Modeling
Dependencies in Large Insurance Claims

Dissertation
zur Erlangung des Doktorgrades der Fakultät für
Mathematik, Informatik und Naturwissenschaften
der Universität Hamburg

vorgelegt im
Department Mathematik
von

Peter Müller

aus Hoyerswerda

Hamburg
2008

Als Dissertation angenommen vom Department Mathematik
der Universität Hamburg

auf Grund der Gutachten von

Prof. Dr. Holger Drees
und Prof. Dr. Johan Segers

Hamburg, den 11. Juli 2008

Prof. Dr. Jens Struckmeier
Leiter des Departments Mathematik

Peter Müller

Modeling Dependencies in Large Insurance Claims

supervised by

Prof. Dr. Holger Drees
Center of Mathematical Statistics and Stochastic Processes
Department of Mathematics
Faculty of Mathematics, Informatics and Natural Sciences

University of Hamburg

2008

Contents

Preface

Motivation

Modeling insurance claim sizes often requires to consider small and medium sized claims on the one hand, and large claims on the other hand separately. The essential reason is that usually (parametric) models that fit the bulk of the data well do not accurately describe the behavior of the very large claims. Here extreme value theory offers models and estimators for the upper tail of the claim size distribution.

If a risk manager has to assess the joint exposure to extreme risks in two different lines of business, then it is not sufficient to model just the upper tails of the marginal d.f.s of claim sizes in these business lines, since a possible dependence between claim sizes of a policyholder in these lines of business must be taken into account. Indeed, to neglect the impact of such a dependence on the probability of a jointly (i.e. occurring in both lines of business) large claim may lead to a substantial underestimation of the real risk, see e.g. the results in Chapter 5. Again, it is not advisable to use parametric models of the dependence structure, since extreme losses often exhibit a different dependence than small claim sizes. Unfortunately, the classical multivariate extreme value theory does also not allow the accurate estimation of the probability of jointly large claims if the claim sizes are asymptotically independent (see Section 1.2.3), i.e., loosely speaking, if the probability that a large claim in one line of business causes a large claim in the other line of business vanishes for increasing claim sizes.

To analyze the full dependence structure separately from the marginal distributions, one often considers the so-called *copula*, i.e., the bivariate distribution function of the claim sizes after standardization of the marginals to uniform random variables. Recently, a large variety of parametric families of copulas (like t-copulas, Gumbel-copulas or Clayton-copulas) have been proposed as models for the dependence structure between different financial risks, see e.g. Chapter 5 of McNeil et al. (2005). However, as mentioned above, small and medium claims often show a different stochastic behavior than the extreme claims. This difference is not restricted to the marginal distributions, but may also become manifest in the dependence structure. Since, in most instances, no parametric copula can be selected on the basis of physical reasons, it seems advisable to use more flexible models for the dependence structure between extreme claims in different lines of business. (For a detailed discussion of the drawbacks of parametric copula modeling in extreme value theory see Mikosch (2005).)

The present work is mainly devoted to the development of such a flexible model, the model fitting and the validation of its central assumptions. Ledford and Tawn (1996, 1997, 1998) proposed a model for the joint tail distribution of bivariate claim sizes which both overcomes the obstacles of the classical multivariate extreme value theory and offers far greater flexibility than a parametric copula approach. We somewhat relax the assumptions of the models of Ledford and Tawn and establish an estimation procedure that allows to prove useful asymptotic properties of estimators of the model parameters, in particular of the estimator of the probability of jointly large claims. This allows to accurately estimate the quantities of interest and to construct confidence intervals for them. As we have indicated above, it is a crucial point to validate the assumptions when modeling the stochastic behavior of jointly large claims. The development of a validation tool for our model is one of the main results of this thesis.

The applicability of our results is not restricted to issues of insurance or financial mathematics. For example, similar structured problems are to be found in hydrology, meteorology or teletraffic problems (such as on/off times of computers or transfer times of files), see e.g. Section 6.1.1 of de Haan and Ferreira (2006), Examples 1.9 – 1.11 of Coles (2001) or Section 8.1 of Beirlant et al. (2004).

Outline

The main intention of Chapter 1 is to introduce notions, concepts and results of univariate extreme value theory and classical multivariate extreme value theory that will be necessary or useful in this work. Section 1.2.3 exhibits the abovementioned limited value of classical multivariate extreme value theory in the case of asymptotic independence. In this view, it serves as motivation for the development of models which are more appropriate in this situation.

In Chapter 2, we present the original and the Extended Ledford and Tawn Model(s). We indicate similarities and differences of these models, where our particular interest lies in the Extended Ledford and Tawn Model, which will be introduced in detail. The central assumption, a bivariate second order condition, and its immediate implications are examined. We conclude that for the bivariate tail of the claim size distribution a scaling law, which plays a crucial role in this thesis, holds.

Chapter 3 is devoted to the fitting of the Extended Ledford and Tawn Model. We establish the asymptotic behavior of the estimators of the model parameters. We explain how the scaling law can be used to construct an estimator for the probability of jointly large claims, prove asymptotic normality of this estimator and investigate the influence of the interplay of the errors when estimating the model parameters.

In Chapter 4 we develop a method to validate the scaling law for given data. We prove that under some regularity conditions an empirical process which measures random deviations from this scaling law converges weakly to a centered Gaussian process. An immediate consequence of this result is used to construct a test statistic to discriminate between presence and absence of the scaling law. We suggest a method to validate the scaling law by a simple three-dimensional plot.

The applicability of our results is demonstrated in Chapter 5. Data of Danish fire insurance claims and of medical claims from US health insurances are investigated. After briefly discussing two ways of an appropriate choice of the data used for the analysis when a certain (and common) type of censorship exists, we analyze the resulting data in these ways. We estimate the model parameters and the probability of jointly large claims, validate the scaling law and compare the results among themselves, with empirical estimates and estimates when independence is assumed.

In the final Chapter 6, we investigate a copula model that is related to the Extended Ledford and Tawn Model. We show that although the presented approach requires similar conditions as the Extended Ledford and Tawn Model, useful information, which the Extended Ledford and Tawn Model exploits, is lost. We see that due to this fact the approach does not overcome the main problems and drawbacks of the widely-used copula approaches and of classical multivariate extreme value theory.

In each chapter, the proofs of the results are deferred to the respective final section.

Acknowledgment

First and foremost, I gratefully acknowledge the guidance, the valuable discussions and proposals of my supervisor Holger Drees. His patience and support even when I came to see him about the most basic questions, and his seemingly uncountable set of ideas for solving the arising problems were essential for the accomplishment of this work.

My thanks also address Johan Segers who offered to co-examine this thesis. His valuable remarks lead to a considerable improvement of this work.

My special acknowledgments go to my family and my friends. By the pleasant time outside the university, they gave me the strength, the motivation and the balance which are so important when writing such a thesis.

Chapter 1

Preliminaries

In this introductory chapter, we *briefly* present some concepts and results of univariate extreme value theory (EVT) and classical multivariate EVT (CMEVT). Of course, we discuss a small selection of issues only, so that the reader should not expect a closed presentation of (CM)EVT. This selection has been made according to two criteria. We only present concepts and results which are (i) required in theory and applications in this work or (ii) useful for the purpose of comparison.

Section 1.1 is devoted to univariate EVT. We state the well-known main results on the limiting distribution of sample maxima (Section 1.1.1) and of peaks over a certain threshold (Section 1.1.2). In Section 1.1.3, we touch on regular variation which plays an essential role in EVT. In addition to stating the basic definition and one of many important links to EVT, we discuss some aspects of so-called second-order regular variation conditions. Section 1.1.4 is devoted to two popular estimators of the most important parameter in univariate EVT – the extreme value index.

In Section 1.2, we briefly discuss the necessity of multivariate EVT to assess the joint exposure to extreme risks in two different lines of business. Similar as in Section 1.1 for the univariate theory, we state the closely related basic results on the limiting distribution of componentwise maxima (Section 1.2.1) and of claims that exceed a certain threshold (Section 1.2.2). The final Section 1.2.3 shows that CMEVT is not the appropriate framework for all issues when modeling multivariate extremes. We see that in the common case of asymptotic independence of risks in two different lines of business, CMEVT is of little help for estimating the upper tail of the joint claim size distribution.

Most of the material treated in this chapter can be found in any recent book on EVT. We mainly borrow from Beirlant et al. (2004) and in particular from de Haan and Ferreira (2006). Further references are given when appropriate.

1.1 Some Elements of Univariate Extreme Value Theory

Throughout this section, denote by F_1 the distribution function (d.f.) of the claim size X or of X_i belonging to an i.i.d. sample $X_1, X_2, \ldots, X_n$ of claim sizes.

1.1.1 Sample Maxima

In this section, we focus on the statistical behavior of $\max\{X_1, X_2, \ldots, X_n\}$. It can be shown that the only possible limits of the d.f. of the suitably standardized sample maximum, i.e.

$$P\left\{\frac{\max_{i=1,\ldots,n} X_i - b_n}{a_n} \le x\right\} \xrightarrow{n\to\infty} G(x) \tag{1.1}$$

for every continuity point x of G, are the so-called generalized extreme value (GEV) distributions with d.f.

$$G(x) = \exp\big(-(1+\gamma x)^{-1/\gamma}\big) \qquad \text{if } 1+\gamma x > 0. \tag{1.2}$$

Here, $a_n > 0$ and $b_n \in \mathbb{R}$ are normalizing constants and the shape parameter $\gamma \in \mathbb{R}$ of G denotes the so-called extreme value index, a key quantity in EVT. If $\gamma = 0$, then the right-hand side of (1.2) is interpreted as $\exp(-e^{-x})$. We say that F_1 belongs to the domain of attraction of G, if (1.1) is satisfied. There is an equivalent formulation in terms of the generalized inverse function of $1/(1-F_1)$. For any non-decreasing function f denote by $f^{\leftarrow}$ its left-continuous generalized inverse, i.e.

$$f^{\leftarrow}(x) := \inf\big\{y : f(y) \ge x\big\},$$

where, as usual, $\inf\emptyset = \infty$, and let

$$U := \left(\frac{1}{1-F_1}\right)^{\leftarrow}.$$

The d.f. F_1 belongs to the domain of attraction of G, if and only if

$$\lim_{r\to\infty} \frac{U(rx)-U(r)}{\tilde{a}(r)} = \frac{x^\gamma - 1}{\gamma}, \qquad x > 0, \tag{1.3}$$

with a positive function $\tilde{a}$, where for $\gamma = 0$ the right-hand side is interpreted as $\log x$. For the case $\gamma > 0$, there is another useful equivalent statement in terms of regular variation. We give it in Section 1.1.3.

1.1.2 Peaks Over Threshold

In order to assess the risk associated with a policy, insurers are often interested in the probability that a certain threshold is exceeded rather than in the distribution of the maximum of claim sizes. In this situation, the so-called peaks-over-threshold (POT) models of EVT offer the appropriate framework. The relation of the EVT conditions (1.1) for sample maxima and for peaks over a certain threshold is close. Condition (1.1) holds if and only if the conditional distribution function of the suitably standardized claim size X given that it exceeds a very large threshold converges, i.e.

$$P\left(\frac{X-r}{a(r)} \le x \,\middle|\, X > r\right) \xrightarrow{P\{X>r\}\to 0} H(x) \tag{1.4}$$

with a positive normalizing constant $a(r)$ depending on the threshold r. It can be shown that the only possible limits in (1.4) are the generalized Pareto distributions (GPD) with d.f.

$$H(x) = 1 - (1+\gamma x)^{-1/\gamma} \qquad \text{if } 1 + \gamma x > 0, \tag{1.5}$$

where the right-hand side of (1.5) is interpreted as $1 - \exp(-x)$ if the shape parameter γ of H equals 0.

Most critical for an insurance company is the case of a heavy-tailed claim size d.f. F_1. As usual, we identify this situation with the case $\gamma > 0$, since then the right endpoint of H is infinity and because $1 - H(x) \sim (\gamma x)^{-1/\gamma}$ as $x \to \infty$, the GPD essentially decreases like a power function. Here, $f(x) \sim g(x)$ stands for asymptotic equivalence of f and g, i.e. $\lim_x f(x)/g(x) = 1$. Furthermore, if $\gamma > 0$, all moments of X of order larger than $1/\gamma$ do not exist. As mentioned, we give a further statement equivalent to (1.4) if $\gamma > 0$ in terms of regular variation in Section 1.1.3.

Due to (1.4), the upper tail of the claim size distribution may be approximated by the GPD with additional location and scale parameters. Suggested by the POT-convergence (1.4), we base the estimation of the tail of the claim size distribution $P\{X > x\} = 1 - F_1(x)$ on the approximation

$$P(X > x \mid X > r) \approx 1 - H\Big(\frac{x-r}{a(r)}\Big) = \Big(1 + \gamma\frac{x-r}{a(r)}\Big)^{-1/\gamma} \tag{1.6}$$

for sufficiently large $r > 0$. Hence, with $p_r := P\{X > r\}$,

$$1 - F_1(x) \approx p_r\Big(1 + \gamma\frac{x-r}{a(r)}\Big)^{-1/\gamma}. \tag{1.7}$$

If we continue (1.7) to

$$\begin{aligned} 1 - F_1(x) &\approx \bigg(\Big(1 + \gamma\frac{x-r}{a(r)}\Big)p_r^{-\gamma}\bigg)^{-1/\gamma} \\ &= \Big(1 + \gamma\frac{x-r}{a(r)p_r^{\gamma}} + p_r^{-\gamma} - 1\Big)^{-1/\gamma} \\ &= \left(1 + \gamma\frac{x - r + \frac{a(r)p_r^{\gamma}}{\gamma}(p_r^{-\gamma} - 1)}{a(r)p_r^{\gamma}}\right)^{-1/\gamma}, \end{aligned}$$

we obtain the parameterization we use in the following, namely

$$1 - F_1(x) \approx \Big(1 + \gamma\frac{x-\mu}{\varsigma}\Big)^{-1/\gamma} \tag{1.8}$$

for sufficiently large $x > 0$, where location parameter $\mu \in \mathbb{R}$ and scale parameter $\varsigma > 0$ are defined as

$$\mu := \frac{a(r)}{\gamma}\big(p_r^{\gamma} - 1\big) + r \qquad \text{and} \qquad \varsigma := a(r)p_r^{\gamma}. \tag{1.9}$$

Motivated by (1.8), we define the generalized Pareto estimator of the tail of the survival function $1 - F_1$ by

$$1 - \hat{F}_{1,n}(x) := \Big(1 + \hat{\gamma}_n\frac{x - \hat{\mu}_n}{\hat{\varsigma}_n}\Big)^{-1/\hat{\gamma}_n} \qquad \text{if } 1 + \hat{\gamma}_n\frac{x-\hat{\mu}_n}{\hat{\varsigma}_n} > 0, \tag{1.10}$$

where $\hat{\gamma}_n$, $\hat{\mu}_n$ and $\hat{\varsigma}_n$ are suitable estimators for γ, μ and ς. We discuss the estimation of γ, μ and ς in such a location-scale setting in Section 1.1.4.

1.1.3 Univariate Extreme Value Theory and Regular Variation

The theory of regular variation is extensive and its links to EVT are numerous. In this section, we only give the basic definition of regular variation and one these links; both is necessary for further developments. However, we frequently refer to results on regular variation throughout this work. Further, we briefly discuss so-called second-order regular variation conditions, since such conditions are important for various aspects addressed in this work. We refer to the book of Bingham et al. (1987) as the standard reference on regular variation and to (Appendix B of) de Haan and Ferreira (2006) for the aspects which are most important for EVT including second-order conditions.

A positive, Lebesgue measurable function $f : \mathbb{R}_+ \to \mathbb{R}_+$ is called *regularly varying (in $a \in [0, \infty]$) with index $\alpha \in \mathbb{R}$*, if

$$\lim_{r \to a} \frac{f(rx)}{f(r)} = x^{\alpha}.$$

If $\alpha = 0$, then f is said to be *slowly varying.* Clearly, the simplest examples of functions which vary regularly (in all $a \in [0, \infty]$) with index α are the power functions $f(x) = x^{\alpha}$, $x > 0$.

One of the many links between regular variation and EVT is that if $\gamma > 0$, then condition (1.4) holds if and only if the survival function $1 - F_1$ is regularly varying (at ∞) with index $-1/\gamma$, i.e.

$$\lim_{r \to \infty} \frac{1 - F_1(rx)}{1 - F_1(r)} = x^{-1/\gamma}, \qquad x > 0. \tag{1.11}$$

Note that for $x > 0$

$$\begin{aligned} P\left(\frac{X - r}{a(r)} \le x \,\middle|\, X > r \right) &= \frac{P\{X \le r + a(r)x\} - P\{X \le r\}}{P\{X > r\}} \\ &= 1 - \frac{1 - F_1(r + a(r)x)}{1 - F_1(r)}, \end{aligned}$$

so that (1.11) is the same as (1.4) with $a(r) = \gamma r$ and $1 + \gamma x$ replaced with x. Typical examples of such survival functions behave (asymptotically for large claim sizes X) as $x^{-1/\gamma}$ or $x^{-1/\gamma} \log^{\rho} x$ for some $\rho \in \mathbb{R}$. (See Beirlant et al. (2004), Table 2.1, for explicit examples.)

Many statistical results, including the asymptotic normality of the estimators considered in Section 1.1.4, require a more sophisticated asymptotic condition on U than the domain of attraction condition (1.3). For the Hill estimator of $\gamma > 0$, considered in Section 1.1.4, we assume

$$\lim_{r\to\infty} \frac{\frac{U(rx)}{U(r)} - x^\gamma}{A_1(r)} = x^\gamma \frac{x^{\rho_1} - 1}{\rho_1}, \qquad x > 0, \tag{1.12}$$

where $\rho_1 \leq 0$ (if $\rho_1 = 0$, then $(x^{\rho_1} - 1)/\rho_1$ is interpreted as $\log x$), A_1 and A_2 (below) are some positive or negative functions with $A_1(r), A_2(r) \to 0$ as $r \to \infty$. For the maximum likelihood estimator of $\gamma > -1/2$ (the maximum likelihood estimator behaves irregular if $\gamma \leq -1/2$), also considered in Section 1.1.4, the less restrictive condition

$$\lim_{r\to\infty} \frac{\frac{U(rx)-U(r)}{\tilde{a}(r)} - \frac{x^\gamma - 1}{\gamma}}{A_2(r)} = K(x), \qquad x > 0, \tag{1.13}$$

suffices, where $K \not\equiv 0$ is not a multiple of $(x^\gamma - 1)/\gamma$. Note that each of these *second-order conditions* implies the domain of attraction condition (1.3). The conditions (1.12) and (1.13) are still quite general, e.g. they are satisfied by all usual distributions that satisfy (1.3). Besides the possibility to prove asymptotic normality of estimators of the extreme value index, second-order conditions are useful for further purposes. For example, analyses of the parameter ρ_1 allow conclusions on the speed of convergence to asymptotic normality of the estimators presented in Section 1.1.4. Thereby, it is possible to prove optimality results for adaptive parameter choices. For a more thorough discussion of (1.12), (1.13), related conditions and their implications for estimators of γ, some of which are summarized in the following Section 1.1.4, see Chapter 3 of de Haan and Ferreira (2006).

We conclude this section with the remark that it is a second-order condition in the bivariate case that enables us to prove the main results of Chapters 3 and 4 of this work.

1.1.4 The Hill Estimator and the Maximum Likelihood Estimator of the Extreme Value Index

The extreme value index γ is a key quantity in EVT. We have already seen that γ is the shape parameter of the limiting distributions (1.2) and (1.5). The particular importance of γ, however, is due to its crucial role when analyzing the limiting behavior of tail quantile and tail probability estimators, see e.g. Chapter 4 of de Haan and Ferreira (2006). This section is devoted to two of the many estimators of γ. We summarize some of the properties of these estimators, in particular if (1.12) and (1.13), respectively, are assumed. The according results can be found in Sections 3.2 and 3.4 of de Haan and Ferreira (2006).

The Hill estimator of $\gamma > 0$ may be motivated through various different methods. We give a brief derivation by a regular variation argument. Further possibilities to construct the Hill estimator can be found in Section 4.2.1 of Beirlant et al. (2004). They show that this estimator is a natural and self-evident choice.

The regular variation of the quantile function $F_1^{\leftarrow}$ at 1 with index $-\gamma < 0$, i.e.

$$\lim_{t\downarrow 0} \frac{F_1^{\leftarrow}(1-tx)}{F_1^{\leftarrow}(1-t)} = x^{-\gamma}, \qquad x > 0, \tag{1.14}$$

is equivalent to (1.11). Hence, for $j \in \mathbb{N}$ such that j/n is sufficiently small,

$$\begin{aligned} \gamma &= \int_0^1 \log t^{-\gamma}\, dt \;\approx\; \frac{1}{j}\sum_{i=1}^{j} \log(i/j)^{-\gamma} \approx \frac{1}{j}\sum_{i=1}^{j} \log \frac{F_1^{\leftarrow}(1-i/n)}{F_1^{\leftarrow}(1-j/n)} \\ &\approx \frac{1}{j}\sum_{i=1}^{j} \log \frac{X_{n-i+1:n}}{X_{n-j:n}} =: \hat{\gamma}_n^H, \end{aligned}$$

where $X_{i:n}$, $i = 1, \ldots, n$, denote the order statistics pertaining to X_i, $i = 1, \ldots, n$. More precisely, $j = j_n$ is an intermediate sequence, i.e. $j \to \infty$ and $j/n \to 0$, which indeed depends on the sample size n. Observe that for reasons of convenience, our notation of the Hill estimator $\hat{\gamma}_n^H$ does not allow for its dependence on j.

An important property of $\hat{\gamma}_n^H$ is its asymptotic normality. If the second order condition (1.12) holds for $\gamma > 0$, a function A_1 regularly varying with index $\rho_1 \leq 0$ and j satisfying $\sqrt{j}A_1(n/j) \to 0$, then $\sqrt{j}\left(\hat{\gamma}_n^H - \gamma\right)$ is asymptotically centered normal with variance

$$\sigma^2_{\hat{\gamma}_n^H} = \gamma^2. \tag{1.15}$$

If $\sqrt{j}A_1(n/j)$ converges to some positive number, a bias enters, see Theorem 3.2.5 of de Haan and Ferreira (2006).

In Chapter 5, in order to choose a suitable number of order statistics j used for $\hat{\gamma}_n^H$, we consider a so-called Hill plot, which displays the estimates $\hat{\gamma}_n^H$ versus j. A small value j results in a high variance of $\hat{\gamma}_n^H$, while too large a j may cause a large bias. (Here, under second-order regular variation conditions an asymptotically optimal choice of j is possible.) This issue is sometimes referred to as *bias-variance trade-off.* A typical (nice) Hill plot exhibits heavy fluctuations for small values of j (due to large variance), followed by a rather stable region where both variance and bias are moderate, before the bias causes a clear upward or downward trend of the curve. Thus, j ought to be chosen in the region where the plot is rather stable. See also the applications in Chapter 5. For a more detailed discussion on this matter see e.g. Sections 6.4 and 6.5 of Embrechts et al. (1997) or Drees et al. (2000).

We now specify our choice of the estimators of μ and ς of (1.9) when we use $\hat{\gamma}_n^H$ for the estimation of γ. Recall that we obtained (1.11) (and therefore the motivation for the Hill estimator (1.14)) from (1.4) with assigning $a(r) = \gamma r$. Further, since for the Hill estimator $\hat{\gamma}_n^H$ we use those X_i with $X_i \geq X_{n-j:n}$, a reasonable choice for r is $X_{n-j:n}$. Therefore, $p_r = P\{X > r\} \approx P\{X > X_{n-j:n}\} \approx n^{-1} \sum_{i=1}^n \mathbb{1}\{X_i \geq X_{n-j:n}\}$ is estimated by $(j+1)/n$. Hence, when we use the Hill estimator $\hat{\gamma}_n^H$, the parameters μ and ς of (1.9) are estimated as

$$\hat{\mu}_n^H := X_{n-j:n}\big((j+1)/n\big)^{\hat{\gamma}_n^H} \quad \text{and} \quad \hat{\varsigma}_n^H := \hat{\mu}_n^H \hat{\gamma}_n^H. \tag{1.16}$$

In this case, the resulting estimator (1.10) of the tail probability $1 - F_1(x)$ reads as

$$1 - \hat{F}_{1,n}^H(x) = \frac{j+1}{n} \left(\frac{x}{X_{n-j:n}} \right)^{-1/\hat{\gamma}_n^H}.$$

Since the class of d.f.s that belong to the domain of attraction of G cannot be parameterized by a finite number of parameters, a straightforward maximum likelihood estimator does not exist. However, relation (1.4) suggests that the tail of X follows a GPD (1.5), which is used as an approximate model. Applying the maximum likelihood procedure to the largest observations using the GPD, cf. Section 3.4 of de Haan and Ferreira (2006), leads to what is generally called the maximum likelihood estimator of γ in EVT. The maximum likelihood estimators $\big(\hat{\gamma}_n^{ML}, \hat{\alpha}_n^{ML}\big)$ of $0 \neq \gamma > -1/2$ and the scale parameter $\alpha := a(r)$, where for the same reasons as above we again use $r = X_{n-j:n}$, are then obtained by the solution of the equations

$$
\begin{aligned}
&\frac{1}{j}\sum_{i=1}^{j}\log\Big(1+\frac{\gamma}{\alpha}\big(X_{n-i+1:n}-X_{n-j:n}\big)\Big)=\gamma,\\
&\frac{1}{j}\sum_{i=1}^{j}\frac{1}{1+\frac{\gamma}{\alpha}\big(X_{n-i+1:n}-X_{n-j:n}\big)}=\frac{1}{\gamma+1}.
\end{aligned}
\tag{1.17}
$$

As for the Hill estimator, $j = j_n$ is an intermediate sequence and for the choice of an appropriate j the same remarks as made for $\hat{\gamma}_n^H$ apply.

The maximum likelihood estimator $\hat{\gamma}_n^{ML}$ has larger asymptotic variance than the Hill estimator $\hat{\gamma}_n^H$. If the second order condition (1.13) holds for $\gamma > -1/2$, a function A_2 regularly varying with index $\rho_2 \leq 0$ and j satisfies $\sqrt{j}A_2(n/j) \to 0$, then $\sqrt{j}\left(\hat{\gamma}_n^{ML} - \gamma\right)$ is asymptotically centered normal with variance

$$\sigma^2_{\hat{\gamma}_n^{ML}} = (1+\gamma)^2. \tag{1.18}$$

If $\sqrt{j}A_2(n/j)$ converges to some positive number, a bias enters, see Theorem 3.4.2 of de Haan and Ferreira (2006).

As above, we estimate p_r by $(j+1)/n$ so that our estimators of the parameters μ and ς appear as

$$
\begin{aligned}
\hat{\mu}_n^{ML} &:= \frac{\hat{\alpha}_n^{ML}}{\hat{\gamma}_n^{ML}}\left(\left(\frac{j+1}{n}\right)^{\hat{\gamma}_n^{ML}}-1\right)+X_{n-j:n} \quad \text{and}\\
\hat{\varsigma}_n^{ML} &:= \hat{\alpha}_n^{ML}\left(\frac{j+1}{n}\right)^{\hat{\gamma}_n^{ML}},
\end{aligned}
\tag{1.19}
$$

when we use the maximum likelihood estimator $\hat{\gamma}_n^{ML}$. In this case, the resulting estimator (1.10) of the tail probability $1 - F_1(x)$ reads as

$$1-\hat{F}_{1,n}^{ML}(x)=\frac{j+1}{n}\left(\frac{\hat{\gamma}_n^{ML}}{\hat{\alpha}_n^{ML}}\left(x+X_{n-j:n}\right)+1\right)^{-1/\hat{\gamma}_n^{ML}}.$$

Henceforth we denote by $\hat{\gamma}_n$ either the Hill estimator $\hat{\gamma}_n^H$ or the maximum likelihood estimator $\hat{\gamma}_n^{ML}$, by $\sigma^2_{\hat{\gamma}_n}$ the pertaining asymptotic variance and by $1 - \hat{F}_{1,n}(x)$ the pertaining estimator of $1 - F_1(x)$. This estimator is asymptotically centered normal if the second order condition (1.12) holds with $\sqrt{j}A_1(n/j)$ converging to some real value, and $x = x_n \to \infty$ such that

$$n(1 - F_1(x_n)) = o(j) \qquad \text{and} \qquad \log\big(n(1 - F_1(x_n))\big) = o\big(j^{1/2}\big). \tag{1.20}$$

More precisely, from Theorem 4.4.7 of de Haan and Ferreira (2006) we obtain that

$$\frac{\gamma j^{1/2}}{\log \frac{j}{n(1-F_1(x_n))}} \left(\frac{1 - \hat{F}_{1,n}(x_n)}{1 - F_1(x_n)} - 1 \right) \xrightarrow{D} \mathcal{N}\big(0, \sigma^2_{\hat{\gamma}_n}\big), \tag{1.21}$$

where $\xrightarrow{D}$ denotes convergence in distribution as $n \to \infty$. This implies that

$$\frac{1 - \hat{F}_{1,n}(x_n)}{1 - F_1(x_n)} - 1 = O_P\left(j^{-1/2} \log \frac{j}{n(1 - F_1(x_n))} \right). \tag{1.22}$$

The conditions (1.20) require the (sequence of) thresholds for which an event is considered “extreme” not to converge too fast or too slow to infinity. The assumption $n(1-F_1(x_n)) = o(j)$ also ensures that $1-F_1(x) = 1-F_1(x_n) \to 0$, which is necessary (and reasonable) when applying asymptotic methods to extrapolate outside the range of available data. In their Section 4.3, p. 134, de Haan and Ferreira (2006) briefly discuss this issue.

1.2 Some Elements of Classical Multivariate Extreme Value Theory

Univariate EVT and tail models do not suffice if a risk manager has to assess the joint exposure to extreme risks in two different lines of business. Suppose that the claim sizes of one policyholder in these lines of business are described by a bivariate random vector (X, Y). Then it is not sufficient to model just the upper tails of the marginal d.f.s of X and Y, because there might be a non-negligible dependence between the two insured risks. Indeed, if one assumes independence and therefore calculates the probability of a *jointly* extreme event like

$$p := P\{X > u_1, Y > u_2\} \tag{1.23}$$

as $P\{X > u_1\}P\{Y > u_2\}$, then one underestimates the risk if the two claim sizes are actually positively dependent; see e.g. the results in Chapter 5. In particular, this is to be expected when the insured risks in the different lines of business are exposed to the same physical cause of damage like storms in motor insurance and residential building insurance, or fire in residential building insurance and household insurance. However, it is worth mentioning that a non-negligible dependence between the claim sizes may also occur when there is no obvious physical mechanism causing the damages in both lines of business.

A very simple measure of the dependence between the claim sizes is the correlation between X and Y. However, the correlation does not help to determine probabilities of type (1.23). Moreover, a single figure (like the correlation) cannot capture all essential features of a usually quite complex dependence structure. Indeed, it often leads to quite misleading interpretations; see e.g. Embrechts et al. (2002) for examples.

Quite flexible dependence models for extremes are offered by the classical multivariate EVT (CMEVT). This section is mainly devoted to summarize some of the results of CMEVT as far as it is necessary to display problems when estimating probabilities like (1.23) and to compare CMEVT to the original and the Extended Ledford and Tawn Model developed in Chapter 2. In particular, we see that in the common case of asymptotic independence of X and Y, CMEVT is of little help for estimating p. Section 1.2.3 is devoted to this case.

1.2.1 Componentwise Maxima

Henceforth, let $\{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}$ be an i.i.d. sample of claim sizes, and denote by F_1 and F_2 the marginal d.f.s of X_i and Y_i, $i = 1, \dots, n$, respectively.

By the term "classical" MEVT, we mean the material presented e.g. in Chapter 5 of Resnick (1987) or Chapter 8 of Beirlant et al. (2004), where it is supposed that the distribution of suitable standardized componentwise maxima converges to a d.f. with non-degenerate marginals. More precisely, in CMEVT it is assumed that

$$P\left\{\frac{\max_{i=1,\dots,n} X_i - b_n}{a_n} \le x, \frac{\max_{i=1,\dots,n} Y_i - d_n}{c_n} \le y\right\} \xrightarrow{n\to\infty} G(x, y), \tag{1.24}$$

for some normalizing constants $a_n, c_n > 0$, $b_n, d_n \in \mathbb{R}$, where the *multivariate extreme value d.f.* G has non-degenerate marginals $G_1(x) := G(x, \infty)$ and $G_2(y) := G(\infty, y)$. Usually, modeling the joint tail distribution of X and Y by means of (1.24) is separated into two parts describing the tails of the marginal d.f.s on the one hand and the tail dependence structure between them on the other hand. The former part is a matter of univariate EVT, since for suitably chosen normalizing constants, the marginals of G are GEV distributions according to (1.2). The second part, estimating the multivariate extreme value dependence structure, is the crucial point here. Section 9.3 of Beirlant et al. (2004) presents some methods which allow inference for this issue.

As in univariate EVT, threshold models are closely related. We establish the connection in the following section. Thereby, we extend the univariate theory briefly presented in Section 1.1.2. We give conditions that are equivalent to (1.24). They display the abovementioned possibility to separate the modeling of the joint tail distribution of X and Y into specifying the marginal d.f.s and the dependence structure. The convergence of the latter is thereby obtained by the convergence of the so-called tail dependence function. Further, we infer a representation for threshold probabilities like p.

1.2.2 Threshold Models

We will see that condition (1.4) for each of the marginals and a certain regularity condition on the probability that *at least one* standardized claim size is large, are jointly equivalent to (1.24). To this end, let F denote the bivariate d.f. of (X, Y) and $F_1(x) := F(x, \infty)$ and $F_2(y) := F(\infty, y)$ the marginal d.f.s of X and Y, respectively, which, for simplicity, are assumed continuous for the rest of this chapter. Define the *tail dependence function* D by

$$D(u, v) := P\{U < u \textbf{ or } V < v\}, \qquad u, v \in [0, 1].$$

where $U := 1 - F_1(X)$ and $V := 1 - F_2(Y)$ are uniform random variables. Further, for any non-increasing function f denote by $f^{\leftarrow}$ its left-continuous generalized inverse, i.e.

$$f^{\leftarrow}(x) := \sup\{y : f(y) \geq x\}.$$

The following result is a consequence of Proposition 5.10 of Resnick (1987), where the marginals are standardized to standard Fréchet distribution. See also Beirlant et al. (2004), Section 8.3.2 and keep in mind that (1.1) is equivalent to (1.4).

1.2.1 Proposition. The following statements are equivalent:

(i) (1.24) holds.

(ii) The conditions

$$\begin{aligned} P\left(\frac{X-r}{a_1(r)} \le x \;\middle|\; X>r\right) &\xrightarrow{P\{X>r\}\to 0} 1-(1+\gamma_1 x)^{-1/\gamma_1}, \\ P\left(\frac{Y-r}{a_2(r)} \le y \;\middle|\; Y>r\right) &\xrightarrow{P\{Y>r\}\to 0} 1-(1+\gamma_2 y)^{-1/\gamma_2}, \\ t^{-1}D(tx,ty) \quad &\xrightarrow{t\downarrow 0} \quad \ell(x,y) \end{aligned} \tag{1.25}$$

with positive normalizing constants a_1 and a_2, are jointly satisfied for all $x, y \ge 0$, where ℓ is the *stable tail dependence function* defined by

$$\ell(x,y) := -\log G\big((-\log G_1)^{\leftarrow}(x), (-\log G_2)^{\leftarrow}(y)\big).$$

△

We now summarize some properties of ℓ which will be useful later. Note that due to Proposition 1.2.1 and the equivalence of (1.1) and (1.4), the marginals G_i, $i=1,2$, of G take for suitably chosen normalizing constants the form

$$G_i(x) = \exp\big(-(1+\gamma_i x)^{-1/\gamma_i}\big) \qquad \text{if } 1+\gamma_i x>0, \qquad i=1,2,$$

so that ℓ may be represented as

$$\ell(x,y) = -\log G\left(\frac{x^{-\gamma_1}-1}{\gamma_1}, \frac{y^{-\gamma_2}-1}{\gamma_2}\right).$$

Due to the standardization of the marginals of the tail dependence function D, the "marginals" of ℓ are standardized:

$$\ell(x,0) = \lim_{t\downarrow 0} \frac{D(tx,0)}{t} = \lim_{t\downarrow 0} \frac{P\{U<tx\}}{t} = x.$$

By

$$\begin{aligned}\ell(sx, sy) &= \lim_{t\downarrow 0} \frac{D(tsx, tsy)}{t} = s \lim_{t\downarrow 0} \frac{D(tsx, tsy)}{ts} \\ &= s \lim_{r\downarrow 0} \frac{D(rx, ry)}{r} = s\ell(x, y),\end{aligned}$$

$s > 0$, we see that ℓ is homogeneous of order 1. Thus, according to (1.25), i.e. $D(tx, ty) \sim t\ell(x, y) = \ell(tx, ty)$, we may approximate $D(u, v)$ by $\ell(u, v)$ for small u, v. Therefore, one may construct an estimator for the probability (1.23) from estimators of the marginal d.f.s and an estimator of ℓ using the approximation

$$\begin{aligned}p &= P\{X > u_1\} + P\{Y > u_2\} - P\{X > u_1 \textbf{ or } Y > u_2\} \\ &\approx 1 - F_1(u_1) + 1 - F_2(u_2) - \ell(1 - F_1(u_1), 1 - F_2(u_2)). \qquad (1.26)\end{aligned}$$

We will see in the following section that (1.26) is of little help for estimating probabilities like p if X and Y are asymptotically independent.

1.2.3 Asymptotic Independence

Suppose that the random variables X and Y are standardized to uniform random variables $U = 1 - F_1(X)$ and $V = 1 - F_2(Y)$. Then, X and Y (or U and V) are called *asymptotically independent* if

$$\lim_{t\downarrow 0} P(U < t \mid V < t) = \lim_{t\downarrow 0} t^{-1} P\{U < t \textbf{ and } V < t\} = 0. \qquad (1.27)$$

Loosely speaking, this means that the probability of a joint occurrence of large values of X (i.e., small values of U) and large values of Y (i.e., small values of V) vanishes asymptotically.

Asymptotic independence is a rather common and therefore important situation, since real world data often exhibit this property, see e.g. Chapter 5. However, within the class of multivariate extreme value distributions, the only possible type of asymptotic independence is perfect independence. In the case of asymptotic independence, observed data are likely to exhibit reasonably strong dependence. Then, the class of multivariate extreme value distributions is not the appropriate model in this case. The aim of this section is to briefly illustrate this. Thereby, we indicate the necessity of models which overcome this drawback and in turn motivate the investigations starting with Chapter 2 of this work. For further discussion of this issue see e.g. Section 9.5 of Beirlant et al. (2004), Section 5 of Draisma et al. (2004), Section 4.4 of Coles et al. (1999) or the results in Chapter 5.

The approach (1.26) fails in the situation of asymptotic independence of U and V, since from (1.27) we obtain for the stable tail dependence function

$$\begin{aligned} \ell(x,y) &= \lim_{t\downarrow 0} t^{-1} P\{U < tx \textbf{ or } V < ty\} \\ &= \lim_{t\downarrow 0} t^{-1} \big(P\{U < tx\} + P\{V < ty\} - P\{U < tx, V < ty\}\big) \\ &= \lim_{t\downarrow 0} t^{-1} \big(tx + ty - P\{U < tx, V < ty\}\big) \\ &= \lim_{t\downarrow 0} x + y - t^{-1} P\{U < tx, V < ty\} = x + y. \end{aligned} \tag{1.28}$$

This means for the probability of a jointly large claim

$$\begin{aligned} p &= P\{X > u_1\} + P\{Y > u_2\} - P\{X > u_1 \textbf{ or } Y > u_2\} \\ &\approx 1 - F_1(u_1) + 1 - F_2(u_2) - \ell(1 - F_1(u_1), 1 - F_2(u_2)) = 0, \end{aligned}$$

which is of little help for estimating p. In other words, from (1.28) we obtain that $t^{-1}P\{U < tx, V < ty\}$ has a degenerate limit 0, i.e. we only know that $P\{U < tx, V < ty\} = o(t)$. What we need is a different standardization in order to obtain a non-degenerate limit.

The approach to solving this problem presented in the following chapter is to specify the speed of convergence in (1.27). To this end, we impose a bivariate second-order condition (cf. Section 1.1.3) on the joint survivor function $P\{U < x, V < y\}$.

Chapter 2

Ledford's and Tawn's Approach and a Related Model

In Section 1.2.3, we have indicated problems which arise in CMEVT in the case of asymptotic independence. We will see that the models discussed in this chapter overcome these problems. We are concerned with the original and the Extended Ledford and Tawn Model(s), where the latter is the object of all further investigations.

In Section 2.1, we start with a brief presentation of the relation between the simpler model of Ledford and Tawn (1996) and the later model of Ledford and Tawn (1997, 1998). Although the theory and applications from Section 2.2 until Chapter 5 hardly ever refer to the original Ledford and Tawn Model(s) (OLTM), Section 2.1 presents the OLTM in some detail. This is mainly done to ease the comparison to the Extended Ledford and Tawn Model (ELTM), which is considered from Section 2.2 on. For example, we specify the assumptions of the OLTM which are not necessary in the ELTM, indicate why these assumptions are made and why they are not required in the ELTM. Thereby, we set the stage to clarify in Section 2.2 in which sense the term "Extended" is to be understood. However, details on e.g. statistical aspects of the OLTM are almost completely omitted. The ELTM is introduced in Section 2.2. Its central condition and the immediate implications, established in Section 2.2.1, play a crucial role in the rest of this work. The scaling law deduced in Section 2.2.2 will also accompany us throughout. The chapter is concluded with the proof of a corollary of Section 2.2.1.

2.1 The Ledford and Tawn Approach

Instead of referring to results that are based on (1.24), Ledford and Tawn (1996, 1997, 1998) model the asymptotic form of the joint survivor function directly. This allows to specify the speed of convergence in (1.27) and hence to overcome the problem illustrated in Section 1.2.3.

In Ledford and Tawn (1996), it is assumed that the joint survivor function varies regularly with index $-1/\eta \in [-2, -1]$ along the diagonal. More precisely, the main assumption is that for standard Fréchet variables Z_1 and Z_2, i.e. $P\{Z_j \le z\} = \exp(-1/z)$, $j = 1, 2$,

$$P\{Z_1 > r, Z_2 > r\} \sim \mathcal{L}(r) r^{-1/\eta} \tag{2.1}$$

as $r \to \infty$, where $\mathcal{L}$ is a slowly varying function. Since the boundary cases $\eta = 1/2$ and $\eta = 1$ correspond to exact independence and perfect dependence, respectively, Ledford and Tawn (1996) named η *coefficient of tail dependence.* (A more thorough interpretation of η in the extended model is given in Section 2.2.1.) In order to ease comparison with the ELTM, we rephrase (2.1) in terms of the joint distribution of $U = 1 - \exp(-1/Z_1)$ and $V = 1 - \exp(-1/Z_2)$. Essentially, (2.1) means the same for the distribution tail of (U, V), since (2.1) is equivalent to

$$P\big\{U < 1 - \exp(-1/r), V < 1 - \exp(-1/r)\big\} \sim \mathcal{L}(r) r^{-1/\eta}$$

and by $t = 1 - \exp(-1/r)$ and $-\log(1-t) = t + O(t^2)$ as $t \downarrow 0$ to

$$\begin{aligned} P\{U < t, V < t\} &\sim \mathcal{L}\left(\frac{1}{-\log(1-t)}\right)\big(-\log(1-t)\big)^{1/\eta} \\ &\sim \mathcal{L}^*(t) t^{1/\eta} \end{aligned} \tag{2.2}$$

as $t \downarrow 0$, where the function $\mathcal{L}^*$ is slowly varying at 0.

Ledford and Tawn (1996) solely model the joint survivor function along the diagonal and $\eta \in [1/2, 1]$ only allows for positive dependence. These drawbacks are resolved in the models of Ledford and Tawn (1997, 1998). The basic condition as stated in (2.3) of Ledford and Tawn (1998) is that there exist some positive normalizing function d and some non-vanishing function ψ such that

$$\lim_{r\to\infty} \frac{P\{Z_1 > rx, Z_2 > ry\}}{d(r)} = \psi(x, y) \tag{2.3}$$

for all $x, y > 0$. Relation (2.3) may be regarded as a bivariate regularity condition (see e.g. the Appendices in de Haan and Resnick (1993) and Draisma et al. (2003)) for the joint survivor function $P\{Z_1 > x, Z_2 > y\}$. If $\eta > 1/2$, it is equivalent to the condition used by de Haan and Resnick (1993). It is shown in Theorem 1 of Ledford and Tawn (1998) that (2.3) implies

$$P\{Z_1 > x, Z_2 > y\} = \frac{\mathcal{L}(x, y)}{x^{c_1} y^{c_2}}, \tag{2.4}$$

where $c_1, c_2 > 0$, $c_1 + c_2 = 1/\eta \in [1, \infty)$ and $\mathcal{L}(x, y)$ denotes a bivariate slowly varying function with limit g, i.e.

$$\lim_{r\to\infty} \frac{\mathcal{L}(rx, ry)}{\mathcal{L}(r, r)} =: g(x, y)$$

with a function g that is homogeneous of order 0, i.e. $g(sx, sy) = g(x, y)$ for all $s > 0$. We see that $g(x, y)x^{-c_1}y^{-c_2}$ is homogeneous of order $-1/\eta$ and from (2.4) that $P\{Z_1 > x, Z_2 > x\}$ is regularly varying with index $-1/\eta$. (In fact, we obtain an even stronger result. The homogeneity property of $g(x, y)x^{-c_1}y^{-c_2}$ together with (2.4) implies the multivariate regular variation of $P\{Z_1 > x, Z_2 > y\}$, cf. Appendix 1.4 of Bingham et al. (1987) or Resnick (1987), (5.32).) Note that the structural form (2.4) of the joint survivor function is rather arbitrary, since $\mathcal{L}_\kappa(x, y) := (x/y)^\kappa \mathcal{L}(x, y)$ is bivariate slowly varying and $\mathcal{L}(x, y)x^{-c_1}y^{-c_2} = \mathcal{L}_\kappa(x, y)x^{-c_1-\kappa}y^{-c_2+\kappa}$, so that c_1 and c_2 are only subject to the condition $c_1 + c_2 = 1/\eta$. Therefore, in order to develop a method for statistical inference, Ledford and Tawn (1997, 1998) impose further structure on $\mathcal{L}$ which allows the separate identification of c_1 and c_2. We consider the further assumptions made by Ledford and Tawn (1997, 1998) before we rephrase (2.3) in terms of U and V.

Let $w = x/(x + y) \in (0, 1)$. Observe that from (2.3)

$$\lim_{r\to\infty} \frac{P\{Z_1 > rw, Z_2 > r(1-w)\}}{d(r)} = \psi(w, 1-w)$$

and

$$\lim_{r\to\infty} \frac{P\{Z_1 > r(1-w), Z_2 > rw\}}{d(r)} = \psi(1-w, w),$$

such that

$$\frac{\psi(w, 1-w)}{\psi(1-w, w)} = \lim_{r\to\infty} \frac{P\{Z_1 > rw, Z_2 > r(1-w)\}}{P\{Z_1 > r(1-w), Z_2 > rw\}}.$$

Ledford and Tawn (1997, 1998) assume that $\psi(w, 1-w)/\psi(1-w, w)$ is regularly varying at $w = 0$. (This implies regular variation at $w = 1$, which becomes obvious when considering what the regular variation at $w = 0$ means for the reciprocal function.) It follows that $\mathcal{L}$ is what Ledford and Tawn (1997, 1998) call quasi-symmetric, i.e. $g_*(w)/g_*(1-w)$ is slowly varying at $w = 0$ (and $w = 1$), where $g_*(w) := g_*(x/(x+y)) := g(x, y)$. By means of this condition, c_1 and c_2 are separately identifiable beyond the condition $c_1 + c_2 = 1/\eta$. This is necessary for the statistical inference proposed by Ledford and Tawn (1997), see Section 4 of that paper. Since we develop a different statistical methodology, this regular variation condition on ψ is one of the assumption which are not needed in the ELTM.

Furthermore, Ledford and Tawn (1998) assume that $\mathcal{L}$ has an asymptotic expansion of the form

$$\mathcal{L}(x, y) = \mathcal{L}_1(x, y) + \frac{\mathcal{L}_2(x, y)}{x^{d_1} y^{d_2}} + O\big(\gamma(x, y)\big) \tag{2.5}$$

with quasi-symmetric bivariate slowly varying functions $\mathcal{L}_j \not\equiv 0$, $d_1, d_2 \geq 0$, $\mathcal{L}_2(rx, ry) = o\big(\mathcal{L}_1(rx, ry)\big)$ if $d_1 = d_2 = 0$ and with a function γ such that $\gamma(rx, ry) = o\big(\mathcal{L}_2(rx, ry)(rx)^{-d_1}(ry)^{-d_2}\big)$ as $r \to \infty$. Hence, the asymptotic expansion used as the model for the joint survivor function reads as

$$P\{Z_1 > x, Z_2 > y\} = \frac{\mathcal{L}_1(x, y)}{x^{c_1} y^{c_2}} + \frac{\mathcal{L}_2(x, y)}{x^{c_1+d_1} y^{c_2+d_2}} + \ldots, \tag{2.6}$$

which is explicitly assumed in Ledford and Tawn (1997). The statistical methodology developed in Ledford and Tawn (1997) is based on approximating $P\{Z_1 > x, Z_2 > y\}$ by the leading term on the right hand side of (2.6), while the second term allows to assess the quality of this approximation and to identify the rates of convergence. The latter, however, turns out to be difficult in most cases, see Section 8 of Ledford and Tawn (1997). The assumption of the existence of an asymptotic expansion (2.5) for $\mathcal{L}$ is the second condition which is not needed in the ELTM. For a more thorough treatment of the OLTM see also Ledford (1996).

Rephrased in terms of the joint distribution of the uniform random variables U and V, (2.3) reads as

$$\lim_{r\to\infty} \frac{P\big\{U < 1 - \exp(-1/(rx)), V < 1 - \exp(-1/(ry))\big\}}{d(r)} = \psi(x, y)$$
$$\Leftrightarrow \lim_{t\downarrow 0} \frac{P\{U < 1 - \exp(-tx), V < 1 - \exp(-ty)\}}{d(1/t)} = \psi(1/x, 1/y)$$

for all $x, y > 0$. If one assumes that this convergence holds locally uniformly in x and y, i.e. there is a neighborhood of (x, y) where the convergence is uniform, then a simple Taylor expansion shows that this condition is equivalent to

$$\lim_{t\downarrow 0} \frac{P\{U < tx, V < ty\}}{d(1/t)} = \psi(1/x, 1/y) \tag{2.7}$$

for all $x, y > 0$. When we introduce the ELTM in the following section, we merely assume a rate of (uniform) convergence in (2.7). In particular, the regular variation of $\psi(w, 1-w)/\psi(1-w, w)$ and the more complicated asymptotic expansion of $P\{U < 1 - \exp(-1/(tx)), V < 1 - \exp(-1/(ty))\}$ (or $\mathcal{L}$, respectively) are not needed.

2.2 The Extended Ledford and Tawn Model

For the rest of this work, we assume that the marginal d.f.s F_j are *tail continuous*, i.e. there exists $\xi_j < F_j^{\leftarrow}(1)$, such that F_j is continuous on $[\xi_j, \infty)$, $j = 1, 2$. Hence the d.f.s of $U = 1 - F_1(X)$ and $V = 1 - F_2(Y)$ are equal to the uniform d.f. on a small neighborhood of 0.

2.2.1 The Central Condition and its Implications

We weaken the model assumptions of Ledford and Tawn (1998) in that we only specify a rate of (uniform) convergence in (2.7). (This ELTM was already considered by Draisma et al. (2004); in particular, see Remark 2.1 of that paper.)

2.2.1 Condition. Assume that

$$\frac{P\{U < tx, V < ty\}}{q(t)} - c(x,y) = O\big(q_1(t)\big) \tag{2.8}$$

as $t \downarrow 0$ for all $x, y \geq 0$ and uniformly on $\{(x,y) | \max(x,y) = 1\}$. Here c is some non-degenerate function and q and q_1 are positive functions that tend to 0 as $t \downarrow 0$, where q_1 is regularly varying at 0 with some index $\tau \geq 0$, i.e.

$$\lim_{t\downarrow 0} \frac{q_1(tx)}{q_1(t)} = x^\tau.$$

$\triangle$

As usual, (2.8) means that

$$\limsup_{t\downarrow 0} \left| \frac{P\{U < tx, V < ty\}}{q(t)} - c(x,y) \right| \Big/ q_1(t) < \infty,$$

for all $x, y \geq 0$, i.e. $q_1(t)$ is an asymptotic upper bound of (the absolute value of) the difference on the left hand side of (2.8). The function q_1 is sometimes called the (maximum) rate of convergence of $P\{U < tx, V < ty\}/q(t)$ to $c(x,y)$. Note that since $q_1(t) \to 0$ as $t \downarrow 0$,

$$\limsup_{t\downarrow 0} \left| \frac{P\{U < tx, V < ty\}}{q(t)} - c(x,y) \right| = 0$$

$$\Rightarrow \lim_{t\downarrow 0} \frac{P\{U < tx, V < ty\}}{q(t)} = c(x,y), \tag{2.9}$$

i.e. (2.8) implies (2.7), where $d(1/t)$ and $\psi(1/x, 1/y)$ play the roles of $q(t)$ and $c(x,y)$.

Similar as (2.3), Condition (2.8) is essentially a bivariate second-order condition for the so-called survival copula

$$Q(x,y) := P\{U < x, V < y\}.$$

We consider a copula approach related to the ELTM in Section 6. For more information on such conditions, see (the Appendices of) de Haan and Resnick (1993) and Draisma et al. (2003).

Observe that we may take $c(1,1) = 1$ in Condition 2.2.1 without loss of generality (w.l.o.g). We obtain from Condition 2.2.1 with $x = y = 1$ that

$$\begin{aligned}\frac{Q(tx,ty)}{Q(t,t)} - c(x,y) &= \left(\frac{Q(tx,ty)}{q(t)} - c(x,y)\right)\frac{q(t)}{Q(t,t)} + c(x,y)\left(\frac{q(t)}{Q(t,t)} - 1\right)\\ &= O\big(q_1(t)\big)\frac{q(t)}{Q(t,t)} + O\big(q_1(t)\big) = O\big(q_1(t)\big),\end{aligned} \tag{2.10}$$

and that conversely, if $Q(tx,ty)/Q(t,t) - c(x,y) = O\big(q_1(t)\big)$ holds, then Condition 2.2.1 is satisfied, so that we may take w.l.o.g.

$$q(t) = Q(t,t) = P\{U < t, V < t\}.$$

Then, for $x, y \in [0,1]$, convergence (2.9) reads as

$$\lim_{t\downarrow 0} P(U < tx, V < ty \mid U < t, V < t) = c(x,y),$$

i.e. the conditional probability $P(U < tx, V < ty \mid U < t, V < t)$ converges to the limit $c(x,y)$. In this sense, (2.9) is a natural analog to the univariate EVT condition (1.4).

For sufficiently small x and y, the link between the survival copula Q and the tail dependence function D which plays a crucial role in CMEVT (cf. Section 1.2) is given by

$$Q(x,y) = x + y - D(x,y).$$

Further, if

$$l := \lim_{t\downarrow 0}\frac{q(t)}{t} \tag{2.11}$$

exists (which is always satisfied with $l = 0$ if U and V are asymptotically independent, since l is precisely the left hand side of (1.27)), then (2.9) implies (1.25), i.e. the condition of the Ledford and Tawn model(s) implies the convergence of the dependence structure in terms of the tail dependence function. To see this, note that (2.9) yields

$$\ell(x,y) = \lim_{t\downarrow 0}\frac{D(tx,ty)}{t} = x + y - \lim_{t\downarrow 0}\frac{Q(tx,ty)}{t} = x + y - lc(x,y) \tag{2.12}$$

for sufficiently small x and y.

We summarize further comments on Condition 2.2.1 in the following

2.2.2 Remarks.

(i) Compared to the assumptions of the OLTM, Condition 2.2.1 is a relaxation in that we do not require the regular variation of $\psi(w,1-w)/\psi(1-w,w)$ at $w=0$ and the more complicated expansion of $P\{U < 1-\exp(-1/(tx)), V < 1-\exp(-1/(ty))\}$. On the other hand, we slightly strengthen Ledford's and Tawn's conditions in that we require uniformity on $\{(x,y) | \max(x,y)=1\}$ (cf. the following remark). However, this slight strengthening is of course less severe than the previously mentioned additional assumptions of the OLTM.

(ii) Since we will not make explicit use of the assumed uniformity of (2.8) on $\{(x,y) | \max(x,y)=1\}$, we remark that, nonetheless, it is an essential element of the developed theory. Many of the results we refer to need this assumption. In particular, the results associated with the asymptotic normality of estimators of the coefficient of tail dependence in Sections 2 and 6 of Draisma et al. (2004), which are used several times, require this uniformity. The results of Section 2 of Draisma et al. (2004) on the estimation of η in the ELTM are summarized in Section 3.1 of this work.

(iii) Condition 2.2.1 can be described as "hidden regular variation" with rate. The notion of hidden regular variation is first introduced in Resnick (2001). Further theory is developed in subsequent papers, see e.g. Maulik and Resnick (2004). Section 2.4 of Heffernan and Resnick (2005) contains a brief comparison with the OLTM.

$\triangle$

The following corollary, proved in Section 2.3, states the most important implications of Condition 2.2.1.

2.2.3 Corollary. From Condition 2.2.1 we obtain that

(i) the function q is regularly varying at 0 with some index $1/\eta \in [1,\infty)$, i.e.

$$\lim_{t\downarrow 0} \frac{q(tx)}{q(t)} = x^{1/\eta}, \tag{2.13}$$

(ii) c is homogeneous of order $1/\eta$, i.e.

$$c(sx,sy) = s^{1/\eta} c(x,y), \tag{2.14}$$

for all $s \geq 0$ and thus continuous on $[0, \infty)^2 \setminus (0, 0)$.

