UNIVERSITÄT HAMBURG

Modeling Dependencies in Large Claims

H. Drees, P. Müller

Preprint–No. 2007-02 January 2007

FACHBEREICH MATHEMATIK

SCHWERPUNKT MATHEMATISCHE STATISTIK
UND STOCHASTISCHE PROZESSE

BUNDESSTR. 55, D – 20146 HAMBURG

Modeling Dependencies in Large Claims

Holger Drees [a], Peter Müller [b, *]

[a] *holger.drees@math.uni-hamburg.de*

[b] *peter.mueller@math.uni-hamburg.de*

University of Hamburg, Department of Mathematics, Center of Mathematical Statistics and Stochastic Processes, Bundesstr. 55, 20146 Hamburg, Germany

Abstract

We consider an extended version of a model for the joint tail distribution of a bivariate random vector proposed by Ledford and Tawn (1997), which essentially assumes an asymptotic power scaling law for the probability that both components of the vector are jointly large. First it is illustrated that classical multivariate extreme value theory does not provide suitable estimators of the probability of jointly extreme events in the case of asymptotic independence. Then we introduce the Extended Ledford & Tawn Model and discuss how to fit the model. Since the estimators of the model parameters rely on the aforementioned scaling law, it is crucial to check this model assumption. To this end, we devise a graphical tool that analyzes the differences between certain empirical probabilities and model based estimates of the same probabilities. The asymptotic normality of these differences allow the construction of statistical tests for the model assumption. The results are applied to claims of a Danish fire insurance and to medical claims from US health insurances.

Key words: asymptotic normality; bivariate tail estimation; dependent catastrophic risks; extreme value theory; model validation

1 Introduction

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 univariate extreme value theory (EVT) offers models and estimators for the upper tail of the claim size distribution. To this end, denote the distribution function (d.f.) of the claim sizes by F_1. Then it can be shown that the only

* corresponding author

possible limits of the conditional d.f. of the suitably standardized claims size X given that it exceeds an increasing threshold t, i.e.

$$P\left(\frac{X-t}{a(t)} \le x \,\middle|\, X > t\right) = 1 - \frac{1-F_1(t+a(t)x)}{1-F_1(t)} \longrightarrow H_\gamma(x) \qquad (1)$$

as $P\{X > t\} \to 0$, are the generalized Pareto distributions (GPD) with d.f.

$$H_\gamma(x) = 1 - (1+\gamma x)^{-1/\gamma} \quad \text{if } 1+\gamma x > 0. \qquad (2)$$

Here, $\gamma \in \mathbb{R}$ denotes the so-called extreme value index and $a(t) > 0$ is a normalizing constant depending on the threshold t. Therefore, these GPD with additional location and scale parameters define a natural model for the upper tail of the claim size distribution. See the monograph by Embrechts et al. (1997) for an illuminative introduction to EVT and tail modeling (Sections 3.4 and 6.5 for GPD models) and their applications to risk management.

Most critical for an insurance company is the case of a heavy-tailed claim size d.f. F_1, i.e. when γ is positive. A necessary and sufficient condition for this case is the regular variation of the survival function $1 - F_1$ with index $-1/\gamma$, i.e.

$$\lim_{t\to\infty} \frac{1-F_1(tx)}{1-F_1(t)} = x^{-1/\gamma}, \qquad x > 0, \qquad (3)$$

which is the same as (1) with $a(t) = \gamma t$ and x replaced with $1 + \gamma x$. Then necessarily all moments of the claim size distribution up to the order $1/\gamma$ (exclusively) are finite, while the moments of order larger than $1/\gamma$ do not exist. 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 Embrechts et al. (1997), Table 3.4.2, for explicit examples.)

However, 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 customer 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

$$p := P\{X > u_1, Y > u_2\} \qquad (4)$$

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 Section 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 (4). 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.

To analyze the full dependence structure separately from the marginal distributions, one often considers the so-called *copula*, i.e., the bivariate d.f. of the claim sizes after standardization of the marginals to uniform random variables. More precisely, 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, with generalized inverse functions (quantile functions) $F_i^{\leftarrow}$, $i = 1, 2$. In the rest of this section, for simplicity, suppose that F_1 and F_2 are continuous. Then the copula C pertaining to F is given by

$$C(u, v) := F(F_1^{\leftarrow}(u), F_2^{\leftarrow}(v)) = P\{\bar{U} \le u, \bar{V} \le v\}, \qquad u, v \in [0, 1], \tag{5}$$

where $\bar{U} := F_1(X)$ and $\bar{V} := F_2(Y)$ are uniform random variables.

Recently, a large variety of parametric families of copulas (like t-copulas or Clayton copulas) have been proposed as models for the dependence structure between different financial risks (see McNeil et al. (2005), Chapter 5). However, as mentioned above, the bulk of 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).)

Quite flexible dependence models for extremes are offered by the classical multivariate EVT, as presented in Chapter 5 of Resnick (1987) or Chapter 8 of Beirlant et al. (2004), where a certain regularity condition on the probability that *at least one* standardized claim size is large is assumed. To this end, let $U := 1 - \bar{U} = 1 - F_1(X)$ and $V := 1 - \bar{V} = 1 - F_2(Y)$, and define the *tail dependence function* D by

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

Then it is assumed that $t^{-1}D(tx, ty)$ converges to a non-degenerate limiting function:

$$\lim_{t \downarrow 0} t^{-1} D(tx, ty) = \ell(x, y), \qquad x, y \geq 0 \tag{6}$$

(see e.g. Beirlant et al. (2004), Section 8.3.2). Since then the so-called *stable tail dependence function* ℓ is necessarily homogeneous of order 1, i.e. $\ell(tx, ty) = t\ell(x, y)$, according to (6) one may approximate $D(u, v)$ by $\ell(u, v)$ for small u and v. Hence one may construct an estimator for the probability (4) from estimators of the marginal d.f.s and an estimator of ℓ using the approximation

$$\begin{aligned} p &= P\{X > u_1\} + P\{Y > u_2\} - P\{X > u_1 \textbf{ or } Y > u_2\} \\ &\approx 1 - F_1(u_1) + 1 - F_2(u_2) - \ell(1 - F_1(u_1), 1 - F_2(u_2)). \end{aligned} \tag{7}$$

Unfortunately, this approach fails if X and Y are *asymptotically independent* (or *tail independent*), i.e. if 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 in the sense that

$$P(U < t \mid V < t) = t^{-1} P\{U < t \textbf{ and } V < t\} \xrightarrow{t \downarrow 0} 0. \tag{8}$$

For, in this case,

$$\begin{aligned} \ell(x, y) &= \lim_{t \downarrow 0} t^{-1} \left(P\{U < tx\} + P\{V < ty\} - P\{U < tx \textbf{ and } V < ty\} \right) \\ &= x + y, \end{aligned}$$

and thus (7) reads as $p \approx 0$, which is too crude an approximation to be useful for estimating p. To overcome this problem, Ledford and Tawn (1997) proposed a model to specify the speed of convergence in (8).

The paper is organized as follows. Section 2 introduces a model which implements the aforementioned idea and slightly extends the approach by Ledford and Tawn (1997). The scaling law which is central to this model is established. In Section 3 we briefly explain how to fit the model. Moreover, an estimator of the probability that both components X and Y are large is presented. Section 4 contains the main result about the asymptotic normality of certain statistics, which measure the deviations from this scaling law. Based on this result, we suggest a method to validate the model. The applicability of this tool is demonstrated by the examples of claims of a Danish fire insurance, and of medical claims from US health insurances. All proofs are deferred to the final Section 6.

2 The Extended Ledford & Tawn Model

Throughout the paper, for simplicity, we assume that the marginal d.f.s F_i are continuous on $[\xi_i, \infty)$ for some $\xi_i < F_i^{\leftarrow}(1)$, $i = 1, 2$. 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.

We consider an extension of the model by Ledford and Tawn (1997), introduced by Draisma et al. (2004). Instead of the original condition supposed in the latter paper, however, the following weaker assumption, mentioned in Remark 2.1 of Draisma et al. (2004), is sufficient for our purposes.

2.1 Condition.

$$\frac{P\{U < tx, V < ty\}}{P\{U < t, V < t\}} - c(x, y) = O(q_1(t)) \tag{9}$$

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_1 a positive function that tends to 0 as $t \downarrow 0$ and that is regularly varying at 0 with some index $\tau \geq 0$, i.e.

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

△

For $x, y \in [0, 1]$, Condition (9) may be interpreted as the convergence of the conditional d.f. $P(U < tx, V < ty \mid U < t, V < t)$ to the limiting d.f. c. Hence, it is a natural analog to the univariate extreme value condition (1). Essentially, relation (9) is a bivariate regular variation condition (see Resnick (1987), (5.32)) for the so-called *survival copula* $Q(x, y) := P\{U < x, V < y\} = x + y - D(x, y)$, where D denotes the tail dependence function.

A first consequence of (9) is $c(1, 1) = 1$. Moreover, choosing $x = y$, one obtains that the function $q(t) := P\{U < t, V < t\}$ is regularly varying at 0, i.e.

