# Local robust and asymptotically unbiased estimation of conditional Pareto-type tails

Goedele Dierckx
Hogeschool-Universiteit Brussel

Yuri Goegebeur
University of Southern Denmark

Armelle Guillou
Strasbourg University

# Topics

1. Introduction
   - Density power divergence
   - Pareto-type distributions
2. Estimation procedure
3. Asymptotic properties
4. Simulation results

# 1. Introduction: density power divergence

- **Basu, Harris, Hjort and Jones (1998):** density power divergence between density functions $f$ and $g$

$$\Delta_\alpha(f,g) := \begin{cases} \int_{\mathbb{R}} \left[g^{1+\alpha}(y) - \left(1+\frac{1}{\alpha}\right) g^\alpha(y) f(y) + \frac{1}{\alpha} f^{1+\alpha}(y)\right] dy, & \alpha > 0, \\ \int_{\mathbb{R}} \log \frac{f(y)}{g(y)} f(y) dy, & \alpha = 0. \end{cases}$$

- Assume that the density function $g$ depends on a parameter vector $\theta$

  Let $Y_1, \ldots, Y_n$ be a independent and identically distributed (i.i.d.) random variables according to density function $f$.

  The **minimum density power divergence (MDPD) estimator** is the value of $\theta$ minimizing the empirical density power divergence. For $\alpha > 0$:

$$\widehat{\Delta}_\alpha(\theta) := \int_{\mathbb{R}} g^{1+\alpha}(y) dy - \left(1 + \frac{1}{\alpha}\right) \frac{1}{n} \sum_{i=1}^{n} g^\alpha(Y_i),$$

and for $\alpha = 0$:

$$\widehat{\Delta}_0(\theta) := -\frac{1}{n}\sum_{i=1}^{n} \log g(Y_i),$$

Note: for $\alpha = 0$ the method corresponds to fitting the density function $g$ with the maximum likelihood method.

- The parameter $\alpha$ controls the **trade-off between efficiency and robustness** of the MDPD estimator:

  the estimator becomes more efficient but less robust against outliers as $\alpha$ gets closer to zero,

  whereas for increasing $\alpha$ the robustness increases and the efficiency decreases.

- We want to use the MDPD method to obtain a **robust nonparametric and asymptotically unbiased estimation method for Pareto-type distributions when there are random covariates.**

# 1. Introduction: Pareto-type distribution

- A distribution function $F$ is said to be of *Pareto-type* if for some $\gamma > 0$

$$1 - F(y) = y^{-\frac{1}{\gamma}} \ell(y), \qquad y > 0, \tag{1}$$

with $\ell$ a *slowly varying function at infinity* :

$$\frac{\ell(\lambda y)}{\ell(y)} \to 1 \text{ as } y \to \infty, \qquad \forall \lambda > 0.$$

- $\gamma$: extreme-value index **First order tail parameter**

- Example: strict Pareto, $F$, Burr, $|t|$, log-gamma, . . .

- Estimation of $\gamma$ has received a lot of attention. **Classical estimators are non-robust and typically show an asymptotic bias.** See Beirlant *et al.* (2004) and de Haan and Ferreira (2006)

- Robust estimation:

  Juárez and Schucany (2004), Kim and Lee (2008), Vandewalle and Beirlant (2007), Pend and Welsh (2001), Hubert *et al.* (2012), . . .

  Dierckx *et al.* (2013): fit the extended Pareto distribution with the MDPD technique

  – Robust and asymptotically unbiased
  – Asymptotic properties are available

- Extension to **regression case:** assume that together with $Y$ we observe a **random covariate** $X$

  $F(y;x)$: conditional distribution function of the response variable $Y$ given $X = x$,

  $b(x)$: density function of $X \in \mathbb{R}^p$.

  $F(y;x)$ is assumed to be of Pareto-type, i.e. there exists a positive function $\gamma(x)$ such that $\bar{F}(y;x) := 1 - F(y;x)$ is of the form

$$\bar{F}(y;x) = y^{-1/\gamma(x)}\ell(y;x), \qquad y > 0, \tag{2}$$

  $\gamma(x)$ describes the tail heaviness of $F(y;x)$ and has to be adequately estimated from the data.

  We use here a **nonparametric approach based on local estimation.**

  Local estimation: Daouia *et al.* (2011)

  Local asymptotically unbiased estimation: Goegebeur *et al.* (2013)

  **But:** these procedures are not robust!

## 2. Estimation procedure

The theoretical study of estimators for $\gamma(x)$ generally requires a **second order condition.**

**Condition** $(\mathcal{R})$. *Let* $\gamma(x) > 0$ *and* $\rho(x) < 0$ *be constants. The conditional distribution function* $F(y;x)$ *is such that* $y^{1/\gamma(x)}\bar{F}(y;x) \to C(x) \in (0,\infty)$ *as* $y \to \infty$ *and the function* $\delta(.;x)$ *defined via*

$$\bar{F}(y;x) = C(x)y^{-1/\gamma(x)}(1+\gamma(x)^{-1}\delta(y;x)),$$

*is ultimately nonzero, of constant sign and* $|\delta| \in RV_{\rho(x)/\gamma(x)}$*, i.e.*

$$\frac{\delta(ty;x)}{\delta(t;x)} \to y^{\rho(x)/\gamma(x)} \text{ as } t \to \infty,\ \forall y > 0.$$

Taking this second order structure into account during the estimation phase allows to obtain bias-corrected estimators.

