# Nonparametric estimation of risk measures of collective risks

Henryk Zähle
Saarland University

**LUH-Kolloquium "Versicherungs- und Finanzmathematik"**
**Hannover**
**November 5, 2015**

**Based on joint work with**

Volker Krätschmer (Essen), Alexandra Lauer (Saarbrücken), and Alexander Schied (Mannheim)

# Contents

# 1. Introduction

This talk will be concerned with statistical inference for law-invariant risk measures (premium principles).

In the Introduction, I will

- recall the definition of risk measures,
- give some examples for risk measures,
- present the basic statistical issue.

**Definition**

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be atomless and $\mathcal{X} \subseteq L^0 = L^0(\Omega, \mathcal{F}, \mathbb{P})$ be a vector space containing the constants. Let $\rho : \mathcal{X} \to \mathbb{R}$ be a map and consider the following conditions:

(1) monotonicity: $\rho(X_1) \leq \rho(X_2)$ for all $X_1, X_2 \in \mathcal{X}$ with $X_1 \leq X_2$.

(2) cash additivity: $\rho(X + m) = \rho(X) + m$ for all $X \in \mathcal{X}$ and $m \in \mathbb{R}$.

(3) subadditivity: $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$ for all $X_1, X_2 \in \mathcal{X}$.

(4) positive homogeneity: $\rho(\lambda X) = \lambda \rho(X)$ for all $X \in \mathcal{X}$ and $\lambda \geq 0$.

$\rho$ is a monetary risk measure if (1)–(2) hold.

$\rho$ is a coherent risk measure if (1)–(4) hold.

$\rho$ is law-invariant if $\rho(X_1) = \rho(X_2)$ whenever $\mathbb{P}_{X_1} = \mathbb{P}_{X_2}$.

**Example 1**

The Value at Risk at level $\alpha \in (0,1)$

$$\mathrm{V@R}_\alpha(X) := F_X^{\leftarrow}(\alpha) = \inf\{x \in \mathbb{R} : F_X(x) \geq \alpha\}$$

is a law-invariant and positively homogeneous monetary risk measure on $\mathcal{X} = L^0$. But it is not subadditive, hence not coherent.

## Example 2

The Average Value at Risk at level $\alpha \in (0,1)$

$$\mathrm{AV@R}_\alpha(X) := \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{V@R}_s(X)\,ds$$

is a law-invariant coherent risk measure on $\mathcal{X} = L^1$.
If $F_X$ is continuous at $\mathrm{V@R}_\alpha(X)$, then

$$\mathrm{AV@R}_\alpha(X) = \mathbb{E}[X | X \geq \mathrm{V@R}_\alpha(X)].$$

**Example 3**

The expectiles-based risk measure at level $\alpha \in [1/2, 1)$

$$\mathrm{Ept}_\alpha(X) \;:=\; \mathbb{U}_\alpha(X)^{-1}(0)$$

is a law-invariant coherent risk measure on $\mathcal{X} = L^1$.
Here we use the notation

$$\mathbb{U}_\alpha(X)(m) := \mathbb{E}[U_\alpha(X - m)], \qquad m \in \mathbb{R}$$

for

$$U_\alpha(x) := \begin{cases} \alpha\, x & , \quad x \geq 0 \\ (1-\alpha)\, x & , \quad x < 0 \end{cases} .$$

If $\alpha = 1/2$, then $\mathrm{Ept}_\alpha(X) = \mathbb{E}[X]$.

## Example 4

The one-sided moment-based risk measure for $p \in [1, \infty)$ and $a \in [0, 1]$

$$\mathrm{OsM}_{p,a}(X) := \mathbb{E}[X] + a\,\mathbb{E}\big[((X - \mathbb{E}[X])^+)^p\big]^{1/p}$$

is a law-invariant coherent risk measure on $\mathcal{X} = L^p$.

## Example 5

Let $g$ be a convex distortion function, i.e. a convex nondecreasing function $g : [0, 1] \to [0, 1]$ with $g(0) = 0$ and $g(1) = 1$. The distortion risk measure associated with $g$

$$\rho_g(X) \;:=\; -\int_{-\infty}^{0} g(F_X(x))\,dx + \int_{0}^{\infty} \big(1 - g(F_X(x))\big)\,dx$$

is a law-invariant coherent risk measure on $\mathcal{X} = \mathcal{X}_g := \{\cdots\}$.
For right-continuous $g$, we have the representations

$$\rho_g(X) \;=\; \int_0^1 \mathrm{V@R}_s(X)\,dg(s) \;=\; \int_{-\infty}^{\infty} x\; d(g \circ F_X)(x).$$

If specifically $g(t) = \max\{(t-\alpha)/(1-\alpha); 0\}$, then $\rho_g = \mathrm{AV@R}_\alpha$.
But there is no distor. function $g$ such that $\rho_g = \mathrm{Ept}_\alpha$ or $\rho_g = \mathrm{OsM}_{p,a}$.

