposes of multivariate extreme value theory (EVT) is to estimate the probability of an extreme set when i.i.d. observations are available from a distribution in the domain of attraction of an extreme value distribution. In order to do so one has to estimate the dependence structure of the limiting distribution . In fact, as in the one-dimensional case, one does the extrapolation via the associated generalized Pareto distribution (GPD) rather than the extreme value distribution itself: the GPD is $1 + \log G(\boldsymbol{x})$ for all $\boldsymbol{x}$ with $0 < G(\boldsymbol{x}) < 1$ with G the extreme value distribution. We shall concentrate on the dependence structure abstracted from the form of the marginal distributions: all marginal distributions will have the form $1 - 1/x$ ($x \geq 1$) which is the one-dimensional GPD with extreme value index 1.

We are going to compare two estimators for the dependence structure:

(i) the maximum likelihood estimator derived in a specific parametric method called the logistic method which is generated by one parameter $\alpha \in (0, 1]$ (cf. Tawn (1988)), but used also for distributions outside this class.

(ii) the nonparametric estimator proposed and studied by Huang and Mason (see Huang (1992)).

How do we compare the two estimation procedures? For this we need to be more concrete about the model.

We restrict ourselves to the two-dimensional space and generate i.i.d. random vectors

$$(X_1, Y_1),\ (X_2, Y_2),\ \dots,\ (X_n, Y_n)$$

with distribution function F which is in the domain of attraction of some extreme value distribution G, and G being of Fréchet marginal distributions with index 1, i.e. $G(x, \infty) = G(\infty, x) = \exp(-1/x)$ for $x > 0$. This distribution function can be written

$$G(x, y) = \exp\left(-l(x^{-1}, y^{-1})\right), \qquad x, y > 0,$$

where l is a homogeneous function of order 1, i.e. for $a, x, y > 0$

$$l(ax, ay) = a\, l(x, y).$$

Moreover there is a measure Λ such that

$$l(x, y) = \Lambda\left(\{(s, t) \in \mathbb{R}_+^2 |\ 0 \le s \le x \text{ or } 0 \le t \le y\}\right)$$

and for each Borel set $B \subset \mathbb{R}_+^2$ with $\Lambda(\partial B) = 0$ and for each $a > 0$

$$\Lambda(aB) = a\Lambda(B).$$

We introduce the level sets of the function l:

$$Q_c := \{(s, t) \in \mathbb{R}_+^2 |\ l(s, t) = c\}$$

where $c > 0$. In fact this "Q-curve" is the graph of a function q, say, which is characterized by

(1) $q(0) = c,\ q(c) = 0$
(2) q is concave.

Conversely any function q with the indicated properties gives rise, via the function l, to an extreme value distribution (page 40 in Huang (1992)).

Next we introduce the spectral measure. In order to be able to connect the two estimation procedures we shall provide two forms of the spectral measure.

(i) Each extreme value distribution gives rise to an unique probability distribution P concentrated on [0, 1] with mean 1/2 as follows (cf. Pickands (1981)):

$$G(x, y) = \exp\Big(-2 \int_0^1 \big(\frac{w}{x} \vee \frac{1-w}{y}\big)\, H(d\,w)\Big), \tag{5.1.1}$$

where H is the distribution function of the probability distribution P, and $\vee$ means maximum.

(ii) Each extreme value distribution gives rise to an unique finite measure on the interval $[0, \pi/2]$ with distribution function Φ such that

$$\int_0^{\pi/2} (1 \wedge \tan\theta)\ \Phi(d\,\theta) = \int_0^{\pi/2} (1 \wedge \cot\theta)\, \Phi(d\,\theta) = 1.$$

The functions G and Φ are related in the following way:

$$G(x,y) = \exp\Big(- \int_0^{\pi/2} (\frac{1 \wedge \tan\theta}{x} \vee \frac{1 \wedge \cot\theta}{y})\ \Phi(d\,\theta)\Big), \tag{5.1.2}$$

where $\wedge$ means minimum.

The connection between the two forms of the spectral measure, namely those with distribution functions H and Φ (there are more!) is given in Lemma 5.2.1 below.

What is our comparison criterion for the two estimation procedures? There will be two:

(i) Comparing the estimators for the Q-curve with the real Q-curve.

(ii) Comparing the estimators for the spectral measure with the real spectral measure (For this purpose we choose the spectral measure with distribution function Φ).

The comparison will be done by simulation. Generating samples from a multidimensional distribution is not completely straightforward. We developed two methods to generate distributions in the domain of attraction of an extreme value distribution in Lemma 5.2.2 and 5.2.3.

The two estimation methods give markedly different results for distributions that are not very smooth and symmetric.

5.2 The estimators; some useful Lemmas

First we describe the estimators for the function l. Recall that

$$l(x,y) = 2\int_0^1 ((xw) \vee (y(1-w)))\ H(d\,w) \tag{5.2.1}$$

$$= \int_0^{\pi/2} ((x(1 \wedge \tan\theta)) \vee (y(1 \wedge \cot\theta)))\ \Phi(d\,\theta). \tag{5.2.2}$$

Maximum likelihood (ML) methods in two-dimensional extreme value theory have been well developed recently. The logistic model has appeared in many articles: see, for example, Tawn (1988), Coles and Tawn(1994). Assume (X, Y) has the distribution function F satisfying

$$F(x, y) = \exp\left(-l(x^{-1}, y^{-1})\right) \quad \text{for } x > u,\, y > u$$

for large u, where function l has the following form

$$l(x, y) = (x^{1/\alpha} + y^{1/\alpha})^{\alpha}, \qquad \alpha \in (0,\, 1].$$

Then for each $x, y > 0$ and $n = 1, 2, ...$

$$F^n(nx, ny) = G(x, y) = \exp\left(-l(x^{-1}, y^{-1})\right).$$

In particular both marginal distribution functions are $\exp(-1/x),\ x > 0$. Hence F is in the domain of attraction of G. The index α reflects the degree of dependence. $\alpha \to 0$ and $\alpha = 1$ correspond to perfect dependence and exact independence respectively. Coles, Heffernan and Tawn (1999) derived the corresponding distribution function $H(w)\,(= H(w, \alpha))$

$$H(w, \alpha) = \frac{1}{2}\left((w^{(1-\alpha)/\alpha} - (1-w)^{(1-\alpha)/\alpha}) \cdot (w^{1/\alpha} + (1-w)^{1/\alpha})^{\alpha-1} + 1\right).$$

In ML-estimation, the contribution to the likelihood function of each observation is defined by

$$L(X_i, Y_i;\, \alpha) = \begin{cases} F(u, u) & \text{for } X_i \le u,\, Y_i \le u, \\ (\partial F(x, u)/\partial x)_{x=X_i} & \text{for } X_i > u,\, Y_i \le u, \\ (\partial F(u, y)/\partial y)_{y=Y_i} & \text{for } X_i \le u,\, Y_i > u, \\ \left(\partial^2 F(x, y)/\partial x \partial y\right)_{x=X_i,\, y=Y_i} & \text{for } X_i > u,\, Y_i > u\ . \end{cases}$$

In our simulation, we choose $u = (X_{n-k,n} + Y_{n-k,n})/2$, where $X_{n-k,n}$ and $Y_{n-k,n}$ are the k-th largest order statistics of $X_1, ..., X_n$ and $Y_1, ..., Y_n$ respectively. By maximizing the likelihood function

$$L((X_i, Y_i)_{1 \le i \le n};\, \alpha) := \prod_{i=1}^{n} L(X_i, Y_i;\, \alpha),$$

the ML-estimator $\hat{\alpha}$ of α, is obtained. Then the ML-estimator of $l(x, y)$ is

$$\hat{l}_{ML}(x, y) = (x^{1/\hat{\alpha}} + y^{1/\hat{\alpha}})^{\hat{\alpha}} \tag{5.2.3}$$

and the ML-estimator of $\Phi(\theta)$ is

$$\hat{\Phi}_{ML}(\theta) = 2\int_0^{1/(1+\cot\theta)} \left(w \vee (1-w)\right) H(d\,w,\, \hat{\alpha}) \tag{5.2.4}$$

by Lemma 5.2.1 (below).

Nonparametric estimation was proposed and studied by Huang and Mason (cf. Huang (1992) and Drees and Huang (1998)). The nonparametric estimator of $l(x,y)$ is

$$\hat{l}_{NP}(x,y) := \frac{1}{k}\sum_{i=1}^{n} I_{\{R_i^X > n-kx \text{ or } R_i^Y > n-ky\}} \tag{5.2.5}$$

and the nonparametric estimator of $\Phi(\theta)$ is

$$\hat{\Phi}_{NP}(\theta) := \frac{1}{k}\sum_{i=1}^{n} I_{\{R_i^X \vee R_i^Y \geq n+1-k \text{ and } n+1-R_i^Y \leq (n+1-R_i^X)\tan\theta\}} \tag{5.2.6}$$

where (X,Y), (X_1,Y_1), ..., (X_n,Y_n) are i.i.d. *r.v.*, and (X,Y) is in the domain of attraction of G, $R_i^X = \text{rank}(X_i)$ among $(X_1,X_2,...,X_n)$ and $R_i^Y = \text{rank}(Y_i)$ among $(Y_1,Y_2,...,Y_n)$, and k satisfies $k\to\infty$, $k/n\to 0$ (Einmahl, de Haan and Piterbarg (2001)). The only requirement is that the underlying distribution function is in some domain of attraction.

Second we introduce some useful Lemmas. Lemma 5.2.1 gives the relation between H and Φ. Lemma 5.2.2 and Lemma 5.2.3 can be used to generate distributions which are in the domain of attraction of a multivariate extreme distribution G.

Lemma 5.2.1. *Suppose (5.2.1) and (5.2.2) hold, then*

$$\Phi(\theta) = 2\int_0^{1/(1+\cot\theta)} (w \vee (1-w))\, H(dw). \tag{5.2.7}$$

Proof. Let $w = \frac{1}{1+\cot\theta}$, then by (5.2.1)

$$\begin{aligned}
l(x,y) &= 2\int_0^{\frac{\pi}{2}} \Big(\frac{x}{1+\cot\theta} \vee \big(y(1-\frac{1}{1+\cot\theta})\big)\Big)\, H\big(d(\frac{1}{1+\cot\theta})\big)\\
&= 2\int_0^{\frac{\pi}{2}} ((x\sin\theta)\vee(y\cos\theta))\,\frac{1}{\sin\theta+\cos\theta}\, H\big(d(\frac{1}{1+\cot\theta})\big)\\
&= 2\int_0^{\frac{\pi}{2}} \big((x(1\wedge\tan\theta))\vee(y(1\wedge\cot\theta))\big)\,\frac{\sin\theta\vee\cos\theta}{\sin\theta+\cos\theta}\, H\big(d(\frac{1}{1+\cot\theta})\big)\\
&=: \int_0^{\frac{\pi}{2}} \big((x(1\wedge\tan\theta))\vee(y(1\wedge\cot\theta))\big)\,\Phi(d\theta).
\end{aligned}$$

Then

$$\Phi(\theta) = 2\int_0^{\theta} (\sin u+\cos u)^{-1}(\sin u\vee\cos u)\, H\big(d(\frac{1}{1+\cot u})\big).$$

Let $w = 1/(1+\cot u)$, by $(\sin u+\cos u)^{-1}(\sin u\vee\cos u) = w\vee(1-w)$, then

$$\Phi(\theta) = 2\int_0^{1/(1+\cot\theta)} (w\vee(1-w))\, H(dw).$$

□

Lemma 5.2.2. *Suppose V is a r.v. with distribution function $F_V(v) = 1 - \frac{2}{v}$, $v \geq 2$, W is a r.v. with values in $[0, 1]$ and distribution function $H(w)$, $EW = \frac{1}{2}$, and V, W are independent. Define $X := VW$, $Y := V(1-W)$. Then (X, Y) is in the domain of attraction of the extreme value distribution G and G has form (5.1.1).*

Proof. For each $x \geq 2,\ y \geq 2$,

$$\begin{aligned}
&P\big(X \geq x \text{ or } Y \geq y\big)\\
&= P\big(VW \geq x \text{ or } V(1-W) \geq y\big)\\
&= P\Big(V \geq \big(\frac{x}{W} \wedge \frac{y}{1-W}\big)\Big)\\
&= E_W\Big(\frac{2}{\frac{x}{W} \wedge \frac{y}{1-W}}\Big)\\
&= 2E_W\Big(\frac{W}{x} \vee \frac{1-W}{y}\Big)\\
&= 2\int_0^1 \big(\frac{w}{x} \vee \frac{1-w}{y}\big)\, H(dw).
\end{aligned}$$

Let $F(x,y) := P(X \leq x,\, Y \leq y)$, It's easy to check that for $x, y > 0$

$$\lim_{t\to\infty} t(1 - F(tx,\, ty)) = 2\int_0^1 \big(\frac{w}{x} \vee \frac{1-w}{y}\big)\, H(dw).$$

Hence (X, Y) or F is in the domain of attraction of G. □

Lemma 5.2.3. *Suppose V is a r.v. with distribution function $F_V = 1 - \frac{1}{v}$, $v \geq 1$, A is a r.v. with distribution function F_A satisfying $EA = 1$, $A \geq 0$ a.s. and V, A are independent. Define $X := V$, $Y := AV$, then (X, Y) is in the domain of attraction of an extreme value distribution G, and G has form (5.1.1) with*

$$H(d\,w) = -\frac{1}{2w} F_A\big(d(\frac{1}{w} - 1)\big) \tag{5.2.8}$$

or the form (5.1.2) with

$$\Phi(d\,\theta) = -(1 \vee \cot\theta) F_A(d\,\cot\theta). \tag{5.2.9}$$

Proof. For each $x \geq 1$, $y > 0$

$$
\begin{aligned}
& P(X \geq x \text{ or } Y \geq y) \\
= & P\big(V \geq x \text{ or } V \geq \frac{y}{A}\big) \\
= & P\big(V \geq (x \wedge \frac{y}{A})\big) \\
= & E_{A}\big(\frac{1}{x \wedge (y/A)}\big) \\
= & \int_0^\infty \big(\frac{1}{x} \vee \frac{a}{y}\big)\, dF_{A}(a) \\
= & 2\int_0^\infty \big(\frac{1}{x(1+a)} \vee \frac{a}{(1+a)y}\big)\frac{1+a}{2}\, dF_{A}(a).
\end{aligned}
$$

Let $F(x,y) := P(X \leq x, Y \leq y)$, it's easy to check that for x, $y > 0$

$$\lim_{t\to\infty} t(1 - F(tx,ty)) = 2\int_0^\infty \big(\frac{1}{x(1+a)} \vee \frac{a}{(1+a)y}\big)\frac{1+a}{2}\, dF_{A}(a).$$

Hence, (X, Y) is in the domain of attraction of G, and

$$G(x,y) = \exp\Big(-2\int_0^\infty \big(\frac{1}{x(1+a)} \vee \frac{a}{(1+a)y}\big)\frac{1+a}{2}\, dF_{A}(a)\Big). \qquad (5.2.10)$$