$\triangle$

2.2.4 Remark. Recall that w.l.o.g. $q(t) = P\{U < t, V < t\}$ to see by (2.2) that the reciprocal η of the index of the regular variation of q indeed plays the role of Ledford's and Tawn's coefficient of tail dependence. $\triangle$

We continue with considering the behavior of the model in the case of asymptotic independence and thereby give a more thorough justification to name η coefficient of tail dependence. From the regular variation (2.13) of q, the function

$$L(t) := t^{-1/\eta} q(t)$$

is slowly varying in 0. Thus, for sufficiently small t,

$$P(U < t \mid V < t) = t^{-1} q(t) = t^{1/\eta - 1} L(t), \tag{2.15}$$

which, by Proposition 1.5.1 of Bingham et al. (1987), converges to 0 if $\eta < 1$. Therefore, η and the function L possess the interpretation

$$\begin{aligned} \eta < 1 \textbf{ or } L(t) \xrightarrow{t \downarrow 0} 0 \quad &\iff \quad \text{asymptotic independence,} \\ \eta = 1 \textbf{ and } L(t) \overset{t \downarrow 0}{\not\to} 0 \quad &\iff \quad \text{asymptotic dependence.} \end{aligned}$$

Further, we see that if the limit of (2.15) as $t \downarrow 0$ exists, η is a measure of the speed of convergence of $P(U < t \mid V < t)$. If U and V are exactly independent, we see from the regular variation (2.13) of q that

$$x^{1/\eta} = \lim_{t \downarrow 0} \frac{q(tx)}{q(t)} = \lim_{t \downarrow 0} \frac{P\{U < tx, V < tx\}}{P\{U < t, V < t\}} = \lim_{t \downarrow 0} \frac{t^2 x^2}{t^2} = x^2,$$

so that $\eta = 1/2$. Then (2.8) holds with $c(x, y) = xy$ and $q_1 \equiv 0$.

The cases $\eta \in (0, 1/2)$ and $\eta \in (1/2, 1)$ correspond to asymptotically vanishing negative dependence and to asymptotically vanishing positive dependence, respectively. As an example, consider the case of (asymptotically vanishing) positive dependence with $x \in (0, 1)$. Then, the ratio $q(tx)/q(t)$ equals the conditional probability $P(U < tx, V < tx \mid U < t, V < t)$. Due to the (asymptotically vanishing) positive dependence of U and V, this conditional probability must be greater than in the case of exact independence of U and V, i.e. $x^{1/\eta} = \lim_{t\downarrow 0} q(tx)/q(t) > x^2$, so that $\eta > 1/2$. The other cases can be treated analogously.

Next we reinterpret the convergence (2.9) as a scaling law. This scaling law plays an important role for the methods of statistical inference and model validation developed in Chapters 3 and 4.

2.2.2 Reinterpretation as Scaling Law

Note that (2.9) and the homogeneity of c imply

$$\lim_{t\downarrow 0} \frac{P\{U < tsx, V < tsy\}}{P\{U < t, V < t\}} = c(sx, sy) = s^{1/\eta} c(x, y) \tag{2.16}$$

for $s, x, y \geq 0$. Combine (2.9) and (2.16) to obtain

$$\lim_{t\downarrow 0} \frac{P\{U < tsx, V < tsy\}}{P\{U < tx, V < ty\}} = s^{1/\eta}. \tag{2.17}$$

We obtain an analogous result for more general sets. The measure μ defined by

$$\mu\big([0, x] \times [0, y]\big) := c(x, y) \tag{2.18}$$

inherits the homogeneity of c, i.e. $\mu(tM) = t^{1/\eta}\mu(M)$ for all bounded measurable sets $M \subset [0, \infty)^2$ and all $t > 0$. Further, recall that if $x, y \in [0, 1]$, (2.9) may be considered as the convergence of the conditional probability $P(U < tx, V < ty \mid U < t, V < t)$ to $c(x, y)$, so that weak convergence holds for the pertaining measures. Thus, by the Portmanteau theorem (see e.g. Theorem 2.1 of Billingsley (1968)) and similar arguments as before,

$$\lim_{t\downarrow 0} \frac{P\{(U, V) \in tsB\}}{P\{(U, V) \in tB\}} = s^{1/\eta} \tag{2.19}$$

for all $s \in [0, 1]$, provided that μ assigns positive mass to $B \subset [0, 1]^2$ and has no mass on the boundary of B.

Thus, the approximate scaling law $P\{(U, V) \in sA\}/P\{(U, V) \in A\} \approx s^{1/\eta}$ holds for suitable sets $A \subset [0, 1]^2$ nearby the origin. This means that contracting a set $A \subset [0, 1]^2$ by a *contraction factor* $s \in [0, 1]$ leads to a

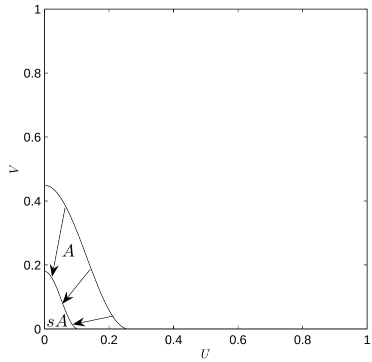


Figure 2.1: The scaling law assumed by the model. The ratio of the probabilities of the events sA and A is supposed to be approximately equal to $s^{1/\eta}$.

decrease of the pertaining probability by the factor $s^{1/\eta}$. Figure 2.1 illustrates the scaling law at work.

Having introduced this scaling law, we sketch an idea of how to estimate the probability of a jointly large claim like $p = P\{X > u_1, Y > u_2\}$ in the ELTM. In fact, this idea is used in Section 3.3. Let

$$U_i := 1 - F_1(X_i) \quad \text{and} \quad V_i := 1 - F_2(Y_i), \quad i = 1, \dots, n, \tag{2.20}$$

respectively. By means of the scaling law (2.17), we see that

$$\begin{aligned} p &= P\big\{U < 1 - F_1(u_1),\ V < 1 - F_2(u_2)\big\} \\ &\approx r^{-1/\eta} \cdot P\big\{U < r(1 - F_1(u_1)),\ V < r(1 - F_2(u_2))\big\} \\ &\approx r^{-1/\eta} \cdot \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\big\{U_i < r(1 - F_1(u_1)),\ V_i < r(1 - F_2(u_2))\big\}} \end{aligned} \tag{2.21}$$

for "some appropriate $r > 0$". We now consider the idea behind this approximation to see what "some appropriate $r > 0$" means. Usually, we are interested in the probability of a jointly large claim $\{U < 1 - F_1(u_1),\ V < 1 - F_2(u_2)\}$ with u_1 and u_2 so large that only a few or no observations with both components above the pertaining threshold are available. The idea of (2.21) is to inflate the set $\{U < 1 - F_1(u_1),\ V < 1 - F_2(u_2)\}$ with factor r to obtain a reasonable number of observations to base an estimator of p on. Then, by the scaling law, the (estimated) probability of this inflated set is scaled down by (an estimate of) the factor $r^{-1/\eta}$. Hence, $r > 0$ must be chosen sufficiently small to satisfy the approximate scaling law and sufficiently large to obtain a reasonable number for the sum on the right hand side of (2.21). We have presented the way we estimate $1 - F_i(u_i)$, $i = 1, 2$, in Sections 1.1.2 and 1.1.4. Further, in order to obtain a suitable estimator of p based on (2.21), we substitute U_i and V_i by empirical counterparts and present suitable estimators of η in Section 3.1. The sketched procedure indicates that the analysis of the total estimation error of an estimator of p is threefold – we must allow for the errors occurring when estimating η and the marginals as well as for the error when applying the approximate scaling law. This is done in Section 3.3. Further, in Chapter 4, we establish a model validation tool that helps to decide whether the scaling law holds.

Before we conclude this chapter with the proof of Corollary 2.2.3, we make some remarks on the idea behind (2.21) in view of related models. Representation (2.21) emphasizes that the ELTM – like the CMEVT by means of Proposition 1.2.1 – separates modeling the joint tail distribution of X and Y into two parts describing the tails of the marginal d.f.s F_1 and F_2 on the one hand and the tail dependence structure (by η) on the other hand. The reader who is familiar with other methods of modeling multivariate distributions may recognize similarities with (parametric) copula approaches which recently have enjoyed great popularity. (Recall that the OLTM and the ELTM may also be considered as a (non-parametric) copula approach, since they assume a condition on the copula Q, cf. Sections 2.1 and 2.2.1.) In contrast to most copula approaches, however, the ELTM focusses on the distribution tail(s) with respect to the marginal distributions and the dependence structure and therefore allows a separate modeling of large claims with respect to the marginal distributions and the dependence structure.

We present a copula approach which is rather closely related to the ELTM and continue the discussion more thoroughly in Section 6.

2.3 Proof of Corollary 2.2.3

For (i), choose $y = x$ in (2.9) to obtain

$$c(tx, tx) = \lim_{s\downarrow 0} \frac{q(stx)}{q(s)} = \lim_{s\downarrow 0} \frac{q(stx)}{q(st)} \frac{q(st)}{q(s)} = c(x, x)c(t, t),$$

which is Cauchy's power equation with solution $c(x, x) = x^\alpha$ for some $\alpha \in \mathbb{R}$. Hence, by (2.9),

$$\lim_{t\downarrow 0} \frac{q(tx)}{q(t)} = \lim_{t\downarrow 0} \frac{P\{U < tx, V < tx\}}{q(t)} = c(x, x) = x^\alpha.$$

To see that the index α of the regular variation of q is indeed not less than 1, note that

$$q(t) = P\{U < t, V < t\} \leq P\{U < t\} = t$$

for sufficiently small $t > 0$.

To prove (ii), note that by the regular variation of q,

$$\begin{aligned} c(sx, sy) &= \lim_{t\downarrow 0} \frac{P\{U < tsx, V < tsy\}}{P\{U < t, V < t\}} \\ &= \lim_{t\downarrow 0} \frac{P\{U < tsx, V < tsy\}}{P\{U < ts, V < ts\}} \frac{P\{U < ts, V < ts\}}{P\{U < t, V < t\}} \\ &= c(x, y) \lim_{t\downarrow 0} \frac{q(ts)}{q(t)} = s^{1/\eta} c(x, y) \end{aligned}$$

for all $s > 0$ and all $x, y \geq 0$, which proves the homogeneity of c. Further, for $x, y \in [0, \infty)^2 \setminus (0, 0)$, assume that c is not continuous. Then the measure μ, introduced in (2.18), has positive mass on some point (x, y), i.e. $\mu\{(x, y)\} > 0$. Let $t = \max(x, y) > 0$ and denote by $E := \{(\bar{x}, \bar{y}) : \max(\bar{x}, \bar{y}) = t\}$ the boundary of the square $(0, t]^2 \cup \{(0, t)\} \cup \{(t, 0)\}$. Then, since $c(t, t) = t^{1/\eta}$ is continuous for all $t > 0$, $\mu(E) = c(t, t) - c(t-, t-) = 0$, where $f(x-, y-)$ denotes the left hand limit of f at (x, y) in the sense of Definition A.1. Hence, since $(x, y) \in E$, $\mu\{(x, y)\} = 0$.

Chapter 3

Estimation in the Extended Ledford and Tawn Model

In this chapter we present methods to estimate the model parameters and the probability of jointly large claims p, defined by (1.23). From (2.21), it is immediately clear that we need a suitable estimator for the coefficient of tail dependence η. We address this issue in Section 3.1. The presentation of the results on asymptotic normality of the considered estimators of η is rather brief, since with the exception of one minor correction they are known from Draisma et al. (2004). Thereafter, we establish asymptotic expansions for these estimators which are necessary to prove the main results of Section 3.3 and Chapter 4. In Section 3.2 we attend to the estimation of c. We prove that the suitably standardized process of the difference of c and the proposed estimator of c converges to a centered Gaussian process. We conclude asymptotic normality for the proposed estimator of $c(x, y)$ in both cases, asymptotic independence and asymptotic dependence. Similar as with the estimation of γ (and η, as we see in Section 3.1), the estimator of c is based on a fraction of – in a certain sense – largest order statistics of the joint tail of U and V. We discuss the important issue of an appropriate choice of the parameter that fixes this fraction in Section 3.2.2, where we consider a random variable which is suggested by the estimation method of η as a natural candidate for this parameter and show that asymptotic normality still holds for this choice. In Section 3.3 we establish asymptotic normality of an estimator of p if we use the estimator (1.10) for the marginal tails and the estimators for η considered in Section 3.1. As before, the number of observations used for the estimation of p has to be suitably chosen. We discuss this issue in Section 3.3.2 and devise a graphical tool to treat this problem appropriately. The proofs of all results of this chapter are deferred to the final Section 3.4.

3.1 Estimation of the Coefficient of Tail Dependence η

As before, let $\{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}$ be a sample of i.i.d. claim sizes and denote by F_1 and F_2 the marginal d.f.s of X_i and Y_i, $i = 1, \dots, n$, respectively. Further, recall from Section 2.2 the definitions (2.20) of U_i and V_i, respectively, that we assumed F_j, $j = 1, 2$, to be tail continuous and that we may take w.l.o.g. $q(t) = Q(t,t) = P\{U < t, V < t\}$ in Condition 2.2.1.

3.1.1 The Hill Estimator and the Maximum Likelihood Estimator

In (2.21) we gave an idea of how to obtain estimators of p. This equation reveals that the estimation of η is a crucial step if we want to estimate p, see also Section 3.3. We relate this problem to the estimation of the extreme value index in a (generalized) Pareto model.

As in Draisma et al. (2004), we consider the random variables

$$T_i := \min\left(\frac{1}{U_i}, \frac{1}{V_i}\right),$$

$i = 1, \dots, n$. Denote by F_T the d.f. of T_i, $i = 1, \dots, n$. We obtain the following simple

3.1.1 Corollary. The function $1 - F_T(t)$ equals $q\,(1/t)$ and is regularly varying at ∞ with index $-1/\eta$. $\triangle$

Since Corollary 3.1.1 shows that the survival function $1 - F_T$ satisfies the condition (1.11) with index $-1/\eta$, we consider estimators of η of the same type as those in Section 1.1.4. Recall that in this setting, the Hill estimator $m^{-1} \sum_{i=1}^{m} \log\left(T_{n-i+1:n}/T_{n-m:n}\right)$, where $T_{i:n}$, $i = 1, \dots, n$, are the order statistics pertaining to T_i, and the maximum likelihood estimator as the first component of the solution of (1.17) with $X_{i:n}$ replaced with $T_{i:n}$ are popular estimators for the extreme value index η with good asymptotic properties, cf. Section 1.1.4. However, observe that these direct analogs to $\hat{\gamma}_n^H$ and $\hat{\gamma}_n^{ML}$ are of little use, since the marginal d.f.s F_1 and F_2, which are necessary to calculate $T_{n-i:n}$, $i = 0, \dots, m$, are unknown. Therefore, we replace U_i and V_i, $i = 1, \dots, n$, by empirical counterparts. Let

$$\hat{U}_i := 1 - \frac{R_i^X}{n+1} \qquad \text{and} \qquad \hat{V}_i := 1 - \frac{R_i^Y}{n+1}, \tag{3.1}$$

where R_i^X and R_i^Y denote the ranks of X_i and Y_i among $(X_1, X_2, \dots, X_n)$ and $(Y_1, Y_2, \dots, Y_n)$, respectively. Further, let

$$T_i^{(n)} := \min\left(\frac{1}{\hat{U}_i}, \frac{1}{\hat{V}_i}\right) = \min\left(\frac{n+1}{n+1-R_i^X}, \frac{n+1}{n+1-R_i^Y}\right)$$

and denote by $T_{i:n}^{(n)}$, $i = 1, \dots, n$, the pertaining order statistics. Hence, we define the Hill estimator of η by

$$\hat{\eta}_n^H := \frac{1}{m} \sum_{i=1}^m \log \frac{T_{n-i+1:n}^{(n)}}{T_{n-m:n}^{(n)}}, \tag{3.2}$$

which also has good asymptotic properties, see Section 3.1.2. Here, $m := m_n$ is (as j in the univariate case) an intermediate sequence (that is, $m \to \infty$ and $m/n \to 0$ as $n \to \infty$) of the number of order statistics used to estimate η and indeed depends on the sample size n. We have discussed an appropriate choice of j in Section 1.1.4. Although the situation is slightly different for m, since the random variables $T_i^{(n)}$ are not i.i.d., the heuristic is the same.

Similarly, we replace the order statistics $X_{i:n}$ with $T_{i:n}^{(n)}$, $i = 1, \dots, n$, (j with m and γ with η, if you like) in (1.17) to obtain the maximum likelihood estimator (MLE) $\hat{\eta}_n^{ML}$ of η as the first component of the solution of the resulting system of equations.

Observe that $\hat{\eta}_n^H$ and $\hat{\eta}_n^{ML}$ are based on the $m+1$ largest order statistics $T_i^{(n)} \geq T_{n-m:n}^{(n)}$. Since $T_i^{(n)} = 1/\max\big(\hat{U}_i, \hat{V}_i\big)$, it follows that we use those data points $\big(\hat{U}_i, \hat{V}_i\big)$ which lie within the square $\big(0, 1/T_{n-m:n}^{(n)}\big]^2$ to obtain $\hat{\eta}_n^H$ and $\hat{\eta}_n^{ML}$. (This is also clear from the simple equality $m + 1 = \sum_{i=1}^n \mathbb{1}\big\{\hat{U}_i \leq 1/T_{n-m:n}^{(n)}, \hat{V}_i \leq 1/T_{n-m:n}^{(n)}\big\}$, which holds if there is exactly one i_0 with $T_{i_0}^{(n)} = T_{n-m:n}^{(n)}$, see also Footnote 1 in Section 3.2.2.) Figure 3.1 indicates such a square for an example which is examined in detail in Chapter 5.

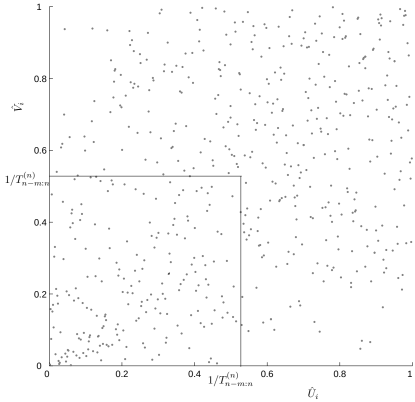


Figure 3.1: The points within the square $\left(0, 1/T^{(n)}_{n-m:n}\right]^2$ are used by the Hill estimator $\hat{\eta}_n^H$ and the MLE $\hat{\eta}_n^{ML}$.

3.1.2 Asymptotic Results on the Hill Estimator and the Maximum Likelihood Estimator

In this section, we summarize results on the asymptotic normality of $\hat{\eta}_n^H$ and $\hat{\eta}_n^{ML}$ and give confidence intervals for η and a test of the hypothesis $\eta = 1$. With the exception of one minor correction, these results are known from Draisma et al. (2004), in particular, see Theorems 2.1 and 2.2 and Section 4 of that paper. Therefore, the summary of these results is rather brief. Furthermore, we establish asymptotic representations of $\hat{\eta}_n^H$ and $\hat{\eta}_n^{ML}$ which are useful to prove the main results of Chapter 4.

Henceforth, by $\hat{\eta}_n$ we mean either the Hill estimator $\hat{\eta}_n^H$ or the MLE $\hat{\eta}_n^{ML}$.

Asymptotic Normality and Confidence Intervals

Suppose that

$$l = \lim_{t\downarrow 0} \frac{q(t)}{t} = \lim_{t\downarrow 0} t^{1/\eta-1} L(t),$$

defined in (2.11), exists. Recall that l is precisely the left hand side of (1.27). Therefore, this condition is always satisfied with $l = 0$ if U and V are asymptotically independent.

3.1.2 Theorem. (Theorem 2.1 of Draisma et al. (2004)) Assume that Condition 2.2.1 holds and that the function c has first-order partial derivatives $c_x := \partial c/\partial x$ and $c_y := \partial c/\partial y$ if U and V are asymptotically dependent. Assume further that m is an intermediate sequence such that $\sqrt{m} q_1\big(q^{\leftarrow}(m/n)\big) \to 0$ as $n \to \infty$. Then $\sqrt{m}\big(\hat{\eta}_n - \eta\big)$ converges to a normal distribution $\mathcal{N}\big(0, \sigma^2_{\hat{\eta}_n}\big)$, where

$$\sigma^2_{\hat{\eta}_n} := \begin{cases} \sigma^2_{\hat{\eta}_n^H} := \eta^2\big(1-l\big)\big(1-2lc_x(1,1)c_y(1,1)\big) & \text{if } \hat{\eta}_n = \hat{\eta}_n^H, \\ \\ \sigma^2_{\hat{\eta}_n^{ML}} := (1+\eta)^2\big(1-l\big)\big(1-2lc_x(1,1)c_y(1,1)\big) & \text{if } \hat{\eta}_n = \hat{\eta}_n^{ML}. \end{cases} \tag{3.3}$$

$\triangle$

3.1.3 Remarks.

(i) The convergence

$$\sqrt{m} q_1\big(q^{\leftarrow}(m/n)\big) \to 0 \tag{3.4}$$

requires that m may not converge too fast to ∞. It plays a similar role as the condition $\sqrt{j} A_i(n/j) \to 0$, $i = 1, 2$, for $\hat{\gamma}_n^H$ and $\hat{\gamma}_n^{ML}$ (cf. Section (1.1.4)), respectively, and is a rather weak assumption, see Remark 2.1(i) of Draisma et al. (2004).

(ii) Observe that $\sigma^2_{\hat{\eta}_n^H} = \eta^2$ and $\sigma^2_{\hat{\eta}_n^{ML}} = (1+\eta)^2$ if $l = 0$, i.e. iff U and V are asymptotically independent (which is always satisfied if $\eta < 1$). Moreover, $\sigma^2_{\hat{\eta}_n^H} \le \eta^2 \le 1$ and $\sigma^2_{\hat{\eta}_n^{ML}} \le (1+\eta)^2 \le 4$ is always true, since $l \in [0, 1]$. Observe further that $\sigma^2_{\hat{\eta}_n^H} \le \sigma^2_{\hat{\eta}_n^{ML}}$ so that, as a rule, we prefer to work with $\hat{\eta}_n^H$ in the applications in Chapter 5.

(iii) Note the similarity to the variance of the Hill estimator and the maximum likelihood estimator of γ: from (1.15) and (1.18) we know that $\sigma^2_{\hat{\gamma}_n^H} = \gamma^2$ and $\sigma^2_{\hat{\gamma}_n^{ML}} = (1+\gamma)^2$, which coincides with $\sigma^2_{\hat{\eta}_n^H} = \eta^2$ and $\sigma^2_{\hat{\eta}_n^{ML}} = (1+\eta)^2$ in the case of asymptotic independence.

$\triangle$

The following theorem provides consistent estimators for the unknown quantities in the asymptotic variance $\sigma^2_{\hat{\eta}_n}$.

3.1.4 Theorem. (slight correction of Theorem 2.2 of Draisma et al. (2004)) Let

$$\hat{l} := \frac{m}{n} T^{(n)}_{n-m:n},$$

$$\hat{c}_x(1,1) := \frac{\hat{k}^{5/4}}{n} \left(T^{(n,\hat{k}^{-1/4})}_{n-m:n} - T^{(n)}_{n-m:n} \right)$$

with $\hat{k} := m/\hat{l}$ and $T^{(n,u)}_{i:n}$, $i = 1, \ldots, n$, denoting the order statistics of

$$T^{(n,u)}_i := \min\left(\frac{1+u}{\hat{U}_i}, \frac{1}{\hat{V}_i} \right).$$

Define $\hat{c}_y(1,1)$ analogously to $\hat{c}_x(1,1)$ with the roles of $\hat{U}_i$ and $\hat{V}_i$ interchanged. If the conditions of Theorem 3.1.2 hold then

$$\hat{l} \rightarrow l$$

in probability. If, in addition, *U and V are asymptotically dependent*, then

$$\hat{c}_x(1,1) \rightarrow c_x(1,1) \qquad \text{and} \qquad \hat{c}_y(1,1) \rightarrow c_y(1,1)$$

in probability. Moreover, the estimator

$$\hat{\sigma}^2_{\hat{\eta}_n} := \begin{cases} \hat{\sigma}^2_{\hat{\eta}^H_n} := \left(\hat{\eta}^H_n\right)^2 \left(1 - \hat{l}\right) \left(1 - 2\hat{l}\hat{c}_x(1,1)\hat{c}_y(1,1)\right) & \text{if } \hat{\eta}_n = \hat{\eta}^H_n, \\ \\ \hat{\sigma}^2_{\hat{\eta}^{ML}_n} := \left(1 + \hat{\eta}^{ML}_n\right)^2 \left(1 - \hat{l}\right) \left(1 - 2\hat{l}\hat{c}_x(1,1)\hat{c}_y(1,1)\right) & \text{if } \hat{\eta}_n = \hat{\eta}^{ML}_n \end{cases}$$

is consistent for $\sigma^2_{\hat{\eta}_n}$ for all $\eta \in (0, 1]$. △

3.1.5 Remarks.

(i) The "slight correction" is that U and V must be *asymptotically dependent* to prove consistency of $\hat{c}_x(1,1)$ and $\hat{c}_y(1,1)$. We show in Section 3.4.1 that $\eta = 1$ is not sufficient in the reasoning of the proof of Theorem 2.2 of Draisma et al. (2004).

(ii) Note that if c is (totally) differentiable (which is satisfied if c_x and c_y are continuous), differentiating $c(tx, ty) = t^{1/\eta} c(x, y)$ with respect to t yields

$$x c_x(tx, ty) + y c_y(tx, ty) = t^{1/\eta - 1} c(x, y)/\eta. \tag{3.5}$$

In particular, $c_x(1,1) + c_y(1,1) = 1/\eta$, so that $1 - \hat{c}_x(1,1)$ is consistent for $c_y(1,1)$ if $l > 0$, but in general $\hat{c}_x(1,1) + \hat{c}_y(1,1) \neq 1$.

△

From the Theorems 3.1.2 and 3.1.4 we conclude that

$$\left[\hat{\eta}_n - m^{-1/2} \hat{\sigma}_{\hat{\eta}_n} \Phi^{\leftarrow}(1-\alpha/2), \hat{\eta}_n + m^{-1/2} \hat{\sigma}_{\hat{\eta}_n} \Phi^{\leftarrow}(1-\alpha/2)\right], \tag{3.6}$$

where $\Phi^{\leftarrow}$ is the quantile function of a standard normal distribution, is a two-sided confidence interval for η with asymptotic confidence level $1-\alpha$. An analogous one-sided statistical test rejects the null hypothesis $\eta = 1$, if

$$\frac{\sqrt{m}\,(1 - \hat{\eta}_n)}{\hat{\sigma}_{\hat{\eta}_n}} > \Phi^{\leftarrow}(1-\alpha). \tag{3.7}$$

Asymptotic Representation of $\hat{\eta}_n$

Denote by Q_n the tail empirical quantile function pertaining to $T_i^{(n)}$, $i = 1, \ldots, n$, that is,

$$Q_n(t) := T^{(n)}_{n - \lfloor mt \rfloor : n}, \qquad 0 < t \leq 1, \tag{3.8}$$

where we denote by $\lfloor a \rfloor$ the largest integer less than or equal to $a \in \mathbb{R}$.

Further, we define a Gaussian process which possesses a somewhat different (covariance) structure in the case of asymptotic independence than in the case of asymptotic dependence of U and V. As in Section 6 of Draisma et al. (2004), let W_1 and W_2 be Gaussian processes with mean zero and covariance structure given by

$$E\big[W_1(x_1, y_1) W_1(x_2, y_2)\big] = c(x_1 \wedge x_2, y_1 \wedge y_2) \tag{3.9}$$

and

$$\begin{aligned} E\big[W_2(x_1, y_1) W_2(x_2, y_2)\big] &= x_1 \wedge x_2 + y_1 \wedge y_2 - lc(x_1, y_1) \\ &\quad - lc(x_2, y_2) + lc(x_1 \vee x_2, y_1 \vee y_2), \end{aligned} \tag{3.10}$$

respectively. Define the Gaussian process W by

$$W = W_1 \tag{3.11}$$

if $l = 0$ (i.e. iff U and V are asymptotically independent) and by

$$\begin{aligned} W(x,y) = {} & \frac{1}{\sqrt{l}}\big[W_2(x,0) + W_2(0,y) - W_2(x,y)\big] \\ & -\sqrt{l}c_x(x,y)W_2(x,0) - \sqrt{l}c_y(x,y)W_2(0,y), \end{aligned} \tag{3.12}$$

if $l > 0$ (i.e. iff U and V are asymptotically dependent). This definition of W suggests that there are always two parts to prove a result in which W is involved, where the part for asymptotically dependent U and V might be substantially more complicated. In the proofs of the main results of Section 3.2 and Chapter 4 we see that this is true. In the proof of the following Lemma, however, we can refer to an existing result to overcome this obstacle. As usual, we call a process $\tilde{Z}$ a *version* of Z, if $\tilde{Z}$ and Z have the same finite dimensional distributions.

3.1.6 Lemma. Assume that the conditions of Theorem 3.1.2 hold. Then there are versions of $\hat{\eta}_n$ and W such that

$$\hat{\eta}_n = \eta + m^{-1/2} \int_0^1 \eta t^{-(\eta+1)} W(t^\eta, t^\eta)\nu_\eta(dt) + o_P(m^{-1/2}), \tag{3.13}$$

where the measure ν_η is defined by

$$\nu_\eta(dt) :=$$

$$\begin{cases} \nu_\eta^H(dt) \quad := t^\eta dt - \varepsilon_1(dt) & \text{if } \hat{\eta}_n = \hat{\eta}_n^H, \\ \nu_\eta^{ML}(dt) := \left(\dfrac{\eta+1}{\eta}\right)^2 \left(t^\eta - (2\eta+1)t^{2\eta}\right) dt + \dfrac{\eta+1}{\eta}\, \varepsilon_1(dt) & \text{if } \hat{\eta}_n = \hat{\eta}_n^{ML} \end{cases} \tag{3.14}$$

with ε_1 denoting the Dirac measure concentrated in 1. △

3.1.7 Corollary. Under the conditions of Lemma 3.1.6,

$$\frac{1}{\hat{\eta}_n} = \frac{1}{\eta} - \frac{m^{-1/2}}{\eta} \int_0^1 t^{-(\eta+1)} W(t^\eta, t^\eta) \nu_\eta(dt) + o_P(m^{-1/2}). \tag{3.15}$$

△

3.1.8 Remarks.

(i) Observe (from the proof of Lemma 3.1.6 and Corollary 3.1.7) that the existence of a suitable version of $\hat{\eta}_n$ follows from the existence of a suitable version of Q_n.

(ii) Further, note the important role of Lemma 6.2 of Draisma et al. (2004) in the proof of Lemma 3.1.6 and Corollary 3.1.7. First, the versions of Q_n and W are exactly those mentioned in Lemma 6.2 of Draisma et al. (2004). Second, Lemma 6.2 of Draisma et al. (2004) is the abovementioned result that holds for both cases, asymptotic independence and asymptotic dependence of U and V.

△

3.2 Estimation of the Limiting Function c

From (2.9) for some sufficiently small r,

$$c(x,y) \approx \frac{P\{U < rx, V < ry\}}{P\{U < r, V < r\}} \approx \frac{\sum_{i=1}^{n} \mathbb{1}\{U_i < rx, V_i < ry\}}{\sum_{i=1}^{n} \mathbb{1}\{U_i < r, V_i < r\}}.$$

Similar as when estimating η, we replace $U_i = 1 - F_1(X_i)$ and $V_i = 1 - F_2(Y_i)$ with the rank standardized approximations $\hat{U}_i = 1 - R_i^X/(n+1)$ and $\hat{V}_i =$

$1 - R_i^Y/(n+1)$ when estimating c. Hence, define

$$\check{c}_n(x,y) := \frac{\sum_{i=1}^{n} \mathbb{1}\Big\{\hat{U}_i < rx, \hat{V}_i < ry\Big\}}{\sum_{i=1}^{n} \mathbb{1}\Big\{\hat{U}_i < r, \hat{V}_i < r\Big\}}$$

for $x, y > 0$ and $r > 1/T_{n:n}^{(n)} = \min_i \big(1/T_i^{(n)}\big) = \min_i \big(\max\big(\hat{U}_i, \hat{V}_i\big)\big)$. In this section we establish the convergence of the suitably standardized process $\check{c}_n - c$. Section 3.2.1 contains the according theorem, a corollary that establishes the asymptotic normality of $\check{c}_n(x,y)$ and a result on estimators required to estimate the asymptotic variance of $\check{c}_n(x,y)$ if U and V are asymptotically dependent. In Section 3.2.2, we see that this result remains true if we choose the random variable $1/T_{n-m:n}^{(n)}$ for r. This is a natural choice, since then, similar as when estimating η by $\hat{\eta}_n^H$ and $\hat{\eta}_n^{ML}$, respectively, the m largest order statistics of T_i, $i = 1, \ldots, n$, are used to estimate c. Further, this choice proves useful and natural for estimating p (Section 3.3) and when validating the scaling law with the method proposed in Chapter 4.

3.2.1 Uniform and Pointwise Limits

Recall the definitions (2.11) and (3.9) – (3.12) of l and W, respectively. The following theorem establishes the convergence of the suitably standardized process $\check{c}_n - c$ to a centered Gaussian process in $D^*\big([0,\infty)^2\big)$, which is a modification of the bivariate Skorohod space $D\big([0,\infty)^2\big)$, both are defined in Appendix A.

3.2.1 Theorem. Assume that Condition 2.2.1 holds with a function c that has first order partial derivatives c_x and c_y. Suppose that $r \to 0$ such that $\tilde{m} := nq(r)$ satisfies $\sqrt{\tilde{m}}\, q_1(r) \to 0$ as $n \to \infty$. Then,

$$\sqrt{\tilde{m}}\big(\check{c}_n - c\big) \to W - cW(1,1)$$

weakly in $D^*\big([0,\infty)^2\big)$. $\triangle$

3.2.2 Remark. Observe that since $r = q^{\leftarrow}(\tilde{m}/n)$, the condition $\sqrt{\tilde{m}}\, q_1(r) \to 0$ as $n \to \infty$ is an analog to (3.4). $\triangle$

It follows that $\sqrt{\tilde{m}}\big(\check{c}_n(x, y) - c(x, y)\big)$ is an asymptotically centered normal random variable for all $x, y > 0$.

3.2.3 Corollary. Under the conditions of Theorem 3.2.1,

$$\sqrt{\tilde{m}}\big(\check{c}_n(x, y) - c(x, y)\big) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2_{x,y}\right)$$

for all $x, y \in (0, 1]$, where

$$\begin{aligned}
\sigma^2_{x,y} &= c(x, y) - 2lc_x(x, y)c(x, y) - 2lc_y(x, y)c(x, y) \\
&\quad + lxc^2_x(x, y) + lyc^2_y(x, y) + 2l^2 c_x(x, y)c_y(x, y)c(x, y) \\
&\quad + 2c(x, y)\Big(lc_x(x, y)\big[c(x, 1) - xc_x(1, 1) - lc_y c(x, 1)\big] \\
&\qquad\qquad - c(x, y)(1 - l) + lc_y(x, y)\big[c(1, y) - yc_y(1, 1) - lc_x c(1, y)\big]\Big) \\
&\quad + c^2(x, y)\Big(1 - 2l\big(c_x(1, 1) + c_y(1, 1)\big) + l\big(c_x(1, 1) + c_y(1, 1)\big)^2 \\
&\qquad\qquad - 2lc_x(1, 1)c_y(1, 1)(1 - l)\Big).
\end{aligned} \tag{3.16}$$

$\triangle$

3.2.4 Remarks.

(i) Observe that

$$\sigma^2_{x,y} = c(x, y)\big(1 - c(x, y)\big) \tag{3.17}$$

if $l = 0$, i.e. iff U and V are asymptotically independent. Further, if $c_x(1, 1) + c_y(1, 1) = 1$ when $l > 0$ (cf. Remark 3.1.5 (ii)), then the last two lines of (3.16) simplify to $c^2(x, y)\big(1 - 2lc_x(1, 1)c_y(1, 1)\big)(1 - l)$.

(ii) For arbitrary $x, y > 0$, Corollary 3.2.3 holds with a more complicated asymptotic variance, see (3.49) and (3.51).

△

From Corollary 3.2.3 and (3.17) we conclude that if $l = 0$,

$$\begin{aligned}\Big[\ &\check{c}_n(x,y) - m^{-1/2}\sqrt{\check{c}_n(x,y)\big(1-\check{c}_n(x,y)\big)}\ \Phi^{\leftarrow}(1-\alpha/2),\\ &\check{c}_n(x,y) + m^{-1/2}\sqrt{\check{c}_n(x,y)\big(1-\check{c}_n(x,y)\big)}\ \Phi^{\leftarrow}(1-\alpha/2)\Big]\end{aligned} \tag{3.18}$$

is a two-sided confidence interval for $c(x, y)$ with asymptotic confidence level $1 - \alpha$.

In order to construct confidence intervals for c if $l > 0$, we need consistent estimators for $c_x(x, y)$ and $c_y(x, y)$, respectively. Recall the homogeneity (2.14) of c and observe that

$$c_x(tx,ty) = \lim_{\varepsilon\downarrow 0}\frac{c(tx+t\varepsilon,ty)-c(tx,ty)}{t\varepsilon} = t^{1/\eta-1}c_x(x,y) \tag{3.19}$$

to see that the derivatives c_x and c_y are homogeneous of order $1/\eta - 1$. Hence, it is sufficient to obtain consistent estimators for $c_x(x, 1)$ and $c_y(x, 1)$. (Recall further that $\eta = 1$ if U and V are asymptotically dependent, so that c is homogeneous of order 1 and c_x and c_y are homogeneous of order 0.) In the following Theorem we present such estimators which are readily obtained by a generalization of $\hat{c}_x(1, 1)$ and $\hat{c}_y(1, 1)$, respectively, which were established in Theorem 3.1.4. Recall the definitions of $\hat{l}$, $\hat{k}$ and $T_{i:n}^{(n,u)}$ of Theorem 3.1.4.

3.2.5 Theorem. Define

$$\hat{c}_x(x,1) := \frac{\hat{k}^{5/4}}{n}\left(T_{n-m:n}^{(n,x-1+\hat{k}^{-1/4})} - T_{n-m:n}^{(n,x-1)}\right)$$

and $\hat{c}_y(x, 1)$ analogously with the roles of $\hat{U}_i$ and $\hat{V}_i$ interchanged. If U and V are asymptotically dependent and the conditions of Theorem 3.2.1 hold, then

$$\hat{c}_x(x,1) \to c_x(x,1) \qquad \text{and} \qquad \hat{c}_y(x,1) \to c_y(x,1)$$

in probability. Moreover, let

$$\hat{c}_x(x,y) := y^{1/\hat{\eta}_n - 1}\hat{c}_x(x/y,1), \qquad \hat{c}_y(x,y) := y^{1/\hat{\eta}_n - 1}\hat{c}_y(x/y,1),$$
$$\hat{c}_x := \hat{c}_x(1,1), \qquad \hat{c}_y := \hat{c}_y(1,1).$$

Further, let

$$\hat{\sigma}^2_{x,y} := \check{c}_n(x,y)(1-\check{c}_n(x,y))$$

if U and V are asymptotically independent and

$$\begin{aligned}\hat{\sigma}^2_{x,y} := {} & \check{c}_n(x,y) - 2\hat{l}\hat{c}_x(x,y)\check{c}_n(x,y) - 2\hat{l}\hat{c}_y(x,y)\check{c}_n(x,y) \\ & + \hat{l}x\hat{c}_x^2(x,y) + \hat{l}y\hat{c}_y^2(x,y) + 2\hat{l}^2\hat{c}_x(x,y)\hat{c}_y(x,y)\check{c}_n(x,y) \\ & -2\check{c}_n(x,y)\Big[\check{c}_n(x,y)(1-\hat{l}) - \hat{l}\hat{c}_x(x,y)\big[\check{c}_n(x,1) - x\hat{c}_x - \hat{l}\hat{c}_y\check{c}_n(x,1)\big] \\ & \qquad -\hat{l}\hat{c}_y(x,y)\big[\check{c}_n(1,y) - y\hat{c}_y - \hat{l}\hat{c}_x\check{c}_n(1,y)\big]\Big] \\ & + \check{c}_n^2(x,y)(1-2\hat{l}\hat{c}_x\hat{c}_y)(1-\hat{l})\end{aligned}$$

if U and V are asymptotically dependent. Then $\hat{\sigma}^2_{x,y}$ is a consistent estimator for $\sigma^2_{x,y}$. △

3.2.6 Remark. We see in the following Section 3.2.2 that Theorem 3.2.1 and Corollary 3.2.3 still hold when $\check{c}_n$ is substituted with $\hat{c}_n$, defined in (3.21). Thus, Theorem 3.2.5 also holds when $\check{c}_n$ is replaced with $\hat{c}_n$. △

From Corollary 3.2.3 and Theorem 3.2.5 we conclude that

$$\Big[\check{c}_n(x,y) - m^{-1/2}\hat{\sigma}_{x,y}\Phi^{\leftarrow}(1-\alpha/2),\ \check{c}_n(x,y) + m^{-1/2}\hat{\sigma}_{x,y}\Phi^{\leftarrow}(1-\alpha/2)\Big]. \tag{3.20}$$

is a two-sided confidence interval for $c(x,y)$ with asymptotic confidence level $1-\alpha$.

3.2.2 Choice of the Tail-Observations Used

To estimate c, we have to choose an appropriate r. We now motivate such a choice by reconsidering the estimators of η presented in Section 3.1. In particular, this choice will prove useful when validating the scaling law with the method proposed in Chapter 4. Recall that both the Hill estimator $\hat{\eta}_n^H$ and the MLE $\hat{\eta}_n^{ML}$ are based on the $m+1$ largest order statistics $T^{(n)}_{n-j+1:n}$, $j = 1, \ldots, m+1$, of the $T_i^{(n)}$. This means that we use those $(\hat{U}_i, \hat{V}_i)$ for the estimation of η which lie within the square $\big(0, 1/T^{(n)}_{n-m:n}\big]^2$. Hence, from this point of view, a natural choice is the random variable $r = 1/T^{(n)}_{n-m:n}$. Therefore, we define

$$\hat{c}_n(x,y) := \frac{\sum_{i=1}^{n} \mathbb{1}\Big\{\hat{U}_i < x/T^{(n)}_{n-m:n}, \hat{V}_i < y/T^{(n)}_{n-m:n}\Big\}}{\sum_{i=1}^{n} \mathbb{1}\Big\{\hat{U}_i < 1/T^{(n)}_{n-m:n}, \hat{V}_i < 1/T^{(n)}_{n-m:n}\Big\}} \tag{3.21}$$

for all $x, y > 0$. Theorem 3.2.1 suggests that one can prove that $\sqrt{\tilde{m}}\big(\hat{c}_n - c\big)$ converges to a centered Gaussian process. We show that this also holds when $\tilde{m}$ is replaced with m. First, however, we give some comments on the relation of the particular choice $\tilde{m} = nq\big(1/T^{(n)}_{n-m:n}\big)$ and m, which reveal that the similarity of the notations of these quantities is not by accident. By the previously introduced choice of r,

$$\tilde{m} = nq\big(1/T^{(n)}_{n-m:n}\big) = nP\left\{U < r, V < r\right\}\Big|_{r=1/T^{(n)}_{n-m:n}}$$

also becomes a random variable. If we assume that there is exactly one i_0 with[1] $T^{(n)}_{i_0} = T^{(n)}_{n-m:n}$, we have for the non-random value m the simple equality

$$m = \sum_{i=1}^{n} \mathbb{1}\Big\{\hat{U}_i < 1/T^{(n)}_{n-m:n}, \hat{V}_i < 1/T^{(n)}_{n-m:n}\Big\}.$$

[1] Under the assumption that F_1 and F_2 are tail continuous, the probability that there are no ties between $T^{(n)}_{i:n}$, $i = n-m, \ldots, n$, (and hence exactly one such i_0) converges to 1; see also the comments made in the proof of Theorem 3.2.1.

and thus

$$\frac{\tilde{m}}{m} = \left. \frac{nP\{U < r, V < r\}}{\sum_{i=1}^{n} 1\!\!1\{\hat{U}_i < r, \hat{V}_i < r\}} \right|_{r=1/T^{(n)}_{n-m:n}} .$$

After these considerations, it is not surprising that a by-product of the proof of the following Theorem 3.2.7 is that $\tilde{m}/m \to 1$ in probability as $n \to \infty$.

Before we state the announced main result of this section, recall Remark 3.1.5 (ii), where we obtained that

$$xc_x(x,y) + yc_y(x,y) = c(x,y)/\eta \tag{3.22}$$

if c is (totally) differentiable.

3.2.7 Theorem. Assume that Condition 2.2.1 holds with a function c that has first order partial derivatives c_x and c_y which satisfy (3.22). Suppose, as in Theorem 3.1.2, that $\sqrt{m}\, q_1\big(q^{\leftarrow}(m/n)\big) \to 0$ as $n \to \infty$. Then,

$$\sqrt{m}\big(\hat{c}_n - c\big) \to W - cW(1,1)$$

weakly in $D^*\big([0,\infty)^2\big)$. $\triangle$

Of course, the analog to Corollary 3.2.3 holds.

3.2.8 Corollary. Under the conditions of Theorem 3.2.7,

$$\sqrt{m}\big(\hat{c}_n(x,y) - c(x,y)\big) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2_{x,y}\right)$$

for all $x, y \in (0,1]$ with the variance $\sigma^2_{x,y}$ defined by (3.16). $\triangle$

3.2.9 Remark. Remark 3.2.4 (ii) applies analogously. $\triangle$

The consistency of $\hat{c}_n(x,y)$ is important for the construction of the graphical tool for model validation in Chapter 4 and is used in the applications of Chapter 5.

Finally, we conclude from Theorem 3.2.7 that when we substitute $\check{c}$ with $\hat{c}$ in (3.18) and (3.20), the resulting intervals are still two-sided confidence intervals for $c(x,y)$ with asymptotic confidence level $1-\alpha$.

3.3 Estimation of the Probability of Jointly Large Claims p

By (2.21), namely

$$p \approx r^{-1/\eta} \cdot \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{U_i < r(1 - F_1(u_1)),\ V_i < r(1 - F_2(u_2))\} \tag{3.23}$$

for all r sufficiently small to satisfy the approximate scaling law and sufficiently large to obtain a reasonable number for the sum on the right hand side of (3.23), we have started to motivate an estimator for p. If we accept (3.23) as the starting point for constructing estimators for p, we have to substitute U_i and V_i as well as F_1 and F_2 by suitable estimators. Recall from Sections 3.1 and 3.2, when estimating η and c, we have replaced U_i and V_i by their empirical counterparts $\hat{U}_i = 1-R_i^X/(n+1)$ and $\hat{V}_i = 1-R_i^Y/(n+1)$. In Section 1.1.2, we briefly discussed the generalized Pareto model in the univariate EVT. As there, we assume that F_1 and F_2 belong to the domain of attraction of a univariate extreme value distribution. Denote by $\hat{F}_{1,n}$ and $\hat{F}_{2,n}$ the generalized Pareto estimators (1.10) of the tails of the marginal d.f.s F_1 and F_2. Then, continuing (3.23),

$$p \approx \frac{r^{-1/\hat{\eta}_n}}{n} \sum_{i=1}^{n} \mathbb{1}\Big\{\hat{U}_i < r\big(1 - \hat{F}_{1,n}(u_1)\big),\ \hat{V}_i < r\big(1 - \hat{F}_{2,n}(u_2)\big)\Big\} =: \hat{p}_n \tag{3.24}$$

which motivates $\hat{p}_n$ as an estimator for p. The procedure to obtain $\hat{p}_n$ may thus be summarized as follows. First, estimate the distribution of the marginals by an appropriate choice of j_i, $i = 1, 2$, where j_i denotes the number of order statistics used for the estimators $\hat{F}_{i,n}(u_i)$, $i = 1, 2$, cf. Sections 1.1.2 and 1.1.4. Second, calculate $(\hat{U}_i, \hat{V}_i)$ and estimate η by an appropriate choice of m, cf. Section 3.1. Third, choose an appropriate r and calculate $\hat{p}_n$. When estimating $1 - F_1(u_1)$ and η, estimation errors occur. In Section 3.3.1, we establish the asymptotic normality of $\hat{p}_n$ if the one or the other of these estimation errors dominates. The appropriate choice of r is discussed in Section 3.3.2, where we devise a graphical tool to select r. First, however, we look at a generalization of $\hat{p}_n$ and an estimator of p which is based on a different standardization of the observations X_i and Y_i when approximating U_i and V_i in (3.23).

The concept of the construction of $\hat{p}_n$ is readily extended to estimators for the probability of more general jointly large events like $P\{(X, Y) \in C\}$

for tail events C that satisfy[2] $(x, y) \in C \Rightarrow [x, \infty] \times [y, \infty] \subset C$. Define the transformed failure set $D := \big\{\big(1 - F_1(x), 1 - F_2(y)\big) \mid (x, y) \in C\big\}$ which is estimated by $\hat{D}_n := \big\{\big(1 - \hat{F}_{1,n}(x), 1 - \hat{F}_{2,n}(y)\big) \mid (x, y) \in C\big\}$. Then, one may motivate an estimator for $P\{(X, Y) \in C\}$ with tail event C by

$$P\{(X, Y) \in C\} = P\{(U, V) \in D\} \approx r^{-1/\eta} P\{(U, V) \in rD\}$$

$$\approx \frac{r^{-1/\hat{\eta}_n}}{n} \sum_{i=1}^{n} \mathbb{1}_{\big\{\big(\hat{U}_i, \hat{V}_i\big) \in r\hat{D}_n\big\}}.$$

Besides $\hat{p}_n$, there is another estimator worth considering. Instead of using the approximations $\hat{U}_i$ and $\hat{V}_i$ to U_i and V_i that are based on ranks, one can use estimators $\tilde{U}_i := 1 - \hat{F}_{1,n}(X_i)$ and $\tilde{V}_i := 1 - \hat{F}_{2,n}(Y_i)$, which are based on the GPD approximations of the marginal tails. The resulting estimator of p is

$$\tilde{p}_n := \frac{r^{-1/\hat{\eta}_n}}{n} \sum_{i=1}^{n} \mathbb{1}_{\big\{\tilde{U}_i < r\big(1 - \hat{F}_{1,n}(u_1)\big),\ \tilde{V}_i < r\big(1 - \hat{F}_{2,n}(u_2)\big)\big\}}.$$

For the generalization of $\tilde{p}_n$ to an estimator of $P\{(X, Y) \in C\}$ with C as defined above, consistency is proved in Section 3 of Draisma et al. (2004).

3.3.1 Asymptotic Normality and Confidence Intervals

Under weak assumptions on F_1 and F_2, we prove asymptotic normality of $\hat{p}_n$ and construct confidence intervals for p. Denote by $\sigma^2_{\hat{\gamma}_{i,n}}$ the asymptotic variances of estimators $\hat{\gamma}_{i,n}$ of the extreme value indices γ_i, $i = 1, 2$, of X and Y, respectively.

3.3.1 Theorem. Assume that the conditions of Theorem 3.1.2 hold with a function q_1 that is regularly varying with an index $\tau > 0$ and that (1.21) when x_n, j and $\sigma^2_{\hat{\gamma}_n}$ are replaced with $u_i = u_{i,n}$, j_i and $\sigma^2_{\hat{\gamma}_{i,n}}$, $i = 1, 2$, holds. Assume further that as $n \to \infty$, the asymptotic properties

[2] This property is rather a technical condition than a restriction. It merely means that we only allow events C, which possess the property that if a claim (x, y) pertains to C then a claim with both components larger must pertain to C, too.

(i) $u_i = u_{i,n} \to F_i^{\leftarrow}(1), \qquad i = 1, 2,$

(ii) $\dfrac{1 - F_2(u_2)}{1 - F_1(u_1)} \to \lambda \in (0, \infty),$

(iii) $r = r_n \to \infty,$

(iv) $nq\big(r(1 - F_1(u_1))\big) \to \infty$ and

(v) $q_1\big(r(1 - F_1(u_1))\big) = o\Big(\big(n\; q(r(1 - F_1(u_1)))\big)^{-1/2}\Big)$

are satisfied. Moreover, suppose that either

$$m \log^{-2} r = O\big(n\; q(r(1 - F_1(u_1)))\big) \tag{3.25}$$

or

$$j_i \log^{-2} \left(\frac{j_i}{n(1 - F_i(u_i))}\right) = O\Big(\big(n\; q(r(1 - F_i(u_i)))\big)^{1/2}\Big), \quad i = 1 \textbf{ or } 2, \tag{3.26}$$

holds. Then,

$$\frac{\eta^2 m^{1/2}}{\log r} \log \frac{\hat{p}_n}{p} \sim m^{1/2}(\hat{\eta}_n - \eta) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2_{\hat{\eta}_n}\right) \tag{3.27}$$

if

$$j_i^{-1} \log^2 \left(\frac{j_i}{n(1 - F_i(u_i))}\right) = O(m^{-1} \log^2 r), \qquad i = 1, 2, \tag{3.28}$$

whereas

$$\begin{aligned} &\frac{c(1,\lambda)}{c_x(1,\lambda)} \cdot \frac{\gamma_1 j_1^{1/2}}{\log \frac{j_1}{n(1-F_1(u_1))}} \log \frac{\hat{p}_n}{p} \\ &\qquad \sim \frac{\gamma_1 j_1^{1/2}}{\log \frac{j_1}{n(1-F_1(u_1))}} \left(\frac{1 - \hat{F}_{1,n}(u_1)}{1 - F_1(u_1)} - 1\right) \xrightarrow{D} \mathcal{N}\big(0, \sigma^2_{\hat{\gamma}_{1,n}}\big) \end{aligned} \tag{3.29}$$

if

$$j_1 = o(j_2) \qquad \text{and} \qquad j_1 \log^{-2} \left(\frac{j_1}{n(1 - F_1(u_1))}\right) = o(m \log^{-2} r) \tag{3.30}$$

and

$$\lambda \frac{c(1,\lambda)}{c_y(1,\lambda)} \cdot \frac{\gamma_2 j_2^{1/2}}{\log \frac{j_2}{n(1-F_2(u_2))}} \log \frac{\hat{p}_n}{p}$$

$$\sim \frac{\gamma_2 j_2^{1/2}}{\log \frac{j_2}{n(1-F_2(u_2))}} \left(\frac{1-\hat{F}_{2,n}(u_2)}{1-F_2(u_2)} - 1 \right) \xrightarrow{D} \mathcal{N}\big(0, \sigma^2_{\hat{\gamma}_{2,n}}\big) \qquad (3.31)$$

if

$$j_2 = o(j_1) \qquad \text{and} \qquad j_2 \log^{-2}\left(\frac{j_2}{n(1-F_2(u_2))}\right) = o(m \log^{-2} r). \qquad (3.32)$$

$\triangle$

We shed light on this seemingly messy bunch of conditions in the following Remarks 3.3.2, where we start with noting that the assumptions (i) – (v) are natural. Further, we comment on the conditions (3.25), (3.26), (3.28), (3.30) and (3.32) which are concerned with the interplay of the estimation (and to some extent approximation) errors occurring when using $\hat{p}_n$ to estimate p.

3.3.2 Remarks.

(i) Assumption (i) ensures that $1 - F_i(u_i) \to 0$ and $p = p_n \to 0$, which is necessary (and reasonable) when applying asymptotic methods to extrapolate outside the range of available data. In their Section 4.3, p. 134, de Haan and Ferreira (2006) briefly discuss this issue in a univariate setting (where $n(1 - F_i(u_i)) = o(j_i)$, $i = 1, 2$, is the appropriate condition, see also Section 1.1.4, in particular (1.20)).

(ii) While assumption (i) relates $u_{i,n}$ and n, assumption (ii) claims that the speed of convergence of $u_{1,n}$ and $u_{2,n}$ to infinity may not be too different with respect to the pertaining marginal exceedance probabilities. In other words, as Draisma et al. (2004) have formulated in a more general setting, assumption (ii) "essentially means that the convergence of the failure set in the x- and the y-direction is balanced".

(iii) In view of the motivation (2.21) for $\hat{p}_n$, the convergence $r = r_n \to \infty$ is a natural consequence. Loosely speaking, the larger $u_{i,n}$, the fewer observations of a jointly large claim we have and the larger r must be to obtain a reasonable number of observations used for $\hat{p}_n$. (See also the following Section 3.3.2 and the applications in Chapter 5.) Therefore, the convergence $r \to \infty$ is a consequence of $u_{i,n} \to \infty$ and thus a reasonable and necessary condition in an asymptotic setting.

(iv) Recall that $q(t) = P\{U < t, V < t\}$ to see that assumption (iv) is also a natural consequence of the motivation (2.21) for $\hat{p}_n$. Roughly speaking, it means that the number of observations used to estimate $P\{U < r(1-F_1(u_1)), V < r(1-F_2(u_2))\}$ converges to ∞, which is again reasonable and necessary in an asymptotic setting.

(v) Assumption (v) requires the sequence r not to converge too fast to ∞ in relation to the marginal exceedance probability (and thus to u_1). In particular, together with assumption (iv), assumption (v) implies $r(1 - F_1(u_1)) \to 0$.

(vi) Theorem 3.3.1 further requires that (3.25) or (3.26) holds. In (3.25) (and analogously in (3.26)), the intermediate sequence of the number m of order statistics on which $\hat{\eta}_n$ is based, may be at most of the order of the intermediate sequence $n\ q(r(1 - F_1(u_1))$. This sequence may be considered as pertaining to the error that is specific to $\hat{p}_n$, i.e. the error that is not caused by the estimation of η or the marginal distributions. For the role of this sequence, see also the proof of Theorem 3.3.1.

(vii) Assumptions (3.28) and (3.30) then determine (the variance of) the limiting distribution. If the error when estimating η dominates the errors when estimating the marginal distributions, i.e. if (3.28) holds, then this is reflected in the according asymptotic variance of the limiting distribution. The analogous (converse) interpretation is valid for (3.30), where $j_1 = o(j_2)$ of (3.30) requires that when estimating the marginal distributions, the error when estimating $F_1(u_1)$ dominates the error when estimating $F_2(u_2)$. In $j_2 = o(j_1)$ of (3.32), it is the other way around.