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

for some η, which must belong to $(0, 1]$ because $q(t) \leq P\{U < t\} = t$ for sufficiently small $t > 0$. Note that by the regular variation of q

$$
\begin{aligned}
c(tx,ty) &= \lim_{s\downarrow 0} \frac{P\{U<stx, V<sty\}}{P\{U<st, V<st\}} \cdot \frac{P\{U<st, V<st\}}{P\{U<s, V<s\}} \\
&= c(x,y)\cdot \lim_{s\downarrow 0} \frac{q(st)}{q(s)} \\
&= t^{1/\eta} c(x,y) \qquad (10)
\end{aligned}
$$

for all $t, x, y \geq 0$, i.e. c is homogeneous of order $1/\eta$.

Note that $P(U < t \mid V < t) = t^{-1}q(t)$ is regularly varying with index $1/\eta - 1$. Hence, η measures the speed of convergence in (8) and is therefore called *coefficient of tail dependence*. If $\eta < 1$, then $\lim_{t\downarrow 0} t^{-1}q(t) = 0$, i.e. U and V (and thus X and Y) are asymptotically independent. Conversely, asymptotic dependence holds if $\eta = 1$ *and* $t^{-1}q(t)$ converges to some $l > 0$ as $t \downarrow 0$. If U and V are exactly independent, then (9) holds with $\eta = 1/2$, $c(x,y) = xy$ and $q_1 \equiv 0$. Loosely speaking, 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.

Next we will reinterpret the central Condition 2.1 as a scaling law. Note that (9) and the homogeneity of c imply

$$
\frac{P\{U<stx, V<sty\}}{P\{U<t, V<t\}} = c(sx,sy) + O(q_1(t)) = s^{1/\eta}c(x,y) + O(q_1(t))
$$

for $x, y \geq 0$. Thus, again by (9)

$$
\lim_{t\downarrow 0} \frac{P\{U<stx, V<sty\}}{P\{U<tx, V<ty\}} = s^{1/\eta}. \qquad (11)
$$

More generally, for a bounded measurable set $B \subset [0,1]^2$ one has

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

provided that the measure with d.f. c assigns positive mass to B and has no mass on the boundary of B.

Thus, in this extension of the Ledford & Tawn Model, the approximative 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 decrease of the pertaining probability by factor $s^{1/\eta}$. Figure 1 illustrates the scaling law at work.

This scaling law will be the basis for the test of the model assumptions proposed in Section 4.

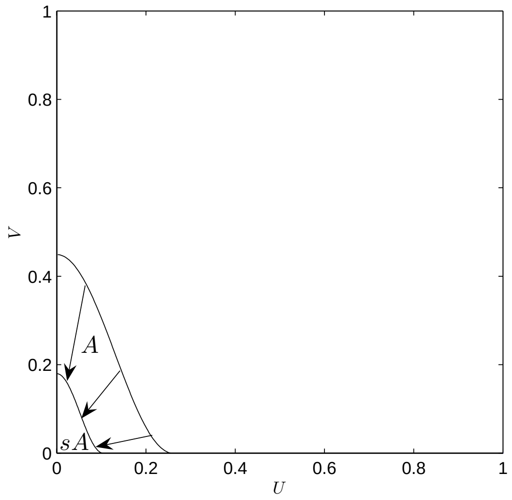

Fig. 1. The scaling law assumed by the model. The ratio of the probabilities of the events sA and A is supposed to be approximately equal to $s^{1/\eta}$.

3 Model Fitting

In this Section we briefly explain how to fit the Extended Ledford & Tawn Model to data. Section 3.1 is essentially a review of some results about (asymptotic) properties of an estimator for the coefficient of tail dependence η. Section 3.2 introduces a simple new estimator for $c(x, y)$ and proves its asymptotic normality. In Section 3.3 we discuss estimators for the probability of a jointly large event $p = P\{X > u_1, Y > u_2\}$. As Sections 3.1 and 3.3 mainly summarize results of Sections 2 and 3 of Draisma et al. (2004), refer to this article for the details.

3.1 Estimation of the Coefficient of Tail Dependence

The asymptotic scaling law (11) reveals that the estimation of η is a crucial step if we want to estimate p, see also Section 3.3. We will relate this problem to the estimation of the extreme value index in a generalized Pareto model.

Let $\{(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\}$ be an i.i.d. sample, and denote by F_1 and F_2 the marginal d.f.s of X_i and Y_i, $i = 1, \ldots, n$, respectively. As Draisma et al. (2004), we consider an estimator for η based on the random variables

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

where $U_i := 1 - F_1(X_i)$ and $V_i := 1 - F_2(Y_i)$, $i = 1, \ldots, n$, respectively. The d.f. F_T of T_i satisfies $1 - F_T(t) = q(1/t)$. Hence, the survival function $1 - F_T$

is regularly varying with index $-1/\eta$ and we are in the situation of (3). In this setting, a popular estimator for the extreme value index η is the Hill estimator $m^{-1}\sum_{j=1}^{m}\log(T_{n-j+1:n}/T_{n-m:n})$, where $T_{i:n}$, $i=1,\ldots,n$, denote the order statistics pertaining to $T_i, i=1,\ldots,n$. This estimator possesses good asymptotic properties, see Section 6.4 of Embrechts et al. (1997).

Since the marginal d.f.s F_1 and F_2 are unknown and hence the random variables T_i cannot be observed, we replace U_i and V_i by empirical counterparts. To avoid division by 0, let $\hat{U}_i := 1 - R_i^X/(n+1)$ and $\hat{V}_i := 1 - R_i^Y/(n+1)$, where R_i^X and R_i^Y denote the ranks of X_i and Y_i among $(X_1, X_2, \ldots, X_n)$ and $(Y_1, Y_2, \ldots, Y_n)$, respectively. Further, let $T_i^{(n)} := \min(1/\hat{U}_i, 1/\hat{V}_i)$, $i=1,\ldots,n$, and denote by $T_{i:n}^{(n)}$, $i=1,\ldots,n$, the pertaining order statistics. In analogy to the above estimator for η, we define the Hill estimator

$$\hat{\eta}_n := \frac{1}{m}\sum_{j=1}^{m}\log\frac{T_{n-j+1:n}^{(n)}}{T_{n-m:n}^{(n)}}, \tag{12}$$

which also possesses good asymptotic properties, reviewed below. Here, $m = m_n$ is an intermediate sequence (i.e. $m\to\infty$ and $m/n\to 0$ as $n\to\infty$). For an appropriate choice of m in applications see the comments and references given in Section 5.

We now give a short summary of the asymptotic properties of $\hat{\eta}_n$ including confidence intervals for η and a test of the hypothesis $\eta = 1$. The results are taken from Theorems 2.1 and 2.2 and Section 4 of Draisma et al. (2004). Suppose that $l := \lim_{t\downarrow 0} q(t)/t$ and the partial derivatives $c_x := \partial c/\partial x$ and $c_y := \partial c/\partial y$ of c exist. Assuming further that m tends to ∞ not too fast such that $\sqrt{m}q_1(q^{\leftarrow}(m/n)) \to 0$, one can prove that $\sqrt{m}(\hat{\eta}_n - \eta)$ converges to a normal distribution $\mathcal{N}(0, \sigma^2_{\hat{\eta}_n})$ as $n\to\infty$, where

$$\sigma^2_{\hat{\eta}_n} = \eta^2\,(1-l)\,(1-2lc_x(1,1)c_y(1,1))\,. \tag{13}$$

Note that l is precisely the limit of left hand side of (8) and that $\sigma^2_{\hat{\eta}_n} = \eta^2$ if $l=0$, i.e. iff U and V are asymptotically independent. Note further that $\sigma^2_{\hat{\eta}_n} \le \eta^2 \le 1$ is always true, since $l\in[0,1]$. Let $\hat{l} := mT^{(n)}_{n-m:n}/n$, denote by $T_{i:n}^{(n,u)}$, $i=1,\ldots,n$, the order statistics pertaining to $T_i^{(n,u)} := \min((1+u)/\hat{U}_i, 1/\hat{V}_i)$, $i=1,\ldots,n$, and define $\hat{c}_x(1,1) := \hat{k}^{5/4}(T_{n-m:n}^{(n,\hat{k}^{-1/4})} - T_{n-m:n}^{(n)})/n$ with $\hat{k} := m/\hat{l}$. 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. Then, the estimator

$$\hat{\sigma}^2_{\hat{\eta}_n} := \hat{\eta}_n^2\left(1-\hat{l}\right)\left(1-2\hat{l}\hat{c}_x(1,1)\hat{c}_y(1,1)\right) \tag{14}$$

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

$$\left[\hat{\eta}_n - m^{-1/2}\hat{\sigma}_{\hat{\eta}_n}\Phi^{\leftarrow}(1-\alpha/2), \hat{\eta}_n + m^{-1/2}\hat{\sigma}_{\hat{\eta}_n}\Phi^{\leftarrow}(1-\alpha/2)\right], \tag{15}$$

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

$$\frac{\sqrt{m}\,(1-\hat{\eta}_n)}{\hat{\sigma}_{\hat{\eta}_n}} > \Phi^{\leftarrow}(1-\alpha). \tag{16}$$

3.2 Estimation of c

From (9), for sufficiently small r,

$$c(x,y) \approx \frac{P\{U<rx, V<ry\}}{P\{U<r, V<r\}}.$$

Similar as above, we will estimate this probability by an empirical counterpart based on the rank standardized approximations $\hat{U}_i$ and $\hat{V}_i$ to U_i and V_i, respectively.

