Submitted to the Bernoulli
arXiv: arXiv:1107.0614

Estimating Failure Probabilities

HOLGER DREES[1] and LAURENS DE HAAN[2]

[1] *University of Hamburg, Department of Mathematics, Bundesstr. 55, 20146 Hamburg, Germany. E-mail:* holger.drees@math.uni-hamburg.de

[2] *Erasmus University Rotterdam, Department of Economics, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands. E-mail:* ldehaan@ese.eur.nl

In risk management often the probability must be estimated that a random vector falls into an extreme failure set. In the framework of bivariate extreme value theory, we construct an estimator for such failure probabilities and analyze its asymptotic properties under natural conditions. It turns out that the estimation error is mainly determined by the accuracy of the statistical analysis of the marginal distributions if the extreme value approximation to the dependence structure is at least as accurate as the generalized Pareto approximation to the marginal distributions. Moreover, we establish confidence intervals and briefly discuss generalizations to higher dimensions and issues arising in practical applications as well.

Keywords: asymptotic normality, exceedance probability, failure set, homogeneity, multivariate extremes, out of sample extrapolation, peaks over threshold.
Primary 62G32; secondary 62H12.

1. Introduction

1.1. Motivation

Suppose an insurance company has contracts in two related lines of business with all customers of an insurance portfolio (e.g., fire insurance and business interruption insurance for industrial customers). On top of quota reinsurances for both lines of business (possibly with different quotas) the remaining total loss from each incidence is covered by an excess of loss reinsurance (CAT-XL) that pays for the part of the total loss which exceeds a given high retention level R. If X and Y denote the original losses from a fire in both lines of business and $1-\alpha_X$ and $1-\alpha_Y$ the corresponding quotas, then a claim occurs in the XL-reinsurance if $\alpha_X X + \alpha_Y Y$ exceeds R. For the purpose of risk management the reinsurer might be interested in the probability that the insurance company will file a claim in case of a fire. If the retention level is high, then the claim probability cannot be estimated using simple empirical estimates, because in the past the retention has rarely (or never) been exceeded.

In this paper a more general setting is considered. We are interested in estimating the probability that a pair of random variables (X, Y) will take on a value in some given "extreme" set. Similar problems arise naturally in many fields. For example, a coastal dike may fail if the vector build from the still water level and the wave heights lie in a certain failure set D (cf. Coles and Tawn, 1994, Bruun and Tawn, 1998, and de Haan and de Ronde, 1998). A financial option (like a down-and-out-put) may become worthless if the price vector of underlyings enters such a “failure set”. Finally, (part of) the principal of a catastrophe bond gets lost for the investors if a vector of triggers becomes too extreme.

As there are insufficiently many observations available in the extreme failure set D to use standard statistical methods, extreme value theory is needed to estimate the failure probability $P\{(X, Y) \in D\}$.

1.2. Extreme value approximations

The basic idea of multivariate extreme value theory is to assume that the suitably standardized componentwise maxima of the observed random vectors converge to a non-degenerate limit distribution. This assumption is equivalent to the convergence of suitably standardized quantile functions of all marginal distributions and a condition on the dependence structure in extreme regions.

To be more precise, denote the marginal distribution functions of X and Y by F_1 and F_2, respectively, and let $U_i(t) := F_i^{\leftarrow}(1 - 1/t)$ with $H^{\leftarrow}(s) := \inf\{x \in \mathbb{R} \mid H(x) \geq s\}$ denoting the generalized inverse of an increasing function H. We assume that there exist real constants γ_i, positive functions a_i and real functions b_i such that for $x > 0$ and $i \in \{1, 2\}$

$$\lim_{t \to \infty} \frac{U_i(tx) - b_i(t)}{a_i(t)} = \frac{x^{\gamma_i} - 1}{\gamma_i}. \tag{1.1}$$

For $\gamma_i = 0$ read the right-hand side as $\log x$. Note that the right-hand side is the U-function of the generalized Pareto distribution (GPD) with distribution function $1 - (1 + \gamma_i x)^{-1/\gamma_i}$ for $1 + \gamma_i x > 0$, that is to be interpreted as the standard exponential distribution function for $\gamma_i = 0$. The parameter γ_i is the so-called extreme value index of the ith marginal. If it is positive, then the support of F_i is unbounded from above and $1 - F_i(t)$ roughly decays like the power function with exponent $1/\gamma_i$, while for $\gamma_i < 0$ the right endpoint $x_i^* := F_i^{\leftarrow}(1)$ of the support is finite and $1 - F_i(x)$ roughly behaves like a multiple of $(x_i^* - x)^{-1/\gamma_i}$ as $x \uparrow x_i^*$.

The aforementioned extremal dependence condition can be given in terms of the standardized random variables $1 - F_1(X)$ and $1 - F_2(Y)$, that are uniformly distributed on $[0, 1]$ if the marginal distributions are continuous. More precisely, we assume the existence of a measure ν such that for ν-continuous Borel sets $B \subset [0, \infty)^2$ bounded away from the origin

$$\lim_{t \to \infty} tP\{(X, Y) \in U(tB)\} = \nu(B). \tag{1.2}$$

Here and in what follows, for functions h_1, h_2 which are defined on subsets of the reals, we define a function h on a subset of $\mathbb{R}^2$ by $h(x_1, x_2) := (h_1(x_1), h_2(x_2))$. The so-called exponent measure ν describes the asymptotic dependence structure between extreme observations X and Y. Its homogeneity property

$$\nu(tB) = t^{-1}\nu(B), \tag{1.3}$$

which holds for all Borel sets $B \subset [0, \infty)^2$ and all $t > 0$, will be pivotal for the construction of our estimator of the failure probability. (Seen from a different angle, we assume an approximate scaling law for the joint distribution of $U^{\leftarrow}(X, Y)$; cf. Anderson (1994).) In addition, we need certain smoothness assumptions to ensure that ν does not have mass on the coordinate axes and not too much mass in their neighborhoods (cf. condition (D2) in Section 2.2). Further details about the extreme value assumptions can be found in de Haan and Ferreira (2006), Sections 1.2 and 6.1, or Beirlant et al. (2004), Chapters 2 and 8.

1.3. Construction of estimators of extreme failure probabilities

We are interested in the situation that at most a few observations lie in the extreme failure set D which implies that in our mathematical framework the failure set $D = D_n$ must depend on the sample size n such that the failure probability

$$p_n := P\{(X, Y) \in D_n\}$$

tends to 0. To motivate an estimator of p_n based on independent copies (X_i, Y_i), $1 \le i \le n$, of (X, Y) first note that from (1.2) we obtain the approximation

$$\frac{n}{k} P\Big\{\frac{k}{n} U^{\leftarrow}(X, Y) \in B\Big\} \approx \nu(B) \tag{1.4}$$

for any sequence $k = k_n \to \infty$ such that $k/n \to 0$. To estimate p_n using this approximation, we must replace $U^{\leftarrow}$ and ν with suitable estimators.

According to (1.1), we may approximate $U_i((n/k)x)$ for sufficiently large n by

$$T_{n,i}(x) := a_i(n/k)\frac{x^{\gamma_i} - 1}{\gamma_i} + b_i(n/k) \tag{1.5}$$

and estimate it by

$$\hat{T}_{n,i}(x) := \hat{a}_i(n/k)\frac{x^{\hat{\gamma}_i} - 1}{\hat{\gamma}_i} + \hat{b}_i(n/k), \tag{1.6}$$

where $\hat{a}_i(n/k), \hat{b}_i(n/k)$ and $\hat{\gamma}_i$ are suitable estimators for $a_i(n/k), b_i(n/k)$ and γ_i, respectively. Likewise, the generalized inverse functions $(k/n)U_i^{\leftarrow}(x)$ can be estimated by

$$\hat{T}_{n,i}^{\leftarrow}(x) := \Big(1 + \hat{\gamma}_i \frac{x - \hat{b}_i(n/k)}{\hat{a}_i(n/k)}\Big)^{1/\hat{\gamma}_i}. \tag{1.7}$$

Here and in the sequel, $(1 + \gamma y)^{1/\gamma}$ is defined as e^y if $\gamma = 0$. For $1 + \gamma y < 0$ (or $1 + \gamma y = 0$ and $\gamma < 0$) the term $(1 + \gamma y)^{1/\gamma}$ is not well defined. If γ is positive and $y < -1/\gamma$, then it may be interpreted as 0, while for $\gamma < 0$ and $y > -1/\gamma$ it may be defined to be ∞. However, we will see that the precise definition of $(1 + \gamma y)^{1/\gamma}$ for very small and for negative values of $1 + \gamma y$ is not important in the present setting (provided it is taken to be a non-decreasing function of y), because the sets on which $\hat{T}^{\leftarrow}_{n,i}$, $i \in \{1,2\}$, are not well defined are asymptotically negligible.

If, in (1.4), we substitute $\hat{T}^{\leftarrow}_n(x_1,x_2) := \big(\hat{T}^{\leftarrow}_{n,1}(x_1), \hat{T}^{\leftarrow}_{n,2}(x_2)\big)$ for the marginal transformation $(k/n)U^{\leftarrow}$ and replace the probability in the left-hand side of (1.4) by its empirical counterpart, we arrive at the following estimator of ν

$$\hat{\nu}_n(B) := \frac{1}{k}\sum_{i=1}^n \varepsilon_{\hat{T}^{\leftarrow}_n(X_i,Y_i)}(B), \tag{1.8}$$

with ε_x denoting the Dirac measure with mass 1 at x.

Now, again interpreting convergence (1.2) (for $t = e_n$) as an approximation, we may estimate the failure probability as follows:

$$\begin{aligned} p_n &= P\{(X,Y)\in D_n\} \\ &= P\big\{(X,Y)\in U(e_n\cdot e_n^{-1}U^{\leftarrow}(D_n))\big\} \\ &\approx \frac{1}{e_n}\nu\big(e_n^{-1}U^{\leftarrow}(D_n)\big) \qquad (1.9)\\ &\approx \frac{1}{e_n}\nu\Big(\frac{n}{ke_n}\hat{T}^{\leftarrow}_n(D_n)\Big) \\ &\approx \frac{1}{e_n}\hat{\nu}_n\Big(\frac{n}{ke_n}\hat{T}^{\leftarrow}_n(D_n)\Big) \\ &=: \hat{p}_n. \qquad (1.10) \end{aligned}$$

The basic idea of this estimator is to blow up the failure set, after a standardization of the marginals, such that it contains sufficiently many observations to allow the estimation of its probability by an empirical probability. Note that, for given marginal transformations, the estimator $\hat{p}_n$ depends on the tuning parameters k and e_n only via their product ke_n, which controls the factor by which the transformed failure set is blown up; see Subsection 2.5 for a detailed discussion. This factor should be chosen by the statistician such that two contrary effects are balanced. On the one hand, ke_n must not be too small, such that the inflated standardized failure set $n/(ke_n)\hat{T}^{\leftarrow}_n(D_n)$ contains sufficiently many marginally transformed observations $\hat{T}^{\leftarrow}_n(X_i,Y_i)$, and thus the empirical probability $\hat{\nu}_n\big(n/(ke_n)\hat{T}^{\leftarrow}_n(D_n)\big)$ is an accurate estimate of its expectation. On the other hand, the set $e_n^{-1}U^{\leftarrow}(D_n)$ must be sufficiently extreme to justify approximation (1.9). In Subsection 2.5 we discuss a heuristic tool to ensure this balance.

1.4. Alternative approaches

An estimator related to $\hat{p}_n$ has been suggested and analyzed by de Haan and Sinha (1999) in a much more restrictive framework. In particular, specific estimators for the marginal parameters have been considered which use the same number k_n of largest order statistics for both marginal fits, which is inefficient if the GPD approximation (cf. (2.6) below) is less accurate for one of the marginal distributions. Likewise, the flexibility of the estimator is increased in the present paper by allowing that the blow-up factor e_n deviates from the unknown model constant d_n defined below, while de Haan and Sinha (1999) used a consistent estimator of d_n that was made identifiable in a quite arbitrary way by fixing some point on the boundary of some set S, which together with the factor d_n determines the failure set D_n (see (2.1)). In our simulation study it turns out that the inferior performance of the estimator proposed by de Haan and Sinha is mainly caused by this often inappropriate choice of e_n.

Moreover, the shape of the failure set considered by de Haan and Sinha is restricted. E.g. the case $q(\infty) = 0$ (in our notation; cf. condition (Q2) below) is ruled out by condition (2.9) of that paper. The model assumption

$$D_n := \{(s,t) \mid f(s/x_n, t/y_n) \geq 1\}$$

for some function f and sequences of normalizing constants x_n and y_n seems quite restrictive and unnatural, because it allows the failure set to tend towards the "north-east" only by a linear scaling of both marginals. This parametrization does not fit well to extreme value theory if the extreme value indices are not positive, which is usually the case in environmetrics, one of the most important fields of application of our theory besides financial risk management.

Even more troublesome is the fact that by assumption (1.5) of de Haan and Sinha (1999) the failure set is described in terms of the number k_n of largest order statistics that is picked by the statistician. Hence the model parametrization depends on the statistical procedure used to analyze the model, which makes it extremely difficult to interpret.

Finally, while the influence of each marginal transformation is clearly separated in the description of the limiting distribution in our main Theorem 2.1, in Theorem 4.1 of de Haan and Sinha (1999) the marginal parameters are seemingly intermingled. Therefore, the generalization of the present results to higher dimension is much more straightforward than those of de Haan and Sinha (see the discussion in Subsection 2.6).

An alternative to our genuinely multivariate estimator can be constructed by the so-called structural variable approach if the failure set is of the form $D_n = \{(s,t) \mid h(s,t) \geq t_n\}$ for some known function h and threshold t_n. Then one may apply techniques from univariate extreme value theory to the pseudo-observations $h(X_i, Y_i)$, $1 \leq i \leq n$ (cf. Coles (2001), §8.2.4 and page 156, or Bruun and Tawn (1998)). However, even for this class of failure sets, an analysis of the dependence structure between the two components of the observed vectors is of independent interest, and it seems more natural to use the same approach for model fitting and for the estimation of quantities like failure probabilities. Moreover, often one wants to estimate the failure probability for several different sets (e.g., to find the cheapest construction to ensure a certain level of safety); in this case it is both more efficient and more natural to use estimators in a unified framework as considered in the present paper.

In the multivariate approach, Coles and Tawn (1994) and Bruun and Tawn (1998) used parametric models for the dependence structure in the closely related problem to estimate a parameter defining a failure set such that the corresponding failure probability equals a given value. However, usually there is no physical reason for such parametric models.

By using them nevertheless, one trades a modeling error, which is difficult to assess, for an estimation error, which can be quantified at least asymptotically (see Theorem 2.1 below). Having said this, it may be sensible to use a parametric estimator of the failure probability if experience strongly suggest that a simple model describes the data well. In that case, our approach may be used as a countercheck of the model assumptions.

Note that our assumptions rule out that the exponent measure ν puts mass on the coordinate axes. In particular, X and Y are assumed asymptotically dependent in the sense of multivariate extreme value theory in that $\lim_{t\to\infty} P(X > U_1(t) \mid Y > U_2(t)) > 0$. In the case of asymptotic independent coordinates X and Y, consistency of an analogous estimator for the failure probability was proved by Draisma et al. (2004), while its asymptotic normality was established by Müller (2008).

1.5. Outline

The paper is organized as follows: In Section 2 we first introduce and discuss in detail the framework in which we then prove asymptotic normality of our estimator of the failure probability. Moreover, we propose a consistent estimator of the limiting variance, derive an asymptotic confidence interval, discuss the role of ke_n and propose a heuristic approach for choosing this factor. In Section 3 we apply the theory to the motivating example given at the beginning, while the finite sample performance of the estimator is investigated in Section 4. All proofs are collected in Section 5.

2. Main results

2.1. Analysis of the estimation error

The main goal of the present paper is to establish the asymptotic normality of the estimator $\hat{p}_n$ under conditions on the underlying distribution and the failure set which are easy to interpret and relatively simple to verify. To achieve this objective, we first decompose the estimation error into 6 parts. Loosely speaking, the one that usually dominates the others (term IV in equation (2.5) below) is due to the marginal fitting, two terms (II and III) are related to the bias and the random error of the estimator of the exponent measure, respectively, term VI stems from the approximation error in (1.4), while the remaining two are related to a technical truncation argument.

To derive this decomposition, recall that, in our asymptotic framework, the failure set D_n must become more extreme in the sense that it moves in the north-east direction as the sample size n increases to ensure that it contains at most a few observations. To make both coordinates comparable, we standardize the marginals using $U^{\leftarrow}$ and assume that $U^{\leftarrow}(D_n)$ is essentially an increasing multiple of a fixed set S. That way we ensure that none of the coordinates dominates the other. More precisely, we assume that for different sample sizes the failure sets are of the type

$$D_n = U(d_n S) \cap \mathbb{R}^2 = \{(U_1(d_n x), U_2(d_n y)) \mid (x, y) \in S\} \cap \mathbb{R}^2 \tag{2.1}$$

for a fixed set $S \subset [0,\infty)^2$ and constants $d_n > 0$ tending to ∞. Note that from the analog to (1.9) where e_n is replaced with d_n one obtains $d_n \approx \nu(S)/p_n$ (see Lemma 5.9 for a precise proof of the assertion $p_n d_n \to \nu(S)$). Hence the model constants d_n determine at which rate the failure probabilities tend to 0.

The crucial idea in the analysis of the asymptotic behavior of $\hat{p}_n$ is to approximate the estimator by the empirical measure of a *random transformation* $H_n(S)$ of the set S (with H_n defined in (2.2) below) under the following analog to $\hat{\nu}_n$ (defined in (1.8)) with the fitted GPDs replaced by the "true" ones:

$$\nu_n(B) := \frac{1}{k}\sum_{i=1}^{n} \varepsilon_{T_n^{\leftarrow}(X_i,Y_i)}(B).$$

Since the GPD approximation of the marginals is accurate only in the upper tail (and to avoid the aforementioned problem with the definition of $T_n^{\leftarrow}$), we must first show that asymptotically it does not matter if we replace S with a suitably defined subset S_n^* that is bounded away from the coordinate axes. For this set, we may use the approximation

$$\hat{p}_n \approx \frac{1}{e_n}\nu_n\Big(\frac{d_n}{e_n}H_n(S_n^*)\Big).$$

where the random transformation H_n of the marginals is defined by

$$H_n(x) := \frac{e_n}{d_n} T_n^{\leftarrow} \circ \hat{T}_n \circ (\hat{T}_n^{(c)})^{\leftarrow} \circ U(d_n x) \tag{2.2}$$

with

$$c = c_n := \frac{k}{n} e_n \tag{2.3}$$

and

$$\hat{T}_n^{(c)}(x,y) = \hat{T}_n(c_n x, c_n y). \tag{2.4}$$

Check that by (1.1) one has $H_n(x) \approx (e_n/d_n)(T_n^{(c)})^{\leftarrow} \circ U(d_n x) \approx (e_n/d_n)(T_n^{(c)})^{\leftarrow} \circ T_n((k/n)d_n x) \approx x$ (cf. Lemma 5.1).

Now, using the homogeneity of ν, we may break the estimation error into 6 parts as follows:

$$\begin{aligned}
\hat{p}_n - p_n \quad = \quad & \hat{p}_n - \frac{1}{e_n}\nu_n\Big(\frac{d_n}{e_n}H_n(S_n^*)\Big) \\
& + \frac{1}{e_n}\big(\nu_n(B) - E\nu_n(B)\big)|_{B=(d_n/e_n)H_n(S_n^*)} \\
& + \frac{1}{e_n}\big(E\nu_n(B) - \nu(B)\big)|_{B=(d_n/e_n)H_n(S_n^*)} \\
& + \frac{1}{d_n}\big(\nu(H_n(S_n^*)) - \nu(S_n^*)\big) \\
& + \frac{1}{d_n}\big(\nu(S_n^*) - \nu(S)\big)
\end{aligned}$$

$$
\begin{aligned}
&+ \nu(d_n S) - p_n \\
=: \quad & I + II + III + IV + V + VI.
\end{aligned} \tag{2.5}
$$

It will turn out that, under suitable conditions, part IV dominates all the other terms. Its asymptotic behavior is largely determined by the asymptotics of the marginal estimators if ν is sufficiently smooth.