$\triangle$

Depending on the choice of the estimator for η and whether U and V are asymptotically (in)dependent, we obtain more or less complicated looking confidence intervals. In general, it follows from Theorem 3.3.1 that if (3.27) applies,

$$\left[\hat{p}_n \exp\left(- \frac{\log r}{m^{1/2}} \cdot \frac{\sigma_{\hat{\eta}_n}}{\hat{\eta}_n^2}\ \Phi^{\leftarrow}(1-\alpha/2)\right), \hat{p}_n \exp\left(\frac{\log r}{m^{1/2}} \cdot \frac{\sigma_{\hat{\eta}_n}}{\hat{\eta}_n^2}\ \Phi^{\leftarrow}(1-\alpha/2)\right)\right], \tag{3.33}$$

and if (3.29) applies,

$$\Big[\hat{p}_n \exp\Big(-\frac{\hat{c}_x(1,\hat{\lambda}_n)}{\hat{c}_n(1,\hat{\lambda}_n)} \cdot \frac{\log \frac{j_1}{n(1-\hat{F}_{1,n}(u_1))}}{j_1^{1/2}} \cdot \frac{\sigma_{\hat{\gamma}_{1,n}}}{\hat{\gamma}_{1,n}}\, \Phi^{\leftarrow}(1-\alpha/2)\Big),$$

$$\hat{p}_n \exp\Big(\frac{\hat{c}_x(1,\hat{\lambda}_n)}{\hat{c}_n(1,\hat{\lambda}_n)} \cdot \frac{\log \frac{j_1}{n(1-\hat{F}_{1,n}(u_1))}}{j_1^{1/2}} \cdot \frac{\sigma_{\hat{\gamma}_{1,n}}}{\hat{\gamma}_{1,n}}\, \Phi^{\leftarrow}(1-\alpha/2)\Big)\Big], \tag{3.34}$$

where $\hat{\lambda}_n := (1-\hat{F}_{2,n}(u_2))/(1-\hat{F}_{1,n}(u_1))$, is a confidence interval for p with asymptotic confidence level $1-\alpha$. (An analogous confidence interval is valid if (3.31) applies.) Recall that under the conditions of Theorem 3.3.1, $\sigma_{\hat{\gamma}^H_{1,n}} = \gamma_1$, $\sigma_{\hat{\gamma}^H_{2,n}} = \gamma_2$ and that if U and V are asymptotically independent, $\sigma_{\hat{\eta}^H_n} = \eta$. Hence, in this case, (3.33) and (3.34) read as

$$\Big[\hat{p}_n \exp\Big(-\frac{\log r}{m^{1/2}} \cdot \frac{\Phi^{\leftarrow}(1-\alpha/2)}{\hat{\eta}^H_n}\Big), \hat{p}_n \exp\Big(\frac{\log r}{m^{1/2}} \cdot \frac{\Phi^{\leftarrow}(1-\alpha/2)}{\hat{\eta}^H_n}\Big)\Big] \tag{3.35}$$

and

$$\Big[\hat{p}_n \exp\Big(-\frac{\hat{c}_x(1,\hat{\lambda}_n)}{\hat{c}_n(1,\hat{\lambda}_n)} \cdot \frac{\log \frac{j_1}{n(1-\hat{F}_{1,n}(u_1))}}{j_1^{1/2}} \cdot \Phi^{\leftarrow}(1-\alpha/2)\Big),$$

$$\hat{p}_n \exp\Big(\frac{\hat{c}_x(1,\hat{\lambda}_n)}{\hat{c}_n(1,\hat{\lambda}_n)} \cdot \frac{\log \frac{j_1}{n(1-\hat{F}_{1,n}(u_1))}}{j_1^{1/2}} \cdot \Phi^{\leftarrow}(1-\alpha/2)\Big)\Big]. \tag{3.36}$$

If (3.31) applies,

$$\Big[\hat{p}_n \exp\Big(-\hat{\lambda}_n \frac{\hat{c}_y(1,\hat{\lambda}_n)}{\hat{c}_n(1,\hat{\lambda}_n)} \cdot \frac{\log \frac{j_2}{n(1-\hat{F}_{2,n}(u_2))}}{j_2^{1/2}} \cdot \Phi^{\leftarrow}(1-\alpha/2)\Big),$$

$$\hat{p}_n \exp\Big(\hat{\lambda}_n \frac{\hat{c}_y(1,\hat{\lambda}_n)}{\hat{c}_n(1,\hat{\lambda}_n)} \cdot \frac{\log \frac{j_2}{n(1-\hat{F}_{2,n}(u_2))}}{j_2^{1/2}} \cdot \Phi^{\leftarrow}(1-\alpha/2)\Big)\Big]. \tag{3.37}$$

is a confidence intervals for p with asymptotic confidence level $1-\alpha$ in this case.

3.3.2 Choice of the Tail-Observations Used

We have discussed how to choose suitable parameters j_i for the estimators $\hat{F}_i(u_i)$, $i = 1, 2$, and m for the estimators $\hat{\eta}_n^H$ and $\hat{\eta}_n^{ML}$ to calculate $\hat{p}_n$ and the above confidence intervals. In this section we discuss an appropriate choice of r. We have indicated in Section 2.2.2 how estimators like $\hat{p}_n$ and $\tilde{p}_n$, i.e. based on (3.23), work. In the following, this is described more thoroughly and a graphical tool to choose an appropriate r is devised.

Usually, we are interested in the probability of a jointly large claim $\{U < 1 - F_1(u_1), V < 1 - F_2(u_2)\}$ where u_1 and u_2 are so large that only few or no observations at all are available, i.e. the sum $\sum_{i=1}^n \mathbb{1}\{\hat{U}_i < 1 - \hat{F}_{1,n}(u_1), \hat{V}_i < 1 - \hat{F}_{2,n}(u_2)\}$ is a very small number or even equals 0. Then, r should be chosen such that sufficiently many observations lie within the inflated set $\{\hat{U}_i < r(1 - \hat{F}_{1,n}(u_1)), \hat{V}_i < r(1 - \hat{F}_{2,n}(u_2))\}$ (see Figure 3.2), which is used by $\hat{p}_n$, cf. (3.24). Hence, because of assumption (i) of Theorem 3.3.1, the expected number of

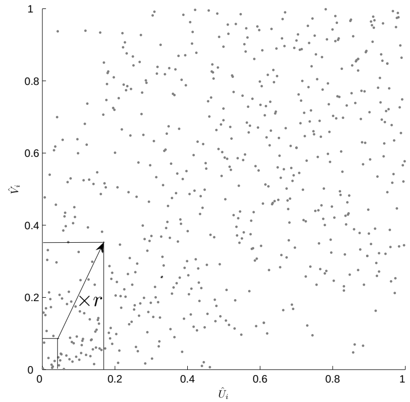


Figure 3.2: The inflation of the set $\{\hat{U}_i < 1 - F_1(u_1), \hat{V}_i < 1 - F_2(u_2)\}$ with factor r. According to the scaling law assumed by the model (cf. Section 2.2.2, in particular Figure 2.1), the probability of the event associated with this set is then estimated by multiplying the empirical probability of the event associated with the inflated set with $r^{-1/\hat{\eta}_n}$.

observations within the inflated set $\{\hat{U}_i < r(1-\hat{F}_{1,n}(u_1)), \hat{V}_i < r(1-\hat{F}_{2,n}(u_2))\}$ must converge to infinity, which in turn is reflected in assumption (iv) of that theorem. As a simplified rule, the larger u_i, the larger r has to be chosen. Then, by means of the scaling law assumed by the model (cf. Section 2.2.2), we scale down the "empirical probability" $n^{-1}\sum_{i=1}^n \mathbb{1}\{\hat{U}_i < r(1-\hat{F}_{1,n}(u_1)), \hat{V}_i < r(1-\hat{F}_{2,n}(u_2))\}$ with factor $r^{-1/\hat{\eta}_n}$. Hence, on the other hand, we must note that r may not be too large, since then the scaling law does not hold. In the following Chapter 4, we establish a model validation tool that helps to decide whether the scaling law holds. This tool checks the presence of the scaling law in an arbitrary (rectangular) region.

We devise the following graphical tool to find an appropriate r. For $\iota = 1, \dots, n$, we choose $r = r(\iota)$ as

$$r(\iota) := \inf\left\{ r > 0 \;\middle|\; \sum_{i=1}^n \mathbb{1}_{\left\{\hat{U}_i < r\left(1-\hat{F}_{1,n}(u_1)\right), \hat{V}_i < r\left(1-\hat{F}_{2,n}(u_2)\right)\right\}} \geq \iota \right\}$$

and plot

$$\left(\iota, \frac{(r(\iota))^{-1/\hat{\eta}_n}}{n} \cdot \iota\right). \tag{3.38}$$

Here, the parameter r is reduced to an auxiliary value: plot (3.38) yields $\hat{p}_n$ depending on ι, the number of observations used for the estimation. For an extreme value statistician, this is a familiar situation. For the according plots when estimating the extreme value index γ (Section 1.1.4) and the coefficient of tail dependence η (Section 3.1), we know that a typical (nice) plot exhibits heavy fluctuations for a small number of observations used (due to large variance), followed by a rather stable region where both variance and bias are moderate, before the bias causes a clear upward or downward trend of the curve. We apply this heuristic to plot (3.38), too. Thus, ι ought to be chosen in the region where the plot is rather stable. This graphical tool is applied in Chapter 5.

3.4 Proofs

3.4.1 Proofs for Section 3.1

Proof of Corollary 3.1.1

The first part of the assertion is simply obtained by

$$1 - F_T(t) = P\{T_i > t\} = P\left\{\min\left(1/U_i, 1/V_i\right) > t\right\}$$

$$= P\left\{1/U_i > t, 1/V_i > t\right\} = P\left\{U < 1/t, V < 1/t\right\} = q\left(1/t\right).$$

Further, the regular variation of q at 0 with index $1/\eta$ means that

$$\begin{aligned}
&\lim_{r\to\infty} \frac{q(t/r)}{q(1/r)} = t^{1/\eta} && \text{for all } t > 0\\
\Leftrightarrow &\lim_{r\to\infty} \frac{q(1/(tr))}{q(1/r)} = t^{-1/\eta} && \text{for all } t > 0\\
\Leftrightarrow &\lim_{r\to\infty} \frac{1 - F_T(tr)}{1 - F_T(r)} = t^{-1/\eta} && \text{for all } t > 0.
\end{aligned}$$

Proof of Theorem 3.1.4

Observe that we only need to show that the reasoning in the proof of Theorem 2.2 of Draisma et al. (2004) requires that U and V are asymptotically dependent. Therefore, we give some details of the important steps of this proof.

We adopt the first lines from the proof of Theorem 2.2 of Draisma et al. (2004), p. 272 until

$$\frac{k}{n} T^{(n,u)}_{n-\lfloor mt \rfloor:n} = \left(\frac{t}{c(1+u,1)}\right)^{-\eta} + O_P(m^{-1/2}). \tag{3.39}$$

Here, as in Draisma et al. (2004), k is defined by $k := \lceil nq^{\leftarrow}(m/n) \rceil$ and for $a \in \mathbb{R}$ we denote by $\lfloor a \rfloor$ the largest integer less than or equal to a and by $\lceil a \rceil$ the smallest integer greater than or equal to a. From

$$\hat{c}_x(1,1) = \frac{\hat{k}^{5/4}}{n} \left(T^{(n,\hat{k}^{-1/4})}_{n-m:n} - T^{(n)}_{n-m:n}\right)$$

follows for $\eta = 1$

$$\hat{c}_x(1,1) = \frac{\hat{k}^{5/4}}{n} \cdot \frac{n}{k} \left(c(1+\hat{k}^{-1/4},1) - c(1,1) + O_P(m^{-1/2}) \right)$$

$$= \frac{\hat{k}}{k} \left(\frac{c(1+\hat{k}^{-1/4},1) - c(1,1)}{\hat{k}^{-1/4}} + O_P(k^{1/4} m^{-1/2}) \right) \qquad (3.40)$$

since $k/\hat{k} \to 1$ in probability. Now note that *in the case of asymptotic dependence*, i.e. $q(t)/t \to l > 0$, we obtain by Theorem 1.5.12 of Bingham et al. (1987) that

$$\frac{q^{\leftarrow}(lt)}{t} \to 1$$

so that

$$\frac{q^{\leftarrow}(t)}{t} \to \frac{1}{l} > 0 \qquad (3.41)$$

as $t \downarrow 0$. Hence, with $m/n \to 0$ as $n \to \infty$ in mind, we can conclude from $k = \lceil n/m \; q^{\leftarrow}(m/n)m \rceil$ that $k/m \to 1/l$, so $k^{1/4}/m^{1/2} \to 0$ and therefore $\hat{c}_x(1,1) \to c_x(1,1)$ in probability. The consistency of $\hat{c}_y(1,1)$ can be proved along the same lines.

Proof of Remark 3.1.5(i)

The condition $\eta = 1$ is not sufficient to prove the consistency of $\hat{c}_x(1,1)$ (and $\hat{c}_y(1,1)$) as the following example reveals. Let $q^{\leftarrow}(t) = t^{\eta}|\log t|$ and $m = \log\log n$. (Note that $q^{\leftarrow}$ is regularly varying at 0 with index η. The existence of a generalized inverse q that varies regularly with index $1/\eta$ follows from Theorem 1.5.12 of Bingham et al. (1987).) Assume $\eta = 1$ and note that $q^{\leftarrow}(t)/t \to \infty$, so that due to (3.41), U and V must be asymptotically independent. Then

$$\begin{aligned} k &\sim n q^{\leftarrow}(m/n) \\ &\sim m \; \big|\log|(\log\log n)/n|\big| = (\log\log n) \; |\log\log\log n - \log n| \\ &\sim (\log\log n) \; \log n, \end{aligned}$$

so that $k^{1/4}/m^{1/2} \to \infty$, so that we cannot conclude consistency of $\hat{c}_x(1,1)$ from (3.40).

Proof of Lemma 3.1.6 and Corollary 3.1.7 for the Hill Estimator $\hat{\eta}^H$

The proof is based on results of Drees (1998b), in particular Example 3.1, p. 103.

We may write

$$\hat{\eta}_n^H = \int_0^1 \log \frac{Q_n(t)}{Q_n(1)}\, dt$$

and define functions $y_n : (0,1] \to \mathbb{R}$ by

$$y_n(t) := \sqrt{m}\left(q^{\leftarrow}(m/n)Q_n(t) - t^{-\eta}\right). \tag{3.42}$$

Further, let the function $y : (0,1] \to \mathbb{R}$ be defined by

$$y(t) := \eta t^{-(\eta+1)} W(t^\eta, t^\eta). \tag{3.43}$$

Hence,

$$\begin{aligned}
\hat{\eta}_n^H - \eta &= \int_0^1 \log \frac{t^{-\eta} + m^{-1/2} y_n(t)}{1 + m^{-1/2} y_n(1)}\, dt - \eta \\
&= \int_0^1 \log t^\eta \frac{t^{-\eta} + m^{-1/2} y_n(t)}{1 + m^{-1/2} y_n(1)}\, dt \\
&= \int_0^1 \log \frac{1 + m^{-1/2} y_n(1) + m^{-1/2} t^\eta y_n(t) - m^{-1/2} y_n(1)}{1 + m^{-1/2} y_n(1)}\, dt \\
&= \int_0^1 \log \left(1 + m^{-1/2} \frac{t^\eta y_n(t) - y_n(1)}{1 + m^{-1/2} y_n(1)}\right) dt.
\end{aligned}$$

In order to be able to proceed with analogous arguments as in Example 3.1 of Drees (1998b), we need the following Lemma. As usual, let $C(0,1]$ and $D(0,1]$ denote the continuous functions on $(0,1]$ and the right continuous functions with left limits existing at each point of $(0,1]$, respectively.

3.4.1 Lemma. For each $\varepsilon > 0$, the functions y_n and y satisfy

$$y_n \in D_{\eta,h} := \Big\{ z : (0,1] \to \mathbb{R} \;\Big|\; \lim_{t \downarrow 0} t^\eta h(t) z(t) = 0,\ \left(t^\eta h(t) z(t)\right)_{t \in (0,1]} \in D(0,1] \Big\},$$

$$y \in C_{\eta,h} := \left\{ z \in D_{\eta,h} \mid z_{|(0,1]} \in C(0,1] \right\}$$

and

$$t^{-\eta} + m^{-1/2} y_n \in \bar{D}_{\eta,h} := \{ z \in D_{\eta,h} \mid z \text{ positive and non-increasing} \},$$

where $h(t) := t^{\varepsilon+1/2}$. △

Proof. Clearly, $y_n(t) = \sqrt{m}\big(q^{\leftarrow}(m/n)Q_n(t) - t^{-\eta}\big) \in D(0,1]$ and, for sufficiently large n, $t^{-\eta} + m^{-1/2} y_n(t) = q^{\leftarrow}(m/n)Q_n(t)$ is positive and non-increasing. By (3.9) – (3.12) and the homogeneity (2.14) of c, we obtain that $\big(W(t^\eta, t^\eta)\big)_{t>0}$ is a Brownian motion for $l = 0$ and can be represented as sum of Brownian motions if $l > 0$. For the latter, note that $\big(W_2(t,0)\big)_{t>0}$ and $\big(W_2(0,t)\big)_{t>0}$ are Brownian motions in t and that this is also true for the process $\big(W_2(t,t)\big)_{t>0}$, since $E[W_2(t_1,t_1)W_2(t_2,t_2)] = (2-l)(t_1 \wedge t_2)$. Hence, $\big(W(t^\eta, t^\eta)\big)_{t>0}$ has almost surely continuous sample paths, so that $y(t) = \eta t^{-(\eta+1)} W(t^\eta, t^\eta) \in C(0,1]$ and thus it only remains to show that $t^\eta h(t) y_n(t)$, $t^\eta h(t) y(t)$ and $t^\eta h(t)\big(t^{-\eta} + m^{-1/2} y_n(t)\big)$ converge to 0 as $t \downarrow 0$. Observe that $Q_n(t) \to T_{n:n}$ as $t \downarrow 0$ so that

$$\lim_{t \downarrow 0} t^\eta h(t) y_n(t) = \lim_{t \downarrow 0} t^{\varepsilon+1/2} \sqrt{m} \, \big(q^{\leftarrow}(m/n)Q_n(t)t^\eta - 1\big) = 0.$$

Further, the law of the iterated logarithm for Brownian motion yields

$$\lim_{t \downarrow 0} t^\eta h(t) y(t) = \lim_{t \downarrow 0} t^{\varepsilon-1/2} \eta W(t^\eta, t^\eta) = 0$$

and finally

$$\lim_{t \downarrow 0} t^\eta h(t)\big(t^{-\eta} + m^{-1/2} y_n(t)\big) = \lim_{t \downarrow 0} t^{\eta+1/2+\varepsilon} \; q^{\leftarrow}(m/n)Q_n(t) = 0.$$

□

Moreover, according to Lemma 6.2 of Draisma et al. (2004), there exist suitable versions of Q_n and W such that

$$\sup_{0<t\le 1} t^{\eta+1/2+\varepsilon} |y_n(t) - y(t)| = o_P(1) \tag{3.44}$$

as $n \to \infty$ for all $\varepsilon > 0$. Thus,

$$\left| \int_0^1 t^\eta y_n(t) dt - \int_0^1 t^\eta y(t) dt \right| \leq \int_0^1 t^\eta |y_n(t) - y(t)| dt = o_P(1). \quad (3.45)$$

By Lemma 3.4.1 and (3.44) we may proceed as follows. Exactly as in Drees (1998b), Example 3.1, p. 103, where z, β and λ_n play the roles of Q_n, η and $m^{-1/2}$, we conclude that[3]

$$\begin{aligned} \hat{\eta}_n^H - \eta &= \int_0^1 m^{-1/2} \frac{t^\eta y_n(t) - y_n(1)}{1 + m^{-1/2} y_n(1)} \, dt + o_P(m^{-1/2}) \\ &= \frac{m^{-1/2}}{1 + m^{-1/2} y_n(1)} \int_0^1 (t^\eta y_n(t) - y_n(1)) dt + o_P(m^{-1/2}) \\ &= m^{-1/2} \int_0^1 (t^\eta y_n(t) - y_n(1)) dt + o_P(m^{-1/2}), \end{aligned}$$

such that (3.45) yields

$$\begin{aligned} \hat{\eta}_n^H - \eta = m^{-1/2} &\int_0^1 (t^\eta y(t) - y(1)) dt + o_P(m^{-1/2}) \\ = m^{-1/2} &\int_0^1 y(t) \nu_\eta^H(dt) + o_P(m^{-1/2}), \end{aligned}$$

which proves Lemma 3.1.6. The Taylor expansion of the reciprocal of (3.13) gives

$$\begin{aligned} \frac{1}{\hat{\eta}_n^H} &= \frac{1}{\eta\big(1 + m^{-1/2} \int_0^1 t^{-(\eta+1)} W(t^\eta, t^\eta) \nu_\eta^H(dt)\big) + o_P(m^{-1/2})} \\ &= \frac{1}{\eta} \left(1 - m^{-1/2} \int_0^1 t^{-(\eta+1)} W(t^\eta, t^\eta) \nu_\eta^H(dt)\right) + o_P(m^{-1/2}), \end{aligned}$$

as asserted in Corollary 3.1.7.

[3] When referring to Drees (1998b), note that $z_\beta(t) = t^{-\beta}$, if $\beta > 0$ (p.99 of Drees (1998b)), such that $T_H(z_\eta + m^{-1/2} y_n) = \hat{\eta}_n^H$.

Proof of Lemma 3.1.6 and Corollary 3.1.7 for the Maximum Likelihood Estimator $\hat{\eta}_n^{ML}$

The proof is based on results of Drees (1998a), Example 4.1, p. 197. The properties of the functional T_{ML}, which is considered subsequently, are shown there.

Since the MLE $\hat{\eta}_n^{ML}$ has no explicit representation, we need to consider the ML-functional $T_{ML} : \bar{D}_{\eta,h} \to \mathbb{R}$ (denoted as T in Example 4.1 of Drees (1998a)), which, applied to the tail empirical quantile function Q_n, yields the MLE $\hat{\eta}_n^{ML}$. With the scale invariance of T_{ML} and the definition (3.42) of y_n, we may write

$$\begin{aligned}\hat{\eta}_n^{ML} &= T_{ML}(Q_n) = T_{ML}\left(\left(\frac{t^{-\eta}}{\eta} + \frac{q^{\leftarrow}(m/n)\ Q_n(t)}{\eta} - \frac{t^{-\eta}}{\eta}\right)_{t\in(0,1]}\right)\\ &= T_{ML}\left(\left(\frac{t^{-\eta}}{\eta} + \frac{m^{-1/2}y_n(t)}{\eta}\right)_{t\in(0,1]}\right). \qquad (3.46)\end{aligned}$$

Similar as in the proof of the Hill case, consider y_n and y as defined in (3.43) with the versions of Q_n and W of Lemma 6.2 of Draisma et al. (2004), such that (3.44) holds. Further, we know from Drees (1998a) that $T_{ML}\big((t^{-\eta}/\eta)_{t\in(0,1]}\big) = \eta$. Then it is shown in Example 4.1 of Drees (1998a) that we may continue (3.46) to

$$\begin{aligned}\hat{\eta}_n^{ML} &= T_{ML}\left(\left(\frac{t^{-\eta}}{\eta}\right)_{t\in(0,1]}\right) + \frac{m^{-1/2}}{\eta}T'_{ML,(t^{-\eta}/\eta)_{t\in(0,1]}}(y) + o_P(m^{-1/2})\\ &= \eta + \frac{m^{-1/2}}{\eta}\frac{(\eta+1)^2}{\eta}\int_0^1 \big(t^{\eta} - (2\eta+1)t^{2\eta}\big)\big(y(t) - y(1)\big)dt + o_P(m^{-1/2}),\end{aligned}$$

where $T'_{ML,(t^{-\eta}/\eta)_{t\in(0,1]}}$ is the Hadamard derivative of T_{ML} tangentially to $C_{\eta,h}$ at $(t^{-\eta}/\eta)_{t\in(0,1]}$. (Note that we proved in Lemma 3.4.1, that y_n, y and $t^{-\eta}/\eta + m^{-1/2}y_n(t)/\eta$ are in the spaces required in Drees (1998a).)

Now observe that

$$\int_0^1 \big(t^{\eta} - (2\eta+1)t^{2\eta}\big)dt = -\frac{\eta}{\eta+1}$$

so that

$$\hat{\eta}_n^{ML} = \eta + m^{-1/2} \int_0^1 y(t)\nu_\eta^{ML}(dt) + o_P(m^{-1/2}).$$

3.4.2 Proofs for Section 3.2

We first recapitulate Lemma 6.1 of Draisma et al. (2004), since we need this result several times in this section. Define $S(x,y) := \sum_{i=1}^n \mathbb{1}\{U_i \le x, V_i \le y\}$.

3.4.2 Lemma. (Lemma 6.1 of Draisma et al. (2004)) Under the conditions of Theorem 3.2.1,

$$\sqrt{\tilde{m}}\left(\frac{S(U_{\lfloor nq^{\leftarrow}(\tilde{m}/n)x\rfloor:n}, V_{\lfloor nq^{\leftarrow}(\tilde{m}/n)y\rfloor:n})}{\tilde{m}} - c(x,y)\right)_{(x,y)\in[0,\infty)^2} \longrightarrow W \quad (3.47)$$

weakly in the bivariate Skorohod space $D\big([0,\infty)^2\big)$. $\triangle$

Proof of Theorem 3.2.1

Recall that for $a \in \mathbb{R}$ we denote by $\lfloor a \rfloor$ the largest integer less than or equal to a and by $\lceil a \rceil$ the smallest integer greater than or equal to a. Let $X_{i:n}$, $Y_{i:n}$, $U_{i:n}$ and $V_{i:n}$, $i = 1,\dots,n$, denote the order statistics pertaining to X_i, Y_i, U_i and V_i, $i = 1,\dots,n$, with the conventions $X_{0:n} := Y_{0:n} := -\infty$, $U_{0:n} := V_{0:n} := 0$, $X_{j:n} := Y_{j:n} := \infty$ and $U_{j:n} := V_{j:n} := 1$ for $j > n$. The first aim is to apply Lemma 3.4.2 to the numerator $\sum_{i=1}^n \mathbb{1}\{\hat{U}_i < rx, \hat{V}_i < ry\}$ and the denominator $\sum_{i=1}^n \mathbb{1}\{\hat{U}_i < r, \hat{V}_i < r\}$ of $\check{c}_n(x,y)$.

Recall that F_i is assumed continuous on $[\xi_i, \infty)$ for some $\xi_i < F_i^{\leftarrow}(1)$, $i = 1,2$. If $X_{n-k:n} \ge \xi_1$, then $X_{n-k:n} < X_{n-k+1:n} < \ldots < X_{n:n}$ almost surely (a.s.). Since $r \to 0$, and thus $X_{\lfloor n-(n+1)r\xi_0+1\rfloor:n} \to F_1^{\leftarrow}(1)$, it follows that the probability that there are no ties between these order statistics tends to 1, i.e. $P\{X_{\lfloor n-(n+1)r\xi_0+1\rfloor:n} < X_{\lfloor n-(n+1)r\xi_0+1\rfloor+1:n} < \ldots < X_{n:n}\} \to 1$ as $n \to \infty$ for all $\xi_0 > 0$. Furthermore, the condition $\hat{U}_i < rx$ is equivalent to $R_i^X > n-(n+1)rx+1$. If there are no ties, then this in turn is equivalent to $X_i > X_{\lfloor n-(n+1)rx+1\rfloor:n}$ and thus to $U_i < U_{\lceil (n+1)rx\rceil:n}$. Hence, together with analogous arguments for V_i and Y_i we see that $\check{c}_n(x,y)$ equals

$$\tilde{c}_n(x,y) := \frac{S(U_{\lceil (n+1)rx\rceil-1:n}, V_{\lceil (n+1)ry\rceil-1:n})}{S(U_{\lceil (n+1)r\rceil-1:n}, V_{\lceil (n+1)r\rceil-1:n})}$$

for all $x, y \in (0, \xi_0]$ with probability tending to 1 as $n \to \infty$ for all $\xi_0 > 0$, i.e.

$$P\big\{\tilde{c}_n(x,y) = \check{c}_n(x,y), \forall\, x, y \in (0,\xi_0]\big\} \xrightarrow{n\to\infty} 1.$$

It is therefore sufficient to prove the assertion of Theorem 3.2.1 for $\tilde{c}_n$, i.e. $\sqrt{\tilde{m}}\big(\tilde{c}_n - c\big) \to W - cW(1,1)$ weakly in $D^*\big([0,\infty)^2\big)$ uniformly on compact subsets of $[0,\infty)^2$.

In the proof of Theorem 2.3 of Drees and Huang (1998), it is shown that (a version of) W_2 has a.s. continuous sample paths.[4] The continuity of the sample paths of (a version of) W_1 follows with similar arguments as in this proof. (In Section 4.6.4 we also show that if $l = 0$, there are versions of W which are transformed Brownian sheets.) Hence, since (a version of) W has a.s. continuous sample paths, by the Skorohod-Dudley-Wichura representation theorem (see e.g. Shorack and Wellner (1986), p. 47, and cf. Appendix A), (3.47) holds a.s. uniformly on compact subsets of $[0,\infty)^2$ for suitable versions of the processes $\big(S(U_{\lfloor nrx\rfloor:n}, V_{\lfloor nry\rfloor:n})\big)_{(x,y)\in[0,\infty)^2}$ and W. Hence, for these versions,

$$\begin{aligned}
\tilde{c}_n(x,y) &= \frac{S(U_{\lceil (n+1)rx\rceil-1:n}, V_{\lceil (n+1)ry\rceil-1:n})}{S(U_{\lceil (n+1)r\rceil-1:n}, V_{\lceil (n+1)r\rceil-1:n})} \\
&= \frac{c(x,y) + \tilde{m}^{-1/2}W(x,y) + o(\tilde{m}^{-1/2})}{c(1,1) + \tilde{m}^{-1/2}W(1,1) + o(\tilde{m}^{-1/2})} \\
&= \Big(c(x,y) + \tilde{m}^{-1/2}W(x,y) + o(\tilde{m}^{-1/2})\Big) \\
&\qquad\qquad \Big(1 - \tilde{m}^{-1/2}W(1,1) + o(\tilde{m}^{-1/2})\Big) \qquad \text{a.s.,}
\end{aligned}$$

i.e.

$$\sqrt{\tilde{m}}\big(\tilde{c}_n(x,y) - c(x,y)\big) = W(x,y) - c(x,y)W(1,1) + o(1) \qquad \text{a.s.} \qquad (3.48)$$

uniformly on compact subsets of $[0,\infty)^2$.

Proof of Corollary 3.2.3

We start with the case $l = 0$, i.e. with asymptotically independent random variables U and V. Then, (3.48) and the covariance structure (3.9) of $W = W_1$

[4] To see that the process W of Drees and Huang (1998) is a version of our W_2, note that Drees and Huang (1998) denote the stable tail dependence function ℓ by l and recall (2.12).

yield

$$\begin{aligned}
\sigma_{x,y}^2 &= \operatorname{var}\big[W(x,y) - c(x,y)W(1,1)\big] \\
&= E\Big[\big(W(x,y) - c(x,y)W(1,1)\big)^2\Big] \\
&= c(x,y) - 2c(x,y)c(x \wedge 1, y \wedge 1) + c^2(x,y), \qquad (3.49)
\end{aligned}$$

which is the asymptotic variance mentioned in Remark 3.2.4 (ii). The assumption $x, y \in (0,1]$ yields the assertion.

It is convenient to first calculate $E[W(x_1,y_1)W(x_2,y_2)]$ in the case of asymptotic dependence. Recall from (3.19) that the derivatives c_x and c_y are homogeneous of order $1/\eta - 1$ and that $\eta = 1$ if U and V are asymptotically dependent, so that c is homogeneous of order 1 and c_x and c_y are homogeneous of order 0.

$$\begin{aligned}
&E[W(x_1,y_1)W(x_2,y_2)] \\
&= E\Big[l^{-1}\big[W_2(x_1,0)W_2(x_2,0) + W_2(x_1,0)W_2(0,y_2) - W_2(x_1,0)W_2(x_2,y_2) \\
&\qquad\qquad + W_2(0,y_1)W_2(x_2,0) + W_2(0,y_1)W_2(0,y_2) - W_2(0,y_1)W_2(x_2,y_2) \\
&\qquad\qquad - W_2(x_1,y_1)W_2(x_2,0) - W_2(x_1,y_1)W_2(0,y_2) + W_2(x_1,y_1)W_2(x_2,y_2)\big] \\
&\quad - c_x(x_2,y_2)\big[W_2(x_1,0)W_2(x_2,0) + W_2(0,y_1)W_2(x_2,0) - W_2(x_1,y_1)W_2(x_2,0)\big] \\
&\quad - c_y(x_2,y_2)\big[W_2(x_1,0)W_2(0,y_2) + W_2(0,y_1)W_2(0,y_2) - W_2(x_1,y_1)W_2(0,y_2)\big] \\
&\quad - c_x(x_1,y_1)\big[W_2(x_1,0)W_2(x_2,0) + W_2(x_1,0)W_2(0,y_2) - W_2(x_1,0)W_2(x_2,y_2)\big] \\
&\qquad\qquad + lc_x(x_1,y_1)c_x(x_2,y_2)W_2(x_1,0)W_2(x_2,0) \\
&\qquad\qquad + lc_x(x_1,y_1)c_y(x_2,y_2)W_2(x_1,0)W_2(0,y_2) \\
&\quad - c_y(x_1,y_1)\big[W_2(0,y_1)W_2(x_2,0) + W_2(0,y_1)W_2(0,y_2) - W_2(0,y_1)W_2(x_2,y_2)\big] \\
&\qquad\qquad + lc_y(x_1,y_1)c_x(x_2,y_2)W_2(0,y_1)W_2(x_2,0) \\
&\qquad\qquad + lc_y(x_1,y_1)c_y(x_2,y_2)W_2(0,y_1)W_2(0,y_2)\Big]
\end{aligned}$$

$$
\begin{aligned}
&= l^{-1}\big[x_1 \wedge x_2 + lc(x_1, y_2) - \big(x_1 \wedge x_2 - lc(x_2, y_2) + lc(x_1 \vee x_2, y_2)\big) \\
&\qquad\quad + lc(x_2, y_1) + y_1 \wedge y_2 - \big(y_1 \wedge y_2 - lc(x_2, y_2) + lc(x_2, y_1 \vee y_2)\big) \\
&\qquad\quad - \big(x_1 \wedge x_2 - lc(x_1, y_1) + lc(x_1 \vee x_2, y_1)\big) \\
&\qquad\quad - \big(y_1 \wedge y_2 - lc(x_1, y_1) + lc(x_1, y_1 \vee y_2)\big) \\
&\qquad\quad + x_1 \wedge x_2 + y_1 \wedge y_2 - lc(x_1, y_1) - lc(x_2, y_2) + lc(x_1 \vee x_2, y_1 \vee y_2)\big] \\
&\quad - c_x(x_2, y_2)\big[x_1 \wedge x_2 + lc(x_2, y_1) - (x_1 \wedge x_2 - lc(x_1, y_1) + lc(x_1 \vee x_2, y_1))\big] \\
&\quad - c_y(x_2, y_2)\big[lc(x_1, y_2) + y_1 \wedge y_2 - (y_1 \wedge y_2 - lc(x_1, y_1) + lc(x_1, y_1 \vee y_2))\big] \\
&\quad - c_x(x_1, y_1)\big[x_1 \wedge x_2 + lc(x_1, y_2) - (x_1 \wedge x_2 - lc(x_2, y_2) + lc(x_1 \vee x_2, y_2))\big] \\
&\quad + lc_x(x_1, y_1)c_x(x_2, y_2)(x_1 \wedge x_2) + l^2 c_x(x_1, y_1)c_y(x_2, y_2)c(x_1, y_2) \\
&\quad - c_y(x_1, y_1)\big[lc(x_2, y_1) + y_1 \wedge y_2 - (y_1 \wedge y_2 - lc(x_2, y_2) + lc(x_2, y_1 \vee y_2))\big] \\
&\quad + l^2 c_y(x_1, y_1)c_x(x_2, y_2)c(x_2, y_1) + lc_y(x_1, y_1)c_y(x_2, y_2)(y_1 \wedge y_2)
\end{aligned}
$$

$$
\begin{aligned}
&= l^{-1}\big[lc(x_1, y_2) - lc(x_1 \vee x_2, y_2) + lc(x_2, y_1) + lc(x_2, y_2) - lc(x_2, y_1 \vee y_2) \\
&\qquad\quad + lc(x_1, y_1) - lc(x_1 \vee x_2, y_1) - lc(x_1, y_1 \vee y_2)) + lc(x_1 \vee x_2, y_1 \vee y_2)\big] \\
&\quad - c_x(x_2, y_2)\big[lc(x_2, y_1) + lc(x_1, y_1) - lc(x_1 \vee x_2, y_1)\big] \\
&\quad - c_y(x_2, y_2)\big[lc(x_1, y_2) + lc(x_1, y_1) - lc(x_1, y_1 \vee y_2)\big] \\
&\quad - c_x(x_1, y_1)\big[lc(x_1, y_2) + lc(x_2, y_2) - lc(x_1 \vee x_2, y_2)\big] \\
&\quad + lc_x(x_1, y_1)c_x(x_2, y_2)(x_1 \wedge x_2) + l^2 c_x(x_1, y_1)c_y(x_2, y_2)c(x_1, y_2) \\
&\quad - c_y(x_1, y_1)\big[lc(x_2, y_1) + lc(x_2, y_2) - lc(x_2, y_1 \vee y_2)\big] \\
&\quad + l^2 c_y(x_1, y_1)c_x(x_2, y_2)c(x_2, y_1) + lc_y(x_1, y_1)c_y(x_2, y_2)(y_1 \wedge y_2)
\end{aligned}
$$

$$
\begin{aligned}
&= c(x_1, y_2) - c(x_1 \vee x_2, y_2) + c(x_2, y_1) + c(x_2, y_2) - c(x_2, y_1 \vee y_2) \\
&\quad + c(x_1, y_1) - c(x_1 \vee x_2, y_1) - c(x_1, y_1 \vee y_2) + c(x_1 \vee x_2, y_1 \vee y_2) \\
&\quad - lc_x(x_2, y_2)\big[c(x_2, y_1) + c(x_1, y_1) - c(x_1 \vee x_2, y_1)\big] \\
&\quad - lc_y(x_2, y_2)\big[c(x_1, y_2) + c(x_1, y_1) - c(x_1, y_1 \vee y_2)\big] \\
&\quad - lc_x(x_1, y_1)\big[c(x_1, y_2) + c(x_2, y_2) - c(x_1 \vee x_2, y_2) \\
&\qquad\qquad - c_x(x_2, y_2)(x_1 \wedge x_2) - lc_y(x_2, y_2)c(x_1, y_2)\big] \\
&\quad - lc_y(x_1, y_1)\big[c(x_2, y_1) + c(x_2, y_2) - c(x_2, y_1 \vee y_2) \\
&\qquad\qquad - c_y(x_2, y_2)(y_1 \wedge y_2) - lc_x(x_2, y_2)c(x_2, y_1)\big].
\end{aligned}
\tag{3.50}
$$

For the remainder of this proof, let $c_x := c_x(1,1)$ and $c_y := c_y(1,1)$. From (3.48) we obtain

$$\begin{aligned}
\sigma^2_{x,y} &= E\left[W^2(x,y)\right] - 2c(x,y)E\left[W(x,y)W(1,1)\right] + c^2(x,y)E\left[W^2(1,1)\right] \\
&= c(x,y) - lc_x(x,y)c(x,y) - lc_y(x,y)c(x,y) \\
&\quad -lc_x(x,y)[c(x,y) - xc_x(x,y) - lc_y(x,y)c(x,y)] \\
&\quad -lc_y(x,y)[c(x,y) - yc_y(x,y) - lc_x(x,y)c(x,y)] \\
&\quad -2c(x,y)\Big[c(x,1) - c(x\vee 1,1) + c(1,y) + 1 - c(1,y\vee 1) + c(x,y) \\
&\qquad\qquad -c(x\vee 1,y) - c(x,y\vee 1) + c(x\vee 1,y\vee 1) \\
&\qquad\qquad -lc_x\big[c(1,y) + c(x,y) - c(x\vee 1,y)\big] \\
&\qquad\qquad -lc_y\big[c(x,1) + c(x,y) - c(x,y\vee 1)\big] \\
&\qquad -lc_x(x,y)\big[c(x,1) + 1 - c(x\vee 1,1) - c_x(x\wedge 1) - lc_y c(x,1)\big] \\
&\qquad -lc_y(x,y)\big[c(1,y) + 1 - c(1,y\vee 1) - c_y(y\wedge 1) - lc_x c(1,y)\big]\Big] \\
&\quad +c^2(x,y)[1 - lc_x - lc_y - lc_x(1 - c_x - lc_y) - lc_y(1 - c_y - lc_x)] \\
&= c(x,y) - 2lc_x(x,y)c(x,y) - 2lc_y(x,y)c(x,y) \\
&\quad +lxc_x^2(x,y) + lyc_y^2(x,y) + 2l^2c_x(x,y)c_y(x,y)c(x,y) \\
&\quad -2c(x,y)\Big[c(x,1) - c(x\vee 1,1) + c(1,y) + 1 - c(1,y\vee 1) + c(x,y) \\
&\qquad\qquad -c(x\vee 1,y) - c(x,y\vee 1) + c(x\vee 1,y\vee 1) \\
&\qquad\qquad -lc_x\big[c(1,y) + c(x,y) - c(x\vee 1,y)\big] \\
&\qquad\qquad -lc_y\big[c(x,1) + c(x,y) - c(x,y\vee 1)\big] \\
&\qquad -lc_x(x,y)\big[c(x,1) + 1 - c(x\vee 1,1) - c_x(x\wedge 1) - lc_y c(x,1)\big] \\
&\qquad -lc_y(x,y)\big[c(1,y) + 1 - c(1,y\vee 1) - c_y(y\wedge 1) - lc_x c(1,y)\big]\Big] \\
&\quad +c^2(x,y)[1 - 2l(c_x + c_y) + l(c_x + c_y)^2 - 2lc_xc_y(1-l)]
\end{aligned} \tag{3.51}$$

which is the asymptotic variance mentioned in Remark 3.2.4 (ii). In the case

$x, y \in (0,1]$, the second summand can be considerably simplified to obtain

$$\begin{aligned}
\sigma_{x,y}^2 = \; & c(x,y) - 2lc_x(x,y)c(x,y) - 2lc_y(x,y)c(x,y) \\
& + lxc_x^2(x,y) + lyc_y^2(x,y) + 2l^2c_x(x,y)c_y(x,y)c(x,y) \\
& - 2c(x,y)\Big[c(x,y)(1-l) - lc_x(x,y)\big[c(x,1) - xc_x - lc_y c(x,1)\big] \\
& \qquad\qquad\qquad\qquad - lc_y(x,y)\big[c(1,y) - yc_y - lc_x c(1,y)\big]\Big] \\
& + c^2(x,y)[1 - 2l(c_x + c_y) + l(c_x + c_y)^2 - 2lc_xc_y(1-l)].
\end{aligned}$$

Proof of Theorem 3.2.5

It is sufficient to prove the consistency of $\hat{c}_x(x,1)$. From the proof of Theorem 3.1.4, recall (3.39), $\hat{k}/k \to 1$ in probability and $k^{1/4}m^{-1/2} \to 0$. We obtain

$$\begin{aligned}
\hat{c}_x(x,1) &= \frac{\hat{k}^{5/4}}{n} \cdot \frac{n}{k}\left(\frac{k}{n}T_{n-m:n}^{(n,x-1+\hat{k}^{-1/4})} - \frac{k}{n}T_{n-m:n}^{(n,x-1)}\right) \\
&= \big(1 + o_P(1)\big)\left(\frac{c(x+\hat{k}^{-1/4},1) - c(x,1)}{\hat{k}^{-1/4}} + O_P\big(k^{1/4}m^{-1/2}\big)\right) \\
&\to c_x(x,1)
\end{aligned}$$

in probability. Since c_x and c_y are homogeneous of order 0 and because the consistency of $\hat{l}$, $\hat{c}_x(1,1)$, $\hat{c}_y(1,1)$ and $\check{c}(x,y)$ has been established in Theorem 3.1.4 and Corollary 3.2.3, the consistency of $\hat{\sigma}_{x,y}^2$ follows.

Proof of Theorem 3.2.7 and Corollary 3.2.8

Similar as in the proof of Theorem 3.2.1, we argue that

$$\begin{aligned}\hat{c}_n(x,y) &= \frac{\sum_{i=1}^{n} \mathbb{1}\left\{\hat{U}_i < x/T^{(n)}_{n-m:n}, \hat{V}_i < y/T^{(n)}_{n-m:n}\right\}}{\sum_{i=1}^{n} \mathbb{1}\left\{\hat{U}_i < 1/T^{(n)}_{n-m:n}, \hat{V}_i < 1/T^{(n)}_{n-m:n}\right\}} \\ &= \frac{S\left(U_{\lceil x(n+1)/T^{(n)}_{n-m:n}\rceil -1:n}, V_{\lceil y(n+1)/T^{(n)}_{n-m:n}\rceil -1:n}\right)}{S\left(U_{\lceil (n+1)/T^{(n)}_{n-m:n}\rceil -1:n}, V_{\lceil (n+1)/T^{(n)}_{n-m:n}\rceil -1:n}\right)} \end{aligned} \tag{3.52}$$

for all $x, y \in (0, \xi_0]$ with probability tending to 1 as $n \to \infty$ for all $\xi_0 > 0$.

Further, again similar as in the proof of Theorem 3.2.1 we argue that according to Lemma 3.4.2 and the Skorohod-Dudley-Wichura representation theorem, there is a version $\big(S_n^*(x,y)\big)_{(x,y)\in[0,\infty)^2}$ of the stochastic process $\big(S(U_{\lfloor nq^{\leftarrow}(m/n)x\rfloor:n}, V_{\lfloor nq^{\leftarrow}(m/n)y\rfloor:n})\big)_{(x,y)\in[0,\infty)^2}$ and a version of W such that

$$\sqrt{m}\left(\frac{S_n^*(x,y)}{m} - c(x,y)\right)_{(x,y)\in[0,\infty)^2} \longrightarrow \big(W(x,y)\big)_{(x,y)\in[0,\infty)^2} \tag{3.53}$$

holds a.s. uniformly on compact subsets of $[0,\infty)^2$. Note that by the existence of such versions it is *not* ensured that (3.53) holds a.s. with the process $S_n^*(x,y)$ replaced with $S\big(U_{\lfloor nq^{\leftarrow}(m/n)x\rfloor:n}, V_{\lfloor nq^{\leftarrow}(m/n)y\rfloor:n}\big)$ or even with $S\big(U_{\lfloor nx/T^{(n)}_{n-m:n}\rfloor:n}, V_{\lfloor ny/T^{(n)}_{n-m:n}\rfloor:n}\big)$ as it would be useful to determine the distribution of the right hand side of (3.52). It is not even ensured that $S_n^*(x,y)$ can be represented in one of these forms. However, we next show that the version of $T^{(n)}_{n-m:n}$ which is defined via the process S_n^* can be used to obtain a version of the right hand side of (3.52).

First note that

$$
\begin{aligned}
T_i^{(n)} \geq x \Leftrightarrow \min\left(\frac{n+1}{n+1-R_i^X}, \frac{n+1}{n+1-R_i^Y}\right) \geq x \\
\Leftrightarrow \frac{n+1-R_i^X}{n+1} \leq \frac{1}{x}, \frac{n+1-R_i^Y}{n+1} \leq \frac{1}{x} \\
\Leftrightarrow R_i^X \geq \lceil (n+1)(1-1/x) \rceil,\ R_i^Y \geq \lceil (n+1)(1-1/x) \rceil.
\end{aligned}
$$

If there are no ties, then $R_i^X \geq \lceil (n+1)(1-1/x) \rceil$ is equivalent to $X_i \geq X_{\lceil n+1-(n+1)/x \rceil : n}$ and thus to $U_i \leq U_{\lfloor (n+1)/x \rfloor : n}$. Hence, with similar arguments for V_i and Y_i, we see that for the (discrete and non-increasing) process

$$
\begin{aligned}
T_{n-j:n}^{(n)} &= \sup\left\{x \in (0,\infty) \;\middle|\; \sum_{i=1}^n \mathbb{1}_{\left\{T_i^{(n)} \geq x\right\}} \geq j+1\right\} \\
&= \sup\left\{x \in (0,\infty) \mid S(U_{\lfloor (n+1)/x \rfloor : n}, V_{\lfloor (n+1)/x \rfloor : n}) \geq j+1\right\}, \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad j = 0, \ldots, n-1.
\end{aligned}
$$

Define the (non-decreasing) process $\bar{S}_n^*$ by $\bar{S}_n^*(x) := S_n^*(x,x)$, where we assume that $S_n^*(x,x)$ (and thus $\bar{S}_n^*$) is right-continuous. It follows that the (right-continuous non-increasing) process[5]

$$
Q_n^*(t) := \sup\left\{x \in (0,\infty) \mid \bar{S}_n^*\left(\tfrac{n+1}{nq^{\leftarrow}(m/n)x}\right) \geq \lfloor mt \rfloor + 1\right\},\ t \in [0,1],
$$

possesses the same distribution as the tail empirical quantile process $Q_n = \left(T_{n-\lfloor mt \rfloor : n}^{(n)}\right)_{t \in [0,1]}$. In particular, note that $Q_n^*(1)$ has the same distribution as $T_{n-m:n}^{(n)}$. Thus,

$$
\left(S_n^*\left(\frac{(n+1)x}{nq^{\leftarrow}(m/n)Q_n^*(1)}-, \frac{(n+1)y}{nq^{\leftarrow}(m/n)Q_n^*(1)}-\right)\right)_{(x,y)\in(0,\infty)^2} \tag{3.54}
$$

is a version of the process given in the numerator of the right hand side of (3.52), where $f(x-,y-)$ denotes the left hand limit of f at (x,y) in the sense

[5] Here and below it is useful to note that $\bar{S}_n^*\big((n+1)/(nq^{\leftarrow}(m/n)x)\big)$ is left-continuous and non-increasing in x.

of Definition A.1. In order to determine the distribution of (3.54), we now establish an asymptotic expansion of $1/Q_n^*(1)$.

Note that for the (left-continuous non-increasing) process

$$
\begin{aligned}
&(Q_n^*)^{\leftarrow}(x) \\
&:= \sup\left\{t \in [0,1] \mid Q_n^*(t) \geq x\right\} \\
&= \sup\left\{t \in [0,1] \;\middle|\; \sup\left\{u \in (0,\infty) \;\middle|\; \bar{S}_n^*\Big(\tfrac{n+1}{nq^{\leftarrow}(m/n)u}\Big) \geq \lfloor mt \rfloor + 1\right\} \geq x\right\} \\
&= \sup\left\{t \in [0,1] \;\middle|\; \bar{S}_n^*\Big(\tfrac{n+1}{nq^{\leftarrow}(m/n)x}\Big) \geq \lfloor mt \rfloor + 1\right\} = \frac{1}{m}\, \bar{S}_n^*\left(\tfrac{n+1}{nq^{\leftarrow}(m/n)x}\right),
\end{aligned}
$$

so that

$$
(Q_n^*)^{\leftarrow}\left(\frac{1}{q^{\leftarrow}(m/n)x}\right) = \frac{1}{m}\bar{S}_n^*\left(\frac{n+1}{n}\, x\right)
$$

and thus by (3.53)

$$
\sqrt{m}\left((Q_n^*)^{\leftarrow}\Big(\tfrac{1}{q^{\leftarrow}(m/n)x}\Big) - x^{1/\eta}\right)_{x\in(0,\infty)} \to \big(W(x,x)\big)_{x\in(0,\infty)} \quad \text{a.s.}
$$

Now one can proceed as in the proof of Lemma 6.2 of Draisma et al. (2004), i.e.

$$
\sqrt{m}\Big((Q_n^*)^{\leftarrow}\Big(\tfrac{x^{-\eta}}{q^{\leftarrow}(m/n)}\Big) - x\Big)_{x\in(0,\infty)} \to \big(W(x^\eta, x^\eta)\big)_{x\in(0,\infty)} \quad \text{a.s.}
$$

$$
\sqrt{m}\Big(\big(q^{\leftarrow}(m/n)Q_n^*(t)\big)^{-1/\eta} - t\Big)_{t\in(0,\infty)} \to \big(-W(t^\eta, t^\eta)\big)_{t\in(0,\infty)} \quad \text{a.s.}
$$

in the Skorohod space consisting of all functions on $(0,\infty)$ which are left-continuous with right limits existing at each point of $(0,\infty)$. In the last step we used the obvious modification of the Lemma of Vervaat (1972) for this space. Hence, still as in the proof of Lemma 6.2 of Draisma et al. (2004),

$$
\begin{aligned}
q^{\leftarrow}(m/n)Q_n^*(t) &= \Big(t - m^{-1/2}W(t^\eta, t^\eta) + o(m^{-1/2})\Big)^{-\eta} \\
&= t^{-\eta}\left(1 + m^{-1/2}\eta t^{-1}W(t^\eta, t^\eta) + o(m^{-1/2})\right) \\
&= t^{-\eta} + m^{-1/2}\eta t^{-(\eta+1)}W(t^\eta, t^\eta) + o(m^{-1/2}) \quad \text{a.s.} \qquad (3.55)
\end{aligned}
$$

uniformly for t belonging to some compact interval bounded away from 0. Since $1/(1+z) = 1 - z + o(z)$ as $z \to 0$, this yields

$$\frac{1}{q^{\leftarrow}(m/n)Q_n^*(1)} = 1 - m^{-1/2}\eta W(1,1) + o(m^{-1/2}) \quad \text{a.s.} \tag{3.56}$$

for $t = 1$, which is the desired asymptotic expansion of $1/Q_n^*(1)$.

Hence, by (3.53) with x and y replaced with $x\big(1 - m^{-1/2}\eta W(1,1) + o(m^{-1/2})\big)$ and $y\big(1 - m^{-1/2}\eta W(1,1) + o(m^{-1/2})\big)$, and the homogeneity of c of order $1/\eta$ and of c_x and c_y of order $1/\eta - 1$,

$$\begin{aligned}
&m^{-1}S_n^*\left(\frac{(n+1)x}{nq^{\leftarrow}(m/n)Q_n^*(1)}-, \frac{(n+1)y}{nq^{\leftarrow}(m/n)Q_n^*(1)}-\right)\\
&= c\left(x\big(1-m^{-1/2}\eta W(1,1)+o(m^{-1/2})\big), y\big(1-m^{-1/2}\eta W(1,1)+o(m^{-1/2})\big)\right)\\
&\qquad\qquad +m^{-1/2}W(x,y) + o(m^{-1/2}) \text{ a.s.}\\
&= c(x,y) - m^{-1/2}\eta W(1,1)xc_x(x,y) - m^{-1/2}\eta W(1,1)yc_y(x,y)\\
&\qquad\qquad +m^{-1/2}W(x,y) + o(m^{-1/2}) \text{ a.s.}
\end{aligned}$$

uniformly on compact subsets of $[0,\infty)^2$. By (3.22), we continue this to

$$\begin{aligned}
&m^{-1}S_n^*\left(\frac{(n+1)x}{nq^{\leftarrow}(m/n)Q_n^*(1)}-, \frac{(n+1)y}{nq^{\leftarrow}(m/n)Q_n^*(1)}-\right)\\
&= c(x,y) - m^{-1/2}W(1,1)c(x,y) + m^{-1/2}W(x,y) + o(m^{-1/2})\\
&= c(x,y) + m^{-1/2}\big(W(x,y) - W(1,1)c(x,y)\big) + o(m^{-1/2}) \quad \text{a.s.}
\end{aligned} \tag{3.57}$$

and thus

$$m/S_n^*\left(\frac{(n+1)}{nq^{\leftarrow}(m/n)Q_n^*(1)}-, \frac{(n+1)}{nq^{\leftarrow}(m/n)Q_n^*(1)}-\right) = 1 + o(m^{-1/2}) \quad \text{a.s.} \tag{3.58}$$

Now combine (3.57) and (3.58) and recall that (3.54) is a version of the numerator of the right-hand side of (3.52) to see that

$$\sqrt{m}\big(\hat{c}_n(x,y) - c(x,y)\big) = W(x,y) - c(x,y)W(1,1) + o(1)$$

holds a.s. uniformly on compact subsets of $[0,\infty)^2$ for versions of Q_n and W. If we compare this with (3.48), the assertion follows.

3.4.3 Proofs for Section 3.3

Proof of Theorem 3.3.1.

We argue as at the beginning of the proof of Theorem 3.2.1 to obtain that

$$\begin{aligned}
\hat{p}_n &= \frac{r^{-1/\hat{\eta}_n}}{n} \cdot S\Big(U_{\lceil (n+1)r(1-\hat{F}_{1,n}(u_1))\rceil - 1:n}, V_{\lceil (n+1)r(1-\hat{F}_{2,n}(u_2))\rceil - 1:n}\Big) \\
&= \frac{r^{-1/\hat{\eta}_n}}{n} \cdot S\left(U_{\lceil kx\rceil - 1:n}, V_{\lceil ky\rceil - 1:n}\right)\Big|_{\substack{x=(n+1)r(1-\hat{F}_{1,n}(u_1))/k \\ y=(n+1)r(1-\hat{F}_{2,n}(u_2))/k}}
\end{aligned}$$

for any $k > 0$ with probability tending to 1 as $n \to \infty$. Thus,

$$\begin{aligned}
&\log \frac{\hat{p}_n}{p} \\
&= \left(\frac{1}{\eta} - \frac{1}{\hat{\eta}_n}\right)\log r + \log \frac{n^{-1} S\left(U_{\lceil kx\rceil - 1:n}, V_{\lceil ky\rceil - 1:n}\right)}{q(k/n)c(x,y)}\Bigg|_{\substack{x=(n+1)r(1-\hat{F}_{1,n}(u_1))/k \\ y=(n+1)r(1-\hat{F}_{2,n}(u_2))/k}} \\
&\quad + \log \frac{c\Big((n+1)r(1-\hat{F}_{1,n}(u_1))/k, (n+1)r(1-\hat{F}_{2,n}(u_2))/k\Big)}{c\Big((n+1)r(1-F_1(u_1))/k, (n+1)r(1-F_2(u_2))/k\Big)} \\
&\quad + \log\left[r^{-1/\eta}\frac{c\Big((n+1)r(1-F_1(u_1))/k, (n+1)r(1-F_2(u_2))/k\Big)q(k/n)}{p}\right] \\
&=: T_1 + T_2 + T_3 + T_4.
\end{aligned}$$

First, note that

$$T_1 = o_P\Big(j_1^{-1/2}\log^2\Big(\frac{j_1}{n(1-F_1(u_1))}\Big)\Big) \qquad \text{and}$$

$$T_1 = o_P\Big(j_2^{-1/2}\log^2\Big(\frac{j_2}{n(1-F_2(u_2))}\Big)\Big)$$

if assumptions (3.30) and (3.32), respectively, hold.