Consider the **extended Pareto distribution** (Beirlant *et al.*, *2004*, Beirlant *et al.*, 2009), with distribution function given by

$$G(z;\gamma,\delta,\rho) = \begin{cases} 1-[z(1+\delta-\delta z^{\rho/\gamma})]^{-1/\gamma}, & z>1, \\ 0, & z\leq 1, \end{cases} \qquad (3)$$

and density function

$$g(z;\gamma,\delta,\rho) = \begin{cases} \frac{1}{\gamma}\, z^{-1/\gamma-1}[1+\delta(1-z^{\rho/\gamma})]^{-1/\gamma-1}[1+\delta(1-(1+\rho/\gamma)z^{\rho/\gamma})], & z>1, \\ 0, & z\leq 1, \end{cases}$$

where $\gamma > 0$, $\rho < 0$, and $\delta > \max\{-1, \gamma/\rho\}$.

For distribution functions satisfying $(\mathcal{R})$, one can **approximate** the conditional distribution function of $Z := Y/u$, given that $Y > u$, where $u$ denotes a threshold value, by the extended Pareto distribution:

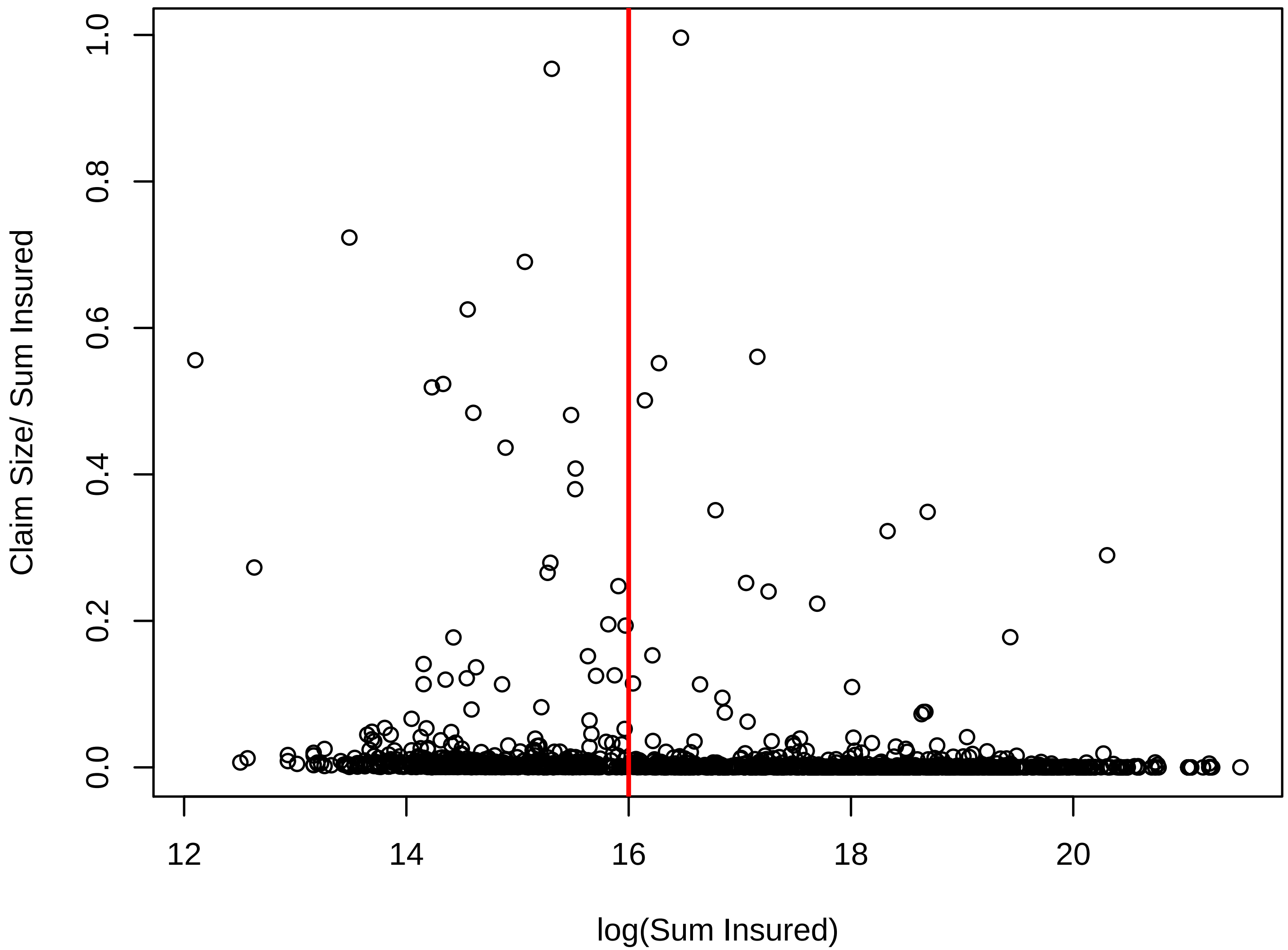
$$\frac{\bar{F}(uz;x)}{\bar{F}(u;x)} \approx \bar{G}(z;\gamma(x),\delta(u;x),\rho(x))$$