Under very weak conditions on the set S, we will show that the terms I and V are negligible, if S_n^* is defined suitably. If d_n/e_n is bounded and bounded away from 0, then using methods from empirical process theory the second term can be shown to be asymptotically negligible. Part VI is a bias term which is negligible if d_n is sufficiently large (depending on the rate of convergence in (1.2)). Similarly, the term III, which equals $\big((n/k)P\{T_n^{\leftarrow}(X,Y) \in \tilde{B}\} - \nu(\tilde{B})\big)/d_n$ for $\tilde{B} = H_n(S_n^*)$, describes a bias term which is asymptotically negligible if both the approximation (1.4) and the marginal approximation $U((n/k)B) \approx T_n(B)$ are sufficiently accurate.

2.2. Conditions for asymptotic normality

We will make the following assumptions about the marginal distributions and the estimators of the marginal parameters:

(M1) There exist constants $x_i^0 < F_i^{\leftarrow}(1)$ such that F_i is continuous and strictly increasing on $[x_i^0, F_i^{\leftarrow}(1)] \cap \mathbb{R}$ for $i \in \{1,2\}$.

(M2) For all $i \in \{1,2\}$, there exist normalizing functions $a_i > 0$, $b_i \in \mathbb{R}$ and $A_i \neq 0$ and constants $\rho_i < 0$ such that for all $x > 0$

$$\lim_{t\to\infty} \frac{\frac{U_i(tx)-b_i(t)}{a_i(t)} - \frac{x^{\gamma_i}-1}{\gamma_i}}{A_i(t)} = \bar{\psi}_{\gamma_i,\rho_i}(x) := \begin{cases} \frac{x^{\gamma_i+\rho_i}}{\gamma_i+\rho_i}, & \gamma_i+\rho_i \neq 0 \\ \log x, & \gamma_i+\rho_i = 0 \end{cases}$$

(M3)

$$k^{1/2}\Big(\frac{\hat{a}_i(n/k)}{a_i(n/k)} - 1, \frac{\hat{b}_i(n/k) - b_i(n/k)}{a_i(n/k)}, \hat{\gamma}_i - \gamma_i\Big)_{1\le i\le 2} \longrightarrow (\alpha_i, \beta_i, \Gamma_i)_{1\le i\le 2}$$

weakly.

Condition (M1) is not crucial, but it is assumed to simplify the proofs and the formulation of some technical results (cf. de Haan and Ferreira, 2006, Theorem B.3.13).

(M2) is the usual second order condition with the additional restriction that the second order parameters ρ_i are negative. Again, one may drop the latter assumption at the cost of additional technical complications. According to Corollary 2.3.7 of de Haan and Ferreira (2006) we may and will assume that the normalizing constants are chosen such that the following uniform version holds: For all $\varepsilon, \delta > 0$ there exists t_0 such that

$$\left| \frac{\frac{U_i(tx)-b_i(t)}{a_i(t)} - \frac{x^{\gamma_i}-1}{\gamma_i}}{A_i(t)} - \bar{\psi}_{\gamma_i,\rho_i}(x) \right| \le \delta x^{\gamma_i+\rho_i} \max(x^{\varepsilon}, x^{-\varepsilon}) =: \delta x^{\gamma_i+\rho_i\pm\varepsilon} \tag{2.6}$$

provided $t, tx > t_0$. In fact, the main results hold under the following weaker assumption:

$$\Big|\frac{U_i(tx) - b_i(t)}{a_i(t)} - \frac{x^{\gamma_i} - 1}{\gamma_i}\Big| = O\big(A_i(t)x^{\gamma_i+\rho_i\pm\varepsilon}\big) \tag{2.7}$$

as $t \to \infty$ uniformly for $x \geq t_0/t$. Under condition (M2), A_i is regularly varying with index ρ_i.

Condition (M3) gives a lower bound on the rate at which the marginal estimators converge. Here some of the limiting random variables may be equal to 0 almost surely. In particular, this will usually be the case, if the ith marginal estimators use k_i largest order statistics and $k = o(k_i)$. However, typically at least some of the limiting random variables are non-degenerate and jointly normally distributed. In the sequel, we will choose versions such that the convergence in (M3) holds in probability.

The failure set D_n has to satisfy the following conditions.

(Q1) There exists a set

$$S = \big\{(x, y) \subset [0,\infty)^2 \mid y \geq q(x)\ \forall\, x \in [0,\infty)\big\} \subset [0,\infty)^2$$

and constants $d_n > 0$ tending to ∞ such that

$$D_n = U(d_n S) \cap \mathbb{R}^2 = \{(U_1(d_n x), U_2(d_n y)) \mid (x, y) \in S\} \cap \mathbb{R}^2.$$

Here the function $q : [0,\infty) \to [0,\infty]$, which describes the boundary of the "archetypal failure set" S, is assumed monotonically decreasing and continuous from the right with $q(0) > 0$.

(Q2)

$$\begin{aligned} x^{(1-\gamma_1)/2}|\log x| &= O(q(x)) \qquad \text{as} \qquad x \downarrow x_l := \inf\{x \geq 0 \mid q(x) < \infty\} \\ y^{(1-\gamma_2)/2}|\log y| &= O(q^{\leftarrow}(y)) \qquad \text{as} \qquad y \downarrow q(\infty) := \lim_{x\to\infty} q(x). \end{aligned}$$

In particular, condition (Q1) ensures that one may define the generalized (right-continuous) inverse function (of a decreasing function) in the usual way:

$$q^{\leftarrow}(v) := \inf\{x > 0 \mid q(x) \leq v\}$$

with the convention $\inf \emptyset = \infty$. Roughly speaking, the conditions (Q2) ensure that the archetypal failure set S does not have too much mass in a neighborhood of the axes where the estimated marginal transformation often perform poorly. It is always fulfilled if $\gamma_1 \leq 1$ or $x_l > 0$, resp., if $\gamma_2 \leq 1$ or $q(\infty) > 0$.

Moreover, we need some conditions on the extremal dependence between X and Y which is asymptotically described by the exponent measure ν defined in (1.2). In view of (1.1), one may replace the standardization by U with a standardization using T_n. To bound the bias terms III and VI in (2.5), we must specify the rate of the resulting convergence towards ν:

(D1) There exist an exponent measure ν on $[0,\infty)^2$ and a function $A_0(t)>0$ converging to 0 as t tends to ∞ such that

$$t_n P\Big\{\Big(\Big(1+\gamma_1\frac{X-b_1(t_n)}{a_1(t_n)}\Big)^{1/\gamma_1},\Big(1+\gamma_2\frac{Y-b_2(t_n)}{a_2(t_n)}\Big)^{1/\gamma_2}\Big)\in B\Big\}-\nu(B)=O(A_0(t_n))$$

uniformly for all sets $B\in\mathcal{B}_{t_n,M}$ for $t_n=n/k$ and for $t_n=d_n$ and arbitrary $M>0$.

Here, $\mathcal{B}_{t_n,M}$ consists of all sets of the form $\big\{(\tilde H^{(n,i)}_{\vartheta_i,\chi_i,\xi_i}(x_i))_{i\in\{1,2\}}\mid (x_1,x_2)\in C\big\}$ with $C=S\cap[u,\infty)\times[v,\infty)$ or $C=[x_l,u)\times[q(u-),\infty)$ or $C=[q^{\leftarrow}(v),\infty)\times[q(\infty),v)$ for some $u,v>0$ and some $\vartheta_i,\chi_i,\xi_i\in[-M,M]$ if $t_n=n/k$, and $\mathcal{B}_{t_n,M}$ comprises all sets of the form $\big\{((1+\gamma_i(U_i(d_nx_i)-b_i(d_n))/a_i(d_n))^{1/\gamma_i})_{i\in\{1,2\}}\mid (x_1,x_2)\in C\big\}$ with $C=[x_l,u)\times[q(u-),\infty)$ or $C=[q^{\leftarrow}(v),\infty)\times[q(\infty),v)$ for some $u,v>0$ if $t_n=d_n$. Here, for $i\in\{1,2\}$,

$$\tilde H^{(n,i)}_{\vartheta_i,\chi_i,\xi_i}(x) \tag{2.8}$$
$$:=\Big[1+\gamma_i\Big(\frac{c_n^{-(\gamma_i-k^{-1/2}\vartheta_i)}-1}{\gamma_i-k^{-1/2}\vartheta_i}(1+k^{-1/2}\xi_i)+c_n^{-(\gamma_i+k^{-1/2}\chi_i)}\frac{U_i(d_nx)-b_i(n/k)}{a_i(n/k)}\Big)\Big]^{1/\gamma_i}.$$

In (D1) the rectangles can also be replaced with the subsets $S\cap\big((0,u)\times(0,\infty)\big)$, resp. $S\cap\big((0,\infty)\times(0,v)\big)$. It is easy to see that condition (D1) is met if (X,Y) has a density f and ν a density η which satisfy the following approximation

$$\sup_{(x,y)\in(0,\infty)^2,x\vee y\ge1}\frac{1}{w(x,y)}\Big|ta_1(t)a_2(t)x^{\gamma_1-1}y^{\gamma_2-1} f\Big(a_1(t)\frac{x^{\gamma_1}-1}{\gamma_1}+b_1(t),a_2(t)\frac{y^{\gamma_2}-1}{\gamma_1}+b_2(t)\Big)-\eta(x,y)\Big|=O(A_0(t))$$

for some weight function w which is Lebesgue-integrable on $\{(x,y)\in(0,\infty)^2,x\vee y\ge1\}$. This sufficient condition applies e.g. to the bivariate Cauchy distribution restricted to $(0,\infty)^2$ and to densities of the form $f(x,y)=1/(1+x^\alpha+y^\beta)$ with $\alpha,\beta>1$ such that $1/\alpha+1/\beta<1$.

The dependence may also be described by the pertaining spectral measure Φ on $[0,\pi/2]$ defined by

$$\Phi([0,\vartheta])=\nu\Big\{(x,y)\in[0,\infty)^2\;\Big|\;x^2+y^2>1,\arctan\frac{y}{x}\le\vartheta\Big\},\quad \vartheta\in[0,\pi/2].$$

(D2) The spectral measure has a continuous Lebesgue density φ on $[0,\pi/2]$ such that $\inf_{\delta\le t\le\pi/2-\delta}\varphi(t)>0$ for all $\delta>0$ and

$$\lim_{\lambda\to1}\limsup_{t\downarrow0}\Big|\frac{\varphi(\lambda t)}{\varphi(t)}-1\Big|+\Big|\frac{\varphi(\pi/2-\lambda t)}{\varphi(\pi/2-t)}-1\Big|=0. \tag{2.9}$$

This assumption rules out that the spectral measure (and hence the exponent measure) puts mass on the coordinate axes. In particular, X and Y must not be asymptotically independent (in the sense of multivariate extreme value theory), because then the spectral measure is concentrated on $\{0, \pi/2\}$. Condition (2.9) is satisfied if φ is extended regularly varying at 0 and at $\pi/2$ (cf. Bingham et al., 1987, Section 2.0) or if the function $\log \circ\varphi\circ\exp$ has a bounded derivative.

Condition (D2) is somewhat restrictive in that it requires the spectral density to be bounded. Thus the exponent measure has a Lebesgue density η given by

$$\eta(x,y) = (x^2+y^2)^{-3/2}\varphi\Big(\arctan\frac{y}{x}\Big), \quad x,y>0, \tag{2.10}$$

which tends to 0 at the rate $(|x|+|y|)^{-3}$ as $|x|\to\infty$ or $|y|\to\infty$.

Finally, we impose the following conditions on the sequences d_n, e_n and $k=k_n$:

(S1) $k\to\infty$, $n=O(e_n)$ (so that $k=O(c_n)$ with $c_n=e_nk/n\to\infty$), $d_n\asymp e_n$ (i.e. $0<\liminf d_n/e_n \le \limsup_{n\to\infty} d_n/e_n<\infty$), and $w_n(\gamma_i)=o(k^{1/2})$ for $i\in\{1,2\}$ with

$$w_n(\gamma_i) := \begin{cases} \log c_n, & \gamma_i>0 \\ \frac{1}{2}\log^2 c_n, & \gamma_i=0 \\ \big(d_nk/n\big)^{-\gamma_i}, & \gamma_i<0. \end{cases}$$

(S2) $A_i(n/k)=o(k^{-1/2}w_n(\gamma_i))$ for $i\in\{1,2\}$ and $A_0(n/k)=o\big(k^{-1/2}\max(w_n(\gamma_1),w_n(\gamma_2))\big)$

(S3) $k^{1/2}=O(c_n\vee c_n^{\gamma_i})$ if $\gamma_i\ge 0$ for $i\in\{1,2\}$,
$k^{1/2}=o(c_n^{1-\gamma_1})$ if $\gamma_1<0$ and $x_l=0$, and
$k^{1/2}=o(c_n^{1-\gamma_2})$ if $\gamma_2<0$ and $q(\infty)=0$.

Recall that d_n is a constant determined by the model, which describes the rate at which the failure probability p_n tends to 0, while e_n is chosen by the statistician such that the inflated failure set contains sufficiently many observations. It seems natural to choose e_n of the same order as d_n, because this way one compensates for the shrinkage of D_n. More precisely, $d_n\asymp e_n$ if and only if the expected number of transformed observations in the inflated transformed failure set is of the same order as k, which can easily be checked in practical applications. To see this, note that by (1.4), (1.3) and (2.1) this expected number equals

$$\begin{aligned} nP\Big\{\hat{T}_n^{\leftarrow}(X,Y)\in \frac{n}{ke_n}\hat{T}_n^{\leftarrow}(D_n)\Big\} &\approx nP\Big\{\frac{k}{n}U^{\leftarrow}(X,Y)\in\frac{1}{e_n}U^{\leftarrow}(D_n)\Big\} \\ &\approx k\nu\Big(\frac{d_n}{e_n}S\Big) = k\frac{e_n}{d_n}\nu(S). \end{aligned} \tag{2.11}$$

However, the condition $d_n\asymp e_n$ can be substantially weakened at the price that one needs different conditions for different combinations of signs of γ_1 and γ_2.

The first condition of (S1) ensures that the expected number of marginally standardized observations in the *inflated* standardized failure region tends to ∞, whereas the second condition means that the expected number of observations in the failure region remains bounded as $n\to\infty$. The last condition of (S1) is needed to ensure consistency of the estimator in the sense that $\hat{p}_n/p_n\to 1$. It can only be satisfied if $\min(\gamma_1,\gamma_2)>-1/2$.

This restriction on the extreme value indices usually arises if one wants to prove asymptotic normality for estimators of tail probabilities; cf., e.g., de Haan and Ferreira (2006), Remark 4.4.3, or Drees et al. (2006), Remark 2.2.

From (S2) it follows that the bias is asymptotically negligible, while (S3) will imply that the part of the set S near the axes (corresponding to observations where one of the coordinates is much larger than the other) does not play an important role asymptotically. Similarly as above, these conditions may also be substantially weakened at the price of much more complicated conditions on the behavior of q depending on γ_1, γ_2 and η.

2.3. Asymptotic approximation of the estimator $\hat{p}_n$

Under the above condition we establish the following approximation to the estimation error of $\hat{p}_n$ in terms of the limiting random variables of the marginal estimators.

Theorem 2.1. *If the conditions (M1)–(M3), (D1), (D2), (Q1), (Q2) and (S1)–(S3) are fulfilled, then*

$$
\begin{aligned}
&k^{1/2} d_n(\hat{p}_n - p_n)\\
&= w_n(\gamma_1)\begin{cases} -\frac{\Gamma_1}{\gamma_1}\int_{q(\infty)}^{\infty} q^{\leftarrow}(v)\eta(q^{\leftarrow}(v),v)\,dv, & \gamma_1>0\\ \big(\frac{\alpha_1}{\gamma_1}-\beta_1-\frac{\Gamma_1}{\gamma_1^2}\big)\int_{q(\infty)}^{\infty}(q^{\leftarrow}(v))^{1-\gamma_1}\eta(q^{\leftarrow}(v),v)\,dv, & \gamma_1<0\\ -\Gamma_1\int_{q(\infty)}^{\infty} q^{\leftarrow}(v)\eta(q^{\leftarrow}(v),v)\,dv, & \gamma_1=0\end{cases}\\
&\quad + w_n(\gamma_2)\begin{cases} -\frac{\Gamma_2}{\gamma_2}\int_{x_l}^{\infty} q(u)\eta(u,q(u))\,du, & \gamma_2>0\\ \big(\frac{\alpha_2}{\gamma_2}-\beta_2-\frac{\Gamma_2}{\gamma_2^2}\big)\int_{x_l}^{\infty}(q(u))^{1-\gamma_2}\eta(u,q(u))\,du, & \gamma_2<0\\ -\Gamma_2\int_{x_l}^{\infty} q(u)\eta(u,q(u))\,du, & \gamma_2=0\end{cases}\\
&\quad + o_P\big(w_n(\gamma_1)\vee w_n(\gamma_2)\big)
\end{aligned}
\tag{2.12}
$$

Since $p_n d_n \to \nu(S)$, Theorem 2.1 remains true when the left-hand side of (2.12) is replaced with $k^{1/2}\nu(S)(\hat{p}_n/p_n - 1)$.

The weights $w_n(\gamma_1)$ and $w_n(\gamma_2)$ on the right-hand side of (2.12) may be different, and then they converge to ∞ at different rates. More precisely, $w_n(\gamma)$ is a non-increasing function of γ, and it is strictly decreasing on $(-\infty, 0]$. Therefore, the smaller of both marginal extreme value indices γ_1 and γ_2 determines the rate of convergence of $\hat{p}_n$ towards p_n. If at least one of the indices is non-positive and the indices are not equal, then the summand corresponding to the larger index is negligible. (In that case, it may happen that one cannot prove asymptotic normality using Theorem 2.1, because the limiting random variables α_i, β_i and Γ_i pertaining to the smaller extreme value index are equal to 0; cf. the above discussion of condition (M3).)

If both extreme value indices are positive, then both main terms on the right-hand side of (2.12) are of the same order. In that case, $(k^{1/2}d_n/\log c_n)(\hat{p}_n - p_n)$ converge to a limit distribution which typically will be non-degenerate if at least one of the limiting random variables Γ_1 and Γ_2 in (M3) is non-degenerate. If they are jointly normal, then we may derive the asymptotic normality of the estimator for the failure probability p_n.

However, if the fit of the marginal tails by GPDs is much more accurate than the approximation of the dependence structure by the extreme value dependence structure described by the exponent measure, then all limiting random variables in condition (M3) may be equal to 0, because the marginal estimators are based on the largest $k_i \gg k$ order statistics and converge at the rate $k_i^{-1/2} = o(k^{-1/2})$. In that case, Theorem 2.1 merely specifies an upper bound on the estimation error but not which of the terms $I - VI$ dominates the others.

2.4. Asymptotic confidence intervals

Theorem 2.1 can be used to construct asymptotic confidence intervals. To this end, it is advisable to reformulate the assertion as a convergence result on $k^{1/2} e_n(\hat{p}_n - p_n)$, because d_n is unknown. Then one needs consistent estimators for the variance of the random variables occurring on the right-hand side of (2.12) which usually are asymptotically normal, and consistent estimators for e_n/d_n times the integral there.