While for the treatment of T_1 and T_3, k may indeed be any positive value, we need to define $k := \lfloor nr(1 - F_1(u_1))\rfloor$ when looking at T_2 and T_4. Observe that according to assumption (iv), $nq(k/n) \sim nq\big(r(1-F_1(u_1))\big) \to \infty$ and that assumption (v) implies $r(1-F_1(u_1)) \to 0$, so that $q(k/n) \sim q\big(r(1-F_1(u_1))\big) \to 0$ as $n \to \infty$; i.e. $nq(k/n)$ is an intermediate sequence. Hence, it is again Lemma 3.4.2 with (the non-integer) $nq(k/n)$ playing the role of (the integer) $\tilde{m}$, that yields

$$\sqrt{nq(k/n)}\bigg(\frac{n^{-1}S\big(U_{\lceil kx\rceil-1:n}, V_{\lceil ky\rceil-1:n}\big)}{q(k/n)c(x,y)} - 1\bigg) \to \frac{W(x,y)}{c(x,y)}$$

weakly in the bivariate Skorohod space $D^*\big((0,\infty)^2\big)$. With the same arguments as in the proof of Theorem 3.2.1, i.e. according to the Skorohod-Dudley-Wichura representation theorem, this convergence holds a.s. for suitable versions of $\big(S\big(U_{\lceil kx\rceil-1:n}, V_{\lceil ky\rceil-1:n}\big)\big)_{(x,y)\in[0,\infty)^2}$ and W. Hence, for these versions,

$$\begin{aligned} T_2 &= \log\bigg[1 + (nq(k/n))^{-1/2}\frac{W(x,y)}{c(x,y)}\bigg|_{\substack{x=(n+1)r(1-\hat{F}_{1,n}(u_1))/k\\ y=(n+1)r(1-\hat{F}_{2,n}(u_2))/k}} + o\big((nq(k/n))^{-1/2}\big)\bigg] \\ &= (nq(k/n))^{-1/2}\frac{W(x,y)}{c(x,y)}\bigg|_{\substack{x=(n+1)r(1-\hat{F}_{1,n}(u_1))/k\\ y=(n+1)r(1-\hat{F}_{2,n}(u_2))/k}} + o\big((nq(k/n))^{-1/2}\big), \end{aligned}$$

since here x and y are a.s. bounded and bounded away from 0.

For T_3, first observe that

$$\begin{aligned} T_3 &= \log\left[\left(\frac{1-\hat{F}_{2,n}(u_2)}{1-F_2(u_2)}\right)^{1/\eta}\cdot\frac{c\Big(\frac{1-\hat{F}_{1,n}(u_1)}{1-\hat{F}_{2,n}(u_2)},1\Big)}{c\Big(\frac{1-F_1(u_1)}{1-F_2(u_2)},1\Big)}\right] \\ &= \frac{1}{\eta}\log\frac{1-\hat{F}_{2,n}(u_2)}{1-F_2(u_2)} + \log\frac{c\Big(\frac{1-\hat{F}_{1,n}(u_1)}{1-\hat{F}_{2,n}(u_2)},1\Big)}{c\Big(\frac{1-F_1(u_1)}{1-F_2(u_2)},1\Big)} =: T_{31} + T_{32}. \end{aligned}$$

For T_{31}, by (1.21),

$$T_{31} = \frac{1}{\eta}\log\big(1 + O_P(m^{-1/2})\big) = O_P\big(m^{-1/2}\big)$$

if assumption (3.28) holds and for suitable versions

$$T_{31} = \frac{1}{\eta}\Big(\frac{1-\hat{F}_{2,n}(u_2)}{1-F_2(u_2)} - 1\Big) = o_P\Big(\frac{1-\hat{F}_{1,n}(u_1)}{1-F_1(u_1)} - 1\Big)$$

if assumptions (3.30) hold. For T_{32}, with assumption (ii),

$$c\Big(\frac{1-\hat{F}_{1,n}(u_1)}{1-\hat{F}_{2,n}(u_2)}, 1\Big) = c\Big(\frac{1-F_1(u_1)}{1-F_2(u_2)}, 1\Big) + c_x\Big(\frac{1-F_1(u_1)}{1-F_2(u_2)}, 1\Big) \times\Big(\frac{1-\hat{F}_{1,n}(u_1)}{1-\hat{F}_{2,n}(u_2)} - \frac{1-F_1(u_1)}{1-F_2(u_2)}\Big)\big(1+o_P(1)\big).$$

and by (1.21),

$$\frac{1-\hat{F}_{1,n}(u_1)}{1-\hat{F}_{2,n}(u_2)} - \frac{1-F_1(u_1)}{1-F_2(u_2)} = \Big(\frac{1-\hat{F}_{1,n}(u_1)}{1-F_1(u_1)} \cdot \frac{1-F_2(u_2)}{1-\hat{F}_{2,n}(u_2)} - 1\Big)\frac{1-F_1(u_1)}{1-F_2(u_2)}.$$

This equals

$$\Big(\big(1+O_P(m^{-1/2})\big)\big(1+O_P(m^{-1/2})\big) - 1\Big)\frac{1-F_1(u_1)}{1-F_2(u_2)} = O_P\big(m^{-1/2}\big)$$

if assumption (3.28) holds, so that

$$T_3 = O_P\big(m^{-1/2}\big) + \log\frac{c\Big(\frac{1-F_1(u_1)}{1-F_2(u_2)}, 1\Big) + O_P\big(m^{-1/2}\big)}{c\Big(\frac{1-F_1(u_1)}{1-F_2(u_2)}, 1\Big)} = O_P\big(m^{-1/2}\big)$$

in this case. Moreover, for suitable versions, by assumption (ii) and the homogeneity properties (2.14) and (3.19) of c of order $1/\eta$ and of c_x of order $1/\eta - 1$,

$$T_{32} = \log\Big[1 + \frac{c_x(1/\lambda,1)}{c(1/\lambda,1)}\Big(\frac{1-\hat{F}_{1,n}(u_1)}{1-F_1(u_1)}\Big(1+o_P\Big(\frac{1-\hat{F}_{1,n}(u_1)}{1-F_1(u_1)} - 1\Big)\Big) - 1\Big) \cdot\frac{1}{\lambda}\big(1+o_P(1)\big)\Big]$$

$$= \frac{c_x(1,\lambda)}{c(1,\lambda)}\Big(\frac{1-\hat{F}_{1,n}(u_1)}{1-F_1(u_1)} - 1\Big)\big(1+o_P(1)\big) = T_3\big(1+o_P(1)\big)$$

if assumptions (3.30) hold. Further, note that one may also write

$$T_3 = \log\left[\left(\frac{1-\hat{F}_{1,n}(u_1)}{1-F_1(u_1)}\right)^{1/\eta} \cdot \frac{c\Big(1, \frac{1-\hat{F}_{2,n}(u_2)}{1-\hat{F}_{1,n}(u_1)}\Big)}{c\Big(1, \frac{1-F_2(u_2)}{1-F_1(u_1)}\Big)}\right]$$

which leads with analogous arguments as before to

$$T_3 = \lambda \frac{c_y(1,\lambda)}{c(1,\lambda)}\left(\frac{1-\hat{F}_{2,n}(u_2)}{1-F_2(u_2)} - 1\right)\big(1+o_P(1)\big)$$

if assumptions (3.32) hold.

For T_4,

$$\begin{aligned} T_4 &= \log\left[\frac{q(k/n)}{p} \cdot r^{-1/\eta} \cdot c\Big((n+1)r(1-F_1(u_1))/k, (n+1)r(1-F_2(u_2))/k\Big)\right] \\ &= \log\left[\frac{q(k/n)}{p}\left(\frac{n+1}{k}(1-F_1(u_1))\right)^{1/\eta} c\left(1, \frac{1-F_2(u_2)}{1-F_1(u_1)}\right)\right]. \end{aligned}$$

Our basic equation (2.8) with $x=1$, $y = \big(1-F_2(u_2)\big)/\big(1-F_1(u_1)\big) \to \lambda$ and $t = 1-F_1(u_1)$, and assumptions (i) and (ii) yield

$$\frac{p}{q\big(1-F_1(u_1)\big)} - c\left(1, \frac{1-F_2(u_2)}{1-F_1(u_1)}\right) = O\big(q_1(1-F_1(u_1))\big)$$

as $n \to \infty$, so that

$$T_4 = \log\left[\frac{q(k/n)}{\big(k/(n+1)\big)^{1/\eta}} \cdot \frac{\big(1-F_1(u_1)\big)^{1/\eta}}{q\big(1-F_1(u_1)\big)}\big(1+O(q_1(1-F_1(u_1)))\big)\right].$$

(Here, the idea to apply (2.8) once more with $x = y = n(1-F_1(u_1))/k$ and $t = k/n$ does not work, since then $x = y \sim 1/r \to 0$, such that x and y are not bounded away from 0.) Next, recall that $q(tx)/q(t) = x^{1/\eta} + O(q_1(t))$ to see that Corollary 3.12.3 of Bingham et al. (1987) yields[6]

$$q(t) = Ct^{1/\eta}\big(1+O(q_1(t))\big) \tag{3.59}$$

[6] To see that q_1 has positive decrease as required by Corollary 3.12.3 of Bingham et al. (1987), note that a τ-regularly varying function has upper Matuszewska index τ, cf. Sections 2.0 and 2.1 (in particular the definitions on p. 68 and 71 and (2.1.10)) of Bingham et al. (1987).

for some constant $C > 0$. From representation (3.59) we obtain

$$\frac{q(k/n)\big(n(1-F_1(u_1))/k\big)^{1/\eta}}{q(1-F_1(u_1))} = \frac{1+O\big(q_1(k/n)\big)}{1+O\big(q_1(1-F_1(u_1))\big)},$$

so that by assumption (v),

$$\begin{aligned} T_4 &= \log\Big[1+O\big(q_1(k/n)+q_1(1-F_1(u_1))+1/n\big)\Big] \\ &= O\big(q_1(k/n)+q_1(1-F_1(u_1))+1/n\big) = o\big((nq(k/n))^{-1/2}\big). \end{aligned}$$

Note that by the assumption that either (3.25) or (3.26) holds, we have shown so far that

$$\log\frac{\hat{p}_n}{p} = \left(\frac{1}{\eta}-\frac{1}{\hat{\eta}_n}\right)\log r + O_P\big(m^{-1/2}\big)$$

if assumption (3.28) holds, that

$$\log\frac{\hat{p}_n}{p} = \frac{c_x(1,\lambda)}{c(1,\lambda)}\left(\frac{1-\hat{F}_{1,n}(u_1)}{1-F_1(u_1)}-1\right)\big(1+o_P(1)\big)$$

if assumptions (3.30) hold and that

$$\log\frac{\hat{p}_n}{p} = \lambda\frac{c_y(1,\lambda)}{c(1,\lambda)}\left(\frac{1-\hat{F}_{2,n}(u_2)}{1-F_2(u_2)}-1\right)\big(1+o_P(1)\big)$$

if assumptions (3.32) hold. In the first case, we obtain by Theorem 3.1.2 that $\hat{\eta}_n\eta \to \eta^2$ in probability and hence

$$\frac{\eta^2 m^{1/2}}{\log r}\log\frac{\hat{p}_n}{p} \sim m^{1/2}(\hat{\eta}_n-\eta) \xrightarrow{D} \mathcal{N}\big(0,\sigma^2_{\hat{\eta}_n}\big)$$

as asserted. In the second case, we obtain by (1.21) that

$$\begin{aligned} &\frac{c(1,\lambda)}{c_x(1,\lambda)}\cdot\frac{\gamma j_1^{1/2}}{\log\frac{j_1}{n(1-F_1(u_1))}}\log\frac{\hat{p}_n}{p} \\ &\qquad\qquad \sim \frac{\gamma j_1^{1/2}}{\log\frac{j_1}{n(1-F_1(u_1))}}\left(\frac{1-\hat{F}_{1,n}(u_1)}{1-F_1(u_1)}-1\right) \xrightarrow{D} \mathcal{N}\big(0,\sigma^2_{\hat{\gamma}_{1,n}}\big). \end{aligned}$$

The conclusion in the third case is analogous.

Chapter 4

Model Validation

In this chapter we develop a method to validate the scaling law (2.17) for given data. In Section 4.1 we start with a motivation of our construction. We prove that under some regularity conditions an empirical process which measures random deviations from this scaling law converges weakly to a centered Gaussian process if s, x and y (cf. (2.17)) are bounded away from 0. We conclude that these random deviations from the scaling law are asymptotically normal for all s, x and y in both cases – asymptotic independence (Section 4.2) and asymptotic dependence (Section 4.3). In Section 4.4 we address the issue of an appropriate choice of s, x and y when validating the scaling law, construct a test statistic to discriminate between presence and absence of the scaling law and suggest a method to validate the scaling law by a simple three-dimensional plot. Section 4.5 presents some further results which are useful to obtain a uniform test statistic for the entirety of the pseudo-(tail-)observations rather than considering m tests for each of these observations as it is done in Section 4.4. In particular, in Section 4.5.1 it is proven that the weighted limiting process converges uniformly in the required sense. In Section 4.5.2, we establish a similar result for the empirical process, where a stronger uniformity condition than in Section 4.5.1 is required. As before, the proofs of the results of this chapter are deferred to the final Section 4.6.

4.1 A Uniform Result

Let us recall the scaling law (2.17) with a contraction factor $s \in (0,1]$, i.e.

$$\lim_{t \downarrow 0} \frac{P\{U < tsx, V < tsy\}}{P\{U < tx, V < ty\}} = s^{1/\eta}. \tag{4.1}$$

If we switch from this asymptotic equality to an approximate equality for sufficiently small $u = tx$, $v = ty$ and take logarithms, we obtain

$$\log \frac{P\{U < su, V < sv\}}{P\{U < u, V < v\}} \approx \frac{1}{\eta} \log s.$$

Now suppose that we choose s such that pseudo-observations $(\hat{U}_i, \hat{V}_i)$ exist with $\hat{U}_i < su$, $\hat{V}_i < sv$ and estimate $P\{U < su, V < sv\}$ by the empirical probability $n^{-1} \sum_{i=1}^{n} \mathbb{1}\{\hat{U}_i < su, \hat{V}_i < sv\}$. Then, in analogy to the concept of pp-plots, the points

$$\left(\log s, \log \frac{\sum_{i=1}^{n} \mathbb{1}_{\{\hat{U}_i < su, \hat{V}_i < sv\}}}{\sum_{i=1}^{n} \mathbb{1}_{\{\hat{U}_i < u, \hat{V}_i < v\}}} \right) \tag{4.2}$$

must approximately lie on the line through origin with slope $1/\eta$ if the scaling law (4.1) is satisfied. This should be true for arbitrary, sufficiently small $u, v > 0$ and all $s \in (0,1]$.

In order to construct a test statistic to discriminate between presence and absence of the scaling law using this basic fact, we must first replace the unknown slope of the line with a suitable estimator, e.g. the reciprocal $1/\hat{\eta}_n$ of the Hill estimator $\hat{\eta}_n^H$ or the MLE $\hat{\eta}_n^{ML}$. Next we must specify what "sufficiently small" precisely means, i.e. in an asymptotic setting, we must specify the rate at which u and v tend to 0. To this end, recall that $\hat{\eta}_n^H$ and $\hat{\eta}_n^{ML}$ are based on those $T_j^{(n)}$ that satisfy $T_j^{(n)} = 1/\max\left(\hat{U}_j, \hat{V}_j\right) \geq T_{n-m:n}^{(n)}$, and thus on those pseudo-observations $\left(\hat{U}_j, \hat{V}_j\right)$ which fall into the square $\left(0, 1/T_{n-m:n}^{(n)}\right]^2$. They are reasonable estimators of η if and only if the points (4.2) with $u, v \leq 1/T_{n-m:n}^{(n)}$ lie approximately on the line through the origin with slope $1/\eta$. Therefore, it is natural to consider values $u = x/T_{n-m:n}^{(n)}$ and $v = y/T_{n-m:n}^{(n)}$ that converge with rate $1/T_{n-m:n}^{(n)}$ to 0.

Finally, to discriminate between random and systematic deviations of the points (4.2) from the line, the size of the estimated deviation

$$\log \frac{\sum_{i=1}^{n} \mathbb{1}_{\{\hat{U}_i < sx/T^{(n)}_{n-m:n},\, \hat{V}_i < sy/T^{(n)}_{n-m:n}\}}}{\sum_{i=1}^{n} \mathbb{1}_{\{\hat{U}_i < x/T^{(n)}_{n-m:n},\, \hat{V}_i < y/T^{(n)}_{n-m:n}\}}} - \frac{1}{\hat{\eta}_n} \log s \tag{4.3}$$

must be examined in the case that the model assumptions are satisfied, i.e. the scaling law holds. Theorem 4.1.1 establishes the asymptotic behavior of the empirical process of this estimated deviation.

4.1.1 Theorem. Assume that Condition 2.2.1 holds with a function c that has first order partial derivatives c_x and c_y which satisfy (3.22). Suppose, as in Theorem 3.1.2, that m is an intermediate sequence such that $\sqrt{m}\, q_1\big(q^{\leftarrow}(m/n)\big) \to 0$ as $n \to \infty$. Then, there are versions of the estimated deviation (4.3) such that

$$\begin{aligned} &\sqrt{m} \left(\log \frac{\sum_{i=1}^{n} \mathbb{1}_{\{\hat{U}_i < sx/T^{(n)}_{n-m:n},\, \hat{V}_i < sy/T^{(n)}_{n-m:n}\}}}{\sum_{i=1}^{n} \mathbb{1}_{\{\hat{U}_i < x/T^{(n)}_{n-m:n},\, \hat{V}_i < y/T^{(n)}_{n-m:n}\}}} - \frac{1}{\hat{\eta}_n} \log s \right) \\ &\qquad = \frac{W(sx,sy)}{c(sx,sy)} - \frac{W(x,y)}{c(x,y)} + \frac{\log s}{\eta} \left(\int_0^1 t^{-(\eta+1)} W(t^\eta, t^\eta) \nu_\eta(dt) \right) + o_P(1) \end{aligned} \tag{4.4}$$

uniformly for (s, x, y) on compact subsets of $(0, 1] \times (0, \infty)^2$. △

4.1.2 Remark. Likewise one can establish an analogous result if another estimator for η based on a certain fraction of largest order statistics of $T_i^{(n)}$ is used instead of $\hat{\eta}_n^H$ or $\hat{\eta}_n^{ML}$. △

4.2 The Case of Asymptotic Independence

4.2.1 Corollary. Assume that the conditions of Theorem 4.1.1 hold for asymptotically independent random variables U and V. Then,

$$\sqrt{m}\left(\log \frac{\sum_{i=1}^{n} \mathbb{1}_{\{\hat{U}_i < sx/T^{(n)}_{n-m:n},\, \hat{V}_i < sy/T^{(n)}_{n-m:n}\}}}{\sum_{i=1}^{n} \mathbb{1}_{\{\hat{U}_i < x/T^{(n)}_{n-m:n},\, \hat{V}_i < y/T^{(n)}_{n-m:n}\}}} - \frac{1}{\hat{\eta}_n}\log s\right) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2_{x,y,s}\right) \tag{4.5}$$

for all $s, x, y \in (0, 1]$, where

$$\sigma^2_{x,y,s} :=$$

$$\begin{cases} \dfrac{s^{-1/\eta}-1}{c(x,y)} - \dfrac{\log^2 s}{\eta^2} & \text{if } \hat{\eta}_n = \hat{\eta}_n^H, \\[2ex] \dfrac{s^{-1/\eta}-1}{c(x,y)} - \dfrac{(\eta+1)^2\log^2 s}{\eta^4}\Bigg[1 + \dfrac{2(2\eta+1)(s-1)}{c(x,y)\log s} & \\ \qquad \times\left(\eta\displaystyle\int_0^1 t^{\eta-1}c(x\wedge t^\eta, y\wedge t^\eta)dt - c(x,y)\right)\Bigg] & \text{if } \hat{\eta}_n = \hat{\eta}_n^{ML}. \end{cases} \tag{4.6}$$

△

4.2.2 Remarks.

(i) If $\hat{\eta}_n = \hat{\eta}_n^{ML}$, one obtains a simpler expression for $\sigma^2_{x,y,s}$ in the special case $x = y$, since then $\eta\int_0^1 t^{\eta-1}c(x\wedge t^\eta, x\wedge t^\eta)dt - c(x,x) = -x^{(\eta+1)/\eta}/(\eta+1)$. Here, as expected when considering the results of Sections 3.1 and 3.3, we see that $\sigma^2_{x,y,s}$ is larger if $\hat{\eta}_n = \hat{\eta}_n^{ML}$ than if $\hat{\eta}_n = \hat{\eta}_n^H$.

(ii) For arbitrary $x, y > 0$, Corollary 4.2.1 holds with a more complicated asymptotic variance, see (4.31) and (4.36).

△

In order to apply this result in a model check, we discuss the appropriate choice of u, v (or x, y) and s in Section 4.4, where we also give the according confidence intervals for $(\log s)/\eta$.

4.3 The Case of Asymptotic Dependence

Henceforth, we abbreviate

$$c_x := c_x(1,1) \quad \text{and} \quad c_y := c_y(1,1).$$

Further, define functions $\chi_1 : \mathbb{R}_+^3 \to \mathbb{R}$ and $\chi_2 : (0,1]^2 \to \mathbb{R}$ by

$$\begin{aligned}
\chi_1(x,y,r) := {} & c(x,r) - c(r \vee x, r) + c(r,y) + r - c(r, r \vee y) \\
& -c(r \vee x, y) - c(x, r \vee y) + c(r \vee x, r \vee y) \\
& -lc_x\big[c(r,y) - c(r \vee x, y)\big] - lc_y\big[c(x,r) - c(x, r \vee y)\big] \\
& -lc_x(x,y)\big[c(x,r) + r - c(r \vee x, r) - c_x(r \wedge x) - lc_y c(x,r)\big] \\
& -lc_y(x,y)\big[c(r,y) + r - c(r, r \vee y) - c_y(r \wedge y) - lc_x c(r,y)\big] \text{ and}
\end{aligned}$$

$$\chi_2(x,y) := l\Big((lc_y - 1)c_x(x,y)c(x,1) + (lc_x - 1)c_y(x,y)c(1,y)\Big). \tag{4.7}$$

4.3.1 Corollary. Assume that the conditions of Theorem 4.1.1 hold for asymptotically dependent random variables U and V. Then,

$$\sqrt{m}\left(\log \frac{\sum_{i=1}^{n} \mathbb{1}_{\{\hat{U}_i < sx/T_{n-m:n}^{(n)},\, \hat{V}_i < sy/T_{n-m:n}^{(n)}\}}}{\sum_{i=1}^{n} \mathbb{1}_{\{\hat{U}_i < x/T_{n-m:n}^{(n)},\, \hat{V}_i < y/T_{n-m:n}^{(n)}\}}} - \frac{1}{\hat{\eta}_n}\log s\right) \xrightarrow{D} \mathcal{N}\left(0, \sigma_{x,y,s}^2\right) \tag{4.8}$$

for all $s, x, y \in (0,1]$, where

$$
\begin{aligned}
&\sigma^2_{x,y,s} \\
&= (\log^2 s)(1-2lc_xc_y)(1-l) \\
&\quad + \frac{2\log s}{c(x,y)}\Bigg(\int_1^{s^{-1}} \chi_2(x/r,y/r)\ dr - l(\log s)[xc_xc_x(x,y)+yc_yc_y(x,y)] \\
&\qquad\qquad + \chi_2(x,y) - s^{-1}\chi_2(sx,sy)\Bigg) \\
&\quad + \frac{1}{c^2(x,y)}\Bigg((s^{-1}-1)c(x,y) \\
&\qquad\qquad + lc_x(x,y)\big[(s^{-1}-1)xc_x(x,y)+2s^{-1}[c(sx,y)-c(x,y)]\big] \\
&\qquad\qquad + lc_y(x,y)\big[(s^{-1}-1)yc_y(x,y)+2s^{-1}[c(x,sy)-c(x,y)]\big] \\
&\qquad\qquad + 2l^2c_x(x,y)c_y(x,y)\big[(s^{-1}+1)c(x,y)-s^{-1}[c(sx,y)+c(x,sy)]\big]\Bigg)
\end{aligned} \tag{4.9}
$$

if $\hat{\eta}_n = \hat{\eta}_n^H$ and

$$
\begin{aligned}
&\sigma^2_{x,y,s} \\
&= 4(\log^2 s)(1-2lc_xc_y)(1-l) \\
&\quad + \frac{8\log s}{c(x,y)}\Bigg(\int_1^{s^{-1}} \chi_2(x/r,y/r)\ dr - l(\log s)[xc_xc_x(x,y)+yc_yc_y(x,y)] \\
&\qquad + 3\int_0^1 \chi_1(x,y,r)dr - 3s\int_0^{s^{-1}} \chi_1(x,y,r)dr + \frac{s^{-1}\chi_2(sx,sy)-\chi_2(x,y)}{2}\Bigg) \\
&\quad + \frac{1}{c^2(x,y)}\Bigg((s^{-1}-1)c(x,y) \\
&\qquad\qquad + lc_x(x,y)\big[(s^{-1}-1)xc_x(x,y)+2s^{-1}[c(sx,y)-c(x,y)]\big] \\
&\qquad\qquad + lc_y(x,y)\big[(s^{-1}-1)yc_y(x,y)+2s^{-1}[c(x,sy)-c(x,y)]\big] \\
&\qquad\qquad + 2l^2c_x(x,y)c_y(x,y)\big[(s^{-1}+1)c(x,y)-s^{-1}[c(sx,y)+c(x,sy)]\big]\Bigg)
\end{aligned} \tag{4.10}
$$

if $\hat{\eta}_n = \hat{\eta}_n^{ML}$. $\triangle$

4.3.2 Remark. For arbitrary $x, y > 0$, Corollary 4.3.1 holds with a more complicated asymptotic variance, see (4.48) and (4.59). $\triangle$

4.4 Construction of Tests

In order to apply the Corollaries 4.2.1 and 4.3.1 in a model check, one has to choose appropriate values for u and v (or x and y) and for s. Here we propose to imitate the approach of pp- and qq-plots in that we consider (4.2) only for points (su, sv) equal to a (pseudo-)observation. More precisely, we consider those points $(s_j u_j, s_j v_j) = (\hat{U}_j, \hat{V}_j)$ which belong to the open square $\big(0, 1/T^{(n)}_{n-m:n}\big)^2$. Recall that (if no ties occur[7]) these pseudo-observations correspond to the order statistics $T^{(n)}_{n-i+1:n}$, $i = 1, \dots, m$, that are used for the estimation of η. Here, for simplicity, we assume that the standardized observations $(\hat{U}_j, \hat{V}_j)$ have been re-indexed such that $(\hat{U}_j, \hat{V}_j) \in \big(0, 1/T^{(n)}_{n-m:n}\big)^2$ for $j = 1, \dots, m$.

Moreover, we take (u_j, v_j) to be the projection of $(\hat{U}_j, \hat{V}_j)$ onto the (upper or right) boundary of the square $\big(0, 1/T^{(n)}_{n-m:n}\big]^2$, which results in the choice

$$s_j := \frac{\max(\hat{U}_j, \hat{V}_j)}{1/T^{(n)}_{n-m:n}}, \qquad (u_j, v_j) := \frac{(\hat{U}_j, \hat{V}_j)}{s_j} = \frac{1}{T^{(n)}_{n-m:n}} \cdot \frac{(\hat{U}_j, \hat{V}_j)}{\max(\hat{U}_j, \hat{V}_j)};$$

see Figure 4.1 below. This way, we ensure that s_j is indeed a contraction factor (i.e., less than 1) as it was assumed above, and that all reference points (u_j, v_j) used in the graphical check of the scaling law can be parameterized by a single real parameter

$$z_j := \begin{cases} u_j & \in \big(0, 1/T^{(n)}_{n-m:n}\big], \qquad \text{if } u_j \le v_j = 1/T^{(n)}_{n-m:n} \\ \big(2/T^{(n)}_{n-m:n} - v_j\big) & \in \big(1/T^{(n)}_{n-m:n}, 2/T^{(n)}_{n-m:n}\big] \quad \text{if } v_j < u_j = 1/T^{(n)}_{n-m:n}. \end{cases}$$

(The parameter z_j equals the distance between the points $(0, 1/T^{(n)}_{n-m:n})$ and (u_j, v_j) measured along the boundary of the square $(0, 1/T^{(n)}_{n-m:n}]^2$.) Note that if we plot (4.2) for all points (u_j, v_j) and contraction factors s_j, $j = 1, \dots, m$,

[7] See Footnote 1. If ties occur in applications, we simply exclude all $(\hat{U}_i, \hat{V}_i)$ with $\max(\hat{U}_i, \hat{V}_i) = 1/T^{(n)}_{n-m:n}$ (and reduce the number of points (4.2) is plotted for).

we lose the information which point of the plot corresponds to which pseudo-observation. To avoid that, we instead use the three-dimensional plot

$$\left(z_j, \log s_j, \log \frac{\sum_{i=1}^{n} \mathbb{1}_{\{\hat{U}_i < s_j u_j, \hat{V}_i < s_j v_j\}}}{\sum_{i=1}^{n} \mathbb{1}_{\{\hat{U}_i < u_j, \hat{V}_i < v_j\}}} \right)_{j=1,\ldots,m}, \tag{4.11}$$

where the additional argument z_j determines the reference point (u_j, v_j). Figure 4.1 illustrates our construction for an example that will be examined in detail in Chapter 5. The points of the plot (4.11) should then approximately lie on the reference plane $(z, u) \mapsto (z, u, u/\eta)$ if the scaling law holds. The following statistical tests formalize this requirement, where we consider the case that U and V are asymptotically independent and that $\hat{\eta}_n = \hat{\eta}_n^H$. The other cases can be treated analogously.

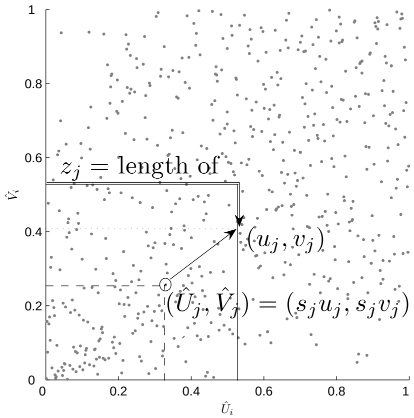


Figure 4.1: The square with side length $1/T_{n-m:n}$ contains the m points used for the estimation of η. The scaling law (2.17) is checked within this square.

Let

$$\Delta_j := \log \frac{\sum_{i=1}^{n} \mathbb{1}_{\{\hat{U}_i < s_j u_j, \hat{V}_i < s_j v_j\}}}{\sum_{i=1}^{n} \mathbb{1}_{\{\hat{U}_i < u_j, \hat{V}_i < v_j\}}} - \frac{1}{\hat{\eta}_n^H} \log s_j \tag{4.12}$$

denote the estimated difference between the third coordinate of the j^{th} point of (4.11) and the third coordinate of the corresponding point (with the same first and second coordinates) on the reference plane. Then, according to Corollary 4.2.1, $\sqrt{m}\Delta_j/\hat{\sigma}_j$ is approximately $\mathcal{N}(0,1)$-distributed, where

$$\hat{\sigma}_j^2 := \frac{s_j^{-1/\hat{\eta}_n^H} - 1}{\hat{c}_n(u_j, v_j)} - \frac{\log^2 s_j}{\left(\hat{\eta}_n^H\right)^2}.$$

Hence, we reject the scaling law on the approximate confidence level $1 - \alpha$, if

$$\left| \frac{\sqrt{m}\Delta_j}{\hat{\sigma}_j} \right| > \Phi^{\leftarrow}(1 - \alpha/2) \tag{4.13}$$

or, equivalently, if the last coordinate of the j^{th} point of the plot (4.11) does not belong to the confidence interval

$$\left[\frac{\log s_j}{\hat{\eta}_n^H} - m^{-1/2}\hat{\sigma}_j \; \Phi^{\leftarrow}(1 - \alpha/2), \frac{\log s_j}{\hat{\eta}_n^H} + m^{-1/2}\hat{\sigma}_j \; \Phi^{\leftarrow}(1 - \alpha/2) \right]. \tag{4.14}$$

To get an overall picture from all m tests, a percentage of points whose last component does not belong to the pertaining confidence interval much greater than α can be interpreted as an indicator that the scaling law is not fulfilled in the square $\left(0, 1/T_{n-m:n}^{(n)}\right]^2$.

One can also incorporate the information provided by the statistical test into the plot (4.11) by indicating for each point either (i) whether its last coordinate belongs to the corresponding confidence interval or (ii) the $(1-p)$-value

$$2 \left| \Phi\left(\frac{\sqrt{m}\Delta_j}{\hat{\sigma}_j} \right) - 0.5 \right| \tag{4.15}$$

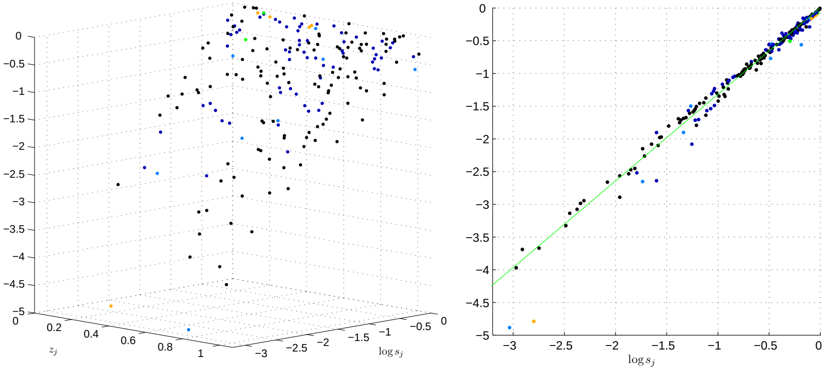


Figure 4.2: Plot (4.11) with $m = 200$ for the Danish fire insurance (see Section 5.1) from two perspectives. The right hand perspective also displays the reference plane. Here, only 5 (orange, cf. Figure 5.7) of 200 points lie outside the approximate 95 % confidence interval (4.14).

of the pertaining test. We give an example in Figure 4.2, which implements (ii) and shows plot (4.11) from two perspectives for data that will be examined in detail in Chapter 5.

We now comment on two issues of our method that were omitted so far and thereby motivate the investigations of Section 4.5. First, the quantities s_j and $(\hat{U}_j, \hat{V}_j)/\max(\hat{U}_j, \hat{V}_j)$, which play the roles of s and (x, y) of the Corollaries 4.2.1 and 4.3.1, are, in contrast to s and (x, y), random variables in our tests. Second, since the tests for each of the observations are (partly) based on the same data, they are not independent. Hence, our method is a heuristic based on the Corollaries 4.2.1 and 4.3.1 rather than a direct application of these results. The statement that $\sqrt{m}\Delta_j/\hat{\sigma}_j$ is approximately $\mathcal{N}(0, 1)$-distributed is not only to be understood in the sense of using an asymptotic equality as an approximation, but also under consideration of these issues. The second issue can be resolved if one considers a test for the entirety of the pseudo-observations rather than m tests for each individual of them. We now restate some results from the proof of Theorem 4.1.1, before we establish a corollary of that theorem which is useful in order to develop such a uniform test. From the proof of Theorem 4.1.1 we know that for the versions of Q_n^*, S_n^*, W and $\hat\eta_n$ as considered there and for s, x, y such that $\hat U_i$ and $\hat V_i$ exist with $\hat U_i < sx/T_{n-m:n}^{(n)}$ and $\hat V_i < sy/T_{n-m:n}^{(n)}$, cf. (4.28),

$$\log \underbrace{\frac{S_n^*\left(\frac{(n+1)sx}{nq^{\leftarrow}(m/n)Q_n^*(1)}-,\ \frac{(n+1)sy}{nq^{\leftarrow}(m/n)Q_n^*(1)}-\right)}{S_n^*\left(\frac{(n+1)x}{nq^{\leftarrow}(m/n)Q_n^*(1)}-,\ \frac{(n+1)y}{nq^{\leftarrow}(m/n)Q_n^*(1)}-\right)}}_{=:\tilde S_n^*(x,y)} - \frac{1}{\hat\eta_n}\log s$$

$$= m^{-1/2}\left[\frac{W(sx,sy)}{c(sx,sy)} - \frac{W(x,y)}{c(x,y)} + \frac{\log s}{\eta}\int_0^1 t^{-(\eta+1)}W(t^\eta,t^\eta)\nu_\eta(dt)\right] + o_P(m^{-1/2}) \tag{4.16}$$

uniformly for all (s,x,y) on compact subsets of $(0,1]\times(0,\infty)^2$, where (cf. (3.54) and (4.27))

$$\left(\frac{\tilde S_n^*(sx,sy)}{\tilde S_n^*(x,y)}\right)_{s,x,y\in(0,1]\times(0,\infty)^2}$$

$$\stackrel{D}{=} \left(\frac{S\left(U_{\lceil sx(n+1)/T_{n-m:n}^{(n)}\rceil-1:n},\ V_{\lceil sy(n+1)/T_{n-m:n}^{(n)}\rceil-1:n}\right)}{S\left(U_{\lceil x(n+1)/T_{n-m:n}^{(n)}\rceil-1:n},\ V_{\lceil y(n+1)/T_{n-m:n}^{(n)}\rceil-1:n}\right)}\right)_{s,x,y\in(0,1]\times(0,\infty)^2} \tag{4.17}$$

and (cf. (3.52) and (4.26))

$$\frac{S\left(U_{\lceil sx(n+1)/T_{n-m:n}^{(n)}\rceil-1:n},\ V_{\lceil sy(n+1)/T_{n-m:n}^{(n)}\rceil-1:n}\right)}{S\left(U_{\lceil x(n+1)/T_{n-m:n}^{(n)}\rceil-1:n},\ V_{\lceil y(n+1)/T_{n-m:n}^{(n)}\rceil-1:n}\right)} = \frac{\sum_{i=1}^n \mathbb{1}_{\{\hat U_i < sx/T_{n-m:n}^{(n)},\ \hat V_i < sy/T_{n-m:n}^{(n)}\}}}{\sum_{i=1}^n \mathbb{1}_{\{\hat U_i < x/T_{n-m:n}^{(n)},\ \hat V_i < y/T_{n-m:n}^{(n)}\}}}$$

for all $x, y \in (0, \xi_0]$ with probability tending to 1 for all $\xi_0 > 0$. Denote by

$$L(s,x,y) := \frac{W(sx,sy)}{c(sx,sy)} - \frac{W(x,y)}{c(x,y)} + \frac{\log s}{\eta}\left(\int_0^1 t^{-(\eta+1)} W(t^\eta, t^\eta)\nu_\eta(dt)\right) \tag{4.18}$$

the limiting process of $\sqrt{m}\big(\log\big(\tilde{S}_n^*(sx,sy)/\tilde{S}_n^*(x,y)\big) - (\log s)/\hat{\eta}_n\big)$.

Since $\log\big(\tilde{S}_n^*(sx,sy)/\tilde{S}_n^*(x,y)\big)$ equals $-\infty$ for sufficiently small $s > 0$, $\sqrt{m}\big(\log\big(\tilde{S}_n^*(sx,sy)/\tilde{S}_n^*(x,y)\big) - (\log s)/\hat{\eta}_n\big)$ cannot converge uniformly to L for all $(s,x,y) \in (0,1] \times (0,\infty)^2$. However, we immediately obtain the following corollary from Theorem 4.1.1.

4.4.1 Corollary. Under the conditions of Theorem 4.1.1,

$$\lim_{n\to\infty} P\left\{\sup_{(s,x,y)\in K} g_1(s,x,y)\left|m^{1/2}\left(\log\frac{\tilde{S}_n^*(sx,sy_0)}{\tilde{S}_n^*(x,y_0)} - \frac{\log s}{\hat{\eta}_n}\right) - L(s,x,y)\right| > \varepsilon\right\} = 0 \tag{4.19}$$

for all compact subsets K of $(0,1] \times (0,\infty)^2$ and all bounded functions g_1. $\triangle$

According to Corollary 4.4.1, a uniform test might be constructed as follows. Denote by $\bar{L} := \sup_{(s,x,y)\in K} |g_1(s,x,y)L(s,x,y)|$ the supremum of the weighted limiting process L over some compact subset K of $(0,1] \times (0,\infty)^2$, where g_1 is some appropriate weight function. Then, $\sup_j |g_1(s_j, \tilde{u}_j, \tilde{v}_j)\sqrt{m}\Delta_j|$ with $\tilde{u}_j = u_j T^{(n)}_{n-m:n}$, $\tilde{v}_j = v_j T^{(n)}_{n-m:n}$ and j such that $(s_j, \tilde{u}_j, \tilde{v}_j) \in K$ is approximately distributed as $\bar{L}$. Hence, we reject the scaling law on the approximate confidence level $1-\alpha$, if $\sup_j |g_1(s_j, \tilde{u}_j, \tilde{v}_j)\sqrt{m}\Delta_j|$ is larger than the $(1-\alpha/2)$-quantile of the d.f. of $\bar{L}$, which we calculate by simulation of L. However, the choice of K, i.e. in particular "how far" s, x and y must be bounded away from 0, is rather arbitrary. Hence, in order to construct a uniform test, we would like to generalize (4.19) to hold without the restriction that s, x and y must be bounded away from 0. We address this issue and discuss the necessity to include a weight function like g_1 (which here might be chosen as $g_1 \equiv 1$) in the following Section 4.5.

In Section 2.2.2 we have indicated that Condition 2.2.1 implies that the scaling law holds not only for rectangular sets, but also for more general sets, cf. (2.19). We conclude this section with the remark that an appropriate generalization of the results of this chapter may yield a similar tool to validate the scaling law for these more general sets.

4.5 Further Uniform Results

In the previous sections, we have proven the asymptotic normality of differences between empirical probabilities and model based estimates of the same probabilities, cf. Corollaries 4.2.1 and 4.3.1. This asymptotic normality allows the construction of a statistical test for the scaling law for each pseudo-observation $(\hat{U}_i, \hat{V}_i)$, see Section 4.4. If the percentage of the tests that fail is much greater than the nominal size, this is interpreted as an indicator that the scaling law is not fulfilled. Note, however, that the statistical test is constructed for a single observation rather than for the entirety of the observations and that the tests for each of the observations are not independent, since they are (partly) based upon the same data. In Section 4.4 we have also indicated that a uniform test based on Theorem 4.1.1 is not entirely satisfying, since due to the assumptions of that theorem, s, x and y must be bounded away from 0. In this section, we search for a possibility to validate the scaling law on a "global" level, i.e. we aim to develop a statistical test for the entirety of the pseudo-observations rather than for each individual of them. Therefore, a uniform version of the test constructed in Section 4.4 without the restriction that s, x and y must be bounded away from 0 is desired.

Since we desire a uniform version of the method of validating the scaling law as constructed in Section 4.4, we still use the pseudo-observations within the square $\big(0, 1/T^{(n)}_{n-m:n}\big]^2$. Recall from Section 4.4 that for the construction of the tests, we take (u_j, v_j) in the definition (4.12) of Δ_j to be the projection onto the (upper or right) boundary of this square, see also Figure 4.1. Hence, we divide this square by the line $y = x$, start with considering the upper of the two triangles and conclude for the lower. According to this construction, it is then sufficient to fix $y = y_0 \leq 1$ and to assume that $x \leq y_0$. Further, for the reasons given before Corollary 4.4.1, i.e. since individually every summand of $\log\big(\tilde{S}^*_n(sx, sy)/\tilde{S}^*_n(x, y)\big) - (\log s)/\hat{\eta}_n$ and L diverges for $s \downarrow 0$ or $x \downarrow 0$, we cannot expect that (4.19) holds for all bounded functions g_1 when giving up that s and x must be bounded away from 0. Thus, g_1 must converge "appropriately" to 0 as $s \downarrow 0$ or $x \downarrow 0$. Considering the definition (4.18) of L and recalling from (3.9) that $W = W_1$ is a Gaussian process with variance

$EW^2(x,y) = c(x,y)$ if U and V are asymptotically independent, $g_1(s,x,y_0) = \big(c(sx,sy_0)\big)^{1/2+\delta}$ with $0<\delta<1/2$ might be a reasonable choice. Hence, in order to construct such a uniform test, we would ideally like to prove that for all $t_0>0$ and all $\varepsilon>0$,

$$\lim_{n\to\infty} P\Bigg\{ \sup_{\substack{0<s\wedge x\le t_0\\ 0<s\vee x\le y_0}} \big(c(sx,sy_0)\big)^{1/2+\delta} 1\!\!1_{(0,\infty)}(\tilde S_n^*(sx,sy_0)) \left| m^{1/2}\left(\log\frac{\tilde S_n^*(sx,sy_0)}{\tilde S_n^*(x,y_0)} - \frac{\log s}{\hat\eta_n}\right) - L(s,x,y_0)\right| > \varepsilon \Bigg\} = 0. \tag{4.20}$$

Recall that Corollary 4.4.1 yields (4.20) if s and x are bounded away from 0, so that we obtain (4.20) without this restriction if

$$\lim_{\vartheta\downarrow 0} P\left\{ \sup_{\substack{0<s\wedge x\le\vartheta\\ 0<s\vee x\le y_0}} \big(c(sx,sy_0)\big)^{1/2+\delta} \big|L(s,x,y_0)\big| > \varepsilon\right\} = 0 \tag{4.21}$$

and

$$\lim_{\vartheta\downarrow 0}\limsup_{n\to\infty} P\Bigg\{ \sup_{\substack{0<s\wedge x\le\vartheta\\ 0<s\vee x\le y_0}} m^{1/2}\big(c(sx,sy_0)\big)^{1/2+\delta} 1\!\!1_{(0,\infty)}(\tilde S_n^*(sx,sy_0)) \left|\log\frac{\tilde S_n^*(sx,sy_0)}{\tilde S_n^*(x,y_0)} - \frac{\log s}{\hat\eta_n}\right| > \varepsilon\Bigg\} = 0 \tag{4.22}$$

jointly hold. We establish (4.21) if $W = W_1$ is centered Gaussian with covariance structure (3.9), i.e. if U and V are asymptotically independent, in Section 4.5.1 and discuss (4.22) in Section 4.5.2.

4.5.1 Uniform Convergence of the Weighted Limiting Process

4.5.1 Theorem. Assume that U and V are asymptotically independent and Condition 2.2.1 holds. Then, for the centered Gaussian process $W = W_1$ with covariance structure (3.9) and all $0 < y_0 \le 1$, $0 < \delta < 1/2$ and $\varepsilon > 0$,

$$\lim_{\vartheta \downarrow 0} P\left\{ \sup_{\substack{0 < s \wedge x \le \vartheta \\ 0 < s \vee x \le y_0}} \bigl(c(sx, sy_0)\bigr)^{1/2+\delta} \left| \frac{W(sx, sy_0)}{c(sx, sy_0)} \right| > \varepsilon \right\} = 0. \tag{4.23}$$

△

4.5.2 Remark. Observe by the definition (4.18) of L that (4.23) implies (4.21). △

Recall that we considered the upper triangle of the square $\bigl(0, 1/T_{n-m:n}^{(n)}\bigr]^2$, which enabled us to fix $y = y_0$. When considering the lower triangle of $\bigl(0, 1/T_{n-m:n}^{(n)}\bigr]^2$ we may fix $x = x_0$ instead of y. The analogous result of Theorem 4.5.1 holds.

4.5.3 Theorem. Assume that U and V are asymptotically independent and Condition 2.2.1 holds. Then, for the centered Gaussian process $W = W_1$ with covariance structure (3.9) and all $0 < x_0 \le 1$, $0 < \delta < 1/2$ and $\varepsilon > 0$,

$$\lim_{\vartheta \downarrow 0} P\left\{ \sup_{\substack{0 < s \wedge y \le \vartheta \\ 0 < s \vee y \le x_0}} \bigl(c(sx_0, sy)\bigr)^{1/2+\delta} \left| \frac{W(sx_0, sy)}{c(sx_0, sy)} \right| > \varepsilon \right\} = 0.$$

△

4.5.2 The Empirical Process

In view of Corollary 4.4.1 and after proving (4.21), it is plausible that one may also obtain (4.22) (and hence (4.20)). In this section we show that (4.22) holds with $(c(sx, sy_0))^{1/2+\delta}$ replaced with a smaller weight function. (In order to avoid redundancy, we restrict ourselves to fixed $y = y_0$ in this section. Similar as in the previous section, the analogous results with fixed $x = x_0$ instead of y also hold.) Since the new weight function converges faster to 0 as $s \downarrow 0$ or $x \downarrow 0$ than $(c(sx, sy_0))^{1/2+\delta}$, this is a weaker result than (4.22). In addition, we require a stronger uniformity condition than assumed previously.

First, however, the following lemma shows that no additional conditions are necessary to substitute the term $(\log s)/\hat{\eta}_n$ on the left hand side of (4.22) with a handier expression.

4.5.4 Lemma. Assume that the conditions of Theorem 3.1.2 hold. Then, for all $0 < y_0 \leq 1$, $0 < \delta < 1/2$ and $\varepsilon > 0$,

$$\lim_{\vartheta \downarrow 0} \lim_{n \to \infty} P\left\{ \sup_{\substack{0 < s \wedge x \leq \vartheta \\ 0 < s \vee x \leq y_0}} m^{1/2} (c(sx, sy_0))^{1/2+\delta} \left| \frac{\log s}{\hat{\eta}_n} - \frac{\log s}{\eta} \right| > \varepsilon \right\} = 0.$$

$\triangle$

After this lemma, we would like to prove a similar result for the term $\log\big(\tilde{S}_n^*(sx, sy_0)/\tilde{S}_n^*(x, y_0)\big)$ on the left hand side of (4.22). We obtain the following theorem, where, as mentioned above, we assume a stronger uniformity condition and a smaller weight function than previously. We briefly comment on these conditions in Remarks 4.5.6.

4.5.5 Theorem. Assume that the conditions of Theorem 3.1.2 hold with Condition 2.2.1 satisfied uniformly for all $x, y \in [0, 1]$. Further, suppose $\log^6 n = o(m)$ and that for all $y_0 > 0$ there exist some constant $C \geq 0$ such that

$$\limsup_{t \downarrow 0} \sup_{0 < x, y \leq y_0} \frac{q(t) c(x, y)}{Q(tx, ty)} = C \tag{4.24}$$

and define $\ell_2(t) := \log(\max(2, |\log t|))$. Then, for all $\varepsilon, \delta > 0$,

$$\begin{aligned} \lim_{\vartheta \downarrow 0} \limsup_{n \to \infty} P\Bigg\{ \sup_{\substack{0 < s \wedge x \leq \vartheta \\ 0 < s \vee x \leq y_0}} & m^{1/2} \min\big(c(sx, sy_0), (sx \wedge sy_0)^{1/2} (\ell_2(sx \wedge sy_0))^{-1/2-\delta}\big) \\ & \times \mathbb{1}_{(0,\infty)}(\tilde{S}_n^*(sx, sy_0)) \left| \log \frac{\tilde{S}_n^*(sx, sy_0)}{\tilde{S}_n^*(x, y_0)} - \frac{\log s}{\hat{\eta}_n} \right| > \varepsilon \Bigg\} = 0, \end{aligned} \tag{4.25}$$

i.e. Equation (4.22) holds with weight function $(c(sx, sy_0))^{1/2+\delta}$ replaced with $\min\big(c(sx, sy_0), (sx \wedge sy_0)^{1/2} (\ell_2(sx \wedge sy_0))^{-1/2-\delta}\big)$. $\triangle$

4.5.6 Remarks.

(i) In view of the similarity to our central Condition 2.2.1, assumption (4.24) does not seem unnatural. However, neither Condition 2.2.1 implies (4.24), nor it is obvious that (4.24) holds for a wide range of bivariate distributions.

(ii) For large parts of the proof of Theorem 4.5.5 the weight function $c(sx, sy_0)$ suffices. However, in general one may expect (4.22) to hold with weight function $c(sx, sy_0)$ only if one assumes that $m^{1/2}c(sx, sy_0) \to 0$ as $n \to \infty$, see Remark 4.6.2 in the proof of Theorem 4.5.5.

$\triangle$

4.6 Proofs

4.6.1 Proof of Theorem 4.1.1

First, we argue as in the proof of Theorem 3.2.1 to see that

$$\log \frac{\sum_{i=1}^{n} \mathbb{1}\{\hat{U}_i < sx/T^{(n)}_{n-m:n}, \hat{V}_i < sy/T^{(n)}_{n-m:n}\}}{\sum_{i=1}^{n} \mathbb{1}\{\hat{U}_i < x/T^{(n)}_{n-m:n}, \hat{V}_i < y/T^{(n)}_{n-m:n}\}}$$

$$= \log \frac{S\left(U_{\lceil sx(n+1)/T^{(n)}_{n-m:n}\rceil -1:n}, V_{\lceil sy(n+1)/T^{(n)}_{n-m:n}\rceil -1:n}\right)}{S\left(U_{\lceil x(n+1)/T^{(n)}_{n-m:n}\rceil -1:n}, V_{\lceil y(n+1)/T^{(n)}_{n-m:n}\rceil -1:n}\right)} \tag{4.26}$$

for all $x, y \in (0, \xi_0]$ with probability tending to 1 as $n \to \infty$ for all $\xi_0 > 0$.

Observe that the denominator of the right hand side of (4.26) equals the numerator of the right hand side of (3.52), and that in the numerator of the right hand side of (4.26) we merely have factor s in addition. Hence, it is along the same lines as we obtained (3.57) that we can conclude that a version of the right hand side of (4.26) is

$$
\begin{aligned}
&\log \frac{S_n^*\left(\frac{(n+1)sx}{nq^{\leftarrow}(m/n)Q_n^*(1)}-, \frac{(n+1)sy}{nq^{\leftarrow}(m/n)Q_n^*(1)}-\right)}{S_n^*\left(\frac{(n+1)x}{nq^{\leftarrow}(m/n)Q_n^*(1)}-, \frac{(n+1)y}{nq^{\leftarrow}(m/n)Q_n^*(1)}-\right)} \\
&= \log \frac{c(sx,sy) + m^{-1/2}\big(W(sx,sy) - W(1,1)c(sx,sy)\big) + o(m^{-1/2})}{c(x,y) + m^{-1/2}\big(W(x,y) - W(1,1)c(x,y)\big) + o(m^{-1/2})} \\
&= \frac{1}{\eta}\log s + \log\left(1 + m^{-1/2}\left(\frac{W(sx,sy)}{c(sx,sy)} - \frac{W(x,y)}{c(x,y)}\right) + o(m^{-1/2})\right) \\
&= \frac{1}{\eta}\log s + m^{-1/2}\left(\frac{W(sx,sy)}{c(sx,sy)} - \frac{W(x,y)}{c(x,y)}\right) + o(m^{-1/2}) \text{ a.s.} \qquad (4.27)
\end{aligned}
$$

uniformly for (s,x,y) on compact subsets of $(0,1] \times (0,\infty)^2$ for the versions Q_n^*, S_n^* and W as considered in the proof of Theorem 3.2.7.

The next step is to substitute $1/\eta$ by (versions of) $1/\hat{\eta}_n$ in (4.27). To this end, note from the proof of Theorem 3.2.7 and from Remarks 3.1.8 that the versions Q_n^*, S_n^* and W of (4.27) are exactly those of Corollary 3.1.7.

Thus, by (4.27) and Corollary 3.1.7 we obtain

$$
\begin{aligned}
&\log \frac{S_n^*\left(\frac{(n+1)sx}{nq^{\leftarrow}(m/n)Q_n^*(1)}-, \frac{(n+1)sy}{nq^{\leftarrow}(m/n)Q_n^*(1)}-\right)}{S_n^*\left(\frac{(n+1)x}{nq^{\leftarrow}(m/n)Q_n^*(1)}-, \frac{(n+1)y}{nq^{\leftarrow}(m/n)Q_n^*(1)}-\right)} - \frac{1}{\hat{\eta}_n}\log s \qquad (4.28) \\
&= m^{-1/2}\left[\frac{W(sx,sy)}{c(sx,sy)} - \frac{W(x,y)}{c(x,y)} + \frac{\log s}{\eta}\left(\int_0^1 t^{-(\eta+1)} W(t^\eta, t^\eta)\nu_\eta(dt)\right)\right] \\
&\qquad + o_P(m^{-1/2})
\end{aligned}
$$

uniformly for (s,x,y) on compact subsets of $(0,1] \times (0,\infty)^2$.

4.6.2 Proof of Corollary 4.2.1

We have to establish the distribution of the right hand side of (4.28) for $s,x,y \in (0,1]$. We start with the Hill case.

The Case $\hat{\eta}_n = \hat{\eta}_n^H$

Here, according to (3.14), $\nu_\eta(dt) = \nu_\eta^H(dt) = t^\eta dt - \varepsilon_1(dt)$. We abbreviate

$$Z^H(t) := \frac{\log s}{\eta}\big(t^{-1}W(t^\eta, t^\eta) - W(1,1)\big) + \frac{W(sx, sy)}{c(sx, sy)} - \frac{W(x,y)}{c(x,y)},$$

so that it remains to show that $\int_0^1 Z^H(t)dt$ is $\mathcal{N}\big(0, \sigma^2_{x,y,s}\big)$-distributed.