## Example 5

Let $g$ be a convex distortion function, i.e. a convex nondecreasing function $g : [0,1] \to [0,1]$ with $g(0) = 0$ and $g(1) = 1$. The distortion risk measure associated with $g$

$$\rho_g(X) \;:=\; -\int_{-\infty}^{0} g(F_X(x))\,dx + \int_0^{\infty} \big(1 - g(F_X(x))\big)\,dx$$

is a law-invariant coherent risk measure on $\mathcal{X} = \mathcal{X}_g := \{\cdots\}$.
For right-continuous $g$, we have the representations

$$\rho_g(X) \;=\; \int_0^1 \mathrm{V@R}_s(X)\,dg(s) \;=\; \int_{-\infty}^{\infty} x\; d(g \circ F_X)(x).$$

Distortion risk measures associated with convex distortion functions are the building blocks of rather general law-invariant coherent risk measures (including $\mathrm{Ept}_\alpha$ and $\mathrm{OsM}_{p,a}$) . . .

**Theorem**

Let $\rho$ be a law-invariant coherent risk measure on $\mathcal{X} = L^p$ for some $p \in [1, \infty]$. Then there is some set $\mathcal{G}_\rho$ of continuous convex distortion functions such that

$$\rho(X) = \sup_{g \in \mathcal{G}_\rho} \rho_g(X) \qquad \text{for all } X \in \mathcal{X}$$

"Kusuoka representation".

Kusuoka (2001)
Krätschmer/H. Z. (2011)
Belomestny/Krätschmer (2012)

For every law-invariant $\rho : \mathcal{X} \to \mathbb{R}$ we may define a map

$$\mathcal{R}_\rho : \mathcal{M}(\mathcal{X}) \longrightarrow \mathbb{R} \qquad \text{by} \qquad \mathcal{R}_\rho(\mathfrak{m}) := \rho(X_\mathfrak{m}),$$

where $X_\mathfrak{m} \in \mathcal{X}$ has law $\mathfrak{m}$ and

$$\mathcal{M}(\mathcal{X}) := \{\mathbb{P}_X : X \in \mathcal{X}\}.$$

We call $\mathcal{R}_\rho$ risk functional associated with $\rho$.

## Statistical issue

We consider a "homogeneous" insurance collective

- $X_1, \dots, X_n \sim \mu$ i.i.d. individual claims in the next insurance period.
  $\sum_{i=1}^n X_i \sim \mu^{*n}$ total claim in the next insurance period.
  Individual claim distribution $\mu$ is unknown.

- $Y_1, \dots, Y_{u_n} \sim \mu$ i.i.d. indiv. claims in the previous insur. period(s),
  $u_n/n \sim c \in (0, \infty)$ (e.g. $u_n = n$)

and are interested in information on the individual premium

$$\frac{1}{n}\rho(\textstyle\sum_{i=1}^n X_i) = \frac{1}{n}\mathcal{R}_\rho(\mu^{*n}).$$

## Statistical issue

Let

$$\Omega := \mathbb{R}^{\mathbb{N}}, \qquad \mathcal{F} := \mathcal{B}(\mathbb{R})^{\otimes \mathbb{N}}, \qquad \mathbb{P}^{\mu} := \mu^{\otimes \mathbb{N}}$$

and note that

$$\big(\Omega, \mathcal{F}, \{\mathbb{P}^{\mu} : \mu \in \mathcal{M}(\mathcal{X})\}\big)$$

is the corresponding nonparametric statistical model. The observation variable $Y_i$ is defined to be the $i$-th coordinate projection on $\Omega = \mathbb{R}^{\mathbb{N}}$.