We will outline how to estimate the term $I_2 := (e_n/d_n) \int_{x_l}^{\infty} q(u)\eta(u, q(u))\, du$, that is needed in the case $\gamma_2 \geq 0$. To avoid the estimation of the density η of ν, we approximate the integral by the ν-measure of a shrinking set as follows. Because η is continuous, for small ℓ_n one has

$$\frac{e_n}{d_n} \int_{x_l}^{\infty} q(u)\eta(u, q(u))\, du \approx \frac{e_n}{d_n} \int_{x_l}^{\infty} \frac{1}{2\ell_n} \int_{(1-\ell_n)q(u)}^{(1+\ell_n)q(u)} \eta(u, v)\, dv\, du = \frac{1}{2\ell_n}(\nu(S_{n,2}^-) - \nu(S_{n,2}^+))$$

with

$$S_{n,2}^{\pm} := \Big\{ \frac{d_n}{e_n}(u, (1 \pm \ell_n)v) \mid (u, v) \in S \Big\}.$$

Now one can proceed similarly as in (1.9) (using (1.4) and (2.1)) to construct an estimator of $(e_n/d_n) \int_{x_l}^{\infty} q(u)\eta(u, q(u))\, du$:

Corollary 2.2. *Let*

$$\hat{S}_{n,2}^{\pm} := \Big\{ (u, (1 \pm \ell_n)v) \mid (u, v) \in \frac{n}{ke_n} \hat{T}_n^{\leftarrow}(D_n) \Big\}$$

for some sequence $\ell_n \downarrow 0$ such that $k^{-1/2}(w_n(\gamma_1) \vee w_n(\gamma_2)) = o(\ell_n)$. Suppose that all conditions of Theorem 2.1 are fulfilled and, in addition, that an analog to condition (D1) holds where c_n is replaced with $c_n/(1 \pm \ell_n)$. Then

$$\hat{I}_{n,2} := \frac{\hat{\nu}_n(\hat{S}_{n,2}^-) - \hat{\nu}_n(\hat{S}_{n,2}^+)}{2\ell_n} = \frac{e_n}{d_n} \int_{x_l}^{\infty} q(u)\eta(u, q(u))\, du (1 + o_P(1)).$$

In a completely analogous way one can estimate $I_1 := (e_n/d_n) \int_{q(\infty)}^{\infty} q^{\leftarrow}(v)\eta(q^{\leftarrow}(v), v)\, dv$ by $\hat{I}_{n,1} := (\hat{\nu}_n(\hat{S}_{n,1}^-) - \hat{\nu}_n(\hat{S}_{n,1}^+))/(2\ell_n)$ with

$$\hat{S}_{n,1}^{\pm} := \Big\{ ((1 \pm \ell_n)u, v) \mid (u, v) \in \frac{n}{ke_n} \hat{T}_n^{\leftarrow}(D_n) \Big\}.$$

Now suppose that both extreme value indices γ_i are positive and that we estimate them by the Hill estimator, i.e., $\hat\gamma_1 = k_1^{-1}\sum_{i=1}^{k_1}\log(X_{n-i+1:n}/X_{n-k_1:n})$ with $X_{n-i+1:n}$ denoting the ith largest order statistic among $X_1,\dots,X_n$, and likewise $\hat\gamma_2 = k_2^{-1}\sum_{i=1}^{k_2}\log(Y_{n-i+1:n}/Y_{n-k_2:n})$. It is well known that $k_i^{1/2}(\hat\gamma_i-\gamma_i)\to\mathcal{N}_{(0,\gamma_i^2)}$ if condition (M2) holds and $k_i^{1/2}A_i(n/k_i)\to 0$. In particular, $\Gamma_i=0$ if $k=o(k_i)$. However, if $k_i/k\to\kappa_i\in(0,\infty)$ for both $i=1$ and $i=2$, then the joint distribution of Γ_1 and Γ_2 is needed for the construction of confidence intervals.

In the case $k_1=k_2=k$, de Haan and Resnick (1993) derived a representation of Γ_i in terms of a Gaussian process under slightly different conditions than used in the present paper. One may mimic their approach to show that under our conditions, $(\Gamma_i/\gamma_i)_{i\in\{1,2\}}$ has the same distribution as $\big(\big(\int_1^\infty t^{-1}W_i(t/\kappa_i)\,dt - W_i(1/\kappa_i)\big)/\kappa_i\big)_{i\in\{1,2\}}$ where (W_1,W_2) is a bivariate centered Gaussian process with covariance function given by $Cov(W_1(s),W_1(t))=\nu\big((s\vee t,\infty)\times(0,\infty)\big)$, $Cov(W_2(s),W_2(t))=\nu\big((0,\infty)\times(s\vee t,\infty)\big)$ and $Cov(W_1(s),W_2(t))=\nu\big((s,\infty)\times(t,\infty)\big)$. Direct calculations show that thus $(\Gamma_i/\gamma_i)_{i\in\{1,2\}}$ is a centered Gaussian vector with marginal variances $1/\kappa_i$ and covariance $\nu\big((\kappa_2,\infty)\times(\kappa_1,\infty)\big)$. Hence, with $z_{1-\alpha/2}$ denoting the standard normal $(1-\alpha/2)$-quantile and $\hat\sigma^2 := \hat I_{n,1}^2/\kappa_1 + \hat I_{n,2}^2/\kappa_2 + 2\hat\nu_n\big((\kappa_2,\infty)\times(\kappa_1,\infty)\big)\hat I_{n,1}\hat I_{n,2}$,

$$\Big[\hat p_n - k^{-1/2}e_n^{-1}\log c_n\hat\sigma z_{1-\alpha/2}, \hat p_n + k^{-1/2}e_n^{-1}\log c_n\hat\sigma z_{1-\alpha/2}\Big] \tag{2.13}$$

is a two-sided confidence interval for p_n with asymptotic confidence level $1-\alpha$. (This formula is also applicable if one of the κ_i equals ∞.)

As an alternative to the above approach, one may estimate the density of the spectral measure Φ (cf. Cai et al., 2011) and construct both an estimator for the integrals and for the joint distribution of the limiting random variables on the right-hand side of (2.12) from it.

2.5. Choice of the blow-up factor

Our estimation procedure consists of two steps. First the marginal parameters are estimated using a certain fraction of largest order statistics, and both the observations and the failure set are marginally standardized accordingly. In the second step the transformed failure set is blown up by a factor chosen by the statistician, and the failure probability is estimated by a suitable fraction of the empirical probability of the inflated set. As the choice of a suitable sample fraction used in the marginal fitting has been extensively discussed in literature (see, e.g., Beirlant et al. (2004), Section 5.8), here we discuss how to choose the blow-up factor in the second step. For simplicity, in the concrete calculations we focus on the case that both extreme value indices are positive, but the general remarks apply to the other cases as well.

The estimation of the marginal parameters $\gamma_i, a_i(n/k)$ and $b_i(n/k)$ yield approximations of the marginal distribution functions of the type

$$\hat F_i(x) := 1-\Big(1+\hat\gamma_i\frac{x-\hat\mu_i}{\hat\sigma_i}\Big)^{-1/\hat\gamma_i}, \quad i=1,2, \tag{2.14}$$

which are sufficiently accurate for x satisfying $1 - F_i(x) \le k_i/n$. The corresponding estimator $\hat U_i^{\leftarrow} := 1/(1-\hat F_i)$ can also be interpreted as an estimator $(n/k)T_{n,i}^{\leftarrow}$ for different values of k. However, if one starts with a given approximation of the marginal tails as in (2.14), then the number k does not have any operational meaning. In that case it seems more natural to reformulate our estimator $\hat p_n$, the main result (2.12) and the resulting confidence interval (2.13) in terms of $\hat U_i^{\leftarrow}$.

For a fixed estimator $\hat U$ of U, the estimator of the failure probability

$$\hat p_n = \frac{1}{e_n}\hat\nu_n\big(\frac{n}{ke_n}\hat T_n^{\leftarrow}(D_n)\big) = \frac{1}{ke_n}\sum_{i=1}^n \varepsilon_{\hat U^{\leftarrow}(X_i,Y_i)}\Big(\frac{n}{ke_n}\hat U^{\leftarrow}(D_n)\Big)$$

depends on the constants k and e_n only via their product ke_n. At first glance, this seems peculiar, because in Theorem 2.1 the estimation error seemingly depends on k and e_n in completely different ways. However, according to the discussion in Subsection 2.4, for $\gamma_1, \gamma_2 > 0$, approximation (2.12) can be rewritten as

$$\hat p_n - p_n = (ke_n)^{-1/2}\log\frac{ke_n}{n}N(1+o_P(1)) \tag{2.15}$$

for a centered Gaussian random variable N with variance

$$\begin{aligned}\sigma_N^2 &= \frac{1}{e_n}\Big(\frac{k}{k_1}I_1^2 + \frac{k}{k_2}I_2^2 + 2\nu\Big(\Big(\frac{k_2}{k},\infty\Big)\times\Big(\frac{k_1}{k},\infty\Big)\Big)I_1I_2\Big)\\ &= \frac{ke_n}{k_1}\Big(\frac{I_1}{e_n}\Big)^2 + \frac{ke_n}{k_2}\Big(\frac{I_2}{e_n}\Big)^2 + 2\nu\Big(\Big(\frac{k_2}{ke_n},\infty\Big)\times\Big(\frac{k_1}{ke_n},\infty\Big)\Big)\frac{I_1}{e_n}\frac{I_2}{e_n}\Big),\end{aligned}$$

where $I_1 := (e_n/d_n)\int_{q(\infty)}^\infty q^{\leftarrow}(v)\eta\big(q^{\leftarrow}(v),v\big)\,dv$ and $I_2 := (e_n/d_n)\int_{x_l}^\infty q(u)\eta\big(u,q(u)\big)\,du$. Thus I_i/e_n does not depend on e_n, and the distribution of the approximating Gaussian random variable on the right-hand side of (2.15) depends on k and e_n only via their product.

Moreover, also the estimators

$$\begin{aligned}\frac{\hat I_{n,1}}{e_n} &= \frac{\hat\nu_n(\hat S_{n,1}^-) - \hat\nu_n(\hat S_{n,1}^+)}{2\ell_n e_n}\\ &= \frac{1}{2\ell_n}\cdot\frac{1}{ke_n}\sum_{i=1}^n \varepsilon_{\hat U^{\leftarrow}(X_i,Y_i)}\Big(\Big\{((1-\ell_n)u,v)\mid (u,v)\in\frac{n}{ke_n}\hat U^{\leftarrow}(D_n)\Big\}\setminus\\ &\qquad\Big\{((1+\ell_n)u,v)\mid (u,v)\in\frac{n}{ke_n}\hat U^{\leftarrow}(D_n)\Big\}\Big)\end{aligned}$$

and likewise $\hat I_{n,2}/e_n$ depend on the product ke_n only. Finally, the covariance term $\nu\big(k_2/(ke_n),\infty)\times(k_1/(ke_n),\infty)\big) = k^2e_n/(\lambda k_1k_2)\nu\big((k/(\lambda k_1),\infty)\times(k/(\lambda k_2),\infty)\big)$ can be estimated by

$$\frac{k^2e_n}{\lambda k_1k_2}\hat\nu_n\Big(\Big(\frac{k}{\lambda k_1},\infty\Big)\times\Big(\frac{k}{\lambda k_2},\infty\Big)\Big) = \frac{ke_n}{\lambda k_1k_2}\sum_{i=1}^n\varepsilon_{\hat U^{\leftarrow}(X_i,Y_i)}\Big(\Big(\frac{n}{\lambda k_1},\infty\Big)\times\Big(\frac{n}{\lambda k_2},\infty\Big)\Big).$$

Here the choice $\lambda \in (0,1]$ ensures that $\hat{U}^{\leftarrow}$ is used only on the range where it is a sufficiently accurate estimator of the true function $U^{\leftarrow}$.

To sum up, all estimates only depend on ke_n, but not on the numbers k and e_n separately. This product should be chosen as large as possible under the constraints that both marginal approximations of $U_i^{\leftarrow}$ by $\hat{U}_i^{\leftarrow}$ and the approximation of the joint distribution of the standardized vector (cf. (1.2)) are reliable. To ensure the former constraint, for the vast majority of the observations (X_i, Y_i), the indicator of the set $\{\hat{U}^{\leftarrow}(X_i, Y_i) \in n/(ke_n)\hat{U}^{\leftarrow}(D_n)\}$ should not depend on the particular values of $\hat{U}_1^{\leftarrow}(X_i)$ or $\hat{U}_2^{\leftarrow}(Y_i)$ if these are smaller than n/k_1 or n/k_2 (either because the other component of the vector is so large that the observations lie in the failure set anyway, or because the other component is so small so that the indicator is 0 even if the maximal value n/k_i is attained). For instance, if we consider failure sets of the type $D_n := \{(x,y) \mid \alpha_1 x + \alpha_2 y > R\}$, then ke_n should be smaller than $\min_{i=1,2} k_i \hat{U}_i^{\leftarrow}(R/\alpha_i)$, because otherwise for sure $\hat{U}^{\leftarrow}(x,y) \in (n/ke_n)\hat{U}^{\leftarrow}(D_n)$ for some values (x,y) for which $\hat{U}^{\leftarrow}(x,y)$ is not a reliable estimate of $U^{\leftarrow}(x,y)$.

However, the above crude upper bound for ke_n is not sufficient to ensure that $\hat{p}_n$ is a reliable estimate of p_n, because the dependence structure must be accurately described by the exponent measure ν, too. To determine a range of reasonable values for ke_n, we propose (in analogy to the well-known Hill plot used for selecting a reasonable sample fraction in the marginal fitting), to plot $\hat{p}_n$ versus ke_n and then to choose ke_n in a range where this curve seems stable. In the data example discussed in Section 3, this approach seems to work pretty well. (Motivated by the discussion by Drees et al. (2000), it might also be worthwhile to use a log-scale for ke_n in order to get a clearer picture about a good choice for this factor, but (unlike for the so-called AltHill plot) a sound theoretical justification for this modification is yet lacking.)

2.6. Generalization to higher dimensions

We conclude this section by indicating how to generalize the main result to $\mathbb{R}^d$-valued vectors $\mathbf{X}_i = (X_{i,1}, \ldots, X_{i,d})$ of arbitrary dimension $d \geq 2$, albeit a detailed discussion is beyond the scope of this paper. An inspection of the proof of Lemma 5.3 reveals that the generalized inverse $q^{\leftarrow}$ of the function q is used to describe the boundary of the set S as a function of the second coordinate. If $d > 2$ (and hence the generalized inverse is not defined), then an analogous description is needed for all coordinates, i.e. we need d different representations of the set S of the form

$$S = \big\{\mathbf{x} \in [0,\infty)^d \mid x_i \geq q_i(\mathbf{x}_{-i})\big\}, \quad 1 \leq i \leq d, \tag{2.16}$$

where $\mathbf{x}_{-i} \in [0,\infty)^{d-1}$ denotes the vector $\mathbf{x}$ with ith coordinate removed and q_i are suitable $[0,\infty]$-valued functions that are decreasing in each argument. Then one may proceed as in the case $d = 2$ by separately examining the influence of the transformation of each marginal on the ν-measure of the (suitably restricted) set S. Under suitable integrability conditions on the functions q_i and obvious generalizations of the conditions

(M1)–(M3), (D1), (D2) and (S1)–(S3), it can be shown that

$$\begin{aligned} & k^{1/2} d_n(\hat{p}_n - p_n) \\ = \quad & \sum_{i=1}^d w_n(\gamma_i) \begin{cases} -\frac{\Gamma_i}{\gamma_i} \int q_i(v) \eta(\tilde{q}_i(v)) 1_{(0,\infty)}(q_i(v))\, \boldsymbol{\lambda}^{d-1}(dv), & \gamma_i > 0 \\ \left(\frac{\alpha_i}{\gamma_i} - \beta_- \frac{\Gamma_i}{\gamma_i^2}\right) \int (q_i(v))^{1-\gamma_1} \eta(\tilde{q}_i(v)) 1_{(0,\infty)}(q_i(v))\, \boldsymbol{\lambda}^{d-1}(dv), & \gamma_i < 0 \\ -\Gamma_i \int q_i(v) \eta(\tilde{q}_i(v)) 1_{(0,\infty)}(q_i(v))\, \boldsymbol{\lambda}^{d-1}(dv), & \gamma_i = 0 \end{cases} \\ & + o_P(w_n(\gamma_i)) \end{aligned} \tag{2.17}$$

Here $\boldsymbol{\lambda}^{d-1}$ denotes the Lebesgue measure on $[0,\infty)^{d-1}$ and $\tilde{q}_i(v)$ is the vector in $[0,\infty)^d$ whose ith coordinate equals $q_i(v)$ and the other $d-1$ coordinates are those of v.

If the boundary ∂S of the set S is sufficiently smooth, then the integrals on the right-hand side of (2.17) can be represented more naturally as integrals w.r.t. certain differential forms (see, e.g., Schreiber, 1977, for an informal introduction to differential forms). More precisely, assume that there exists a set $D \subset [0,\infty)^{d-1}$ and a continuously differentiable function $q : D \to [0,\infty)$, such that $\partial S = \{\Psi(u) := (u, q(u)) \mid u \in D\}$. Then the right-hand side of (2.17) equals

$$\begin{aligned} & \sum_{i=1}^d w_n(\gamma_i) \begin{cases} -\frac{\Gamma_i}{\gamma_i} \int_\Psi pr_i \cdot \eta \, dx_1 \wedge \cdots \wedge dx_{i-1} \wedge dx_{i+1} \wedge \cdots \wedge dx_d, & \gamma_i > 0 \\ \left(\frac{\alpha_i}{\gamma_i} - \beta_- \frac{\Gamma_i}{\gamma_i^2}\right) \int_\Psi (pr_i)^{1-\gamma_i} \cdot \eta \, dx_1 \wedge \cdots \wedge dx_{i-1} \wedge dx_{i+1} \wedge \cdots \wedge dx_d, & \gamma_i < 0 \\ -\Gamma_i \int_\Psi pr_i \cdot \eta \, dx_1 \wedge \cdots \wedge dx_{i-1} \wedge dx_{i+1} \wedge \cdots \wedge dx_d, & \gamma_i = 0 \end{cases} \\ & + o_P(w_n(\gamma_i)) \end{aligned}$$

with pr_i denoting the projection to the ith coordinate, which is the integral of a $(d-1)$-form over δS. This representation reflects most clearly the fact that the ith term results from the change of the boundary surface of S by the marginal transformation $H_{n,i}$. Such a representation can be derived for more general differentiable manifolds ∂S.

3. Analysis of insurance claims

In this section, we discuss issues arising in the analysis of a well-known data set of claims to Danish fire insurances. The data set contains losses to building(s), losses to contents and losses to profits (caused by the same fire) observed in the period 01/1980 - 12/2002, discounted to 07/1985. The claims are recorded only if the sum of all components exceeds 1 million Danish Kroner (DKK). Due to this recording method, there is an artificial negative dependence between the components, since if one component is smaller than 1 million DKK, the sum of the others must be accordingly larger. To avoid this effect, we therefore consider only those claims for which at least one component exceeds 1 million DKK, which leads to a sample of 3976 claims. Moreover, we focus on the losses to buildings, denoted by X_i as a multiple of one million DKK, and the losses to contents Y_i. A more detailed description of the data can be found in Müller (2008) and Drees and Müller (2008).