3.1 Theorem. Assume that (9) holds with asymptotically independent random variables U and V and with a function c that has first order partial derivatives. 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

$$\sqrt{\tilde{m}} \left(\frac{\sum_{i=1}^{n} \mathbb{1}_{\left\{\hat{U}_i < rx, \hat{V}_i < ry\right\}}}{\sum_{i=1}^{n} \mathbb{1}_{\left\{\hat{U}_i < r, \hat{V}_i < r\right\}}} - c(x,y) \right) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2_{x,y}\right) \tag{17}$$

for all $x, y \in (0,1]$, where $\xrightarrow{D}$ denotes convergence in distribution as $n \to \infty$ and

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

$\triangle$

3.2 Remarks.

(i) In Section 6, we actually prove the stronger result that the process $\sqrt{\tilde{m}}\Big(\sum_{i=1}^{n} \mathbb{1}_{\left\{\hat{U}_i<rx,\hat{V}_i<ry\right\}}/\sum_{i=1}^{n} \mathbb{1}_{\left\{\hat{U}_i<r,\hat{V}_i<r\right\}} - c(x,y)\Big)_{(x,y)\in[0,\infty)^2}$ converges weakly in the bivariate Skorohod space $D([0,\infty)^2)$ to a centered Gaussian process.

(ii) For arbitrary $x, y > 0$, (17) holds with

$$\sigma^2_{x,y} = c(x,y) - 2c(x,y)c(x \wedge 1, y \wedge 1) + c^2(x,y).$$

(iii) A similar result (with a more complicated asymptotic variance) holds in the case of asymptotic dependence; see the accompanying technical note Drees and Müller (2006).

△

To estimate c, we have to choose an appropriate r. We will now motivate such a choice by a more detailed examination of the Hill estimator $\hat{\eta}_n$, introduced in Section 3.1. In particular, this choice will prove useful when validating the scaling law (11) with the method proposed in Section 4. Recall that the Hill estimator $\hat{\eta}_n$ is based on the $m+1$ largest order statistics $T^{(n)}_{n-j+1:n}$, $j = 1, \ldots, m+1$, of the pseudo-observations $T^{(n)}_i = 1/\max(\hat{U}_i, \hat{V}_i)$. This means that we use those $(\hat{U}_i, \hat{V}_i)$ for the estimation of η that lie within the square $(0, 1/T^{(n)}_{n-m:n}]^2$. Hence, from this point of view, it seems natural to work with the random value $r = 1/T^{(n)}_{n-m:n}$ and to define

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

In the accompanying technical note Drees and Müller (2006) it is shown that $\hat{c}_n$ satisfies $\sqrt{\tilde{m}}(\hat{c}_n(x,y) - c(x,y)) \xrightarrow{D} \mathcal{N}(0, \sigma^2_{x,y})$, which also holds when $\tilde{m}$ is replaced with m; see the following comments on the relation of the particular choice $\tilde{m} = nq(1/T^{(n)}_{n-m:n})$ and m. In particular, $\hat{c}_n(x,y)$ is consistent for $c(x,y)$. Remarks 3.2 apply analogously.

The similarity of the notations of the number of order statistics m used for the estimation of η (Section 3.1) and $\tilde{m}$ is not by accident. By the aforementioned random choice of r,

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

also becomes a random variable. If the claim sizes have no ties [1], we have for

[1] Under the assumption that F_1 and F_2 are tail continuous, the probability that there are no ties converges to 1; see also the comments made in Section 6.

the deterministic value m the simple equality

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

and thus

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

In the accompanying technical note Drees and Müller (2006) it is shown that $\tilde{m}/m \to 1$ in probability as $n \to \infty$ and therefore

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

The choice $r = 1/T_{n-m:n}^{(n)}$ will be important for the construction of the graphical tool for model validation in Section 4 and will be used in the applications of Section 5.

3.3 Probability of a Jointly Large Event

To motivate an estimator for the probability of a jointly large event like $p = P\{X > u_1, Y > u_2\}$, recall the scaling law (11) to see that

$$\begin{aligned} p &= P\Big\{U < 1 - F_1(u_1),\ V < 1 - F_2(u_2)\Big\} \\ &\approx r^{-1/\eta} \cdot P\Big\{U < r(1 - F_1(u_1)),\ V < r(1 - F_2(u_2))\Big\} \end{aligned} \tag{18}$$

for all $r \in (0,1]$. This representation emphasizes that the Extended Ledford & Tawn Model 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 and the tail dependence structure (by η). In contrast to the most copula approaches (see Section 1), however, the Extended Ledford & Tawn Model focusses on the distribution tail(s) and therefore allows a separate modeling of large claims with respect to the marginal distributions *and* the dependence structure. Also in Section 1, we briefly discussed the generalized Pareto model in the univariate EVT. Denote by $\hat{F}_{1,n}$ and $\hat{F}_{2,n}$ estimators of the tails of the marginal d.f.s F_1 and F_2 that are based on the GPD approximation (1). Then, continuing (18),

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

motivates an estimator for p. In a forthcoming paper, we will establish consistency and asymptotic normality of $\hat{p}_n$ under additional weak assumptions on the marginal d.f.s F_i and the thresholds u_i, $i = 1, 2$. Using the asymptotic normality of $\hat{p}_n$, one can easily construct confidence intervals for p. (Because one has to distinguish several cases depending on the rates of convergence of the marginal estimators $\hat{F}_{i,n}(u_i)$ and of $\hat{\eta}_n$, the exact statement of the asymptotic normality is somewhat lengthy. Its formulation clearly goes beyond the scope of the present paper that is focussed on the fitting and validation of the model, not on the estimation of tail probabilities.)

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 events C that satisfy[2] $(x, y) \in C \Rightarrow [x, \infty] \times [y, \infty] \subset C$. The construction of such an estimator for $P\{(X, Y) \in C\}$ can be found in Section 3 of Draisma et al. (2004), where the following estimator $\tilde{p}_n$ is generalized to an estimator for $P\{(X, Y) \in C\}$.

Instead of 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

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

For the pertaining more general estimator of $P\{(X, Y) \in C\}$ with C as defined above, consistency is proved in Section 3 of Draisma et al. (2004).

4 Model Validation

In this section we develop a method to validate the scaling law (11) for given data. We prove that under some regularity conditions the random deviations from this scaling law are asymptotically normal, construct pointwise confidence intervals and suggest a method to validate the scaling law by means of a simple three-dimensional plot.

Let us recall the scaling law (11), i.e. $P\{U < su, V < sv\}/P\{U < u, V < v\} \approx s^{1/\eta}$ for all $s \in (0, 1]$ and sufficiently small $u, v > 0$. Taking logarithms leads

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

to

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

Now suppose that we choose s such that $(\hat{U}_i, \hat{V}_i)$ 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

$$\left(\log s, \log \frac{\sum_{i=1}^n \mathbb{1}\{\hat{U}_i < su, \hat{V}_i < sv\}}{\sum_{i=1}^n \mathbb{1}\{\hat{U}_i < u, \hat{V}_i < v\}} \right) \tag{19}$$

must approximately lie on the line through the origin with slope $1/\eta$ if the scaling law (11) 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 of the Hill estimator (12). 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 the Hill estimator $\hat{\eta}_n$ is based on the $m+1$ largest order statistics $T_j^{(n)} = 1/\max(\hat{U}_j, \hat{V}_j) \geq T_{n-m:n}^{(n)}$, and thus on those pseudo-observations $(\hat{U}_j, \hat{V}_j)$ which fall into the square $(0, 1/T_{n-m:n}^{(n)}]^2$. It is a reasonable estimator for η if and only if the points (19) with $x = y \leq 1/T_{n-m:n}^{(n)}$ 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 point (19) from the line, the size of the estimated deviation must be examined in the case that the model assumptions are satisfied, i.e. the scaling law holds. The following theorem establishes the asymptotic behavior of this estimated deviation.

4.1 Theorem. Assume that (9) holds with asymptotically independent random variables U and V and with a function c that has first order partial derivatives c_x and c_y. Suppose (as in Theorem 3.1) that m is an intermediate sequence such that $\sqrt{m}\, q_1(q^{\leftarrow}(m/n)) \to 0$ as $n \to \infty$. Then

$$\sqrt{m}\left(\log\frac{\sum_{i=1}^{n}\mathbb{1}\{\hat U_i < sx/T^{(n)}_{n-m:n}, \hat V_i < sy/T^{(n)}_{n-m:n}\}}{\sum_{i=1}^{n}\mathbb{1}\{\hat U_i < x/T^{(n)}_{n-m:n}, \hat V_i < y/T^{(n)}_{n-m:n}\}} - \frac{1}{\hat\eta_n}\log s\right) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2_{x,y,s}\right) \tag{20}$$

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

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

△

4.2 Remarks.

(i) For arbitrary $x, y > 0$, (20) holds with

$$\begin{aligned}\sigma^2_{x,y,s} = & \frac{s^{-1/\eta}-1}{c(x,y)} + \frac{\log^2 s}{\eta^2} \\ & + \frac{2\log s}{c(x,y)}\Bigg(\int_1^{s^{-1/\eta}} t^{-1}c(x\wedge t^\eta, y\wedge t^\eta)dt \\ & \qquad - s^{-1/\eta}c(sx\wedge 1, sy\wedge 1) + c(x\wedge 1, y\wedge 1)\Bigg).\end{aligned}$$

(ii) A similar result (with a more complicated asymptotic variance) holds in the case of asymptotic dependence; see the accompanying technical note Drees and Müller (2006).

