Validation of the Ledford & Tawn Model

Holger Drees Peter Müller

University of Hamburg

Extreme Value Analysis
Gothenburg, August 15-19, 2005

Outline

Modelling Dependence

(X, Y), (X_i, Y_i) $\mathbb{R}^2$–valued, iid with d.f. F

$F_1(x) := F(x, \infty)$, $F_2(y) := F(\infty, y)$ assumed continuous in right tail

Example: (X, Y) claim sizes in two lines of business of insurance company

We assume that marginal df's modelled using univariate extreme value statistics

To model dependence structure standardize margins to uniform df:

$$U := 1 - F_1(X), \qquad V := 1 - F_2(Y)$$

Aim: Model df of (U, V) (survival copula) on neighborhood of origin

Basic Model Assumption

$$P\Big(\frac{U}{t} < x,\ \frac{V}{t} < y \,\Big|\, U < t, V < t\Big) = \frac{P\{U < tx,\ V < ty\}}{P\{U < t,\ V < t\}} \xrightarrow{t\downarrow 0} c(x,y)$$

uniformly on $\{(x,y) \mid \max(x,y) = 1\}$ for some non-degenerate function c

Consequences:

- c homogeneous of order $1/\eta$ for some $\eta \in (0,1]$:

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

- $t \mapsto P\{U < t,\ V < t\}$ regularly varying at 0 with exponent $1/\eta$

Coefficient of Tail Dependence η

- If $\eta < 1$, then for some slowly varying function l

$$P(U < t \mid V < t) = \frac{P\{U < t,\ V < t\}}{t} = t^{1/\eta - 1} l(t) \xrightarrow{t \downarrow 0} 0$$

i.e., *asymptotic independence*

- Roughly speaking

$\eta = 1$:	asymptotic dependence
$\eta \in (1/2, 1)$:	positive dependence, vanishes asymptotically
$\eta = 1/2$:	independence
$\eta \in (0, 1/2)$:	negative dependence, vanishes asymptotically

Scaling Law

In the Ledford & Tawn model the following scaling law holds:

$$\frac{P\{U < tx,\ V < ty\}}{P\{U < x, V < y\}} \approx t^{1/\eta}$$

for small x, y, because

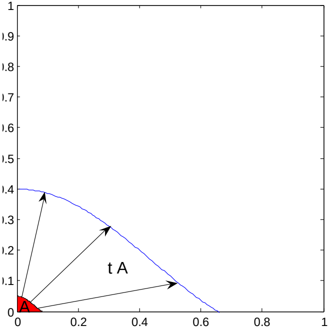
$$\frac{P\{U < tx,\ V < ty\}}{P\{U < x, V < x\}} \approx c(t, ty/x) = t^{1/\eta} c(1, y/x)$$

$$\frac{P\{U < x,\ V < y\}}{P\{U < x, V < x\}} \approx c(1, y/x)$$

Scaling Law

More generally: For sets A nearby origin

$$\frac{P\{(U,V) \in tA\}}{P\{(U,V) \in A\}} \approx t^{1/\eta}$$



Blowing up set A by factor t increases probability by factor $t^{1/\eta}$

Estimating the Coefficient of Tail Dependence

survival function $1 - F_T$ of

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

is regularly varying with exp. $-1/\eta$, since $P\{T_i > t\} = P\{U_i < 1/t, V_i < 1/t\}$.

$\rightsquigarrow$ approximate U_i, V_i with

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

and apply Hill estimator to $m = m_n$ largest order statistics of $\hat{T}_i := \min\left(\frac{1}{\hat{U}_i}, \frac{1}{\hat{V}_i}\right)$

$\rightsquigarrow$ $\hat{\eta}_n$

Draisma et al. (2004): asympt. normality, if $m_n \to \infty$ not too fast, c smooth.