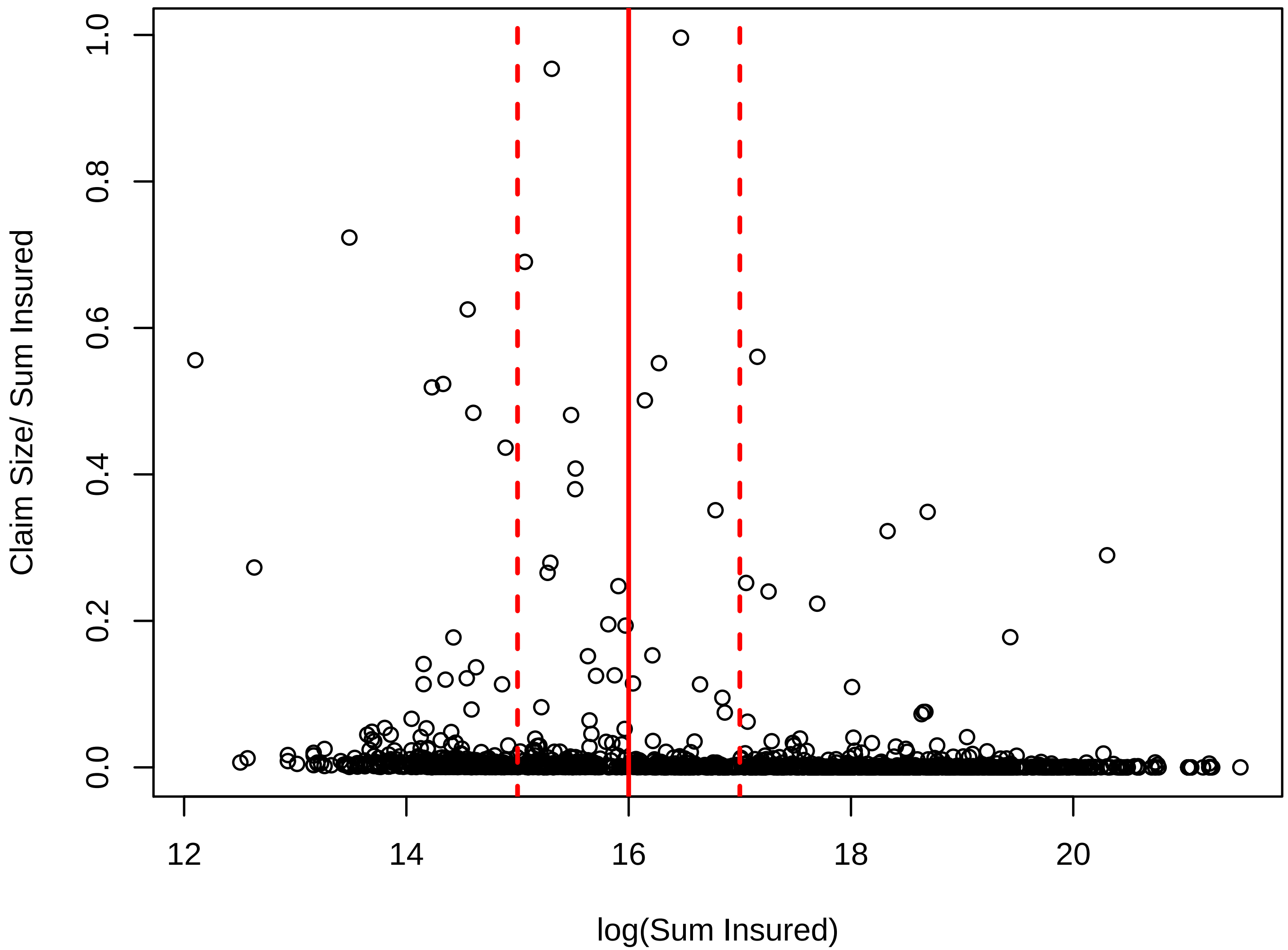
for large $u$.