(iii) Likewise one can establish analogous results if another estimator for η based on a certain fraction of largest order statistics of $T_i^{(n)}$ is used instead of the Hill estimator.

△

In order to apply this result 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 (19) only for points (su, sv) equal to a (pseudo-)observation. More precisely, we consider those points $(s_ju_j, s_jv_j) = (\hat U_j, \hat V_j)$ which belong to the open square $(0, 1/T^{(n)}_{n-m:n})^2$.

Recall that (if no ties occur[3]) these pseudo-observations correspond to the order statistics $T^{(n)}_{n-i+1:n}$, $i = 1, \ldots, m$, 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 $(\hat{U}_j, \hat{V}_j) \in (0, 1/T^{(n)}_{n-m:n})^2$ for $j = 1, \ldots, m$.

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 $(0, 1/T^{(n)}_{n-m:n}]^2$, 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 2 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

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

(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 (19) for all points (u_j, v_j) and contraction factors s_j, $j = 1, \ldots, m$, we lose the information which point of the plot corresponds to which pseudo-observation. To avoid that, we instead use the three-dimensional plot

$$\left(z_j, \log s_j, \log \frac{\sum_{i=1}^n \mathbb{1}_{\{\hat{U}_i < s_j u_j, \hat{V}_i < s_j v_j\}}}{\sum_{i=1}^n \mathbb{1}_{\{\hat{U}_i < u_j, \hat{V}_i < v_j\}}} \right)_{j=1,\ldots,m}, \tag{21}$$

where the additional argument z_j determines the reference point (u_j, v_j). The right hand plot of Figure 2 illustrates our construction. The points of the plot (21) 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.

Let $\Delta_j := \log(\sum_{i=1}^n \mathbb{1}_{\{\hat{U}_i < s_j u_j, \hat{V}_i < s_j v_j\}} / \sum_{i=1}^n \mathbb{1}_{\{\hat{U}_i < u_j, \hat{V}_i < v_j\}}) - 1/\hat{\eta}_n \log s_j$ denote the estimated difference between the third coordinate of the j^{th} point of (21) and the third coordinate of the corresponding point (with the same first and second coordinates) on the reference plane. Then, according to Theorem 4.1, $\sqrt{m}\Delta_j/\hat{\sigma}_j$ with $\hat{\sigma}_j^2 := (s_j^{-1/\hat{\eta}_n} - 1)/\hat{c}(u_j, v_j) - (\log^2 s_j)/\hat{\eta}_n^2$ is approximately $\mathcal{N}(0, 1)$-distributed. Hence, we reject the scaling law on the approximate con-

[3] 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 (19) is plotted for).

fidence level $1-\alpha$, if

$$\left|\frac{\sqrt{m}\Delta_j}{\hat{\sigma}_j}\right| > \Phi^{\leftarrow}(1-\alpha/2) \tag{22}$$

or, equivalently, if the last coordinate of the j^{th} point of the plot (21) does not belong to the confidence interval $[\log s_j/\hat{\eta}_n - m^{-1/2}\hat{\sigma}_j\Phi^{\leftarrow}(1-\alpha/2), \log s_j/\hat{\eta}_n + m^{-1/2}\hat{\sigma}_j\,\Phi^{\leftarrow}(1-\alpha/2)]$. 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 $(0, 1/T^{(n)}_{n-m:n}]^2$.

One can also incorporate the information provided by the statistical test into the plot (21) by indicating for each point either (i) whether its last coordinate belongs to the corresponding confidence interval or (ii) the $(1-p)$-value

$$2\left|\Phi\left(\frac{\sqrt{m}\Delta_j}{\hat{\sigma}_j}\right) - 0.5\right| \tag{23}$$

of the pertaining test. Since only a very restricted variety of colors is available here, the plots given in the following applications merely implement (i). [4]

5 Applications

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 are widely used 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 at least 1 million Danish Kroner (DKK). For the period 01/1980 - 12/1990 and 01/1980 - 12/1993, respectively, these data were used 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. As remarked by the authors, this is appropriate e.g. for stress testing rather than for modeling multivariate extremes as realistic as possible.

[4] Files where also plots with discretized $(1-p)$-values are implemented are available for download at *http://www.math.uni-hamburg.de/home/drees/extrdep/mdilc.html*; see also Section 5.

The extended data set we consider contains 6,870 recorded claims of the period 01/1980 - 12/2002. Note that due to the recording method, there is an artificial negative dependence between the components, since if one component is smaller than 1 million DKK, the sum of the others must be accordingly larger. We eliminate this artificial dependence by consulting a claim only if *both* components X_i (Building) *and* Y_i (Content) attain at least 1 million DKK. Hence, in the sequel, we analyze the conditional distribution of (X, Y) given $\{X \geq 1, Y \geq 1\}$, where here and below we refer to claims in millions of DKK. The sample size of the remaining data is $n = 588$. We discounted the claim sizes to 7/1985 prices according to the Danish Consumer Price Index (DCPI)[5] on a monthly basis. Figure 2 displays the scatterplot of the data and of the pseudo-observations $(\hat{U}_i, \hat{V}_i)$.

To choose a suitable number of order statistics m used for the Hill estimator $\hat{\eta}_n$, we consider a so-called Hill plot, which displays the estimates $\hat{\eta}_n$ versus m. A small value m results in a high variance of $\hat{\eta}_n$, while too large an m may cause a large bias. A typical (nice) Hill plot exhibits heavy fluctuations for small values of m (due to large variance), followed by a rather stable region where both variance and bias are moderate, before the bias causes an almost monotone decrease or increase of the curve. Thus, m ought to be chosen in the region where the plot is rather stable. For a more extended discussion on this matter see e.g. Sections 6.4 and 6.5 of Embrechts et al. (1997) or Drees et al. (2000). For the present data set, the (heuristic) analysis of the Hill plot (Figure 3) suggests $m = 200$, which yields $1/T^{(n)}_{n-m:n} = 0.53$ (indicated in the right hand plot of Figure 2) and $\hat{\eta}_n = 0.76$. The 95% confidence interval (15) for η is $[0.66, 0.87]$. The pertaining test (16) clearly rejects the null hypothesis $\eta = 1$ on a 95% confidence level with $(1-p)$-value > 0.9999, so that we assume asymptotic independence.

Only 5 of the points (21), i.e. 2.5 %, lie outside their 95% confidence interval, so that our method accepts the presence of the scaling law (11). The pertaining plots (21) are shown in Figure 4, where gray points lie inside their confidence interval, black points outside. A more detailed three-dimensional presentation including a discretized specification of the $(1 - p)$-values (23) of the respective test (22) for all $j = 1, \ldots, m$ is available at *http://www.math.uni-hamburg.de/home/drees/extrdep/mdilc.html* (plot generating MatLab files and an avi file with a rotating plot).

Table 1 compares estimates and 95% confidence intervals for the conditional probability of the jointly large event $P(X > u_1, Y > u_2 \mid X \geq 1, Y \geq 1)$.

[5] *http://www.dst.dk/Statistik/seneste/Indkomst/Priser/Forbrugerprisindeks.aspx*

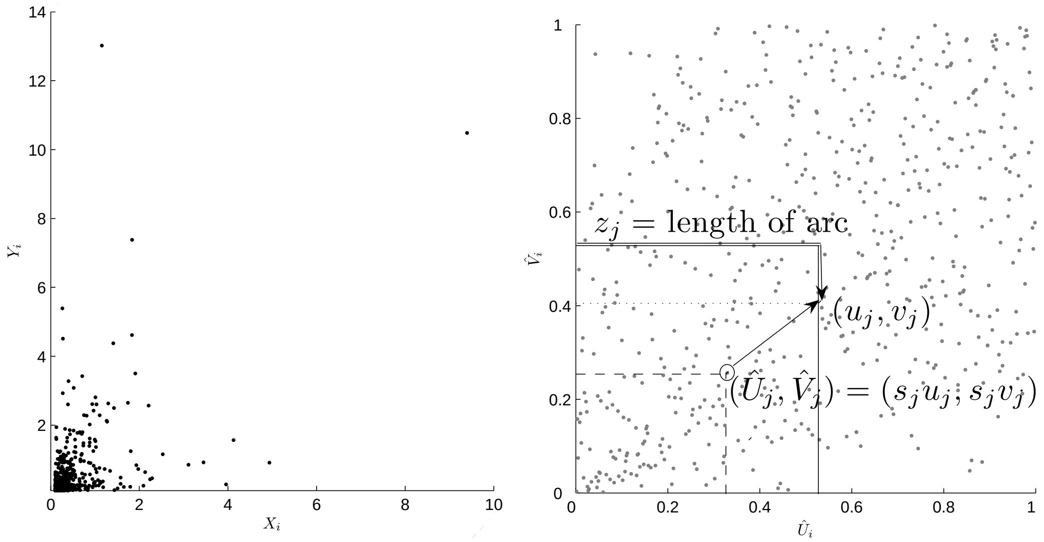

Fig. 2. Scatterplot X_i vs. Y_i and $\hat{U}_i$ vs. $\hat{V}_i$. The square with side length $1/T^{(n)}_{n-m:n} = 0.53$ contains the $m = 200$ points used for the estimation of η. The scaling law (11) is checked within this square.