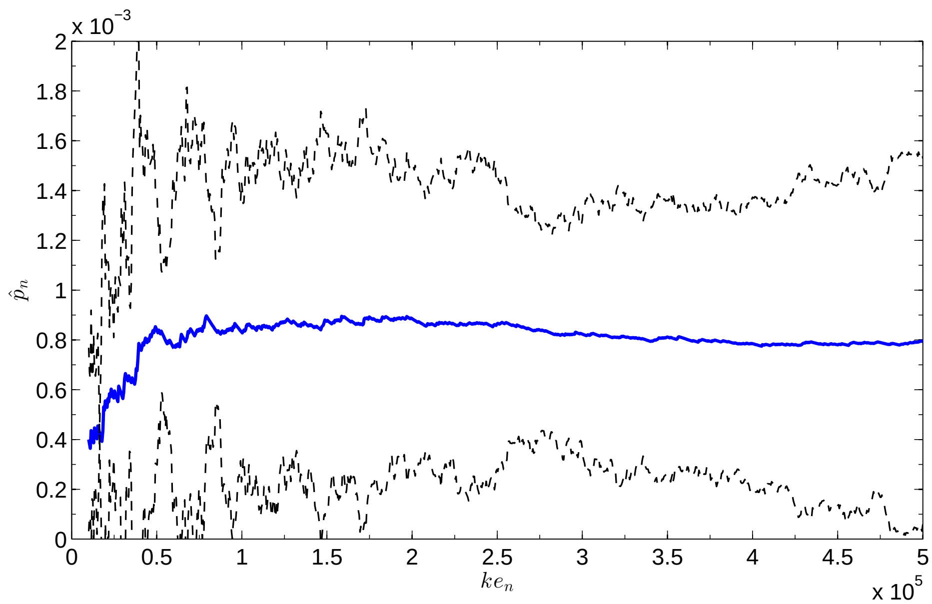

Figure 1. $\hat{p}_n$ (solid blue line) and confidence intervals (black dashed line) versus ke_n for Danish fire insurance claims

As described in the introduction, we assume that a quota reinsurance pays $(1-\alpha_X)X_i$ for each loss X_i to the building and $(1-\alpha_Y)Y_i$ for each loss Y_i of content, while an XL-reinsurance pays if the remaining costs $\alpha_X X_i + \alpha_Y Y_i$ exceed a retention level R. We want to estimate the probability $p_n := P(D_n)$ with $D_n := \{\alpha_X X_i + \alpha_Y Y_i > R\}$ that a fire results in a claim to the XL-reinsurance for $\alpha_X = 1$, $\alpha_Y = 0.5$ and $R = 100$. (More precisely, we estimate the conditional probability given that $\max(X_i, Y_i) > 1$.)

Müller (2008), Section 5.1.2, fitted GPD's to the marginal distributions using the Hill estimators based on the $k_1 = 900$ and $k_2 = 600$ largest observations to obtain a tail approximation of the type (2.14) with parameters $\hat{\gamma}_1 = 0.57$, $\hat{\sigma}_1 = 0.54$, $\hat{\mu}_1 = 0.91$, $\hat{\gamma}_2 = 0.72$, $\hat{\sigma}_2 = 0.47$ and $\hat{\mu}_2 = 0.15$. Moreover, he showed that the components of the claim vector are apparently asymptotically dependent.

As suggested in Subsection 2.5, in Figure 1 we plot the estimate of the failure probability versus ke_n for values of ke_n ranging from 10^4 to $5 \cdot 10^5$. In the present situation, the crude upper bound on ke_n discussed in Subsection 2.5 is about $1.7 \cdot 10^6$, but the curve of probability estimates shows a clear downward trend for $ke_n > 2 \cdot 10^5$, which is most likely due to a deviation of the dependence structure from its limit. On the other hand, for values smaller than $5 \cdot 10^4$ the curve is very unstable, too, because the random error is too large as just a few observations fall into the inflated failure set (e.g., about 25 if $ke_n \approx 3 \cdot 10^4$). This lower bound on ke_n reflects the condition in the asymptotic framework that n is of smaller order than ke_n (see condition (S1)). In view of this plot, the choice $ke_n = 2 \cdot 10^5$ seems reasonable.

In addition, Figure 1 shows a two-sided confidence interval with nominal size 0.95, again as a function of ke_n. Here we have chosen $\ell_n = 0.1$ and $\lambda = 1$ in the estimator of the variance σ_N^2 described above; other values of λ between 1/2 and 1 yield approximately the same estimates, while smaller values of ℓ_n lead to larger fluctuations in the confidence bounds, that however are still of a similar size.

For $ke_n = 2 \cdot 10^5$ one obtains a point estimate for p_n of about $8.8 \cdot 10^{-4}$ and a confidence interval $[2.2 \cdot 10^{-4}, 1.54 \cdot 10^{-3}]$. At first glance, this confidence interval seems rather wide. However, here we estimate the probability of a very rare event which has occurred only twice in the observational period of more than 20 years. Indeed the empirical probability of the event is about $5 \cdot 10^{-4}$, and the Clopper-Pearson confidence interval $[6 \cdot 10^{-5}, 1.8 \cdot 10^{-3}]$ (again with nominal size 0.95) is even wider. It is worth mentioning that both the empirical point estimate and the Clopper-Pearson confidence interval are exactly the same if one wants to estimate the probability that a claim occurs to the XL-reinsurance for any retention level R between 77 and 145 million DKK! Moreover, for retention level above 152 million DKK the point estimate would be 0 and thus useless for purposes of risk management.

4. Simulation study

In a small simulation study we examine the finite sample behavior of our estimator of a failure probability. In particular, we want to compare its performance with that of the estimator proposed by de Haan and Sinha (1999). Moreover, we demonstrate that often the fit of the marginal distributions is the main source of the random error, as indicated by the main Theorem 2.1.

We consider two different models for the dependence structure of (X, Y).

- For the first model class we assume that (X, Y) has a Gumbel copula, i.e. for some $\vartheta \in (1, \infty)$

  $$F(F_1^{\leftarrow}(u), F_2^{\leftarrow}(v)) = C_\vartheta^{Gum}(u, v) = \exp\Big(-\big(|\log u|^\vartheta + |\log v|^\vartheta\big)^{1/\vartheta}\Big), \quad 0 < u, v \le 1.$$

  Note that C_ϑ^{Gum} is the Copula of a bivariate extreme value distribution. The corresponding exponent measure is given by

  $$\nu_\vartheta^{Gum}\big((u, \infty) \times (v, \infty)\big) = \frac{1}{u} + \frac{1}{v} + \big(u^{-\vartheta} + v^{-\vartheta}\big)^{1/\vartheta}, \quad u, v > 0,$$

  and the spectral density by

  $$\varphi_\vartheta^{Gum}(\arctan t) = (\vartheta - 1)\big(1 + t^2\big)^{3/2} t^{-(\vartheta+1)}\big(1 + t^{-\vartheta}\big)^{1/\vartheta - 2}.$$

  Hence

  $$\varphi_\vartheta^{Gum}(t) \sim (\vartheta - 1)t^{\vartheta-2} \qquad \text{and} \qquad \varphi_\vartheta^{Gum}\big(\pi/2 - t\big) \sim (\vartheta - 1)t^{\vartheta-2} \text{ as } t \downarrow 0,$$

and condition (D2) is obviously fulfilled for $\vartheta \geq 2$, whereas for $\vartheta \in (1,2)$ the spectral density is not bounded and thus (D2) is not satisfied.
In the simulation results presented below, the Gumbel copula with $\vartheta = 5$ is combined with generalized extreme value (GEV) marginals $F_i(x) = \exp\big(-(1+\gamma_i x)^{-1/\gamma_i}\big)$ for $1+\gamma_i x > 0$ with extreme value index $\gamma_1 = \gamma_2 \in \{-1/4, 0, 1/4\}$. (Simulations with GPD marginals yielded very similar results which are thus not presented here.)

- The second model class is related to the model suggested by Schlather (2002). Similarly as in Example 3.8 Segers (2012), we define $(X,Y) = \sqrt{2\pi}(SZ, TZ)$ where Z is a unit Fréchet random variable (i.e. $P\{Z \leq x\} = \exp(-1/x)$ for $x > 0$) independent from the vector (S,T) which has the distribution of a centered normal vector with variances 1 and covariance $\varrho \in (-1,1)$ conditioned on both coordinates being positive. By Lemma 3.1 of Segers (2012), the distribution of (X,Y) belongs to the max domain of attraction of an extreme value distribution with unit Fréchet marginals. Direct calculations show that the stable tail dependence function is given by

$$\ell(x,y) := \nu\Big(\Big(\frac{1}{x},\infty\Big)\times\Big(\frac{1}{y},\infty\Big)\Big) = \frac{1}{1+\rho}\big(\rho(x+y)+(x^2+y^2-2\rho xy)^{1/2}\big), \quad x,y \geq 0.$$

Hence, the exponent measure has the density

$$\eta(u,v) = \frac{1-\varrho^2}{1+\rho}\big(u^2+v^2-2\varrho uv\big)^{-3/2},$$

and the pertaining spectral density

$$\varphi(t) = \frac{1-\varrho^2}{1+\rho}\Big(1-2\varrho\frac{\tan t}{1+\tan^2 t}\Big)$$

is strictly positive and continuous on $[0,\pi/2]$, so that condition (D2) is fulfilled. We have simulated data sets with correlation $\varrho \in \{-0.8, 0.2, 0.8\}$, but for briefness sake will report results only for the last value, because they look similar in all three cases.
Direct calculations show that the marginal distributions are symmetric with

$$1-F_i(x) = 1/2 - \exp\big(\pi x^{-2}\big)\big(1-\Phi\big(\sqrt{2\pi}x^{-1}\big)\big), \quad x > 0.$$

We want to estimate failure probabilities of linear half spaces of the form $D = \{(x,y) \mid x + y/2 > R\}$ where R has been chosen such that the failure probability p_n (which is determined by simulations) lies between $2\cdot 10^{-4}$ and $5\cdot 10^{-4}$ for each of the above models.

For each setting we have simulated 1000 data sets of size $n = 500$. As $np_n \ll 1$ (with a value of about 0.1 in most settings), the estimation of p_n is a challenging task. In particular, one cannot use empirical probabilities, because in most simulations the failure set will not contain any data point.

To make a comparison with the estimator $\hat{p}_n^{(HS)}$ proposed by de Haan and Sinha (1999) easier, we use their marginal estimators:

$$\begin{aligned}
\hat{\gamma}_1 &:= M_1(X) + 1 - \frac{1}{2} \Big/ \Big(1 - \frac{M_1^2(X)}{M_2(X)}\Big), \\
\hat{a}_1(n/k) &:= X_{n-k:n} \big(3M_1^2(X) - M_2(X)\big)^{1/2} \times \\
&\quad \times \Big(\frac{3}{(1-\hat{\gamma}_1 \vee 0)^2} - \frac{2}{(1-\hat{\gamma}_1 \vee 0)(1-2\hat{\gamma}_1 \vee 0)} \Big)^{-1/2}, \\
\hat{b}_1(n/k) &:= X_{n-k:n}
\end{aligned}$$

with

$$M_r(X) = \frac{1}{k} \sum_{i=1}^{k} \Big(\log \frac{X_{n-i+1:n}}{X_{n-k:n}} \Big)^r.$$

For small values of k it happens (in at most a few percents of the simulations) that the scale estimate $\hat{a}_1(n/k)$ is not defined. In these simulations, instead we use the moment estimator

$$\hat{a}_1(n/k) := \frac{1}{2} X_{n-k:n} M_1(X) \Big/ \Big(1 - \frac{M_1^2(X)}{M_2(X)}\Big)$$

proposed by Dekkers et al. (1989). The estimators of the parameters of the second marginal distribution are defined likewise with X_i replaced with Y_i.

We now discuss our findings for the model with Gumbel copula C_5^{Gum} and Gumbel margins (i.e. GEV margins with $\gamma = 0$) in detail. The results for the other distributions of (X, Y) are then presented more briefly. In this model, the true failure probability p_n is about $2.25 \cdot 10^{-4}$ for $R = 12$.

Figure 2 displays the empirical root mean squared error (RMSE) of our estimator $\hat{p}_n$ as a function of the number k of largest order statistics used for marginal fitting and the product ke_n which determines the blow-up factor. As expected, if either of these values is small, then the RMSE is high due to a large standard error, while for too large a value the bias leads to a large RMSE; in particular the first effect is much more pronounced for the blow-up factor. It can be most clearly seen from the contour lines shown in the left plot of Figure 3 that values of k in the range 50–150 and values of ke_n around $1.5 \cdot 10^6$ yield most accurate estimates.

Next we compare the performance of $\hat{p}_n$ with that of the estimator $\hat{p}_n^{(HS)}$ suggested by de Haan and Sinha (1999). Recall from subsection 1.4 that the main difference between these estimators is that the latter uses a data-driven value $\hat{c}_n$ instead of ke_n/n, which is constructed as an estimator of the (quite arbitrarily fixed) factor d_n; see formula (1.6) of de Haan and Sinha (1999). The graph on the right hand side of Figure 3 shows a contour plot of the ratio of the RMSE of $\hat{p}_n^{(HS)}$ and of the RMSE of $\hat{p}_n$ again as a function of k and ke_n. Obviously, the RMSE of $\hat{p}_n^{(HS)}$ is much larger than that of our estimator for almost all values of k and e_n. Indeed, for the most reasonable choices of k the former is usually as least double as large as the latter.

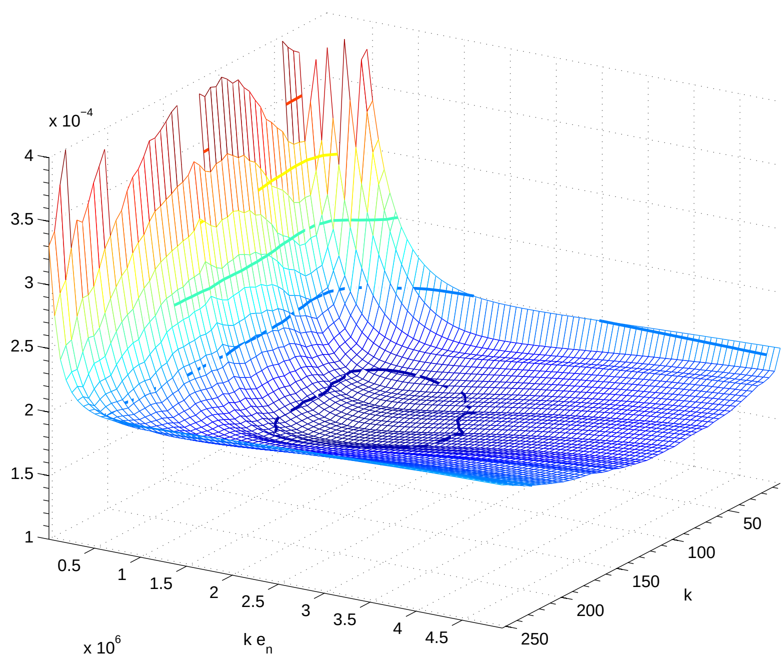

Figure 2. Empirical RMSE of $\hat{p}_n$ for (X, Y) with Gumbel $\vartheta = 5$ copula and Gumbel marginals

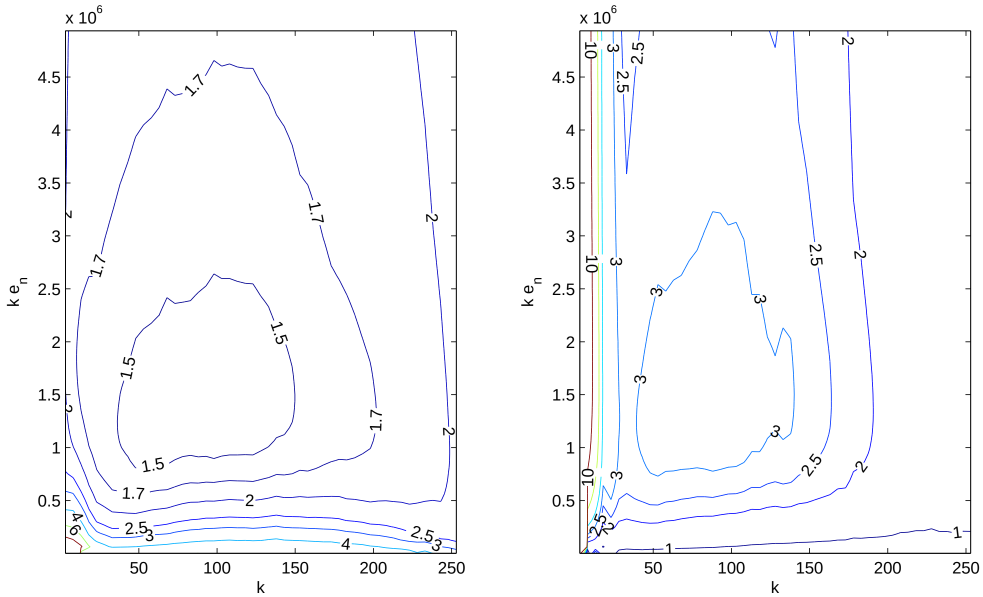

Figure 3. Contour lines of $10^4\times$RMSE of $\hat{p}_n$ (left) and of the ratio between the RMSE of $\hat{p}_n^{(HS)}$ and $\hat{p}_n$ (right)

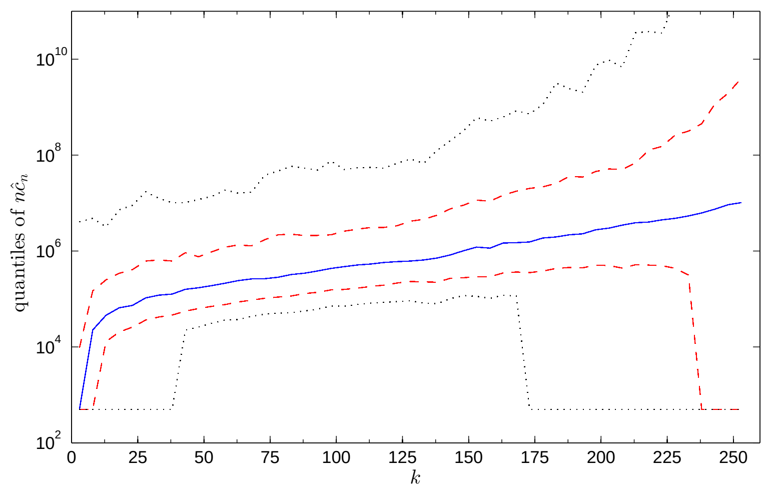

Figure 4. Empirical quantiles of $n\hat{c}_n$ to the levels 0.1, 0.25, 0.5, 0.75 and 0.9

The reason for the inferiority of $\hat{p}_n^{(HS)}$ can be seen from Figure 4 which shows empirical quantiles of $n\hat{c}_n$ (corresponding to ke_n) as a function of k for the levels 0.1, 0.25, 0.5, 0.75 and 0.9. For $k < 200$, in more than half of the simulations $n\hat{c}_n$ is clearly smaller than 10^6, whereas a good choice of ke_n would be in the range $10^6 - 2 \cdot 10^6$. Moreover, the variation of $n\hat{c}_n$ is huge with more than 10% of the estimates exceeding 10^7 for $k \geq 50$. This figure indicates that the estimator employed is rather inaccurate. However, even if the theoretical value of $c_n = kd_n/n$ was known, in general the resulting estimator of the failure probability would not perform well, because d_n is arbitrarily defined by the requirement that (1,1) lies on the boundary of S. In particular, this choice of d_n is not at all related to the accuracy of the approximation (1.4) which strongly influences the part of the bias not resulting from the fitting of the marginal distributions. For example, in the present situation $d_n = U^{\leftarrow}(R/1.5) = \mathrm{e}^{R/1.5} \approx 3000$, which leads to much too small values of ke_n if one chooses $e_n = d_n$.