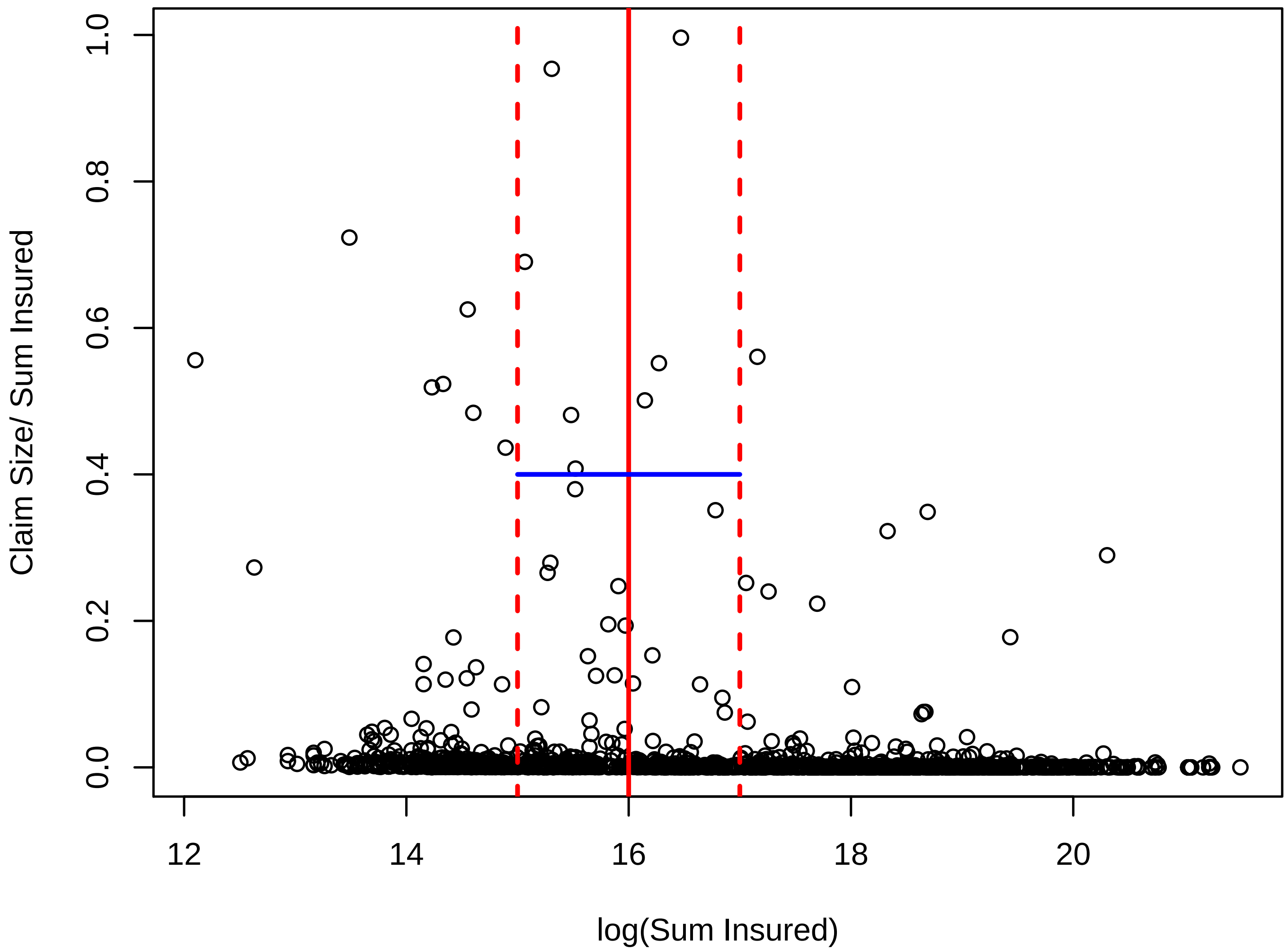
Formally, as shown in Beirlant *et al.* (2009), one has that

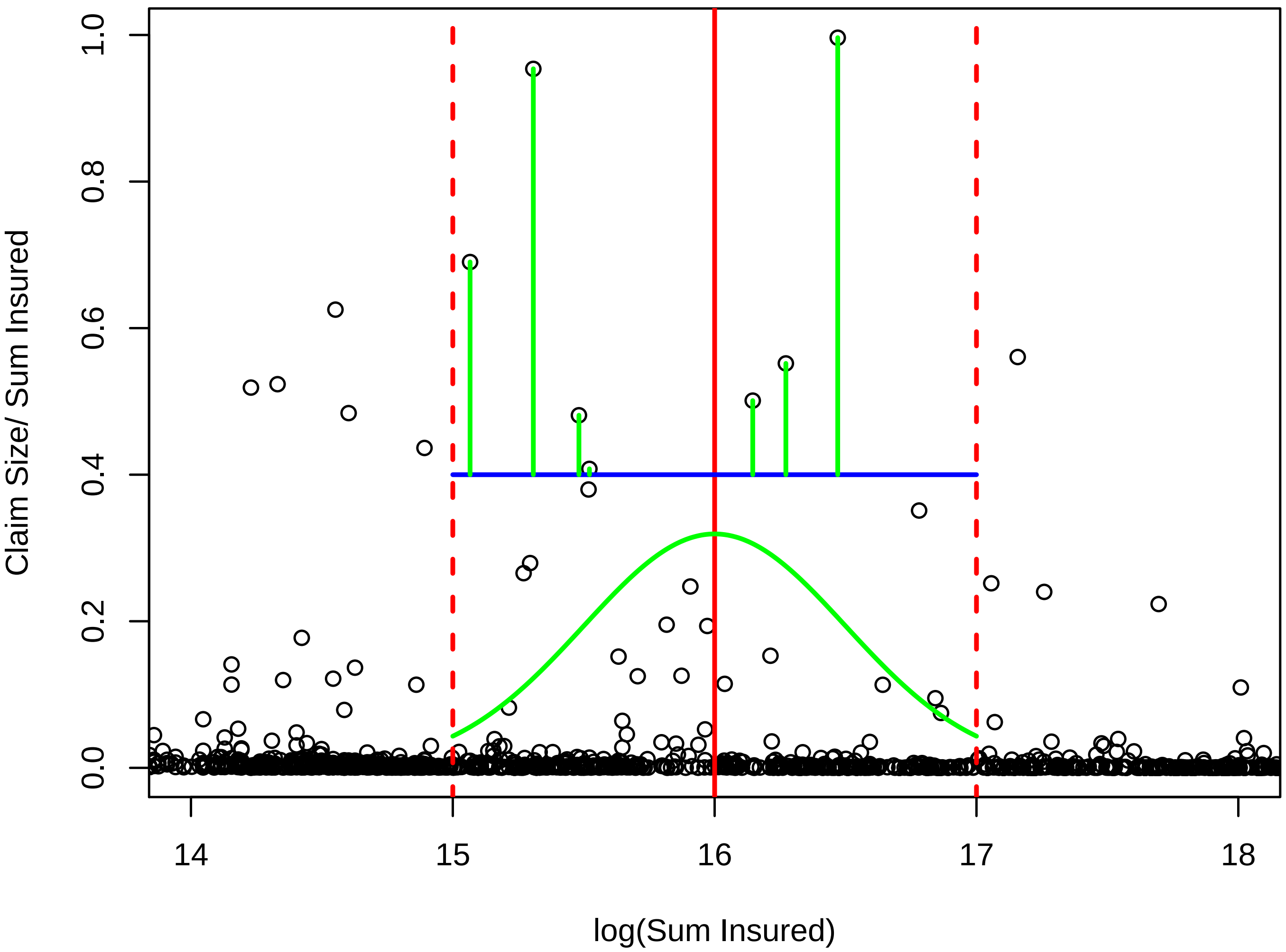
$$\sup_{z\geq 1}\left|\frac{\bar{F}(uz;x)}{\bar{F}(u;x)} - \bar{G}(z;\gamma(x),\delta(u;x),\rho(x))\right| = o(\delta(u;x)), \qquad \text{if } u \to \infty.$$