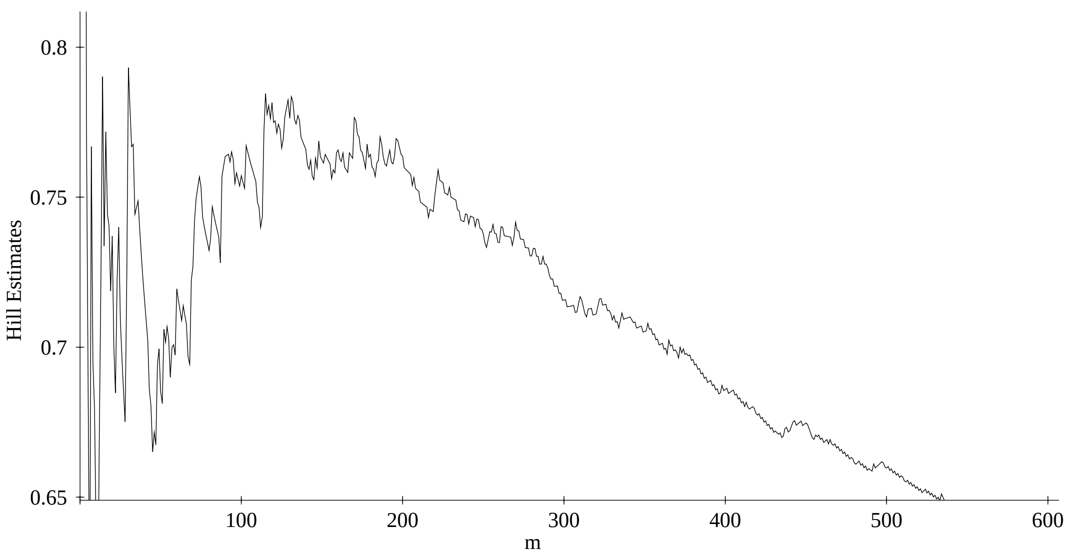

Fig. 3. Hill plot: $\hat{\eta}_n$ as a function of m.

Denote by $\hat{G}_{i,n}(x)$, $i = 1, 2$ the estimates for the conditional marginal d.f.s of X and Y given $\{X \geq 1, Y \geq 1\}$, respectively. For the applications in this section we introduce a location parameter μ and a scale parameter ς by replacing the argument x in (2) by $(x-\mu)/\varsigma$. See Section 6.5 of Embrechts et al. (1997) for a discussion of how to estimate γ, μ and ς in such a location-scale setting. We obtain $1 - \hat{G}_{1,n}(x) = (1 + 0.47(x - 0.96)/1.85)^{-2.14}$ and $1 - \hat{G}_{2,n}(y) = (1 + 0.52(y - 0.72)/2.45)^{-1.93}$. 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.

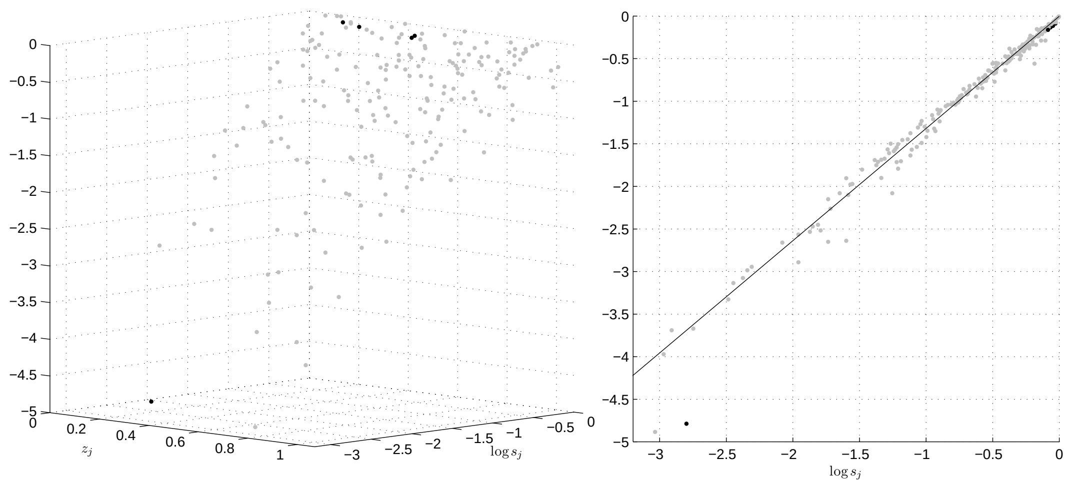

Fig. 4. Plot (21) with $m = 200$, $1/T^{(n)}_{n-m:n} = 0.53$, $\hat{\eta}_n = 0.76$ from two perspectives. The right hand perspective also displays the reference plane. Five (black) points lie outside their 95% confidence interval.

Further, estimates $\hat{p}_i := (1 - \hat{G}_{1,n}(u_1))(1 - \hat{G}_{2,n}(u_2))$ assuming independence of X and Y are given.

u_1	u_2	$\hat{p}_n$ in %	$\hat{p}_e$ in %	$\hat{p}_i$ in %
5	5	12.48 [11.30,13.77]	12.59 [10.01,15.54]	6.38
10	10	3.49 [2.80,4.34]	3.40 [2.09,5.20]	0.96
20	20	0.75 [0.49,1.16]	0.34 [0.04,1.22]	0.10
25	25	0.44 [0.26,0.73]	0.17 [0.00,0.94]	0.046
100	10	0.018 [0.01,0.05]	0 [0,0.63]	0.012

Table 1. Comparison of the estimates $\hat{p}_n$, the empirical probability $\hat{p}_e$ and the estimated probability $\hat{p}_i$ assuming independence of X and Y.

We observe that the empirical probability is close to the estimates obtained in the Extended Ledford & Tawn Model 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 2.) Of course, the empirical probability is of very limited value if the event under consideration has not occurred yet. As expected, the assumption of independence of X and Y yields systematically smaller estimates than the empirical probability (if the latter is positive) and those obtained in the Extended Ledford & Tawn Model.

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 [6]. 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 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 sort out those claims with $\min(X_i, Y_i) < 25,000$ USD. Hence, in the sequel, we analyze the conditional distribution of (X, Y) given $\{X \geq 25, Y \geq 25\}$, where here and below we refer to claims in thousands of USD. The sample size of the remaining data is $n = 7675$.

Figure 5 displays the scatterplot of the data and of the pseudo-observations $(\hat{U}_i, \hat{V}_i)$. The analysis of the Hill plot (Figure 6) leads to $m = 1,000$, which yields $1/T^{(n)}_{n-m:n} \approx 0.34$ (indicated in the right hand plot of Figure 5) and $\hat{\eta}_n \approx 0.59$. The 95% confidence interval (15) is $[0.55, 0.62]$. The pertaining test (16) clearly rejects the hypothesis $\eta = 1$ on a 95% confidence level with $(1 - p)$-value > 0.9999, so that we assume asymptotic independence.

We obtain that 49 of the points (21), i.e. 4.9 %, lie outside their 95% confidence interval so that our method accepts the presence of the scaling law (11). The pertaining plot (21) is shown in Figure 7, where gray points lie inside their confidence interval, black points outside. A more detailed three-dimensional presentation including a discretized specification of the $(1 - p)$-values (23) of the respective test (22) for all $j = 1, \ldots, m$ is available at *http://www.math.uni-hamburg.de/home/drees/extrdep/mdilc.html* (plot generating MatLab files and an avi file with the rotating plot).