In the situation of Theorem 2.1, asymptotically the main source of *random* error is the fitting of the marginal distributions. To check whether this bears out for moderate sample sizes, we have calculated an analog to our estimator of the failure probability where the marginal estimator $T_{n,i}^{\leftarrow}$ is replaced with the true function $(k/n)U^{\leftarrow}$:

$$\hat{p}_n^{(tm)} := \frac{1}{ke_n} \sum_{i=1}^{n} \varepsilon_{U^{\leftarrow}(X_i,Y_i)} \Big(\frac{n}{ke_n} U^{\leftarrow}(D) \Big).$$

The left-hand plot in Figure 5 displays contour lines of the ratio between the standard errors of $\hat{p}_n$ and of $\hat{p}_n^{(tm)}$ as a function of k and ke_n. For almost all values of these tuning parameters, the standard error of the estimator $\hat{p}_n$ with estimated marginals is at least

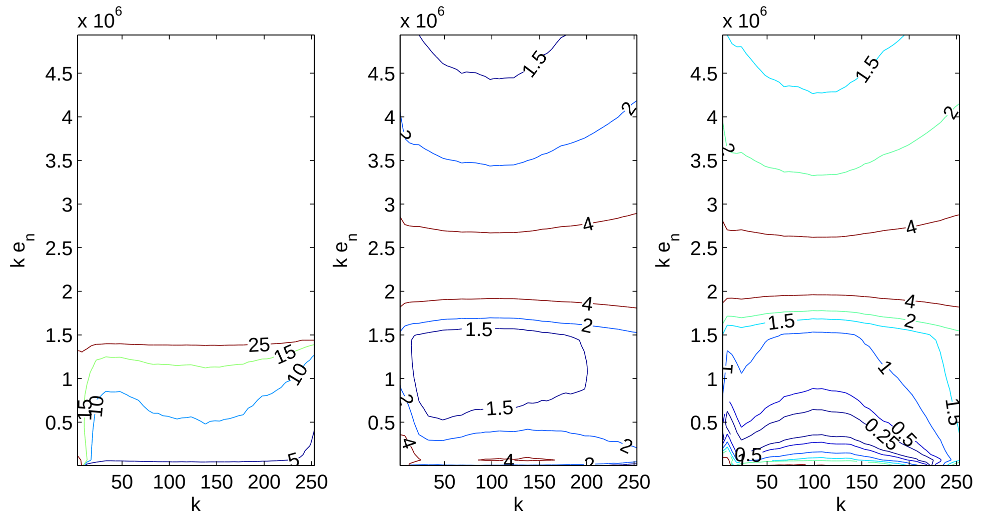

Figure 5. Contour plots of the ratio of the standard errors of $\hat{p}_n$ and of $\hat{p}_n^{(tm)}$ (left), of the corresponding ratios of RMSE (middle) and of the ratio of absolute bias (right)

model	copula C_5^{Gum}		Schlather type model
	$\gamma = 0.25$	$\gamma = -0.25$	$\varrho = 0.8$
R	40	5	5000
$10^4\times$failure probab.	2.21	5.00	3.40

Table 1. Failure probabilities for different models and thresholds R

5 times larger than that of $\hat{p}_n^{(tm)}$, and it is more than 100 times larger if $ke_n > 1.5 \cdot 10^6$. However, if one considers the RMSE, then the picture is less dramatic. While the total error of $\hat{p}_n$ is still at least about 25% higher than that of $\hat{p}_n^{(tm)}$, for the most reasonable choices of k and e_n it is less than three times as large (middle plot in Figure 5). The reason for this different behavior of the RMSE is of course the bias, which is at least 5 times higher in absolute value than the standard error for $\hat{p}_n^{(tm)}$ with $ke_n > 0.5 \cdot 10^6$ and is thus the dominating part of the total error. At first glance one might expect that using the true marginal distributions should also reduce the bias. However, in general this is not true, since the bias resulting from the marginal fit may partly cancel out with the bias caused by the error in approximation (1.4). Indeed, for moderate values of ke_n the absolute bias of $\hat{p}_n$ is often smaller than that of $\hat{p}_n^{(tm)}$ as it can be seen from the contour plot of the ratio of these values in the right-hand plot of Figure 5.

Table 1 lists the true failure probabilities for the four remaining settings not discussed so far.

Figure 6 shows contour lines of the RMSE of $\hat{p}_n$ and of the ratio of the RMSE of $\hat{p}_n^{(HS)}$ and $\hat{p}_n$ in the two left-hand columns (corresponding to the plots in Figure 3) and the ratios of the standard error and the RMSE of $\hat{p}_n$ and $\hat{p}_n^{(tm)}$ (corresponding to the first two plots in Figure 5) in the two right-hand columns. The four models are displayed in the four rows in the same order as in Table 1.

While in most cases the results are qualitatively the same as for the Gumbel model discussed above, there are two remarkable exceptions. First in the model with Gumbel copula and GEV marginals with $\gamma = -0.25$ the estimator suggested by de Haan and Sinha (1999) works pretty well with an RMSE which is at most 25% larger than that of $\hat{p}_n$ for most combinations of k and e_n. (For some not very reasonable combinations it is even smaller.)

Second, in the Schlather type model with $\varrho = 0.8$, due to its rather large bias, the RMSE of the estimator $\hat{p}_n^{(tm)}$ which uses the true marginal distributions is clearly larger than the one of $\hat{p}_n$ if $k \geq 50$ and ke_n is about $8 \cdot 10^5$. However, as the first plot for this distribution shows, these combinations of values for k and ke_n are not good choices for $\hat{p}_n$, because its RMSE is more than 4 times as large as the minimal value. Indeed, for all simulated distributions the bias of $\hat{p}_n^{(tm)}$ was more stable than that of $\hat{p}_n$ (in particular for small value of k), thus indicating again that one should be particularly careful with the estimation of the marginal distributions.

5. Proofs

The proof of Theorem 2.1 is based on the decomposition (2.5) of the estimation error. The asymptotic behavior of the leading term $IV + V$ is established in Corollary 5.5. As a preparation for this result, first we establish an approximation of the random transformation of the marginals defined in (2.2). Thereby we must restrict ourselves to arguments that are neither too small nor too large, which defines a certain subset S_n^* of S. In Lemma 5.2 an upper bound on the difference between the ν-measures of S and S_n^* is derived, while the Lemmas 5.3 and 5.4 analyze the influence of the marginal transformations on the ν-measure of S_n^*.

In Lemma 5.6 an upper bound on the term II of decomposition (2.5) is proved using empirical process theory. Finally, Lemma 5.8 establishes upper bounds on the terms I and III, while Lemma 5.9 takes care of VI.

Lemma 5.1. *Assume that the conditions (M1)–(M3) and (S1) are fulfilled. For $i \in \{1,2\}$, let $\lambda_{n,i} > 0$ be a decreasing and $\tau_{n,i} < \infty$ an increasing sequence, such that the following conditions are met:*

(i) $A_i(n/k)(\lambda_{n,i} d_n k/n)^{\rho_i \pm \varepsilon} = o\big(k^{-1/2} w_n(\gamma_i)\big)$ *for some* $\varepsilon > 0$
(ii) If $\gamma_i > 0$, *then* $k^{-1/2} = o\big((\lambda_{n,i} d_n/e_n)^{\gamma_i}\big)$.
(iii) If $\gamma_i < 0$, *then* $k^{-1/2} = o\big((\tau_{n,i} d_n k/n)^{\gamma_i}\big)$ *and* $\log(d_n/e_n) = o\big((d_n k/n)^{-\gamma_i}\big)$
(iv) If $\gamma_i = 0$, *then* $k^{-1/2} \log \tau_{n,i} \to 0$ *and* $\log(d_n/e_n) = o(\log c_n)$

Then, for $i \in \{1,2\}$,

$$\frac{d_n}{e_n} H_{n,i}(x) = T_{n,i}^{\leftarrow} \circ \hat{T}_{n,i} \circ \hat{T}_{n,i}^{(c)\leftarrow} \circ U_i(d_n x)$$

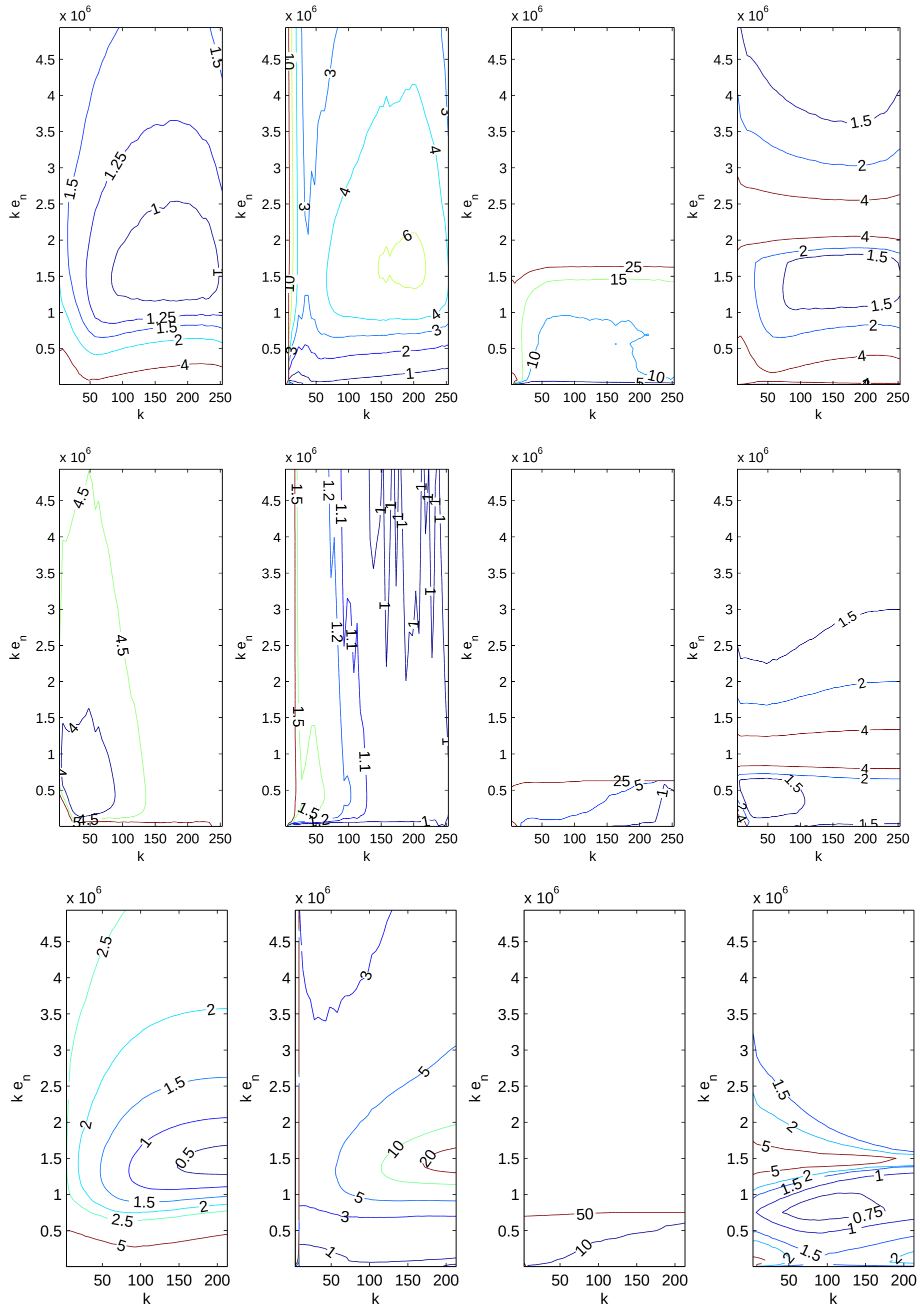

Figure 6. From left to right: contour plots of $10^4\times$RMSE, the ratio of RMSE of $\hat{p}_n^{(HS)}$ and $\hat{p}_n$, the ratio of standard errors of $\hat{p}_n$ and $\hat{p}_n^{(tm)}$, and the ratio of RMSE of $\hat{p}_n$ and $\hat{p}_n^{(tm)}$; from top to bottom: Gumbel $\vartheta = 5$ copula with GEV marginals with $\gamma = 0.25$ resp. $\gamma = -0.25$, Schlather type model with $\varrho = 0.8$

$$= \frac{d_n}{e_n} x \Bigg(1 + \left\{ \begin{array}{ll} -k^{-1/2} \log c_n (\frac{\Gamma_i}{\gamma_i} + o_P(1)) + O_P(k^{-1/2}(x d_n/e_n)^{-\gamma_i}), & \gamma_i > 0 \\ k^{-1/2}(d_n k/n)^{-\gamma_i} \big((\alpha_i/\gamma_i - \beta_i - \Gamma_i/\gamma_i^2 + o_P(1)) x^{-\gamma_i} + o_P(1) \big), & \gamma_i < 0 \\ -k^{-1/2} \log^2 c_n (\Gamma_i/2 + o_P(1)) + O_P(k^{-1/2} \log c_n \log x), & \gamma_i = 0 \end{array} \right. \Bigg)$$

uniformly for $x \in [\lambda_{n,i}, \tau_{n,i}]$.

Proof. For notational simplicity, we omit all indices and arguments of the marginal parameters and normalizing functions and their estimators; e.g., we use $\hat{a}$ as a short form of $\hat{a}_i(n/k)$. Moreover, we drop all indices referring to the ith marginal, i.e., we write U instead of U_i, T_n instead of $T_{n,i}$ and so on.

By (2.7), for all $0 < \varepsilon < |\rho|$,

$$\begin{aligned} \Delta_1(x) &:= \frac{U(d_n x) - b}{a} - \frac{(x d_n k/n)^\gamma - 1}{\gamma} \\ &= O\big(A(n/k)(x d_n k/n)^{\gamma+\rho\pm\varepsilon}\big) = o\big(k^{-1/2} w_n(\gamma)(x d_n k/n)^\gamma\big) \end{aligned} \tag{5.1}$$

uniformly for all $x \geq \lambda_n$, where in the last step we have used condition (i). Now one can conclude that $U(d_n x) \in \hat{T}_n((0,\infty))$ for all $x \in [\lambda_n, \tau_n]$ with probability tending to 1. For example, if $\gamma > 0$, then we have to show that $U(d_n x) > \hat{b} - \hat{a}/\hat{\gamma}$ for all $x \geq \lambda_n$ or, equivalently, (using (M3)) that $\Delta_1(\lambda_n)$ is larger than

$$\frac{\hat{b} - b}{a} - \frac{\hat{a}}{a\hat{\gamma}} - \frac{(\lambda_n d_n k/n)^\gamma - 1}{\gamma} = -\frac{1}{\gamma}\Big(\frac{d_n k}{n}\lambda_n\Big)^\gamma + O(k^{-1/2})$$

which follows immediately from (5.1), (S1) and (ii).

Hence

$$T_n^{\leftarrow} \circ \hat{T}_n \circ \hat{T}_n^{(c)\leftarrow} \circ U(d_n x) = \left[1 + \frac{\gamma}{a} \Big(\hat{a} \frac{\big(c_n^{-1}\big(1 + \hat{\gamma}\frac{U(d_n x) - \hat{b}}{\hat{a}}\big)^{1/\hat{\gamma}}\big)^{\hat{\gamma}} - 1}{\hat{\gamma}} + \hat{b} - b \Big) \right]^{1/\gamma} =: \tilde{H}(x)$$

if the expression in brackets is strictly positive, which will indeed follow from the calculations below.

Now direct calculations show that

$$\tilde{H}(x) = \left[1 + \gamma \Big(c_n^{-\hat{\gamma}} \frac{U(d_n x) - b}{a} + \frac{c_n^{-\hat{\gamma}} - 1}{\hat{\gamma}} \Big(\frac{\hat{a}}{a} - \frac{\hat{b} - b}{a}\hat{\gamma} \Big) \Big) \right]^{1/\gamma}. \tag{5.2}$$

By assumption (M3)

$$\Delta_2 := \frac{\hat{a}}{a} - \frac{\hat{b} - b}{a}\hat{\gamma} - 1 = k^{-1/2}(\alpha - \gamma\beta + o_P(1)). \tag{5.3}$$

If $\gamma > 0$, then the Taylor expansion

$$c_n^{-\hat\gamma/\gamma} = c_n^{-1}\Big(1 - k^{-1/2}\frac{\Gamma}{\gamma}\log c_n + o_P(k^{-1/2}\log c_n)\Big)$$

together with (5.1), (5.3) and (S1) implies

$$\begin{aligned}
\tilde H(x) &= \left[c_n^{-\hat\gamma}\Big(\Big(\frac{d_n k}{n}x\Big)^{\gamma} - 1 + \gamma\Delta_1(x) + \frac{\gamma}{\hat\gamma}(1+\Delta_2)\Big) + 1 - \frac{\gamma}{\hat\gamma} - \frac{\gamma}{\hat\gamma}\Delta_2\right]^{1/\gamma} \\
&= c_n^{-\hat\gamma/\gamma}\frac{d_n k}{n}x \times \\
&\quad \times\left[1 + O_P\Big(\big(|\Delta_1(x)| + k^{-1/2}\big)\Big(\frac{d_n k}{n}x\Big)^{-\gamma}\Big) + O_P\Big(k^{-1/2}c_n^{\hat\gamma}\Big(\frac{d_n k}{n}x\Big)^{-\gamma}\Big)\right]^{1/\gamma} \\
&= \frac{d_n}{e_n}x\Big(1 - k^{-1/2}\frac{\Gamma}{\gamma}\log c_n + o_P(k^{-1/2}\log c_n)\Big) \times \\
&\quad \times\left[1 + o_P(k^{-1/2}\log c_n) + O_P\Big(k^{-1/2}\Big(\frac{d_n}{e_n}x\Big)^{-\gamma}\Big)\right],
\end{aligned}$$

from which the assertion follows readily.

If $\gamma < 0$, then similar arguments prove

$$\begin{aligned}
\tilde H(x) &= \frac{d_n}{e_n}x\big(1 + O_P(k^{-1/2}\log c_n)\big) \times \\
&\quad \times\left[1 + k^{-1/2}\frac{1}{\gamma}\Big(\alpha - \gamma\beta - \frac{\Gamma}{\gamma} + o_P(1)\Big)\Big(\frac{d_n k}{n}x\Big)^{-\gamma} + o_P(k^{-1/2}w_n(\gamma)))\right],
\end{aligned}$$

and hence the assertion, because the assumption (iii) ensures that $\log c_n = o(w_n(\gamma))$.

Finally, for $\gamma = 0$, the Taylor expansion

$$c_n^{-\hat\gamma} = 1 - \hat\gamma\log c_n + \frac{1}{2}\hat\gamma^2\log^2 c_n + O_P(\hat\gamma^3\log^3 c_n)$$

yields

$$\begin{aligned}
\tilde H(x) &= \exp\Big[\big(1 - \hat\gamma\log c_n + O_P(k^{-1}\log^2 c_n)\big)\Big(\log\Big(\frac{d_n k}{n}x\Big) + \Delta_1(x)\Big) + \\
&\quad + \big(-\log c_n + \frac{1}{2}\hat\gamma\log^2 c_n + O_P(k^{-1}\log^3 c_n)\big)(1+\Delta_2)\Big] \\
&= \frac{d_n k}{c_n n}x\exp\Big[-\hat\gamma\log c_n\Big(\log c_n + \log\Big(\frac{d_n}{e_n}x\Big)\Big) + o_P(k^{-1/2}\log^2 c_n) + \frac{1}{2}\hat\gamma\log^2 c_n\Big] \\
&= \frac{d_n}{e_n}x\Big[1 - \frac{1}{2}(\Gamma + o_P(1))k^{-1/2}\log^2 c_n + O_P(k^{-1/2}\log c_n\log x)\Big],
\end{aligned}$$

which concludes the proof. □

In what follows we denote by $\lambda_{n,1} \searrow x_l$, $\lambda_{n,2} \searrow q(\infty) := \lim_{x\to\infty} q(x)$ and $\tau_{n,i} \uparrow \infty$, $i \in \{1,2\}$, sequences which satisfy the conditions of Lemma 5.1. (These sequences will be specified in the proof of Corollary 5.5.) Note that in particular constant sequences $\lambda_{n,i}, \tau_{n,i} \in (0,\infty)$ satisfy the conditions of Lemma 5.1, provided

$$A_i(n/k)c_n^{\rho_i+\varepsilon} = o\big(k_n^{-1/2} w_n(\gamma_i)\big) \quad \text{for } i \in \{1,2\} \text{ and some } \varepsilon > 0 \tag{5.4}$$

and (S1) holds. Therefore we may and will choose

$$\begin{aligned} \lambda_{n,1} &= x_l \quad && \text{if } x_l := \inf\{x \ge 0 \mid q(x) < \infty\} > 0, \\ \lambda_{n,2} &= q(\infty) \quad && \text{if } q(\infty) > 0, \end{aligned} \tag{5.5}$$

We want to apply the approximations just established to points (x,y) on the boundary of S. To ensure that $x \in [\lambda_{n,1}, \tau_{n,1}]$ and $y \in [\lambda_{n,2}, \tau_{n,2}]$, we consider a subset S_n^* of S that is bounded away from the coordinate axes. More precisely, we define

$$S_n^* := S \cap \big([u_n^*, \infty) \times [v_n^*, \infty)\big)$$

with

$$u_n^* := \lambda_{n,1} \vee q^{\leftarrow}(\tau_{n,2}), \quad v_n^* := \lambda_{n,2} \vee q(\tau_{n,1}).$$

The following lemma implies that the ν-measure of the set $S \setminus S_n^*$ is asymptotically negligible.