We are interested in the following aspect of the parameter $\mu$:

$$T_n(\mu) := \frac{1}{n}\mathcal{R}_{\rho}(\mu^{*n}).$$

# 2. Choice of estimators

**Approach**

- choose reasonable estimator $\widehat{\mu_{u_n}^{*n}}$ for $\mu^{*n}$ based on $Y_1, \ldots, Y_{u_n}$
- use $\widehat{T}_n := \frac{1}{n}\mathcal{R}_\rho(\widehat{\mu_{u_n}^{*n}})$ as estimator for $T_n(\mu) := \frac{1}{n}\mathcal{R}_\rho(\mu^{*n})$

**Examples**

The Central Limit Theorem and Glivenko–Cantelli suggest respectively

$$\widehat{\mu_{u_n}^{*n}} := \mathcal{N}_{n\widehat{m}_{u_n}, n\widehat{s}_{u_n}^2} \qquad \text{and} \qquad \widehat{\mu_{u_n}^{*n}} := \widehat{\mu}_{u_n}^{*n},$$

where

$$\begin{aligned} \widehat{m}_{u_n} &= \text{empirical mean of } Y_1, \ldots, Y_{u_n} \\ \widehat{s}_{u_n} &= \text{empirical standard deviation of } Y_1, \ldots, Y_{u_n} \\ \widehat{\mu}_{u_n} &= \text{empirical probability measure of } Y_1, \ldots, Y_{u_n} = \frac{1}{u_n}\sum_{i=1}^{u_n} \delta_{Y_i}. \end{aligned}$$

# 2. Choice of estimators

**Approach**

- choose reasonable estimator $\widehat{\mu_{u_n}^{*n}}$ for $\mu^{*n}$ based on $Y_1, \ldots, Y_{u_n}$
- use $\widehat{T}_n := \frac{1}{n}\mathcal{R}_\rho(\widehat{\mu_{u_n}^{*n}})$ as estimator for $T_n(\mu) := \frac{1}{n}\mathcal{R}_\rho(\mu^{*n})$

**Note**

The normal approximation with estimated parameters

$$\widehat{\mu_{u_n}^{*n}} := \mathcal{N}_{n\widehat{m}_{u_n}, n\widehat{s}^2_{u_n}}$$

is easy to compute. However the total claim distribution $\mu^{*n}$ is typically skewed to the right, whereas the normal distribution is symmetric.

# 2. Choice of estimators

**Approach**

- choose reasonable estimator $\widehat{\mu_{u_n}^{*n}}$ for $\mu^{*n}$ based on $Y_1, \ldots, Y_{u_n}$
- use $\widehat{T}_n := \frac{1}{n}\mathcal{R}_\rho(\widehat{\mu_{u_n}^{*n}})$ as estimator for $T_n(\mu) := \frac{1}{n}\mathcal{R}_\rho(\mu^{*n})$

**Note**

The normal approximation with estimated parameters

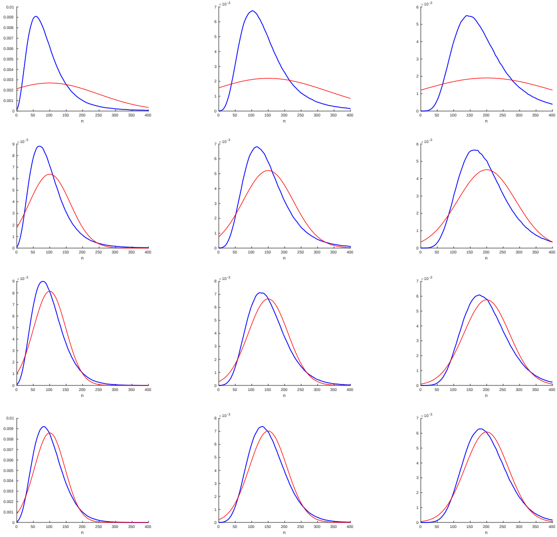
$$\widehat{\mu_{u_n}^{*n}} := \mathcal{N}_{n\widehat{m}_{u_n}, n\widehat{s}^2_{u_n}}$$

is easy to compute. However the total claim distribution $\mu^{*n}$ is typically skewed to the right, whereas the normal distribution is symmetric.

*For instance,* $\mu = (1-p)\delta_0 + p\mathrm{P}_{a,b}$ *for* $p = 0.1$ *and* $\mathrm{P}_{a,b} = f_{a,b}\ell$ *Pareto*

$$f_{a,b}(x) := ab^{-1}\big(b^{-1}x + 1\big)^{-(a+1)}\,\mathbb{1}_{(0,\infty)}(x) \qquad (a, b > 0).$$