Table 2 compares probability estimates and 95% confidence intervals for the probability of the jointly large event $P(X > u_1, Y > u_2 \mid X \geq 25, Y \geq 25)$. Denote by $\hat{G}_{i,n}(x)$, $i = 1, 2$, the estimates for the conditional marginal d.f.s of X and Y given $\{X \geq 25, Y \geq 25\}$, respectively. Then, for $\hat{p}_n$ and $\hat{p}_i$ we used the univariate GPD fits $1 - \hat{G}_{1,n}(x) = (1 + 0.26(x - 22.95)/58.62)^{-3.80}$ and

[6] *http://www.soa.org/ccm/content/research-publications/experience-studies-tools/1991-92-group-medical-insurance-large-claims-database/*

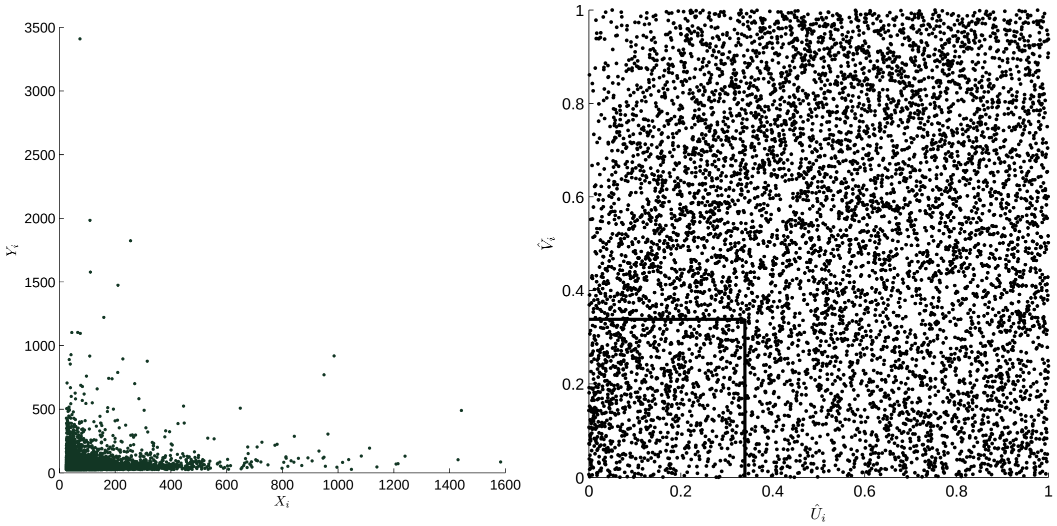

Fig. 5. Scatterplot X_i vs. Y_i and $\hat{U}_i$ vs. $\hat{V}_i$. The square with side length $1/T^{(n)}_{n-m:n} = 0.34$ contains the $m = 1,000$ points used for the estimation of η. The scaling law (11) is checked within this square.

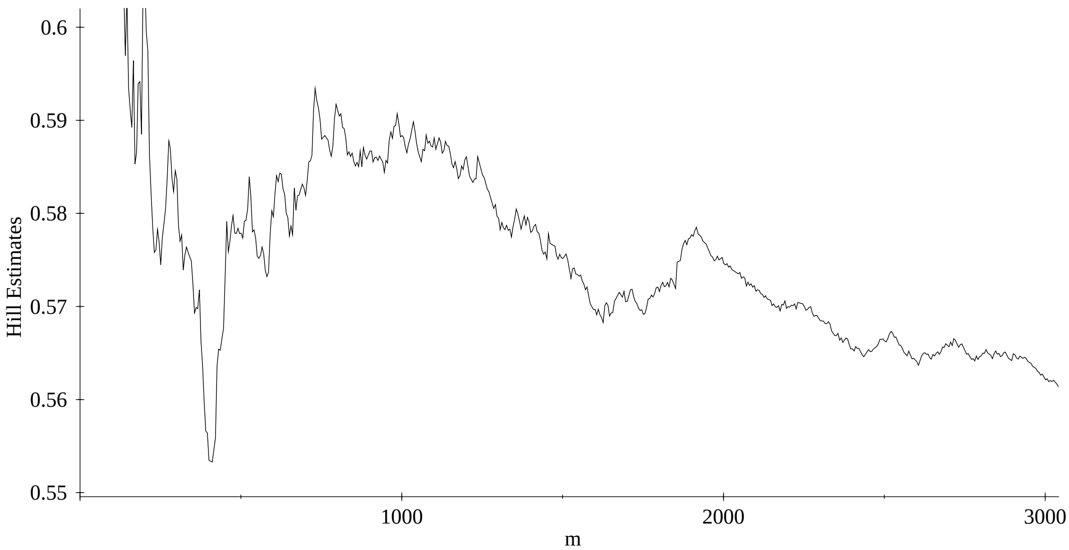

Fig. 6. Hill plot: $\hat{\eta}_n$ as a function of m.

$$1 - \hat{G}_{2,n}(y) = (1 + 0.41(y + 2.59)/28.72)^{-2.45}.$$

The observations made in Table 1 are confirmed. In fact, the results in Table 2 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.

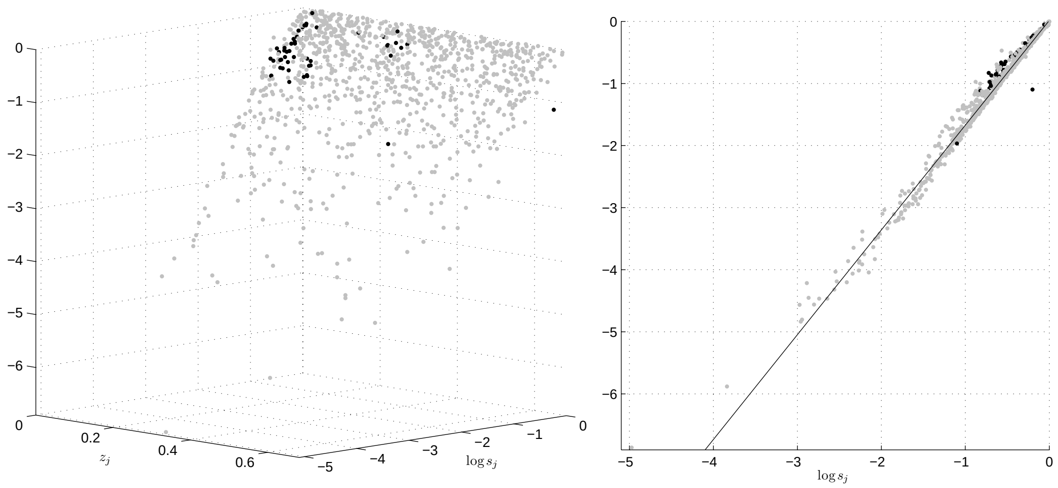

Fig. 7. Plot (21) with $m = 1,000$, $1/T^{(n)}_{n-m:n} \approx 0.34$, $\hat{\eta}_n \approx 0.59$ from two perspectives. The right hand perspective also displays the reference plane. 49 (black) points lie outside their 95% confidence interval.

u_1	u_2	$\hat{p}_n$ in %	$\hat{p}_e$ in %	$\hat{p}_i$ in %
100	100	3.88 [3.81,3.95]	3.82 [3.40,4.27]	3.57
200	200	0.59 [0.53,0.64]	0.61 [0.45,0.81]	0.39
400	400	0.052 [0.04,0.07]	0.07 [0.02,0.15]	0.022
600	400	0.026 [0.02,0.03]	0.05 [0.01,0.13]	0.007

Table 2. Comparison of the estimates $\hat{p}_n$, the empirical probability $\hat{p}_e$ and the estimated probability $\hat{p}_i$ assuming independence of X and Y.

For the purpose of comparison, Figure 8 displays the plot (21) for $m = 3,800$, that is, about half of all data points are used for the model fitting. (A more sophisticated plot is provided by files on the aforementioned website.) Since data is used which cannot be considered extreme, it seems quite likely that the scaling law (11), that was motivated by asymptotic arguments for extreme observations, does not hold on the much larger square with side length $1/T^{(n)}_{n-m:n} \approx 0.70$. Indeed, we observe that 1,729 points, about 46%, lie outside their 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 $m = 3,800$ is too large, the resulting point estimate $\hat{\eta}_n \approx 0.55$ belongs to the confidence interval for η obtained with the choice $m = 1,000$, i.e. the difference is not statistically significant.)

This example demonstrates that our method cannot only be used to decide whether the Extended Ledford & Tawn Model is appropriate or not in *some* tail region, but also, loosely speaking, to determine the maximum tail region for which the application of the Extended Ledford & Tawn Model is still reasonable.

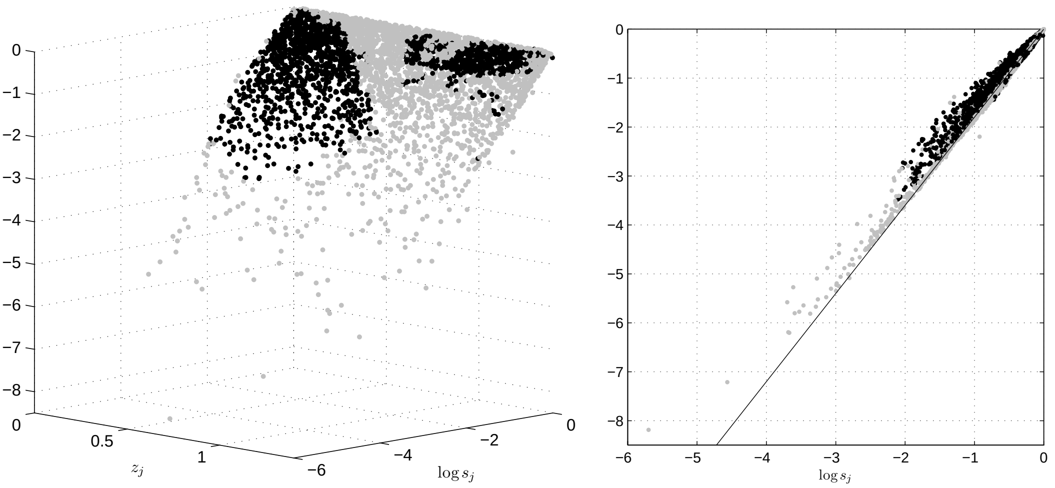

Fig. 8. Plot (21) with $m = 3,800$, $1/T^{(n)}_{n-m:n} = 0.70$, $\hat{\eta}_n = 0.55$ from two perspectives. The right hand perspective also displays the reference plane. 1,729 (black) points lie outside their 95% confidence interval.