Lemma 5.2.

$$\nu(S) - \nu(S_n^*) = O\Big(\frac{\lambda_{n,1} - x_l}{q^2(\lambda_{n,1}-)} + \frac{q(\tau_{n,1}) - q(\infty)}{\tau_{n,1}^2} + \frac{\lambda_{n,2} - q(\infty)}{(q^{\leftarrow}(\lambda_{n,2}))^2} + \frac{q^{\leftarrow}(\tau_{n,2}) - x_l}{\tau_{n,2}^2}\Big)$$

with $q(x-) := \lim_{t\uparrow x} q(t)$.

Proof. First note that $S \subset [x_l, \infty) \times [q(\infty), \infty)$ implies

$$\nu(S) - \nu\big(S \cap ([0, u_n^*) \times [0,\infty))\big) \le \nu\big([x_l, \lambda_{n,1}) \times [q(\lambda_{n,1}-), \infty)\big) + \nu\big([x_l, q^{\leftarrow}(\tau_{n,2})) \times [\tau_{n,2}, \infty)\big).$$

The spectral density φ is assumed continuous and hence it is bounded. From (2.10) we conclude that for arbitrary $0 \le u_0 \le u_1$ and $v_0 > 0$

$$\nu\big([u_0, u_1) \times [v_0, \infty)\big) = O\Big(\int_{u_0}^{u_1} \int_{v_0}^{\infty} (u^2 + v^2)^{-3/2}\, dv\, du\Big) = O\Big(\frac{u_1 - u_0}{v_0^2}\Big)$$

and thus

$$\nu(S) - \nu\big(S \cap ([0, u_n^*) \times [0,\infty))\big) = O\Big(\frac{\lambda_{n,1} - x_l}{q^2(\lambda_{n,1}-)} + \frac{q^{\leftarrow}(\tau_{n,2}) - x_l}{\tau_{n,2}^2}\Big).$$

Likewise, one can show that

$$\nu\big(S \cap ([0, u_n^*) \times [0,\infty))\big) - \nu(S_n^*) = O\Big(\frac{q(\tau_{n,1}) - q(\infty)}{\tau_{n,1}^2} + \frac{\lambda_{n,2} - q(\infty)}{(q^{\leftarrow}(\lambda_{n,2}))^2}\Big).$$

A combination of these two bounds yields the assertion. □

On the set S_n^* we can now use the approximation from Lemma 5.1 to first examine the influence of the transformation $H_{n,2}$ of the second coordinate on the ν-measure of S_n^*. In a second step we then similarly determine how the ν-measure of this transformed set is altered by the transformation $H_{n,1}$ of the first coordinate. Hereby note that by Lemma 5.1 the marginal transformations are invertible with probability tending to 1.

Lemma 5.3. *Let $H_n(x,y) := (H_{n,1}(x), H_{n,2}(y)) := \frac{e_n}{d_n} T_n^{\leftarrow} \circ \hat{T}_n \circ \hat{T}_n^{(c)\leftarrow} \circ U(d_n x, d_n y)$. Suppose that the conditions (D2) and (Q1) are met.*

Then one has with $q_n(u) := q(u) \vee v_n^*$ and $\tilde{q}_n^{\leftarrow}(v) := q^{\leftarrow}(H_{n,2}^{\leftarrow}(v)) \vee u_n^*$

$$\begin{aligned}
&\Big|\nu(H_n(S_n^*)) - \nu(S_n^*) + \int_{u_n^*}^{\infty} \big(H_{n,2}(q_n(u)) - q_n(u)\big)\eta(u, q_n(u))\, du \\
&\qquad + \int_{H_{n,2}(v_n^*)}^{\infty} \big(H_{n,1}(\tilde{q}_n^{\leftarrow}(v)) - \tilde{q}_n^{\leftarrow}(v)\big)\eta(\tilde{q}_n^{\leftarrow}(v), v)\, dv\Big| \\
&= o\Big(\int_{u_n^*}^{\infty} \big|H_{n,2}(q_n(u)) - q_n(u)\big|\eta(u, q_n(u))\, du \\
&\qquad + \int_{H_{n,2}(v_n^*)}^{\infty} \big|H_{n,1}(\tilde{q}_n^{\leftarrow}(v)) - \tilde{q}_n^{\leftarrow}(v)\big|\eta(\tilde{q}_n^{\leftarrow}(v), v)\, dv\Big)
\end{aligned} \tag{5.6}$$

with probability tending to 1.

Proof. According to the proof of Lemma 5.1, for all $\delta \in (0,1)$, on the set $[\lambda_{n,i}(1-\delta), \tau_{n,i}(1+\delta)]$ the transformation $H_{n,i}$ is continuous and strictly increasing and $H_{n,i}(x) = x(1+o(1))$ with probability tending to 1.

We first quantify the influence of the transformation of the second coordinate. Note that

$$\begin{aligned}
\nu(S_n^*) &= \int_{u_n^*}^{\infty} \int_{q_n(u)}^{\infty} \eta(u,v)\, dv\, du \\
\nu\{(x, H_{n,2}(y)) \mid (x,y) \in S_n^*\} &= \int_{u_n^*}^{\infty} \int_{H_{n,2}(q_n(u))}^{\infty} \eta(u,v)\, dv\, du
\end{aligned}$$

and hence

$$\nu\{(x, H_{n,2}(y)) \mid (x,y) \in S_n^*\} - \nu(S_n^*) = -\int_{u_n^*}^{\infty} \int_{q_n(u)}^{H_{n,2}(q_n(u))} \eta(u,v)\, dv\, du. \tag{5.7}$$

The inner integral equals

$$\begin{aligned}
&\int_{q_n(u)}^{H_{n,2}(q_n(u))} \eta(u,v)\, dv \\
&= \big(H_{n,2}(q_n(u)) - q_n(u)\big)\eta(u, q_n(u))
\end{aligned} \tag{5.8}$$

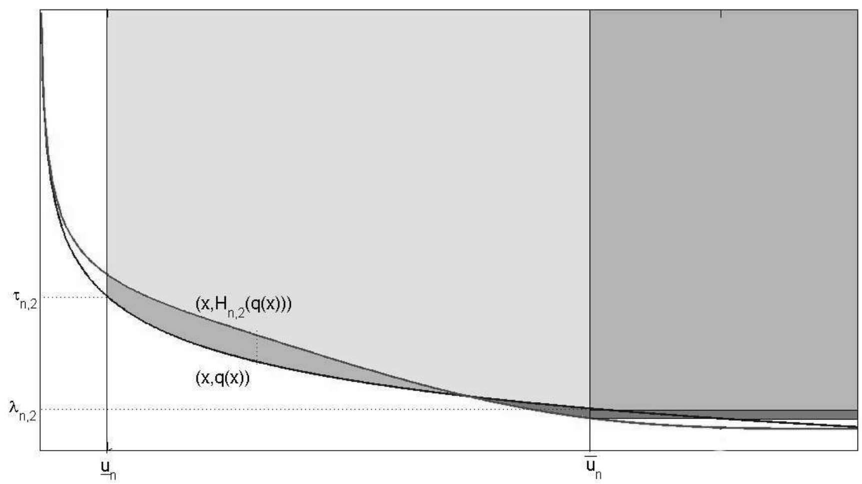

Figure 7. The light and mid grey regions show the approximation S_n^* of the set S, the mid and the dark grey regions the symmetric difference between $\{(x, H_{n,2}(y)) \mid (x, y) \in S_n^*\}$ and S_n^*, where the dark grey region is counted with a positive sign, the mid grey region with a negative sign. (Here it is assumed that $u_n^* = q^{\leftarrow}(\tau_{n,2})$.)

$$+\int_1^{H_{n,2}(q_n(u))/q_n(u)} \Big(\frac{\eta(u, q_n(u)w)}{\eta(u, q_n(u))} - 1\Big)\, dw\; \eta(u, q_n(u))q_n(u).$$

By the assumptions and Lemma 5.1, $H_{n,2}(q_n(u))/q_n(u) \to 1$ uniformly for $u \in (u_n^*, \infty)$ as $q_n(u) \in [\lambda_{n,2}, \tau_{n,2}]$ for $u > u_n^*$.

Next note that for all $0 < \delta < \pi/4$

$$\sup_{0<t\le\tan\delta} \Big|\frac{\arctan(tw)}{\arctan t} - 1\Big| \le \sup_{0<t\le\tan\delta} \frac{t|w-1|}{\arctan t} \le (1 + \tan^2\delta)|w - 1| \to 0$$

as $w \to 1$, and likewise by symmetry

$$\sup_{t\ge\tan(\pi/2-\delta)} \Big|\frac{\pi/2 - \arctan(tw)}{\pi/2 - \arctan t} - 1\Big| = \sup_{0<t\le\tan\delta} \Big|\frac{\arctan(t/w)}{\arctan t} - 1\Big| \to 0.$$

Therefore, by condition (2.9), to each $\varepsilon > 0$ there exists $\delta > 0$ such that for all $t \in (0, \tan\delta] \cup [\tan(\pi/2 - \delta), \infty)$

$$\Big|\frac{\varphi(\arctan(tw))}{\varphi(\arctan t)} - 1\Big| < \varepsilon.$$

Moreover, by the uniform continuity of $\varphi \circ \arctan$ on $[\tan\delta, (\tan(\pi/2 - \delta]$

$$\sup_{\tan\delta\le t\le\tan(\pi/2-\delta)} \Big|\frac{\varphi(\arctan(tw))}{\varphi(\arctan t)} - 1\Big|$$

$$\leq \frac{\sup_{\tan\delta\leq t\leq \tan(\pi/2-\delta)} |\varphi(\arctan(tw))-\varphi(\arctan t)|}{\inf_{\delta\leq u\leq \pi/2-\delta}\varphi(u)} \to 0$$

as $w\to 1$. Using (2.10) and

$$\frac{1+t^2}{1+t^2w^2} = 1+\frac{1-w^2}{t^{-2}+w^2}\to 1$$

as $w\to 1$ uniformly for $t>0$, we conclude that

$$\frac{\eta(u,vw)}{\eta(u,v)} = \Big(\frac{1+(v/u)^2}{1+(v/u)^2w^2}\Big)^{3/2}\frac{\varphi\big(\arctan\frac{vw}{u}\big)}{\varphi\big(\arctan\frac{v}{u}\big)} \;\to\; 1$$

as $w\to 1$ uniformly for $u,v>0$. Thus,

$$\int_1^{H_{n,2}(q_n(u))/q_n(u)} \Big(\frac{\eta(u,q_n(u)w)}{\eta(u,q_n(u))}-1\Big)\,dw = o\big(H_{n,2}(q_n(u))/q_n(u)-1\big)$$

which, combined with (5.7) and (5.8), yields

$$\begin{aligned}\nu\{(x,H_{n,2}(y)) \mid (x,y)\in S_n^*\} - \nu(S_n^*) + \int_{u_n^*}^\infty \big(H_{n,2}(q_n(u))-q_n(u)\big)\eta(u,q_n(u))\,du \\ = \quad o\Big(\int_{u_n^*}^\infty \big|H_{n,2}(q_n(u))-q_n(u)\big|\eta(u,q_n(u))\,du\Big). \qquad (5.9)\end{aligned}$$

One can derive an analogous approximation of the difference between $\nu\{(x,H_{n,2}(y)) \mid (x,y)\in S\}$ and $\nu\{(H_{n,1}(x),H_{n,2}(y)) \mid (x,y)\in S\}$ by similar arguments if one interchanges the order of integration:

$$\begin{aligned}\Big|\nu(H_n(S_n^*)) - \nu\{(x,H_{n,2}(y)) \mid (x,y)\in S_n^*\} \\ + \int_{H_{n,2}(v_n^*)}^\infty \big(H_{n,1}(\tilde q_n^{\leftarrow}(v))-\tilde q_n^{\leftarrow}(v)\big)\eta(\tilde q_n^{\leftarrow}(v),v)\,dv\Big| \\ = \quad o\Big(\int_{H_{n,2}(v_n^*)}^\infty \big|H_{n,1}(\tilde q_n^{\leftarrow}(v))-\tilde q_n^{\leftarrow}(v)\big|\eta(\tilde q_n^{\leftarrow}(v),v)\,dv\Big). \qquad (5.10)\end{aligned}$$

Summing up (5.9) and (5.10), we arrive at the assertion. □

In the next lemma, we calculate the limits of the integrals arising in Lemma 5.3 using the approximation established in Lemma 5.1.

Lemma 5.4. *Suppose that the conditions of Lemma 5.3 and, in addition, the following conditions are fulfilled for some $x_0\in(x_l,q^{\leftarrow}(q(\infty)))$, $y_0\in(q(\infty),q(x_l))$:*

$$\int_{y_0}^\infty (q^{\leftarrow}(v))^{1-\gamma_1}v^{-3}\,dv<\infty \quad \text{or} \quad \lambda_{n,1}^{1-\gamma_1}=o(\log c_n) \qquad (5.11)$$

$$\int_{x_0}^{\infty} (q(u))^{1-\gamma_2} u^{-3}\, du < \infty \qquad \text{or} \quad \lambda_{n,2}^{1-\gamma_2} = o(\log c_n) \tag{5.12}$$

Then the following approximations hold true:

(i)

$$\frac{k^{1/2}}{w_n(\gamma_2)} \int_{u_n^*}^{\infty} \big(H_{n,2}(q_n(u)) - q_n(u)\big)\eta(u, q_n(u))\, du$$
$$\to \begin{cases} -\frac{\Gamma_2}{\gamma_2} \int_{x_l}^{\infty} q(u)\eta(u, q(u))\, du, & \gamma_2 > 0 \\ \big(\frac{\alpha_2}{\gamma_2} - \beta_2 - \frac{\Gamma_2}{\gamma_2^2}\big) \int_{x_l}^{\infty} (q(u))^{1-\gamma_2}\eta(u, q(u))\, du, & \gamma_2 < 0 \\ -\Gamma_2 \int_{x_l}^{\infty} q(u)\eta(u, q(u))\, du, & \gamma_2 = 0 \end{cases}$$

Moreover,

$$\int_{u_n^*}^{\infty} \big|H_{n,2}(q_n(u)) - q_n(u)\big|\eta(u, q_n(u))\, du = O(k^{-1/2} w_n(\gamma_2)).$$

(ii)

$$\frac{k^{1/2}}{w_n(\gamma_1)} \int_{H_{n,2}(v_n^*)}^{\infty} \big(H_{n,1}(\tilde{q}_n^{\leftarrow}(v)) - \tilde{q}_n^{\leftarrow}(v)\big)\eta(\tilde{q}_n^{\leftarrow}(v), v)\, dv$$
$$\to \begin{cases} -\frac{\Gamma_1}{\gamma_1} \int_{q(\infty)}^{\infty} q^{\leftarrow}(v)\eta(q^{\leftarrow}(v), v)\, dv, & \gamma_1 > 0 \\ \big(\frac{\alpha_1}{\gamma_1} - \beta_1 - \frac{\Gamma_1}{\gamma_1^2}\big) \int_{q(\infty)}^{\infty} (q^{\leftarrow}(v))^{1-\gamma_1}\eta(q^{\leftarrow}(v), v)\, dv, & \gamma_1 < 0 \\ -\Gamma_1 \int_{q(\infty)}^{\infty} q^{\leftarrow}(v)\eta(q^{\leftarrow}(v), v)\, dv, & \gamma_1 = 0 \end{cases}$$

Furthermore,

$$\int_{H_{n,2}(v_n^*)}^{\infty} \big|H_{n,1}(\tilde{q}_n^{\leftarrow}(v)) - \tilde{q}_n^{\leftarrow}(v)\big|\eta(\tilde{q}_n^{\leftarrow}(v), v)\, dv = O(k^{-1/2} w_n(\gamma_1)).$$

***Proof.* ad (i)**: Because the spectral density φ is bounded, there exists a constant $K > 0$ such that

$$\eta(u, q(u)) \le K\big(u^2 + (q(u))^2\big)^{-3/2} \le K\big(u^{-3} \wedge (q(u))^{-3}\big) \quad \forall\, u > 0. \tag{5.13}$$

Hence $q_n(u)\eta(u, q_n(u)) \le K(q(u))^{-2}$ for $u \in [x_l, x_0]$ and n sufficiently large and $q_n(u)$ $\eta(u, q_n(u)) \le K q(x_0) u^{-3}$ for $u > x_0$. Therefore,

$$\lim_{n\to\infty} \int_{u_n^*}^{\infty} q_n(u)\eta(u, q_n(u))\, du = \int_{x_l}^{\infty} q(u)\eta(u, q(u))\, du < \infty \tag{5.14}$$

by the dominated convergence theorem and $u_n^* \downarrow x_l$.

Now, we distinguish three cases.

If $\underline{\gamma_2 > 0}$, then by Lemma 5.1 and $d_n \asymp e_n$

$$\begin{aligned}
&\frac{k^{1/2}}{\log c_n}\int_{u_n^*}^{\infty}\big(H_{n,2}(q_n(u)) - q_n(u)\big)\eta(u, q_n(u))\,du \\
&= -\Big(\frac{\Gamma_2}{\gamma_2} + o_P(1)\Big)\int_{u_n^*}^{\infty} q_n(u)\eta(u, q_n(u))\,du \\
&\quad + O_P\Big(\frac{1}{\log c_n}\int_{u_n^*}^{\infty}(q_n(u))^{1-\gamma_2}\eta(u, q_n(u))\,du\Big).
\end{aligned}$$

Because of (5.13) and (5.12)

$$\begin{aligned}
&\int_{u_n^*}^{\infty}(q_n(u))^{1-\gamma_2}\eta(u, q_n(u))\,du \\
&\le K(q(x_0))^{-2-\gamma_2}(x_0 - x_l) + K\int_{x_0}^{\infty}(q(u)\vee\lambda_{n,2})^{1-\gamma_2}u^{-3}\,du \\
&= o(\log c_n). \qquad (5.15)
\end{aligned}$$

Hence, in view of (5.14), we have

$$\begin{aligned}
&\int_{u_n^*}^{\infty}\big(H_{n,2}(q_n(u)) - q_n(u)\big)\eta(u, q_n(u))\,du \\
&= -k^{-1/2}\log c_n\frac{\Gamma_2}{\gamma_2}\int_{x_l}^{\infty} q(u)\eta(u, q(u))\,du + o_P(k^{-1/2}\log c_n).
\end{aligned}$$

If $\underline{\gamma_2 < 0}$, then the assertion follows similarly from Lemma 5.1 and (5.13).