If transforming $\frac{1}{1+a}$ into w, $G(x,y)$ has the form of (5.1.1) with

$$H(d\,w) = -\frac{1}{2w}F_A\big(d(\frac{1}{w} - 1)\big).$$

By Lemma 5.2.1, It's easy to check that $G(x,y)$ has the form of (5.1.2) with

$$\Phi(d\,\theta) = -(1 \vee \cot\theta)F_{A}(d\,\cot\theta).$$

□

Remark 5.2.1. *Multivariate extensions of the constructions of Lemma 5.2.2 and Lemma 5.2.3 are as follows. Consider d-dimensional case with $d \in \mathbb{N}$ and $d > 2$. In Lemma 5.2.2, let V be as before and take a random vector $(W_1, W_2, ..., W_d)$, independent of V, with distribution function H satisfying*

$$\int_0^1 \cdots \int_0^1 w_i\, H(dw_1, dw_2, ..., dw_d) = 1/d, \qquad i = 1, 2..., d.$$

Take $(X_1, X_2, ..., X_d) := V \cdot (W_1, W_2, ..., W_d)$, then

$$P\big(\bigcup_{i=1}^{d}\{X_i \geq x_i\}\big) = d\int_{w_1+w_2+...+w_d \leq 1} \big(\bigvee_{i=1}^{d} \frac{w_i}{x_i}\big) H(dw_1, dw_2, ..., dw_d).$$

In Lemma 5.2.3, take $(X_1, X_2, ..., X_d) := V \cdot (1, A_1, ..., A_{d-1})$*, where* V *is as before,* $(A_1, A_2, ..., A_{d-1})$ *is independent of* V *and* $A_i \geq 0$*,* $EA_i = 1$ *for* $i = 1, 2, ..., d-1$*. Then*

$$P\big(\bigcup_{i=1}^{d}\{X_i \geq x_i\}\big) = E\Big(\frac{1}{x_1} \bigvee \big(\bigvee_{i=2}^{d} \frac{A_{i-1}}{x_i}\big)\Big).$$

Remark 5.2.2. *Lemma 5.2.2 and Lemma 5.2.3 allow for the construction of non-symmetric models. In particular, if A from Lemma 5.2.3 has a two-points distribution, i.e. there exist* $0 < a_1 < 1 < a_2 < \infty$ *and* $0 < p_1,\ p_2 < 1$, $p_1 + p_2 = 1$ *such that*

$$P(A = a_i) = p_i,\ i = 1, 2 \quad \text{and} \quad EA = 1$$

then

$$l(x, y) = p_1(x \vee (a_1 y)) + p_2(x \vee (a_2 y))$$

and the set $\{(x, y) \in \mathbb{R}_+^2 \mid l(x, y) = 1\}$ *are the union of three subsets*

$$\{(x,\ y) \in \mathbb{R}_+^2 \mid 0 \leq x \leq a_1,\ y = 1\}$$

$$\{(x,\ y) \in \mathbb{R}_+^2 \mid (y-1)/(x-a_1) = (1-1/a_2)/(a_1-1),\ a_1 < x < 1\}$$

$$\{(x,\ y) \in \mathbb{R}_+^2 \mid 0 \leq y \leq 1/a_2,\ x = 1\}.$$

Moreover, if A has a continuous distribution but the main probability mass is on two small intervals around the two points a_1, a_2 *such that the density of A on some interval* $[b_1,\ b_2]$ $([b_1, b_2] \subset [a_1, a_2])$ *is very small, then the curve of* $\{(x,\ y) \in \mathbb{R}_+^2 \mid l(x, y) = 1\}$ *is a smooth curve but the basic shape is similar as in the case two-points distribution.*

Example 5.2.1. *If A has a standard exponent distribution, i.e.* $F_A(a) = 1 - e^{-a}$*,* $a > 0$*, then* $EA = 1$ *and*

$$H(dw) = \frac{1}{2} w^{-3} e^{1-\frac{1}{w}} dw,$$
$$\Phi(d\theta) = (\sin\theta)^{-3}(\sin\theta \vee \cos\theta) e^{-\cot\theta} d\theta.$$

Example 5.2.2. *If A has a Gamma(*α*) distribution, i.e.* $F_A(a) = e^{-(\Gamma(-\frac{1}{\alpha}+1)a)^{-\alpha}}$*,* $a > 0$*. Let* $\alpha = 2$ *(note that* $\Gamma(1/2) = \sqrt{\pi}$*), then* $EA = 1$ *and*

$$H(dw) = \frac{1}{\pi}(1-w)^{-3} e^{-(\sqrt{\pi}(\frac{1}{w}-1))^{-2}} dw,$$
$$\Phi(d\theta) = (\sin\theta)^{-3}(\sin\theta \vee \cos\theta)\frac{2}{\pi}(\cot\theta)^{-3} e^{-(\sqrt{\pi}\cot\theta)^{-2}} d\theta.$$

Example 5.2.3. *If A has a $R(\mu)$ distribution, i.e. the density function of A is $f_A(a) = \frac{a}{\mu^2} e^{-\frac{a^2}{2\mu^2}}$, $a > 0$. ($EA = \sqrt{\frac{\pi}{2}}\mu$, $Var\, A = \frac{4-\pi}{2}\mu^2$). Let $\mu = \sqrt{\frac{2}{\pi}}$, then $EA = 1$ and*

$$H(dw) = \frac{\pi}{w^3}(\frac{1}{w} - 1)e^{-\frac{\pi}{4}(\frac{1}{w}-1)^2} dw,$$
$$\Phi(d\theta) = (\sin\theta)^{-3}(\sin\theta \vee \cos\theta)\frac{\pi}{2}\cot\theta\, e^{-\frac{\pi}{4}\cot^2\theta} d\theta.$$

5.3 Simulation

We have compared the performance of the two estimation methods by simulation. We consider three distributions obtained by the method of Lemma 5.2.2, three distributions obtained by the method of Lemma 5.2.3, as well as the logistic model on which the parametric method is based. Hence in total we consider seven distributions (models).

A. In Lemma 5.2.2 we consider a random variables W with density

$$h(w) = \begin{cases} \frac{b}{a} w & w \in [0, a], \\ \frac{b}{1-a}(1 - w) & w \in (a, p], \\ \frac{b}{1-a}(1 - w) + \frac{c}{1-p}(w - p) & w \in (p, 1], \end{cases}$$

with $0 < a < p < 1$, positive real b and c. $EW = 1/2$ implies

$$b = \frac{1 + 2p}{1 + p - a}, \qquad c = \frac{2 - b}{1 - p}.$$

Consider two specific choices:

- $\mathbf{A_1}$: $a = 1/8 \quad b = 3/2 \quad c = 4/3 \quad p = 5/8$,
- $\mathbf{A_2}$: $a = 2/8 \quad b = 18/11 \quad c = 32/33 \quad p = 5/8$.

B. In Lemma 5.2.2 we consider a discrete distribution for W:

$$P(W = 3/8) = 3/4, \qquad P(W = 7/8) = 1/4.$$

C. In Lemma 5.2.3 we specify:

- $\mathbf{C_1}$: A has a standard exponent distribution (Example 5.2.1),
- $\mathbf{C_2}$: A has a Gamma(α) distribution with $\alpha = 2$ (Example 5.2.2).

D. Now we consider non-symmetric models constructed as in Lemma 5.2.3. Using the technique of Remark 5.2.2 we construct the distribution of A in the following way. First choose $0 < a_1 < 1 < a_2 < \infty$; second choose positive p_1 and p_2 such that

$$p_1 + p_2 = 1, \quad a_1 p_1 + a_2 p_2 = 1,$$

i.e.

$$p_1 = \frac{a_2 - 1}{a_2 - a_1}, \quad p_2 = \frac{1 - a_1}{a_2 - a_1};$$

third choose two independent random variables A_1 and A_2 on $[0, \infty]$ such that $E\,A_1 = a_1$ and $E\,A_2 = a_2$; last define $A = p_1 A_1 + p_2 A_2$.

We consider specification: both with A_1, A_2 having $\mathrm{R}(\mu)$ distributions (see Example 5.2.3), A_1, A_2 independent, $A_1 \sim \mathrm{R}(\mu_1)$, $A_2 \sim \mathrm{R}(\mu_2)$ and $a_1 = 1/5$, $a_2 = 2$, $p_1 = 5/9$, $p_2 = 4/9$, $\mu_1 = \frac{1}{5}\sqrt{\frac{2}{\pi}}$, $\mu_2 = 2\sqrt{\frac{2}{\pi}}$.

E. In Lemma 5.2.2 we consider a random variable W with distribution function

$$H(w, \alpha) = \frac{1}{2}\left((w^{(1-\alpha)/\alpha} - (1-w)^{(1-\alpha)/\alpha}) \cdot (w^{1/\alpha} + (1-w)^{1/\alpha})^{\alpha - 1} + 1 \right)$$

i.e.

$$l(x, y) = (x^{1/\alpha} + y^{1/\alpha})^{\alpha}$$

with $\alpha = 0.5$.

We generate r ($r = 100$) samples (X, Y) for each model with sample size $n = 5000$. There are many methods to determine optimal k. Here, it is reasonable to choose *optimal* k via estimating $l(1,1)$. $l(1,1)$ is a possible measure of dependence in the tail: in case of independence $l(x,y) = x + y$ hence $l(1,1) = 2$; in case of full dependence $l(x,y) = x \vee y$ hence $l(1,1) = 1$. For each model, we first calculate the averages of the two estimators of $l(1,1)$, say $m_{NP}(1,1)$ and $m_{ML}(1,1)$, for varying k from 50 to 500 by increment 50. Here, for each $x, y > 0$

$$m_{NP}(x,y) := \frac{1}{r}\sum_{i=1}^{r} \hat{l}^{(i)}_{NP}(x,y), \qquad m_{ML}(x,y) := \frac{1}{r}\sum_{i=1}^{r} \hat{l}^{(i)}_{ML}(x,y),$$

with $\hat{l}^{(i)}_{NP}$, $\hat{l}^{(i)}_{ML}$ meaning the non-parametric estimator and maximum likelihood estimator of l for ith sample respectively. Then determine the *optimal* k, say k_0, by minimizing the absolute difference between real $l(1,1)$ and its estimator. For each model, different methods may have different k_0. After determining k_0, we calculate the estimator of the dependence function l and the estimator of the spectrum measure Φ for each model.

Table 5.1 (below) lists the mean, sample variance and mean squared error for two estimations for each model. There, $s^2_{NP}(1,1)$ and $\sigma^2_{NP}(1,1)$ are defined by

$$s^2_{NP}(1,1) := \frac{1}{r-1}\sum_{i=1}^{r}\left(\hat{l}^{(i)}_{NP}(1,1) - m_{NP}(1,1)\right)^2,$$
$$\sigma^2_{NP}(1,1) := \frac{1}{r}\sum_{i=1}^{r}\left(\hat{l}^{(i)}_{NP}(1,1) - l(1,1)\right)^2$$

(similar definitions for $s^2_{ML}(1,1)$) and $\sigma^2_{ML}(1,1)$). Table 5.2 and Table 5.3 show the corresponding results when replacing point $(1,1)$ by point $(1,1/2)$ and point $(1/2,1)$ respectively.

For each model we plot two sets of Q-curves:

$$\{(x,y)\in\mathbb{R}^2_+|\ l(x,y)=c\} \quad \text{and} \quad \{(x,y)\in\mathbb{R}^2_+|\ m_{NP}(x,y)=c\}$$
$$\{(x,y)\in\mathbb{R}^2_+|\ l(x,y)=c\} \quad \text{and} \quad \{(x,y)\in\mathbb{R}^2_+|\ \bar{l}_{ML}(x,y)=c\}$$

with $c = 0.2,\ 0.4, ..., 1.0$, and

$$\bar{l}_{ML}(x,y) = (x^{1/\bar{\alpha}} + y^{1/\bar{\alpha}})^{\bar{\alpha}}, \quad \bar{\alpha} = \frac{1}{r}\sum_{i=1}^{r}\alpha_i,$$

where α_i is the ML estimator of α for the ith sample. For each model, the value of $\bar{\alpha}$ is presented in Table 5.1.

For each model we also plot the following two sets of curves:

$$\Phi(\theta) \quad \text{and} \quad \bar{\Phi}_{NP}(\theta), \quad \theta\in[0,\pi/2]$$
$$\Phi(\theta) \quad \text{and} \quad \bar{\Phi}_{ML}(\theta), \quad \theta\in[0,\pi/2],$$

where $\bar{\Phi}_{NP}$ is the average of the estimators of Φ by the non-parametric method and $\bar{\Phi}_{ML}$ is the spectral measure related to $\bar{l}$.

The figures we display below are indicated by a groups of three symbols: first the indication of the model, i.e. $A_1,\ A_2,\ ...,\ E$; then the indication of the object to be estimated, i.e. Φ or Q (for the $Q-$curve); finally the indication of the method of estimation, i.e. ML (maximum likelihood) or NP (nonparametric). Example: $A_1\ Q\ ML$. Note that, for example in Figure 5.1, $A_1\,Q\,NP$, the intersection of the line $y = x$ and the $Q-$curve with indication "1" is $(1/m_{NP}(1,1),\ 1/m_{NP}(1,1))$. Also, in the Figure 5.1, $A_1\,\Phi\,NP$, we find $\bar{\Phi}_{NP}(\pi/2) = m_{NP}(1,1)$.

From Table 5.1-5.3 and Figure 5.1-5.7 it looks to us that the non-parametric method is doing reasonably well throughout. The parametric method, which

	Model	A_1	A_2	B	C_1	C_2	D	E
	$l(1,1)$	1.5179	1.4545	1.3750	1.3679	1.3024	1.4719	1.4142
NP	k_0	400	100	450	100	50	150	250
	$m_{NP}(1,1)$	1.5187	1.4547	1.3743	1.3679	1.3046	1.4700	1.4146
	$s_{NP}(1,1)$	0.0212	0.0432	0.0224	0.0362	0.0595	0.0345	0.0273
	$\sigma_{NP}(1,1)$	0.0211	0.0430	0.0222	0.0360	0.0592	0.0344	0.0272
ML	k_0	400	350	350	150	450	200	50
	$m_{ML}(1,1)$	1.5406	1.4671	1.3749	1.3707	1.3024	1.4727	1.4170
	$s_{ML}(1,1)$	0.0161	0.0150	0.0143	0.0197	0.0095	0.0195	0.0410
	$\sigma_{ML}(1,1)$	0.0279	0.0195	0.0142	0.0198	0.0095	0.0195	0.0409
	$\bar{\alpha}$	0.6235	0.5529	0.4592	0.4548	0.3812	0.5583	0.5023

Table 5.1: Results for point $(1,1)$ with $r = 100$ samples and sample size $n = 5000$.