**Estimation of $\gamma(x)$:** Let $(X_i, Y_i)$, $i = 1, \ldots, n$, be independent realizations of the random vector $(X, Y) \in \mathbb{R}^p \times \mathbb{R}_{+,0}$, where $X$ has a distribution with joint density function $b$, and $\bar{F}(y;x)$ satisfies $(\mathcal{R})$.


Claim Size/ Sum Insured
0.0
0.2
0.4
0.6
0.8
1.0
12
14
16
18
20
log(Sum Insured)


Claim Size/ Sum Insured
0.0
0.2
0.4
0.6
0.8
1.0
12
14
16
18
20
log(Sum Insured)


0.0
0.2
0.4
0.6
0.8
1.0
Claim Size/ Sum Insured
12
14
16
18
20
log(Sum Insured)


Claim Size/ Sum Insured
0.0
0.2
0.4
0.6
0.8
1.0
14
15
16
17
18
log(Sum Insured)

Fit $g$ locally to the relative excesses $Z_i := Y_i/u_n$, $i = 1, \ldots, n$, by **MDPD, adjusted to locally weighted estimation,** i.e. we minimize

$$\widehat{\Delta}_\alpha(\gamma, \delta; \rho) :=$$
$$\frac{1}{n}\sum_{i=1}^{n} K_{h_n}(x - X_i)\left\{\int_1^\infty g^{1+\alpha}(z;\gamma,\delta,\rho)dz - \left(1+\frac{1}{\alpha}\right)g^\alpha(Z_i;\gamma,\delta,\rho)\right\}\mathbf{1}\{Y_i > u_n\},$$

in case $\alpha > 0$ and

$$\widehat{\Delta}_0(\gamma, \delta; \rho) \quad := \quad -\frac{1}{n}\sum_{i=1}^{n} K_{h_n}(x - X_i)\ln g(Z_i;\gamma,\delta,\rho)\mathbf{1}\{Y_i > u_n\},$$

in case $\alpha = 0$,

where

$K_{h_n}(x) := K(x/h_n)/h_n^p$, $K$ is a joint density function on $\mathbb{R}^p$,

$h_n$ is a positive non-random sequence of bandwidths with $h_n \to 0$ if $n \to \infty$,

$u_n$ is a local non-random threshold sequence satisfying $u_n \to \infty$ if $n \to \infty$.

The MDPD estimator for $(\gamma(x), \delta(u_n; x))$ satisfies the **estimating equations**

$$
\begin{aligned}
0 &= \frac{1}{n}\sum_{i=1}^{n} K_{h_n}(x - X_i)\mathbf{1}\{Y_i > u_n\}\int_1^{\infty} g^{\alpha}(z;\gamma,\delta,\rho)\frac{\partial g(z;\gamma,\delta,\rho)}{\partial\gamma}dz \\
&\quad -\frac{1}{n}\sum_{i=1}^{n} K_{h_n}(x - X_i)g^{\alpha-1}(Z_i;\gamma,\delta,\rho)\frac{\partial g(Z_i;\gamma,\delta,\rho)}{\partial\gamma}\mathbf{1}\{Y_i > u_n\}, \quad (4)
\end{aligned}
$$

$$
\begin{aligned}
0 &= \frac{1}{n}\sum_{i=1}^{n} K_{h_n}(x - X_i)\mathbf{1}\{Y_i > u_n\}\int_1^{\infty} g^{\alpha}(z;\gamma,\delta,\rho)\frac{\partial g(z;\gamma,\delta,\rho)}{\partial\delta}dz \\
&\quad -\frac{1}{n}\sum_{i=1}^{n} K_{h_n}(x - X_i)g^{\alpha-1}(Z_i;\gamma,\delta,\rho)\frac{\partial g(Z_i;\gamma,\delta,\rho)}{\partial\delta}\mathbf{1}\{Y_i > u_n\}. \quad (5)
\end{aligned}
$$

Note:

Only $\gamma(x)$ and $\delta(u_n; x)$ are estimated by the MDPD method.