A special case of Proposition 2.2.1 of Shorack and Wellner (1986) states that

$$\int_0^1 Z^H(t)dt \stackrel{D}{=} \mathcal{N}\big(0, \sigma^2_{\int Z^H}\big), \tag{4.29}$$

where $\stackrel{D}{=}$ stands for equality in distribution and

$$\sigma^2_{\int Z^H} = \int_0^1 \int_0^1 \operatorname{cov}\big[Z^H(r), Z^H(t)\big]\, dr dt. \tag{4.30}$$

By means of the covariance structure (3.9) of $W = W_1$, we obtain

$$\begin{aligned}
&\operatorname{cov}\big[Z^H(r), Z^H(t)\big] = E\big[Z^H(r)Z^H(t)\big] \\
&= \frac{\log^2 s}{\eta^2}\big[(rt)^{-1}c(r^\eta \wedge t^\eta, r^\eta \wedge t^\eta) \\
&\qquad -r^{-1}c(r^\eta \wedge 1, r^\eta \wedge 1) - t^{-1}c(t^\eta \wedge 1, t^\eta \wedge 1) + 1\big] \\
&\quad + \frac{\log s}{\eta c(x,y)}\Big[r^{-1}\Big(s^{-1/\eta}c(sx \wedge r^\eta, sy \wedge r^\eta) - c(x \wedge r^\eta, y \wedge r^\eta)\Big) \\
&\qquad -2s^{-1/\eta}c(sx \wedge 1, sy \wedge 1) + 2c(x \wedge 1, y \wedge 1) \\
&\qquad +t^{-1}\Big(s^{-1/\eta}c(sx \wedge t^\eta, sy \wedge t^\eta) - c(x \wedge t^\eta, y \wedge t^\eta)\Big)\Big] \\
&\quad + \frac{s^{-2/\eta}c(sx, sy) - 2s^{-1/\eta}c(sx, sy) + c(x,y)}{c^2(x,y)}.
\end{aligned}$$

To calculate $\sigma^2_{\int Z^H}$, we consider the double integrals of these three summands separately, use the homogeneity (2.14) of c in the third summand and define

$$I := \frac{\log^2 s}{\eta^2} \int_0^1 \int_0^1 \big[(rt)^{-1} c(r^\eta \wedge t^\eta, r^\eta \wedge t^\eta) \\ -r^{-1} c(r^\eta \wedge 1, r^\eta \wedge 1) - t^{-1} c(t^\eta \wedge 1, t^\eta \wedge 1) + 1\big] drdt,$$

$$J(x,y) := \frac{\log s}{\eta c(x,y)} \int_0^1 \int_0^1 \Big[r^{-1}\big(s^{-1/\eta} c(sx \wedge r^\eta, sy \wedge r^\eta) - c(x \wedge r^\eta, y \wedge r^\eta)\big) \\ -2s^{-1/\eta} c(sx \wedge 1, sy \wedge 1) + 2c(x \wedge 1, y \wedge 1) \\ +t^{-1}\Big(s^{-1/\eta} c(sx \wedge t^\eta, sy \wedge t^\eta) - c(x \wedge t^\eta, y \wedge t^\eta)\Big)\Big] drdt,$$

$$K(x,y) := \int_0^1 \int_0^1 \frac{s^{-1/\eta} - 1}{c(x,y)} drdt = \frac{s^{-1/\eta} - 1}{c(x,y)},$$

which means that

$$\sigma^2_{\int Z^H} = I + J(x,y) + K(x,y).$$

The integrals I and J simplify to

$$\begin{aligned} I &= \frac{2\log^2 s}{\eta^2} \int_0^1 \int_0^t \big[(rt)^{-1} c(r^\eta, r^\eta) \\ &\qquad -r^{-1} c(r^\eta, r^\eta) - t^{-1} c(t^\eta, t^\eta) + 1\big] drdt \\ &= \frac{2\log^2 s}{\eta^2} \int_0^1 \int_0^t (t^{-1} - 1 - 1 + 1) drdt \\ &= \frac{2\log^2 s}{\eta^2} \int_0^1 (1-t) dt = \frac{\log^2 s}{\eta^2} \end{aligned}$$

and

$$J(x,y) = \frac{2\log s}{\eta c(x,y)} \int_0^1 \Big[t^{-1}\Big(s^{-1/\eta} c(sx \wedge t^\eta, sy \wedge t^\eta) - c(x \wedge t^\eta, y \wedge t^\eta)\Big) \\ -s^{-1/\eta} c(sx \wedge 1, sy \wedge 1) + c(x \wedge 1, y \wedge 1)\Big] dt.$$

Then, substituting $t = s^{1/\eta}u$ in the first summand yields

$$
\begin{aligned}
&J(x,y)\\
&= \frac{2\log s}{\eta c(x,y)}\bigg(\int_0^{s^{-1/\eta}} s^{-1/\eta}u^{-1}c(x\wedge u^\eta, y\wedge u^\eta)s^{1/\eta}du - \int_0^1 t^{-1}c(x\wedge t^\eta, y\wedge t^\eta)dt\\
&\qquad -s^{-1/\eta}c(sx\wedge 1, sy\wedge 1) + c(x\wedge 1, y\wedge 1)\bigg)\\
&= \frac{2\log s}{\eta c(x,y)}\bigg(\int_1^{s^{-1/\eta}} t^{-1}c(x\wedge t^\eta, y\wedge t^\eta)dt\\
&\qquad -s^{-1/\eta}c(sx\wedge 1, sy\wedge 1) + c(x\wedge 1, y\wedge 1)\bigg)
\end{aligned}
$$

and thus

$$
\begin{aligned}
\sigma^2_{\int Z} = \frac{\log^2 s}{\eta^2} + \frac{s^{-1/\eta}-1}{c(x,y)} + \frac{2\log s}{\eta c(x,y)}\bigg(\int_1^{s^{-1/\eta}} t^{-1}c(x\wedge t^\eta, y\wedge t^\eta)dt\\
-s^{-1/\eta}c(sx\wedge 1, sy\wedge 1) + c(x\wedge 1, y\wedge 1)\bigg),
\end{aligned}
\tag{4.31}
$$

which is the asymptotic variance mentioned in Remark 4.2.2 (ii). The assumption $x, y \in (0,1]$ together with the homogeneity (2.14) of c yields

$$
J(x,y) = \frac{2\log s}{\eta}\int_1^{s^{-1/\eta}} t^{-1}dt = -\frac{2\log^2 s}{\eta^2}
$$

so that

$$
\sigma^2_{\int Z^H} = \frac{s^{-1/\eta}-1}{c(x,y)} - \frac{\log^2 s}{\eta^2}.
$$

It remains to show that $\sigma^2_{x,y,s}$ possesses the form (4.6) if $\hat\eta_n$ is the MLE.

The Case $\hat\eta_n = \hat\eta_n^{ML}$

Here, due to (3.14), $\nu_\eta(dt) = \nu_\eta^{ML}(dt) = \big((\eta+1)/\eta\big)^2 \big(t^\eta - (2\eta+1)t^{2\eta}\big)\, dt + \big((\eta+1)/\eta\big)\varepsilon_1(dt)$, so that the integral of (4.28) reads as

$$\begin{aligned}
&\int_0^1 t^{-(\eta+1)} W(t^\eta, t^\eta)\nu_\eta^{ML}(dt)\\
&= \left(\frac{\eta+1}{\eta}\right)^2 \int_0^1 \big(t^\eta - (2\eta+1)t^{2\eta}\big)t^{-(\eta+1)} W(t^\eta,t^\eta)dt + \frac{\eta+1}{\eta} W(1,1).
\end{aligned}$$

Hence, we may continue (4.28) to

$$\begin{aligned}
&\log \frac{S_n^*\left(\frac{(n+1)sx}{nq^{\leftarrow}(m/n)Q_n^*(1)}-, \frac{(n+1)sy}{nq^{\leftarrow}(m/n)Q_n^*(1)}-\right)}{S_n^*\left(\frac{(n+1)x}{nq^{\leftarrow}(m/n)Q_n^*(1)}-, \frac{(n+1)y}{nq^{\leftarrow}(m/n)Q_n^*(1)}-\right)} - \frac{1}{\hat\eta_n^{ML}}\log s\\
&= m^{-1/2}\Bigg[\frac{W(sx,sy)}{c(sx,sy)} - \frac{W(x,y)}{c(x,y)}\\
&\qquad + \frac{(\eta+1)^2 \log s}{\eta^3}\int_0^1 \big(t^{-1} - (2\eta+1)t^{\eta-1}\big)W(t^\eta,t^\eta)dt\\
&\qquad + \frac{(\eta+1)\log s}{\eta^2} W(1,1)\Bigg] + o_P(m^{-1/2})\\
&= m^{-1/2}\frac{(\eta+1)^2\log s}{\eta^3}\int_0^1 \Bigg[\frac{\eta^3}{(\eta+1)^2\log s}\left(\frac{W(sx,sy)}{c(sx,sy)} - \frac{W(x,y)}{c(x,y)}\right)\\
&\qquad + \big(t^{-1} - (2\eta+1)t^{\eta-1}\big)W(t^\eta,t^\eta) + \frac{\eta}{\eta+1}W(1,1)\Bigg]dt\\
&\qquad + o_P(m^{-1/2}).
\end{aligned} \tag{4.32}$$

Similar to the Hill case we now define

$$\begin{aligned}
Z^{ML}(t) := {}& \big(t^{-1} - (2\eta+1)t^{\eta-1}\big)W(t^\eta,t^\eta) + \frac{\eta}{\eta+1}W(1,1)\\
&+ \frac{\eta^3}{(\eta+1)^2\log s}\left(\frac{W(sx,sy)}{c(sx,sy)} - \frac{W(x,y)}{c(x,y)}\right),
\end{aligned}$$

so that it remains to show that

$$\sigma^2_{x,y,s} = \frac{(\eta+1)^4 \log^2 s}{\eta^6} \; \sigma^2_{\int Z^{ML}}, \tag{4.33}$$

where $\sigma^2_{\int Z^{ML}}$ is the variance of $\int_0^1 Z^{ML}(t)dt$. Note that (4.29) and (4.30) hold for Z^H replaced with Z^{ML}. We start calculating

$$\begin{aligned}
&\operatorname{cov}\big[Z^{ML}(r), Z^{ML}(t)\big] = E\big[Z^{ML}(r)Z^{ML}(t)\big] \\
&= E\left[\big(r^{-1} - (2\eta+1)r^{\eta-1}\big)W(r^\eta, r^\eta)\big(t^{-1} - (2\eta+1)t^{\eta-1}\big)W(t^\eta, t^\eta)\right] \\
&\quad + E\left[\big(r^{-1} - (2\eta+1)r^{\eta-1}\big)W(r^\eta, r^\eta)\frac{\eta}{\eta+1}W(1,1)\right] \\
&\quad + E\left[\frac{\eta^3\big(r^{-1} - (2\eta+1)r^{\eta-1}\big)}{(\eta+1)^2\log s}W(r^\eta, r^\eta)\left(\frac{W(sx,sy)}{c(sx,sy)} - \frac{W(x,y)}{c(x,y)}\right)\right] \\
&\quad + E\left[\frac{\eta}{\eta+1}W(1,1)\big(t^{-1} - (2\eta+1)t^{\eta-1}\big)W(t^\eta, t^\eta)\right] \\
&\quad + E\left[\frac{\eta^3\big(t^{-1} - (2\eta+1)t^{\eta-1}\big)}{(\eta+1)^2\log s}\left(\frac{W(sx,sy)}{c(sx,sy)} - \frac{W(x,y)}{c(x,y)}\right)W(t^\eta, t^\eta)\right] \qquad (4.34) \\
&\quad + E\left[\left(\frac{\eta}{\eta+1}W(1,1) + \frac{\eta^3}{(\eta+1)^2\log s}\left(\frac{W(sx,sy)}{c(sx,sy)} - \frac{W(x,y)}{c(x,y)}\right)\right)^2\right],
\end{aligned}$$

so that by the homogeneity (2.14) of c,

$$
\begin{aligned}
&\sigma^2_{\int Z^{ML}} \\
&= \int_0^1 \int_0^1 \left(r^{-1} - (2\eta+1)r^{\eta-1}\right)\left(t^{-1} - (2\eta+1)t^{\eta-1}\right) c(r^\eta \wedge t^\eta, r^\eta \wedge t^\eta) drdt \\
&\quad +2 \int_0^1 \left(t^{-1} - (2\eta+1)t^{\eta-1}\right) \frac{\eta}{\eta+1} c(t^\eta \wedge 1, t^\eta \wedge 1)\ dt \\
&\quad +2 \int_0^1 \frac{\eta^3\left(t^{-1} - (2\eta+1)t^{\eta-1}\right)}{(\eta+1)^2 c(x,y)\log s} \left(\frac{c(sx \wedge t^\eta, sy \wedge t^\eta)}{s^{1/\eta}} - c(x \wedge t^\eta, y \wedge t^\eta)\right) dt \\
&\quad + \left(\frac{\eta}{\eta+1}\right)^2 + \frac{\eta^6}{(\eta+1)^4 c(x,y)\log^2 s}\left(s^{-1/\eta} - 2 + 1\right) \\
&\quad +2 \frac{\eta^4}{(\eta+1)^3 c(x,y)\log s}\left(s^{-1/\eta} c(sx \wedge 1, sy \wedge 1) - c(x \wedge 1, y \wedge 1)\right). \qquad (4.35)
\end{aligned}
$$

We continue in merely considering the first two summands, since the others are not simplifiable any more for arbitrary $x, y > 0$. Similar to the procedure in the Hill case and once more by the homogeneity (2.14) of c, we split the remaining double integral into

$$
\begin{aligned}
&\int_0^1 \left(t^{-1} - (2\eta+1)t^{\eta-1}\right) \int_0^t \left(r^{-1} - (2\eta+1)r^{\eta-1}\right) r\ drdt \\
&\quad + \int_0^1 \left(t^{-1} - (2\eta+1)t^{\eta-1}\right) \int_t^1 \left(r^{-1} - (2\eta+1)r^{\eta-1}\right) t\ drdt
\end{aligned}
$$

For symmetry reasons, these two double integrals are identical and therefore their sum equals

$$
2\int_0^1 \left[1 - t^\eta(2\eta+1) - \frac{2\eta+1}{\eta+1} t^\eta + \frac{(2\eta+1)^2}{\eta+1} t^{2\eta}\right] dt = 2\left(1 - \frac{2\eta+1}{(\eta+1)^2}\right).
$$

Since the second summand $2\int_0^1 \left(t^{-1} - (2\eta+1)t^{\eta-1}\right)\eta/(\eta+1)c(t^\eta \wedge 1, t^\eta \wedge 1)dt$ of (4.35) equals $2\int_0^1 \eta/(\eta+1)\left(1 - (2\eta+1)t^\eta\right)dt = -2\left(1 - (2\eta+1)/(\eta+1)^2\right)$, summands number 3-6 of that formula remain for $\sigma^2_{\int Z^{ML}}$. Recall from (4.33) that this must be multiplied with $(\eta+1)^4(\log^2 s)/\eta^6$ to obtain $\sigma^2_{x,y,s}$ in the general case $x, y > 0$, i.e.

$$\begin{aligned}
&\sigma^2_{x,y,s}\\
&= \frac{s^{-1/\eta}-1}{c(x,y)} + \frac{(\eta+1)^2\log^2 s}{\eta^4}\Bigg[1+\frac{2\eta}{c(x,y)\log s}\\
&\quad\times\int_0^1 \left(t^{-1}-(2\eta+1)t^{\eta-1}\right)\left(s^{-1/\eta}c(sx\wedge t^\eta, sy\wedge t^\eta)-c(x\wedge t^\eta, y\wedge t^\eta)\right)dt\Bigg]\\
&\quad+\frac{2(\eta+1)(\log s)\left(s^{-1/\eta}c(sx\wedge 1, sy\wedge 1)-c(x\wedge 1, y\wedge 1)\right)}{\eta^2 c(x,y)},
\end{aligned} \tag{4.36}$$

cf. Remark 4.2.2 (ii).

The expression slightly simplifies in the case $x, y \in (0, 1]$. Note that then the last summand of (4.36) equals 0 and for the integral we have seen in the Hill case when calculating (the Lebesgue integral over $[0,1]^2$ of) J that

$$\begin{aligned}
&\int_0^1 t^{-1}\left(s^{-1/\eta}c(sx\wedge t^\eta, sy\wedge t^\eta)-c(x\wedge t^\eta, y\wedge t^\eta)\right)dt\\
&\qquad\qquad= \int_1^{s^{-1/\eta}} t^{-1}c(x\wedge t^\eta, y\wedge t^\eta)dt = -\frac{c(x,y)\log s}{\eta}.
\end{aligned}$$

Further, in a similar manner, i.e. with substitution $t = s^{1/\eta}u$ (and renaming u), one obtains that

$$\begin{aligned}
&\int_0^1 t^{\eta-1}\left(s^{-1/\eta}c(sx\wedge t^\eta, sy\wedge t^\eta)-c(x\wedge t^\eta, y\wedge t^\eta)\right)dt\\
&= s\int_0^{s^{-1/\eta}} t^{\eta-1}c(x\wedge t^\eta, y\wedge t^\eta)dt - \int_0^1 t^{\eta-1}c(x\wedge t^\eta, y\wedge t^\eta)dt\\
&= (s-1)\int_0^1 t^{\eta-1}c(x\wedge t^\eta, y\wedge t^\eta)dt + s\int_1^{s^{-1/\eta}} t^{\eta-1}c(x\wedge t^\eta, y\wedge t^\eta)dt\\
&= (s-1)\left(\int_0^1 t^{\eta-1}c(x\wedge t^\eta, y\wedge t^\eta)dt - \frac{c(x,y)}{\eta}\right)
\end{aligned}$$

Hence, the second summand of (4.36) reads as

$$\frac{(\eta+1)^2\log^2 s}{\eta^4}\left[-1-\frac{2(2\eta+1)(s-1)}{c(x,y)\log s}\left(\eta\int_0^1 t^{\eta-1}c(x\wedge t^\eta, y\wedge t^\eta)dt - c(x,y)\right)\right]$$

which completes the proof.

4.6.3 Proof of Corollary 4.3.1

Similar as in the proof of Corollary 4.2.1, we distiguish the cases that $\hat{\eta}_n$ is the Hill estimator and the MLE, respectively, in separate subsections.

Note that we can copy the proof(s) of the case of asymptotic independence until the calculation of

$$\operatorname{cov}\big[Z^H(r), Z^H(t)\big] = E\big[Z^H(r)Z^H(t)\big] \qquad \text{and}$$

$$\operatorname{cov}\big[Z^{ML}(r), Z^{ML}(t)\big] = E\big[Z^{ML}(r)Z^{ML}(t)\big],$$

respectively. Recall from (3.12) and (3.10) that the particular difficulty in the case of asymptotic dependence is that the (covariance) structure of the Gaussian process W is substantially more complicated. Recall further that we must show that $\sigma^2_{x,y,s} = \sigma^2_{\int Z^H}$ if $\hat{\eta}_n = \hat{\eta}_n^H$ and $\sigma^2_{x,y,s} = \big(16\log^2 s\big)\sigma^2_{\int Z^{ML}}$ (since here $\eta = 1$) if $\hat{\eta}_n = \hat{\eta}_n^{ML}$, where $\sigma^2_{\int Z^H}$ and $\sigma^2_{\int Z^{ML}}$ are the Lebesgue integrals of $\operatorname{cov}\big[Z^H(r), Z^H(t)\big]$ and $\operatorname{cov}\big[Z^{ML}(r), Z^{ML}(t)\big]$ over $[0,1]^2$, respectively.

The case $\hat{\eta}_n = \hat{\eta}_n^H$

$$\begin{aligned}
&\operatorname{cov}\big[Z^H(r), Z^H(t)\big] = E\big[Z^H(r)Z^H(t)\big] \\
&= (\log^2 s) E\Bigg[(rt)^{-1}W(r,r)W(t,t) \\
&\qquad\qquad -W(1,1)\left[r^{-1}W(r,r) + t^{-1}W(t,t)\right] + W^2(1,1)\Bigg] \\
&\quad +(\log s)E\left[\left(\frac{W(sx,sy)}{c(sx,sy)} - \frac{W(x,y)}{c(x,y)}\right)\left(\frac{W(r,r)}{r} - 2W(1,1) + \frac{W(t,t)}{t}\right)\right] \\
&\quad +E\left[\frac{W(sx,sy)}{c(sx,sy)} - \frac{W(x,y)}{c(x,y)}\right]^2 \qquad\qquad (4.37) \\
&=: A(r,t) + B(r,t,x,y) + C(x,y).
\end{aligned}$$

Define

$$\begin{aligned} A_1(r,t) &:= E\left[(rt)^{-1}W(r,r)W(t,t)\right] \\ A_2(r,t) &:= E\left[W(1,1)\big(r^{-1}W(r,r)+t^{-1}W(t,t)\big)\right] \\ A_3 &:= E\left[W^2(1,1)\right] \end{aligned} \tag{4.38}$$

which means that

$$A(r,t) = (\log^2 s)\big(A_1(r,t) - A_2(r,t) + A_3\big).$$

Calculating A_1, A_2 and A_3, it is convenient to first specify $E[W(r,r)W(t,t)]$ for $r \le t$ which is easily done with the help of (3.50). Recall (3.22) and from (3.19) that c_x and c_y are homogeneous of order 0, so that $c_x(r,r) = c_x(t,t) = c_x$ and $c_y(r,r) = c_y(t,t) = c_y$. We obtain

$$\begin{aligned} &E[W(r,r)W(t,t)] \\ &= r - lrc_x - lrc_y - lc_x\big[c(r,t) - rc_x - lc_y c(r,t)\big] - lc_y\big[c(t,r) - rc_y - lc_x c(t,r)\big] \\ &= r\big(1 + l(c_x^2 + c_y^2 - 1)\big) - l\big[c(r,t)c_x(1 - lc_y) + c(t,r)c_y(1 - lc_x)\big]. \end{aligned} \tag{4.39}$$

Hence, $A_1(r,t)$ equals

$$t^{-1}\big(1 + l(c_x^2 + c_y^2 - 1)\big) - (rt)^{-1}l\big[c(r,t)c_x(1 - lc_y) + c(t,r)c_y(1 - lc_x)\big]$$

for $r \le t$ and therefore

$$\int_0^1\int_0^1 A_1(r,t)drdt = 2\int_0^1\int_0^t A_1(r,t)drdt$$

$$= 2\big(1+l(c_x^2+c_y^2-1)\big) - 2l\int_0^1\int_0^t (rt)^{-1}\big[c(r,t)c_x(1-lc_y)+c(t,r)c_y(1-lc_x)\big]drdt$$

$$= 2\big(1+l(c_x^2+c_y^2-1)\big) - 2l\int_0^1 u^{-1}\left[c(u,1)c_x(1 - lc_y) + c(1,u)c_y(1 - lc_x)\right]du,$$

where in the last step we substituted $r = tu$ in the double integral and exploited the homogeneity (2.14) of c. Further, note that

$$\int_0^1\int_0^1 A_2(r,t)drdt$$

$$= 2\int_0^1 \Big(1 + l(c_x^2 + c_y^2 - 1) - r^{-1}l\big[c(r,1)c_x(1 - lc_y) + c(1,r)c_y(1 - lc_x)\big]\Big)dr$$

so that $\int_0^1 \int_0^1 \big(A_1(r,t) - A_2(r,t)\big) drdt = 0$. Moreover,

$$A_3 = 1 + l(c_x^2 + c_y^2 - 1) - l(c_x + c_y - 2lc_xc_y), \tag{4.40}$$

so that we obtain by (3.22) with $x = y = 1$

$$\int_0^1 \int_0^1 A(r,t)\ drdt = (\log^2 s)(1 - 2lc_xc_y)(1 - l) \tag{4.41}$$

for the Lebesgue integral over $[0,1]^2$ of the first summand A of (4.37).

In order to treat the second of the three summands B of (4.37), we first calculate $E\big[W(x,y)W(r,r)\big]$ by means of (3.50), which yields

$$\begin{aligned}
&E[W(x,y)W(r,r)] \\
&= c(x,r) - c(r \vee x, r) + c(r,y) + r - c(r, r \vee y) + c(x,y)(1-l) \\
&\quad -c(r \vee x, y) - c(x, r \vee y) + c(r \vee x, r \vee y) \\
&\quad -lc_x\big[c(r,y) - c(r \vee x, y)\big] - lc_y\big[c(x,r) - c(x, r \vee y)\big] \\
&\quad -lc_x(x,y)\big[c(x,r) + r - c(r \vee x, r) - c_x(r \wedge x) - lc_yc(x,r)\big] \\
&\quad -lc_y(x,y)\big[c(r,y) + r - c(r, r \vee y) - c_y(r \wedge y) - lc_xc(r,y)\big] \\
&= \chi_1(x,y,r) + c(x,y)(1-l).
\end{aligned} \tag{4.42}$$

Thus,

$$\begin{aligned}
&B(r,t,x,y) \\
&= \frac{\log s}{c(x,y)}\Big((sr)^{-1}\chi_1(sx,sy,r) - 2s^{-1}\chi_1(sx,sy,1) + (st)^{-1}\chi_1(sx,sy,t) \\
&\qquad -r^{-1}\chi_1(x,y,r) + 2\chi_1(x,y,1) - t^{-1}\chi_1(x,y,t)\Big)
\end{aligned}$$

and

$$\begin{aligned}
\int_0^1 \int_0^1 B(r,t,x,y)\ drdt = \frac{2\log s}{c(x,y)}\Big(&\chi_1(x,y,1) - s^{-1}\chi_1(sx,sy,1) \\
&+ \int_0^1 \big[(sr)^{-1}\chi_1(sx,sy,r) - r^{-1}\chi_1(x,y,r)\big]dr\Big).
\end{aligned}$$

Substitute $r = su$ (and rename u) in the first of the integral's summands and observe that χ_1 is homogeneous of order 1 to obtain

$$\int_0^1 \int_0^1 B(r,t,x,y)\ drdt = \frac{2\log s}{c(x,y)}\Big(\chi_1(x,y,1) - s^{-1}\chi_1(sx,sy,1) \tag{4.43}$$
$$+ \int_1^{s^{-1}} r^{-1}\chi_1(x,y,r)dr\Big).$$

Recall the definition (4.7) of χ_2 to see that if $x, y \in (0,1]$,

$$\chi_1(x,y,1) - s^{-1}\chi_1(sx,sy,1) = \chi_2(x,y) - s^{-1}\chi_2(sx,sy) \tag{4.44}$$

and

$$\int_1^{s^{-1}} r^{-1}\chi_1(x,y,r)dr$$
$$= \int_1^{s^{-1}} lr^{-1}\Big(c_x(x,y)[c(x,r)(lc_y - 1) + xc_x]$$
$$+c_y(x,y)[c(r,y)(lc_x - 1) + yc_y]\Big)dr$$

$$= -l(\log s)[xc_xc_x(x,y) + yc_yc_y(x,y)]$$
$$+ \int_1^{s^{-1}} l\big[(lc_y - 1)c_x(x,y)c(x/r,1) + (lc_x - 1)c_y(x,y)c(1,y/r)\big]dr, \tag{4.45}$$

so that

$$\int_0^1 \int_0^1 B(r,t,x,y)\ drdt$$
$$= \frac{2\log s}{c(x,y)}\Bigg(\chi_2(x,y) - s^{-1}\chi_2(sx,sy) - l(\log s)[xc_xc_x(x,y) + yc_yc_y(x,y)]$$
$$+ \int_1^{s^{-1}} \chi_2(x/r,y/r)\ dr\Bigg) \tag{4.46}$$

if $x, y \in (0,1]$.

Turning to the third summand C of (4.37), we first see from (3.50) that

$$\begin{aligned}
&E[W(sx,sy)W(x,y)] \\
&= sc(x,y) - lc_x(x,y)\big[sc(x,y) + c(sx,y) - sxc_x(x,y) - lc_y(x,y)c(sx,y)\big] \\
&\qquad -lc_y(x,y)\big[sc(x,y) + c(x,sy) - syc_y(x,y) - lc_x(x,y)c(x,sy)\big].
\end{aligned}$$

(Note that then $E\big[W^2(sx,sy)\big]$ can be easily calculated by substituting s by 1, x by sx and y by sy.) Hence,

$$\begin{aligned}
&\int_0^1\int_0^1 C(x,y)drdt = C(x,y) \\
&= \frac{1}{c^2(x,y)}\Big(s^{-2}E[W^2(sx,sy)] - 2s^{-1}E[W(sx,sy)W(x,y)] + E[W^2(x,y)]\Big) \\
&= \frac{1}{c^2(x,y)}\Big(s^{-1}c(x,y) - ls^{-1}c_x(x,y)[2c(x,y) - xc_x(x,y) - lc_y(x,y)c(x,y)] \\
&\qquad -2c(x,y) - ls^{-1}c_y(x,y)[2c(x,y) - yc_y(x,y) - lc_x(x,y)c(x,y)] \\
&\qquad +2ls^{-1}c_x(x,y)[sc(x,y) + c(sx,y) - sxc_x(x,y) - lc_y(x,y)c(sx,y)] \\
&\qquad +2ls^{-1}c_y(x,y)[sc(x,y) + c(x,sy) - syc_y(x,y) - lc_x(x,y)c(x,sy)] \\
&\qquad +c(x,y) - lc_x(x,y)[2c(x,y) - xc_x(x,y) - lc_y(x,y)c(x,y)] \\
&\qquad\qquad -lc_y(x,y)[2c(x,y) - yc_y(x,y) - lc_x(x,y)c(x,y)]\Big) \\
&= \frac{1}{c^2(x,y)}\Big(s^{-1}c(x,y) - c(x,y) \\
&\qquad -lc_x(x,y)[2s^{-1}c(x,y) - s^{-1}xc_x(x,y) - 2s^{-1}c(sx,y) + xc_x(x,y)] \\
&\qquad -lc_y(x,y)[2s^{-1}c(x,y) - s^{-1}yc_y(x,y) - 2s^{-1}c(x,sy) + yc_y(x,y)] \\
&\qquad +2l^2c_x(x,y)c_y(x,y)\big[s^{-1}c(x,y) - s^{-1}[c(sx,y) + c(x,sy)] + c(x,y)\big]\Big)
\end{aligned}$$

$$
\begin{aligned}
= \frac{1}{c^2(x,y)} \Bigg(& (s^{-1}-1)c(x,y) \\
& + lc_x(x,y)\big[(s^{-1}-1)xc_x(x,y) + 2s^{-1}\big(c(sx,y)-c(x,y)\big)\big] \\
& + lc_y(x,y)\big[(s^{-1}-1)yc_y(x,y) + 2s^{-1}\big(c(x,sy)-c(x,y)\big)\big] \\
& + 2l^2 c_x(x,y)c_y(x,y)\big[(s^{-1}+1)c(x,y) - s^{-1}\big(c(sx,y)+c(x,sy)\big)\big]\Bigg).
\end{aligned} \tag{4.47}
$$

Hence, adding up (4.41), (4.43) and (4.47), which is $\int_0^1\int_0^1 \big(A(r,t)+B(r,t,x,y)+C(x,y)\big)drdt = \sigma^2_{\int Z^H}$ for arbitrary $x,y>0$, yields the asymptotic variance mentioned in Remark 4.3.2, namely

$$
\begin{aligned}
\sigma^2_{\int Z^H} = {} & (\log^2 s)(1-2lc_xc_y)(1-l) \\
& + \frac{2\log s}{c(x,y)}\Big(\chi_1(x,y,1) - s^{-1}\chi_1(sx,sy,1) + \int_1^{s^{-1}} r^{-1}\chi_1(x,y,r)dr\Big) \\
& + \frac{1}{c^2(x,y)}\Bigg((s^{-1}-1)c(x,y) \\
& \qquad + lc_x(x,y)\big[(s^{-1}-1)xc_x(x,y) + 2s^{-1}\big(c(sx,y)-c(x,y)\big)\big] \\
& \qquad + lc_y(x,y)\big[(s^{-1}-1)yc_y(x,y) + 2s^{-1}\big(c(x,sy)-c(x,y)\big)\big] \\
& \qquad + 2l^2 c_x(x,y)c_y(x,y)\big[(s^{-1}+1)c(x,y) - s^{-1}\big(c(sx,y)+c(x,sy)\big)\big]\Bigg).
\end{aligned} \tag{4.48}
$$

Adding up (4.41), (4.46) and (4.47), which is $\int_0^1\int_0^1 \big(A(r,t)+B(r,t,x,y)+C(x,y)\big)drdt = \sigma^2_{\int Z^H}$ if $x,y\in(0,1]$, yields the required result.

The Case $\hat{\eta}_n = \hat{\eta}_n^{ML}$

Similar as in the Hill case, we merely need to calculate (4.34) for $\eta = 1$ and W with covariance structure according to (3.12) and (3.10) and integrate the result over $[0,1]^2$ to obtain $\sigma^2_{\int Z^{ML}}$.

$$\mathrm{cov}\big[Z^{ML}(r), Z^{ML}(t)\big] = E\big[Z^{ML}(r)\ Z^{ML}(t)\big]$$

$$\begin{aligned}
&= E\left[(r^{-1}-3)W(r,r)(t^{-1}-3)W(t,t)\right] \\
&\quad + E\left[(r^{-1}-3)W(r,r)\frac{1}{2}W(1,1)\right] \\
&\quad + E\left[\frac{1}{2}W(1,1)(t^{-1}-3)W(t,t)\right] \\
&\quad + E\left[(r^{-1}-3)W(r,r)\frac{1}{4\log s}\left(\frac{W(sx,sy)}{c(sx,sy)} - \frac{W(x,y)}{c(x,y)}\right)\right] \\
&\quad + E\left[\frac{1}{4\log s}\left(\frac{W(sx,sy)}{c(sx,sy)} - \frac{W(x,y)}{c(x,y)}\right)(t^{-1}-3)W(t,t)\right] \\
&\quad + E\left[\left(\frac{1}{2}W(1,1) + \frac{1}{4\log s}\left(\frac{W(sx,sy)}{c(sx,sy)} - \frac{W(x,y)}{c(x,y)}\right)\right)^2\right] \\
&=: T_1(r,t) + T_2(r) + T_3(t) + T_4(r,x,y) + T_5(t,x,y) + T_6(x,y).
\end{aligned}$$

We add T_1, T_2 and T_3 to obtain

$$\begin{aligned}
T_1(r,t) + T_2(r) + T_3(t) &= A_1(r,t) - 3tA_1(r,t) - 3rA_1(r,t) + 9rtA_1(r,t) \\
&\quad + \frac{1}{2}A_2(r,t) - \frac{3}{2}E\big[W(1,1)\big(W(r,r)+W(t,t)\big)\big].
\end{aligned} \tag{4.49}$$

In the Hill case, we have seen that

$$\begin{aligned}
&\int_0^1\int_0^1 \Big(A_1(r,t) + \frac{1}{2}A_2(r,t)\Big)drdt \\
&= 3\big(1 + l(c_x^2 + c_y^2 - 1)\big) - 3l\int_0^1 r^{-1}[c_x c(r,1)(1-lc_y) + c_y c(1,r)(1-lc_x)]dr.
\end{aligned} \tag{4.50}$$

Further, by (4.39) we readily see that

$$
\begin{aligned}
&\int_0^1 \int_0^1 \frac{3}{2} E\big[W(1,1)\big(W(r,r)+W(t,t)\big)\big] drdt \\
&= 3\int_0^1 \Big(r\big(1+l(c_x^2+c_y^2-1)\big) - l\big[c_x c(r,1)(1-lc_y) + c_y c(1,r)(1-lc_x)\big]\Big) dr \\
&= \frac{3}{2}\big(1+l(c_x^2+c_y^2-1)\big) - 3l\int_0^1 \big[c_x c(r,1)(1-lc_y)+c_y c(1,r)(1-lc_x)\big] dr. \qquad (4.51)
\end{aligned}
$$

With the techniques that we have already seen in the proof of the Hill case (symmetry, substitution)

$$
\begin{aligned}
&\int_0^1 \int_0^1 9rtA_1(r,t) drdt \\
&= 9\int_0^1 \int_0^1 E\big[W(r,r)W(t,t)\big] drdt \\
&= 3\big(1+l(c_x^2+c_y^2-1)\big) - 18l\int_0^1 \int_0^t \big[c_x c(r,t)(1-lc_y) + c_y c(t,r)(1-lc_x)\big] drdt \\
&= 3\big(1+l(c_x^2+c_y^2-1)\big) - 6l\int_0^1 \big[c_x c(r,1)(1-lc_y) + c_y c(1,r)(1-lc_x)\big] dr \qquad (4.52)
\end{aligned}
$$

and

$$
\begin{aligned}
&\int_0^1 \int_0^1 3tA_1(r,t) + 3rA_1(r,t) drdt \\
&= 6\int_0^1 \int_0^t \Big(r^{-1}E\big[W(r,r)W(t,t)\big] + t^{-1}E\big[W(r,r)W(t,t)\big]\Big) drdt \\
&= 6\int_0^1 \int_0^t \Big(1+l(c_x^2+c_y^2-1) + rt^{-1}\big(1+l(c_x^2+c_y^2-1)\big)\Big) drdt \\
&\quad -6l\int_0^1 \int_0^1 \Big(tr^{-1}\big[c_x c(r,1)(1-lc_y) + c_y c(1,r)(1-lc_x)\big] \\
&\qquad\qquad + t\big[c_x c(r,1)(1-lc_y) + c_y c(1,r)(1-lc_x)\big]\Big) drdt \\
&= 3\big(1+l(c_x^2+c_y^2-1)\big) + \frac{3}{2}\big(1+l(c_x^2+c_y^2-1)\big) \\
&\quad -3l\int_0^1 \big(r^{-1}+1\big)\big[c_x c(r,1)(1-lc_y) + c_y c(1,r)(1-lc_x)\big] dr. \qquad (4.53)
\end{aligned}
$$

Hence, the sum of the double integrals of T_1, T_2 and T_3 is (4.50) minus (4.51) plus (4.52) minus (4.53), such that

$$\int_0^1 \int_0^1 \Big(T_1(r,t) + T_2(r) + T_3(t)\Big) dr dt = 0.$$

By the homogeneity (2.14) of c and with (4.42) we obtain that

$$\begin{aligned}
&T_4(r,x,y)\\
&= \frac{1}{4c(x,y)\log s}\Big((sr)^{-1}\big(\chi_1(sx,sy,r) + c(sx,sy)(1-l)\big)\\
&\qquad -r^{-1}\big(\chi_1(x,y,r) + c(x,y)(1-l)\big)\\
&\qquad -3s^{-1}\big(\chi_1(sx,sy,r) + c(sx,sy)(1-l)\big)\\
&\qquad +3\big(\chi_1(x,y,r) + c(x,y)(1-l)\big)\Big)\\
&= \frac{(sr)^{-1}\chi_1(sx,sy,r) - r^{-1}\chi_1(x,y,r) - 3s^{-1}\chi_1(sx,sy,r) + 3\chi_1(x,y,r)}{4c(x,y)\log s}
\end{aligned}$$

which yields

$$\begin{aligned}
&\int_0^1 \int_0^1 T_4(r,x,y) dr dt\\
&= \frac{1}{4c(x,y)\log s}\int_0^1 \Big((sr)^{-1}\chi_1(sx,sy,r) - r^{-1}\chi_1(x,y,r)\\
&\qquad +3\chi_1(x,y,r) - 3s^{-1}\chi_1(sx,sy,r)\Big) dr\\
&= \frac{1}{4c(x,y)\log s}\Big(\int_1^{s^{-1}} r^{-1}\chi_1(x,y,r) dr \\
&\qquad + \int_0^1 3\chi_1(x,y,r) dr - \int_0^{s^{-1}} 3s\chi_1(x,y,r) dr\Big),
\end{aligned} \tag{4.54}$$

where in the last step we substituted $r = su$ and renamed u in the first and the last integral summand. Observe that $T_4 = T_5$ and recall (4.45) to see that if $x, y \in [0,1]$,

$$\int_0^1 \int_0^1 \Big(T_4(r,x,y) + T_5(t,x,y)\Big) drdt$$

$$= \frac{1}{2c(x,y)\log s}\Bigg(\int_1^{s^{-1}} \chi_2(x/r, y/r)\ dr - l(\log s)[xc_x c_x(x,y) + yc_y c_y(x,y)]$$

$$+3\int_0^1 \chi_1(x,y,r)dr - 3s\int_0^{s^{-1}} \chi_1(x,y,r)dr\Bigg). \tag{4.55}$$

We expand T_6 to

$$T_6(x,y) = E\Bigg[\frac{W^2(1,1)}{4}\Bigg] + E\left[\frac{1}{4\log s}\left(\frac{W(sx,sy)}{c(sx,sy)} - \frac{W(x,y)}{c(x,y)}\right)W(1,1)\right] \tag{4.56}$$

$$+E\left[\frac{1}{16\log^2 s}\left(\frac{W^2(sx,sy)}{c^2(sx,sy)} - 2\frac{W(sx,sy)W(x,y)}{c(sx,sy)c(x,y)} + \frac{W^2(x,y)}{c^2(x,y)}\right)\right].$$

Recall (4.38) and (4.40) for the first summand of (4.56), exploit the homogeneity (2.14) of c, consider (4.42) with $r = 1$ for the second summand of (4.56) and recall (4.47) for the third summand of that formula to see that

$$\int_0^1 \int_0^1 T_6(x,y)drdt = T_6(x,y)$$

$$= \frac{1+l(c_x^2+c_y^2-1)-l(c_x+c_y-2lc_xc_y)}{4} + \frac{s^{-1}\chi_1(sx,sy,1) - \chi_1(x,y,1)}{4c(x,y)\log s}$$

$$+\frac{C(x,y)}{16\log^2 s} \tag{4.57}$$

which due to (4.44) simplifies to

$$\int_0^1 \int_0^1 T_6(x,y)drdt = T_6(x,y)$$

$$= \frac{1+l(c_x^2+c_y^2-1)-l(c_x+c_y-2lc_xc_y)}{4} + \frac{s^{-1}\chi_2(sx,sy) - \chi_2(x,y)}{4c(x,y)\log s} + \frac{C(x,y)}{16\log^2 s} \tag{4.58}$$

if $x, y \in [0,1]$.

Recall from the beginning of Section 4.6.3 that the sum of (4.57) and twice (4.54) equals $\sigma^2_{\int ZML}$ for arbitrary $x, y > 0$, and that this sum must be multiplied with $16\log^2 s$ to obtain the asymptotic variance mentioned in Remark 4.3.2, namely

$$
\begin{aligned}
&16\sigma^2_{\int ZML}\log^2 s \\
&= 4(\log^2 s)(1-2lc_xc_y)(1-l) \\
&\quad + \frac{8\log s}{c(x,y)}\Big(\int_1^{s^{-1}} r^{-1}\chi_1(x,y,r)dr + \int_0^1 3\chi_1(x,y,r)dr - \int_0^{s^{-1}} 3s\chi_1(x,y,r)dr \\
&\qquad\qquad + \frac{s^{-1}\chi_1(sx,sy,1)-\chi_1(x,y,1)}{2}\Big) \\
&\quad + \frac{1}{c^2(x,y)}\Big((s^{-1}-1)c(x,y) \\
&\qquad + lc_x(x,y)\big[(s^{-1}-1)xc_x(x,y) + 2s^{-1}\big(c(sx,y)-c(x,y)\big)\big] \\
&\qquad + lc_y(x,y)\big[(s^{-1}-1)yc_y(x,y) + 2s^{-1}\big(c(x,sy)-c(x,y)\big)\big] \\
&\qquad + 2l^2c_x(x,y)c_y(x,y)\big[(s^{-1}+1)c(x,y) - s^{-1}\big(c(sx,y)+c(x,sy)\big)\big]\Big).
\end{aligned}
\tag{4.59}
$$

Add (4.55) and (4.58) and multiply this sum with $16\log^2 s$ to obtain $\sigma^2_{x,y,s}$.

4.6.4 Proofs for Section 4.5

Before we can prove Theorem 4.5.1, we have to make some preparations. Let $(\tilde{W}(B))_{B\in\bar{\mathbb{B}}[0,\infty)^2}$ be a (set indexed) centered Gaussian process defined for all bounded Borel sets $B \in \bar{\mathbb{B}}[0,\infty)^2$ and with covariance structure

$$\operatorname{cov}(\tilde{W}(A), \tilde{W}(B)) = \mu(A\cap B),$$

where the measure μ, defined by $\mu\big([0,x]\times[0,y]\big) = c(x,y)$ for all $x, y \geq 0$, cf. (2.18), inherits the homogeneity (2.14) of c, i.e. $\mu(tB) = t^{1/\eta}\mu(B)$ for all

$t > 0$ and all bounded Borel sets $B \in \bar{\mathbb{B}}[0,\infty)^2$. Let the random variable $W_l(x,y)$ be the process $\tilde{W}$ with the lower triangle of the rectangle $[0,x]\times[0,y]$ as argument, i.e.

$$W_l(x,y) := \tilde{W}\left(\left\{(w,z) \in [0,\infty)^2 \mid w \le x, z \le \frac{y}{x}\, w\right\}\right).$$

Analogously, let

$$W_u(x,y) := \tilde{W}\left(\left\{(w,z) \in [0,\infty)^2 \mid z \le y, w < \frac{x}{y}\, z\right\}\right)$$

be $\tilde{W}$ with the the upper triangle of the rectangle $[0,x]\times[0,y]$ (without the diagonal) as argument. See the left hand plot of Figure 4.3. Next we show that in fact $W \overset{D}{=} W_l + W_u$.

4.6.1 Lemma. The process $W_l + W_u$ is a version of W, i.e.

$$E\left[\big((W_l + W_u)(x_1,y_1)\big)\big((W_l + W_u)(x_2,y_2)\big)\right] = c(x_1 \wedge x_2, y_1 \wedge y_2) \tag{4.60}$$

for all $x_1, x_2, y_1, y_2 \ge 0$. $\triangle$

Proof. Observe that

$$(W_l + W_u)(x,y) = \tilde{W}\big([0,x]\times[0,y]\big)$$

so that the left hand side of (4.60) equals

$$\begin{aligned} E\left[\big(\tilde{W}([0,x_1]\times[0,y_1])\big)\big(\tilde{W}([0,x_2]\times[0,y_2])\big)\right] &= \mu\big([0,x_1\wedge x_2]\times[0,y_1\wedge y_2]\big) \\ &= c(x_1\wedge x_2, y_1\wedge y_2). \end{aligned}$$

$\square$

Proof of Theorem 4.5.1

From Lemma 4.6.1 we see that (4.23) follows from

$$\lim_{\vartheta\downarrow 0} P\left\{ \sup_{\substack{0<s\wedge x\le\vartheta \\ 0<s\vee x\le y_0}} \left| \frac{W_l(sx, sy_0)}{(c(sx,sy_0))^{1/2-\delta}} \right| > \varepsilon \right\} = 0 \tag{4.61}$$

and

$$\lim_{\vartheta \downarrow 0} P\left\{ \sup_{\substack{0<s\wedge x\leq\vartheta \\ 0<s\vee x\leq y_0}} \left| \frac{W_u(sx, sy_0)}{(c(sx, sy_0))^{1/2-\delta}} \right| > \varepsilon \right\} = 0 \tag{4.62}$$

to hold jointly. By suitable decompositions of $\{(s, x) \mid 0 < s \wedge x \leq \vartheta, 0 < s \vee x \leq y_0\}$ we show that the processes W_l and W_u are equal in distribution to transformed Brownian sheets.

We first prove (4.62), which is less complicated than to prove (4.61). Let

$$c_u(x, y) := \mu\left(\left\{(w, z) \in [0, \infty)^2 \mid z \leq y, w < \frac{x}{y}\, z\right\}\right),$$

i.e. $c_u(x, y)$ is the μ-measure of the upper triangle of the rectangle $[0, x] \times [0, y]$. The reason that (4.62) is less complicated to prove is that the covariance of W_u substantially simplifies in the case of $y = y_0$ fixed. This is due to the fact that the intersection of two upper triangles equals the smaller of these upper triangle if $y = y_0$, whereas the intersection of two lower triangles is none of the two if $y = y_0$, see also Figure 4.3. We can write the covariance of $(W_u(x, y_0))_{x\geq 0}$ in the simple form

$$E\big[W_u(x_1, y_0)W_u(x_2, y_0)\big] = c_u(x_1 \wedge x_2, y_0).$$

Define $c_{u,y_0}(x) := c_u(x, y_0)$ and let $c^{\leftarrow}_{u,y_0}(t) := \inf\{x \mid c_{u,y_0}(x) \geq t\}$ be the left-continuous generalized inverse of c_{u,y_0}. Then, $W_u(c^{\leftarrow}_{u,y_0}(t), y_0)$ is a Brownian motion for sufficiently small $t \geq 0$, since due to the continuity of c_u,

$$\begin{aligned} E\big[W_u(c^{\leftarrow}_{u,y_0}(t_1), y_0)W_u(c^{\leftarrow}_{u,y_0}(t_2), y_0)\big] &= c_u(c^{\leftarrow}_{u,y_0}(t_1) \wedge c^{\leftarrow}_{u,y_0}(t_2), y_0) \\ &= t_1 \wedge t_2. \end{aligned} \tag{4.63}$$

(Note that since $c^{\leftarrow}_{u,y_0}(t)$ equals ∞ for some $t \geq 0$ if c_{u,y_0} is bounded, we indeed obtain this result only for sufficiently small $t \geq 0$. Note further that this is sufficient for our purposes, since we prove an asymptotic result as $\vartheta \downarrow 0$.)

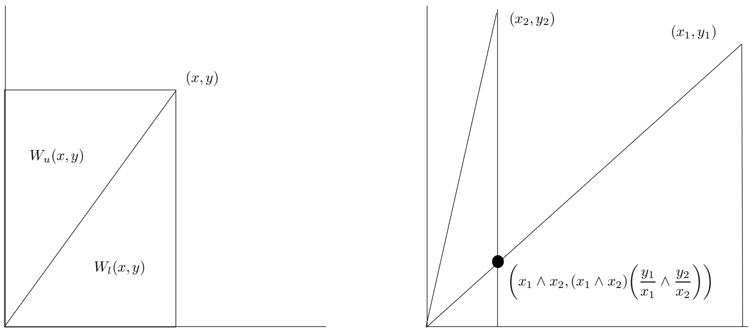


Figure 4.3: Left: the random variables $W_u(x,y)$ and $W_l(x,y)$ are the set indexed process $\tilde{W}$ with the upper and the lower triangle of the rectangle $[0,x]\times[0,y]$ as argument. Right: the covariance of W_l is the μ-measure of the lower triangle of the rectangle $[0,x_1\wedge x_2]\times[0,(x_1\wedge x_2)\big(y_1/x_1\wedge y_2/x_2\big)]$.

Further, observe that

$$\begin{aligned}
&E\big[W_u(s_1x_1,s_1y_0)W_u(s_2x_2,s_2y_0)\big]\\
&=E\Big[\tilde{W}\Big(\big\{(w,z)\mid z\le s_1y_0, w<\frac{x_1}{y_0}\ z\big\}\Big)\tilde{W}\Big(\big\{(w,z)\mid z\le s_2y_0, w<\frac{x_2}{y_0}\ z\big\}\Big)\Big]\\
&=\mu\Big(\big\{(w,z)\mid z\le (s_1\wedge s_2)y_0, w<\frac{x_1\wedge x_2}{y_0}\ z\big\}\Big)\\
&=(s_1\wedge s_2)^{1/\eta}E\big[W_u(x_1,y_0)W_u(x_2,y_0)\big], \qquad (4.64)
\end{aligned}$$

where we used the homogeneity of μ in the last step. We consider the supremum of $|W_u(sx,sy_0)(c(sx,sy_0))^{\delta-1/2}|$ over the set $\{(s,x)\mid 0<x\le\vartheta, 0<s\le y_0\}$ in detail and merely comment on the supremum of that process over $\{(s,x)\mid 0<s\le\vartheta, 0<x\le y_0\}$. Let $\tilde{\vartheta}:=c_{u,y_0}(\vartheta)$. We use the symmetry of

W_u and that $c_u \le c$ to obtain

$$P\left\{\sup_{\substack{0<x\le\vartheta\\0<s\le y_0}}\left|\frac{W_u(sx,sy_0)}{(c(sx,sy_0))^{1/2-\delta}}\right|>\varepsilon\right\}$$

$$\le 2\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}P\left\{\sup_{\substack{x\in(c^{\leftarrow}_{u,y_0}(e^{-(i+1)}\tilde\vartheta),c^{\leftarrow}_{u,y_0}(e^{-i}\tilde\vartheta)]\\s\in(e^{-(j+1)},e^{-j}]}}\frac{W_u(s^\eta x,s^\eta y_0)}{(sc(x,y_0))^{1/2-\delta}}>\varepsilon\right\}$$

$$= 2\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}P\left\{\sup_{\substack{u\in(e^{-1},1]\\s\in(e^{-(j+1)},e^{-j}]}}\frac{W_u(s^\eta c^{\leftarrow}_{u,y_0}(e^{-i}\tilde\vartheta u),s^\eta y_0)}{(sc_{u,y_0}(c^{\leftarrow}_{u,y_0}(e^{-i}\tilde\vartheta u)))^{1/2-\delta}}>\varepsilon\right\}. \quad (4.65)$$

We combine (4.63) and (4.64) to see that for sufficiently small $\tilde\vartheta$ the process $\big(W_u(s^\eta c^{\leftarrow}_{u,y_0}(e^{-i}\tilde\vartheta u),s^\eta y_0)\big)_{s\ge0,u\in(e^{-1},1]}$ has covariance function

$$E\big[W_u\big(s_1^\eta c^{\leftarrow}_{u,y_0}(e^{-i}\tilde\vartheta u_1),s_1^\eta y_0\big)W_u\big(s_2^\eta c^{\leftarrow}_{u,y_0}(e^{-i}\tilde\vartheta u_2),s_2^\eta y_0\big)\big]$$
$$=\tilde\vartheta e^{-i}(s_1\wedge s_2)(u_1\wedge u_2).$$

This means that $\big(W_u(s^\eta c^{\leftarrow}_{u,y_0}(\tilde\vartheta e^{-i}u),s^\eta y_0)\big)_{s\ge0,u\in(e^{-1},1]}$ is a scaled Brownian sheet. More precisely,

$$\big(W_u(s^\eta c^{\leftarrow}_{u,y_0}(\tilde\vartheta e^{-i}u),s^\eta y_0)\big)_{s\ge0,u\in(e^{-1},1]}\overset{D}{=}\big(\tilde\vartheta e^{-i}\big)^{1/2}\mathcal{B}_1$$

with $\mathcal{B}_1$ denoting a standard Brownian sheet. We use this fact, the continuity of c_u and that $0<\delta<1/2$ to obtain from (4.65)

$$P\left\{\sup_{\substack{0<x\le\vartheta\\0<s\le y_0}}\left|\frac{W_u(sx,sy_0)}{(c(sx,sy_0))^{1/2-\delta}}\right|>\varepsilon\right\}$$

$$\le 2\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}P\left\{\sup_{\substack{u\in(e^{-1},1]\\s\in(e^{-(j+1)},e^{-j}]}}\big(\tilde\vartheta e^{-i}\big)^{1/2}\mathcal{B}_1(s,u)>\varepsilon\inf_{\substack{u\in(e^{-1},1]\\s\in(e^{-(j+1)},e^{-j}]}}\big(se^{-i}\tilde\vartheta u\big)^{1/2-\delta}\right\}$$

$$\le 2\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}P\left\{\sup_{\substack{u\in(e^{-1},1]\\s\in(e^{-(j+1)},e^{-j}]}}\mathcal{B}_1(s,u)>K\tilde\vartheta^{-\delta}e^{-j(1/2-\delta)}e^{i\delta}\right\},$$

where K (and K_1 below) is a generic constant (i.e. it may vary from line to line), which may depend on δ (or η), but not on i, j or ϑ. The analogue of the reflection principle which is valid for Brownian sheet, (see e.g. Proposition 3.7 of Walsh (1986)) leads to

$$
\begin{aligned}
&P\left\{\sup_{\substack{0<x\le\vartheta\\0<s\le y_0}}\left|\frac{W_u(sx, sy_0)}{(c(sx, sy_0))^{1/2-\delta}}\right| > \varepsilon\right\}\\
&\le K_1\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} P\left\{G\Big([e^{-(j+1)}, e^{-j}]\times[e^{-1}, 1]\Big) > K\tilde{\vartheta}^{-\delta}e^{-j(1/2-\delta)}e^{i\delta}\right\},
\end{aligned}
$$

where G is Gaussian white noise on $\mathbb{R}^2$, i.e. G is a random set function defined on the Borel sets $A, B \in \mathbb{B}^2$ which have finite Lebesgue measure such that (i) $W(A)$ is a centered normal random variable with variance equal to the Lebesgue measure of A and (ii) from $A \cap B = \emptyset$ follows that $W(A)$ and $W(B)$ are independent. Hence, the random variable $G\big([e^{-(j+1)}, e^{-j}] \times [e^{-1}, 1]\big)$ is centered normal with variance $(e^{-j} - e^{-(j+1)})(1 - e^{-1})$, i.e.

$$
\begin{aligned}
&P\left\{\sup_{\substack{0<x\le\vartheta\\0<s\le y_0}}\left|\frac{W_u(sx, sy_0)}{(c(sx, sy_0))^{1/2-\delta}}\right| > \varepsilon\right\}\\
&\qquad\le K_1\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\left(1-\Phi\left(K\tilde{\vartheta}^{-\delta}\frac{e^{-j(1/2-\delta)}}{(e^{-j}-e^{-(j+1)})^{1/2}}e^{i\delta}\right)\right)\\
&\qquad\le K_1\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\left(1-\Phi\left(K\tilde{\vartheta}^{-\delta}e^{j\delta}e^{i\delta}\right)\right).
\end{aligned}
$$

Mill's ratio (or rather the pertaining inequality, see e.g. Resnick (1992), p. 487) yields

$$
\begin{aligned}
&P\left\{ \sup_{\substack{0<x\le\vartheta\\0<s\le y_0}} \left| \frac{W_u(sx,sy_0)}{(c(sx,sy_0))^{1/2-\delta}} \right| > \varepsilon \right\}\\
&\le K_1 \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \tilde{\vartheta}^{\delta} e^{-j\delta} e^{-i\delta} \exp\Big(- K\tilde{\vartheta}^{-2\delta} e^{2j\delta} e^{2i\delta}\Big)\\
&\le K_1 \tilde{\vartheta}^{\delta} \sum_{i=0}^{\infty} e^{-i\delta} \sum_{j=0}^{\infty} e^{-j\delta}\\
&\le \frac{K_1\tilde{\vartheta}^{\delta}}{(1-e^{-\delta})^2},
\end{aligned}
$$

so that $\vartheta \downarrow 0$ gives the required result.

When considering the supremum of $|W_u(sx,sy_0)(c(sx,sy_0))^{\delta-1/2}|$ over the set $\{(s,x) \mid 0 < s \le \vartheta, 0 < x \le y_0\}$, the decomposition

$$
\begin{aligned}
&\{(s,x) \mid 0 < s \le \vartheta, 0 < x \le y_0\}\\
&\quad = \bigcup_{i=0}^{\infty}\bigcup_{j=0}^{\infty} \Big\{ (s,x) \mid s \in \vartheta\big(e^{-(j+1)}, e^{-j}\big],\\
&\qquad\qquad x \in y_0\big(\overleftarrow{c}_{u,y_0}(e^{-(i+1)} c_{u,y_0}(1)), \overleftarrow{c}_{u,y_0}(e^{-i} c_{u,y_0}(1))\big] \Big\}
\end{aligned}
$$

leads with analogous arguments to the same result.