$n = 100$ $n = 150$ $n = 200$



$a$=2.1, $b$=11

$a$=3, $b$=20

$a$=6, $b$=50

$a$=10, $b$=90

true convolution $\mu^{*n}$ normal approximation $\mathcal{N}_{nm,ns^2}$

**Approach**

- choose reasonable estimator $\widehat{\mu_{u_n}^{*n}}$ for $\mu^{*n}$ based on $Y_1, \ldots, Y_{u_n}$
- use $\widehat{T}_n := \frac{1}{n}\mathcal{R}_\rho(\widehat{\mu_{u_n}^{*n}})$ as estimator for $T_n(\mu) := \frac{1}{n}\mathcal{R}_\rho(\mu^{*n})$

**Note**

The computation of the convolution

$$\widehat{\mu_{u_n}^{*n}} := \widehat{\mu}_{u_n}^{*n}$$

is more time-consuming (use, for instance, the Panjer recursion).
On the other hand, it takes into account the skewness of $\mu^{*n}$.

## Questions

- Consistency:

$$\left(\frac{1}{n}\mathcal{R}_\rho(\widehat{\mu^{*n}_{u_n}}) - \frac{1}{n}\mathcal{R}_\rho(\mu^{*n})\right) \xrightarrow{\text{a.s.}} 0$$

- Asymptotic normality:

$$\sqrt{u_n}\left(\frac{1}{n}\mathcal{R}_\rho(\widehat{\mu^{*n}_{u_n}}) - \frac{1}{n}\mathcal{R}_\rho(\mu^{*n})\right) \xrightarrow{\text{d}} Z \sim \mathcal{N}_{0,\sigma^2(\mu)}$$

- Qualitative robustness:

For every $\varepsilon > 0$ there exist a $\delta > 0$ such that for all $n \in \mathbb{N}$

$$d_{\text{weak}}(\mu,\nu) \leq \delta \quad \Longrightarrow \quad d_{\text{Lévy}}\left(\mathbb{P}^{\mu}_{\frac{1}{n}\mathcal{R}_\rho(\widehat{\mu^{*n}_{u_n}})}, \mathbb{P}^{\nu}_{\frac{1}{n}\mathcal{R}_\rho(\widehat{\mu^{*n}_{u_n}})}\right) \leq \varepsilon$$

## Questions

- Consistency:

$$n^r\left(\frac{1}{n}\mathcal{R}_\rho(\widehat{\mu^{*n}_{u_n}}) - \frac{1}{n}\mathcal{R}_\rho(\mu^{*n})\right) \xrightarrow{\text{a.s.}} 0 \qquad \text{for } r < 1/2$$

- Asymptotic normality:

$$\sqrt{u_n}\left(\frac{1}{n}\mathcal{R}_\rho(\widehat{\mu^{*n}_{u_n}}) - \frac{1}{n}\mathcal{R}_\rho(\mu^{*n})\right) \xrightarrow{\text{d}} Z \sim \mathcal{N}_{0,\sigma^2(\mu)}$$

- Qualitative robustness:

For every $\varepsilon > 0$ there exist a $\delta > 0$ such that for all $n \in \mathbb{N}$

$$d_{\text{weak}}(\mu,\nu) \leq \delta \quad \Longrightarrow \quad d_{\text{Lévy}}\left(\mathbb{P}^{\mu}_{\frac{1}{n}\mathcal{R}_\rho(\widehat{\mu^{*n}_{u_n}})}, \mathbb{P}^{\nu}_{\frac{1}{n}\mathcal{R}_\rho(\widehat{\mu^{*n}_{u_n}})}\right) \leq \varepsilon$$

# 3. Consistency, asymptotic normality

**Basic assumption**

Let $\rho : \mathcal{X} \to \mathbb{R}$ be a law-invariant map, $\lambda > 2$, and assume:

(a) $u_n/n \longrightarrow c \in (0, \infty)$.

(b) $\rho$ is cash additive and positively homogeneous, and $\mathcal{M}_1^\lambda \subseteq \mathcal{M}(\mathcal{X})$.

(c) $\mu \in \mathcal{M}(L^\lambda)$.

(d) For each sequence $(\mathfrak{m}_n) \subset \mathcal{M}_1^\lambda$ with $d_\lambda(\mathfrak{m}_n, \mathcal{N}_{0,1}) \longrightarrow 0$ there exist constants $C, \beta > 0$ such that for all $n \in \mathbb{N}$

$$|\mathcal{R}_\rho(\mathfrak{m}_n) - \mathcal{R}_\rho(\mathcal{N}_{0,1})| \leq C\, d_\lambda(\mathfrak{m}_n, \mathcal{N}_{0,1})^\beta.$$

$d_\lambda(\mu_1, \mu_2) := \sup_{x \in \mathbb{R}} |F_{\mu_1}(x) - F_{\mu_2}(x)|(1 + |x|^\lambda)$

$\mathcal{M}_1^\lambda := \{\mu \in \mathcal{M}_1 : d_\lambda(\mu, \delta_0) < \infty\} \quad (\subset \mathcal{M}(L^p)$ for any $p < \lambda)$.