The rate parameter $\rho(x)$ **will either be fixed or estimated externally in a consistent way.**

# 3. Asymptotic properties

For all $x_1, x_2 \in \mathbb{R}^p$, the Euclidean distance between $x_1$ and $x_2$ is denoted by $d(x_1, x_2)$.

**Assumption** $(\mathcal{B})$ *There exists* $c_b > 0$ *such that* $|b(x_1) - b(x_2)| \leq c_b d(x_1, x_2)$ *for all* $x_1$, $x_2 \in \mathbb{R}^p$.

**Assumption** $(\mathcal{K})$ $K$ *is a bounded density function on* $\mathbb{R}^p$, *with support* $\Omega$ *included in the unit hypersphere in* $\mathbb{R}^p$.

We also need to control the oscillation of $F(y;x)$ when considered as a function of its second argument.

Consider the conditional expectation

$$m(u_n, s, t; x) := \mathbb{E}\left[\left(\frac{Y}{u_n}\right)^s \left(\ln_+ \frac{Y}{u_n}\right)^t \mathbf{1}\{Y > u_n\} \middle| X = x\right],$$

with $s \leq 0$, $t \geq 0$.

**Assumption** $(\mathcal{M})$ *The function* $m(u_n, s, t; x)$ *satisfies that, for* $u_n \to \infty$, $h_n \to 0$, *and some* $S < 0$ *and* $T > 0$,

$$\Phi(u_n, h_n; x) := \sup_{(s,t)\in[S,0]\times[0,T]} \sup_{z\in\Omega} \left| \frac{m(u_n, s, t; x - h_n z)}{m(u_n, s, t; x)} - 1 \right| \to 0 \text{ if } n \to \infty.$$

- **Case 1: $\rho_0(x)$ known**

  **Theorem 1. (Existence and consistency)** *Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a sample of independent copies of the random vector $(X, Y)$ where $Y|X = x$ satisfies $(\mathcal{R})$, $X \sim b$, and assume $(\mathcal{B})$, $(\mathcal{K})$ and $(\mathcal{M})$ hold. For all $x \in \mathbb{R}^p$ where $b(x) > 0$, we have that if $h_n \to 0$, $u_n \to \infty$ with $nh_n^p \bar{F}(u_n; x) \to \infty$, then with probability tending to 1 there exists sequences of solutions $(\hat{\gamma}_n(x), \hat{\delta}_n(x))$ of the estimating equations (4) and (5), with $\rho$ fixed at $\rho_0(x)$, such that $(\hat{\gamma}_n(x), \hat{\delta}_n(x)) \xrightarrow{\mathbb{P}} (\gamma_0(x), 0)$, as $n \to \infty$.*

**Theorem 2. (Asymptotic normality)** *Let* $(X_1, Y_1), \ldots, (X_n, Y_n)$ *be a sample of independent copies of the random vector* $(X, Y)$ *where* $Y|X = x$ *satisfies* $(\mathcal{R})$, $X \sim b$, *and assume* $(\mathcal{B})$, $(\mathcal{K})$ *and* $(\mathcal{M})$ *hold.*

*Consider* $(\hat{\gamma}_n(x), \hat{\delta}_n(x))$, *a consistent sequence of estimators for* $(\gamma_0(x), 0)$ *satisfying (4) and (5), with* $\rho$ *fixed at* $\rho_0(x)$.

*For all* $x \in \mathbb{R}^p$ *where* $b(x) > 0$, *we have that if* $h_n \to 0$, $u_n \to \infty$ *with* $nh_n^p \bar{F}(u_n; x) \to \infty$, $\sqrt{nh_n^p \bar{F}(u_n; x)}\delta(u_n; x) \to \lambda \in \mathbb{R}$, $\sqrt{nh_n^p \bar{F}(u_n; x)}h_n \to 0$, *and* $\sqrt{nh_n^p \bar{F}(u_n; x)}\Phi(u_n, h_n; x) \to 0$, *then*

$$\sqrt{nh_n^p \bar{F}(u_n; x) b(x)} \begin{bmatrix} \hat{\gamma}_n(x) - \gamma_0(x) \\ \hat{\delta}_n(x) - \delta(u_n; x) \end{bmatrix}$$
$$\rightsquigarrow N_2(\mathbf{0}, \mathbb{C}^{-1}(\rho_0(x))\mathbb{B}(\rho_0(x))\boldsymbol{\Sigma}(\rho_0(x))\mathbb{B}'(\rho_0(x))\mathbb{C}^{-1}(\rho_0(x))).$$

$\rightarrow$ **Bias-corrected!**

- **Case 2: $\rho$ fixed at some value $\tilde{\rho}(x) < 0$**

**Proposition 1.** *Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a sample of independent copies of the random vector $(X, Y)$ where $Y|X = x$ satisfies $(\mathcal{R})$ and assume the parameter $\rho$ is fixed at $\tilde{\rho}(x)$ in (4) and (5). Suppose also that $X \sim b$, and assume $(\mathcal{B})$, $(\mathcal{M})$ and $(\mathcal{K})$ hold. For all $x \in \mathbb{R}^p$ where $b(x) > 0$, we have that if $h_n \to 0$, $u_n \to \infty$ with $nh_n^p \bar{F}(u_n; x) \to \infty$, when $n \to \infty$, then with probability tending to 1 there exists sequences of solutions $(\hat{\gamma}_n(x), \hat{\delta}_n(x))$ of the estimating equations (4) and (5) such that $(\hat{\gamma}_n(x), \hat{\delta}_n(x)) \xrightarrow{\mathbb{P}} (\gamma_0(x), 0)$. If additionally $\sqrt{nh_n^p \bar{F}(u_n; x)}\delta(u_n; x) \to \lambda \in \mathbb{R}$, $\sqrt{nh_n^p \bar{F}(u_n; x)}h_n \to 0$, and $\sqrt{nh_n^p \bar{F}(u_n; x)}\Phi(u_n, h_n; x) \to 0$, then*