Graphical Tools

$$\frac{P\{U < tx,\ V < ty\}}{P\{U < x, V < y\}} \approx t^{1/\eta} \qquad \text{for small } x, y$$

Hence

$$\frac{1}{\eta} \log t \approx \log \frac{P\{U < tx,\ V < ty\}}{P\{U < x, V < y\}} \approx \log \frac{\sum_{i=1}^{n} \mathbb{1}_{\{\hat{U}_i < tx,\ \hat{V}_i < ty\}}}{\sum_{i=1}^{n} \mathbb{1}_{\{\hat{U}_i < x,\ \hat{V}_i < y\}}},$$

i.e. points

$$\left(\log t, \log \frac{\sum_{i=1}^{n} \mathbb{1}_{\{\hat{U}_i < tx,\ \hat{V}_i < ty\}}}{\sum_{i=1}^{n} \mathbb{1}_{\{\hat{U}_i < x,\ \hat{V}_i < y\}}} \right)$$

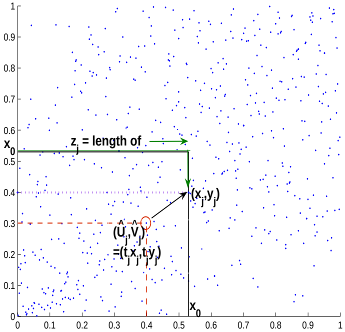
approximately on line through origin with slope $1/\eta$, independent of (x, y).

Planar Log-Log-Plot

Points

$$\left(z_j, \log t_j, \log \frac{\sum_{i=1}^n \mathbb{1}_{\{\hat{U}_i < t_j x_j, \hat{V}_i < t_j y_j\}}}{\sum_{i=1}^n \mathbb{1}_{\{\hat{U}_i < x_j, \hat{V}_i < y_j\}}} \right)$$

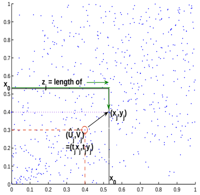
should approximately lie on plane $(z, u) \mapsto (z, u, u/\eta)$ where

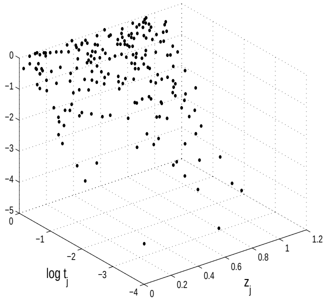


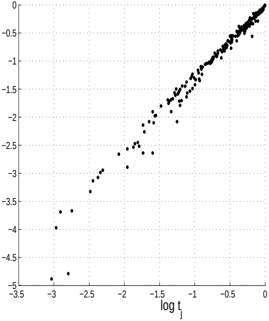
and $x_0 = \frac{1}{T_{n-m_n+1:n}}$

(i.e., consider region used for estimation of η)

Planar Log-Log-Plot

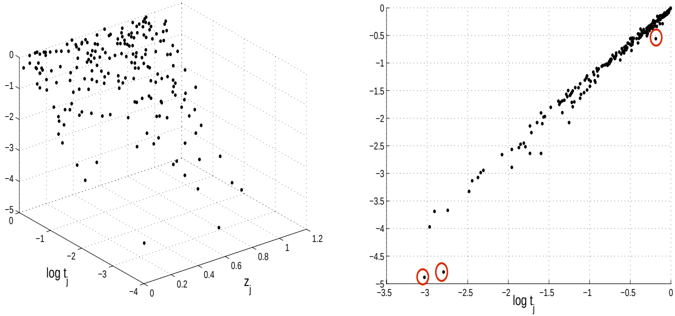






Planar Log-Log-Plot

Which deviations from plane are significant?



Confidence Intervals

Under asymptotic independence and further conditions:

Estimated deviation from plane

$$\log \frac{\sum_{i=1}^{n} \mathbb{1}\{\hat{U}_i < tx, \hat{V}_i < ty\}}{\sum_{i=1}^{n} \mathbb{1}\{\hat{U}_i < x, \hat{V}_i < y\}} - \frac{1}{\hat{\eta}_n} \log t$$

approximately distributed according to $\mathcal{N}(0, m_n^{-1}\sigma^2_{x,y,t})$ with

$$\sigma^2_{x,y,t} = \frac{t^{-1/\eta} - 1}{(F_T^{-1}(1 - m_n/n))^{1/\eta} c(x,y)} - \frac{\log^2 t}{\eta^2}$$