We now prove (4.61). The structure of the proof is very similar to the one of (4.62). However, we need a more sophisticated decomposition of $\big\{(s,x) \mid 0 < s \wedge x \le \vartheta, 0 < s \vee x \le y_0\big\}$, since the covariance of W_l does not substantially simplify (as the covariance of W_u does) in the case of $y = y_0$ fixed. Let

$$
c_l(x,y) := \mu\left(\left\{(w,z) \in [0,\infty)^2 \mid w \le x, z \le \frac{y}{x}\, w\right\}\right),
$$

i.e. $c_l(x,y)$ is the μ-measure of the lower triangle of the rectangle $[0,x]\times[0,y]$. Define $c_{l,1}(\cdot) = c_l(1,\cdot)$. The covariance of W_l is the μ-measure of the lower triangle of the rectangle $[0, x_1\wedge x_2] \times [0, (x_1\wedge x_2)\big(y_1/x_1 \wedge y_2/x_2\big)]$ as indicated in the right hand plot of Figure 4.3, i.e. with the homogeneity of μ,

$$
\begin{aligned}
E\big[W_l(x_1,y_1)W_l(x_2,y_2)\big] &= c_l\left(x_1\wedge x_2,(x_1\wedge x_2)\left(\frac{y_1}{x_1}\wedge\frac{y_2}{x_2}\right)\right)\\
&= (x_1\wedge x_2)^{1/\eta}c_{l,1}\left(\frac{y_1}{x_1}\wedge\frac{y_2}{x_2}\right). \qquad (4.66)
\end{aligned}
$$

In view of this formula, we are interested in a decomposition that gives an easy expression for the first argument sx of W_l in (4.61). Foremost, however, the ratio y_0/x of the first and the second argument should be handy when used as an argument of $c_{l,1}$. Over the set $\big\{(s,x)\mid 0<s\wedge x\le\vartheta, 0<s\vee x\le y_0\big\}$ we have that $sx\in(0,\vartheta]$ and $y_0/x\ge 1$, so that

$$
\begin{aligned}
&\big\{(s,x)\mid 0<s\wedge x\le\vartheta, 0<s\vee x\le y_0\big\}\\
&\qquad\subset\bigcup_{i=0}^{\infty}\Big\{(s,x)\mid sx\in\vartheta\big(e^{-(i+1)},e^{-i}\big],\frac{y_0}{x}\ge 1\Big\}.
\end{aligned}
$$

Hence, first assuming that $c_{l,1}(\infty)=\infty$ and observing that $c_{l,1}^{\leftarrow}(1)\ge 1$ for the left-continuous generalized inverse $c_{l,1}^{\leftarrow}(t):=\inf\{y\mid c_{l,1}(y)\ge t\}$, we have that

$$
\begin{aligned}
&\big\{(s,x)\mid 0<s\wedge x\le\vartheta, 0<s\vee x\le y_0\big\}\\
&\subset\bigcup_{i=0}^{\infty}\Big\{(s,x)\mid sx\in\vartheta\big(e^{-(i+1)},e^{-i}\big],\frac{y_0}{x}\in\big[1,c_{l,1}^{\leftarrow}(1)\big)\Big\} \qquad (4.67)\\
&\quad\cup\bigcup_{i=0}^{\infty}\bigcup_{j=1}^{\infty}\Big\{(s,x)\mid sx\in\vartheta\big(e^{-(i+1)},e^{-i}\big],\frac{y_0}{x}\in\big[c_{l,1}^{\leftarrow}(j),c_{l,1}^{\leftarrow}(j+1)\big)\Big\}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&P\Bigg\{\sup_{\substack{0<s\wedge x\leq\vartheta\\0<s\vee x\leq y_0}}\frac{|W_l(sx,sy_0)|}{(c(sx,sy_0))^{1/2-\delta}}>\varepsilon\Bigg\}\\
&\leq 2\sum_{i=0}^{\infty}P\Bigg\{\sup_{\substack{sx\in\vartheta(e^{-(i+1)},e^{-i}]\\y_0/x\in[1,c^{\leftarrow}_{l,1}(1))}}\frac{W_l(sx,sy_0)}{(c(sx,sy_0))^{1/2-\delta}}>\varepsilon\Bigg\}\\
&\quad+2\sum_{i=0}^{\infty}\sum_{j=1}^{\infty}P\Bigg\{\sup_{\substack{sx\in\vartheta(e^{-(i+1)},e^{-i}]\\y_0/x\in[c^{\leftarrow}_{l,1}(j),c^{\leftarrow}_{l,1}(j+1))}}\frac{W_l(sx,sy_0)}{(c(sx,sy_0))^{1/2-\delta}}>\varepsilon\Bigg\}\\
&\leq 2\sum_{i=0}^{\infty}P\Bigg\{\sup_{\substack{\lambda\in(e^{-1},1]\\\tau\in[c_{l,1}(1),1)}}W_l\big(\vartheta e^{-i}\lambda,\vartheta e^{-i}\lambda c^{\leftarrow}_{l,1}(\tau)\big)\\
&\qquad\qquad\qquad>\varepsilon\inf_{\substack{\lambda\in(e^{-1},1]\\\tau\in[c_{l,1}(1),1)}}\big((\vartheta e^{-i}\lambda)^{1/\eta}c(1,c^{\leftarrow}_{l,1}(\tau))\big)^{1/2-\delta}\Bigg\}\\
&\quad+2\sum_{i=0}^{\infty}\sum_{j=1}^{\infty}P\Bigg\{\sup_{\substack{\lambda\in(e^{-1},1]\\\tau\in[j,j+1)}}W_l\big(\vartheta e^{-i}\lambda,\vartheta e^{-i}\lambda c^{\leftarrow}_{l,1}(\tau)\big)\\
&\qquad\qquad\qquad>\varepsilon\inf_{\substack{\lambda\in(e^{-1},1]\\\tau\in[j,j+1)}}\big((\vartheta e^{-i}\lambda)^{1/\eta}c(1,c^{\leftarrow}_{l,1}(\tau))\big)^{1/2-\delta}\Bigg\}.
\end{aligned}
\tag{4.68}
$$

We continue similar as in the proof of (4.62). According to (4.66) and the continuity of c_l, the process $W_l\big(\vartheta e^{-i}\lambda,\vartheta e^{-i}\lambda c^{\leftarrow}_{l,1}(\tau)\big)_{\lambda,\tau\geq0} =: \big(Z_i(\lambda,\tau)\big)_{\lambda,\tau\geq0}$ has covariance function

$$E\big[Z_i(\lambda_1,\tau_1)Z_i(\lambda_2,\tau_2)\big]=(\vartheta e^{-i})^{1/\eta}(\lambda_1\wedge\lambda_2)^{1/\eta}(\tau_1\wedge\tau_2).$$

This means that Z_i is a version of a transformed Brownian sheet. More precisely,

$$\big(Z_i(\lambda,\tau)\big)_{\lambda,\tau\geq0}\overset{D}{=}\big(\vartheta e^{-i}\big)^{1/2\eta}\big(\mathcal{B}_1(\lambda^{1/\eta},\tau)\big)_{\lambda,\tau\geq0}.$$

We use this fact, the relation $c_{l,1}(\cdot) \le c(1,\cdot)$ and that $0<\delta<1/2$ to continue (4.68) to

$$
\begin{aligned}
&P\Bigg\{ \sup_{\substack{0<s\wedge x\le\vartheta \\ 0<s\vee x\le y_0}} \frac{|W_l(sx,sy_0)|}{(c(sx,sy_0))^{1/2-\delta}} > \varepsilon \Bigg\} \\
&\le 2\sum_{i=0}^{\infty} P\Bigg\{ \sup_{\substack{\lambda\in(e^{-1},1] \\ \tau\in[c_{l,1}(1),1)}} \mathcal{B}_1(\lambda^{1/\eta},\tau) > K(\vartheta e^{-i})^{-1/2\eta}(\vartheta e^{-(i+1)})^{\frac{1/2-\delta}{\eta}} \Bigg\} \\
&\quad +2\sum_{i=0}^{\infty}\sum_{j=1}^{\infty} P\Bigg\{ \sup_{\substack{\lambda\in(e^{-1},1] \\ \tau\in[j,j+1)}} \mathcal{B}_1(\lambda^{1/\eta},\tau) \\
&\qquad\qquad > K(\vartheta e^{-i})^{-1/2\eta}(\vartheta e^{-(i+1)})^{\frac{1/2-\delta}{\eta}}(c(1,\overleftarrow{c_{l,1}}(j)))^{1/2-\delta} \Bigg\} \\
&= 2\sum_{i=0}^{\infty} P\Bigg\{ \sup_{\substack{t\in(e^{-1/\eta},1] \\ \tau\in[c_{l,1}(1),1)}} \mathcal{B}_1(t,\tau) > Ke^{i\delta/\eta}\vartheta^{-\delta/\eta} \Bigg\} \\
&\quad +2\sum_{i=0}^{\infty}\sum_{j=1}^{\infty} P\Bigg\{ \sup_{\substack{t\in(e^{-1/\eta},1] \\ \tau\in[j,j+1)}} \mathcal{B}_1(t,\tau) > Ke^{i\delta/\eta}\vartheta^{-\delta/\eta}(c(1,\overleftarrow{c_{l,1}}(j)))^{1/2-\delta} \Bigg\} \\
&\le K_1\sum_{i=0}^{\infty} P\Big\{ G\big((e^{-1/\eta},1]\times[c_{l,1}(1),1)\big) > Ke^{i\delta/\eta}\vartheta^{-\delta/\eta} \Big\} \\
&\quad +K_1\sum_{i=0}^{\infty}\sum_{j=1}^{\infty} P\Big\{ G\big((e^{-1/\eta},1]\times[j,j+1)\big) > Ke^{i\delta/\eta}\vartheta^{-\delta/\eta} j^{1/2-\delta} \Big\}.
\end{aligned}
$$

Similar as in the proof of (4.62), we argue that since the random variables $G\big((e^{-1/\eta},1]\times[c_{l,1}(1),1)\big)$ and $G\big((e^{-1/\eta},1]\times[j,j+1)\big)$ are centered normal with variance $(1-e^{-1/\eta})(1-c_{l,1}(1))$ and $1-e^{-1/\eta}$, respectively, we obtain by

Mill's inequality and for sufficiently small ϑ (e.g. such that $K\vartheta^{-2\delta/\eta} > 1$),

$$\begin{aligned}
P\Bigg\{ \sup_{\substack{0<s\wedge x\le\vartheta\\ 0<s\vee x\le y_0}} \frac{|W_l(sx, sy_0)|}{(c(sx, sy_0))^{1/2-\delta}} > \varepsilon \Bigg\} \\
&= K_1 \sum_{i=0}^{\infty} \Big(1 - \Phi\Big(K e^{i\delta/\eta}\vartheta^{-\delta/\eta}\Big)\Big) \\
&\quad + K_1 \sum_{i=0}^{\infty}\sum_{j=1}^{\infty} \Big(1 - \Phi\Big(K e^{i\delta/\eta}\vartheta^{-\delta/\eta} j^{1/2-\delta}\Big)\Big) \\
&\le K_1 \sum_{i=0}^{\infty} \vartheta^{\delta/\eta} e^{-i\delta/\eta} \exp\Big(-K\vartheta^{-2\delta/\eta} e^{2i\delta/\eta}\Big) \\
&\quad + K_1 \sum_{i=0}^{\infty}\sum_{j=1}^{\infty} \vartheta^{\delta/\eta} e^{-i\delta/\eta} j^{-(1/2-\delta)} \exp\Big(-K\vartheta^{-2\delta/\eta} e^{2i\delta/\eta} j^{1-2\delta}\Big) \\
&\le K_1 \vartheta^{\delta/\eta} \sum_{i=0}^{\infty} e^{-i\delta/\eta} \\
&\quad + K_1 \vartheta^{\delta/\eta} \sum_{i=0}^{\infty} e^{-i\delta/\eta} \sum_{j=1}^{\infty} j^{-(1/2-\delta)} \exp\Big(-K\vartheta^{-2\delta/\eta} e^{2i\delta/\eta} j^{1-2\delta}\Big) \\
&\le K_1 \vartheta^{\delta/\eta} \sum_{i=0}^{\infty} e^{-i\delta/\eta} \Big(1 + \sum_{j=1}^{\infty} j^{-(1/2-\delta)} \exp\Big(-j^{1-2\delta}\Big)\Big) \\
&\le K_1 \vartheta^{\delta/\eta} \sum_{i=0}^{\infty} e^{-i\delta/\eta} \\
&\le \vartheta^{\delta/\eta} \frac{K_1}{1 - e^{-\delta/\eta}},
\end{aligned}$$

so that $\vartheta \downarrow 0$ gives the required result. To complete the proof of (4.61), it remains to treat the (easier) case $c_{l,1}(\infty) < \infty$. Here, a rougher decomposition is sufficient. By above arguments,

$$
\begin{aligned}
P\Bigg\{ & \sup_{\substack{0<s\wedge x\leq\vartheta\\ 0<s\vee x\leq y_0}} \frac{|W_l(sx,sy_0)|}{(c(sx,sy_0))^{1/2-\delta}} > \varepsilon\Bigg\} \\
&\leq 2\sum_{i=0}^{\infty} P\Bigg\{ \sup_{\substack{sx\in\vartheta(e^{-(i+1)},e^{-i}]\\ y_0/x\geq y_0}} W_l(sx,sy_0) \\
&\qquad\qquad > \varepsilon \inf_{\substack{sx\in(e^{-(i+1)}\vartheta,e^{-i}\vartheta]\\ y_0/x\geq y_0}} \big((sx)^{1/\eta}c(1,y_0/x)\big)^{1/2-\delta}\Bigg\} \\
&= 2\sum_{i=0}^{\infty} P\Bigg\{ \sup_{\substack{\lambda\in(e^{-1},1]\\ \tau\geq y_0}} W_l(\vartheta e^{-i}\lambda,\vartheta e^{-i}\lambda\tau) \\
&\qquad\qquad > \varepsilon(e^{-(i+1)}\vartheta)^{(1/2-\delta)/\eta}(c(1,y_0))^{1/2-\delta}\Bigg\}
\end{aligned}
\tag{4.69}
$$

The remaining part is as before, i.e. we can conclude

$$
\big(W_l(\vartheta e^{-i}\lambda,\vartheta e^{-i}\lambda\tau)\big)_{\lambda,\tau>0} \overset{D}{=} \big(\vartheta e^{-i}\big)^{1/2\eta}\big(\mathcal{B}_1(\lambda^{1/\eta},c_{u,1}(\tau))\big)_{\lambda,\tau>0},
$$

so that (4.69) is continued to

$$
\begin{aligned}
P\Bigg\{ & \sup_{\substack{0<s\wedge x\leq\vartheta\\ 0<s\vee x\leq 1}} \frac{|W_l(sx,sy_0)|}{(c(sx,sy_0))^{1/2-\delta}} > \varepsilon\Bigg\} \\
&\leq K_1\sum_{i=0}^{\infty} P\Big\{G\Big((e^{-1/\eta},1]\times[c_{l,1}(y_0),c_{l,1}(\infty)]\Big) \\
&\qquad\qquad > \varepsilon e^{-(1/2-\delta)/\eta}e^{i\delta/\eta}\vartheta^{-\delta/\eta}(c(1,y_0))^{1/2-\delta}\Big\}.
\end{aligned}
$$

Mill's inequality shows that this converges to 0 as $\vartheta\downarrow 0$.

Proof of Lemma 4.5.4

Note that

$$P\left\{\sup_{\substack{0<s\wedge x\le\vartheta\\0<s\vee x\le 1}} m^{1/2}(c(sx,sy_0))^{1/2+\delta}\left|\frac{\log s}{\hat\eta_n}-\frac{\log s}{\eta}\right|>\varepsilon\right\}$$

$$= P\left\{m^{1/2}\left|\frac{\hat\eta_n-\eta}{\hat\eta_n\eta}\right|\sup_{\substack{0<s\wedge x\le\vartheta\\0<s\vee x\le 1}} s^{(1/2+\delta)/\eta}(c(x,y_0))^{1/2+\delta}(-\log s)>\varepsilon\right\}. \tag{4.70}$$

The function $h(s) := -s^{(1/2+\delta)/\eta}\log s$ is strictly increasing on $(0, e^{-\eta/(1/2+\delta)})$ and strictly decreasing on $(e^{-\eta/(1/2+\delta)}, 1]$. Hence, h is maximal on $(0,\vartheta]$ at $s_\vartheta = e^{-\eta/(1/2+\delta)}\wedge\vartheta$. Thus,

$$\sup_{\substack{0<s\le\vartheta\\0<x\le 1}} -s^{(1/2+\delta)/\eta}(\log s)(c(x,y_0))^{1/2+\delta} = h(s_\vartheta)(c(1,y_0))^{1/2+\delta} =: H_1(\vartheta)$$

and

$$\sup_{\substack{0<x\le\vartheta\\0<s\le 1}} -s^{(1/2+\delta)/\eta}(\log s)(c(x,y_0))^{1/2+\delta} = h\big(e^{-\eta/(1/2+\delta)}\big)(c(\vartheta,y_0))^{1/2+\delta} =: H_2(\vartheta).$$

Now define the function $H := \max\big(H_1,H_2\big)$ and observe that $H(\vartheta)\to 0$, as $\vartheta\downarrow 0$. (Note that $h(s)\to 0$ as $s\downarrow 0$.) Hence, the right hand side of (4.70) equals

$$P\left\{\left|\frac{m^{1/2}(\hat\eta_n-\eta)}{\hat\eta_n\eta}\right|>\frac{\varepsilon}{H(\vartheta)}\right\}. \tag{4.71}$$

According to Theorem 3.1.2, $m^{1/2}(\hat\eta_n-\eta)$ is asymptotically centered normal with variance $\sigma^2_{\hat\eta_n}$. Hence, $\hat\eta_n\eta\to\eta^2$ in probability and Slutsky's theorem yields that $m^{1/2}(\hat\eta_n-\eta)/(\hat\eta_n\eta)$ converges to a centered normal random variable with variance $\sigma^2_{\hat\eta_n}/\eta^4$. This means that (4.71) converges to $2\big(1-\Phi(\varepsilon\eta^2/(\sigma_{\hat\eta_n}H(\vartheta))\big)$ as $n\to\infty$ and the assertion follows.

Proof of Theorem 4.5.5

We start with some steps that simplify the assertion to prove. Note that

$$\left|\log\frac{\tilde S_n^*(sx,sy_0)}{\tilde S_n^*(x,y_0)}-\frac{\log s}{\hat\eta_n}\right| \le \left|\log\frac{\tilde S_n^*(sx,sy_0)}{\tilde S_n^*(x,y_0)}-\frac{\log s}{\eta}\right| + \left|\frac{\log s}{\hat\eta_n}-\frac{\log s}{\eta}\right|$$

and that

$$\lim_{\vartheta\downarrow 0}\limsup_{n\to\infty} P\left\{\sup_{\substack{0<s\wedge x\le\vartheta\\ 0<s\vee x\le y_0}} m^{1/2}c(sx,sy_0)\left|\frac{\log s}{\hat\eta_n}-\frac{\log s}{\eta}\right|>\varepsilon\right\}=0$$

follows from Lemma 4.5.4. Hence, it is sufficient to prove

$$\begin{aligned}\lim_{\vartheta\downarrow 0}\limsup_{n\to\infty} P\Bigg\{&\sup_{\substack{0<s\wedge x\le\vartheta\\ 0<s\vee x\le y_0}} m^{1/2}\min\left(c(sx,sy_0),(sx\wedge sy_0)^{1/2}(\ell_2(sx\wedge sy_0))^{-1/2-\delta}\right)\\ &\times \mathbb{1}_{(0,\infty)}(\tilde S_n^*(sx,sy_0))\left|\log\frac{\tilde S_n^*(sx,sy_0)}{\tilde S_n^*(x,y_0)}-\frac{\log s}{\eta}\right|>\varepsilon\Bigg\}=0.\end{aligned}\tag{4.72}$$

Next recall from (4.17) that

$$\left(\frac{\tilde S_n^*(sx,sy_0)}{\tilde S_n^*(x,y_0)}\right)_{s,x\in(0,1]\times(0,\infty)} \overset{D}{=} \left(\frac{S\left(U_{\lceil sx(n+1)/T^{(n)}_{n-m:n}\rceil-1:n},V_{\lceil sy_0(n+1)/T^{(n)}_{n-m:n}\rceil-1:n}\right)}{S\left(U_{\lceil x(n+1)/T^{(n)}_{n-m:n}\rceil-1:n},V_{\lceil y_0(n+1)/T^{(n)}_{n-m:n}\rceil-1:n}\right)}\right)_{s,x\in(0,1]\times(0,\infty)}.$$

As an empirical counterpart of $Q(u,v)=P\{U<u,V<v\}$ we define[8]

$$Q_n(u,v):=\frac{1}{n}S(u,v)=\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{U_i\le u,V_i\le v\}}.$$

Further, recall from the proof of Theorem 3.2.7 and Corollary 3.2.8 that $T^{(n)}_{n-m:n}\overset{D}{=}Q_n^*(1)$, where Q_n^* is a version of the tail empirical quantile process Q_n, cf. (3.8). Hence, by (3.56) and Corollary 3.1.1, we have $Q_n^*(1)=$

[8] To avoid confusion, note that the function $Q_n(\cdot,\cdot)$ (and versions $Q_n^*(\cdot,\cdot)$, respectively) is different from the tail empirical quantile function $Q_n(\cdot)$ (and versions $Q_n^*(\cdot)$, respectively), defined in (3.8). Further, the notation Q_n (and versions Q_n^*, respectively) always refers to the tail empirical quantile process $\left(Q_n(t)\right)_{0<t\le 1}$.

$F_T^{\leftarrow}(1-m/n)\big(1+m^{-1/2}\eta W(1,1)+o(m^{-1/2})\big)$, so that for all $\varepsilon>0$ and all $0\le x\le y_0$,

$$\tilde{x}:=x\frac{F_T^{\leftarrow}(1-m/n)}{Q_n^*(1)}\quad\in[x(1-\varepsilon),x(1+\varepsilon)]$$

$$\tilde{y}:=y_0\frac{F_T^{\leftarrow}(1-m/n)}{Q_n^*(1)}\quad\in[y_0(1-\varepsilon),y_0(1+\varepsilon)]$$

with probability tending to 1. Moreover, note that $c(s\tilde{x},s\tilde{y})=\big(F_T^{\leftarrow}(1-m/n)/Q_n^*(1)\big)^{1/\eta}c(sx,sy_0)\to c(sx,sy_0)$ as $n\to\infty$. Hence, to obtain (4.72), it suffices to prove

$$\lim_{\vartheta\downarrow 0}\limsup_{n\to\infty}P\Bigg\{\sup_{(s,\tilde{x},\tilde{y})\in I}m^{1/2}\min\big(c(s\tilde{x},s\tilde{y}),(s\tilde{x}\wedge s\tilde{y})^{1/2}(\ell_2(s\tilde{x}\wedge s\tilde{y}))^{-1/2-\delta}\big)$$

$$\times 1\!\!1_{(0,\infty)}\big(Q_n^*\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)\big)\left|\log\frac{Q_n^*\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)}{Q_n^*\big(U_{h_n(\tilde{x}):n},V_{h_n(\tilde{y}):n}\big)}-\frac{\log s}{\eta}\right|>\varepsilon\Bigg\}=0, \tag{4.73}$$

where, in order to ease notation, $I:=\{(s,\tilde{x},\tilde{y})\mid(s\wedge\tilde{x},s\vee\tilde{x},\tilde{y})\in(0,\vartheta]\times(0,y_0]\times[y_0/2,2y_0]\}$, say, and

$$h_n(\tau):=\lceil\tau(n+1)/F_T^{\leftarrow}(1-m/n)\rceil-1.$$

Finally, to prove (4.73), it is sufficient to show that

$$\lim_{\vartheta\downarrow 0}\limsup_{n\to\infty}P\Bigg\{\sup_{(s,\tilde{x},\tilde{y})\in I}m^{1/2}c(s\tilde{x},s\tilde{y})1\!\!1_{(0,\infty)}\big(Q_n^*\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)\big)$$

$$\times\left|\log\frac{Q_n^*\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)}{Q_n^*\big(U_{h_n(\tilde{x}):n},V_{h_n(\tilde{y}):n}\big)}-\log\frac{Q\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)}{Q\big(U_{h_n(\tilde{x}):n},V_{h_n(\tilde{y}):n}\big)}\right|>\varepsilon\Bigg\}=0 \tag{4.74}$$

and

$$\lim_{\vartheta\downarrow 0}\limsup_{n\to\infty}P\Bigg\{\sup_{(s,\tilde{x},\tilde{y})\in I}m^{1/2}\min\big(c(s\tilde{x},s\tilde{y}),(s\tilde{x}\wedge s\tilde{y})^{1/2}(\ell_2(s\tilde{x}\wedge s\tilde{y}))^{-1/2-\delta}\big)$$

$$\times 1\!\!1_{(0,\infty)}\big(Q_n^*\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)\big)\left|\log\frac{Q\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)}{Q\big(U_{h_n(\tilde{x}):n},V_{h_n(\tilde{y}):n}\big)}-\frac{\log s}{\eta}\right|>\varepsilon\Bigg\}=0 \tag{4.75}$$

jointly hold. This, starting with (4.74), will be done subsequently.

With $b = 1/n$, Inequality 10.3.2 (8) of Shorack and Wellner (1986) yields that for all $\lambda \geq 1$,

$$P\left\{ \min_{1 \leq h_n(s\tilde{x}) \leq n} nU_{h_n(s\tilde{x}):n}/h_n(s\tilde{x}) \leq 1/\lambda \right\} \leq \exp(-\tilde{h}(1/\lambda))$$

with $\tilde{h}(\tau) = \tau - \log \tau - 1 \to \infty$ as $\tau \to \infty$ or $\tau \to 0$. Further, again with $b = 1/n$, Inequality 10.3.2 (7) of the same reference yields

$$P\left\{ \max_{1 \leq h_n(s\tilde{x}) \leq n} nU_{h_n(s\tilde{x}):n}/h_n(s\tilde{x}) \geq \lambda \right\} \leq \exp(-\tilde{h}(\lambda)).$$

Hence, for every $\delta_1 > 0$ there exists $\lambda \geq 1$ such that

$$P\left\{ \frac{1}{\lambda} \leq \min_{1 \leq h_n(s\tilde{x}) \leq n} \frac{nU_{h_n(s\tilde{x}):n}}{h_n(s\tilde{x})} \leq \max_{1 \leq h_n(s\tilde{x}) \leq n} \frac{nU_{h_n(s\tilde{x}):n}}{h_n(s\tilde{x})} \leq \lambda \right\} > 1 - \delta_1. \quad (4.76)$$

Of course, an analogous relation is valid for $V_{h_n(s\tilde{y}):n}$.

Further, note that

$$h_n(s\tilde{x}) < 1 \textbf{ or } h_n(s\tilde{y}) < 1 \quad \Rightarrow \quad Q_n^*(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}) = 0$$

$$h_n(s\tilde{x}) \geq 1 \quad \Rightarrow \quad \frac{h_n(s\tilde{x})}{n} \underbrace{F_T^{\leftarrow}(1 - m/n)}_{=:F_T^{\leftarrow}} \geq \frac{s\tilde{x}}{2}$$

so that if $h_n(s\tilde{x}) \geq 1$ and $nU_{h_n(s\tilde{x}):n}/h_n(s\tilde{x}) \geq 1/\lambda$ (and the analogous inequalities for $h_n(s\tilde{y})$ and $V_{h_n(s\tilde{y}):n}$),

$$\begin{aligned} c(s\tilde{x}, s\tilde{y}) &= (2\lambda)^{1/\eta}\, c\big(s\tilde{x}/(2\lambda), s\tilde{y}/(2\lambda)\big) \\ &\leq (2\lambda)^{1/\eta}\, c\left(\frac{1}{\lambda} \frac{h_n(s\tilde{x})}{n} F_T^{\leftarrow}, \frac{1}{\lambda} \frac{h_n(s\tilde{y})}{n} F_T^{\leftarrow} \right) \\ &\leq (2\lambda)^{1/\eta}\, c\big(F_T^{\leftarrow} U_{h_n(s\tilde{x}):n}, F_T^{\leftarrow} V_{h_n(s\tilde{y}):n}\big) \\ &\leq 2C(2\lambda)^{1/\eta}\, \frac{n}{m}\, Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) \end{aligned} \quad (4.77)$$

for sufficiently large n, where the last inequality follows from (4.24) with $t = 1/F_T^{\leftarrow}$.

Hence, by (4.76) and (4.77) we obtain that for all $\delta_1 > 0$ there exists $\lambda \geq 1$ such that

$$\begin{aligned}
&\limsup_{n\to\infty} P\Bigg\{ \sup_{(s,\tilde{x},\tilde{y})\in I} \sqrt{m}\, c(s\tilde{x}, s\tilde{y}) \mathbb{1}_{(0,\infty)}\big(Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)\big) \\
&\qquad \times \Big| \log Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) - \log Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)\Big| > \varepsilon \Bigg\} \\
&\leq \delta_1 + \limsup_{n\to\infty} P\Bigg\{ \sup_{(s,\tilde{x},\tilde{y})\in I} 2C\ (2\lambda)^{1/\eta} \frac{n}{\sqrt{m}}\, Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) \\
&\qquad \times \mathbb{1}_{(0,\infty)}\big(Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)\big) \\
&\qquad \times \Big| \log Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) - \log Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)\Big| > \varepsilon \Bigg\}.
\end{aligned} \tag{4.78}$$

We first consider the case $Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) \leq Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)$ and define the sets

$$\begin{aligned}
I_{1,n} &:= \Bigg\{ (s,\tilde{x},\tilde{y}) \in I \ \Big|\, Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) \\
&\qquad\qquad \leq Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) \leq \varepsilon \frac{\sqrt{m}}{2Cn\log n} \Bigg\} \\
I_{2,n} &:= \Bigg\{ (s,\tilde{x},\tilde{y}) \in I \ \Big|\, \max\left(Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big), \varepsilon \frac{\sqrt{m}}{Cn} \right) \\
&\qquad\qquad \leq Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) \Bigg\} \\
I_{3,n} &:= \Bigg\{ (s,\tilde{x},\tilde{y}) \in I \ \Big|\ \max\left(Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big), \varepsilon \frac{\sqrt{m}}{2Cn\log n} \right) \\
&\qquad\qquad < Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) \leq \varepsilon \frac{\sqrt{m}}{Cn} \Bigg\}.
\end{aligned}$$

We consider the supremum over these three sets separately.

For $(s,\tilde{x},\tilde{y}) \in I_{1,n}$, the rough bounds $Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) \le 1$ and $Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) \ge 1/n$ if $Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) > 0$ suffice to see that

$$
\begin{aligned}
\sqrt{m}\, 2C\, \frac{n}{m}\, Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)\, \mathbb{1}_{(0,\infty)}\Big(Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)\Big)& \\
\times \Big| \log Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) - \log Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)\Big|& \\
\le \frac{2Cn}{\sqrt{m}}\, Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) \log n \le \varepsilon.&
\end{aligned}
$$

For $(s,\tilde{x},\tilde{y}) \in I_{2,n}$, note that $\varepsilon\sqrt{m}/\Big(2CnQ\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)\Big) \le 1/2$, so that by the mean value theorem,

$$
\begin{aligned}
&P\Bigg\{ \sup_{(s,\tilde{x},\tilde{y})\in I_{2,n}} \sqrt{m}\, 2C\frac{n}{m}\, Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) \mathbb{1}_{(0,\infty)}\Big(Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)\Big) \\
&\qquad \times \Big| \log Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) - \log Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)\Big| > \varepsilon \Bigg\} \\
&\le P\Bigg\{ Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) \exp\left(- \frac{\varepsilon\sqrt{m}}{2Cn\, Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)}\right) \\
&\qquad \ge Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) > 0 \quad \text{for some } (s,\tilde{x},\tilde{y}) \in I_{2,n} \Bigg\} \\
&= P\Bigg\{ Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) - Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) \\
&\qquad \ge Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)\left(1 - \exp\left(- \frac{\varepsilon\sqrt{m}}{2Cn\, Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)}\right)\right) \\
&\qquad\qquad \text{for some } (s,\tilde{x},\tilde{y}) \in I_{2,n} \Bigg\}
\end{aligned}
$$

$$\le P\Big\{ Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) - Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)$$
$$> Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)\,\frac{e^{-1/2}\varepsilon\sqrt{m}}{2Cn\,Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)}$$
$$\text{for some } (s,\tilde{x},\tilde{y}) \in I_{2,n}\Big\}$$
$$\le P\Big\{ \sup_{(s,\tilde{x},\tilde{y})\in I} \frac{2Cn}{\sqrt{m}} \Big(Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) - Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)\Big) > \frac{\varepsilon}{e^{1/2}}\Big\}. \tag{4.79}$$

Similarly as in Lemma 3.4.2 (i.e. Lemma 6.1 of Draisma et al. (2004)) one can show the convergence of the process $nm^{-1/2}\big(Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) - Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)\big)_{(s,\tilde{x},\tilde{y})\in I}$ to the process $(-W(s\tilde{x}, s\tilde{y}))_{(s,\tilde{x},\tilde{y})\in I} = (-W_1(s\tilde{x}, s\tilde{y}))_{(s,\tilde{x},\tilde{y})\in I}$ as $n \to \infty$. Hence, due to Theorem 4.5.1, (4.79) converges to 0 as $n \to \infty$ and $\vartheta \downarrow 0$.

For $(s,\tilde{x},\tilde{y}) \in I_{3,n}$, with the same reasoning as we have obtained (4.79),

$$P\Big\{ \sup_{(s,\tilde{x},\tilde{y})\in I_{3,n}} \sqrt{m}\,2C\,\frac{n}{m}\, Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)\,1\!\!1_{(0,\infty)}\big(Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)\big)$$
$$\times \Big|\log Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) - \log Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)\Big| > \varepsilon\Big\}$$
$$\le P\Big\{ Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) - Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)$$
$$\ge Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}:n}\big)\big(1 - e^{-1/2}\big) \text{ for some } (\tilde{x},\tilde{y}) \in I_{3,n}\Big\}.$$

Note that by (4.76) (and the analogous inequality for $h_n(s\tilde{y})$ and $V_{h_n(s\tilde{y}):n}$),

$$P\Big\{ \frac{Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)}{Q\big(\lambda h_n(s\tilde{x})/n, \lambda h_n(s\tilde{y})/n\big)} \le 1\Big\} \ge 1 - 2\delta_1 \quad \text{and}$$
$$P\Big\{ 1 \le \frac{Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)}{Q\big(h_n(s\tilde{x})/(\lambda n), h_n(s\tilde{y})/(\lambda n)\big)}\Big\} \ge 1 - 2\delta_1.$$

Further, Theorem 5.3.14 of Meerschaert and Scheffler (2001) yields that for all $\delta_2 > 0$ there exists $t_0 > 0$, $0 < a < 1$, and $A > 1$ such that

$$at^{1/\eta-\delta_2} \leq \frac{Q(t\tilde{x}, t\tilde{y})}{Q(\tilde{x}, \tilde{y})} \leq At^{1/\eta+\delta_2}$$

for all $t \geq 1$, $\tilde{x}$ and $\tilde{y}$ with $t\tilde{x}, t\tilde{y} \leq t_0$. If we combine these results, we may conclude that for sufficiently large n,

$$\begin{aligned}
\alpha &:= \sup_{(s,\tilde{x},\tilde{y})\in I_{3,n}} Q\big(\lambda h_n(s\tilde{x})/n, \lambda h_n(s\tilde{y})/n\big) \\
&\leq K \sup_{(s,\tilde{x},\tilde{y})\in I_{3,n}} Q\big(h_n(s\tilde{x})/(\lambda n), h_n(s\tilde{y})/(\lambda n)\big) \\
&\leq K\frac{\sqrt{m}}{n}
\end{aligned}$$

and

$$\begin{aligned}
M &:= \inf_{(s,\tilde{x},\tilde{y})\in I_{3,n}} \sqrt{n}\, Q\big(h_n(s\tilde{x})/(\lambda n), h_n(s\tilde{y})/(\lambda n)\big)\big(1-e^{-1/2}\big) \\
&\geq K\sqrt{n} \inf_{(s,\tilde{x},\tilde{y})\in I_{3,n}} Q\big(\lambda h_n(s\tilde{x})/n, \lambda h_n(s\tilde{y})/n\big) \\
&\geq K\frac{\sqrt{m}}{\sqrt{n}\ \log n},
\end{aligned}$$

where K, as in the proof of Theorem 4.5.1, is some generic constant. This yields that $M = o\big(\alpha n^{1/2}\big)$ and $\sqrt{\alpha|\log\alpha|} = O\big((n^{-1}m^{1/2}\log n)^{1/2}\big) = o(M)$, where the last equality follows from the assumption $\log^6 n = o(m)$. Further, recall that the family of all cubes which belong to $(0,\vartheta] \times (0, y_0] \times [y_0/2, 2y_0]$ is a Vapnik-Chervonenkis class, see e.g. Section 26.1 of Shorack and Wellner (1986) or Section 1 of Alexander (1984). Then, Corollary 2.9 of the latter paper yields that for sufficiently large n,

$$P\Big\{\sup_{(s,\tilde{x},\tilde{y})\in I_{3,n}} \sqrt{m}\,2C\frac{n}{m}\,Q\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)\,\mathbb{1}_{(0,\infty)}\big(Q_n^*\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)\big) \times\Big|\log Q_n^*\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)-\log Q\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)\Big|>\varepsilon\Big\}$$

$$\le P\Big\{\sup_{(s,\tilde{x},\tilde{y})\in I_{3,n}} \sqrt{n}\,\big|Q_n^*\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)-Q\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)\big|>M\Big\} +4\delta_1$$

$$\le 16\exp\Big(-\frac{M^2}{4\alpha}\Big)+4\delta_1$$

$$\le 16\exp\Big(-K\frac{m}{n\log^2 n}\cdot\frac{n}{\sqrt{m}}\Big)+4\delta_1$$

where the first summand, by our assumption $\log^6 n = o(m)$, converges to 0 as $n\to\infty$.

We now turn to the case $Q_n^*\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big) > Q\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)$ which is easier to treat. Observe that by the mean value theorem,

$$\Big|\log Q_n^*\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)-\log Q\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)\Big| \le \frac{\big|Q_n^*\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)-Q\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)\big|}{Q_n^*\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)\wedge Q\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)}$$

so that the event considered on the right hand side of (4.78) intersected with event $Q_n^*\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big) > Q\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)$ has probability less than or equal to

$$P\Big\{\sup_{(s,\tilde{x},\tilde{y})\in I} 2C\,(2\lambda)^{1/\eta}\frac{n}{\sqrt{m}}\,\mathbb{1}_{(0,\infty)}\big(Q_n^*\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)\big) \times\big|Q_n^*\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)-Q\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)\big|>\varepsilon, \quad Q_n^*\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)>Q\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)\Big\}.$$

It again follows by Lemma 3.4.2 that this converges to 0 as $n \to \infty$ and $\vartheta \downarrow 0$.

Hence, we have proven so far that

$$\lim_{\vartheta\downarrow 0}\limsup_{n\to\infty} P\Bigg\{\sup_{(s,\tilde{x},\tilde{y})\in I} m^{1/2}c(s\tilde{x},s\tilde{y})\mathbb{1}_{(0,\infty)}\big(Q_n^*\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)\big)$$
$$\times\Big|\log Q_n^*\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)-\log Q\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)\Big|>\varepsilon\Bigg\}=0.$$

Of course, this also holds for $s = 1$ so that we may conclude to (4.74). It remains to show (4.75).

Define $d_n := d_n(\tilde{x},\tilde{y}) := U_{h_n(\tilde{x}):n} \vee V_{h_n(\tilde{y}):n}$ and observe that due to (4.24) and $s \le 1$, the quotient of $Q\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)$ and $q\big(d_n(\tilde{x},\tilde{y})\big)$ $c\big(U_{h_n(s\tilde{x}):n}/d_n,V_{h_n(s\tilde{y}):n}/d_n\big)$ is stochastically bounded away from 0. Since for all $1+x>\epsilon$, the mean value theorem yields $|\log(1+x)| \le x/(1+\theta x)$, $0<\theta<1$, $1+\theta x>\epsilon$, we obtain that

$$m^{1/2}c(s\tilde{x},s\tilde{y})\left|\log\frac{Q\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)}{q\big(d_n(\tilde{x},\tilde{y})\big)c\big(U_{h_n(s\tilde{x}):n}/d_n,V_{h_n(s\tilde{y}):n}/d_n\big)}\right|$$
$$\le\frac{m^{1/2}c(s\tilde{x},s\tilde{y})}{C}\left|\frac{Q\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)}{q\big(d_n(\tilde{x},\tilde{y})\big)c\big(U_{h_n(s\tilde{x}):n}/d_n,V_{h_n(s\tilde{y}):n}/d_n\big)}-1\right|$$
$$=O_P\left(m^{1/2}q_1(d_n)\frac{c(s\tilde{x},s\tilde{y})}{c\big(U_{h_n(s\tilde{x}):n}/d_n,V_{h_n(s\tilde{y}):n}/d_n\big)}\right)$$

with probability tending to 1, uniformly for all $(s,\tilde{x},\tilde{y}) \in I$ that satisfy $Q_n^*\big(U_{h_n(s\tilde{x}):n},V_{h_n(s\tilde{y}):n}\big)>0$, where the last equality follows from the assumption that Condition 2.2.1 holds uniformly for all $x,y\in[0,1]$. Note that although $s\tilde{x}$ and $s\tilde{y}$ are not multiples of $U_{h_n(s\tilde{x}):n}$ and $V_{h_n(s\tilde{y}):n}$, respectively, we may find $0<a<b<\infty$ such that $ac\big(U_{h_n(s\tilde{x}):n}/d_n,V_{h_n(s\tilde{y}):n}/d_n\big)\le c(s\tilde{x},s\tilde{y})\le bc\big(U_{h_n(s\tilde{x}):n}/d_n,V_{h_n(s\tilde{y}):n}/d_n\big)$, so that the regular variation of q_1, the homogeneity (2.14) of c and (4.76) lead to

$$\begin{aligned}
m^{1/2} c(s\tilde{x}, s\tilde{y}) \left| \log \frac{Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)}{q\big(d_n(\tilde{x},\tilde{y})\big) c\big(U_{h_n(s\tilde{x}):n}/d_n, V_{h_n(s\tilde{y}):n}/d_n\big)} \right| & \\
\le O_P\Big(m^{1/2} q_1((\tilde{x}\vee\tilde{y})/F_T^{\leftarrow})\big(d_n F_T^{\leftarrow}\big)^{1/\eta}\Big) & \\
= o_P\left(\frac{q_1((\tilde{x}\vee\tilde{y})/F_T^{\leftarrow})}{q_1(1/F_T^{\leftarrow})}\,(\tilde{x}\vee\tilde{y})^{1/\eta}\right) & \\
= o_P\Big((\tilde{x}\vee\tilde{y})^{1/\eta+\tau-\delta_3}\Big) \qquad \text{for all } \delta_3 > 0 & \\
= o_P(1) &
\end{aligned}$$

uniformly for all $(s,\tilde{x},\tilde{y}) \in I$ with $Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big) > 0$, where the last but one equality is obtained by the Potter bounds. This inequality and the homogeneity (2.14) of c leads us to

$$\begin{aligned}
\limsup_{n\to\infty} P\bigg\{ \sup_{(s,\tilde{x},\tilde{y})\in I} m^{1/2} c(s\tilde{x}, s\tilde{y}) \mathbb{1}_{(0,\infty)}\big(Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)\big) & \\
\times \left| \log \frac{Q\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)}{Q\big(U_{h_n(\tilde{x}):n}, V_{h_n(\tilde{y}):n}\big)} - \log \frac{c\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)}{c\big(U_{h_n(\tilde{x}):n}, V_{h_n(\tilde{y}):n}\big)} \right| > \varepsilon \bigg\} = 0, &
\end{aligned} \tag{4.80}$$

so that it suffices to show that

$$\begin{aligned}
\lim_{\vartheta\downarrow 0} \limsup_{n\to\infty} P\bigg\{ \sup_{(s,\tilde{x},\tilde{y})\in I} m^{1/2} (s\tilde{x}\wedge s\tilde{y})^{1/2} (\ell_2(s\tilde{x}\wedge s\tilde{y}))^{-1/2-\delta} & \\
\times \mathbb{1}_{(0,\infty)}\big(Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)\big) \left| \log \frac{c\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)}{c\big(sU_{h_n(\tilde{x}):n}, sV_{h_n(\tilde{y}):n}\big)} \right| > \varepsilon \bigg\} = 0. &
\end{aligned} \tag{4.81}$$

From the convergence of the uniform tail quantile process to a Brownian motion under a suitable weighted supremum norm, see e.g. Theorem 5.2.5 of Csörgő and Horváth (1993), we know that for any intermediate sequence $k = k_n$ and all $\delta_4 > 0$ that

$$\frac{n}{k}\, U_{\lceil kx\rceil:n} - x = o_P\Big(k^{-1/2} x^{1/2} (\ell_2(x))^{(1+\delta_4)/2}\Big)$$

uniformly for x from a compact set. Hence, due to (4.76), we may conclude that

$$\frac{n}{h_n(1)}\frac{U_{h_n(s\tilde{x}):n}}{s\tilde{x}} = 1 + O_P\Big(\min\Big(1, \Big(\frac{F_T^{\leftarrow}}{ns\tilde{x}}\Big)^{1/2}\Big((\ell_2(s\tilde{x}))^{(1+\delta_4)/2} + \frac{F_T^{\leftarrow}}{ns\tilde{x}}\Big)\Big)\Big)$$

uniformly for $0 < s\tilde{x} \le 2y_0$. By the analogous equation for $V_{h_n(\tilde{y}):n}$ and $\tilde{y}$, the homogeneity (2.14) of c, (4.76) and $\log(1+x) = O(x)$ if x is bounded away from -1 we obtain

$$\begin{aligned}
&\log\frac{c\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)}{c\big(sU_{h_n(\tilde{x}):n}, sV_{h_n(\tilde{y}):n}\big)}\\
&= O_P\Big(\min\Big(1, \Big(\frac{F_T^{\leftarrow}}{ns\tilde{x}}\Big)^{1/2}\Big((\ell_2(s\tilde{x}))^{(1+\delta_4)/2} + \frac{F_T^{\leftarrow}}{ns\tilde{x}}\Big)\\
&\qquad + \min\Big(1, \Big(\frac{F_T^{\leftarrow}}{ns\tilde{y}}\Big)^{1/2}\Big((\ell_2(s\tilde{y}))^{(1+\delta_4)/2} + \frac{F_T^{\leftarrow}}{ns\tilde{y}}\Big).
\end{aligned}$$

Finally, observe that $\big(mF_T^{\leftarrow}/n\big)^{1/2} = \big(m/(q^{\leftarrow}(m/n)n)\big)^{1/2} \le 1$ and that $\lim_{\vartheta\downarrow 0}\sup_{0<s\tilde{x}\le\vartheta}(\ell_2(s\tilde{x}))^{-\delta_4/2} = 0$, so that (4.81) holds.

4.6.2 Remark. After the proof of (4.80), in order to show that (4.22) holds with weight function $c(sx, sy_0)$ (cf. Remark 4.5.6 (ii)), it would suffice to show that

$$\begin{aligned}
\lim_{\vartheta\downarrow 0}\limsup_{n\to\infty} P\Big\{\sup_{(s,\tilde{x},\tilde{y})\in I} & m^{1/2} c(s\tilde{x}, s\tilde{y})\mathbb{1}_{(0,\infty)}\big(Q_n^*\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)\big)\\
&\times\Big|\log\frac{c\big(U_{h_n(s\tilde{x}):n}, V_{h_n(s\tilde{y}):n}\big)}{c\big(sU_{h_n(\tilde{x}):n}, sV_{h_n(\tilde{y}):n}\big)}\Big| > \varepsilon\Big\} = 0.
\end{aligned}\tag{4.82}$$

However, e.g. choosing $s = 1/2$, $\tilde{x} = 2F_T^{\leftarrow}/n$ and $\tilde{y} = y_0$ yields

$$\log\frac{c\big(U_{1:n}, V_{\lceil y_0(n+1)/(2F_T^{\leftarrow})\rceil-1:n}\big)}{c\big(U_{2:n}/2, V_{\lceil y_0(n+1)/(F_T^{\leftarrow})\rceil-1:n}/2\big)}$$

for the logarithm in (4.82), where the quotient of $V_{\lceil y_0(n+1)/(2F_T^{\leftarrow})\rceil-1:n}$ and $V_{\lceil y_0(n+1)/F_T^{\leftarrow}\rceil-1:n}/2$ converges to 1 as $n\to\infty$, but in general $2U_{1:n}/U_{2:n}$ converges to a non-degenerated limiting distribution. Hence, in general one may expect (4.22) to hold with weight function $c(sx, sy_0)$ only if one assumes that $m^{1/2}c(sx, sy_0) \to 0$ as $n \to \infty$, cf. Remark 4.5.6 (ii). $\triangle$

Chapter 5

Applications

As before, let $\{(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\}$ be a sample of i.i.d. claim sizes. Analyzing real data, we often face the problem that, for example, a claim is recorded only if it exceeds a threshold or if the sum $X_i + Y_i$ pertaining to one pair (X_i, Y_i) of claims of two different lines of business exceeds a certain value, see e.g. the data sets of the Danish fire insurance claims and the US medical claims in the following Sections 5.1 and 5.2. This censorship is a very common problem in actuarial practice, especially in reinsurance. In probabilistic terms, such data sets represent samples of some kind of *conditional* distributions of (X, Y), e.g. of $P^{(X,Y)|X>x_0,Y>y_0}$ or of $P^{(X,Y)|X+Y>z_0}$. We must be aware that the way of recording claims may influence the dependence structure of the data. This is not the case if the data set represents a sample of $P^{(X,Y)|X>x_0,Y>y_0}$ – it solely implies that we are estimating $P(X > u_1, Y > u_2 \mid X > x_0, Y > y_0)$ rather than $P\{X > u_1, Y > u_2\}$ when investigating such a data set. We must be more careful in the latter case $P^{(X,Y)|X+Y>z_0}$, however, since this way of recording data causes an artificial negative dependence between the claims in the different business lines. If one of the components is smaller than z_0, the other one must be accordingly larger (to be recorded). There are two different natural ways to correct this artificial dependence: we consult a claim (X_i, Y_i) only if either both components X_i and Y_i separately exceed z_0 (i.e. if $\min(X_i, Y_i) > z_0$) or if at least one component exceeds z_0 (i.e. if $\max(X_i, Y_i) > z_0$).

We discuss the choice of the appropriate method within the following sections, in particular at the end of Section 5.1.

5.1 Danish Fire Insurance

Our first application deals with a well-known data set of Danish fire insurance claims. These data have first been considered by Rytgaard (1996) and since then they have been studied several times in extreme value analysis. The data set contains losses to building(s) X_i, losses to contents Y_i and losses to profits caused by the same fire. We suppose, as an example, that we are interested in a bivariate analysis of X_i and Y_i. The claims are recorded only if the sum of all components attains or exceeds 1 million Danish Kroner (DKK). For the period 01/1980 – 12/1990 and 01/1980 – 12/1993, respectively, these data were considered for a univariate extreme value analysis e.g. by McNeil (1997) and Embrechts et al. (1997) (starting with Example 6.2.9). For the latter period, Blum et al. (2002) investigate dependencies between Y_i and losses to profits by fitting various parametric copulas to the data. (For a discussion of using parametric copulas for fitting multivariate extremes see the Preface, Chapter 6 and Mikosch (2005).)

The extended data set we consider contains 6,870 recorded claims of the period 01/1980 – 12/2002. We discounted the claim sizes to 7/1985 prices according to the Danish Consumer Price Index (DCPI)[9] on a monthly basis. We have discussed above that due to the recording method, there is an artificial negative dependence between the components. We correct this artificial dependence by consulting a claim (X_i, Y_i) only if either both components X_i and Y_i separately attain or exceed 1,000,000 DKK (i.e. $\min(X_i, Y_i) \geq 1,000,000$ DKK) or if at least one component attains or exceeds $1,000,000$ DKK (i.e. $\max(X_i, Y_i) \geq 1,000,000$ DKK). We analyze the resulting data sets in separate sections.

5.1.1 Both Components Exceed 1,000,000 DKK

In this section, we analyze the conditional distribution of (X, Y) given $\{X \geq 1, Y \geq 1\}$, where here and henceforth in Section 5.1 we refer to claims in millions of DKK. The sample size of the remaining data is $n = 588$. Figure 5.1 displays the scatterplot of the data on a linear and on a logarithmic scale. In particular in the plot on the logarithmic scale we see that the observations tend to cluster around the diagonal. This suggests a moderate positive dependence.

[9] *http://www.dst.dk/Statistik/seneste/Indkomst/Priser/Forbrugerprisindeks.aspx*

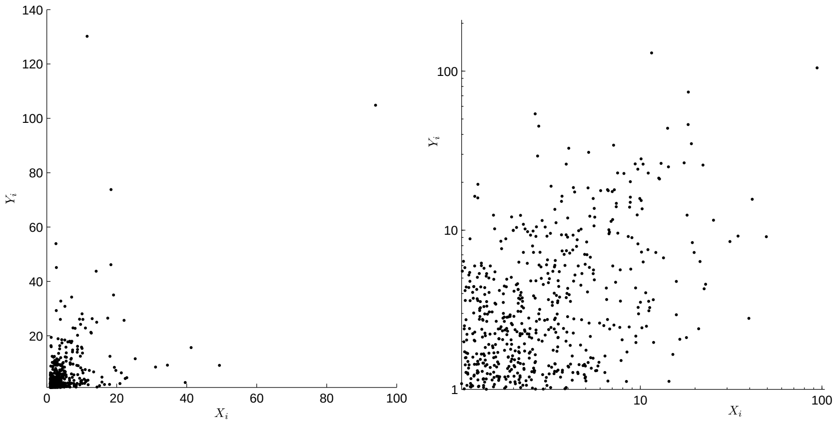


Figure 5.1: Scatterplot Y_i vs. X_i on a linear and on a logarithmic scale.

Fitting of the Marginals

As described in Section 1.1.4, in order to choose a suitable number of order statistics j_i used for the (Hill or ML) estimator $\hat{\gamma}_{i,n}$, we consider a so-called Hill plot, which displays the estimates $\hat{\gamma}_{i,n}$ versus j_i, $i = 1, 2$. We briefly recapitulate the aspects discussed in Section 1.1.4. Recall that a small value j_i results in a high variance of $\hat{\gamma}_{i,n}$, while too large a j_i, $i = 1, 2$, may cause a large bias. A typical (nice) Hill plot exhibits heavy fluctuations for small values of j_i, $i = 1, 2$, (due to a large variance), followed by a rather stable region where both variance and bias are moderate, before the bias causes a clear trend of the curve. Thus, j_i, $i = 1, 2$, ought to be chosen in the region where the plot is rather stable. For a more detailed discussion on this matter see e.g. Sections 6.4 and 6.5 of Embrechts et al. (1997) or Drees et al. (2000). We agree that in plots $\hat{\gamma}_{i,n}$ versus j_i in this chapter, the Hill estimator $\hat{\gamma}_{i,n}^{H}$ is displayed in a solid black line, while the maximum likelihood estimator $\hat{\gamma}_{i,n}^{ML}$, $i = 1, 2$, is displayed in a dotted red line. According to (1.15) and (1.18), the asymptotic variance of $\hat{\gamma}_{i,n}^{H}$ is smaller than the asymptotic variance of $\hat{\gamma}_{i,n}^{ML}$, so that we prefer to work with $\hat{\gamma}_{i,n}^{H}$, $i = 1, 2$, in this chapter. Further, in this section, denote by $\hat{G}_{i,n}^{H}$, $i = 1, 2$, the estimates (1.10), where $\hat{\mu}_n$ and $\hat{\varsigma}_n$ are estimated according to (1.16), for the conditional marginal d.f.s of X and Y given $\{X \geq 1, Y \geq 1\}$, respectively.

For the present data set, the (heuristic) analysis of the Hill plots (Figures 5.2 and 5.3) suggests $j_1 = 350$ and $j_2 = 380$, which yields

$$1 - \hat{G}^H_{1,n}(x) \approx \left(1 + 0.47\frac{x - 0.96}{1.85}\right)^{-2.14}$$

and

$$1 - \hat{G}^H_{2,n}(y) \approx \left(1 + 0.52\frac{y - 0.72}{2.45}\right)^{-1.93}.$$

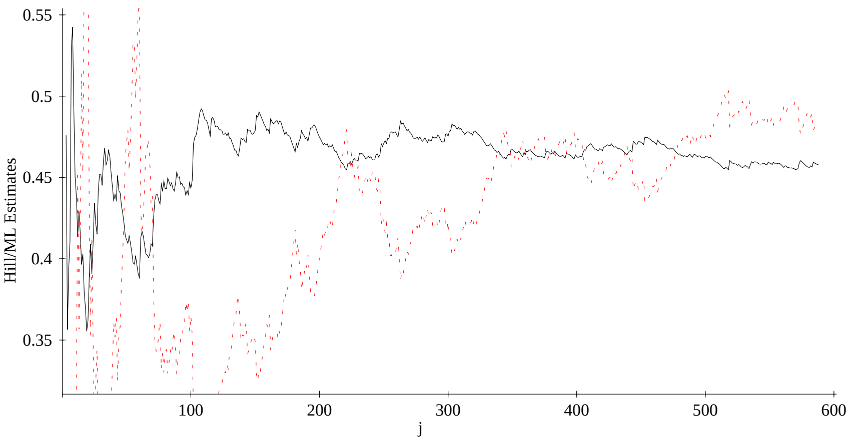


Figure 5.2: Hill plot: $\hat{\gamma}_{1,n}$ as a function of j_1.

In Figure 5.4, the estimates are validated by exponential qq-plots for $\left(1 - \hat{G}^H_{i,n}\right)^{\leftarrow}$, $i = 1, 2$, i.e. by plots

$$\left(-\log\frac{k - 1/2}{j_1}, \log X_{n-k+1:n}\right)_{1 \leq k \leq j_1} \quad \text{and} \quad \left(-\log\frac{k - 1/2}{j_2}, \log Y_{n-k+1:n}\right)_{1 \leq k \leq j_2} \tag{5.1}$$

with the lines $\log X_{n-j_1:n} + \hat{\gamma}_{1,n} \log(2j_1)$ and $\log Y_{n-j_2:n} + \hat{\gamma}_{2,n} \log(2j_2)$, respectively, for comparison.

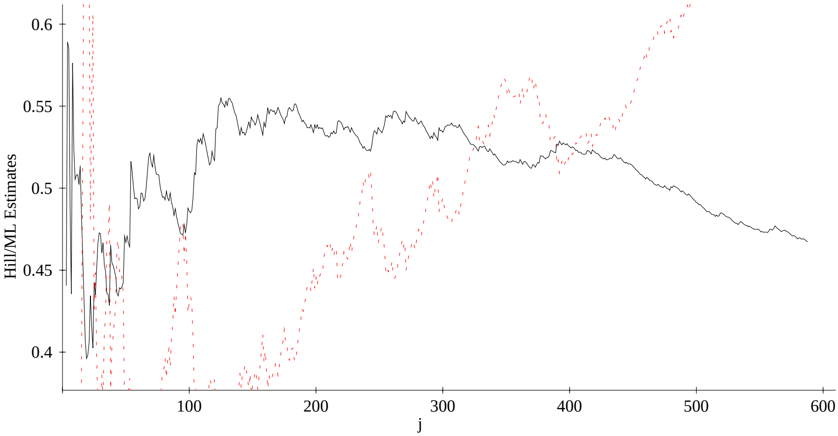


Figure 5.3: Hill plot: $\hat{\gamma}_{2,n}$ as a function of j_2.