$$
r_n \left[ \begin{array}{c} \hat{\gamma}_n(x) - \gamma_0(x) \\ \hat{\delta}_n(x) \end{array} \right] \rightsquigarrow N_2(-\lambda\sqrt{b(x)}\mathbb{C}^{-1}(\tilde{\rho}(x))\mathbb{B}(\tilde{\rho}(x))\tilde{\mathbb{D}},
$$

$$
\mathbb{C}^{-1}(\tilde{\rho}(x))\mathbb{B}(\tilde{\rho}(x))\mathbf{\Sigma}(\tilde{\rho}(x))\mathbb{B}'(\tilde{\rho}(x))\mathbb{C}^{-1}(\tilde{\rho}(x))),
$$

- **Case 3: $\rho_0(x)$ estimated externally in a consistent way.**

  **Theorem 3.** *The result of Theorem 1 and 2 continues to hold if $\rho$ is replaced by an external consistent estimator $\hat{\rho}_n(x)$ in (4) and (5).*

  E.g. use the consistent estimator for $\rho(x)$ from Goegebeur *et al.* (2013).

# 4. Simulation results

Estimators

- Non-robust local estimator

$$\hat{\gamma}_n^{(2)}(x,t,K,K) = \frac{1}{t+1}\frac{\sum_{i=1}^n K_{h_n}(x-X_i)(\ln Y_i - \ln u_n)_+^{t+1}\mathbf{1}\{Y_i > u_n\}}{\sum_{i=1}^n K_{h_n}(x-X_i)(\ln Y_i - \ln u_n)_+^{t}\mathbf{1}\{Y_i > u_n\}}$$

with $t = 0$.

Bias-corrected version

$$\hat{\gamma}_n^{(2)}(x,\beta) = \beta\hat{\gamma}_n^{(2)}(x,0,K,K) + (1-\beta)\hat{\gamma}_n^{(2)}(x,1,K,K)$$

with $\beta = -1$ and $\beta = 1/\hat{\rho}(x)$. See Goegebeur *et al.* (2013) for details

- Robust local MDPD estimators

  Estimator with $\delta = 0$ in $G$ (not bias-corrected)

  Bias-corrected: MDPD with $\gamma$ and $\delta$ estimated jointly

    $\rho(x)$ fixed at -1 and $\rho(x)$ estimated

All kernels are taken to be the bi-quadratic kernel function

$$K(x) = \frac{15}{16}(1 - x^2)^2 \mathbf{1}\{x \in [-1, 1]\}, \qquad x \in \mathbb{R}.$$

For practical implementation we have **two tuning parameters** that have to be determined, namely

- the bandwidth parameter $h_n$
- the threshold $u_n$

Tuning parameter selection methods:

- Oracle strategy: min MSE
- Heuristic, data driven method

# 4. Simulation results: uncontaminated case

We simulate $N = 100$ samples of size $n = 1\,000$, with $X \sim U(0,1)$ and $Y|X = x$ is generated from the following Burr distribution