▸ exact result ▸ proof

$\rightsquigarrow$ test whether deviation of single point of plot is significant

Graphical tool: Use colors to indicate p-values

Data

Claim sizes of US health insurer in 1991

- X_i: hospital
- Y_i: other

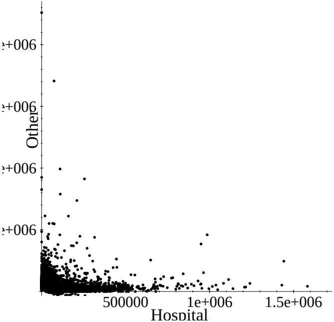
Claims reported only if $X_i + Y_i \geq 25\,000$ (\$)

$\rightsquigarrow$ 92 750 claims

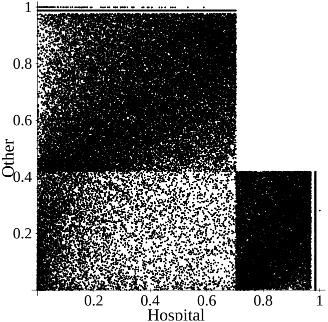
If interested in dependence structure for $(X, Y) \in [25\,000, \infty)^2$,
then suffices to consider only (X_i, Y_i) with $\max(X_i, Y_i) \geq 25\,000$

$\rightsquigarrow$ $n = 62\,822$ claims

Standardize Marginal Distributions

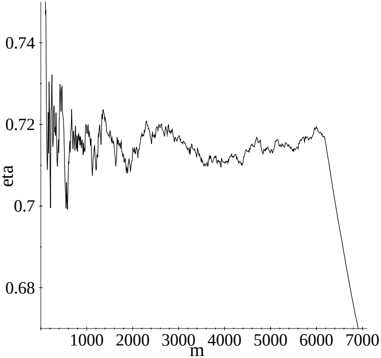


(X_i, Y_i)

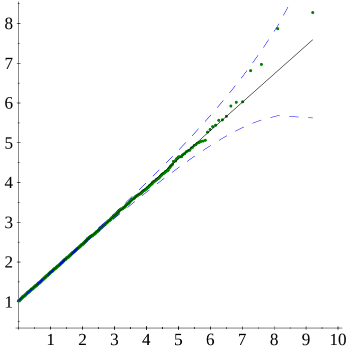


$$(\hat{U}_i, \hat{V}_i) = \left(1 - \frac{R_i^X}{n+1}, 1 - \frac{R_i^Y}{n+1}\right)$$

Estimate η



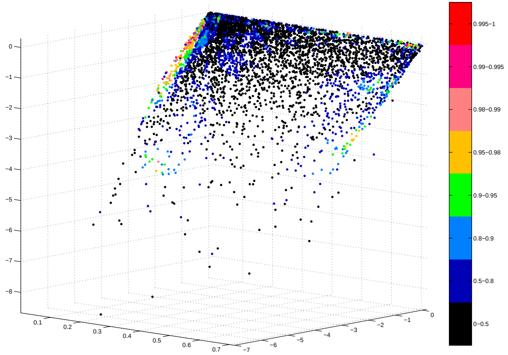
Hill plot



Hill $qq-plot$ with 95%-confidence intervals

$m = 5000 \quad \rightsquigarrow \hat{\eta}_n \approx 0.713 \quad ([0.693, 0.733])$

Model Check



test at 5%-level rejects model for 3.6% of points

Beware!

Due to standardization with marginal df's

η and c do not depend only on large X, Y!

Similar analysis based on $(X_i, Y_i) \in [25\,000, \infty)^2$ (i.e., $\min(X_i, Y_i) \geq 25\,000$) yields

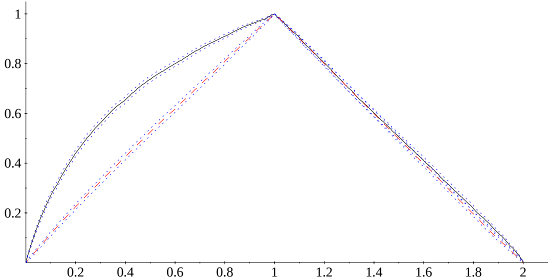
$\hat{\eta}_n \approx 0.58 \quad ([0.55, 0.62])$

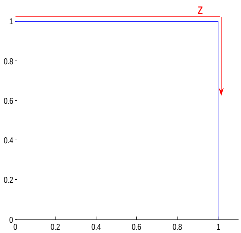
Difference to above estimate $\hat{\eta}_n \approx 0.713$ is statistically significant!