We conclude this section with the general warning that it is an inherent problem of tail modeling that the estimation accuracy decreases when we extrapolate farther into the tail, i.e. if the thresholds u_1 and u_2 increase. Small errors in the estimation of η then have greater impact on estimates of the probability of more extreme events (imaginable as a leverage effect on the reference plane), particularly if u_1 and u_2 lie outside the range of the data.

6 Proofs

Proof of Theorem 3.1. 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. Furthermore, 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 $X_{0:n} := Y_{0:n} := U_{0:n} := V_{0:n} := 0$ and $X_{j:n} := Y_{j:n} := U_{j:n} := V_{j:n} := 1$ for $j > n$. The first aim is to apply Lemma 6.1 of Draisma et al. (2004) to $\sum_{i=1}^n \mathbb{1}_{\{\hat{U}_i < rx, \hat{V}_i < ry\}}$ and $\sum_{i=1}^n \mathbb{1}_{\{\hat{U}_i < r, \hat{V}_i < r\}}$.

Recall that F_i is assumed continuous on $[\xi_i, \infty)$ for some $\xi_i < F_i^{\leftarrow}(1)$, $i = 1, 2$. If $X_{n-k:n} \geq \xi_1$, then $X_{n-k:n} < X_{n-k+1:n} < \dots < X_{n:n}$ almost surely (a.s.). Since $r \to 0$, and thus $X_{\lfloor n-(n+1)rx_0+1 \rfloor:n} \to F_1^{\leftarrow}(1)$, it follows that the probability that there are no ties between these order statistics tends to 1, i.e. $P\{X_{\lfloor n-(n+1)rx_0+1 \rfloor:n} < X_{\lfloor n-(n+1)rx_0+1 \rfloor+1:n} < \dots < X_{n:n}\} \to 1$ as $n \to \infty$ for all $x_0 > 0$. Furthermore, if there are no ties, then the condition $\hat{U}_i < rx$ is equivalent to $R_i^X > n - (n+1)rx + 1$ which in turn is equivalent to

$X_i > X_{\lfloor n-(n+1)rx+1\rfloor:n}$ and to $U_i < U_{\lceil (n+1)rx\rceil:n}$. Hence, together with analogous arguments for V_i and Y_i and the definition $S(x,y) := \sum_{i=1}^n \mathbb{1}_{\{U_i\le x, V_i\le y\}}$ we see that $\sum_{i=1}^n \mathbb{1}_{\{\hat U_i<rx,\hat V_i<ry\}}/\sum_{i=1}^n \mathbb{1}_{\{\hat U_i<r,\hat V_i<r\}}$ 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 $x,y\in(0,x_0]$ with probability tending to 1 as $n\to\infty$ for all $x_0>0$. It is therefore sufficient to prove the assertion of Theorem 3.1 for $\tilde c_n$, i.e. $\sqrt{\tilde m}(\tilde c(x,y)-c(x,y)) \xrightarrow{D} \mathcal{N}(0,\sigma^2_{x,y})$ for all $x,y\in(0,1]$.

Let W be a Gaussian process with mean zero and covariance structure given by

$$E\left[W(x_1,y_1),W(x_2,y_2)\right] = c(x_1\wedge x_2, y_1\wedge y_2). \tag{24}$$

Lemma 6.1 of Draisma et al. (2004) states that under the conditions of Theorem 3.1,

$$\sqrt{\tilde m}\left(\frac{S(U_{\lfloor nrx\rfloor:n}, V_{\lfloor nry\rfloor:n})}{\tilde m} - c(x,y)\right)_{(x,y)\in[0,\infty)^2} \longrightarrow (W(x,y))_{(x,y)\in[0,\infty)^2}. \tag{25}$$

weakly in the bivariate Skorohod space $D([0,\infty)^2)$. Since W possesses a.s. continuous sample paths (cf. the proof of Lemma 6.2 of Draisma et al. (2004)), this convergence holds uniformly on compact subsets of $[0,\infty)^2$. Further, according to the Skorohod-Dudley-Wichura representation theorem (see e.g. Shorack and Wellner (1986), p. 47), this convergence holds a.s. for suitable versions of the process $(S(U_{\lfloor nrx\rfloor:n}, V_{\lfloor nry\rfloor:n}))_{(x,y)\in[0,\infty)^2}$ and W. Hence, for these versions

$$\begin{aligned}\tilde c_n(x,y) &= \frac{c(x,y)+\tilde m^{-1/2}W(x,y)+o(\tilde m^{-1/2})}{c(1,1)+\tilde m^{-1/2}W(1,1)+o(\tilde m^{-1/2})}\\ &= \Big(c(x,y)+\tilde m^{-1/2}W(x,y)+o(\tilde m^{-1/2})\Big)\\ &\qquad\qquad \Big(1-\tilde m^{-1/2}W(1,1)+o(\tilde m^{-1/2})\Big) \text{ a.s.,}\end{aligned}$$

i.e.

$$\sqrt{\tilde m}(\tilde c_n(x,y)-c(x,y)) \to W(x,y)-c(x,y)W(1,1) \text{ a.s.}$$

uniformly on compact subsets of $[0,\infty)^2$. The covariance structure (24) of W yields

$$\begin{aligned}\operatorname{var}\left(W(x,y)-c(x,y)W(1,1)\right) &= E\left[(W(x,y)-c(x,y)W(1,1))^2\right]\\ &= c(x,y)-2c(x,y)c(x\wedge 1, y\wedge 1)+c^2(x,y),\end{aligned}$$

see also Remark 3.2 (ii). The assumption $x, y \in (0,1]$ yields the assertion. $\square$

Proof of Theorem 4.1. First, we argue as in the proof of Theorem 3.1 to see that

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

$$= \log \frac{S\left(U_{\lceil sx(n+1)/T^{(n)}_{n-m:n}\rceil -1:n}, V_{\lceil sy(n+1)/T^{(n)}_{n-m:n}\rceil -1:n}\right)}{S\left(U_{\lceil x(n+1)/T^{(n)}_{n-m:n}\rceil -1:n}, V_{\lceil y(n+1)/T^{(n)}_{n-m:n}\rceil -1:n}\right)} \tag{26}$$

for all $x, y \in (0, x_0]$ with probability tending to 1 as $n \to \infty$ for all $x_0 > 0$. As in (25), according to the Skorohod-Dudley-Wichura representation theorem, there is a process $(S^*_n(x,y))_{(x,y)\in[0,\infty)^2}$ with the same distribution as $(S(U_{\lfloor nrx\rfloor:n}, V_{\lfloor nry\rfloor:n}))_{(x,y)\in[0,\infty)^2}$, where $r = q^{\leftarrow}(m/n)$, and a Brownian motion W such that

$$\sqrt{m}\left(\frac{S^*_n(x,y)}{m} - c(x,y)\right)_{(x,y)\in[0,\infty)^2} \longrightarrow (W(x,y))_{(x,y)\in[0,\infty)^2} \tag{27}$$

holds a.s. uniformly on compact subsets of $[0,\infty)^2$. Let $\bar{S}^*_n(x) := S^*_n(x,x)$. Because of

$$T^{(n)}_{n-k:n} = \inf\left\{x \in (0,\infty) \mid S(U_{\lfloor (n+1)/x\rfloor:n}, V_{\lfloor (n+1)/x\rfloor:n}) \le k\right\},\ k = 0,\dots,n-1,$$

the process

$$Q^*_n(t) := \inf\left\{x \in (0,\infty) \mid \bar{S}^*_n((n+1)/(nrx)) \le \lfloor mt\rfloor\right\}, \quad t \in [0,1],$$

possesses the same distribution as the tail empirical quantile process $Q_n = (T^{(n)}_{n-\lfloor mt\rfloor:n})_{t\in[0,1]}$. Thus,

$$\left(S^*_n\left(\frac{(n+1)sx}{nrQ^*_n(1)}-, \frac{(n+1)sy}{nrQ^*_n(1)}-\right)\right)_{(x,y)\in(0,\infty)^2}$$

is a version of the process given in the numerator of the right hand side of (26), where $f(x-)$ denotes the left-hand limit of f at x.