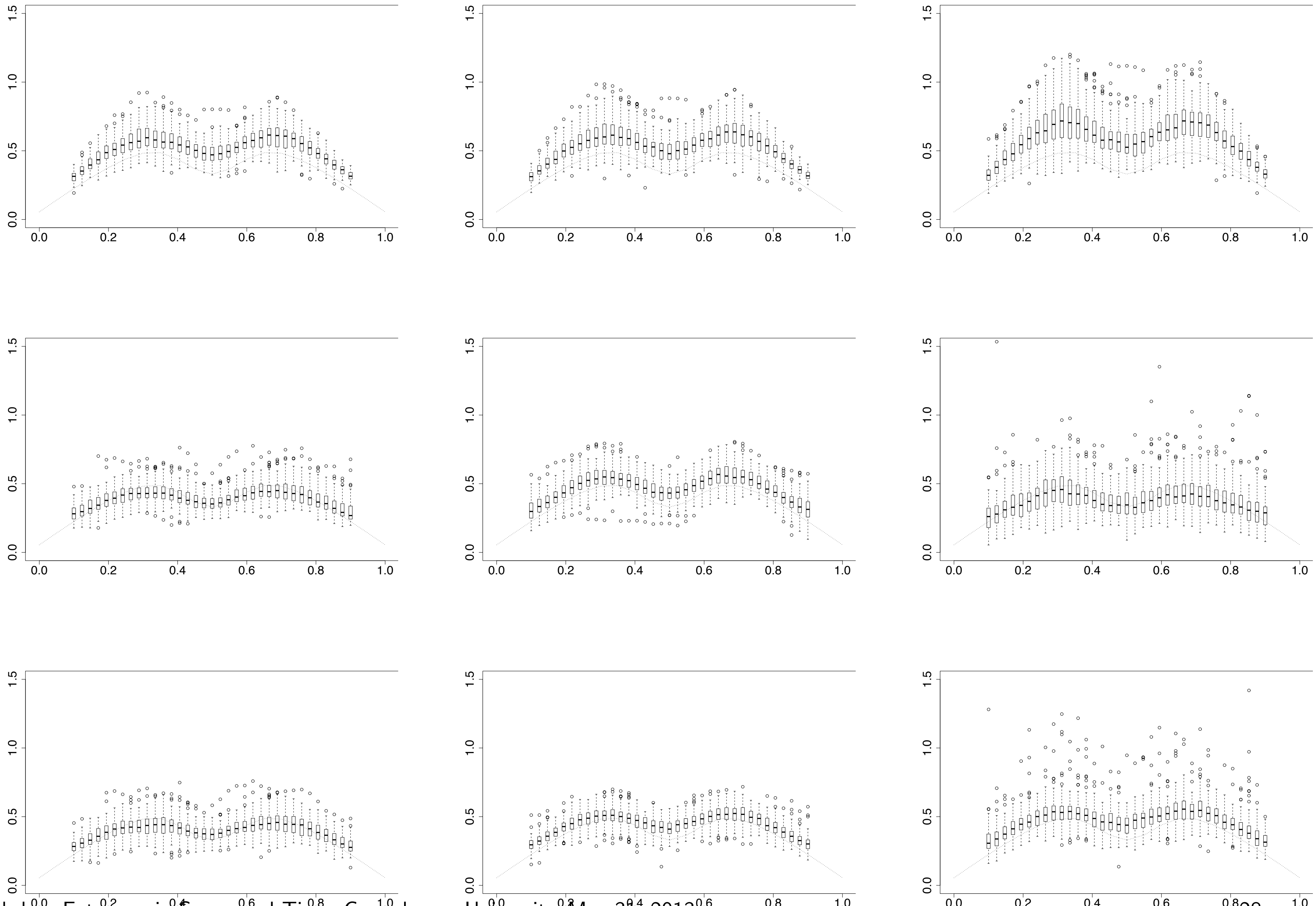
$$1 - F(y;x) = \left(1 + y^{-\rho(x)/\gamma(x)}\right)^{1/\rho(x)}, \qquad y > 0,$$

where

$$\gamma(x) = 0.5\,(0.1 + \sin(\pi x))\left(1.1 - 0.5\exp(-64(x-0.5)^2)\right) \text{ and } \rho(x) = -1.$$

## MSE of the different estimators

| Non Robust/Robust | Estimator | Oracle strategy | Data driven method |
|---|---|---|---|
| non robust | biased | 0.006 | 0.019 |
| non robust | bias-corrected $\rho(x) = -1$ | 0.003 | 0.006 |
| non robust | bias-corrected $\rho(x) = \hat{\rho}(x)$ | 0.007 | 0.007 |
| robust $\alpha = 0.1$ | biased | 0.006 | 0.025 |
| robust $\alpha = 0.1$ | bias-corrected $\rho(x) = -1$ | 0.007 | 0.011 |
| robust $\alpha = 0.1$ | bias-corrected $\rho(x) = \hat{\rho}(x)$ | 0.006 | 0.007 |
| robust $\alpha = 0.5$ | biased | 0.008 | 0.055 |
| robust $\alpha = 0.5$ | bias-corrected $\rho(x) = -1$ | 0.007 | 0.017 |
| robust $\alpha = 0.5$ | bias-corrected $\rho(x) = \hat{\rho}(x)$ | 0.007 | 0.019 |


0.0 0.5 1.0 1.5
0.0 0.2 0.4 0.6 0.8 1.0

# 4. Simulation results: Contaminated case 1

Burr distribution

$$1 - F(y;x) = \left(1 + y^{-\rho(x)/\gamma(x)}\right)^{1/\rho(x)}, \qquad y > 0,$$

where

$$\gamma(x) = 0.5\,(0.1 + \sin(\pi x))\left(1.1 - 0.5\exp(-64(x-0.5)^2)\right) \text{ and } \rho(x) = -1.$$

Contaminated distribution

$$F_\epsilon(y;x) = (1-\epsilon)F(y;x) + \epsilon\tilde{F}(y;x)$$

where $\tilde{F}(y;x) = 1 - \left(\frac{y}{x_c}\right)^{-0.5}, y > x_c$,

We set $\epsilon = 0.01$, $x_c$= 1.2 times the 99.99% quantile of $F(y;x)$

# MSE of the different estimators

| Non Robust/Robust | Estimator | Oracle strategy | Data driven method |
|---|---|---|---|
| non robust | biased | 0.053 | 0.069 |
| non robust | bias-corrected $\rho(x) = -1$ | 0.291 | 0.977 |
| non robust | bias-corrected $\rho(x) = \hat{\rho}(x)$ | 0.447 | 0.470 |
| robust $\alpha = 0.1$ | biased | 0.020 | 0.039 |
| robust $\alpha = 0.1$ | bias-corrected $\rho(x) = -1$ | 0.011 | 0.025 |
| robust $\alpha = 0.1$ | bias-corrected $\rho(x) = \hat{\rho}(x)$ | 0.014 | 0.023 |
| robust $\alpha = 0.5$ | biased | 0.012 | 0.060 |
| robust $\alpha = 0.5$ | bias-corrected $\rho(x) = -1$ | 0.007 | 0.009 |
| robust $\alpha = 0.5$ | bias-corrected $\rho(x) = \hat{\rho}(x)$ | 0.009 | 0.012 |