Also estimators for c show statistically significant differences...

Estimates of c

$$\hat{c}_n(x,y) := \frac{\sum_{i=1}^n \mathbb{1}_{\{\hat{U}_i < xk/n,\, \hat{V}_i < yk/n\}}}{\sum_{i=1}^n \mathbb{1}_{\{\hat{U}_i < k/n,\, \hat{V}_i < k/n\}}}$$





black: $\max(X_i, Y_i) \geq 25\,000$

red: $\min(X_i, Y_i) \geq 25\,000$

Asymptotic Normality of Deviation

If

- $$\sup_{(x,y):\max(x,y)=1} \left| \frac{P\{U < tx,\ V < ty\}}{P\{U < t,\ V < t\}} - c(x,y) \right| = O(q_1(t))$$
- asymptotic independence holds
- $m_n \to \infty$ such that $\sqrt{m_n} q_1\left(\frac{1}{F_T^{-1}(1-m_n/n)}\right) \to 0$
- c partially differentiable,

then for $k_n := \frac{n}{F_T^{-1}(1-m_n/n)}$ and $\tilde{\sigma}^2_{x,y,t} = \frac{t^{-1/\eta}-1}{c(x,y)} - \frac{\log^2 t}{\eta^2}$ one has

$$\sqrt{m_n}\left(\log \frac{\sum_{i=1}^n \mathbb{1}_{\{\hat{U}_i < tx_1 k_n/n,\ \hat{V}_i < ty_1 k_n/n\}}}{\sum_{i=1}^n \mathbb{1}_{\{\hat{U}_i < x_1 k_n/n,\ \hat{V}_i < y_1 k_n/n\}}} - \frac{1}{\hat{\eta}_n} \log t \right) \to \mathcal{N}(0, \tilde{\sigma}^2_{x,y,t})$$

back to confidence intervals

Idea of Proof

Proof is based on approximations of certain empirical processes.

In particular (Draisma et al. (2004)):

$$m_n^{1/2}\left(\frac{\sum_{i=1}^n \mathbb{1}\{\hat{U}_i \le \frac{\lfloor kx \rfloor}{n}, \hat{V}_i \le \frac{\lfloor ky \rfloor}{n}\}}{m_n} - c(x,y)\right) \longrightarrow W(x,y)$$

weakly in $D[0,\infty)^2$, where W is a centered Gaussian process with

$$Cov\big(W(x_1,y_1), W(x_2,y_2)\big) = c\big(\min(x_1,x_2), \min(y_1,y_2)\big)$$

under asymptotic independence.

back to confidence intervals

Idea of Proof

$$m_n(\hat{\eta}_n - \eta) = \int_0^1 \eta t^{-(\eta+1)} W(t^\eta, t^\eta)\,(t^\eta dt - \varepsilon_1(dt)) + o_P(1)$$

$\rightsquigarrow$

$$\begin{aligned}
&\sqrt{m_n}\left(\log \frac{\sum_{i=1}^n \mathbb{1}_{\{\hat{U}_i < tx_1 k/n,\, \hat{V}_i < ty_1 k/n\}}}{\sum_{i=1}^n \mathbb{1}_{\{\hat{U}_i < x_1 k/n,\, \hat{V}_i < y_1 k/n\}}} - \frac{1}{\hat{\eta}_n} \log t \right) \\
= \quad & \sqrt{m_n}\left(\frac{W(tx_1, ty_1)}{c(tx_1, ty_1)} - \frac{W(x_1, y_1)}{c(x_1, y_1)} + \right. \\
& \left. \frac{\log t}{\eta} \int_0^1 \eta t^{-(\eta+1)} W(t^\eta, t^\eta)\,(t^\eta dt - \varepsilon_1(dt)) \right) + o_P(1)
\end{aligned}$$

back to confidence intervals