	Model	A_1	A_2	B	C_1	C_2	D	E
	$l(1,1/2)$	1.1602	1.1187	1.0000	1.0677	1.0785	1.0934	1.1180
NP	$m_{NP}(1,1/2)$	1.1628	1.1224	1.0000	1.0681	1.0842	1.0949	1.1175
	$s_{NP}(1,1/2)$	0.0169	0.0301	0.0000	0.0255	0.0413	0.0261	0.0182
	$\sigma_{NP}(1,1/2)$	0.0171	0.0302	0.0000	0.0254	0.0415	0.0026	0.0181
ML	$m_{ML}(1,1/2)$	1.1941	1.1490	1.0962	1.0940	1.0591	1.1524	1.1201
	$s_{ML}(1,1/2)$	0.0101	0.0090	0.0078	0.0107	0.0045	0.0117	0.0233
	$\sigma_{ML}(1,1/2)$	0.0353	0.0316	0.0305	0.0284	0.0020	0.0602	0.0233

Table 5.2: Results for point $(1,1/2)$ with $r = 100$ samples and sample size $n = 5000$.

	Model	A_1	A_2	B	C_1	C_2	D	E
	$l(1/2,1)$	1.1881	1.1541	1.1563	1.1065	1.0294	1.1704	1.1180
NP	$m_{NP}(1/2,1)$	1.1898	1.1578	1.1571	1.1081	1.0334	1.1697	1.1188
	$s_{NP}(1/2,1)$	0.0184	0.0326	0.0149	0.0291	0.0257	0.0229	0.0200
	$\sigma_{NP}(1/2,1)$	0.0184	0.0327	0.0145	0.0290	0.0259	0.0228	0.0199
ML	$m_{ML}(1/2,1)$	1.1941	1.1490	1.0962	1.0940	1.0591	1.1524	1.1201
	$s_{ML}(1/2,1)$	0.0101	0.0090	0.0078	0.0107	0.0045	0.0117	0.0234
	$\sigma_{ML}(1/2,1)$	0.0117	0.0102	0.0606	0.0164	0.0030	0.0215	0.0234

Table 5.3: Results for point $(1/2,1)$ with $r = 100$ samples and sample size $n = 5000$.

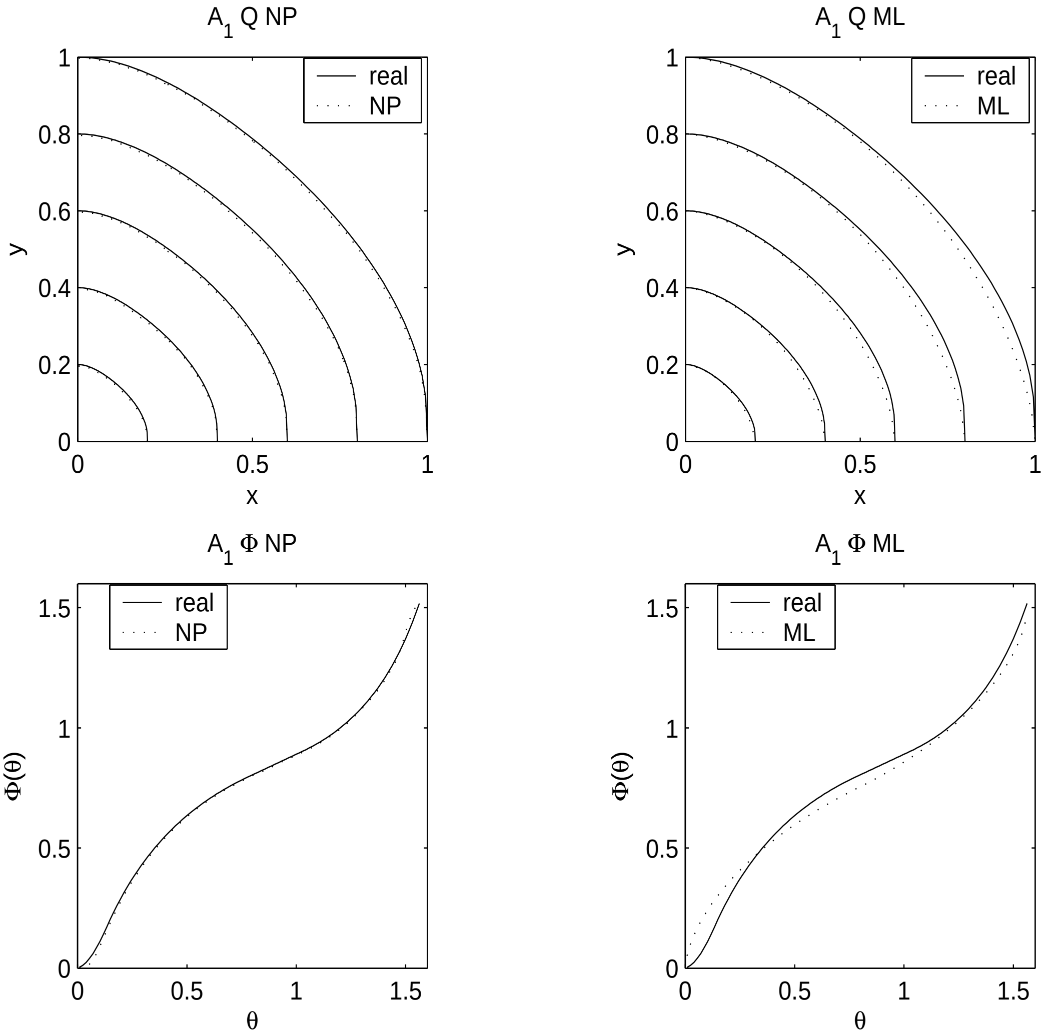

Figure 5.1: Model A_1 with $r = 100$ samples and sample size $n = 5000$.

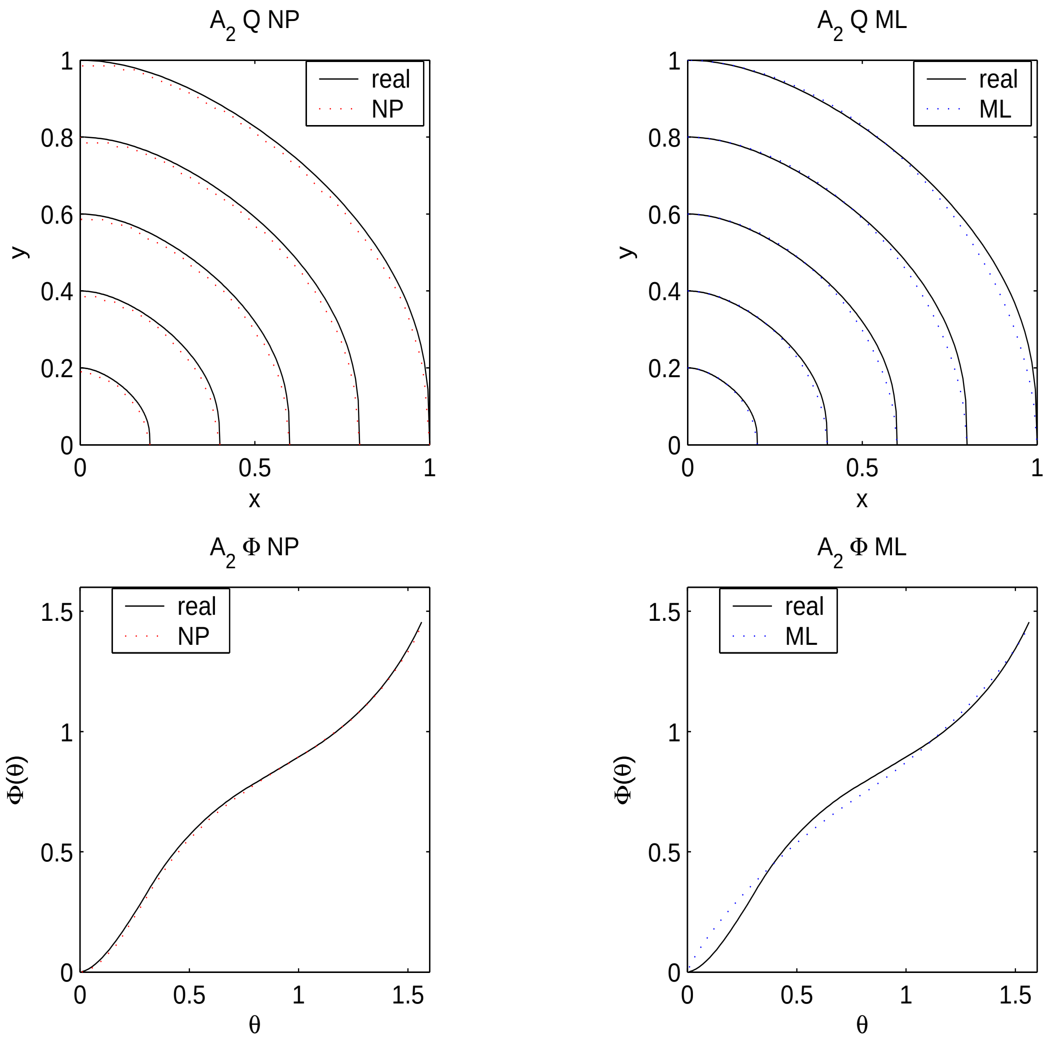

Figure 5.2: Model A_2 with $r = 100$ samples and sample size $n = 5000$.

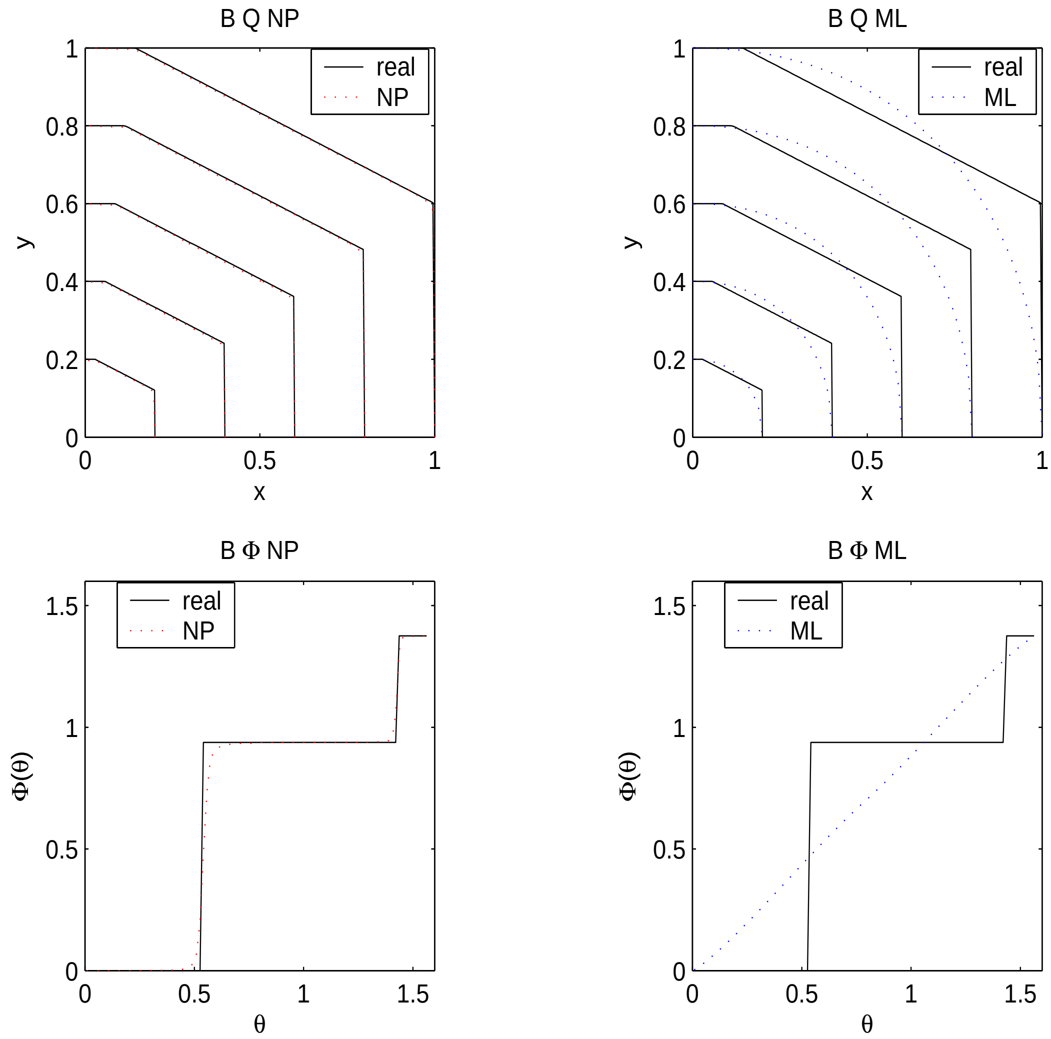

Figure 5.3: Model B with $r = 100$ samples and sample size $n = 5000$.

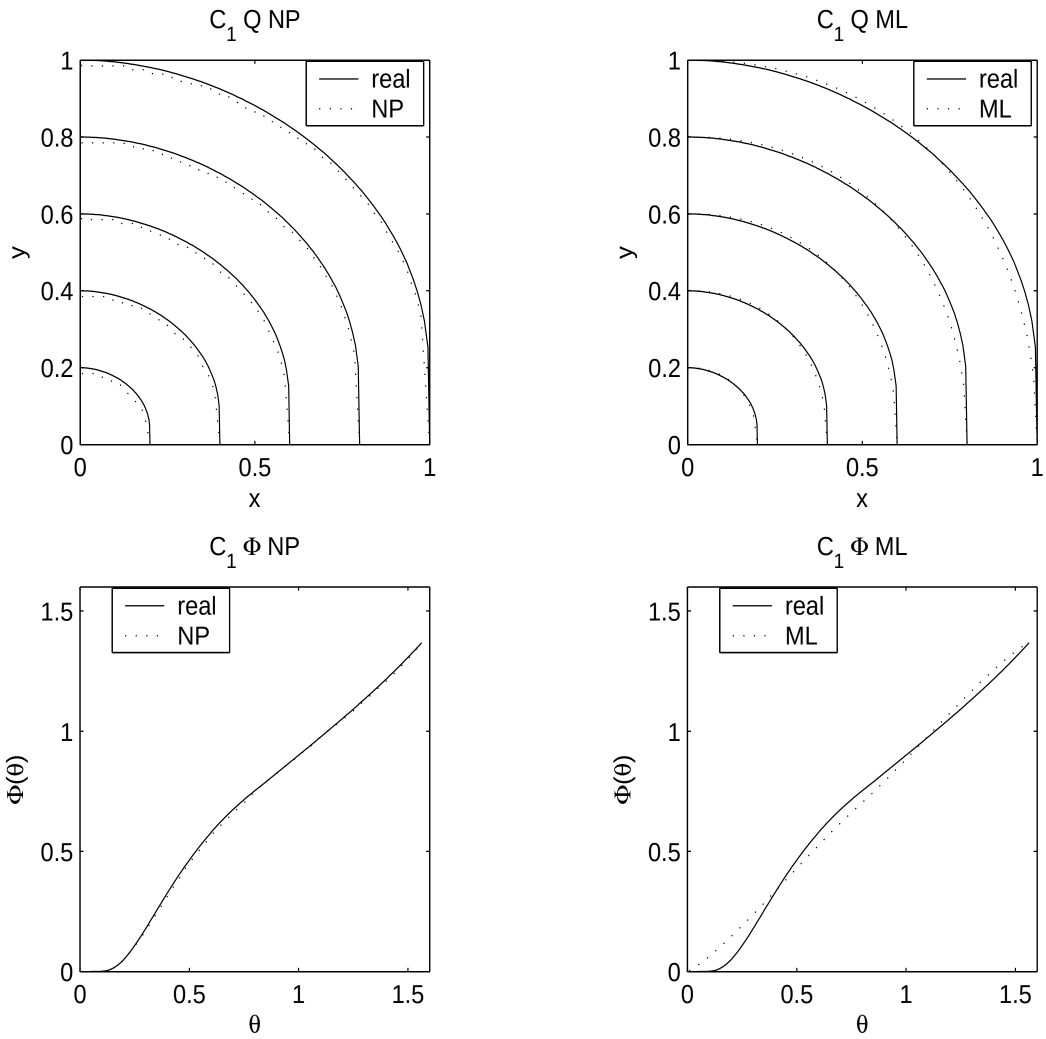

Figure 5.4: Model C_1 with $r = 100$ samples and sample size $n = 5000$.