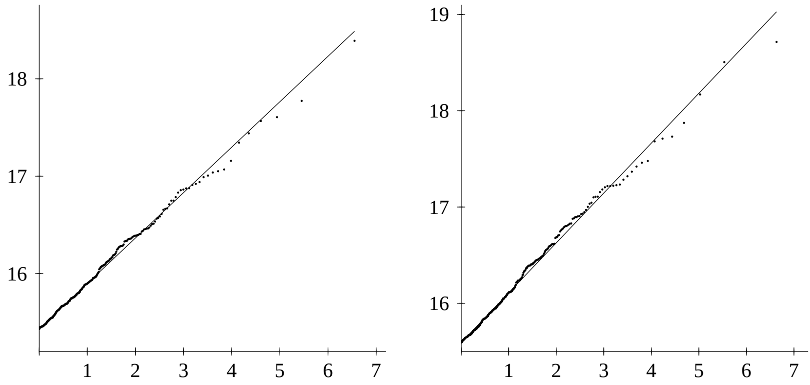


Figure 5.4: Exponential qq-plots for $\left(1 - \hat{G}^H_{1,n}\right)^{\leftarrow}$ and $\left(1 - \hat{G}^H_{2,n}\right)^{\leftarrow}$.

Estimation of η

As we have indicated in Section 3.1, the heuristic when choosing a suitable number m of largest order statistics $T^{(n)}_{i:n}$ used for the estimator $\hat{\eta}_n$ is the same as when choosing a suitable number j of order statistics $X_{i:n}$ used for the estimator $\hat{\gamma}_n$. Therefore, we consider a Hill plot which displays the estimates $\hat{\eta}_n$ versus m. Similar as for the Hill plot $\hat{\gamma}_{i,n}$ versus j_i, $i = 1, 2$, we agree that the Hill estimator $\hat{\eta}^H_n$ is displayed in a solid black line, while the maximum likelihood estimator $\hat{\eta}^{ML}_n$ is displayed in a dotted red line. As mentioned in Remark 3.1.3 (ii), we prefer to work with $\hat{\eta}^H_n$ in this chapter.

For the present data set, the (heuristic) analysis of the Hill plot (Figure 5.5) suggests $m = 200$, which yields $1/T^{(n)}_{n-m:n} = 0.53$ and $\hat{\eta}^H_n = 0.76$. The

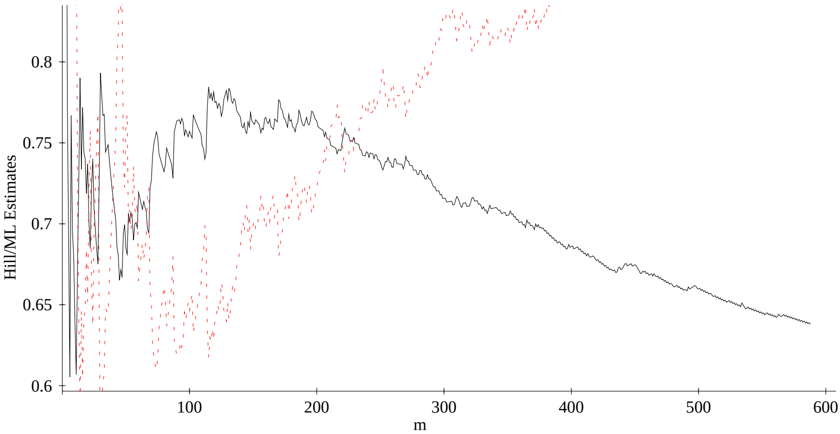


Figure 5.5: Hill plot: $\hat{\eta}_n$ as a function of m.

approximate 95% confidence interval (3.6) for η is $[0.66, 0.87]$. The pertaining test (3.7) clearly rejects the null hypothesis $\eta = 1$ on an approximate 95% confidence level with $(1 - p)$-value > 0.9999, so that we assume asymptotic independence. Hence, the moderate positive dependence that was suggested by the scatterplots of Figure 5.1 is confirmed and vanishes asymptotically.

Figure 5.6 displays the scatterplot of the pseudo-observations $(\hat{U}_i, \hat{V}_i)$, where here and in other scatterplots $\hat{V}_i$ vs. $\hat{U}_i$ in this chapter, we use middle ranks if ties occur. This means that if there are k observations X_i with $X_i < X_{i_1}$, further $n-k-l$ observations with $X_i > X_{i_1}$ and $X_{i_1} = X_{i_2} = \ldots = X_{i_l}$, then we assign $R^X_{i_1} = R^X_{i_2} = \ldots = R^X_{i_l} = k + (l+1)/2$. In the left hand plot of Figure 5.6 the square with side length $1/T^{(n)}_{n-m:n} = 0.53$ is indicated.

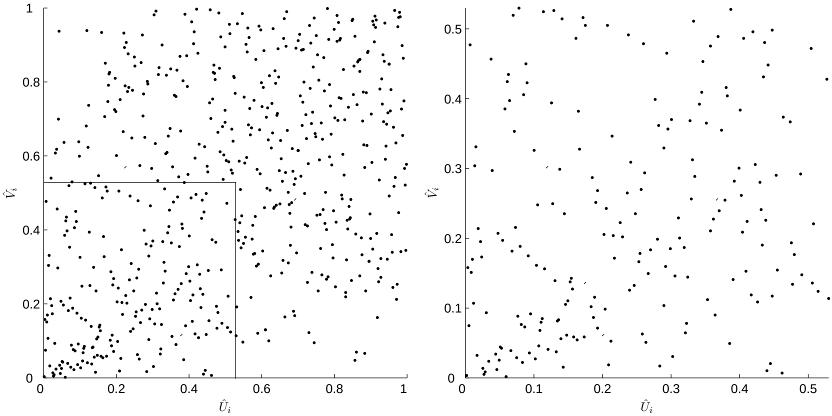


Figure 5.6: Scatterplot $\hat{V}_i$ vs. $\hat{U}_i$. The square with side length $1/T^{(n)}_{n-m:n} = 0.53$ contains the $m = 200$ points used for the estimation of η. The scaling law (2.17) is checked within this square.

Model Validation

Only 5 of the points (4.11), i.e. 2.5%, lie outside their approximate 95% confidence interval (4.14), so that our method accepts the presence of the scaling law (2.17). The pertaining plots (4.11) are shown in Figure 5.8, where the approximate $(1-p)$-values (4.15) of the tests (4.13) are divided into 8 intervals and displayed in the plot (4.11) in color according to the colorbar of Figure 5.7. Note that a point fails the test on an approximate 95% confidence level iff it has a reddish color, i.e. iff it is orange, light red, dark red or pink.[10]

[10] For files with a three-dimensional presentation, see *http://www.math.uni-hamburg.de/home/drees/extrdep/mdilc.html*. There you find avi files which display a

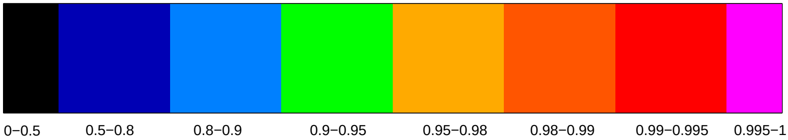


Figure 5.7: The colorbar used to divide the approximate $(1-p)$-values (4.15) of the tests (4.13) into 8 intervals.

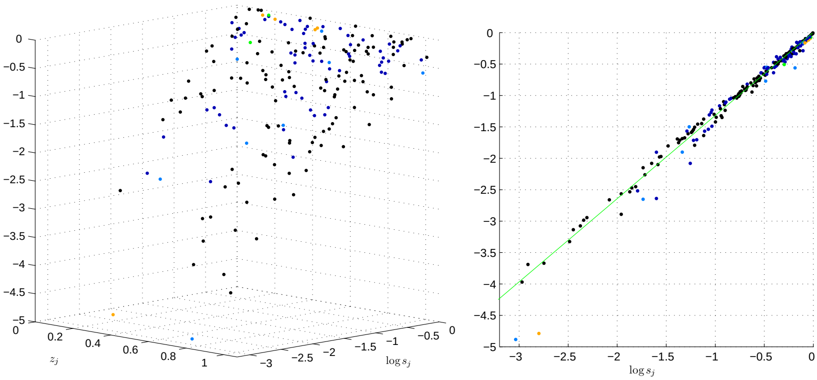


Figure 5.8: Plot (4.11) with $m = 200$, $1/T^{(n)}_{n-m:n} = 0.53$, $\hat{\eta}^H_n = 0.76$ from two perspectives. The right hand perspective also displays the reference plane. 5 (reddish) points lie outside their approximate 95% confidence interval.

For the purpose of comparison, Figure 5.9 displays the plot (4.11) for $m = 500$, that is, about 85% of all data points are used for the model fitting. Since essentially more data is used for the model fitting than suggested by the Hill plot $\hat{\eta}_n$ versus m, it seems quite likely that the scaling law (2.17), that was motivated by asymptotic arguments for extreme observations, does not hold on the much larger square with side length $1/T^{(n)}_{n-m:n} = 0.92$. Indeed, we ob-

---

plot rotation in a film. An m and a mat file additionally allow to rotate the plots in any manner (MatLab required).

serve that 82 points, about 16%, lie outside their approximate 95% confidence interval. Thus, in this case, our method clearly detects the deviations from the scaling law. (Note that although also the Hill plot for η indicates that $m = 500$ is too large, the resulting point estimate $\hat{\eta}_n^H = 0.66$, cf. Figure 5.5, belongs to the confidence interval for η obtained with the choice $m = 200$, i.e. the difference is not statistically significant.) Files with the plot are provided on the website mentioned in Footnote 10.

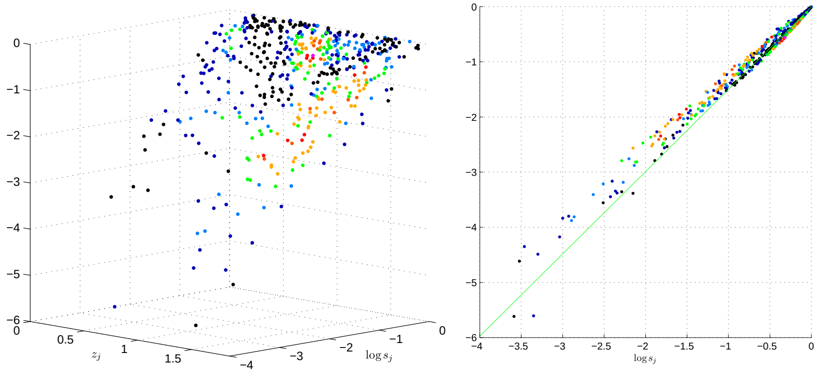


Figure 5.9: Plot (4.11) with $m = 500$, $1/T^{(n)}_{n-m:n} = 0.92$, $\hat{\eta}_n^H = 0.66$ from two perspectives. The right hand perspective also displays the reference plane. 82 (reddish) points lie outside their approximate 95% confidence interval.

This example (and the corresponding example considered in Section 5.2.1) demonstrates that our method cannot only be used to decide whether the ELTM is appropriate or not in *some* tail region, but also, loosely speaking, to determine the maximum tail region for which the application of the ELTM is still reasonable.

Probability of Jointly Large Claims

As we have discussed in Section 3.3.2, the heuristic when choosing a suitable number ι of largest order statistics $T^{(n)}_{i:n}$ used for the estimator $\hat{p}_n$ is the same as when choosing a suitable number m of largest order statistics $T^{(n)}_{i:n}$ used for the estimator $\hat{\eta}_n$ or as when choosing a suitable number j of order statistics $X_{i:n}$

used for the estimator $\hat{\gamma}_n$. Therefore, we consider a similar plot, namely (3.38), which displays $\hat{p}_n$ versus ι. Further, we include the confidence intervals (3.35), (3.36) and (3.37). We agree that the estimator $\hat{p}_n$ is plotted in a solid black line, the confidence intervals (3.35) (valid if the error when estimating η dominates) in a dotted black line, the confidence interval (3.36) (valid if the error when estimating γ_1 dominates) in a dotted red line and the confidence interval (3.37) (valid if the error when estimating γ_2 dominates) in a dotted green line. Recall that under the conditions of Theorem 3.3.1, the confidence intervals (3.35), (3.36) and (3.37) apply asymptotically if (3.28), (3.30) and (3.32), respectively, hold. Loosely speaking, this means that the confidence interval (3.35) applies asymptotically if the sequences $j_i = j_{i,n}$, $i = 1, 2$, converge faster to ∞ than $m = m_n$ as $n \to \infty$, see also (1.22) and Remark 3.3.2 (vii). If j_1 (j_2) converges slower to ∞ than m and j_2 (j_1), then (3.36) ((3.37)) applies asymptotically. Hence, as a heuristic rule, one might use (3.35) if m is the minimum of m, j_1 and j_2 and use (3.36) ((3.37)) if j_1 (j_2) is the minimum of m, j_1 and j_2. However, the confidence intervals correspond to the cases that the pertaining statistical error dominates, cf. Theorem 3.3.1 and Remarks 3.3.2 (vi) and (vii). Thus, also to ensure maximum prudence, we will use the widest confidence interval of the three.

For the present data set, plots for estimates $\hat{p}_n$ and approximate 95% confidence intervals for the conditional probability $P(X > u_1, Y > u_2 \mid X \geq 1, Y \geq 1)$ with $u_1 = u_2 = 10$ and $u_1 = 20$, $u_2 = 30$ are given in Figure 5.10, where we used $\hat{G}^H_{1,n}$ and $\hat{G}^H_{2,n}$ as marginal Pareto estimates.

Table 5.1 compares estimates and (approximate) 95% confidence intervals for the conditional probability $P(X > u_1, Y > u_2 \mid X \geq 1, Y \geq 1)$ for further values of u_1 and u_2, where we also give our choices of ι. The confidence intervals (3.35) are the widest in this example. The confidence intervals for the empirical probability $\hat{p}_e := n^{-1} \sum_{i=1}^{n} \mathbb{1}\{X_i > u_1, Y_i > u_2\}$ are computed according to Clopper and Pearson (1934), see also e.g. Santner and Duffy (1989), p. 35. Further, estimates $\hat{p}_i := (1 - \hat{G}^H_{1,n}(u_1))(1 - \hat{G}^H_{2,n}(u_2))$ assuming independence of X and Y (given that $\{X \geq 1, Y \geq 1\}$) are displayed.

A typical interpretation of an estimate $\hat{p}_n$ may be as follows. The estimate of 0.549% for the probability that X exceeds 20 and Y exceeds 30 given that both exceed 1 means that on average such a claim occurs approximately once every 200 years. However, observe that the upper bound of the approximate 95% confidence interval is almost twice the estimate, so that there is still a risk of approximately 5% that on average such a claim occurs approximately once

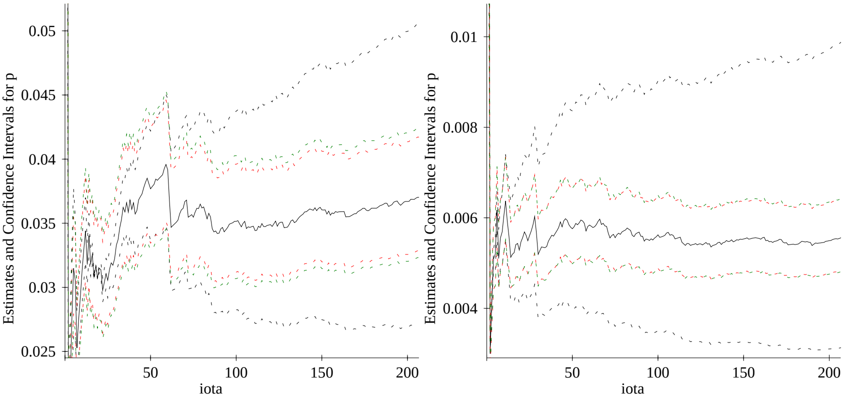


Figure 5.10: Estimation of $P(X > u_1, Y > u_2 \mid X \geq 1, Y \geq 1)$. Plot (3.38) and approximate 95% confidence intervals (3.35), (3.36) and (3.37) with $u_1 = u_2 = 10$ (left) and $u_1 = 20$, $u_2 = 30$ (right).

u_1	u_2	ι	$\hat{p}_n$ (%)	$\hat{p}_e$ (%)	$\hat{p}_i$ (%)
5	5	120	12.49 [11.25, 13.69]	12.59 [10.01, 15.54]	6.377
10	10	120	3.492 [2.80, 4.34]	3.401 [2.09, 5.20]	0.962
20	20	120	0.751 [0.49, 1.15]	0.340 [0.04, 1.22]	0.100
20	30	120	0.549 [0.32, 0.89]	0.170 [0.00, 0.94]	0.051
25	25	120	0.442 [0.26, 0.70]	0.170 [0.00, 0.94]	0.046
100	10	120	0.180 [0.10, 0.45]	0 [0, 0.63]	0.012

Table 5.1. Comparison of the estimates $\hat{p}_n$, the empirical probability $\hat{p}_e$ and the estimated probability $\hat{p}_i$ assuming conditional independence of X and Y.

every 100 years or even more often. We observe that the empirical probability is close to the estimates obtained in the ELTM if the event under consideration has been observed sufficiently often. However, the confidence intervals calculated from the empirical probabilities are typically considerably wider. (The latter fact is demonstrated even more impressively in Table 5.4.) The empirical probability is (substantially) smaller than the estimates $\hat{p}_n$ if only a few observations have been made. For the considered example, the estimate of 0.170% means that the risk that such a claim occurs is three times smaller than the estimate $\hat{p}_n$ suggests, meaning that on average such a claim occurs approximately once every 600 years. Of course, the empirical probability is of very limited value if the event under consideration has not yet occurred. As expected, the assumption of conditional independence of X and Y yields systematically smaller estimates than the empirical probability (if the latter is positive) and those obtained in the ELTM. For the considered example, the risk is estimated as less dangerous by factor 10, meaning that on average such a claim occurs approximately once every 2,000 years.

5.1.2 At Least One Component Exceeds 1,000,000 DKK

In this section, we analyze the conditional distribution of (X, Y) given $\{X \geq 1 \text{ or } Y \geq 1\}$. From the 6,870 pairs of claims of the original data set, there are 2,894 pairs with both components smaller than 1. Hence, the sample size of the data investigated in this section is $n = 3,976$. Figure 5.11 displays the scatterplot of the data on a linear and on a logarithmic scale. Note that claims

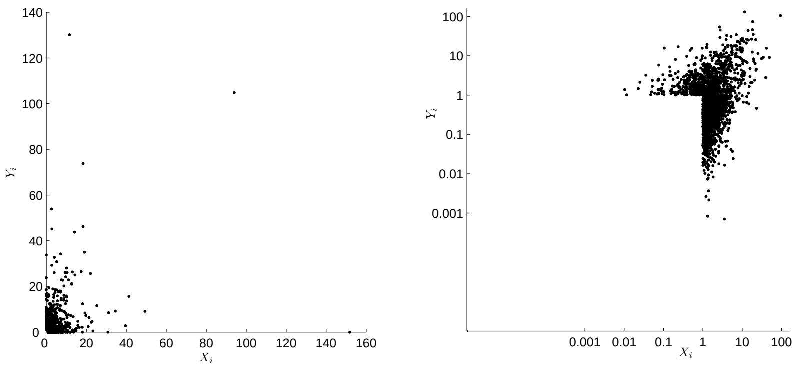


Figure 5.11: Scatterplot Y_i vs. X_i on a linear and on a logarithmic scale.

with a component equal to 0 are not representable on a logarithmic scale, so that 378 pairs (X_i, Y_i) with $X_i = 0$ and 1,016 pairs (X_i, Y_i) with $Y_i = 0$ are not displayed in the right hand plot of Figure 5.11.

Fitting of the Marginals

Here, the (heuristic) analysis of the Hill plots (Figures 5.12 and 5.13) suggests $j_1 = 900$ and $j_2 = 600$ which yields

$$1 - \hat{G}^H_{1,n}(x) \approx \left(1 + 0.57\frac{x - 0.91}{0.54}\right)^{-1.75}$$

and

$$1 - \hat{G}^H_{2,n}(y) \approx \left(1 + 0.72\frac{y - 0.15}{0.47}\right)^{-1.40},$$

with $1 - \hat{G}^H_{1,n}$ and $1 - \hat{G}^H_{2,n}$ denoting the estimates (1.10), where $\hat{\mu}_n$ and $\hat{\varsigma}_n$ are estimated according to (1.16), for the conditional d.f.s of X and Y given $\{X \geq 1 \text{ or } Y \geq 1\}$. However, the choice of j_1 and j_2 is not easy in this case, since the Hill plots are not "nice" in the sense as described in Section 5.1.1. Further, the exponential qq-plots (5.1) for $\left(1 - \hat{G}^H_{i,n}\right)^{\leftarrow}$, $i = 1, 2$, in Figure 5.14 exhibit a rather moderately good fit of the data, in particular for $i = 2$.

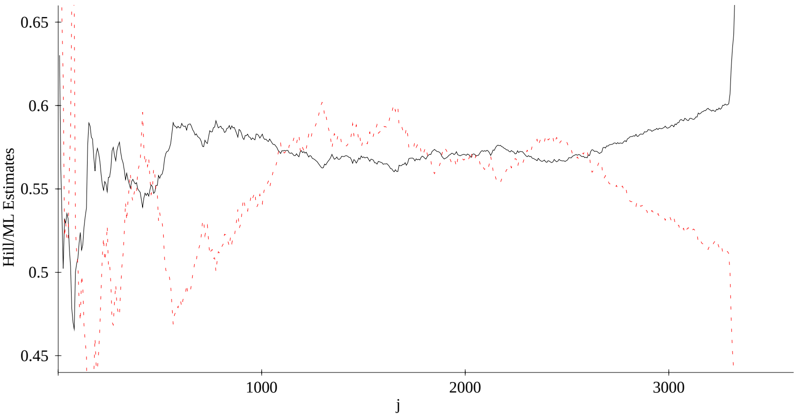


Figure 5.12: Hill plot: $\hat{\gamma}_{1,n}$ as a function of j_1.

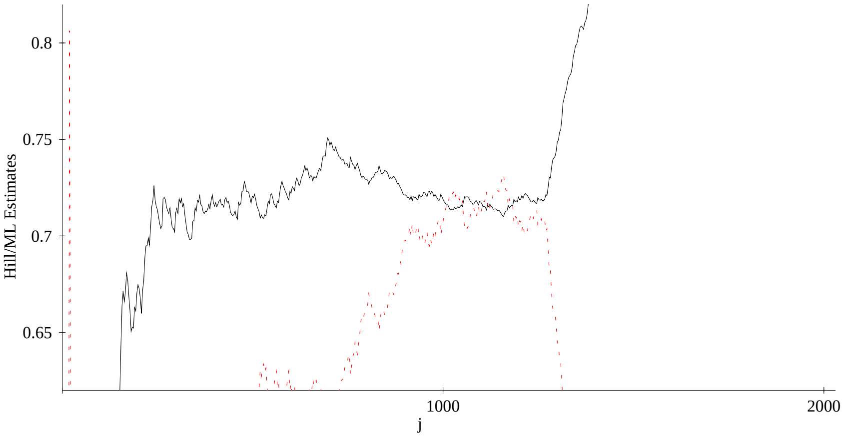


Figure 5.13: Hill plot: $\hat{\gamma}_2$ as a function of j_2.

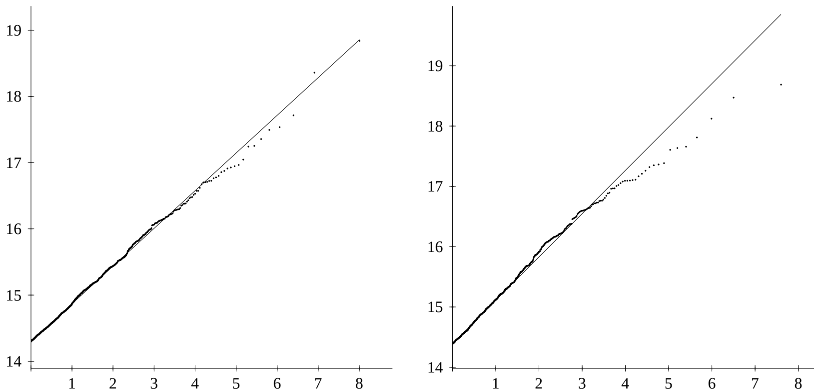


Figure 5.14: Exponential qq-plots for $\left(1-\hat{G}^H_{1,n}\right)^{\leftarrow}$ and $\left(1-\hat{G}^H_{2,n}\right)^{\leftarrow}$.

Estimation of η

The (heuristic) analysis of the Hill plot (Figure 5.15) leads to $m = 200$, which yields $1/T^{(n)}_{n-m:n} = 0.15$ and $\hat{\eta}^H_n = 0.97$, where the Hill estimates for $\eta \leq 1$ that are larger than 1 (for, approximately, $220 < m < 400$) are meaningless. The approximate 95% confidence interval (3.6) for η is $[0.84, 1]$. The pertaining test (3.7) does not reject the hypothesis $\eta = 1$ on an approximate 95% confidence level.

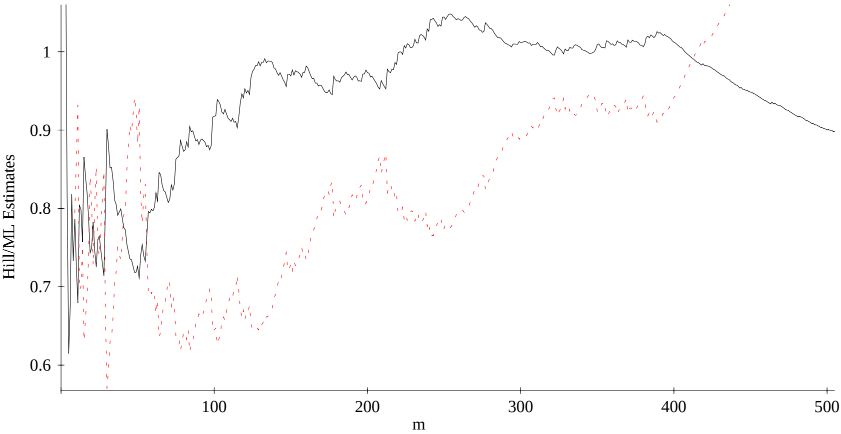


Figure 5.15: Hill plot: $\hat{\eta}_n$ as a function of m.

Figure 5.16 displays the scatterplot of the pseudo-observations $(\hat{U}_i, \hat{V}_i)$ including the square with side length $1/T^{(n)}_{n-m:n} = 0.15$. At first glance, the left hand plot of that figure exhibits some unexpected particularities. To understand these specials, recall that we used ranks R^X_i and R^Y_i to define the pseudo-observations $\hat{U}_i = 1 - R^X_i/(n + 1)$ and $\hat{V}_i = 1 - R^Y_i/(n + 1)$. Recall further that we use middle ranks if ties occur, so that we assign rank 189.5 to the 378 claims $X_i = 0$, such that the corresponding $\hat{U}_i$ equal $1 - 189.5/3,977 \approx 0.95$. Analogously, we treat the 1,016 claims $Y_i = 0$, so that the points closest to $\hat{U}_i = 1$ and $\hat{V}_i = 1$ lie on lines $\hat{U}_i \approx 0.95$ and $\hat{V}_i \approx 0.87$. Further, the structural interruptions ("breaks") at $\hat{U}_i \approx 0.83$ and $\hat{V}_i \approx 0.32$ are due to both our standardization method and that we consider claims that attain or exceed 1 in at

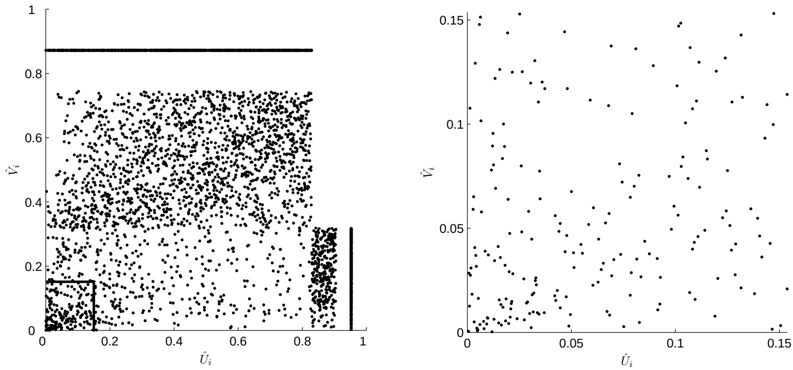


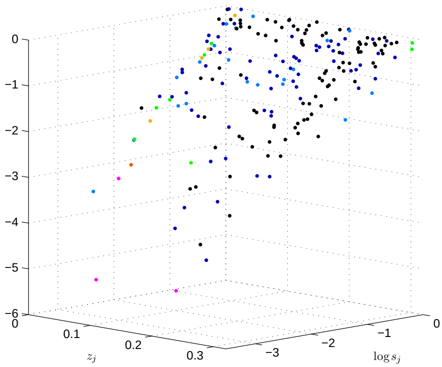
Figure 5.16: Scatterplot $\hat{V}_i$ vs. $\hat{U}_i$. The square with side length $1/T^{(n)}_{n-m:n} = 0.15$ contains the $m = 200$ points used for the estimation of η. The scaling law (2.17) is checked within this square.

least one component. First, we have that $\sum_{i=1}^n \mathbb{1}\{X_i < 1\} = 679$, i.e. the number of claims $X_i < 1$ equals 679, and $\sum_{i=1}^n \mathbb{1}\{Y_i < 1\} = 2,709$. Hence, due to our standardization using ranks, the 679-378=301 pseudo-observations $(\hat{U}_i, \hat{V}_i)$ that correspond to the pairs (X_i, Y_i) with $0 < X_i < 1$ fit into the rectangle $[1 - 679/3,977; 1 - 379/3,977] \times [0; 1 - 2,709/3,977] \approx [0.83; 0.90] \times [0; 0.32]$. Analogously, 2,709-1,016=1,693 pseudo-observations fit into the rectangle $\approx [0; 0.83] \times [0.32; 0.74]$. Compared with the rectangle $\approx [0; 0.83] \times [0; 0.32]$, which contains the 588 pseudo-observations $(\hat{U}_i, \hat{V}_i)$ that correspond to those (X_i, Y_i) with both components larger than 1, these rectangles contain, relative to their size, a rather large number of pseudo-observations $(\hat{U}_i, \hat{V}_i)$, so that abovementioned “breaks” occur. It is clear that an analysis of the bivariate tail of (X, Y) cannot be performed beyond one of these “breaks”. Further, one may not believe that the $m = 200$ points within the square $(0, 1/T^{(n)}_{n-m:n}]^2$ in Figure 5.16 must correspond to the $m = 200$ points within the square $(0, 1/T^{(n)}_{n-m:n}]^2$ in Figure 5.6 and therefore lead to identical results. This would only be the case if the number of pairs (X_i, Y_i) with $X_i < 1$ is identical (or very close) to the number of pairs (X_i, Y_i) with $Y_i < 1$. Geometrically speaking, the rather large difference of these numbers in these data causes a contraction of the square $[0; 1] \times [0; 1]$ containing the 588 pseudo-observations when considering the data with both components larger than 1 to a rectangle with sides length that exhibit a rather large difference when considering the data with at least one component larger than 1.

Model Validation

In the rest of this section, we assume that X and Y are asymptotically independent. However, it must be emphasized that this assumption, due to the results obtained when estimating η, might not apply. This means that the results in the rest of this section must be considered with care.

We obtain that 8 of the points (4.11), i.e. 4%, lie outside their approximate 95% confidence interval (4.14), so that our method accepts the presence of the scaling law (2.17). The pertaining plot (4.11) is shown in Figure 5.17, where the colors of the points are according to Figure 5.7, i.e. reddish (orange, red, pink) points lie outside their confidence interval.



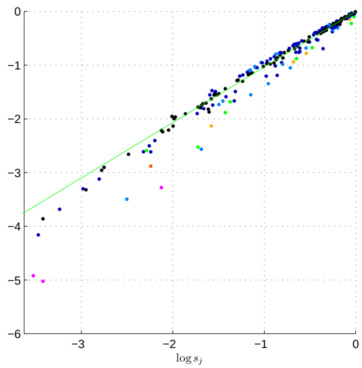


Figure 5.17: Plot (4.11) with $m = 200$, $1/T^{(n)}_{n-m:n} = 0.15$, $\hat{\eta}^H_n = 0.97$ from two perspectives. The right hand perspective also displays the reference plane. 8 (reddish) points lie outside their approximate 95% confidence interval.

Probability of Jointly Large Claims

The plots for estimates $\hat{p}_n$ and approximate 95% confidence intervals for the conditional probability $P(X > u_1, Y > u_2 \mid X \geq 1 \text{ or } Y \geq 1)$ with $u_1 = u_2 = 10$ and $u_1 = 20$, $u_2 = 30$ are given in Figure 5.18, where we used $\hat{G}^H_{1,n}$ and $\hat{G}^H_{2,n}$ as marginal Pareto estimates.

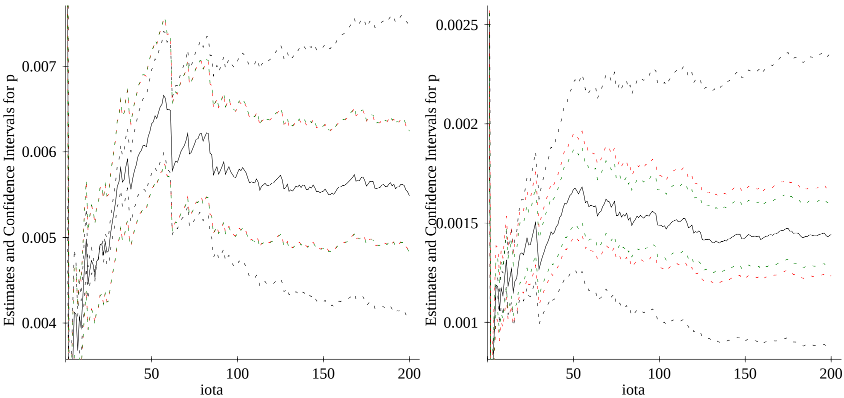


Figure 5.18: Estimation of $P(X > u_1, Y > u_2 \mid X \geq 1 \text{ or } Y \geq 1)$. Plot (3.38) and confidence intervals (3.35), (3.36) and (3.37) with $u_1 = u_2 = 10$ (left) and $u_1 = 20$, $u_2 = 30$ (right).

Table 5.2 compares probability estimates $\hat{p}_n$ and (approximate) 95% confidence intervals for the conditional probability $P(X > u_1, Y > u_2 \mid X \geq 1 \text{ or } Y \geq 1)$ with (confidence intervals for) $\hat{p}_e$ and $\hat{p}_i = (1 - \hat{G}^H_{1,n}(u_1))(1 - \hat{G}^H_{2,n}(u_2))$ for further values of u_1 and u_2. For the reasons given in Section 5.1.1, the confidence intervals (3.35) are again given.

The observations made in Table 5.1 are confirmed.

The $\hat{p}_n$-estimates of $P(X > u_1, Y > u_2 \mid X \geq 1, Y \geq 1)$ in Table 5.1 and of $P(X > u_1, Y > u_2 \mid X \geq 1 \text{ or } Y \geq 1)$ in Table 5.2 are reasonable when compared with each other (and might therefore be considered as confirmation

u_1	u_2	ι	$\hat{p}_n$ (%)	$\hat{p}_e$ (%)	$\hat{p}_i$ (%)
5	5	150	1.697 [1.57,2.09]	1.861 [1.46,2.33]	0.277
10	10	150	0.567 [0.43,0.74]	0.503 [0.31,0.78]	0.035
20	20	150	0.168 [0.11,0.28]	0.503 [0.31,0.78]	0.004
20	30	150	0.127 [0.08,0.21]	0.025 [0.00,0.14]	0.002
25	25	150	0.102 [0.06,0.19]	0.025 [0.00,0.14]	0.002
100	10	150	0.027 [0.01,0.05]	0 [0.00,0.09]	6.7e-4

Table 5.2. Comparison of the estimates $\hat{p}_n$, the empirical probability $\hat{p}_e$ and the estimated probability $\hat{p}_i$ assuming conditional independence of X and Y.

of our calculations). To see this, note that

$$\frac{P\{X > 1, Y > 1\}}{P\{X > 1 \text{ or } Y > 1\}} \approx \frac{\sum_{i=1}^{n} \mathbb{1}\{X_i > 1, Y_i > 1\}}{\sum_{i=1}^{n} \mathbb{1}\{X_i > 1 \text{ or } Y_i > 1\}},$$

so that for $u_1, u_2 \geq 1$, a reasonable estimate of $P(X > u_1, Y > u_2 \mid X \geq 1, Y \geq 1)$ from an estimate of $P(X > u_1, Y > u_2 \mid X \geq 1 \text{ or } Y \geq 1)$ can be motivated by

$$\begin{aligned} P(X > u_1, Y > u_2 \mid X \geq 1, Y \geq 1) &= \frac{P\{X > u_1, Y > u_2\}}{P\{X > 1, Y > 1\}} \\ &\approx \frac{\sum_{i=1}^{n} \mathbb{1}\{X_i > 1 \text{ or } Y_i > 1\}}{\sum_{i=1}^{n} \mathbb{1}\{X_i > 1, Y_i > 1\}} \, \frac{P\{X > u_1, Y > u_2\}}{P\{X > 1 \text{ or } Y > 1\}} \end{aligned}$$

$$= \frac{\sum_{i=1}^{n} \mathbb{1}\{X_i > 1 \text{ or } Y_i > 1\}}{\sum_{i=1}^{n} \mathbb{1}\{X_i > 1, Y_i > 1\}} P(X > u_1, Y > u_2 \mid X > 1 \text{ or } Y > 1).$$

Hence, by multiplying the $\hat{p}_n$-estimates and the pertaining confidence intervals of Table 5.2 with the ratio 3,976/588 of the sample sizes of the data considered here and in Section 5.1.1 we obtain a reasonable estimate $\hat{p}_{o,n}$ and a pertaining confidence interval of $P(X > u_1, Y > u_2 \mid X \geq 1, Y \geq 1)$ (although the latter can only be considered as a very rough heuristic). The results are shown in Table 5.3, where we copied the estimates $\hat{p}_{b,n}$ (renamed to avoid confusion) and the pertaining confidence intervals of $P(X > u_1, Y > u_2 \mid X \geq 1, Y \geq 1)$ obtained in Section 5.1.1 from Table 5.1. Indeed, the estimates $\hat{p}_{b,n}$ and $\hat{p}_{o,n}$ are quite close and may therefore be considered as confirmation of our calculations. Note that although some of the estimates $\hat{p}_{b,n}$ and $\hat{p}_{o,n}$ differ by factor 1.6, these differences are not statistically significant, since the estimates always belong to the confidence interval of the respective other estimate.

u_1	u_2	$\hat{p}_{b,n}$ (%)	$\hat{p}_{o,n}$ (%)
5	5	12.49 [11.25,13.69]	11.47 [10.62,14.13]
10	10	3.492 [2.80,4.34]	3.834 [2.91,5.00]
20	20	0.751 [0.49,1.15]	1.141 [0.74,1.89]
20	30	0.549 [0.32,0.89]	0.859 [0.54,1.42]
25	25	0.442 [0.26,0.70]	0.690 [0.41,1.29]
100	10	0.180 [0.10,0.45]	0.183 [0.07,0.34]

Table 5.3. Comparison of the estimates of $P(X > u_1, Y > u_2 \mid X \geq 1, Y \geq 1)$ when considering the data with both components larger than 1 ($\hat{p}_{b,n}$, cf. Table 5.1) and when considering the data with at least component larger than 1 ($\hat{p}_{o,n}$).

We conclude this section with some remarks on the choice of the data for the statistical analysis. For the Danish fire insurance data, either method – analyzing the pairs (X_i, Y_i) of claims with both components larger than 1 and analyzing the pairs (X_i, Y_i) of claims with at least one component larger than 1 – leads to very similar estimates for the probability of jointly large claims p. However, there might be situations (i.e. data) in which these estimates differ (statistically) significantly, such that a choice between the two methods has to be made. As things stand now, one should then choose the method which yields the better fit of the data in terms of the according tools used in the model, i.e. Hill plots, qq-plots, model validation and confidence intervals. (For the Danish fire insurance data, the method to choose would rather be the analysis of the pairs (X_i, Y_i) of claims with both components larger than 1.) For future research, it might be interesting to theoretically study the impact of choosing one of these methods on the model parameters, in particular on the coefficient of tail dependence η. (Note that the difference of the estimates of η using the different methods is statistically significant.) Further, the influence of thresholds for the sum of the components of a pair of claims which must be exceeded to record the pair is not entirely clear and is therefore an issue to be investigated more thoroughly.

5.2 Medical Claims

In our second application we analyze a data set given in the Society of Actuaries Group Medical Insurance Large Claims Database[11]. The data set contains annual hospital charges X_i and annual other charges Y_i of a risk and refer to the years 1991 and 1992. The claims are recorded only if the sum of both components attains or exceeds at least 25,000 US-Dollar (USD). For a detailed description of the project see Grazier and G'Sell Associates (1997). A summarized description of the data and a univariate extreme value analysis of the total charges $(X_i + Y_i)$ can be found in Cebriàn et al. (2003).

Here, we merely consider the data of the year 1991, for which 92,750 claims with separate information about hospital and other charges are available. For the same reason as with the Danish fire insurance claims we eliminate the artificial negative dependence between the components by consulting a claim (X_i, Y_i) only if either both components X_i and Y_i separately attain or exceed 25,000 USD (i.e. $\min(X_i, Y_i) \geq 25,000$ USD) or if at least one component attains or exceeds 25,000 USD (i.e. $\max(X_i, Y_i) \geq 25,000$ USD). Again, we analyze the resulting data sets in separate sections.

[11] *http://www.soa.org*

5.2.1 Both Components Exceed 25,000 USD

In this section, we analyze the conditional distribution of (X, Y) given $\{X \geq 25, Y \geq 25\}$, where here and henceforth in Section 5.2 we refer to claims in thousands of USD. The sample size of the remaining data is $n = 7,675$. Figure 5.19 displays the scatterplot of the data on a linear and on a logarithmic scale. We do not observe a clear cluster around the diagonal. In the plot on the logarithmic scale the claims seem to exhibit the minor tendency that if one component is large, then the other is rather large than small which suggests a weak positive dependence.

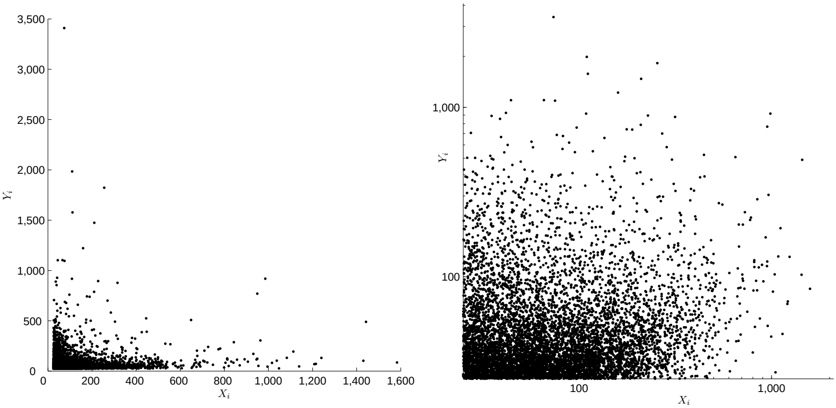


Figure 5.19: Scatterplot Y_i vs. X_i on a linear and on a logarithmic scale.

Fitting of the Marginals

Here, the (heuristic) analysis of the Hill plots (Figures 5.20 and 5.21) suggests $j_1 = j_2 = 800$ which yields

$$1 - \hat{G}^H_{1,n}(x) \approx \left(1 + 0.26\frac{x - 22.00}{58.65}\right)^{-3.78}$$

and

$$1 - \hat{G}^H_{2,n}(y) \approx \left(1 + 0.43\frac{y - 2.43}{26.75}\right)^{-2.33},$$

with $1 - \hat{G}^H_{1,n}$ and $1 - \hat{G}^H_{2,n}$ denoting the estimates (1.10), where $\hat{\mu}_n$ and $\hat{\varsigma}_n$ are estimated according to (1.16), for the conditional d.f.s of X and Y given $\{X \geq 25, Y \geq 25\}$. Here, it is again rather difficult to choose appropriate j_i, in particular for $i = 1$, since a "clear trend" for j_1 greater than some j_0 is not observable. Further, the exponential qq-plot (5.1) for $\left(1 - \hat{G}^H_{1,n}\right)^{\leftarrow}$ in Figure 5.22 exhibits a rather moderately good fit of the data.

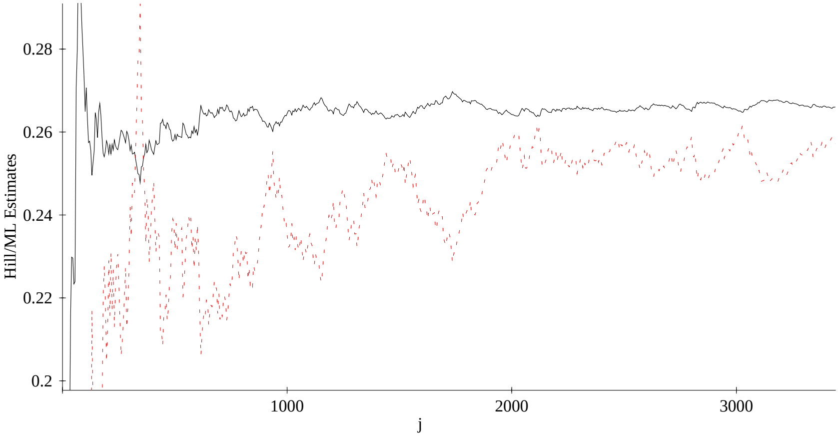


Figure 5.20: Hill plot: $\hat{\gamma}_{1,n}$ as a function of j_1.

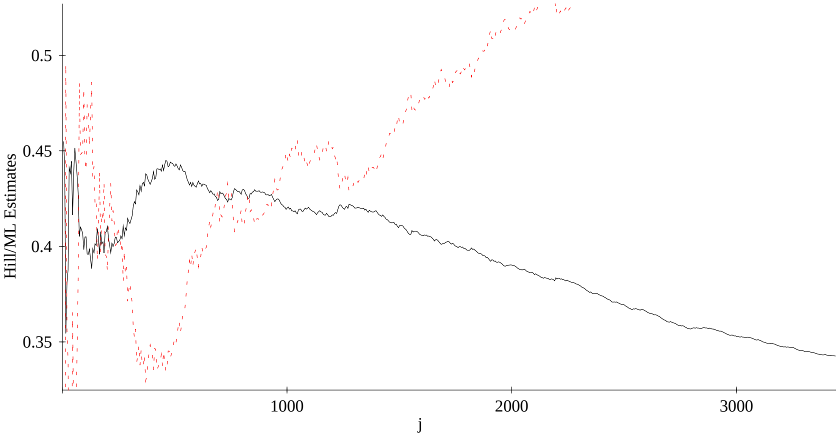


Figure 5.21: Hill plot: $\hat{\gamma}_{2,n}$ as a function of j_2.

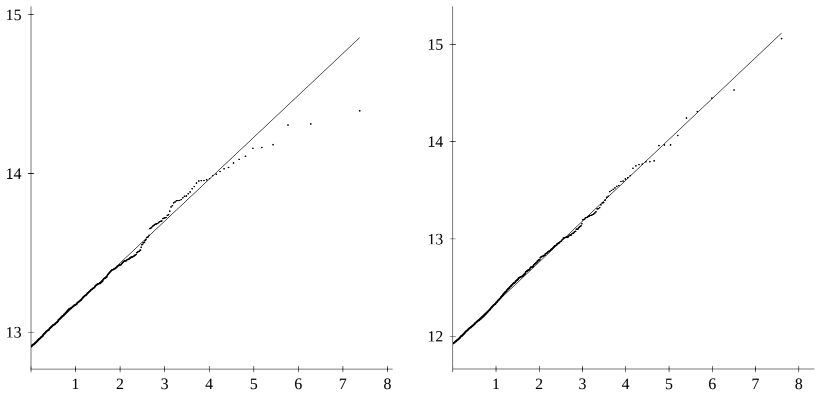


Figure 5.22: Exponential qq-plots for $\left(1-\hat{G}^H_{1,n}\right)^{\leftarrow}$ and $\left(1-\hat{G}^H_{2,n}\right)^{\leftarrow}$.

Estimation of η

The (heuristic) analysis of the Hill plot (Figure 5.23) leads to $m = 1,000$, which yields $1/T^{(n)}_{n-m:n} = 0.34$ (indicated in Figure 5.24, which displays the scatterplot of the pseudo-observations $(\hat{U}_i, \hat{V}_i)$) and $\hat{\eta}^H_n = 0.59$. The approximate 95% confidence interval (3.6) for η is $[0.55, 0.62]$. The pertaining test (3.7) clearly rejects the hypothesis $\eta = 1$ on an approximate 95% confidence level with $(1-p)$-value > 0.9999, so that we assume asymptotic independence. Hence, we indeed identify a weak positive dependence that vanishes asymptotically.

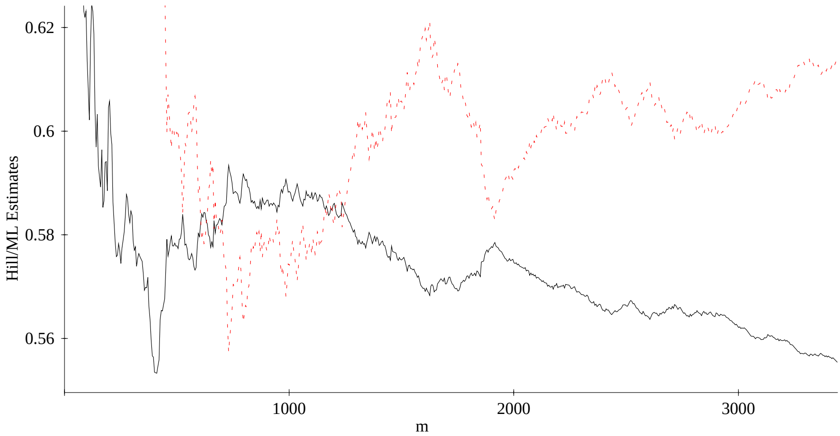


Figure 5.23: Hill plot: $\hat{\eta}_n$ as a function of m.

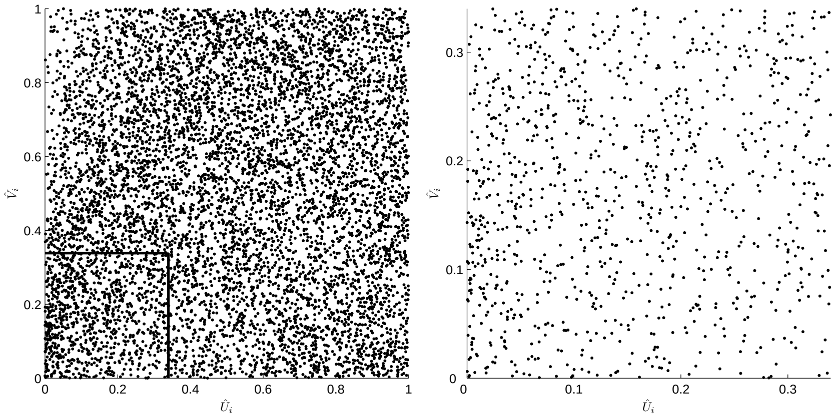


Figure 5.24: Scatterplot $\hat{V}_i$ vs. $\hat{U}_i$. The square with side length $1/T^{(n)}_{n-m:n} = 0.34$ contains the $m = 1,000$ points used for the estimation of η. The scaling law (2.17) is checked within this square.

Model Validation

We obtain that 49 of the points (4.11), i.e. 4.9%, lie outside their approximate 95% confidence interval (4.14), so that our method accepts the presence of the scaling law (2.17). The pertaining plot (4.11) is shown in Figure 5.25, where the colors of the points are according to Figure 5.7, i.e. reddish (orange, red, pink) points lie outside their confidence interval.

For the purpose of comparison, Figure 5.26 displays the plot (4.11) for $m = 3,800$, that is, about half of all data points are used for the model fitting. We make similar observations as in the corresponding example for the Danish fire insurance data in Section 5.1.1. A percentage considerably larger than 5%, namely about 46% (1,729 points), lie outside their approximate 95% confidence interval – evidence that, as expected, the scaling law fails when essentially more data is used for the model fitting than suggested by the Hill plot $\hat{\eta}_n$ versus m. Thus, our method clearly detects the deviations from the scaling law and again demonstrates that it can also be used to determine the maximum tail region for which the application of the ELTM is still reasonable.

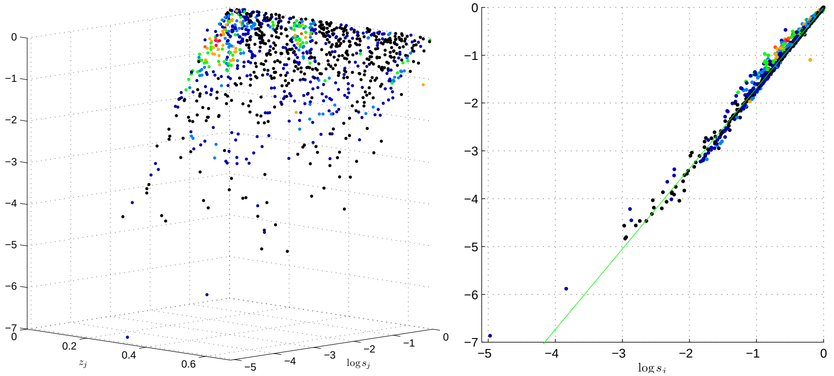


Figure 5.25: Plot (4.11) with $m = 1,000$, $1/T^{(n)}_{n-m:n} = 0.34$, $\hat{\eta}^H_n = 0.59$ from two perspectives. The right hand perspective also displays the reference plane. 49 (reddish) points lie outside their approximate 95% confidence interval.

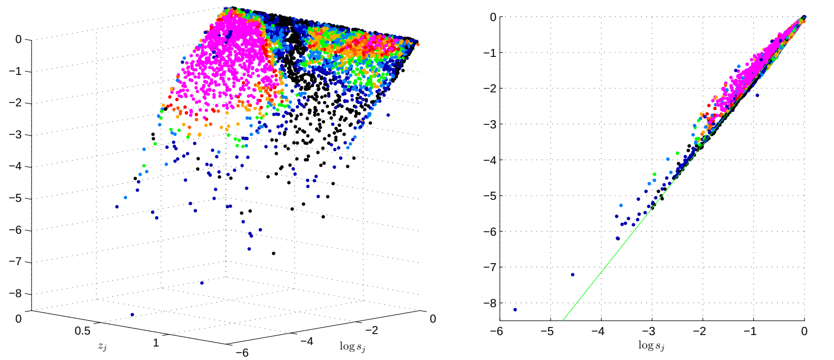


Figure 5.26: Plot (4.11) with $m = 3,800$, $1/T^{(n)}_{n-m:n} = 0.70$, $\hat{\eta}^H_n = 0.55$ from two perspectives. The right hand perspective also displays the reference plane. 1,729 (reddish) points lie outside their approximate 95% confidence interval.

(Moreover, note that although also the Hill plot for η, cf. Figure 5.23, indicates that $m = 3,800$ is too large, the resulting point estimate $\hat{\eta}_n^H = 0.55$ belongs to the confidence interval for η obtained with the choice $m = 1,000$, i.e. the difference is not statistically significant.) Files with the plots of the Figures 5.25 and 5.26 are provided on the website mentioned in Footnote 10.

Probability of Jointly Large Claims

The plots for estimates $\hat{p}_n$ and approximate 95% confidence intervals for the conditional probability $P(X > u_1, Y > u_2 \mid X \geq 25, Y \geq 25)$ with $u_1 = u_2 = 250$ and $u_1 = 600$, $u_2 = 400$ are given in Figure 5.27, where we used $\hat{G}_{1,n}^H$ and $\hat{G}_{2,n}^H$ as marginal Pareto estimates.

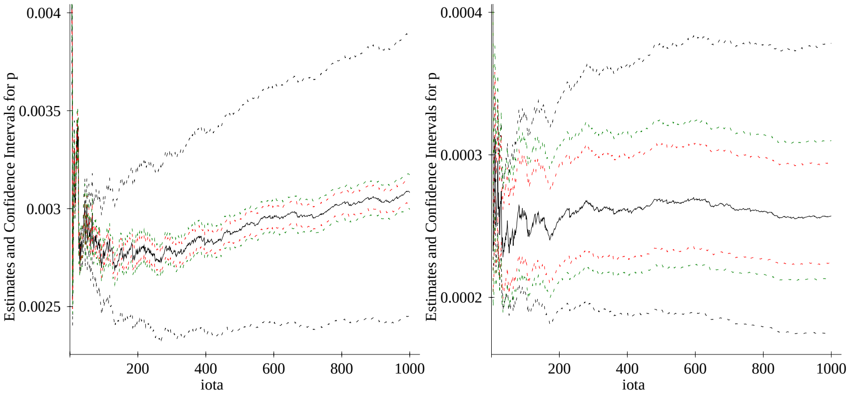


Figure 5.27: Estimation of $P(X > u_1, Y > u_2 \mid X \geq 25, Y \geq 25)$. Plot (3.38) and confidence intervals (3.35), (3.36) and (3.37) with $u_1 = u_2 = 250$ (left) and $u_1 = 600$, $u_2 = 400$ (right).