Finally, in the case $\underline{\gamma_2 = 0}$

$$\begin{aligned}
&\int_{x_l}^{\infty} q_n(u)|\log q_n(u)|\eta(u, q_n(u))\,du \\
&\le K\sup_{x\le x_0}\frac{|\log q(x)|}{(q(x))^2} + K\sup_{x\ge x_0} q(x)|\log q(x)|\int_{x_0}^{\infty}u^{-3}\,du < \infty. \qquad (5.16)
\end{aligned}$$

Hence, similarly as in the first case, we may conclude the assertion from Lemma 5.1.

ad (ii): The second assertion can be proved in a very similar fashion using $q(H_{n,2}(v_n^*)) \to q(\infty)$ and the fact that $\tilde{q}_n^{\leftarrow}(u) \to q^{\leftarrow}(u)$ for Lebesgue-almost all $u > q(\infty)$, because of Lemma 5.1 and the Lebesgue-almost surely continuity of $q^{\leftarrow}$. For that reason, we only give the analog to the bound (5.15) for the integral under consideration in the case $\gamma_1 > 0$.

For $y_0 \in (q(\infty), q(x_l))$ and all sufficiently large n, we have

$$\int_{H_{n,2}(v_n^*)}^{y_0}(\tilde{q}_n^{\leftarrow}(v))^{1-\gamma_1}\eta\big(\tilde{q}_n^{\leftarrow}(v), v\big)\,dv \le K(\tilde{q}_n^{\leftarrow}(y_0))^{-2-\gamma_1}(y_0 - q(\infty)) = O(1).$$

If $\gamma_1 \le 1$, then

$$\int_{y_0}^{\infty} (\tilde{q}_n^{\leftarrow}(v))^{1-\gamma_1} \eta\big(\tilde{q}_n^{\leftarrow}(v), v\big)\, dv \le K(\tilde{q}_n^{\leftarrow}(y_0))^{1-\gamma_1} \int_{y_0}^{\infty} v^{-3}\, dv = O(1).$$

Finally, if $\gamma_1 > 1$, then by the monotonicity of $q^{\leftarrow}$ and the asymptotic behavior of $H_{n,2}$ we have for all $\delta > 0$ and sufficiently large n

$$\begin{aligned} &\int_{y_0}^{\infty} (\tilde{q}_n^{\leftarrow}(v))^{1-\gamma_1} \eta\big(\tilde{q}_n^{\leftarrow}(v), v\big)\, dv \\ &\le K \int_{y_0}^{\infty} \Big(\big(q^{\leftarrow}(v(1+\delta))\big)^{1-\gamma_1} \wedge \lambda_{n,1}^{1-\gamma_1} \Big) v^{-3}\, dv \\ &= O\Big(\int_{y_0(1+\delta)}^{\infty} (q^{\leftarrow}(v))^{1-\gamma_1} v^{-3}\, dv \wedge \lambda_{n,1}^{1-\gamma_1} \Big) = o(\log c_n) \end{aligned}$$

by condition (5.11). □

The following result gives sufficient conditions such that the difference between the ν-measure of S and of the truncated set after the marginal transformations (i.e. $d_n(IV+V)$ in (2.5)) can be approximated by the limiting terms in Lemma 5.4. For the sake of simplicity, we assume that d_n and e_n are of the same order, but it is not difficult to prove similar results under weaker conditions on d_n/e_n. Moreover, one can weaken the condition (S2) and the assumptions (Q2) could be replaced with rather strong conditions on the rate at which k tends to ∞.

Corollary 5.5. *If the conditions (M1)–(M3), (D2), (Q1), (Q2) and (S1)–(S3) are fulfilled, then*

$$\begin{aligned} &\nu(H_n(S_n^*)) - \nu(S) \\ &= k^{-1/2} w_n(\gamma_1) \begin{cases} -\frac{\Gamma_1}{\gamma_1} \int_{q(\infty)}^{\infty} q^{\leftarrow}(v) \eta(q^{\leftarrow}(v), v)\, dv, & \gamma_1 > 0 \\ \big(\frac{\alpha_1}{\gamma_1} - \beta_1 - \frac{\Gamma_1}{\gamma_1^2}\big) \int_{q(\infty)}^{\infty} (q^{\leftarrow}(v))^{1-\gamma_1} \eta(q^{\leftarrow}(v), v)\, dv, & \gamma_1 < 0 \\ -\Gamma_1 \int_{q(\infty)}^{\infty} q^{\leftarrow}(v) \eta(q^{\leftarrow}(v), v)\, dv, & \gamma_1 = 0 \end{cases} \\ &\quad + k^{-1/2} w_n(\gamma_2) \begin{cases} -\frac{\Gamma_2}{\gamma_2} \int_{x_l}^{\infty} q(u) \eta(u, q(u))\, du, & \gamma_2 > 0 \\ \big(\frac{\alpha_2}{\gamma_2} - \beta_2 - \frac{\Gamma_2}{\gamma_2^2}\big) \int_{x_l}^{\infty} (q(u))^{1-\gamma_2} \eta(u, q(u))\, du, & \gamma_2 < 0 \\ -\Gamma_2 \int_{x_l}^{\infty} q(u) \eta(u, q(u))\, du, & \gamma_2 = 0 \end{cases} \\ &\quad + o_P\big(k^{-1/2}(w_n(\gamma_1) \vee w_n(\gamma_2))\big) \end{aligned} \tag{5.17}$$

Proof. In view of the Lemmas 5.2 – 5.4, it suffices to define sequences $\lambda_{n,i}$ and $\tau_{n,i}$, $i \in \{1, 2\}$, such that the conditions (i)–(iv) of Lemma 5.1 and (5.11) and (5.12) are fulfilled and

$$\frac{\lambda_{n,1} - x_l}{q^2(\lambda_{n,1}-)} + \frac{q(\tau_{n,1}) - q(\infty)}{\tau_{n,1}^2} = o\big(k^{-1/2} w_n(\gamma_1)\big),$$

$$\frac{\lambda_{n,2}-q(\infty)}{(q^{\leftarrow}(\lambda_{n,2}))^2}+\frac{q^{\leftarrow}(\tau_{n,2})-x_l}{\tau_{n,2}^2}=o\big(k^{-1/2}w_n(\gamma_2)\big).$$

Note that we can check these conditions for $i=1$ and $i=2$ separately. We focus on the sequences $\lambda_{n,1}$ and $\tau_{n,1}$, since the case $i=2$ can be treated analogously if x_l is replaced with $q(\infty)$ and q with $q^{\leftarrow}$. Again we distinguish three cases depending on the sign of γ_1.

If $\underline{\gamma_1>0}$, then $\tau_{n,1}$ must only satisfy $(q(\tau_{n,1})-q(\infty))/\tau_{n,1}^2=o(k^{-1/2}\log c_n)$, which can easily be fulfilled by letting $\tau_{n,1}$ tend to ∞ sufficiently fast.

The sequence $\lambda_{n,1}$ has to satisfy the conditions (i) and (ii) of Lemma 5.1, (5.11) and $(\lambda_{n,1}-x_l)/q^2(\lambda_{n,1})=o(k^{-1/2}\log c_n)$. If $x_l>0$, then $\lambda_{n,1}=x_l$ does the job, because condition (i) of Lemma 5.1 is implied by (S2).

If $x_l=0$ and $\gamma_1\le 1$, then the integrability condition of (5.11) is trivial. Moreover, $\lambda_{n,1}:=k^{-1/2}(\log c_n)^{1/2}\to 0$ obviously fulfills 5.1 (ii) and $(\lambda_{n,1}-x_l)/q^2(\lambda_{n,1})=O(\lambda_{n,1})=o(k^{-1/2}\log c_n)$. Condition 5.1 (i) follows from (S2) and (S3), which implies $c_n\lambda_{n,1}\to\infty$.

Finally, if $x_l=0$ and $\gamma_1>1$, then $\lambda_{n,1}:=(k^{-1/2}\log c_n)^{1/\gamma_1}$ fulfills 5.1 (ii), 5.1 (i) follows from (S2) and (S3) as above, and (Q2) implies

$$\frac{\lambda_{n,1}-x_l}{q^2(\lambda_{n,1})}=O\Big(\frac{\lambda_{n,1}^{\gamma_1}}{|\log\lambda_{n,1}|^2}\Big)=O\Big(k^{-1/2}\frac{\log c_n}{|\log(k^{-1/2}\log c_n)|^2}\Big)=o(k^{-1/2}\log c_n)$$

by (S1). Furthermore, the integrability condition of (5.11) is fulfilled, because (Q2) implies $(v/\log v)^{2/(1-\gamma_1)}=O(q^{\leftarrow}(v))$ as $v\to\infty$.

Next we consider the case $\underline{-1/2<\gamma_1<0}$, when the integrability condition of (5.11) is trivial. If $x_l>0$, then we can argue as above that $\lambda_{n,1}=x_l$ satisfies all conditions on $\lambda_{n,1}$. If $x_l=0$, then define $\lambda_{n,1}=c_n^{-1}\varphi_n$ for some $\varphi_n\to\infty$ sufficiently slowly, so that 5.1 (i) follows from (S2). Further $(\lambda_{n,1}-x_l)/q^2(\lambda_{n,1})=O(c_n^{-1}\varphi_n)=o(k^{-1/2}c_n^{-\gamma_1})$ follows from assumption (S3).

The conditions on $\tau_{n,1}$ read as $(q(\tau_{n,1})-q(\infty))/\tau_{n,1}^2=o(k^{-1/2}c_n^{-\gamma_1})$ and $k^{-1/2}=o((c_n\tau_{n,1})^{\gamma_1})$ in this case, which are fulfilled by $\tau_{n,1}=k^{1/2}c_n^{\gamma_1}\to\infty$.

In the case $\underline{\gamma_1=0}$ the integrability condition of (5.11) is again trivial and $\lambda_{n,1}=x_l$ if $x_l>0$, and $\lambda_{n,1}=c_n^{-1}\log c_n$ if $x_l=0$ does the job. Moreover, it is easily checked that $\tau_{n,1}=k^{1/4}$ satisfies $(q(\tau_{n,1})-q(\infty))/\tau_{n,1}^2=o(k^{-1/2}\log^2 c_n)$ and condition 5.1 (iv). □

Observe that we have verified stronger conditions on $\lambda_{n,1}$ and $\tau_{n,1}$ than actually necessary, if $w_n(\gamma_1)=o(w_n(\gamma_2))$. A refined analysis would lead to weaker, but more complex conditions on q and k that depend on both the values of γ_1 and γ_2 at the same time. (Also the proof would become more lengthy as one had to consider 9 cases arising from different combinations of signs of γ_1 and γ_2.) Moreover, note that for the above choice of $\lambda_{n,i}$ one has

$$c_n\lambda_{n,i}\to\infty,\quad i\in\{1,2\}, \tag{5.18}$$

and

$$\lambda_{n,i}^{-\gamma_i}=O(k^{1/2}/\log c_n)\quad \text{if } \gamma_i>0,\qquad i\in\{1,2\}. \tag{5.19}$$

Now we use classical empirical process theory to establish a uniform bound on $\nu_n(B) - E\nu_n(B)$ and thus on term II in decomposition (2.5).

Lemma 5.6. *Under the conditions of Theorem 2.1, one has*

$$\nu_n(B) - E\nu_n(B)|_{B=(d_n/e_n)H_n(S_n^*)} = o_P\big(k^{-1/2}(w_n(\gamma_1) \vee w_n(\gamma_2))\big).$$

Proof. Note that by (5.2) one has

$$\frac{d_n}{e_n} H_n(x_1, x_2) = \big(\tilde{H}^{(n,1)}_{\vartheta_1,\chi_1,\xi_1}(x_1), \tilde{H}^{(n,2)}_{\vartheta_2,\chi_2,\xi_2}(x_2)\big)$$

for $(x_1, x_2) \in [u_n^*, \infty) \times [v_n^*, \infty)$ with $\tilde{H}^{(n,i)}_{\vartheta_i,\chi_i,\xi_i}$ defined by (2.8) and

$$-\vartheta_i = \chi_i = k^{1/2}(\hat{\gamma}_i - \gamma_i), \quad \xi_i = k^{1/2}\Big(\frac{\hat{a}_i(n/k)}{a_i(n/k)} - 1 - \frac{\hat{b}_i(n/k) - b_i(n/k)}{a_i(n/k)}\hat{\gamma}_i\Big).$$

Since, according to condition (M3), these random variables are stochastically bounded, it suffices to prove that for all $M > 0$

$$\sup_{\max(|\vartheta_i|,|\chi_i|,|\xi_i|)\le M} \Big|\nu_n\big(E^{(n)}_{(\vartheta_i,\chi_i,\xi_i)_{i=1,2}}\big) - E\nu_n\big(E^{(n)}_{(\vartheta_i,\chi_i,\xi_i)_{i=1,2}}\big)\Big| = o_P\big(k^{-1/2}(w_n(\gamma_1) \vee w_n(\gamma_2))\big)$$

where

$$E^{(n)}_{(\vartheta_i,\chi_i,\xi_i)_{i=1,2}} := \big\{\big(\tilde{H}^{(n,1)}_{\vartheta_1,\chi_1,\xi_1}(x_1), \tilde{H}^{(n,2)}_{\vartheta_2,\chi_2,\xi_2}(x_2)\big) \mid (x_1, x_2) \in S_n^*\big\}.$$

Letting $\theta := (\vartheta_i, \chi_i, \xi_i)_{i=1,2}$ and

$$Z_n(\theta) := \frac{k^{1/2}}{w_n(\gamma_1) \vee w_n(\gamma_2)}\big(\nu_n(E^{(n)}_\theta) - E\nu_n(E^{(n)}_\theta)\big), \quad \theta \in [-M, M]^6,$$

we have to prove that Z_n tends to 0 in probability uniformly. To this end, we establish asymptotic equicontinuity of Z_n, i.e.

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} P\Big\{\sup_{\theta,\psi\in[-M,M]^6, \|\theta-\psi\|_\infty \le \delta} |Z_n(\theta) - Z_n(\psi)| > \eta\Big\} = 0 \quad \forall\, \eta > 0, \tag{5.20}$$

and convergence in probability of $Z_n(\theta)$ for all $\theta \in [-M, M]^6$ (see van der Vaart and Wellner, 2000, Theorem 1.5.7).

For the proof of asymptotic equicontinuity, it is crucial that the functions $\tilde{H}^{(n,i)}_{\vartheta_i,\chi_i,\xi_i}(x_i)$ are decreasing in all three parameters for all $(x_1, x_2) \in [u_n^*, \infty) \times [v_n^*, \infty)$. For ξ_i resp. ϑ_i this monotonicity is an immediate consequence of the facts that $(c_n^{-\gamma} - 1)/\gamma$ is negative and increasing in γ (for $c_n > 1$) and that $(1 + \gamma t)^{1/\gamma}$ is increasing in t. Because $c_n^{-\gamma}$ is a decreasing function of γ, the monotonicity in χ_i follows from (2.7), (5.18) and condition (i) of Lemma 5.1, which imply

$$\frac{U_i(d_n x) - b_i(n/k)}{a_i(n/k)} = \frac{(x_i d_n k/n)^{\gamma_i} - 1}{\gamma_i} + O\big(A_i(n/k)(x_i d_n k/n)^{\gamma_i+\rho_i+\varepsilon}\big)$$

$$\begin{aligned} &= \frac{(x_i c_n d_n/e_n)^{\gamma_i} - 1}{\gamma_i} + o\big((c_n x_i)^{\gamma_i} k^{-1/2} w_n(\gamma_i)\big) \\ &> 0 \end{aligned}$$

for sufficiently large n.

The monotonicity of $H^{(n,i)}_{\cdot,\cdot,\cdot}(x_i)$ implies that the sets $E^{(n)}_{(\vartheta_i,\chi_i,\xi_i)_{i=1,2}}$ are increasing in all parameters. Hence, for arbitrary $\theta,\psi \in [-M,M]^6$

$$|Z_n(\theta) - Z_n(\psi)| \le \frac{k^{1/2}}{w_n(\gamma_1)\vee w_n(\gamma_2)}\big(\nu_n(E^{(n)}_{\theta\vee\psi}\setminus E^{(n)}_{\theta\wedge\psi}) + E\nu_n(E^{(n)}_{\theta\vee\psi}\setminus E^{(n)}_{\theta\wedge\psi})\big)$$

where $\theta\vee\psi$ resp. $\theta\wedge\psi$ denote the coordinatewise maximum resp. minimum of θ and ψ.

To establish asymptotic equicontinuity of Z_n, we cover the parameter space $[-M,M]^6$ with hypercubes $I_l := \times_{i=1}^6 [l_i\delta,(l_i+1)\delta]$, $-\lceil M/\delta\rceil \le l_i \le \lfloor M/\delta\rfloor$, for some small $\delta > 0$ (depending on the value η in (5.20)) to be specified later on. For $\theta,\psi \in [-M,M]^6$ with $\|\theta-\psi\|_\infty \le \delta$ and $l(\theta) := \big(\lfloor\theta_i/\delta\rfloor\big)_{1\le i\le 6}$, one has $\|l(\theta)-l(\psi)\| \le 1$ and thus

$$\begin{aligned} &|Z_n(\theta) - Z_n(\psi)| \\ &\quad\le |Z_n(\theta) - Z_n(l(\theta)\delta)| + |Z_n(\psi) - Z_n(l(\psi)\delta)| + |Z_n(l(\theta)\delta) - Z_n(l(\psi)\delta)| \\ &\quad\le 3 \max_{l\in\{-\lceil M/\delta\rceil,\dots,\lfloor M/\delta\rfloor\}^6} \sup_{t,u\in I_l} |Z_n(t) - Z_n(u)| \\ &\quad\le 3\frac{k^{1/2}}{w_n(\gamma_1)\vee w_n(\gamma_2)} \max_{l\in\{-\lceil M/\delta\rceil,\dots,\lfloor M/\delta\rfloor\}^6} \big(\nu_n(E^{(n)}_{(l+1)\delta}\setminus E^{(n)}_{l\delta}) \\ &\qquad\qquad + E\nu_n(E^{(n)}_{(l+1)\delta}\setminus E^{(n)}_{l\delta})\big) \end{aligned} \tag{5.21}$$

where $(l+1)\delta := ((l_i+1)\delta)_{1\le i\le 6}$. By (D1), the expectation can be approximated as follows:

$$E\nu_n(E^{(n)}_{(l+1)\delta}\setminus E^{(n)}_{l\delta}) = \frac{n}{k}P\{T_n^{\leftarrow}(X,Y)\in E^{(n)}_{(l+1)\delta}\setminus E^{(n)}_{l\delta}\} = \nu\big(E^{(n)}_{(l+1)\delta}\setminus E^{(n)}_{l\delta}\big) + O(A_0(n/k)). \tag{5.22}$$

To bound the right-hand side, first note that by similar calculations as in the proof of Lemma 5.1, one obtains

$$\begin{aligned} &\tilde H^{(n,i)}_{\vartheta_i,\chi_i,\xi_i}(x) \\ &\quad= \frac{d_n}{e_n}x\Bigg(1 + \\ &\qquad + \begin{cases} -k^{-1/2}\log c_n(\frac{\chi_i}{\gamma_i} + o_P(1)) + O_P(k^{-1/2}(xd_n/e_n)^{-\gamma_i}), & \gamma_i > 0 \\ k^{-1/2}(d_nk/n)^{-\gamma_i}\big((\xi_i/\gamma_i + \vartheta_i/\gamma_i^2 + o_P(1))x^{-\gamma_i} + o_P(1)\big), & \gamma_i < 0 \\ -k^{-1/2}\log^2 c_n(\chi_i + \vartheta_i/2 + o_P(1)) + O_P(k^{-1/2}\log c_n \log x), & \gamma_i = 0 \end{cases}\Bigg) \end{aligned}$$

uniformly for $x \in [\lambda_{n,i},\tau_{n,i}]$. That means that under the same conditions as in Lemma 5.1 one can prove an analogous approximation where Γ_i is replaced with χ_i if $\gamma_i > 0$,