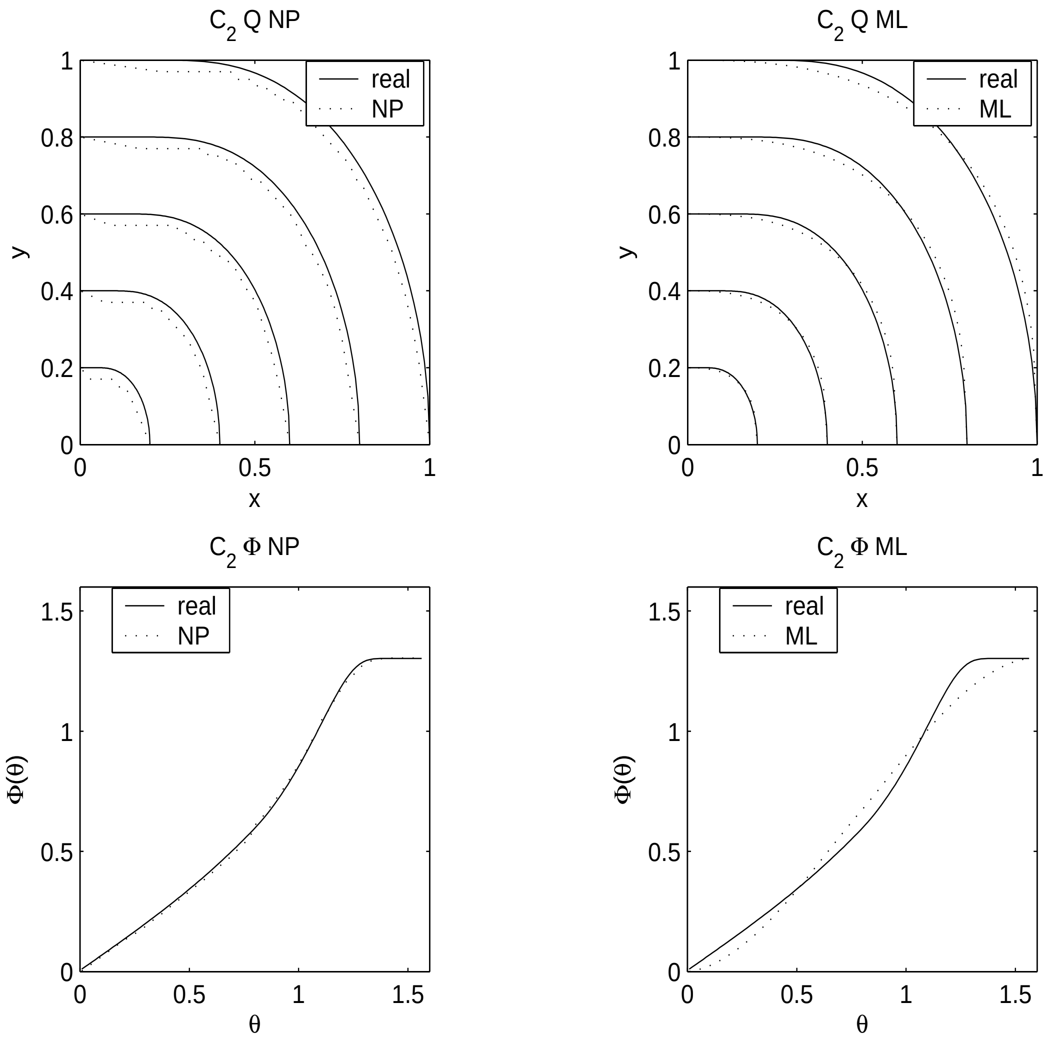

Figure 5.5: Model C_2 with $r = 100$ samples and sample size $n = 5000$.

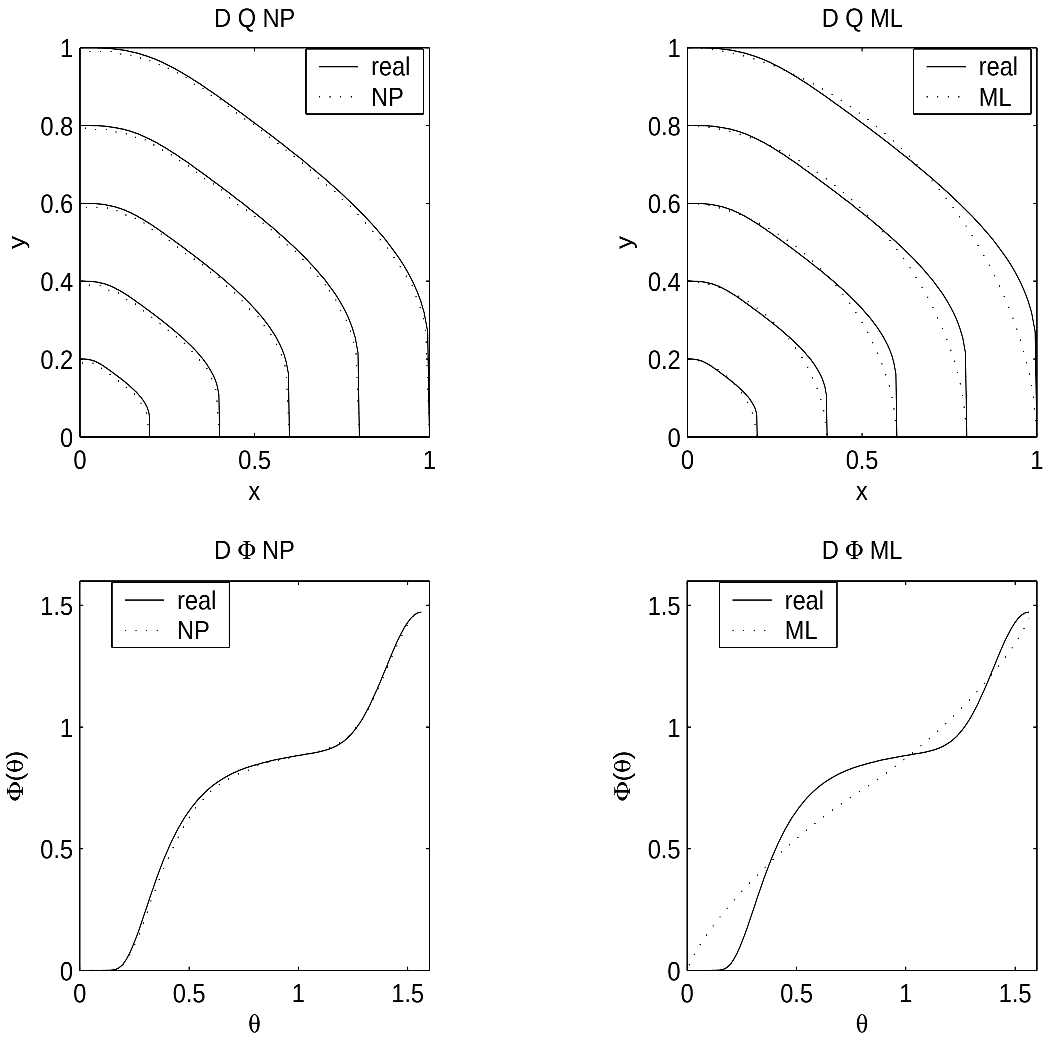

Figure 5.6: Model D with $r = 100$ samples and sample size $n = 5000$.

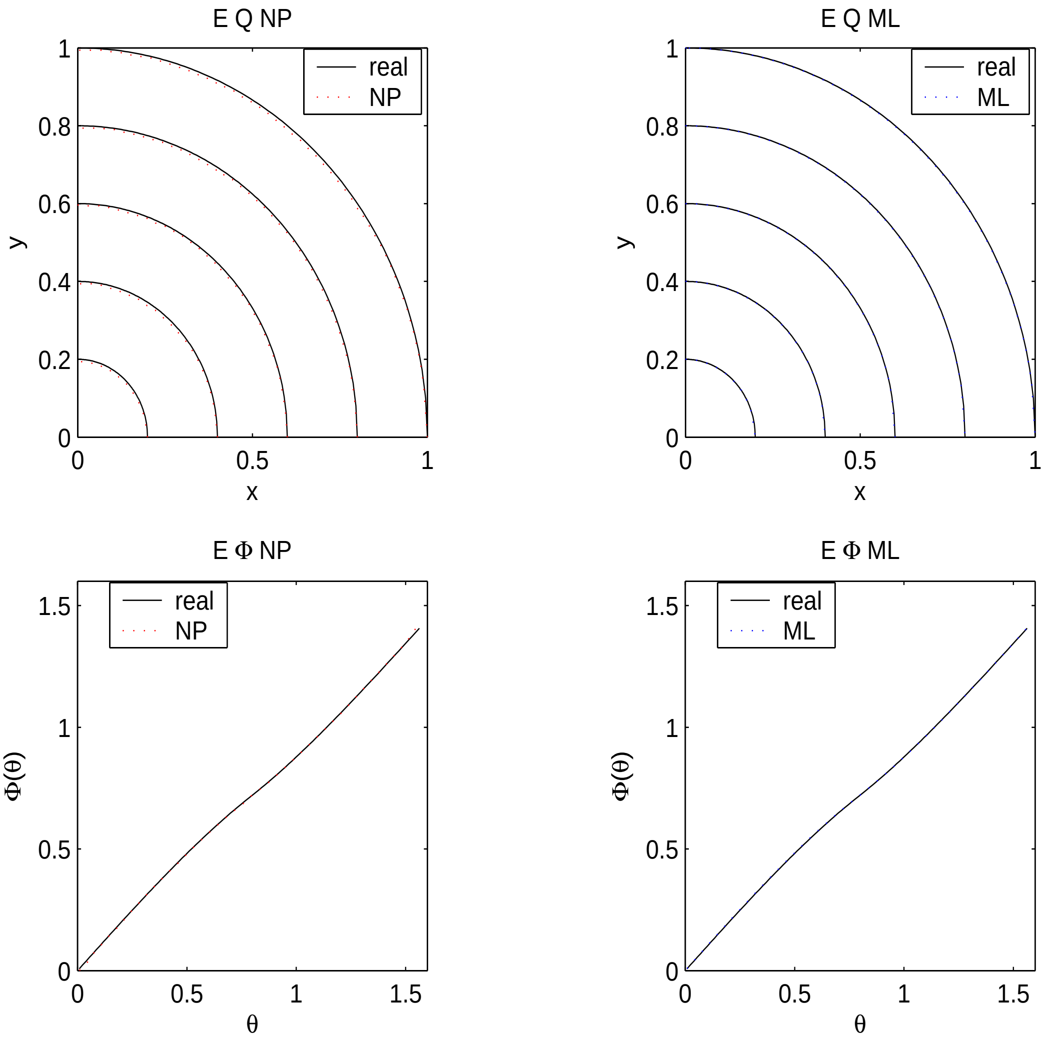

Figure 5.7: Model E with $r = 100$ samples and sample size $n = 5000$.

has been designed only for the logistic model, still gives good results when the spectral distribution is quite smooth and symmetric.

Acknowledgment. We are grateful to two anonymous referees who pointed out errors in several of our calculations.

Chapter 6

Alternative Conditions for Attraction to Stable Vectors

co-authors: Laurens de Haan, Liang Peng and Helena Iglesias Pereira

Prob. Math. Stat. 22 2 (2002) pp. 303-317

Abstract. Relying on Geluk and de Haan (2000) we derive alternative necessary and sufficient conditions for the domain of attraction of a stable distribution in $\mathcal{R}^d$ which are phrased entirely in terms of (joint distributions of) linear combinations of the marginals. The conditions in terms of characteristic functions should be useful for determining rates of convergence, as in de Haan and Peng (1999).

6.1 Introduction and Main Results

Let $\mathbf{X}_1, \mathbf{X}_2, \cdots$ be i.i.d. random vectors taken values in $\mathcal{R}^d$. We consider the sequence $\mathbf{S}_n := \mathbf{X}_1 + \cdots + \mathbf{X}_n$, $n = 1, 2, \cdots$, and suppose that for some sequences of norming constants $a_n > 0$ and $\mathbf{b}_n$ $(n = 1, 2, \cdots)$ the sequence $\mathbf{S}_n/a_n - \mathbf{b}_n$ has a limit distribution with non-degenerate marginals.

The limit distributions are called stable distributions and the set of distributions such that $\mathbf{S}_n/a_n - \mathbf{b}_n$ converges to a particular stable distribution is called its domain of attraction.

The indicated results have been developed a long time ago. The stable distributions have been identified by E. Feldheim in 1937 under the direction of P. Levy and the domain of attraction conditions by E.L. Rvaceva under the direction of B.V. Gnedenko in 1950. A full account of the theory is Rvaceva (1962). For stable stochastic processes see Samorodnitsky and Taqqu (1994).

Here we use the methods of Geluk and de Haan (2000) to arrive at alternative domain of attraction conditions based on the probability distributions of linear combinations of the marginal random variables. However the relation between our conditions and those of Rvaceva (1962) are not easy to derive directly. We can prove only the implication in one direction, for the other direction we use Feller's methods (see Section 6.3).