Table 5.4 compares probability estimates $\hat{p}_n$ and (approximate) 95% confidence intervals for the conditional probability $P(X > u_1, Y > u_2 \mid X \geq 25, Y \geq 25)$ with (confidence intervals for) $\hat{p}_e$ and $\hat{p}_i = (1 - \hat{G}_{1,n}^H(u_1))(1 - \hat{G}_{2,n}(u_2))$ for further values of u_1 and u_2. For the reasons given in Section 5.1.1, we give (3.35) as confidence intervals pertaining to $\hat{p}_n$.

u_1	u_2	ι	$\hat{p}_n$ (%)	$\hat{p}_e$ (%)	$\hat{p}_i$ (%)
200	200	300	0.581 [0.50,0.66]	0.612 [0.45,0.81]	0.388
250	250	300	0.279 [0.23,0.33]	0.326 [0.21,0.48]	0.164
400	400	300	0.052 [0.04,0.06]	0.065 [0.02,0.15]	0.022
600	400	300	0.027 [0.02,0.03]	0.052 [0.01,0.13]	0.007
800	$1,000$	300	0.002 [1e-3,3e-3]	0 [0,0.05]	4.5e-4

Table 5.4. Comparison of the estimates $\hat{p}_n$, the empirical probability $\hat{p}_e$ and the estimated probability $\hat{p}_i$ assuming conditional independence of X and Y.

The observations made in Tables 5.1 and 5.2 are confirmed. In fact, the results in Table 5.4 demonstrate even more impressively that the confidence intervals for $\hat{p}_e$ can be considerably wider than those for $\hat{p}_n$, i.e. the empirical estimates are less accurate. Observe that in contrast to what we have seen for the fire insurance data, in the present example, for large thresholds u_1 and u_2, the empirical probabilities are larger than the estimates $\hat{p}_n$. Of course, for sufficiently large thresholds, the empirical probabilities will be 0 and hence underestimate the real risk, but in general this need not be true for extreme events which have occurred in the past.

5.2.2 At Least One Component Exceeds 25,000 USD

In this section, we analyze the conditional distribution of (X, Y) given $\{X \geq 25 \text{ or } Y \geq 25\}$. From the 92,750 pairs of claims of the original data set, there are 30,010 pairs with both components smaller than 25. Hence, the sample size of the data investigated in this section is $n = 62,740$. Figure 5.28 displays the scatterplot of the data on a linear and on a logarithmic scale. Here, there are 1,972 pairs (X_i, Y_i) with $X_i = 0$ and 1,388 pairs (X_i, Y_i) with $Y_i = 0$, which cannot not be represented in the plot on the logarithmic scale.

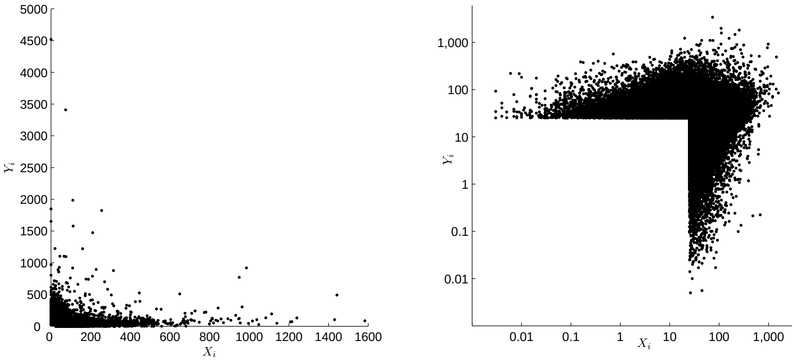


Figure 5.28: Scatterplot Y_i vs. X_i on a linear and on a logarithmic scale.

Fitting of the Marginals

The (heuristic) analysis of the Hill plots (Figures 5.29 and 5.30) suggests $j_1 = j_2 = 1,500$, which yields

$$1 - \hat{G}^H_{1,n}(x) \approx \left(1 + 0.27\frac{x + 30.54}{32.06}\right)^{-3.74}$$

and

$$1 - \hat{G}^H_{2,n}(y) \approx \left(1 + 0.35\frac{y + 15.75}{18.77}\right)^{-2.89},$$

with $1 - \hat{G}^H_{1,n}$ and $1 - \hat{G}^H_{2,n}$ denoting the estimates (1.10), where $\hat{\mu}_n$ and $\hat{\varsigma}_n$ are estimated according to (1.16), for the conditional d.f.s of X and Y given $\{X \geq 25 \text{ or } Y \geq 25\}$. In Figure 5.31, the estimates are validated by exponential qq-plots (5.1) for $\left(1 - \hat{G}^H_{i,n}\right)^{\leftarrow}$, $i = 1, 2$.

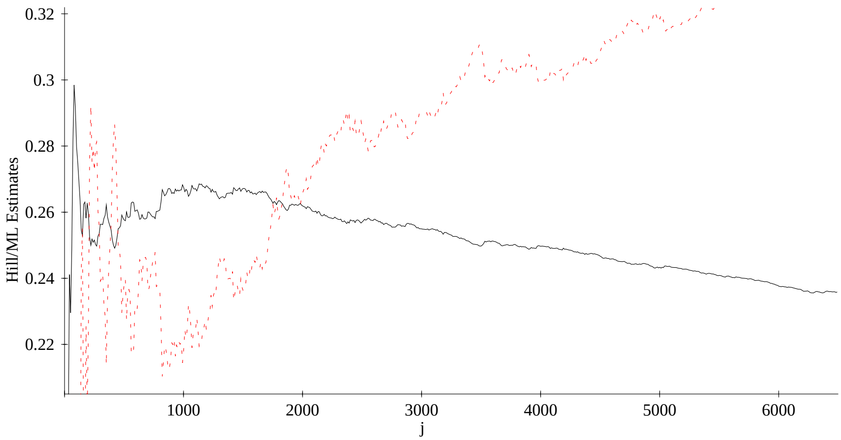


Figure 5.29: Hill plot: $\hat{\gamma}_{1,n}$ as a function of j_1.

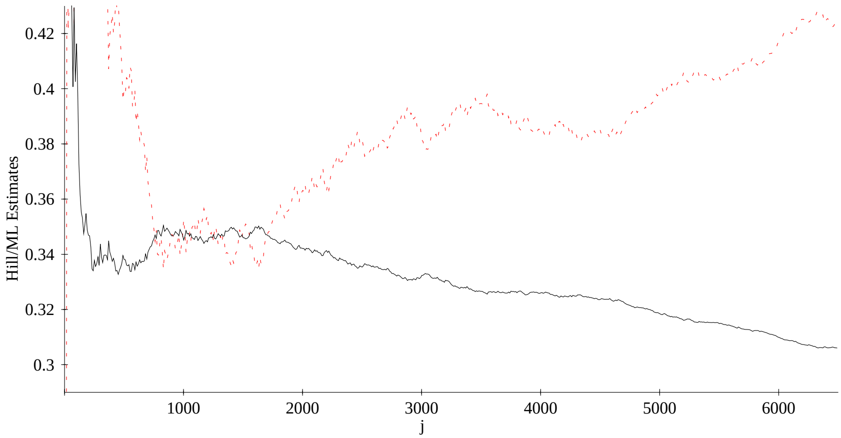


Figure 5.30: Hill plot: $\hat{\gamma}_{2,n}$ as a function of j_2.

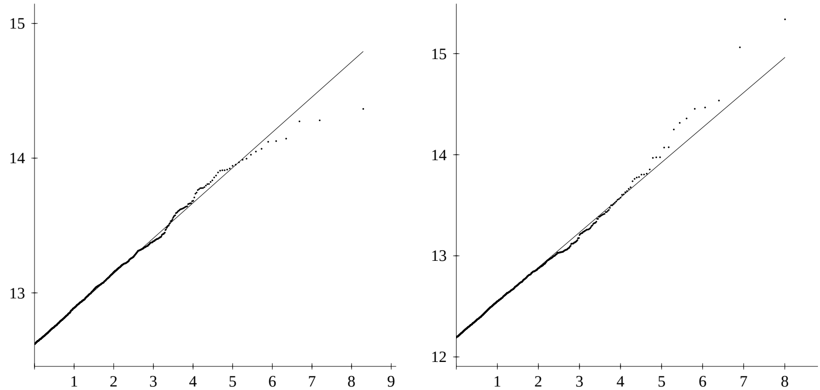


Figure 5.31: Exponential qq-plots for $\left(1-\hat{G}^H_{1,n}\right)^{\leftarrow}$ and $\left(1-\hat{G}^H_{2,n}\right)^{\leftarrow}$.

Estimation of η

The (heuristic) analysis of the Hill plot (Figure 5.32) leads to $m = 4,000$ (where in fact any choice between $m = 2,000$ and $m = 6,000$ is reasonable), which yields $1/T^{(n)}_{n-m:n} = 0.31$ and $\hat{\eta}^H_n = 0.71$. The approximate 95% confidence interval (3.6) for η is $[0.69, 0.73]$. The pertaining test (3.7) clearly rejects the hypothesis $\eta = 1$ on an approximate 95% confidence level with $(1-p)$-value > 0.9999, so that we assume moderate asymptotic independence that vanishes asymptotically.

Figure 5.33 displays the scatterplot of the pseudo-observations $(\hat{U}_i, \hat{V}_i)$ including the square with side length $1/T^{(n)}_{n-m:n} = 0.31$, where the same particularities as in Figure 5.16, examined in Section 5.1.2, occur.

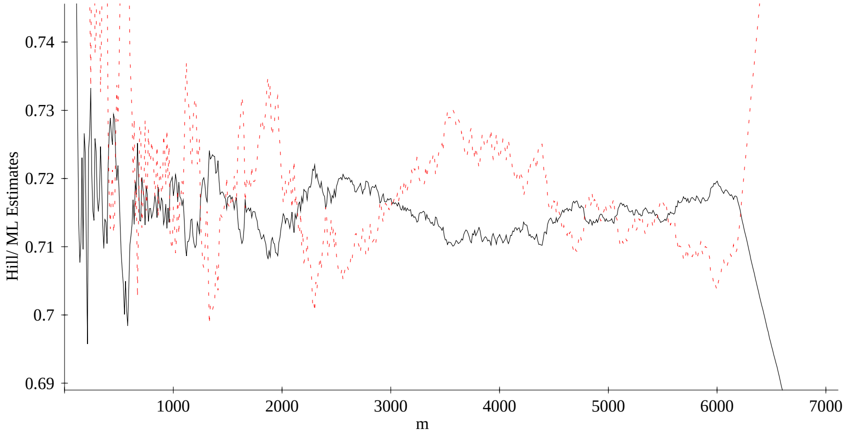


Figure 5.32: Hill plot: $\hat{\eta}_n$ as a function of m.

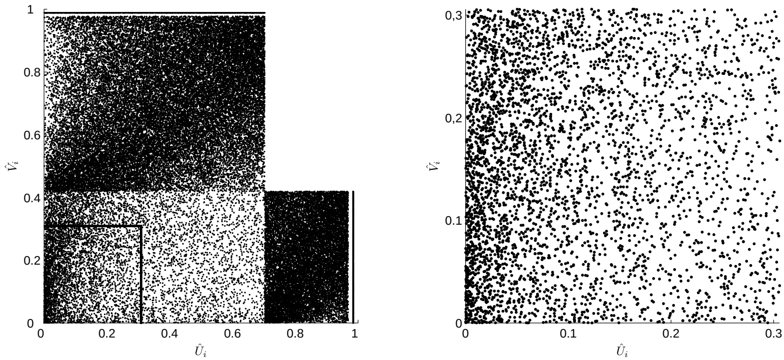


Figure 5.33: Scatterplot $\hat{V}_i$ vs. $\hat{U}_i$. The square with side length $1/T^{(n)}_{n-m:n} = 0.31$ contains the $m = 4,000$ points used for the estimation of η. The scaling law (2.17) is checked within this square.

Model Validation

We obtain that 173 of the points (4.11), i.e. 4.3%, lie outside their approximate 95% confidence interval (4.14), so that our method accepts the presence of the scaling law (2.17). The pertaining plot (4.11) is shown in Figure 5.34, where the colors of the points are according to Figure 5.7, i.e. reddish (orange, red, pink) points lie outside their confidence interval.

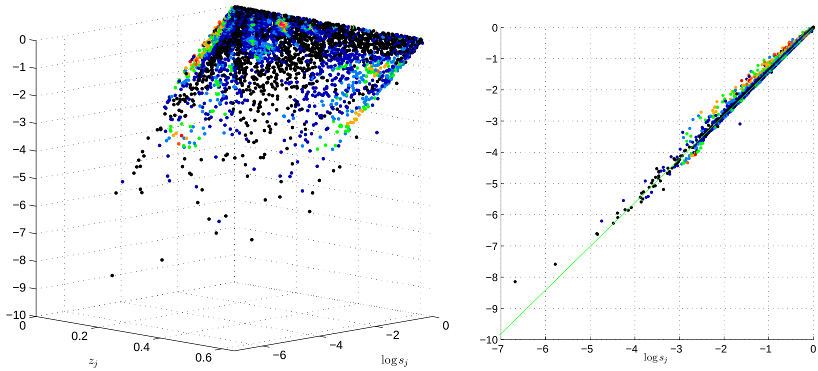


Figure 5.34: Plot (4.11) with $m = 4,000$, $1/T^{(n)}_{n-m:n} = 0.31$, $\hat{\eta}^H_n = 0.71$ from two perspectives. The right hand perspective also displays the reference plane. 173 (reddish) points lie outside their approximate 95% confidence interval.

Probability of Jointly Large Claims

The plots for estimates $\hat{p}_n$ and approximate 95% confidence intervals for the conditional probability $P(X > u_1, Y > u_2 \mid X \geq 25 \text{ or } Y \geq 25)$ with $u_1 = u_2 = 250$ and $u_1 = 600$, $u_2 = 400$ are given in Figure 5.35, where we used $\hat{G}^H_{1,n}$ and $\hat{G}^H_{2,n}$ as marginal Pareto estimates.

Table 5.5 compares probability estimates $\hat{p}_n$ and (approximate) 95% confidence intervals for the conditional probability $P(X > u_1, Y > u_2 \mid X \geq 25 \text{ or } Y \geq 25)$ with (confidence intervals for) $\hat{p}_e$ and $\hat{p}_i = (1 - \hat{G}^H_{1,n}(u_1))(1 - \hat{G}^H_{2,n}(u_2))$ for further values of u_1 and u_2. For the reasons given in Section 5.1.1, we give (3.35) as confidence intervals pertaining to $\hat{p}_n$.

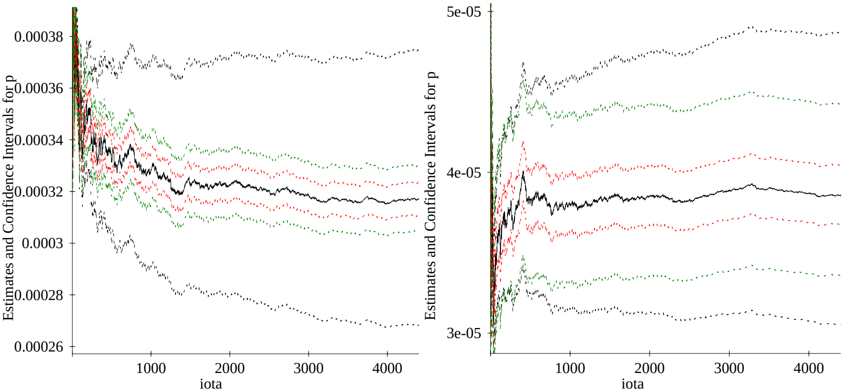


Figure 5.35: Estimation of $P(X > u_1, Y > u_2 \mid X \geq 25 \text{ or } Y \geq 25)$. Plot (3.38) and confidence intervals (3.35), (3.36) and (3.37) with $u_1 = u_2 = 250$ (left) and $u_1 = 600$, $u_2 = 400$ (right).

u_1	u_2	ι	$\hat{p}_n$ (%)	$\hat{p}_e$ (%)	$\hat{p}_i$ (%)
200	200	4,000	0.064 [0.05,0.08]	0.075 [0.06,0.10]	0.020
250	250	4,000	0.032 [0.03,0.04]	0.040 [0.03,0.06]	0.007
400	400	4,000	0.006 [5e-3,7e-3]	0.008 [0.00,0.02]	6.5e-4
600	400	4,000	0.004 [3e-3,5e-3]	0.006 [0.00,0.02]	2.0e-4
800	$1,000$	4,000	3e-4 [2e-4,4e-4]	0 [0,0.01]	7.7e-8

Table 5.5. Comparison of the estimates $\hat{p}_n$, the empirical probability $\hat{p}_e$ and the estimated probability $\hat{p}_i$ assuming conditional independence of X and Y.

The observations made in the Tables 5.1, 5.2 and 5.4 are confirmed.

Like with the Danish fire insurance claims, we obtain that the $\hat{p}_n$-estimates of $P(X > u_1, Y > u_2 \mid X \geq 25, Y \geq 25)$ and of $P(X > u_1, Y > u_2 \mid X \geq 25 \text{ or } Y \geq 25)$ are reasonable when compared with each other. Analogous to what is discussed in Section 5.1.2, when multiplying an estimate of $P(X > u_1, Y > u_2 \mid X \geq 25$ or $Y \geq 25)$ with the ratio 62,740/7,675 to calculate a reasonable estimate of $P(X > u_1, Y > u_2 \mid X \geq 25, Y \geq 25)$, we obtain values which are very close to those of $\hat{p}_n$ in Table 5.4. For example, if $u_1 = u_2 = 250$, we have $0.032 \cdot 62,740/7,675 = 0.262$ as an estimate for $P(X > u_1, Y > u_2 \mid X \geq 25, Y \geq 25)$, which is very close to 0.279 in Table 5.4.

Estimation of the Limiting Function c

In this final subsection of the chapter we apply the estimator $\hat{c}$ to estimate $c(x,1)$ and $c(1,y)$ for $x, y \in [0,1]$ and briefly interpret the results. In order to obtain a plot with easily interpretable curves, we include both estimates in one plot. To this end, define the univariate function $c : [0,2] \to [0,1]$ and its estimator $\hat{c}_n$ (cf. (3.21)) by

$$c(z) = \begin{cases} c(z,1) & \text{if } z \in [0,1], \\ c(1, 2-z) & \text{if } z \in (1,2] \end{cases} \quad \text{and} \quad \hat{c}_n(z) = \begin{cases} \hat{c}_n(z,1) & \text{if } z \in [0,1], \\ \hat{c}_n(1, 2-z) & \text{if } z \in (1,2]. \end{cases}$$

Figure 5.36 illustrates the meaning of the argument z of function c. Figure 5.37 displays $\hat{c}_n(z)$ versus $z \in [0,2]$ and therefore provides estimates for $(c(x,1))_{x\in[0,1]} = (c(z))_{z\in[0,1]}$ and $(c(1,y))_{y\in[0,1]} = (c(z))_{z\in[1,2]}$. The confidence intervalls are computed according to (3.18) with $\check{c}_n(x,y)$ replaced with $\hat{c}_n(x,1)$ and $\hat{c}_n(1,y)$, respectively, see also Section 3.2.2. Note that if U and V (and hence X and Y) were independent, $c(z) = c(z,1) = \lim_{t\downarrow 0} P\{U < tz, V < t\}/P\{U < t, V < t\} = z$ for $z \in [0,1]$. Hence, together with an analogous argument for $z \in (1,2]$, one would expect an approximate triangle with vertices $(0,0)$, $(1,1)$ and $(2,0)$ in the plot of Figure 5.37 in this case. While the right half of the plot, i.e. the estimates of $c(1,y)$, satisfies this expectation, the left half, i.e. the estimates of $c(x,1)$, exhibits a clear deviation from the line $c(z) = z$. Loosely speaking, this means that small values of V may be expected to cause small values of U rather than the other way round. In other words, we may conclude that large other charges cause large hospital charges rather than vice versa.

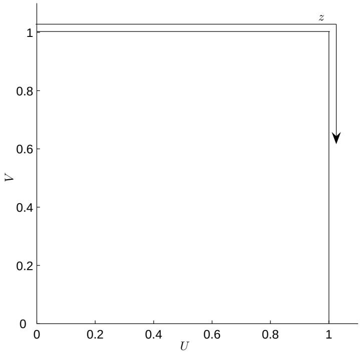


Figure 5.36: Argument z of function c.

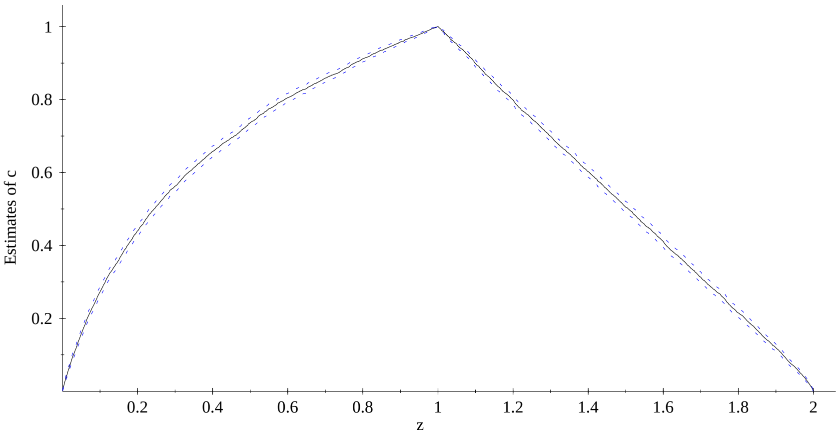


Figure 5.37: Estimation of $c(x, 1)$ and $c(1, y)$ via $\hat{c}_n(z)$ including the confidence intervals (3.18).

Chapter 6

A Related Copula Approach

In this final chapter we investigate a copula model that is related to the ELTM. We show that although the approach introduced by Juri and Wüthrich (2002, 2003) and extended by Charpentier and Juri (2006) requires similar conditions as the ELTM, useful information, which the ELTM exploits, is lost. We see that due to this fact, Juri's and Wüthrich's approach does not overcome the main problems and drawbacks of the widely-used copula approaches (cf. Preface) and CMEVT (see Section 1.2.3). We state the main result of Juri and Wüthrich (2003) in Section 6.1, indicate difficulties with statistical inference and that the selection of the model is similarly arbitrary as described for these widely-used copula models. In Section 6.2, we present the relation of Juri's and Wüthrich's approach to the ELTM and demonstrate that Juri's and Wüthrich's approach exhibits similar problems as the CMEVT in the case of asymptotic independence.

Unless otherwise stated, we refer to the latest paper Charpentier and Juri (2006).

6.1 Juri's and Wüthrich's Approach

The random variables U and V are assumed to be standard uniform with joint d.f. (copula) Q that is strictly increasing in each argument. The results obtained by the approach discussed in this section refer to what Juri and

Wüthrich (2003) call *lower tail dependence copula relative to* Q, defined by

$$Q^{lo}_{u,v}(x,y) := P\big(U \le F^{\leftarrow}_{u,v,1}(x), V \le F^{\leftarrow}_{u,v,2}(y) \mid U \le u, V \le v\big)$$

$$= \frac{Q\left(F^{\leftarrow}_{u,v,1}(x), F^{\leftarrow}_{u,v,2}(y)\right)}{Q(u,v)},$$

where $F^{\leftarrow}_{u,v,1}$ and $F^{\leftarrow}_{u,v,2}$ are the generalized left-continuous inverses of

$$F_{u,v,1}(x) := P(U \le x \mid U \le u, V \le v) = \frac{Q(x,v)}{Q(u,v)} \qquad 0 \le x \le u$$

and

$$F_{u,v,2}(y) := P(V \le y \mid U \le u, V \le v) = \frac{Q(u,y)}{Q(u,v)} \qquad 0 \le y \le v.$$

We remark on the choice of $Q^{lo}_{u,v}$ as the matter of interest in the following Section 6.2.

In essence, Theorem 3.1 of Charpentier and Juri (2006) states that if there is a measurable function $g : (0,\infty)^2 \to (0,\infty)$ such that

$$\lim_{t \downarrow 0} \frac{Q\big(u(t)x, v(t)y\big)}{Q\big(u(t), v(t)\big)} = g(x,y) \tag{6.1}$$

for all $x, y > 0$ with some strictly increasing, continuous, regularly varying auxiliary functions $u, v : (0,\infty) \to (0,\infty)$ that tend to 0 as $t \downarrow 0$, then, for all $(x,y) \in [0,1]^2$,

$$\lim_{t \downarrow 0} Q^{lo}_{u(t),v(t)}(x,y) = g\big(g_1^{\leftarrow}(x), g_2^{\leftarrow}(y)\big), \tag{6.2}$$

where $g_1(\cdot) = g(\cdot, 1)$ and $g_2(\cdot) = g(1, \cdot)$ and g satisfies a homogeneity property. In particular, $g(x,x) = x^\theta$ for some $\theta > 0$ if both u and v are regularly varying with index 1.

Except for the case that Q belongs to the class of strict Archimedean copula, i.e. there is a strictly decreasing, convex function $\psi : [0,1] \to [0,\infty]$ with $\psi(0) = \infty$ and $\psi(1) = 0$ such that $Q(x,y) = \psi^{\leftarrow}\big(\psi(x) + \psi(y)\big)$, it is not specified how to estimate the function g or the parameter θ. If one assumes that Q is strictly Archimedean and satisfies further regularity conditions, then the right hand side of (6.2) is the Clayton copula $Q^{Cl}(x,y) = (x^{-\alpha} + y^{-\alpha} - 1)^{-1/\alpha}$ with $\alpha > 0$ (cf. Theorem 3.4 of Juri and Wüthrich (2003)). In this case, an estimate of the limiting d.f. might be obtained by an estimate of the parameter α. Essentially, Juri and Wüthrich (2003) give a justification to use the Clayton copula to fit bivariate pseudo-observations, if Q is strictly Archimedean. However, in real world problems, it is not clear how to check whether Q is strictly Archimedean, so that the use of the Clayton copula is as arbitrary as to use any copula. Further, since the estimation of tail probabilities in this approach must be based on the limiting distribution $g\big(g_1^{\leftarrow}(x), g_2^{\leftarrow}(y)\big)$ of the lower tail dependence copula $Q^{lo}_{u,v}$, the important case of asymptotic independence can only be treated as perfect independence as we demonstrate by examples in the following section. In this sense, Juri's and Wüthrich's approach exhibits similar problems as the CMEVT, see Section 1.2.3.

6.2 Comparison with the Model of Ledford and Tawn

In order to ease comparison, we rewrite (2.9) with $q(t) = P\{U < t, V < t\}$ of the ELTM in terms of Q, that is

$$\lim_{t \downarrow 0} \frac{Q(tx, ty)}{Q(t,t)} = c(x,y). \tag{6.3}$$

Here, it becomes obvious that both models assume a similar type of conditions: Condition (6.1) with $u = v = \mathrm{id}$ gives (6.3), where c and $1/\eta$ play the roles of g and θ.

However, by the ELTM we are able to describe the behavior of (X,Y) for "large" values by the (GPD-approximations of) the marginal distributions and an approximation c of the copula Q nearby the origin; whereas, as we have indicated in Section 6.1, when working with the limit $g\big(g_1^{\leftarrow}(x), g_2^{\leftarrow}(y)\big)$ of the lower tail dependence copula $Q^{lo}_{u,v}$ (with or without (GPD-approximations of) the marginal distributions), it is not clear how probabilities of jointly large claims are to be estimated. We see in the following Examples 6.2.1 and 6.2.2 that $g\big(g_1^{\leftarrow}(x), g_2^{\leftarrow}(y)\big)$ is not sufficient to model the tail of the bivariate distribution of (X,Y) (or (U,V)). In this sense, the choice of $Q^{lo}_{u,v}$ and $g\big(g_1^{\leftarrow}(x), g_2^{\leftarrow}(y)\big)$ as the matter of interest seems to be at least questionable. The ELTM is able to exploit more (essential) information than the limiting function $g\big(g_1^{\leftarrow}(x), g_2^{\leftarrow}(y)\big)$ of $Q^{lo}_{u,v}$. This is in particular due to the information provided by η – the speed of convergence to asymptotic independence. The following definitions will be convenient. Let

$$c_1(\cdot) := c(\cdot, 1) \qquad \text{and} \qquad c_2(\cdot) := c(1, \cdot).$$

6.2.1 Examples.

(1) Example 2.1 of Draisma et al. (2004) states that the bivariate normal distribution with mean 0, variance 1 and correlation coefficient ρ satisfies Condition 2.2.1 with

$$\eta = \frac{1+\rho}{2} \qquad \text{and} \qquad c(x,y) = (xy)^{1/(1+\rho)}.$$

Hence,

$$c_1(x) = c_2(x) = x^{1/(1+\rho)} \qquad \text{and} \qquad c_1^{\leftarrow}(x) = c_2^{\leftarrow}(x) = x^{1+\rho}.$$

We obtain

$$g\big(g_1^{\leftarrow}(x), g_2^{\leftarrow}(y)\big) = \big(x^{1+\rho}y^{1+\rho}\big)^{1/(1+\rho)} = xy.$$

Comparing c and $g\big(g_1^{\leftarrow}(x), g_2^{\leftarrow}(y)\big)$, we see that it is exactly the information contained in η, i.e. the speed of convergence to asymptotic independence, which is lost when employing Juri's and Wüthrich's approach. It is clear that $g\big(g_1^{\leftarrow}(x), g_2^{\leftarrow}(y)\big) = xy$ is not an appropriate model for the tail of the bivariate distribution of (X,Y) (or (U,V)). An approximation based on $g\big(g_1^{\leftarrow}(x), g_2^{\leftarrow}(y)\big) = xy$ implies perfect independence, so that this approach exhibits similar problems as the CMEVT in the case of asymptotic independence, see Section 1.2.3.

(2) We consider the so-called *logistic model*, a popular paramatric approach within the class of bivariate extreme value distributions, see e.g. Section 8.2 of Coles (2001), in particular (8.10), where the marginals are standardized to standard Frechét distribution, i.e. $F_j(x) = e^{-1/x}$, $x > 0$, $j = 1, 2$. The logistic copula is

$$Q(u,v) = \exp\left\{-\left[\big(\log u^{-1}\big)^r + \big(\log v^{-1}\big)^r\right]^{1/r}\right\}.$$

Since as $t \downarrow 0$

$$\begin{aligned}
Q(tx, ty) &= \exp\left\{-\left[(-\log t - \log x)^r + (-\log t - \log y)^r\right]^{1/r}\right\} \\
&= \exp\left\{\log t\left[\left(1+\frac{\log x}{\log t}\right)^r + \left(1+\frac{\log y}{\log t}\right)^r\right]^{1/r}\right\} \\
&= \exp\left\{\log t\left[1 + r\frac{\log x}{\log t} + O(1/\log^2 t)\right.\right. \\
&\qquad\qquad \left.\left. +1 + r\frac{\log y}{\log t} + O(1/\log^2 t)\right]^{1/r}\right\} \\
&= \exp\left\{\log t\left[2 + r\frac{\log xy}{\log t} + O(1/\log^2 t)\right]^{1/r}\right\} \\
&= \exp\left\{2^{1/r}\log t\left[1 + r\frac{\log xy}{2\log t} + O(1/\log^2 t)\right]^{1/r}\right\} \\
&= \exp\left\{\log t^{2^{1/r}}\left[1 + \frac{\log xy}{2\log t} + O(1/\log^2 t)\right]\right\} \\
&= \exp\left\{\log t^{2^{1/r}} + 2^{1/r-1}\log xy + O(1/\log t)\right\} \\
&= t^{2^{1/r}}(xy)^{2^{1/r-1}}(1 + O(1/\log t)),
\end{aligned}$$

we obtain

$$c(x, y) = (xy)^{2^{1/r-1}}$$

and hence $\eta = 2^{-1/r}$, whereas again

$$g\big(g_1^{\leftarrow}(x), g_2^{\leftarrow}(y)\big) = \left(x^{2^{1-1/r}}y^{2^{1-1/r}}\right)^{2^{1/r-1}} = xy.$$

△

The previous examples exhibit that the information provided by η, i.e. the information about the speed of the convergence to asymptotic independence is lost when applying Juri's and Wüthrich's approach. This is not by accident, since we can calculate $c = g$, if symmetric (i.e. $g_1 = g_2$), from $g\big(g_1^{\leftarrow}(x), g_1^{\leftarrow}(y)\big)$ and η. To see this, observe that $g\big(g_1^{\leftarrow}(x), g_1^{\leftarrow}(x)\big) = \big(g_1^{\leftarrow}(x)\big)^{\theta}$ so that $g_1^{\leftarrow}(x) =$

$\big(g\big(g_1^{\leftarrow}(x), g_1^{\leftarrow}(x)\big)\big)^{1/\theta}$ and hence g_1 and $c(x,y) = g\big(g_1^{\leftarrow}(g_1(x)), g_1^{\leftarrow}(g_1(y))\big)$ can be calculated. The knowledge (estimation, respectively) of $\theta = 1/\eta$ provides the missing information if $c = g$ is symmetric. However, even more information is lost if this symmetry cannot be assumed as the following example shows.

6.2.2 Example. (Asymmetric Example of Tawn (1988)) Tawn (1988), Section 5 specifies the joint survival function $\bar{F}(x,y) = P\{X > x, Y > y\}$ of the asymmetric logistic model if the marginals X and Y are unit exponential, i.e. if $F_j(x) = 1 - e^{-x}$, $x > 0$, $j = 1, 2$, by

$$\bar{F}(x,y) = \exp\left\{-(1-\theta)x - (1-\phi)y - (x^r\theta^r + y^r\phi^r)^{1/r}\right\},$$

where $0 \le \theta, \phi \le 1$, $r \ge 1$. This leads to the copula

$$\begin{aligned}
Q(u,v) &= \bar{F}(-\log u, -\log v) \\
&= \exp\left\{\log u^{1-\theta} + \log v^{1-\phi} - \left[\left(\log u^{-\theta}\right)^r + \left(\log v^{-\phi}\right)^r\right]^{1/r}\right\} \\
&= u^{1-\theta}v^{1-\phi}\exp\left\{-\left[\left(\log u^{-\theta}\right)^r + \left(\log v^{-\phi}\right)^r\right]^{1/r}\right\}.
\end{aligned}$$

We slightly compress the calculation of $Q(tx, ty)$, since it is very similar to the case of the (symmetric) logistic model. Since as $t \downarrow 0$

$$
\begin{aligned}
& Q(tx,ty)(tx)^{-(1-\theta)}(ty)^{-(1-\phi)} \\
&= \exp\left\{ -\left[\left(-\log t^{\theta} - \log x^{\theta} \right)^{r} + \left(-\log t^{\phi} - \log y^{\phi} \right)^{r} \right]^{1/r} \right\} \\
&= \exp\left\{ \log t \left[\theta^{r} \left(1 + \frac{\log x}{\log t} \right)^{r} + \phi^{r} \left(1 + \frac{\log y}{\log t} \right)^{r} \right]^{1/r} \right\} \\
&= \exp\left\{ \log t \left[\theta^{r} + \phi^{r} + r \frac{\theta^{r} \log x + \phi^{r} \log y}{\log t} + O\big(1/\log^{2} t\big) \right]^{1/r} \right\} \\
&= \exp\left\{ (\theta^{r} + \phi^{r})^{1/r} \log t \left[1 + r(\theta^{r} + \phi^{r})^{-1} \frac{\log x^{\theta^{r}} + \log y^{\phi^{r}}}{\log t} \right.\right. \\
&\qquad\qquad \left.\left. + O\big(1/\log^{2} t\big) \right]^{1/r} \right\} \\
&= \exp\left\{ \log t^{(\theta^{r}+\phi^{r})^{1/r}} \left[1 + \frac{\log x^{\theta^{r}(\theta^{r}+\phi^{r})^{-1}} + \log y^{\phi^{r}(\theta^{r}+\phi^{r})^{-1}}}{\log t} \right.\right. \\
&\qquad\qquad \left.\left. + O\big(1/\log^{2} t\big) \right] \right\} \\
&= t^{(\theta^{r}+\phi^{r})^{1/r}} x^{\theta^{r}(\theta^{r}+\phi^{r})^{1/r-1}} y^{\phi^{r}(\theta^{r}+\phi^{r})^{1/r-1}} \big(1 + O(1/\log t)\big),
\end{aligned}
$$

we obtain

$$
c(x,y) = x^{1-\theta+\theta^{r}(\theta^{r}+\phi^{r})^{1/r-1}} y^{1-\phi+\phi^{r}(\theta^{r}+\phi^{r})^{1/r-1}}
$$

and thus $1/\eta = 2 - \theta - \phi + (\theta^{r} + \phi^{r})^{1/r}$, whereas once more

$$
g\big(g_1^{\leftarrow}(x), g_2^{\leftarrow}(y)\big) = xy.
$$

$\triangle$

This example demonstrates that even more than the information provided by η can be lost if only the limit of $Q^{lo}_{u,v}$ is considered. The function c does not only provide the information of η, but also that there is asymmetry in the tails of the bivariate distribution of (X, Y). Neither this, nor the fact that we do not face perfect independence is reflected in $g\big(g_1^{\leftarrow}(x), g_2^{\leftarrow}(y)\big) = xy$.

Appendix A

The Spaces $C\big([0,\infty)^2\big)$, $D\big([0,\infty)^2\big)$ and $D^*\big([0,\infty)^2\big)$

We give appropriate definitions of the spaces $C\big([0,\infty)^2\big)$, $D\big([0,\infty)^2\big)$ and of a simple modification of the latter. We establish some properties of these spaces which are used in some proofs of this thesis. In particular, we prove the separability of $C\big([0,\infty)^2\big)$ when it is endowed with an appropriate metric. This separability is a necessary condition to apply the Skorohod-Dudley-Wichura theorem.

We start with a generalization of the notion of one-sided limits for functions defined on $[0,\infty)$ in order to define the space $D\big([0,\infty)^2\big)$ appropriately. We use the ideas of Neuhaus (1971). For $(x,y) \in [0,\infty)^2$ we define quadrants $Q_{\delta_1\delta_2}(x,y)$, where $\delta_1, \delta_2 \in \{0,1\}$, by

$$Q_{\delta_1\delta_2}(x,y) := I_{\delta_1}(x) \times I_{\delta_2}(y),$$

where the intervals $I_\delta(z)$ are given by

$$I_\delta(z) := \begin{cases} [0,z) & \text{if } \delta = 0 \\ [z,\infty) & \text{if } \delta = 1. \end{cases}$$

A.1 Definition. Let f be a real-valued function defined on $[0,\infty)^2$. If for every sequence $(x_n, y_n) \subset Q_{\delta_1\delta_2}(x,y)$ with $(x_n,y_n) \to (x,y) \in [0,\infty)^2$ the sequence $f(x_n, y_n)$ converges, then we denote the limit by

$$\begin{aligned}
&f(x-, y-) \text{ if } \delta_1 = \delta_2 = 0,\\
&f(x+, y-) \text{ if } \delta_1 = 1, \delta_2 = 0,\\
&f(x-, y+) \text{ if } \delta_1 = 0, \delta_2 = 1 \text{ and}\\
&f(x+, y+) \text{ if } \delta_1 = \delta_2 = 1.
\end{aligned}$$

We call $f(x-, y-)$ and $f(x+, y+)$ the left hand limit and the right hand limit of f at (x, y), respectively. △

By this definition, we are now ready to define $D\big([0,\infty)^2\big)$.

A.2 Definition. The space $D\big([0,\infty)^2\big)$ is the set of all functions $f : [0,\infty)^2 \to \mathbb{R}$ for which $f(x-, y-)$, $f(x+, y-)$ and $f(x-, y+)$ exist and which are right continuous in the sense that $f(x, y) = f(x+, y+)$ for every $(x, y) \in [0,\infty)^2$. The function f belongs to the space $C\big([0,\infty)^2\big)$, iff in addition $f(x, y) = f(x-, y-) = f(x+, y-) = f(x-, y+)$ holds for every $(x, y) \in [0,\infty)^2$. The space $D^*\big([0,\infty)^2\big)$ is defined analogously to $D\big([0,\infty)^2\big)$ with the roles of $f(x-, y-)$ and $f(x+, y+)$ interchanged, where we call $f \in D^*\big([0,\infty)^2\big)$ left continuous. △

Note that $C\big([0,\infty)^2\big) \subset D\big([0,\infty)^2\big)$ as well as $C\big([0,\infty)^2\big) \subset D^*\big([0,\infty)^2\big)$. The following considerations refer to $D\big([0,\infty)^2\big)$. However, they also apply if $D\big([0,\infty)^2\big)$ is substituted with $D^*\big([0,\infty)^2\big)$.

Let f and g be elements of $D\big([0,\infty)^2\big)$. It is easily checked that

$$d_C(f, g) := \sum_{k=1}^{\infty} 2^{-k} \min\left(1, \sup_{(x,y)\in[0,k]^2} |f(x, y) - g(x, y)|\right)$$

defines a metric on $D\big([0,\infty)^2\big)$. The metric d_C is a generalization of the *metric of uniform convergence on compacta*, see Definition 1 of Chapter VI of Pollard (1984). Recall that a subset A of a metric space M is called *dense* in M, if for any element $m \in M$, a neighborhood of m contains at least one element of A. A metric space is called *separable* if it contains a countable, dense subset.

A.3 Lemma. The space $C\big([0,\infty)^2\big)$ endowed with the metric d_C is separable. △

Proof. We show that the polynoms with rational coefficients are dense in $C\big([0,\infty)^2\big)$ with respect to d_C, i.e. for $f \in C\big([0,\infty)^2\big)$ and every $\varepsilon > 0$ there exists such a polynom p with $d_C(p,f) < \varepsilon$.

First, let $k_0 > 0$ such that $2^{-k_0} < \varepsilon/2$ and hence

$$\sum_{k=k_0+1}^{\infty} 2^{-k} = 2^{-(k_0+1)} \sum_{j=0}^{\infty} 2^{-j} < \varepsilon/2.$$

The polynoms with rational coefficients are dense in $C\big([0,k_0)^2\big)$ endowed with the uniform metric, i.e. there exists such a polynom p with

$$\sup_{(x,y)\in[0,k_0]^2} \big|p(x,y) - f(x,y)\big| < \varepsilon/2.$$

Further, observe that $\sum_{k=1}^{k_0} 2^{-k}\ \varepsilon/2 < \varepsilon/2$ and hence

$$d_C(p,f) \leq \sum_{k=1}^{k_0} 2^{-k}\ \frac{\varepsilon}{2} + \sum_{k=k_0+1}^{\infty} 2^{-k} < \varepsilon.$$

$\square$

Summary and Outlook

We summarize this work and its results and suggest possible extensions.

In the Preface, we start to discuss the two main motivations of this work – first, the failing of the classical multivariate extreme value theory when estimating the probability of jointly large claims p in the case of asymptotic independence of the claim sizes, and second, the lack of flexibility (and to some extent the arbitrariness) of many popular copula approaches. These issues are discussed several times in the work, in particular in Section 1.2.3 and Chapter 6, respectively. We extend a model of Ledford and Tawn (1996, 1997, 1998) for the joint tail distribution of bivariate claim sizes which overcomes these obstacles. We already see in Section 2.2.2 that the scaling law which holds in the Extended Ledford and Tawn Model offers an access to the case of asymptotic independence which overcomes the problems of the classical multivariate extreme value theory and preserves flexibility. Essentially, the procedure of estimating p may be summarized as follows. Estimate the distribution of the marginals, standardize the observations to the unit square, estimate the coefficient of tail dependence η, choose an appropriate scaling factor r to estimate the probability of a jointly large claim which has been observed sufficiently often and scale down this probability with the pertaining factor.

In Chapter 3 we establish asymptotic representations for estimators of η. This enables us to prove asymptotic normality of an estimator of p which makes use of the scaling law in the described way. We show that this asymptotic normality holds if either the error when estimating η or when estimating the marginals dominates. We devise a graphical tool to choose an appropriate scaling factor r and an appropriate number of tail-observations used for the estimation of p, respectively. Further, we prove that the suitably standardized process of the difference of the model-relevant limiting function c and the proposed estimator of c converges to a centered Gaussian process. This result also holds if the parameter r to select is a random variable, which is natural and important for applications.

A method to validate the scaling law is developed in Chapter 4. We prove asymptotic normality of random deviations from the scaling law and construct a test that is applied to the standardized observations. In order to obtain a uniform test for the entirety of the (standardized) observations, we prove results on the empirical process and on the Gaussian limiting process of the random deviations from the scaling law.

We demonstrate the usefulness of the developed theory by real-life claims data in Chapter 5. We see that the assumption of independence of the claim sizes may underestimate the real risk, that the empirical probability is useful only if the considered jointly large claim is observed sufficiently often, that the tool for validating the scaling law is able to check the appropriateness of the model and that the latter is the case for the insurance data under investigation.

In Chapter 6 we compare the Extended Ledford and Tawn Model with another model for multivariate tail dependence. We show that this copula approach exploits less relevant information and does not overcome the main problems and drawbacks of many popular copula approaches and the classical multivariate extreme value theory.

We conclude this summary with some remarks on possible model extensions and objectives for future research. As indicated e.g. in Sections 2.2.2 and 3.3, the basic model Condition 2.2.1 implies that the scaling law, and hence the idea of the estimation of p, also holds for more general sets than rectangles, cf. also Figure 2.1. For these more general sets, appropriate generalizations of the results of this work shall be found. Further, as remarked in Section 4.5, since the convergence of the weighted limiting process L works with the weight function $(c(sx, sy_0))^{1/2+\delta}$ and without the stronger uniform condition (4.24), one may assume that an analogous result holds for the empirical process. To find a proof of such a result is another task for future research. Finally, a more thorough investigation of the impact of the data choice when we face some censoring as in Chapter 5 is desirable. We see in the examples in Chapter 5 that if claims are recorded only if their sum exceeds some threshold, the decision whether to use the data when both components exceed this threshold or when at least one component exceeds this threshold influences the estimates of model parameter(s) (statistically) significantly (although the impact on the estimates of p is not significant). It has to be analyzed more thoroughly which data should be used in which situation.

Zusammenfassung

Motivation für diese Arbeit ist insbesondere das Versagen der klassischen multivariaten Extremwerttheorie bei der Schätzung von Wahrscheinlichkeiten p des gleichzeitigen Auftretens von Versicherungsgroßschäden in verschiedenen Risikobeständen, wenn die Schadenhöhen asymptotisch unabhängig sind, sowie die (Willkür und) mangelnde Flexibilität vieler für solche Schätzungen verwendeten Copula-Modelle. Diese Probleme werden mehrfach thematisiert, insbesondere im Vorwort und in den Kapiteln 1.2.3 und 6. Es wird ein Modell von Ledford und Tawn (1996, 1997, 1998) für das gemeinsame Verteilungsende der bivariaten Schadenhöhenverteilung erweitert, das diese Probleme überwindet. Die Grundidee dieses Modells besteht darin, die zu p gehörende Ereignismenge mit Hilfe eines Skalierungsfaktors r so zu vergrößern, dass in der entstehenden Ereignismenge hinreichend viele Beobachtungen für eine stabile Schätzung mit Hilfe empirischer Wahrscheinlichkeiten vorhanden sind. Das Skalierungsgesetz, das im Erweiterten Modell von Ledford und Tawn gilt, besagt, dass mit Hilfe des zentralen Modellparameters, dem Koeffizienten η der Abhängigkeit im gemeinsamen Verteilungsende, die Wahrscheinlichkeit p aus der (empirischen) Wahrscheinlichkeit der vergrößerten Ereignismenge durch Multiplikation mit (einer Schätzung von) $r^{-1/\eta}$ erhalten wird. Schon in Kapitel 2.2.2 wird deutlich, dass im Erweiterten Modell von Ledford und Tawn die Probleme der klassischen multivariaten Extremwerttheorie nicht auftreten und dennoch hohe Flexibilität gewährleistet ist. Das Vorgehen zur Schätzung einer Wahrscheinlichkeit wie p kann im Wesentlichen wie folgt zusammengefasst werden. Schätze die Marginalverteilungen, standardisiere die Beobachtungen auf das Einheitsquadrat, schätze η, wähle einen geeigneten Skalierungsfaktor r um die Wahrscheinlichkeit eines gemeinsamen (Großschaden-) Ereignisses, das hinreichend oft beobachtet werden konnte, zu schätzen und skaliere diese

Wahrscheinlichkeit mit einer Schätzung von $r^{-1/\eta}$.

In Kapitel 3 werden asymptotische Darstellungen für Schätzer von η entwickelt. Mit Hilfe dieser Darstellungen wird dann asymptotische Normalität eines Schätzers von p gezeigt, der auf dem Skalierungsgesetz beruht. Es wird bewiesen, dass diese asymptotische Normalität gilt wenn der statistische Fehler zur Schätzung von η oder der statistische Fehler zur Schätzung der Marginalverteilungen dominiert. Ein grafisches Analysewerkzeug zur Wahl eines geeigneten Skalierungsfaktors r beziehungsweise zur Wahl der Anzahl der zur Schätzung von p verwendeten Daten wird entwickelt. In Kapitel 3 wird außerdem bewiesen, dass der geeignet standardisierte Prozess der Differenz der für das Modell relevanten Grenzfunktion c und eines Schätzers von c gegen einen zentrierten Gauß-Prozess konvergiert. Dieses Resultat wird dann auf den Fall eines zufälligen Skalierungsfaktors r erweitert, was ein natürlicher und für Anwendungen wichtiger Fall ist.

Eine Methode zur Validierung des für das Modell zentralen Skalierungsgesetzes wird in Kapitel 4 entwickelt. Es wird bewiesen, dass die zufälligen Abweichungen vom Skalierungsgesetz asymptotisch normalverteilt sind. Ein auf diesem Resultat basierender statistischer Test wird konstruiert. Um einen gleichmäßigen Test für die Gesamtheit der standardisierten Beobachtungen zu erhalten, werden außerdem Resultate zum empirischen Prozess und zum Gauß'schen Grenzprozess der zufälligen Abweichungen vom Skalierungsgesetz bewiesen.

Die Nützlichkeit der entwickelten Theorie wird in Kapitel 5 an Hand von Schadendatensätzen aus der Versicherungspraxis demonstriert. Es wird gezeigt, dass die Annahme der Unabhängigkeit der Schadenhöhen in verschiedenen Risikobeständen das tatsächliche Risiko unterschätzen kann, dass die empirische Wahrscheinlichkeit meist nur dann nützlich ist, wenn das betrachtete (Großschaden-) Ereignis hinreichend oft beobachtet wurde, dass das vorgeschlagene Verfahren zur Validierung des Skalierungsgesetzes die Eignung des Modells überprüfen kann und dass das Skalierungsgesetz für die betrachteten Schadendaten erfüllt ist.

In Kapitel 6 wird das Erweiterte Modell von Ledford und Tawn mit einem anderen Modell für Abhängigkeiten im multivariaten Verteilungsende verglichen. Es wird gezeigt, dass dieser Copula-Ansatz viele für die Schätzung von p wesentliche Informationen nicht nutzt und damit die Schwierigkeiten und Nachteile der vieler anderer Copula-Modelle und der klassischen mulitvariaten Extremwerttheorie nicht überwinden kann.

Bibliography

Alexander, K.S. (1984): Probability inequalities for empirical processes and a law of the iterated logarithm. *The Annals of Probability* 12 (4), 1041-1067.

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004): *Statistics of Extremes.* Wiley & Sons, Chichester.

Billingsley, P. (1968): *Convergence of Probability Measures.* Wiley, New York.

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

Blum, P., Dias, A. and Embrechts, P. (2002): The ART of dependence modelling: the latest advances in correlation analysis. In Lane, M. (Ed.): *Alternative Risk Strategies.* Risk Books, London, 339-356.

Cebriàn, A.C., Denuit, M. and Lambert, P. (2003): Generalized pareto fit to the Society of Actuaries' large claims database. *North American Actuarial Journal* 7 (3), 18-36.

Charpentier, A. and Juri A. (2006): Limiting dependence structures for tail events, with applications to credit derivatives. *Journal of Applied Probability* 43, 563-586.

Clopper, C.J. and Pearson E.S. (1934): The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 26, 404-413.

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

Coles, S.G., Heffernan, J.E. and Tawn J.A. (1999): Dependence measures for extreme value analyses. *Extremes* 2 (4), 339-365.

Csörgő, M. and Horváth, L. (1993): *Weighted Approximations in Probability and Statistics.* Wiley, New York.

Draisma, G., Drees, H., Ferreira, A. and de Haan, L. (2003): Bivariate tail estimation: dependence in asymptotic independence, technical report. EURANDOM, Eindhoven.

Draisma, G., Drees, H., Ferreira, A. and de Haan, L. (2004): Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli* 10 (2), 251-280.

Drees, H. (1998a): On smooth statistical tail functionals. *Scandinavian Journal of Statistics* 25, 187-210.

Drees, H. (1998b): A general class of estimators of the extreme value index. *Journal of Statistical Planning and Inference* 66, 95-112.

Drees, H., de Haan, L. and Resnick, S.I. (2000): How to make a Hill plot. *The Annals of Statistics* 28 (1), 254-274.

Drees, H. and Huang, X. (1998): Best attainable rates of convergence for estimators of the stable tail dependence function. *Journal of Multivariate Analysis* 64 (1), 25-47.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997): *Modelling Extremal Events for Insurance and Finance.* Springer, Berlin.

Embrechts, P., McNeil, A.J. and Straumann, D. (2002): Correlation and dependence in risk management: properties and pitfalls. In Dempster, M.A.H. (Ed.): *Risk Management: Value at Risk and Beyond.* Cambridge University Press, Cambridge.

Grazier, K.L. and G'Sell Associates (1997): *Group Medical Insurance Large Claims Data Base Collection and Analysis.* Society of Actuaries, Monograph M-HB97-1, Schaumburg, Illinois.

de Haan, L. and Ferreira, A. (2006): *Extreme Value Theory, An Introduction.* Springer, New York.

de Haan, L. and Resnick S. (1993): Estimating the limit distribution of multivariate extremes. *Communications in Statistics – Stochastic Models* 9 (2), 275-309.

de Haan, L. and Stadtmüller, U. (1996): Generalized regular variation of second order. *Journal of Australian Mathematical Society* 61, 381-395.

Heffernan, J.E. and Resnick S.I. (2005): Hidden regular variation and the rank transform. *Advances in Applied Probability* 37 (2), 393-414.

Juri, A. and Wüthrich, M.V. (2002): Copula convergence theorems for tail events. *Insurance: Mathematics and Economics* 30 (3), 405-420.

Juri, A. and Wüthrich, M.V. (2003): Tail dependence from a distributional point of view. *Extremes* 6 (3), 213-246.

Ledford, A.W. (1996): *Dependence within Extreme Values: Theory and Applications.* PhD Thesis, Lancaster University, Lancaster.

Ledford, A.W. and Tawn, J.A. (1996): Statistics for near independence in multivariate extreme values. *Biometrika* 83 (1), 169-187.

Ledford, A.W. and Tawn, J.A. (1997): Modelling dependence within joint tail regoins. *Journal of the Royal Statistical Society* 59 (2), 475-499.

Ledford, A.W. and Tawn, J.A. (1998): Concomitant tail behaviour for extremes. *Advances in Applied Probability* 30, 197-215.

Maulik, K. and Resnick S.I. (2004): Characterizations and examples of hidden regular variation. *Extremes* 7 (1), 31-67.

McNeil, A.J. (1997): Estimating the tails of loss severity distributions using extreme value theory. *Astin Bulletin* 27 (1), 117-137.

McNeil, A.J., Frey, R. and Embrechts, P. (2005): *Quantitative Risk Management.* Princeton Series in Finance.

Meerschaert, M.M. and Scheffler, H.-P. (2001): *Limit Distributions for Sums of Independent Random Vectors.* Wiley, New York.

Mikosch, T. (2005): How to model multivariate extremes if one must. *Statistica Neerlandica* 59, 324-338.

Neuhaus, G. (1971): On weak convergence of stochastic processes with multidimensional time parameter. *The Annals of Mathematical Statistics* 42 (4), 1285-1295.

Peng, L. (1999): Estimation of the coefficient of tail dependence in bivariate extremes. *Statistics & Probability Letters* 43, 399-409.

Pollard, D. (1984): *Convergence of Stochastic Processes.* Springer, New York.

Resnick, S.I. (1987): *Extreme Values, Regular Variation and Point Processes.* Springer, New York.

Resnick, S.I. (2001): Hidden regular variation, second order regular variation and asymptotic independence. *Extremes* 5, 303-336.

Rytgaard, M. (1996): Simulation experiments on the mean residual life function $m(x)$. In *Proceedings of the XXVII Astin Colloquium* (1), 59-81. Copenhagen, Denmark.

Santner, T.J. and Duffy, D.E. (1989): *The Statistical Analysis of Discrete Data.* Springer, New York.

Shorack, G.R. and Wellner, J.A. (1986): *Empirical Processes with Applications to Statistics.* Wiley, New York.

Tawn, J.A. (1988): Bivariate extreme value theory: models and estimation. *Biometrika* 75 (3), 397-415.

Vervaat, W. (1972): Functional central limit theorems for processes with positive drift and their inverses. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 23, 245-253.

Walsh, J.B. (1986): Martingales with a multidimensional parameter and stochastic integrals in the plane. In *Lectures in Probability and Statistics (Santiago de Chile, 1986)* 329-491. Springer, Berlin.

Curriculum Vitae

Peter Müller

geboren am 19. Mai 1978 in Hoyerswerda

http://www.PeterMueller195.de

Wissenschaftliche Ausbildung

01/2004 – 07/2008	Tätigkeit als Wissenschaftlicher Mitarbeiter und Promotionsstudium am Schwerpunkt Mathematische Statistik und Stochastische Prozesse der Universität Hamburg, Gesamtnote *magna cum laude*
09/2002 – 12/2002	Studienaufenthalt am Department for Actuarial Science der Heriot-Watt-University Edinburgh, U.K.
10/1999 – 09/2002	Studentische Mitarbeit u. a. Seminarleitung und bei Projekten des Lehrstuhls für Finanzwirtschaft und Finanzdienstleistungen der TU Dresden
10/1997 – 12/2003	Studium der Wirtschaftsmathematik an der TU Dresden mit den Schwerpunkten Finanz- und Versicherungsmathematik, sowie Finanzwirtschaft, Diplomarbeit *Modeling Credit Risk with Markov Chains*, Gesamtnote *sehr gut*

Schulbildung, Wehrdienst

Bis 08/1997	Abitur am Konrad-Zuse-Gymnasium Hoyerswerda; anschließend Wehrdienst

Weitere Berufliche Tätigkeiten

Seit 04/2008	Banken-, Versicherungs-, und Wertpapieraufsicht in der Querschnittsabteilung Risikomodellierung der Bundesanstalt für Finanzdienstleistungsaufsicht (BaFin)
11/2005 – 02/2006	Aktuarielles Trainee-Programm bei einem internationalen Risk-Consulting-Unternehmen (IRMG Luxemburg und IRMG London, Aon Corp.)
10/2004 – 01/2005	Aktuarielle Projektarbeit für IRMG Switzerland
04/2000 – 09/2000	Praktikum bei der Investmentbank Concord Effekten AG, Frankfurt am Main