Next note that

$$(Q^*_n)^{\leftarrow}(x) := \inf\{t \in [0,1] \mid Q^*_n(t) \le x\} = \frac{1}{m}\bar{S}^*_n\left(\frac{n+1}{nrx}\right)$$

and thus by (27)

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

Now one can proceed as in the proof of Lemma 6.2 of Draisma et al. (2004) to conclude

$$rQ_n^*(t) = t^{-\eta} + m^{-1/2}\eta t^{-(\eta+1)}W(t^\eta,t^\eta) + o(m^{-1/2}) \quad \text{ a.s.}$$

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 $1/(rQ_n^*(1)) = 1-m^{-1/2}\eta W(1,1)+o(m^{-1/2})$ a.s. for $t = 1$. Further, consider $(c(tx+t\varepsilon,ty)-c(tx,ty))/(t\varepsilon)$ as $\varepsilon \downarrow 0$ to see that c_x and c_y are homogeneous of order $1/\eta - 1$. Hence, by (27) with x replaced with $x(1 - m^{-1/2}\eta W(1,1) + o(m^{-1/2}))$, and the homogeneity of c and of c_x and c_y

$$\begin{aligned}
&m^{-1}S_n^*\left(\frac{(n+1)sx}{nrQ_n^*(1)}-, \frac{(n+1)sy}{nrQ_n^*(1)}-\right)\\
&= c\left(sx(1-m^{-1/2}\eta W(1,1)+o(m^{-1/2})), sy(1-m^{-1/2}\eta W(1,1)+o(m^{-1/2}))\right)\\
&\qquad\qquad + m^{-1/2}W(sx,sy) + o(m^{-1/2})\\
&= c(sx,sy) - m^{-1/2}\eta W(1,1)sxc_x(sx,sy) - m^{-1/2}\eta W(1,1)syc_y(sx,sy)\\
&\qquad\qquad + m^{-1/2}W(sx,sy) + o(m^{-1/2})\\
&= s^{1/\eta}c(x,y)\bigg[1 - m^{-1/2}\eta W(1,1)x\frac{c_x(x,y)}{c(x,y)} - m^{-1/2}\eta W(1,1)y\frac{c_y(x,y)}{c(x,y)}\\
&\qquad\qquad + m^{-1/2}\frac{W(sx,sy)}{c(sx,sy)} + o(m^{-1/2})\bigg] \text{ a.s.}
\end{aligned}$$

Further, we use $1/(1-z) = 1 + z + o(z)$ as $z \to 0$ once more to see that

$$\begin{aligned}
&m/S_n^*\left(\frac{(n+1)x}{nrQ_n^*(1)}-, \frac{(n+1)y}{nrQ_n^*(1)}-\right)\\
&= \frac{1}{c(x,y)}\bigg[1 + m^{-1/2}\eta W(1,1)x\frac{c_x(x,y)}{c(x,y)} + m^{-1/2}\eta W(1,1)y\frac{c_y(x,y)}{c(x,y)}\\
&\qquad\qquad - m^{-1/2}\frac{W(x,y)}{c(x,y)} + o(m^{-1/2})\bigg] \text{ a.s.}
\end{aligned}$$

With $\log(1 + z) = z + o(z)$ as $z \to 0$ it follows

$$\log \frac{S_n^*\left(\frac{(n+1)sx}{nrQ_n^*(1)}-,\frac{(n+1)sy}{nrQ_n^*(1)}-\right)}{S_n^*\left(\frac{(n+1)x}{nrQ_n^*(1)}-,\frac{(n+1)y}{nrQ_n^*(1)}-\right)} \tag{28}$$

$$= \frac{1}{\eta}\log s + m^{-1/2}\left(\frac{W(sx,sy)}{c(sx,sy)} - \frac{W(x,y)}{c(x,y)}\right) + o(m^{-1/2}) \text{ a.s.}$$

To continue the proof of Theorem 4.1, we need an asymptotic representation of the corresponding version $1/\hat{\eta}_n^*$.

6.1 Lemma. For above processes S_n^* and W and the corresponding version $\hat{\eta}_n^* := m^{-1}\sum_{i=1}^m \log(Q_n^*((i-1)/m)/Q_n^*(1))$ of $\hat{\eta}_n$,

$$\hat{\eta}_n^* = \eta\left(1 + m^{-1/2}\left(\int_0^1 t^{-1}W(t^\eta,t^\eta)dt + W(1,1)\right)\right) + o_P(m^{-1/2}). \tag{29}$$

Proof. The proof is based on results of Drees (1998), Example 3.1, p. 103. We may write

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

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

$$y_n(t) := \sqrt{m}\left(rQ_n^*(t) - t^{-\eta}\right).$$

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

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

Hence,

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

Exactly as in Drees (1998), Example 3.1, p. 103, where $Q_n^*(t)$, η and $m^{-1/2}$ play the roles of $z(t)$, β and λ_n, we conclude that

$$\hat{\eta}_n^* - \eta = \int_0^1 m^{-1/2} \frac{t^\eta y_n(t) - y_n(1)}{1 + m^{-1/2} y_n(1)} dt + o(m^{-1/2}).$$

According to (the proof of) Lemma 6.2 of Draisma et al. (2004),

$$\sup_{0<t\leq 1} t^{\eta-1/2+\varepsilon} |y_n(t) - y(t)| \xrightarrow{P} 0, \qquad (30)$$

for all $\varepsilon > 0$, where $\xrightarrow{P}$ denotes convergence in probability as $n \to \infty$. Thus,

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

□

6.2 Corollary.

$$\frac{1}{\hat{\eta}_n^*} = \frac{1}{\eta} - \frac{m^{-1/2}}{\eta} \left(\int_0^1 t^{-1} W(t^\eta, t^\eta) dt + W(1,1) \right) + o_P(m^{-1/2}). \qquad (31)$$

Proof. The Taylor expansion of the reciprocal of (29) gives

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

□

By (28) and Corollary 6.2, we obtain

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

Define $Z(t) := \eta^{-1}(\log s)\,(t^{-1}W(t^\eta, t^\eta) - W(1,1)) + W(sx,sy)/c(sx,sy) - W(x,y)/c(x,y)$, so that it remains to show that $\int_0^1 Z(t)dt \sim \mathcal{N}(0, \sigma^2_{x,y,s})$.

A special case of Proposition 2.2.1 of Shorack and Wellner (1986) states that $\int_0^1 Z(t)dt \sim \mathcal{N}(0, \sigma^2_{\int Z})$, with $\sigma^2_{\int Z} = \int_0^1 \int_0^1 \operatorname{cov}(Z(r), Z(t))drdt$. By (24) we obtain

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

Straightforward calculations lead to

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

see also Remark 4.2 (i). Hence, the assumption $x, y \in (0, 1]$ together with the homogeneity (10) of c yields the required result. □

Although the proof of the case of asymptotic dependent variables U and V requires significantly messier calculations, the proceeding is analogous. The additional effort is mainly due to the substantially more complicated (covariance) structure of the process W. The proof is provided in the accompanying technical note Drees and Müller (2006).

Acknowledgements

The extended data set of the Danish fire insurance claims was reported to us by Mette Havning and Mogens Steffensen.

References

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of Extremes.* Wiley, Chichester.

Blum, P., Dias, A. and Embrechts, P. (2002): *The ART of Dependence Modelling: The Latest Advances in Correlation Analysis* in *Alternative Risk Strategies*, edited by Morton Lane, 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

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

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., de Haan, L. and Resnick, S.I. (2000): *How to Make a Hill Plot*, The Annals of Statistics 28 (1), 254-274

Drees, H. (1998): *A General Class of Estimators of the Extreme Value Index*, Journal of Statistical Planning and Inference 66, 95-112

Drees, H. and Müller, P. (2006): *Technical Note on Modeling Dependencies in Large Insurance Claims, Technical Supplement and The Case of Asymptotic Dependence*, http://www.math.uni-hamburg.de/drees/extrdep/mdilc.html

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997): *Modelling Extremal Events for Insurance and Finance*, Springer, New York

Embrechts, P., McNeil, A.J. and Straumann, D. (2002): *Correlation and Dependence in Risk Management: Properties and Pitfalls*, in *Risk Management: Value at Risk and Beyond*, edited by Dempster, M.A.H., Cambridge University Press, Cambridge

Grazier, K.L. and G'Sell Associates (1997): *Group Medical Insurance Large Claims Data Base Collection and Analysis*, SOA Monograph M-HB97-1, Society of Actuaries, Schaumburg, Illinois

Ledford, A.W. and Tawn, J.A. (1997): *Modelling Dependence within Joint Tail Regions*, Journal of Royal Statistical Society 59(2), 475-499

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.

Mikosch, T. (2005): *How to model multivariate extremes if one must.* Statistica Neerlandica 59, 324-338.

Resnick, S.I. (1987): *Extreme Values, Regular Variation and Point Processes*, Springer-Verlag, New York

Rytgaard, M. (1996): *Simulation Experiments on the Mean Residual Life*

Function $m(x)$, in Proceedings of the XXVII Astin Colloquium, Copenhagen, Denmark (1), 59-81
Santner, T.J. and Duffy, D.E. (1989): *The Statistical Analysis of Discrete Data*, Springer Verlag, New York
Shorack, G.R. and Wellner, J.A. (1986): *Empirical Processes with Applications to Statistics*, Wiley Series in Probability and Mathematical Statistics, New York