We start by stating the general form of the characteristic function ψ of a stable distribition: for $0 < \alpha < 2$ we have

$$\psi(\boldsymbol{\theta}) = \exp\Big(-\int_{\mathbf{S}} \big(|\boldsymbol{\theta}^T\mathbf{u}|^\alpha + i\boldsymbol{\theta}^T\mathbf{u}(1-\alpha)\tan\frac{\pi\alpha}{2}\frac{|\boldsymbol{\theta}^T\mathbf{u}|^{\alpha-1}-1}{\alpha-1}\big)\,\mu(d\mathbf{u})\Big), \quad (6.1.1)$$

where

$$\boldsymbol{\theta} = (\theta_1, \cdots, \theta_d)^T, \qquad \mathbf{u} = (u_1, \cdots, u_d)^T,$$
$$\mathbf{S} := \{\mathbf{x} = (x_1, \cdots, x_d)^T : \mathbf{x}^T\mathbf{x} = 1\},$$

and μ is a positive and finite measure on $\mathbf{S}$, or any other distribution of the same type.

For $\alpha = 2$ we have

$$\psi(\boldsymbol{\theta}) = \exp\{-q(\boldsymbol{\theta})\}, \quad (6.1.2)$$

where $q(\boldsymbol{\theta}) = \boldsymbol{\theta}^T\mathcal{Q}\boldsymbol{\theta}$ and $\mathcal{Q}$ is symmetric and positive definite, or any other distribution of the same type.

For $\alpha = 1$ the function ψ is to be understood by continuity; so $(|t|^{\alpha-1} - 1)/(\alpha-1)$ becomes $\log|t|$ and $(1-\alpha)\tan(\pi\alpha/2)$ becomes $2/\pi$ for $\alpha = 1$.

We shall now state our results. For ease of writing only we restrict ourselves to the two-dimensional case. So let $(X_1, X_2), (X_{11}, X_{21}), \cdots$ be i.i.d. random vectors with distribution function F and characteristic function ϕ. As in Geluk and de Haan (2000) we define for $t > 0$ and $\theta_1\theta_2 \neq 0$

$$U_{(\theta_1,\theta_2)}(t) := \mathrm{Re}\phi(\theta_1/t, \theta_2/t), \qquad V_{(\theta_1,\theta_2)}(t) := \mathrm{Im}\phi(\theta_1/t, \theta_2/t),$$

$$c_\alpha := \int_1^\infty x^{-\alpha}\cos x\,dx + \int_0^1 x^{-\alpha}(\cos x - 1)\,dx = \Gamma(1-\alpha)\sin\frac{\pi\alpha}{2} - \frac{1}{1-\alpha}.$$

Theorem 6.1.1. *Let the random vector* (W_1, W_2) *have characteristic function* ψ *from (6.1.1) for some* $0 < \alpha < 2$. *The following statements are equivalent:*
A. *There exist sequences* $a_n > 0$, b_n *and* d_n $(n = 1, 2, \cdots)$ *such that*

$$\Big(\sum_{j=1}^n X_{1j}/a_n - b_n, \sum_{j=1}^n X_{2j}/a_n - d_n\Big) \xrightarrow{d} (W_1, W_2).$$

B. *For all* (θ_1, θ_2),

$$\lim_{t\to\infty} \frac{P(\theta_1X_1 + \theta_2X_2 > t)}{P(|X_1 + X_2| > t)} = \frac{\int_{\mathbf{S}} |\theta_1u_1 + \theta_2u_2|^\alpha(1 + \mathrm{sign}(\theta_1u_1 + \theta_2u_2))\,\mu(du_1, du_2)}{2\int_{\mathbf{S}} |u_1 + u_2|^\alpha\mu(du_1, du_2)}, \quad (6.1.3)$$

$$\lim_{t\to\infty} \frac{\int_0^1 \left(\Delta_{(\theta_1,\theta_2)}(ts) - \theta_1\Delta_{(1,0)}(ts) - \theta_2\Delta_{(0,1)}(ts)\right) ds}{P(|X_1+X_2|>t)}$$
$$= -\frac{\int_{\mathbf{S}} \left(\frac{|\theta_1 u_1+\theta_2 u_2|^{\alpha-1}-1}{\alpha-1}(\theta_1 u_1+\theta_2 u_2) - \frac{|u_1|^{\alpha-1}-1}{\alpha-1}\theta_1 u_1 - \frac{|u_2|^{\alpha-1}-1}{\alpha-1}\theta_2 u_2\right)\mu(du_1,du_2)}{\int_{\mathbf{S}}|u_1+u_2|^\alpha \mu(du_1,du_2)}, \tag{6.1.4}$$

where

$$\Delta_{(\theta_1,\theta_2)}(t) = P(\theta_1 X_1+\theta_2 X_2 > t) - P(\theta_1 X_1 + \theta_2 X_2 < -t).$$

C. *For all* (θ_1,θ_2),

$$\lim_{t\to\infty} \frac{1-U_{(\theta_1,\theta_2)}(t)}{1-U_{(1,1)}(t)} = \frac{\int_{\mathbf{S}}|\theta_1 u_1+\theta_2 u_2|^\alpha \mu(du_1,du_2)}{\int_{\mathbf{S}}|u_1+u_2|^\alpha\mu(du_1,du_2)}, \tag{6.1.5}$$

$$\lim_{t\to\infty} \frac{V_{(\theta_1,\theta_2)}(t) - \theta_1 V_{(1,0)}(t) - \theta_2 V_{(0,1)}(t)}{1-U_{(1,1)}(t)}$$
$$= -(1-\alpha)\tan\frac{\pi\alpha}{2}\left(\int_{\mathbf{S}}|u_1+u_2|^\alpha\mu(du_1,du_2)\right)^{-1}\times$$
$$\times\int_{\mathbf{S}}\left(\frac{|\theta_1 u_1+\theta_2 u_2|^{\alpha-1}-1}{\alpha-1}(\theta_1 u_1+\theta_2 u_2) - \frac{|u_1|^{\alpha-1}-1}{\alpha-1}\theta_1 u_1 - \frac{|u_2|^{\alpha-1}-1}{\alpha-1}\theta_2 u_2\right)\times$$
$$\times\,\mu(du_1,du_2). \tag{6.1.6}$$

Remark 6.1.1. *The condition in (6.1.4) can be replaced by*

$$\lim_{t\to\infty}\frac{E(\theta_1X_1+\theta_2X_2)I(|\theta_1X_1+\theta_2X_2|<t) - \theta_1 EX_1 I(|X_1|<t) - \theta_2 EX_2 I(|X_2|<t)}{tP(|X_1+X_2|>t)}$$
$$= -\frac{\int_{\mathbf{S}}\left(\frac{|\theta_1 u_1+\theta_2 u_2|^{\alpha-1}-1}{\alpha-1}(\theta_1 u_1+\theta_2 u_2) - \frac{|u_1|^{\alpha-1}-1}{\alpha-1}\theta_1 u_1 - \frac{|u_2|^{\alpha-1}-1}{\alpha-1}\theta_2 u_2\right)\mu(du_1,du_2)}{\int_{\mathbf{S}}|u_1+u_2|^\alpha\mu(du_1,du_2)}$$
$$-\frac{\int_{\mathbf{S}}\left(|\theta_1 u_1+\theta_2 u_2|^{\alpha-1}(\theta_1 X_1+\theta_2 X_2) - \theta_1|u_1|^{\alpha-1}u_1 - \theta_2|u_2|^{\alpha-1}u_2\right)\mu(du_1,du_2)}{\int_{\mathbf{S}}|u_1+u_2|^\alpha\,\mu(du_1,du_2)}.$$

Remark 6.1.2. *From Theorem 6.1.1 we conjecture that requiring a rate of convergence in (6.1.3) and (6.1.4) will lead to a uniform rate of convergence in statement* **A**. *This will be a part of our future research.*

For $0<\alpha<2$, $\alpha\neq 1$, the conditions in Theorem 6.1.1 can be simplified as follows.

Theorem 6.1.2. *Let the random vector* (W_1, W_2) *have characteristic function* ψ *from (6.1.1) for some* $0 < \alpha < 2$, $\alpha \neq 1$. *The following statements are equivalent:*
A. *There exist sequences* $a_n > 0$, b_n *and* d_n $(n = 1, 2, \cdots)$ *such that*

$$\Big(\sum_{j=1}^{n} X_{1j}/a_n - b_n, \sum_{j=1}^{n} X_{2j}/a_n - d_n\Big) \xrightarrow{d} (W_1, W_2).$$

B. *For all* (θ_1, θ_2),

$$\lim_{t\to\infty} \frac{P(\theta_1 X_1 + \theta_2 X_2 > t)}{P(|X_1 + X_2| > t)} = \frac{\int_{\mathbf{S}} |\theta_1 u_1 + \theta_2 u_2|^\alpha (1 + \operatorname{sign}(\theta_1 u_1 + \theta_2 u_2))\, \mu(du_1, du_2)}{2 \int_{\mathbf{S}} |u_1 + u_2|^\alpha \mu(du_1, du_2)}.$$

C. *For all* $(\theta_1, \theta_2) \neq (0, 0)$,

$$\lim_{t\to\infty} \frac{1 - U_{(\theta_1,\theta_2)}(t)}{1 - U_{(1,1)}(t)} = \frac{\int_{\mathbf{S}} |\theta_1 u_1 + \theta_2 u_2|^\alpha \mu(du_1, du_2)}{\int_{\mathbf{S}} |u_1 + u_2|^\alpha \mu(du_1, du_2)},$$

and

$$\lim_{t\to\infty} \frac{V_{(\theta_1,\theta_2)}(t)}{1 - U_{(1,1)}(t)} = \tan\frac{\alpha\pi}{2} \frac{\int_{\mathbf{S}} |\theta_1 u_1 + \theta_2 u_2|^\alpha \operatorname{sign}(\theta_1 u_1 + \theta_2 u_2)\, \mu(du_1, du_2)}{\int_{\mathbf{S}} |u_1 + u_2|^\alpha\, \mu(du_1, du_2)} \quad \textit{if} \quad 0 < \alpha < 1,$$

$$\lim_{t\to\infty} \frac{tV_{(\theta_1,\theta_2)}(t) - \theta_1 E(X_1) - \theta_2 E(X_2)}{t[1 - U_{(1,1)}(t)]} = \tan\frac{\alpha\pi}{2} \frac{\int_{\mathbf{S}} |\theta_1 u_1 + \theta_2 u_2|^\alpha \operatorname{sign}(\theta_1 u_1 + \theta_2 u_2)\, \mu(du_1, du_2)}{\int_{\mathbf{S}} |u_1 + u_2|^\alpha\, \mu(du_1, du_2)} \quad \textit{if} \quad 1 < \alpha < 2.$$

Now we consider the normal limit distribution.

Theorem 6.1.3. *Let the random vector* (W_1, W_2) *have characteristic function* ψ *from (6.1.2). The following statements are equivalent:*
A. *There exist sequences* $a_n > 0$, b_n *and* d_n $(n = 1, 2, \cdots)$ *such that*

$$\Big(\sum_{j=1}^{n} X_{1j}/a_n - b_n, \sum_{j=1}^{n} X_{2j}/a_n - d_n\Big) \xrightarrow{d} (W_1, W_2).$$

B. *For all* (θ_1, θ_2),

$$\lim_{t\to\infty} \frac{\int_0^t P\big((\theta_1 X_1 + \theta_2 X_2)^2 > s\big)\, ds}{\int_0^t P\big((X_1 + X_2)^2 > s\big)\, ds} = \frac{q(\theta_1, \theta_2)}{q(1, 1)}. \tag{6.1.7}$$

C. *For all* (θ_1, θ_2),

$$\lim_{t\to\infty} \frac{1 - U_{(\theta_1,\theta_2)}(t)}{1 - U_{(1,1)}(t)} = \frac{q(\theta_1, \theta_2)}{q(1,1)} \tag{6.1.8}$$

and

$$\lim_{t\to\infty} \frac{E(\theta_1 X_1 + \theta_2 X_2) - tV_{(\theta_1,\theta_2)}(t)}{1 - U_{(1,1)}(t)} = 0. \tag{6.1.9}$$

Remark 6.1.3. *Relation (6.1.7) is equivalent to*

$$\lim_{t\to\infty} \frac{E(\theta_1 X_1 + \theta_2 X_2)^2 I(|\theta_1 X_1 + \theta_2 X_2| \le t)}{E(X_1 + X_2)^2 I(|X_1 + X_2| \le t)} = \frac{q(\theta_1, \theta_2)}{q(1,1)}. \tag{6.1.10}$$

Section 6.2 contains proofs. In Section 6.3 we explore the relation between relations **B** of Theorem 6.1.1 and the well known condition of Rvaceva (1962):

$$\lim_{t\to\infty} \frac{P(\sqrt{X_1^2 + X_2^2} > tx, \arctan(X_2/X_1) \in A)}{P(\sqrt{X_1^2 + X_2^2} > t)} = x^{-\alpha} \frac{\mu(A)}{\mu(\mathbf{S})} \tag{6.1.11}$$

for each $x > 0$ and each Borel subset A of $\mathbf{S}$ which is a continuity set for μ.

6.2 Proofs

Lemma 6.2.1. *If $f(t) \in RV_0$ and there exists $\{a_n\}$ such that $a_n \to \infty$, $a_{n+1}/a_n \to 1$ and $f(a_n) \to c$ as $n \to \infty$, then $\lim_{t\to\infty} f(t) = c$.*

Proof. For any $\varepsilon, \delta > 0$, there exists $t_0 = t_0(\varepsilon, \delta) > 0$ such that

$$|f(tx)/f(t) - 1| \le \varepsilon max(x^{\delta}, x^{-\delta})$$

for all $t, tx \ge t_0$. For any sequence $\{t_n\}$ such that $t_n \to \infty$ as $n \to \infty$, there exists $\{k_n\}$ such that $a_{k_n} \le t_n \le a_{k_{n+1}}$. Let $x_n = t_n/a_{k_n}$. Then $\lim_{n\to\infty} x_n = 1$. Hence there exists N such that for all $n \ge N$

$$|f(t_n)/f(a_{k_n}) - 1| \le \varepsilon 2^{\delta},$$

i.e. $|f(t_n)/f(a_{k_n}) - 1| \to 0$ as $n \to \infty$. Since $|f(a_{k_n}) - c| \to 0$ as $n \to \infty$, we have

$$\lim_{n\to\infty} |f(t_n) - c| = \lim_{n\to\infty} \left| \frac{f(t_n)}{f(a_{k_n})} f(a_{k_n}) - c \right| = 0.$$

Hence the lemma. □

Lemma 6.2.2. *Let X be a random variable. Define $U(t) = \mathrm{Re}Ee^{iX/t}$ for $t \neq 0$. The following are equivalent:*

1. *The function $P(|X| > t)$ is regularly varying with index $\alpha \in (0, 2)$.*
2. *The function $1 - U(t)$ is regularly varying with index $\alpha \in (0, 2)$.*

Moreover, both imply

$$\lim_{t\to\infty} \frac{1 - U(t)}{P(|X| > t)} = \Gamma(1-\alpha)\cos\frac{\pi\alpha}{2},$$

to be interpreted as $\pi/2$ for $\alpha = 1$.