$\alpha_i/\gamma_i - \beta_i - \Gamma_i/\gamma_i^2$ is replaced with $\xi_i/\gamma_i + \vartheta_i/\gamma_i^2$ if $\gamma_i < 0$, and Γ_i is replaced with $2\chi_i + \vartheta_i$ in the case $\gamma_i = 0$. Hence, we may also conclude a corresponding analog to Corollary 5.5, i.e. $\nu((e_n/d_n)E^{(n)}_{(\vartheta_i,\chi_i,\xi_i)_{i=1,2}}) - \nu(S)$ equals the right-hand side of (5.17) with the above substitutions. Because all integrals are finite, there exists a constant $K > 0$ such that for sufficiently large n

$$\nu(E^{(n)}_{(l+1)\delta}) - \nu(E^{(n)}_{l\delta}) \le \frac{e_n}{d_n} K\delta k^{-1/2}(w_n(\gamma_1) \vee w_n(\gamma_2))$$

uniformly for all $l \in \{-\lceil M/\delta\rceil \dots \lfloor M/\delta\rfloor\}^6$. A combination with (5.22), $e_n \asymp d_n$ and condition (S2) shows that to each $\eta > 0$ there exists $\delta > 0$ such that for sufficiently large n

$$E\nu_n(E^{(n)}_{(l+1)\delta} \setminus E^{(n)}_{l\delta}) \le \frac{\eta}{12} k^{-1/2}(w_n(\gamma_1) \vee w_n(\gamma_2)). \tag{5.23}$$

In view of (5.21), we obtain

$$\begin{aligned}
&P\Big\{\sup_{\theta,\psi\in[-M,M]^6,\|\theta-\psi\|_\infty\le\delta} |Z_n(\theta) - Z_n(\psi)| > \eta\Big\} \\
&\le P\Big\{\max_{l\in\{-\lceil M/\delta\rceil\dots\lfloor M/\delta\rfloor\}^6} \big(\nu_n(E^{(n)}_{(l+1)\delta} \setminus E^{(n)}_{l\delta}) + E\nu_n(E^{(n)}_{(l+1)\delta} \setminus E^{(n)}_{l\delta})\big) \\
&\qquad\qquad > \frac{\eta}{3} k^{-1/2}(w_n(\gamma_1) \vee w_n(\gamma_2))\Big\} \\
&\le \sum_{l\in\{-\lceil M/\delta\rceil\dots\lfloor M/\delta\rfloor\}^6} P\Big\{\big|\nu_n(E^{(n)}_{(l+1)\delta} \setminus E^{(n)}_{l\delta}) - E\nu_n(E^{(n)}_{(l+1)\delta} \setminus E^{(n)}_{l\delta})\big| \\
&\qquad\qquad > \frac{\eta}{6} k^{-1/2}(w_n(\gamma_1) \vee w_n(\gamma_2))\Big\}.
\end{aligned}$$

Therefore the asserted asymptotic equicontinuity (5.20) follows from (5.23) and Chebyshev's inequality applied to the binomial random variables $k\nu_n(E^{(n)}_{(l+1)\delta} \setminus E^{(n)}_{l\delta})$:

$$\begin{aligned}
&P\Big\{\big|\nu_n(E^{(n)}_{(l+1)\delta} \setminus E^{(n)}_{l\delta}) - E\nu_n(E^{(n)}_{(l+1)\delta} \setminus E^{(n)}_{l\delta})\big| > \frac{\eta}{6} k^{-1/2}(w_n(\gamma_1) \vee w_n(\gamma_2))\Big\} \\
&\le \frac{kE\nu_n(E^{(n)}_{(l+1)\delta} \setminus E^{(n)}_{l\delta})}{(\eta/6)^2 k((w_n(\gamma_1) \vee w_n(\gamma_2)))^2} \to 0 .
\end{aligned}$$

uniformly for all $l \in \{-\lceil M/\delta\rceil \dots \lfloor M/\delta\rfloor\}^6$.

It remains to prove that $Z_n(\theta) \to 0$ in probability for all $\theta \in [-M, M]^6$. This, however, follows similarly by Chebyshev's inequality, (D1) and the aforementioned analog to Corollary 5.5:

$$\begin{aligned}
&P\{|Z_n(\vartheta)| > \eta\} \\
&= P\Big\{k\big|\nu_n(E^{(n)}_\theta) - E\nu_n(E^{(n)}_\theta)\big| > \eta k^{1/2}(w_n(\gamma_1) \vee w_n(\gamma_2))\Big\} \\
&\le \frac{nP\{T_n^{\leftarrow}(X,Y) \in E^{(n)}_\theta\}}{\eta^2 k(w_n(\gamma_1) \vee w_n(\gamma_2))^2}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\nu(E_\vartheta^{(n)}) + O(A_0(n/k))}{\eta^2(w_n(\gamma_1)\vee w_n(\gamma_2))^2}\\
&= \frac{\nu(S) + o(1)}{\eta^2(w_n(\gamma_1)\vee w_n(\gamma_2))^2}\\
&\to 0.
\end{aligned}
$$

□

Remark 5.7. *Two remarks on this proof are in place. At first glance it seems peculiar that in the definition of $\tilde H^{(n,i)}_{\vartheta_i,\chi_i,\xi_i}$ both parameters $-\vartheta_i$ and χ_i take over the role of $k^{1/2}(\hat\gamma_i - \gamma_i)$ in the definition of $\tilde H$. This, however, is necessary to ensure the crucial monotonicity property of $\tilde H^{(n,i)}_{\vartheta_i,\chi_i,\xi_i}$ in the case $\gamma_i > 0$.*

Secondly, we used the (slightly old-fashioned) classical approach to establish asymptotic equicontinuity instead of the often more elegant approach via bracketing numbers (see van der Vaart and Wellner (2000), Theorem 2.11.9), because the same approximation error of order $O(A_0(n/k))$ in (D1) always enters the upper bound on $E\nu_n(E^{(n)}_{(l+1)\delta}\setminus E^{(n)}_{l\delta})$, thus impeding the calculation of bracketing numbers for radii of smaller order.

Next we show that the terms I and III in decomposition (2.5) are negligible.

Lemma 5.8. *If the conditions of Theorem 2.1 are fulfilled, then*

$$
\begin{aligned}
\hat p_n - \frac{1}{e_n}\nu_n\Big(\frac{d_n}{e_n}H_n(S_n^*)\Big) &= o_P\big(d_n^{-1}k^{-1/2}(w_n(\gamma_1)\vee w_n(\gamma_2))\big) && (5.24)\\
\frac{1}{e_n}\big(E\nu_n(B)-\nu(B)\big)\big|_{B=(d_n/e_n)H_n(S_n^*)} &= o_P\big(d_n^{-1}k^{-1/2}(w_n(\gamma_1)\vee w_n(\gamma_2))\big). && (5.25)
\end{aligned}
$$

Proof. As $\hat p_n = \nu_n\big((d_n/e_n)H_n(S)\big)/e_n$, the left-hand side of (5.24) is non-negative with expectation

$$
\begin{aligned}
&\frac{n}{ke_n}P\Big\{T_n^{\leftarrow}(X,Y)\in\frac{d_n}{e_n}H_n(S\setminus S_n^*)\Big\}\\
&\le \frac{n}{ke_n}P\Big\{T_n^{\leftarrow}(X,Y)\in\frac{d_n}{e_n}H_n\big((0,u_n^*)\times[q(u_n^*-),\infty)\cup[q^{\leftarrow}(v_n^*),\infty)\times[q(\infty),v_n^*)\big)\Big\}\\
&= \frac{1}{d_n}\Big(\nu\big(H_n\big((0,u_n^*)\times[q(u_n^*-),\infty)\cup[q^{\leftarrow}(v_n^*),\infty)\times[q(\infty),v_n^*)\big)\\
&\qquad\qquad +o\big(k^{-1/2}(w_n(\gamma_1)\vee w_n(\gamma_2))\big)\Big)
\end{aligned}
$$

where we have used (D1) and (S2). Now assertion (5.24) follows from Lemma 5.2 and the proof of Corollary 5.5.

Likewise, by conditions (D1), (S2) and $d_n \asymp e_n$, the left-hand side of (5.25) equals

$$
\frac{1}{e_n}\Big(\frac{n}{k}P\{T_n^{\leftarrow}(X,Y)\in B\}-\nu(B)\Big)\Big|_{B=(d_n/e_n)H_n(S_n^*)}
$$

$$= O_P\big(e_n^{-1}A_0(n/k)\big)$$
$$= o_P\big(d_n^{-1}k^{-1/2}(w_n(\gamma_1)\vee w_n(\gamma_2))\big).$$

□

Finally, we derive a bound on term VI in decomposition (2.5).

Lemma 5.9. *Under the assumptions of Theorem 2.1 one has*

$$\nu(d_nS) - p_n = o\big(d_n^{-1}k^{-1/2}(w_n(\gamma_1)\vee w_n(\gamma_2))\big).$$

Proof. With $\lambda_{n,i},\tau_{n,i}$ as in Lemma 5.1, we define for $x\in[\lambda_{n,i},\tau_{n,i}]$

$$H^*_{n,i}(x) := \Big(1+\gamma_i\frac{U_i(d_nx)-b_i(d_n)}{a_i(d_n)}\Big)^{1/\gamma_i}.$$

According to de Haan and Ferreira (2006), Theorem 2.3.6 and 2.3.7 one can choose $a_i(t)$ as a multiple of t^{γ_i} and $b_i(t)=U_i(t)+O(a_i(t)A_i(t))$. Thus, for $\Delta_1(x)$ defined in the proof of Lemma 5.1

$$\begin{aligned}
&\frac{U_i(d_nx)-b_i(d_n)}{a_i(d_n)}\\
&= \frac{a_i(n/k)}{a_i(d_n)}\Big(\frac{U_i(d_nx)-b_i(n/k)}{a_i(n/k)} - \frac{b_i(d_n)-b_i(n/k)}{a_i(n/k)}\Big)\\
&= \Big(\frac{n}{kd_n}\Big)^{\gamma_i}\Big(\frac{(xd_nk/n)^{\gamma_i}-1}{\gamma_i}+\Delta_1(x)+\frac{(d_nk/n)^{\gamma_i}-1}{\gamma_i}+\Delta_1(1)\Big)+O(A_i(d_n))\\
&= \frac{x^{\gamma_i}-1}{\gamma_i}+O\Big(A_i(n/k)\Big(\frac{d_nk}{n}\Big)^{\rho_i+\varepsilon}(x^{\gamma_i+\rho_i+\varepsilon}+1)\Big)+o\Big(A_i(n/k)\Big(\frac{d_nk}{n}\Big)^{\rho_i+\varepsilon}\Big),
\end{aligned}$$

where in the last step we have used (5.1), (5.18) and the Potter bound for the regularly varying function A_0 (de Haan and Ferreira (2006), Prop. B.1.9 5.). We conclude that

$$1+\gamma_i\frac{U_i(d_nx)-b_i(d_n)}{a_i(d_n)} = x^{\gamma_i}\Big(1+O\Big(A_i(n/k)\Big(\frac{xd_nk}{n}\Big)^{\rho_i+\varepsilon}\Big)+O\Big(A_i(n/k)\Big(\frac{d_nk}{n}\Big)^{\rho_i+\varepsilon}x^{-\gamma_i}\Big)\Big).$$

Check that the first remainder term is of smaller order than $k^{-1/2}w_n(\gamma_i)$ by condition (i) of Lemma 5.1. Moreover, for $\gamma_i>0$, (5.19) and again condition (i) of Lemma 5.1 imply

$$A_i(n/k)\Big(\frac{d_nk}{n}\Big)^{\rho_i+\varepsilon}x^{-\gamma_i} = O\Big(A_i(n/k)\Big(\frac{d_nk}{n}\Big)^{\rho_i+\varepsilon}k^{1/2}/\log c_n\Big)\to 0,$$

while for $\gamma_i<0$ this convergence follows from the conditions (i) and (iii) of Lemma 5.1, and for γ_i it is obvious from condition (i).

This shows that $H^*_{n,i}(x)$ is indeed well defined with

$$H^*_{n,i}(x) = x\left(1+\left\{\begin{array}{ll} o(k^{-1/2}\log c_n)+O\big(A_i(n/k)(d_nk/n)^{\rho_i+\varepsilon}x^{-\gamma_i}\big), & \gamma_i>0\\ o\big(k^{-1/2}(d_nk/n)^{-\gamma_i}(1+x^{-\gamma_i})\big), & \gamma_i<0\\ o(k^{-1/2}\log^2 c_n), & \gamma_i=0\end{array}\right.\right)$$

uniformly for $x \in [\lambda_{n,i}, \tau_{n,i}]$. Notice that this representation is of similar type as the approximation derived in Lemma 5.1 with all leading terms equal to 0 (though in the case $\gamma_i > 0$ the second remainder term has a slightly different form). Therefore, we may proceed as before to conclude

$$\begin{aligned}
&\nu(H_n^*(S_n^*)) - \nu(S) \\
&= o\big(k^{-1/2}(w_n(\gamma_1) \vee w_n(\gamma_2))\big) + \sum_{i=1}^{2} O\big(A_i(n/k)(d_n k/n)^{\rho_i+\varepsilon}\big) 1_{\{\gamma_i>0\}} \\
&= o\big(k^{-1/2}(w_n(\gamma_1) \vee w_n(\gamma_2))\big),
\end{aligned}$$

where the last equality follows from Lemma 5.1 (i) (cf. Corollary 5.5).

To complete the proof, we must show that

$$p_n - \nu\big(d_n H_n^*(S_n^*)\big) = o\big(d_n^{-1} k^{-1/2}(w_n(\gamma_1) \vee w_n(\gamma_2))\big).$$

This, however, follows from assumption (D1) (with $t = d_n$) in a similar way as (5.24). $\square$

Proof of Theorem 2.1. The assertion is a direct consequence of (2.5), Corollary 5.5 and of the Lemmas 5.8, 5.6 and 5.9. $\square$

Proof of Corollary 2.2. First note that, similarly as for $\hat{p}_n$, one obtains the representation $\hat{\nu}_n(\hat{S}^+_{n,2}) = \nu_n\Big(\frac{d_n}{e_n} H_n^+(S)\Big)$ with $H_n^+(x,y) := \big(H_{n,1}(x), H^+_{n,2}(y)\big)$,

$$H^+_{n,2}(y) := \frac{e_n}{d_n} T_n^{\leftarrow} \circ \hat{T}_n \circ (\hat{T}_n^{(c^+)})^{\leftarrow} \circ U(d_n y)$$

and $c^+ := c_n^+ := (1+\ell_n)n/(ke_n)$. Thus Lemma 5.1 (with e_n replaced by $e_n/(1+\ell_n)$) yields the approximation

$$H^+_{n,2}(y) = (1+\ell_n) y \left(1 + \left\{ \begin{array}{ll} -k^{-1/2}\log c_n(\Gamma_2/\gamma_2 + o_P(1)) + O_P(k^{-1/2}(yd_n/e_n)^{-\gamma_2}), & \gamma_2 > 0 \\ k^{-1/2}(d_n k/n)^{-\gamma_2}\big((\alpha_2/\gamma_2 - \beta_2 - \Gamma_2/\gamma_2^2 + o_P(1))y^{-\gamma_2} + o_P(1)\big), & \gamma_2 < 0 \\ -k^{-1/2}\log^2 c_n(\Gamma_2/2 + o_P(1)) + O_P(k^{-1/2}\log c_n \log y), & \gamma_2 = 0 \end{array} \right. \right)$$

Now the very same arguments as used in the analysis of $\hat{p}_n$ show that

$$\hat{\nu}_n(\hat{S}^+_{n,2}) = \nu(S^+_{n,2}) + O_P\big(k^{-1/2}(w_n(\gamma_1) \vee w_n(\gamma_2))\big).$$

Together with an analogous approximation for $\hat{\nu}_n(\hat{S}^-_{n,2})$ and our assumption on ℓ_n, we may conclude that

$$\frac{d_n}{e_n}\hat{I}_{n,2} \quad = \quad \frac{d_n}{e_n}\frac{\nu(S^-_{n,2}) - \nu(S^+_{n,2})}{2\ell_n} + o_P(1)$$

$$
\begin{aligned}
&= \int_{x_l}^{\infty} (2\ell_n)^{-1} \int_{(1-\ell_n)q(u)}^{(1+\ell_n)q(u)} \eta(u,v)\,dv\,du \\
&\to \int_{x_l}^{\infty} q(u)\eta(u,q(u))\,du.
\end{aligned}
$$

In the last step we have used the fact that, on the range of integration, $\eta(u,v)$ is continuous and bounded by a multiple of $u^{-3} \vee (q(u))^{-3}$ (cf. (2.10)), so that the integrand of the outer integral can easily be bounded by an integrable function and convergence follows by the dominated convergence theorem. □

Acknowledgment: The research was partially supported by grant FCT/PTDC/MAT/112770/2009 of the Portuguese National Foundation for Science and Technology FCT. We thank two anonymous referees and the associate editor for their thoughtful comments which helped to improve the presentation significantly.

References

Anderson, C.W. (1994). Discussion of a paper by Coles and Tawn. *J. Royal Statist. Soc. B* **43**, 31–33.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., de Waal, D., and Ferro, C. (2004). *Statistics of Extremes.* Wiley.

Bingham, N.H., C.M. Goldie and J.L. Teugels (1987). *Regular Variation.* Cambridge University Press, Cambridge.

Bruun, J.T., and Tawn, J.A. (1998). Comparison of approaches or estimating the probability of coastal flooding. *Appl. Statist.* **47**, 405–423.

Cai, J.-J., Einmahl, J.H.J., and de Haan, L. (2011). Estimation of extreme risk regions under multivariate regular variation. *Ann. Statist.*, **39**, 1803–1826.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values.* Springer.

Coles, S.G., and Tawn, J.A. (1994). Statistical methods for multivariate extremes: an application to structural design. *J. Royal Statist. Soc. B* **43**, 1–31.

Davison, A.C. (1994). Discussion of a paper by Coles and Tawn. *J. Royal Statist. Soc. B* **43**, 33–34.

Dekkers, A.L.M., Einmahl, J.H.J. and de Haan, L. (1989). A moment estimator for the index of an extreme value distribution. *Ann. Statist.* **17**, 1833–1855.

Draisma, G., Drees, H., Ferreira, A., and de Haan, L. (2004). Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli* **10**, 251–280.

Drees, H., de Haan, L., and Li, D. (2006). Approximations to the tail empirical distribution function with application to testing extreme value conditions. *J. Statist. Plann. Inference* **136**, 3498–3538.

Drees, H., de Haan, L., and Resnick, S. (2000). How to make a Hill plot. *Ann. Statist.* **28**, 254–274.

Drees, H., and Müller, P. (2008). Fitting and validation of a bivariate model for large claims. *Insurance: Math. Econ.* **42**, 638–650.

de Haan, L., and Ferreira, A. (2006). *Extreme Value Theory.* Springer.

de Haan, L., and Resnick, S.I. (1993). Estimating the limit distribution of multivariate extremes. *Stoch. Models* **9**, 275–309.

de Haan, L., and de Ronde, J. (1998). Sea and wind: multivariate extremes at work. *Extremes* **1**, 7–45.

de Haan, L., and Sinha, A.K. (1999). Estimating the probability of a rare event. *Ann. Statist.* **27**, 732–759.

Müller, P. (2008). *Modeling Dependencies in Large Insurance Claims.* Dissertation, University of Hamburg.

Schlather, M. (2002). Models for max-stable random fields. *Extremes* **5**, 33–44.

Schreiber, M. (1977). *Differential Forms: a Heuristic Introduction.* Springer.

Segers, J. (2012) Max-stable models for multivariate extremes. *REVSTAT* **10**, 61-82.

van der Vaart, A.W., and Wellner, J.A. (2000). *Weak Convergence and Empirical Processes* (corr. 2nd printing). Springer.