Proof. This is just part of the proof of (ii) ⇔ (iii) of Theorem 1 of Geluk and de Haan (2000). □

Proof of Theorem 6.1.1.
A ⇒ **C**. By the continuity theorem for characteristic functions statement **A** is equivalent to

$$\lim_{n\to\infty} \phi^n(\theta_1/a_n, \theta_2/a_n)e^{-ib_n\theta_1}e^{-id_n\theta_2} = \psi(\theta_1, \theta_2) \tag{6.2.1}$$

locally uniformly. Feller (1971, ChXVII, Section 1, Theorem 1) shows that this is equivalent to

$$\lim_{n\to\infty} n\big(\phi(\theta_1/a_n, \theta_2/a_n) - 1\big) - ib_n\theta_1 - id_n\theta_2 = \log\psi(\theta_1, \theta_2)$$

locally uniformly or

$$\lim_{n\to\infty} n\big(1 - U_{(\theta_1,\theta_2)}(a_n)\big) = \int_{\mathbf{S}} |\theta_1u_1 + \theta_2u_2|^\alpha \mu(du_1, du_2), \tag{6.2.2}$$

$$\begin{aligned}&\lim_{n\to\infty} nV_{(\theta_1,\theta_2)}(a_n) - \theta_1b_n - \theta_2d_n \\ &= -(1-\alpha)\tan\frac{\alpha\pi}{2}\int_{\mathbf{S}} \frac{|\theta_1u_1 + \theta_2u_2|^{\alpha-1} - 1}{\alpha - 1}(\theta_1u_1 + \theta_2u_2)\,\mu(du_1, du_2).\end{aligned} \tag{6.2.3}$$

From relation (6.2.3) we have

$$\begin{aligned}\lim_{n\to\infty} nV_{(1,0)}(a_n) - b_n &= -(1-\alpha)\tan\frac{\alpha\pi}{2}\int_{\mathbf{S}} \frac{|u_1|^{\alpha-1} - 1}{\alpha - 1}u_1\,\mu(du_1, du_2), \\ \lim_{n\to\infty} nV_{(0,1)}(a_n) - d_n &= -(1-\alpha)\tan\frac{\alpha\pi}{2}\int_{\mathbf{S}} \frac{|u_2|^{\alpha-1} - 1}{\alpha - 1}u_2\,\mu(du_1, du_2).\end{aligned} \tag{6.2.4}$$

Combination of (6.2.3) and (6.2.4) gives

$$\begin{aligned}&\lim_{n\to\infty} n\big(V_{(\theta_1,\theta_2)}(a_n) - \theta_1V_{(1,0)}(a_n) - \theta_2V_{(0,1)}(a_n)\big) \\ &= -(1-\alpha)\tan\frac{\pi\alpha}{2}\int_{\mathbf{S}} \Big(\frac{|\theta_1u_1 + \theta_2u_2|^{\alpha-1} - 1}{\alpha - 1}(\theta_1u_1 + \theta_2u_2) \\ &\qquad - \frac{|u_1|^{\alpha-1} - 1}{\alpha - 1}\theta_1u_1 - \frac{|u_2|^{\alpha-1} - 1}{\alpha - 1}\theta_2u_2\Big)\,\mu(du_1, du_2).\end{aligned} \tag{6.2.5}$$

We are now going to use one-dimensional results. It follows from (6.2.1) and Theorem 1 of Geluk and de Haan (2000) that

$$1 - U_{(\theta_1,\theta_2)}(t) \in RV_{-\alpha}, \tag{6.2.6}$$

$$\lim_{t\to\infty} \frac{txV_{(\theta_1,\theta_2)}(tx) - tV_{(\theta_1,\theta_2)}(t)}{t\big(1 - U_{(\theta_1,\theta_2)}(t)\big)}$$
$$= (1-\alpha)\tan\frac{\alpha\pi}{2}\,\frac{|x|^{1-\alpha}-1}{1-\alpha}\,\frac{\int_{\mathbf{S}} |\theta_1 u_1 + \theta_2 u_2|^{\alpha}\mathrm{sign}(\theta_1 u_1 + \theta_2 u_2)\,\mu(du_1,du_2)}{\int_{\mathbf{S}} |\theta_1 u_1 + \theta_2 u_2|^{\alpha}\,\mu(du_1,du_2)}. \tag{6.2.7}$$

Since (6.2.6) holds in particular for $(\theta_1,\theta_2) = (1,1)$, we get

$$\frac{1 - U_{(\theta_1,\theta_2)}(t)}{1 - U_{(1,1)}(t)} \in RV_0 \tag{6.2.8}$$

By (6.2.2), (6.2.8) and Lemma 6.2.1

$$\lim_{t\to\infty} \frac{1 - U_{(\theta_1,\theta_2)}(t)}{1 - U_{(1,1)}(t)} = \lim_{n\to\infty} \frac{1 - U_{(\theta_1,\theta_2)}(a_n)}{1 - U_{(1,1)}(a_n)} = \frac{\int_{\mathbf{S}} |\theta_1 u_1 + \theta_2 u_2|^{\alpha}\mu(du_1,du_2)}{\int_{\mathbf{S}} |u_1 + u_2|^{\alpha}\mu(du_1,du_2)}, \tag{6.2.9}$$

i.e. (6.1.5) is proved. Now (6.2.7) allows us to replace the argument a_n in (6.2.5) by $a_n x$ in each of the three terms separately. This results in

$$\lim_{n\to\infty} xn\big(V_{(\theta_1,\theta_2)}(xa_n) - \theta_1 V_{(1,0)}(xa_n) - \theta_2 V_{(0,1)}(xa_n)\big)$$
$$= -(1-\alpha)\tan\frac{\alpha\pi}{2}|x|^{1-\alpha}\int_{\mathbf{S}}\Big(\frac{|\theta_1 u_1 + \theta_2 u_2|^{\alpha-1}-1}{\alpha-1}(\theta_1 u_1 + \theta_2 u_2)$$
$$- \frac{|u_1|^{\alpha-1}-1}{\alpha-1}\theta_1 u_1 - \frac{|u_2|^{\alpha-1}-1}{\alpha-1}\theta_2 u_2\Big)\,\mu(du_1,du_2)$$

for each $x > 0$. By Lemma 9 in Geluk and de Haan (2000), this implies

$$V_{(\theta_1,\theta_2)}(t) - \theta_1 V_{(1,0)}(t) - \theta_2 V_{(0,1)}(t) \in RV_{-\alpha}$$

and we have

$$\frac{V_{(\theta_1,\theta_2)}(t) - \theta_1 V_{(1,0)}(t) - \theta_2 V_{(0,1)}(t)}{1 - U_{(1,1)}(t)} \in RV_0. \tag{6.2.10}$$

Using (6.2.2), (6.2.5), (6.2.10) and Lemma 6.2.1, we now get (6.1.6).

C $\Rightarrow$ **A**. By taking $(\theta_1,\theta_2) = (x,x)$ for some $x > 0$ in (6.1.5), we find that $1 - U_{(1,1)}(t)$ is regularly varying with index $-\alpha$. Hence we can define sequences $a_n > 0$, b_n and d_n such that

$$\lim_{n\to\infty} n\big(1 - U_{(1,1)}(a_n)\big) = \int_{\mathbf{S}} |u_1 + u_2|^{\alpha}\,\mu(du_1,du_2),$$

$$b_n := nV_{(1,0)}(a_n) + (1-\alpha)\tan\frac{\alpha\pi}{2}\int_{\mathbf{S}} \frac{|u_1|^{\alpha-1}-1}{\alpha-1} u_1\,\mu(du_1, du_2),$$

$$d_n := nV_{(0,1)}(a_n) + (1-\alpha)\tan\frac{\alpha\pi}{2}\int_{\mathbf{S}} \frac{|u_2|^{\alpha-1}-1}{\alpha-1} u_2\,\mu(du_1, du_2).$$

Combining the definition of a_n with relation (6.1.5) we get for any (θ_1, θ_2)

$$\lim_{n\to\infty} n\big(1 - U_{(\theta_1,\theta_2)}(a_n)\big) = \int_{\mathbf{S}} |\theta_1 u_1 + \theta_2 u_2|^\alpha\,\mu(du_1, du_2). \tag{6.2.11}$$

Further, combining (6.1.6) and the definitions of a_n, b_n and d_n, we get for any (θ_1, θ_2)

$$\begin{aligned}
&\lim_{n\to\infty} \big(nV_{(\theta_1,\theta_2)}(a_n) - \theta_1 b_n - \theta_2 d_n\big)\\
&= \lim_{n\to\infty} n\big(V_{(\theta_1,\theta_2)}(a_n) - \theta_1 V_{(1,0)}(a_n) - \theta_2 V_{(0,1)}(a_n)\big)\\
&\qquad - (1-\alpha)\tan\frac{\alpha\pi}{2}\int_{\mathbf{S}} \frac{|u_1|^{\alpha-1}-1}{\alpha-1}\theta_1 u_1\,\mu(du_1, du_2)\\
&\qquad - (1-\alpha)\tan\frac{\alpha\pi}{2}\int_{\mathbf{S}} \frac{|u_2|^{\alpha-1}-1}{\alpha-1}\theta_2 u_2\,\mu(du_1, du_2)\\
&= -(1-\alpha)\tan\frac{\pi\alpha}{2}\int_{\mathbf{S}} \Big(\frac{|\theta_1 u_1 + \theta_2 u_2|^{\alpha-1}-1}{\alpha-1}(\theta_1 u_1 + \theta_2 u_2)\\
&\qquad\quad - \frac{|u_1|^{\alpha-1}-1}{\alpha-1}\theta_1 u_1 - \frac{|u_2|^{\alpha-1}-1}{\alpha-1}\theta_2 u_2\Big)\,\mu(du_1, du_2)\\
&\qquad - (1-\alpha)\tan\frac{\alpha\pi}{2}\int_{\mathbf{S}} \frac{|u_1|^{\alpha-1}-1}{\alpha-1}\theta_1 u_1\,\mu(du_1, du_2)\\
&\qquad - (1-\alpha)\tan\frac{\alpha\pi}{2}\int_{\mathbf{S}} \frac{|u_2|^{\alpha-1}-1}{\alpha-1}\theta_2 u_2\,\mu(du_1, du_2)\\
&= -(1-\alpha)\tan\frac{\alpha\pi}{2}\int_{\mathbf{S}} \frac{|\theta_1 u_1 + \theta_2 u_2|^{\alpha-1}-1}{\alpha-1}(\theta_1 u_1 + \theta_2 u_2)\,\mu(du_1, du_2).
\end{aligned} \tag{6.2.12}$$

Hence by (6.2.11) and (6.2.12) statement **A** holds.

B $\Leftrightarrow$ **C**. By Lemma 6.2.2, (6.1.5) is equivalent to

$$\lim_{t\to\infty} \frac{P(|\theta_1 X_1 + \theta_2 X_2| > t)}{P(|X_1 + X_2| > t)} = \frac{\int_{\mathbf{S}} |\theta_1 u_1 + \theta_2 u_2|^\alpha \mu(du_1, du_2)}{\int_{\mathbf{S}} |u_1 + u_2|^\alpha \mu(du_1, du_2)}. \tag{6.2.13}$$

C $\Rightarrow$ **B**. Application of (6.1.6) to $V_{(\theta_1/x,\theta_2/x)}(t) = V_{(\theta_1,\theta_2)}(tx)$ and $V_{(\theta_1,\theta_2)}(t)$ and combination of the results gives for $x > 0$

$$\begin{aligned}
&\lim_{t\to\infty} \frac{txV_{(\theta_1,\theta_2)}(tx) - tV_{(\theta_1,\theta_2)}(t)}{t\big(1 - U_{(\theta_1,\theta_2)}(t)\big)}\\
&\quad = (1-\alpha)\tan\frac{\alpha\pi}{2}\,\frac{|x|^{1-\alpha}-1}{1-\alpha}\,\frac{\int_{\mathbf{S}} |\theta_1 u_1 + \theta_2 u_2|^\alpha \operatorname{sign}(\theta_1 u_1 + \theta_2 u_2)\,\mu(du_1, du_2)}{\int_{\mathbf{S}} |\theta_1 u_1 + \theta_2 u_2|^\alpha\,\mu(du_1, du_2)}.
\end{aligned}$$

We also know $1-U_{(\theta_1,\theta_2)}(t) \in RV_{-\alpha}$ by (6.1.5). Hence the conditions of Theorem 1, part (iii), of Geluk and de Haan (2000) are fulfilled. Thus for any $(\theta_1,\theta_2) \neq (0,0)$ the random variable $\theta_1 X_1 + \theta_2 X_2$ is in the domain of attraction of a stable law. Then Theorem 1 of Geluk and de Haan (2000), part (ii), and relation (10), give

$$\begin{aligned} &\lim_{t\to\infty} \frac{P(\theta_1 X_1 + \theta_2 X_2 > t)}{P(|\theta_1 X_1 + \theta_2 X_2| > t)} \\ &= \frac{\int_{\mathbf{S}} |\theta_1 u_1 + \theta_2 u_2|^\alpha \big(1 + \operatorname{sign}(\theta_1 u_1 + \theta_2 u_2)\big)\, \mu(du_1, du_2)}{2\int_{\mathbf{S}} |\theta_1 u_1 + \theta_2 u_2|^\alpha\, \mu(du_1, du_2)} \end{aligned} \tag{6.2.14}$$

and

$$\begin{aligned} &\lim_{t\to\infty} \frac{V_{(\theta_1,\theta_2)}(t) - t^{-1}\int_0^t \Delta_{(\theta_1,\theta_2)}(s)\, ds}{P(|\theta_1 X_1 + \theta_2 X_2| > t)} \\ &= c_\alpha \frac{\int_{\mathbf{S}} |\theta_1 u_1 + \theta_2 u_2|^\alpha \operatorname{sign}(\theta_1 u_1 + \theta_2 u_2)\, \mu(du_1, du_2)}{\int_{\mathbf{S}} |\theta_1 u_1 + \theta_2 u_2|^\alpha\, \mu(du_1, du_2)}. \end{aligned} \tag{6.2.15}$$

If we combine (6.2.14) with (6.2.13), we get (6.1.4). If we combine (6.2.15) with (6.2.13), we get

$$\begin{aligned} &\lim_{t\to\infty} \frac{V_{(\theta_1,\theta_2)}(t) - t^{-1}\int_0^t \Delta_{(\theta_1,\theta_2)}(s)\, ds}{P(|X_1 + X_2| > t)} \\ &= c_\alpha \frac{\int_{\mathbf{S}} |\theta_1 u_1 + \theta_2 u_2|^\alpha \operatorname{sign}(\theta_1 u_1 + \theta_2 u_2)\, \mu(du_1, du_2)}{\int_{\mathbf{S}} |u_1 + u_2|^\alpha\, \mu(du_1, du_2)}. \end{aligned}$$

This, combined with (6.1.6) and Lemma 6.2.2, leads directly to (6.1.4).

B $\Rightarrow$ **C**. Clearly from (6.1.3) we have (6.2.13), hence (6.1.5). Further (6.1.3) implies that any random variable $\theta_1 X_1 + \theta_2 X_2$ is in the domain of attraction of a stable law (see Geluk and de Haan (2000), Theorem 1, part (ii). Next relation (10) of the same theorem, combined with (6.2.13), gives

$$\begin{aligned} &\lim_{t\to\infty} \frac{V_{(\theta_1,\theta_2)}(t) - \int_0^1 \Delta_{(\theta_1,\theta_2)}(st)\, ds}{P(|X_1 + X_2| > t)} \\ &= c_\alpha \frac{\int_{\mathbf{S}} |\theta_1 u_1 + \theta_2 u_2|^\alpha \operatorname{sign}(\theta_1 u_1 + \theta_2 u_2)\, \mu(du_1, du_2)}{\int_{\mathbf{S}} |u_1 + u_2|^\alpha\, \mu(du_1, du_2)}. \end{aligned}$$

This, with (6.1.4), leads directly to (6.1.6). □

Proof of Theorem 6.1.2.

A $\Rightarrow$ **B**. It follows from corresponding part of Theorem 6.1.1.

B $\Rightarrow$ **C**. Suppose $0 < \alpha < 1$. Statement **B** implies for $(\theta_1,\theta_2) \neq (0,0)$ that

$$P(|\theta_1 X_1 + \theta_2 X_2| > t) \in RV_{-\alpha} \tag{6.2.16}$$

and

$$\lim_{t\to\infty} \frac{P(|\theta_1 X_1 + \theta_2 X_2| > t)}{P(|X_1 + X_2| > t)} = \frac{\int_{\mathbf{S}} |\theta_1 u_1 + \theta_2 u_2|^\alpha \, \mu(du_1, du_2)}{\int_{\mathbf{S}} |u_1 + u_2|^\alpha \, \mu(du_1, du_2)}, \tag{6.2.17}$$

and hence

$$\begin{aligned} &\lim_{t\to\infty} \frac{P(\theta_1 X_1 + \theta_2 X_2 > t)}{P(|\theta_1 X_1 + \theta_2 X_2| > t)} \\ &= \frac{\int_{\mathbf{S}} |\theta_1 u_1 + \theta_2 u_2|^\alpha \big(1 + \operatorname{sign}(\theta_1 u_1 + \theta_2 u_2)\big) \, \mu(du_1, du_2)}{2 \int_{\mathbf{S}} |\theta_1 u_1 + \theta_2 u_2|^\alpha \, \mu(du_1, du_2)}. \end{aligned} \tag{6.2.18}$$

Relations (6.2.16) and (6.2.18) imply that any linear combination $\theta_1 X_1 + \theta_2 X_2$ with $(\theta_1, \theta_2) \neq (0, 0)$ is in the domain of attraction of a stable distribution. Hence by Theorem 1, part (iii), of Geluk and de Haan (2000) we have

$$\lim_{t\to\infty} \frac{V_{(\theta_1,\theta_2)}(t)}{1 - U_{(1,1)}(t)} = \frac{\int_{\mathbf{S}} |\theta_1 u_1 + \theta_2 u_2|^\alpha \operatorname{sign}(\theta_1 u_1 + \theta_2 u_2) \, \mu(du_1, du_2)}{\int_{\mathbf{S}} |u_1 + u_2|^\alpha \, \mu(du_1, du_2)}.$$

Also, relations (6.2.16) and (6.2.17) imply, in virtue of Lemma 6.2.2, that

$$\lim_{t\to\infty} \frac{1 - U_{(\theta_1,\theta_2)}(t)}{1 - U_{(1,1)}(t)} = \frac{\int_{\mathbf{S}} |\theta_1 u_1 + \theta_2 u_2|^\alpha \, \mu(du_1, du_2)}{\int_{\mathbf{S}} |u_1 + u_2|^\alpha \, \mu(du_1, du_2)}.$$

This completes the proof for $0 < \alpha < 1$. The case $1 < \alpha < 2$ is similar.

C $\Rightarrow$ **A**. Suppose $0 < \alpha < 1$. Define the sequence $\{a_n\}$ by

$$\lim_{n\to\infty} n\big(1 - U_{(1,1)}(a_n)\big) = \int_{\mathbf{S}} |u_1 + u_2|^\alpha \, \mu(du_1, du_2).$$

This makes sense since $1 - U_{(1,1)}(t) \in RV_{-\alpha}$. Then by statement **C**,

$$\lim_{n\to\infty} n\big(1 - U_{(\theta_1,\theta_2)}(a_n)\big) = \int_{\mathbf{S}} |\theta_1 u_1 + \theta_2 u_2|^\alpha \, \mu(du_1, du_2).$$

Further by statement **C**,

$$\lim_{n\to\infty} n V_{(\theta_1,\theta_2)}(a_n) = \tan\frac{\alpha\pi}{2} \int_{\mathbf{S}} |\theta_1 u_1 + \theta_2 u_2|^\alpha \operatorname{sign}(\theta_1 u_1 + \theta_2 u_2) \, \mu(du_1, du_2).$$

Since (6.2.2) and (6.2.3) are fulfilled, the proof is complete for $0 < \alpha < 1$. The case $1 < \alpha < 2$ is similar. □

Proof of Theorem 6.1.3.
A $\Leftrightarrow$ **C**. From the equality

$$n\big(\phi(\theta_1/a_n, \theta_2/a_n) - 1 - i\mu_1\theta_1 - i\mu_2\theta_2\big) = -q(\theta_1, \theta_2)$$

with $\mu_1 = E(X_1)$ and $\mu_2 = E(X_2)$ we get

$$\lim_{n\to\infty} n\big(1 - U_{(\theta_1,\theta_2)}(a_n)\big) = q(\theta_1,\theta_2) \tag{6.2.19}$$

and

$$\lim_{n\to\infty} n\big(V_{(\theta_1,\theta_2)}(a_n) - \mu_1\theta_2 - \mu_2\theta_2\big) = 0. \tag{6.2.20}$$

As in the proof of Theorem 6.1.1 relation (6.2.19) implies

$$\lim_{t\to\infty} \frac{1 - U_{(\theta_1,\theta_2)}(t)}{1 - U_{(1,1)}(t)} = \frac{q(\theta_1,\theta_2)}{q(1,1)}.$$

Similarly, from (6.2.19) and (6.2.20) we get

$$\lim_{t\to\infty} \frac{tV_{(\theta_1,\theta_2)}(t) - \mu_1\theta_1 - \mu_2\theta_2}{1 - U_{(1,1)}(t)} = 0 \qquad \text{for all } (\theta_1,\theta_2).$$

The converse implication is easy.

C $\Rightarrow$ **B**. The distribution of any $\theta_1X_1 + \theta_2X_2$ satisfies the conditions of part (iii) of Theorem 2 of Geluk and de Haan (2000). Relation (13) of this Theorem states that

$$\lim_{t\to\infty} \frac{\int_0^t sP(|\theta_1X_1 + \theta_2X_2| > s)\,ds}{t^2\big(1 - U_{(\theta_1,\theta_2)}(t)\big)} = 1 \qquad \text{for all } (\theta_1,\theta_2) \neq (0,0).$$

Hence, by statement **C**,

$$\lim_{t\to\infty} \frac{\int_0^t sP(|\theta_1X_1 + \theta_2X_2| > s)\,ds}{\int_0^t sP(|X_1 + X_2| > s)\,ds} = \frac{q(\theta_1,\theta_2)}{q(1,1)} \qquad \text{for all } (\theta_1,\theta_2) \neq (0,0). \tag{6.2.21}$$

B $\Rightarrow$ **C**. Condition **B** implies that each linear combination $\theta_1X_1+\theta_2X_2$ ($(\theta_1,\theta_2) \neq (0,0)$) is in the domain of attraction of a normal distribution (see Theorem 2, part (ii) of Geluk and de Haan (2000)). Hence by that theorem

$$\lim_{t\to\infty} \frac{\int_0^t sP(|\theta_1X_1 + \theta_2X_2| > s)\,ds}{t^2\big(1 - U_{(\theta_1,\theta_2)}(t)\big)} = 1$$

and

$$\lim_{t\to\infty} \frac{tV_{(\theta_1,\theta_2)}(t) - \mu_1\theta_1 - \mu_2\theta_2}{t\big(1 - U_{(\theta_1,\theta_2)}(t)\big)} = 0.$$

These two relations in combination with **B** imply **C**. □

Proof of Remark 6.1.3.

Relation (6.1.7) implies that

$$\int_0^t sP(|\theta_1 X_1 + \theta_2 X_2| > s)\, ds = \frac{1}{2}\int_0^{t^2} P((\theta_1 X_1 + \theta_2 X_2)^2 > s)\, ds$$

is slowly varying.

Now, for any probability distribution function G the slow variation of $\int_0^t (1 - G(s))\, ds$ is equivalent to

$$t(1 - G(t)) / \int_0^t (1 - G(s))\, ds \to 0,$$

as $t \to \infty$, since on the one hand for any $0 < x < 1$

$$\frac{t(1 - G(t))}{\int_0^t (1 - G(s))\, ds} \le \frac{1}{1-x} \frac{\int_{tx}^t (1 - G(s))\, ds}{\int_0^t (1 - G(s))\, ds}$$

and on the other hand for any $0 < x < 1$

$$\log \int_0^t (1 - G(s))\, ds - \log \int_0^{tx} (1 - G(s))\, ds = \int_{tx}^t \Big(\frac{s(1 - G(s))}{\int_0^s (1 - G(u))\, du} \Big) \frac{ds}{s}.$$

Hence, since

$$\begin{aligned} &\int_0^t sP(|\theta_1 X_1 + \theta_2 X_2| > s)\, ds \\ &= \frac{1}{2} t^2 P(|\theta_1 X_1 + \theta_2 X_2| > t) + E|\theta_1 X_1 + \theta_2 X_2| I(|\theta_1 X_1 + \theta_2 X_2| \le t), \end{aligned}$$

by the result just proved, (6.1.7) and (6.1.10) are equivalent. □

6.3 Rvaceva's results

In this section we give a direct proof of the implication: Rvaceva's condition (i.e. (6.1.11)) for $0 < \alpha < 2$ implies our condition (i.e. (6.1.4)). We have not been able to prove the converse implication for $\alpha = 1$. For $\alpha \neq 1$ the implication follows from the work of Basrak, Davis and Mikosch (2000). For completeness we include a proof of the necessity of Rvaceva's condition based on Feller's proof (Feller(1971)) for one-dimensional case.

Proof of (6.1.11) $\Rightarrow$ (6.1.4). For $\theta_1^2 + \theta_2^2 = 1$, by Rvaceva's condition (6.1.11), we have

$$\begin{aligned}
&\frac{E(\theta_1 X_1+\theta_2 X_2)I(|\theta_1 X_1+\theta_2 X_2|\le t, X_1^2+X_2^2>t^2)}{P(|X_1+X_2|>t)}\\
&=\int_0^1 \frac{P(ts<\theta_1 X_1+\theta_2 X_2<t, X_1^2+X_2^2>t^2)}{P(|X_1+X_2|>t)}\,ds\\
&\quad-\int_{-1}^0 \frac{P(-t<\theta_1 X_1+\theta_2 X_2<ts, X_1^2+X_2^2>t)}{P(|X_1+X_2|>t)}\,ds \qquad (6.3.1)\\
&\to \int_0^1 \nu\{(x_1,x_2): s<\theta_1 x_1+\theta_2 x_2<1, x_1^2+x_2^2>1\}\,ds\\
&\quad-\int_{-1}^0 \nu\{(x_1,x_2): -1<\theta_1 x_1+\theta_2 x_2<s, x_1^2+x_2^2>1\}\,ds,
\end{aligned}$$

where ν is defined by

$$\nu\{(x_1,x_2): x_1^2+x_2^2>y^2, \arctan(x_2/x_1)\in A\}=\alpha^{-1}y^{-\alpha}\mu(A)$$

for $y>0$ and any continuity set A of μ. The right hand side of (6.3.1) equals

$$\begin{aligned}
&\int_{R^2}(\theta_1 x_1+\theta_2 x_2)I(|\theta_1 x_1+\theta_2 x_2|\le 1, x_1^2+x_2^2>1)\,\nu(dx_1,dx_2)\\
&=\int_S(\theta_1 u_1+\theta_2 u_2)\int_{1<r<1/|\theta_1 u_1+\theta_2 u_2|} r^{-\alpha}\,dr\mu(du_1,du_2)\\
&=-\int_S(\theta_1 u_1+\theta_2 u_2)\frac{|\theta_1 u_1+\theta_2 u_2|^{\alpha-1}-1}{\alpha-1}\,\mu(du_1,du_2).
\end{aligned}$$

Now we can proceed to prove (6.1.4). Since $|\theta_1 X_1+\theta_2 X_2|<t$ implies $X_1^2+X_2^2<t^2$, we have

$$\begin{aligned}
&E(\theta_1 X_1+\theta_2 X_2)I(|\theta_1 X_1+\theta_2 X_2|\le t)\\
&\quad-E\theta_1 X_1 I(|\theta_1 X_1|\le t)-E\theta_2 X_2 I(|\theta_2 X_2|\le t)\\
&=E(\theta_1 X_1+\theta_2 X_2)I(X_1^2+X_2^2\le t^2)\\
&\quad+E(\theta_1 X_1+\theta_2 X_2)I(|\theta_1 X_1+\theta_2 X_2|\le t, X_1^2+X_2^2>t^2)\\
&\quad-E\theta_1 X_1 I(X_1^2+X_2^2\le t^2)+E\theta_1 X_1 I(|\theta_1 X_1|\le t, X_1^2+X_2^2>t^2)\\
&\quad-E\theta_2 X_2 I(X_1^2+X_2^2\le t^2)+E\theta_2 X_2 I(|\theta_2 X_2|\le t, X_1^2+X_2^2>t^2)\\
&=E(\theta_1 X_1+\theta_2 X_2)I(|\theta_1 X_1+\theta_2 X_2|\le t, X_1^2+X_2^2>t^2)\\
&\quad-E\theta_1 X_1 I(|\theta_1 X_1|\le t, X_1^2+X_2^2>t^2)-E\theta_2 X_2 I(|\theta_2 X_2|\le t, X_1^2+X_2^2>t^2).
\end{aligned}$$

If we divide this expression by $P(|X_1+X_2|>t)$, it converges, by the result just proved, to

$$\begin{aligned}
&-\int_{\mathbf{S}}(\theta_1 u_1+\theta_2 u_2)\frac{|\theta_1 u_1+\theta_2 u_2|^{\alpha-1}-1}{\alpha-1}\,\mu(du_1,du_2)\\
&\quad-\int_{\mathbf{S}}\theta_1 u_1\frac{|\theta_1 u_1|^{\alpha-1}-1}{\alpha-1}\,\mu(du_1,du_2)-\int_{\mathbf{S}}\theta_2 u_2\frac{|\theta_2 u_2|^{\alpha-1}-1}{\alpha-1}\,\mu(du_1,du_2),
\end{aligned}$$

which is equivalent to (6.1.4) (see Remark 6.1.1). □

For completeness we add a proof of the implication: the statement of Theorem 6.1.1 **A** implies (6.1.11) (Rvaceva's condition), following the line of reasoning of Feller (1971, Ch. XVII). We start from

$$\begin{aligned}&\lim_{n\to\infty} n\big(\phi(\theta_1/a_n,\theta_2/a_n)-1-i\theta_1\mathrm{Im}\phi(1/a_n,0)-i\theta_2\mathrm{Im}\phi(0,1/a_n)\big)\\&=\log\psi(\theta_1,\theta_2),\end{aligned}\tag{6.3.2}$$

locally uniformly. Denote the left-hand side by $\psi_n(\theta_1,\theta_2)$ and define

$$\psi_n^*(\theta_1,\theta_2)=\psi_n(\theta_1,\theta_2)-\frac{1}{4}\iint_{|s_1|<1,|s_2|<1}\psi_n(\theta_1+s_1,\theta_2+s_2)\,ds_1ds_2.$$

An easy calculation shows that

$$\psi_n^*(\theta_1,\theta_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{i(\theta_1x_1+\theta_2x_2)}(x_1^2+x_2^2)K(x_1,x_2)n\,F(a_ndx_1,a_ndx_2)$$

with

$$K(x_1,x_2)=\frac{1-\frac{\sin x_1}{x_1}\frac{\sin x_2}{x_2}}{x_1^2+x_2^2}.$$

Note that $\lim_{x_1,x_2\to 0}K(x_1,x_2)=1/6$ and $\lim_{x_1^2+x_2^2\to\infty}(x_1^2+x_2^2)K(x_1,x_2)=1$. Relation (6.3.2) implies

$$\begin{aligned}&\lim_{n\to\infty}\psi_n^*(\theta_1,\theta_2)\\&=\log\psi(\theta_1,\theta_2)-\frac{1}{4}\iint_{|s_1|<1,|s_2|<1}\log\psi(\theta_1+s_1,\theta_2+s_2)\,ds_1ds_2\end{aligned}\tag{6.3.3}$$

locally uniformly. Relation (6.3.3) for $\theta_1=\theta_2=0$ implies that $\lim_{n\to\infty}M_n^*(\mathcal{R}^2)$ exists. Define

$$M_n^*(dx_1,dx_2)=n(x_1^2+x_2^2)K(x_1,x_2)\,F(a_ndx_1,a_ndx_2).$$

By the continuity theorem for characteristic function the sequence of probability distributions $M_n^*/M_n^*(\mathcal{R}^2)$ converges in distribution to some probability distribution. It follows the two properties of K that

$$\lim_{n\to\infty}nE(X_1^2+X_2^2)I(X_1^2+X_2^2\le a_nx)$$

exists for all $x>0$ and that

$$\lim_{n\to\infty}nP((X_1,X_2)\in a_nA_{x_1,x_2})$$

converges for all but denumerably many real $(x_1,x_2)\neq(0,0)$ with $A_{x_1,x_2}:=\{(ax_1,bx_2):a,b>1\}$. The latter condition is easily seen to imply Rvaceva's condition (6.1.11).

Bibliography

[1] Basrak, B., Davis, R.A. and Mikosch, T. (2002). A characterization of multivariate regular variation. *Ann. Appl. Probab., Vol.* **29** *No. 2, 766-801.*

[2] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[3] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation.* Cambridge University Press, Cambridge.

[4] Cheng, S. and Jiang, C. (2001). The Edgeworth expansion for distributions of extreme values. *Sci. China* **44**, 427-437.

[5] Coles, S.G., Heffernan, J. and Tawn, J.A. (1999). Dependence measures for extreme value analyses. *Extremes* **2**, *339-365.*

[6] Coles, S.G. and Tawn, J.A. (1994). Statistical methods for multivariate extremes: an application to structal design(with discussion). *Applied Statistics* **43**, *1-48.*

[7] Csörgő, M. and Horváth, L. (1993). *Weighted approximations in probability and statistics.* Wiley, Chichester.

[8] Dietrich, D., de Haan, L. and Hüsler, J. (2002). Testing extreme value conditions. *Extremes* **5**, 71–85.

[9] Drees, H. (1998). On smooth statistical tail functionals. *Scand. J. Statist.* **25**, 187-210.

[10] Drees, H., Ferreira, A. and de Haan, L. (2004). On maximum likelihood estimation of the extreme value index. To appear in *Ann. Appl. Probab.*

[11] Drees, H., de Haan, L. and Li, D. (2003). On large deviations for extremes. *Statist. Probab. Letters*, **64**, 51–62.

[12] Drees, H., de Haan, L. and Li, D. (2004). Approximations to the tail empirical distribution function with application to testing extreme value conditions. Submitted.

[13] Drees, H. and Huang, X. (1998). Best attainable rates of convergence for estimators of the stable tail dependence function. *Journal of Multivariate Analysis* **64**, *25-47.*

[14] Einmahl, J. (1992). Limit theorems for tail processes with application to intermediate quantile estimation. *J. Statist. Plann. Inference*, **32**, 137-145.

[15] Einmahl, J.H.J. (1997). Poisson and Gaussian approximation of weighted local empirical processes. *Stoch. Proc. Appl.* **70**, 31-58.

[16] Einmahl, J.H.J., de Haan, L. and Piterbarg, V.I. (2001). Nonparametric estimation of the spetral measure of an extreme value distribution. *The Annals of Statistics, Vol. 29, No.* **5**, *1401-1423.*

[17] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications, Volume II.* John Wiley & Sons, Inc.

[18] Geluk, J.G. and de Haan, L. (2000). Stable probability distributions and their domains of attraction: a direct approach. *Probability and Mathematical Statistics,* **20***(1), 169 - 188.*

[19] Grimshaw, S.D. (1993). Computing maximum likelihood estimates for the generalized pareto distribution. *Technometrics* **35**, 185–191.

[20] de Haan, L., Li, D., Peng, L. and Pereira, H.I. (2002). Alternative conditions for attraction to stable vectors. *Prob. Math. Stat.* **22***(2), 303-317.*

[21] de Haan, L. and Peng, L. (1999). Exact rates of convergence to a stable law. *J. London Math. Soc. (2),* **59**, *1134 - 1152.*

[22] de Haan, L. and Resnick, S. (1996). Second order regular variation and rates of convergence in extreme value theory. *Ann. Prob.* **24**, 97–124.

[23] de Haan, L. and Rootzén, H. (1993). On the estimation of high quantiles. *J. Statist. Planning Inference* **35**, 1-13.

[24] de Haan, L. and Sinha, A.K. (1999). Estimating the probability of a rare event. *Ann. Statist.*, **27**, 732-759.

[25] de Haan, L. and Stadtmüller, U. (1996). Generalized regular variation of second order. *J. Austral. Math. Soc. Ser. A* **61**, 381-395.

[26] Huang, X. (1992). *Statistics of bivariate extremes.* Thesis, Erasmus University Rotterdam. Tinbergen Institute Series no. **22**.

[27] Neuhaus, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. *Ann. Math. Statist.*, **42**, 1285-1295.

[28] Orey, S. and Pruitt, W. (1973). Sample functions of the N-parameter Wiener process. *Ann. Probab.*, **1**, 138-163.

[29] Paulauskas, V. and Račkauskas, A. (1989). *Approximation Theory in the Central Limit Theorem. Exact Results in Banach Spaces.* Kluwer, Dordrecht.

[30] Pickands, J. (1981). Multivariate extreme value distributions. *Bull. Int. Statist. Inst. 859-878.*

[31] Rvaceva, E.L. (1962). On domains of attraction of multi-dimensional distributions. *Selected Translations in Mathematical Statistics and Probability,* **2**, *183 - 205.*

[32] Resnick, S.I. (1987). *Extreme Values, Regular Variation, and Point Processes.* Springer, New York.

[33] Samorodnitsky, G. and Taqqu, M.S. (1994). *Stable Non-Gaussian Random Processes.* Chapman & Hall.

[34] Schlather, M. (2001). Examples for the coefficient of tail dependence and the domain of attraction of a bivariate extreme value distribution. *Statist. Probab. Lett.*, **53**, 325-329.

[35] Shorack, G. and Wellner, J. (1986). *Empirical Processes with Applications to Statistics*, Wiley, New York.

[36] Smith, R.L. (1987). Estimating tails of probability distributions. *Ann. Statist.* **15**, 1174-1207.

[37] Tawn, J.A. (1988). Bivariate extreme value theory: model and estimation. *Biometrika* **75**, *397-415.*

[38] van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*, Springer, New York.

[39] Vervaat, W. (1972). Functional central limit theorems for processes with positive drift and their inverses. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **23**, 245–253.

Nederlandse samenvatting

Het onderwerp van dit proefschrift is extreme waarden theorie, het onderdeel van de statistiek dat gaat over het schatten van de kans op een zeldzame gebeurtenis, één die bijvoorbeeld sinds mensenheugenis niet heeft plaatsgevonden. Het schatten van zo'n zeldzame gebeurtenis is alleen mogelijk onder bepaalde zwakke voorwaarden op het onderliggende kansmodel. Ofschoon het schatten van zo'n zeldzame gebeurtenis eigenlijk alleen zinvol is onder deze voorwaarden (dat wil zeggen dat de voorwaarden eigenlijk onvermijdelijk zijn), is het toch nuttig om na te gaan of in de praktijk aan deze voorwaarden voldaan is.

Een dergelijke toetsing van de voorwaarden wordt behandeld in de hoofdstukken 2 en 4 van het proefschrift. De afleiding van de toets geeft bovendien aanleiding tot het ontwikkelen van enige interessante theoretische resultaten over de rand van de empirische verdelingsfunktie. Een afsplitsing van enige van deze resultaten is te vinden in hoofdstuk 3.

Dan zijn er nog twee hoofdstukken die enigszins los staan van de andere, één over het vergelijken van schatters in meerdimensionale extreme waarden theorie en één over voorwaarden voor de zwakke convergentie van partiële sommen van onafhankelijke en gelijk verdeelde stochastische grootheden.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

291. P. DE BRUIN, *Essays on modeling nonlinear time series.*
292. J. ARDTS, *All is well that begins well: A longitudinal study of organisational socialisation.*
293. J.E.M. VAN NIEROP, *Advanced choice models.*
294. D.J. VAN VUUREN, *The market for passenger transport by train. An empirical analysis.*
295. A. FERRER CARBONELL, *Quantitative analysis of well-being with economic applications.*
296. L.M. VINHAS DE SOUZA, *Beyond transition: Essays on the monetary integration of the accession countries in Eastern Europe.*
297. J. LEVIN, *Essays in the economics of education.*
298. E. WIERSMA, *Non-financial performance measures: An empirical analysis of a change in a firm's performance measurement system.*
299. M. MEKONNEN AKALU, *Projects for shareholder value: A capital budgeting perspective.*
300. S. ROSSETTO, *Optimal timing of strategic financial decisions.*
301. P.W. VAN FOREEST, *Essays in financial economics.*
302. A. SIEGMANN, *Optimal financial decision making under loss averse preferences.*
303. A. VAN DER HORST, *Government interference in a dynamic economy.*
304. A.P. RUSSO, *The sustainable development of heritage cities and their regions: Analysis, policy, governance.*
305. G.A.W. GRIFFIOEN, *Technical analysis in financial markets.*
306. I.S. LAMMERS, *In conflict, een geschiedenis van kennismanagement.*
307. O.L. LISTES, *Stochastic programming approaches for strategic logistics problems.*
308. A.T. DE BLAEIJ, *The value of a statistical life in road safety.*
309. S.H.K. WUYTS, *Partner selection in business markets. A structural embeddedness perspective.*
310. H.C. DEKKER, *Control of inter-organizational relationships: The effects of appropriation concerns, coordination requirements and social embeddedness.*
311. I.V. OSSOKINA, *Environmental policy and environment-saving technologies. Economic aspects of policy making under uncertainty.*
312. D. BROUNEN, *Real estate securitization and corporate strategy: From bricks to bucks.*
313. J.M.P. DE KOK, *Human resource management within small and medium-sized enterprises.*
314. T. VERHAGEN, *Towards understanding online purchase behavior.*
315. R. HOEKSTRA, *Structural change of the physical economy. Decomposition analysis of physical and hybrid-units input-output tables.*
316. R.K. AIDIS, *By law and by custom: Factors affecting small and medium*

sized enterprises during the transition in Lithuania.

317. S. MANZAN, *Essays in nonlinear economic dynamics.*
318. K. OLTMER, *Agricultural policy, land use and environmental effects: Studies in quantitative research synthesis.*
319. H.J. HORBEEK, *The elastic workfloor. About the implementation of internal flexibility arrangements.*
320. P. HOUWELING, *Empirical studies on credit markets.*
321. E. MENDYS, *Essays on network economics.*
322. J. NOAILLY, *Coevolutionary modeling for sustainable economic development.*
323. Y.V. VELD-MERKOULOVA, *Essays on futures markets and corporate spin-offs.*
324. F.J. WILLE, *Auditing using Bayesian decision analysis.*
325. D. HUISMAN, *Integrated and dynamic vehicle and crew scheduling*.
326. S. WANG, *Global climate change policies: An analysis of CDM policies with an adapted GTAP model.*
327. M.W. VAN GELDEREN, *Opportunity entry performance. Studies of entrepreneurship and small business*.
328. W. VAN WINDEN, *Essays on urban ICT policies.*
329. G.J. KULA, *Optimal retirement decision*.
330. R.J. IMESON, *Economic analysis and modeling of fisheries management in complex marine ecosystems.*
331. M. DEKKER, *Risk, resettlement and relations: Social security in rural Zimbabwe.*
332. A. MULATU, *Relative stringency of environmental regulation and international competitiveness.*
333. C.M. VAN VEELEN, *Survival of the fair: Modelling the evolution of altruism, fairness and morality.*
334. R. PHISALAPHONG, *The impact of economic integration programs on inward foreign direct investment.*
335. A.H. NÖTEBERG, *The medium matters: The impact of electronic communication media and evidence strength on belief revision during auditor-client inquiry.*
336. M. MASTROGIACOMO, *Retirement, expectations and realizations. Essays on the Netherlands and Italy.*
337. E. KENJOH, *Balancing work and family life in Japan and four European countries: Econometric analyses on mothers' employment and timing of maternity.*
338. A.H. BRUMMANS, *Adoption and diffusion of EDI in multilateral networks of organizations.*
339. K. STAAL, *Voting, public goods and violence.*
340. R.H.J. MOSCH, *The economic effects of trust. Theory and empirical evidence.*
341. F. ESCHENBACH, *The impact of banks and asset markets on economic growth and fiscal stability.*