# 3. Consistency, asymptotic normality

**Basic assumption**

Let $\rho : \mathcal{X} \to \mathbb{R}$ be a law-invariant map, $\lambda > 2$, and assume:

(a) $u_n/n \longrightarrow c \in (0, \infty)$.

(b) $\rho$ is cash additive and positively homogeneous, and $\mathcal{M}_1^\lambda \subseteq \mathcal{M}(\mathcal{X})$.

(c) $\mu \in \mathcal{M}(L^\lambda)$.

(d) For each sequence $(\mathfrak{m}_n) \subset \mathcal{M}_1^\lambda$ with $d_\lambda(\mathfrak{m}_n, \mathcal{N}_{0,1}) \longrightarrow 0$ there exist constants $C, \beta > 0$ such that for all $n \in \mathbb{N}$

$$|\mathcal{R}_\rho(\mathfrak{m}_n) - \mathcal{R}_\rho(\mathcal{N}_{0,1})| \leq C\, d_\lambda(\mathfrak{m}_n, \mathcal{N}_{0,1})^\beta.$$

The number $u_n$ of observed individual claims is of the same "dimension" as the number $n$ of individual risks in the collective. In other words:

Claims could be observed only in the last **few** years.

# 3. Consistency, asymptotic normality

**Basic assumption**

Let $\rho : \mathcal{X} \to \mathbb{R}$ be a law-invariant map, $\lambda > 2$, and assume:

(a) $u_n/n \longrightarrow c \in (0, \infty)$.

(b) $\rho$ is cash additive and positively homogeneous, and $\mathcal{M}_1^\lambda \subseteq \mathcal{M}(\mathcal{X})$.

(c) $\mu \in \mathcal{M}(L^\lambda)$.

(d) For each sequence $(\mathfrak{m}_n) \subset \mathcal{M}_1^\lambda$ with $d_\lambda(\mathfrak{m}_n, \mathcal{N}_{0,1}) \longrightarrow 0$ there exist constants $C, \beta > 0$ such that for all $n \in \mathbb{N}$

$$|\mathcal{R}_\rho(\mathfrak{m}_n) - \mathcal{R}_\rho(\mathcal{N}_{0,1})| \leq C\, d_\lambda(\mathfrak{m}_n, \mathcal{N}_{0,1})^\beta.$$

For instance, $\rho = \mathrm{V@R}_\alpha$, $\mathrm{AV@R}_\alpha$, $\mathrm{Ept}_\alpha$, $\mathrm{OsM}_p$, $\rho_g$, ...

# 3. Consistency, asymptotic normality

**Basic assumption**

Let $\rho : \mathcal{X} \to \mathbb{R}$ be a law-invariant map, $\lambda > 2$, and assume:

(a) $u_n/n \longrightarrow c \in (0, \infty)$.

(b) $\rho$ is cash additive and positively homogeneous, and $\mathcal{M}_1^\lambda \subseteq \mathcal{M}(\mathcal{X})$.

(c) $\mu \in \mathcal{M}(L^\lambda)$.

(d) For each sequence $(\mathfrak{m}_n) \subset \mathcal{M}_1^\lambda$ with $d_\lambda(\mathfrak{m}_n, \mathcal{N}_{0,1}) \longrightarrow 0$ there exist constants $C, \beta > 0$ such that for all $n \in \mathbb{N}$

$$|\mathcal{R}_\rho(\mathfrak{m}_n) - \mathcal{R}_\rho(\mathcal{N}_{0,1})| \leq C\, d_\lambda(\mathfrak{m}_n, \mathcal{N}_{0,1})^\beta.$$

For instance, $\mathcal{X} = L^p$ and
the individual claims $X_i \sim \mu$ lie in $L^\lambda$ for some $\lambda > p$.

# 3. Consistency, asymptotic normality

### Basic assumption

Let $\rho : \mathcal{X} \to \mathbb{R}$ be a law-invariant map, $\lambda > 2$, and assume:

(a) $u_n/n \longrightarrow c \in (0, \infty)$.

(b) $\rho$ is cash additive and positively homogeneous, and $\mathcal{M}_1^\lambda \subseteq \mathcal{M}(\mathcal{X})$.

(c) $\mu \in \mathcal{M}(L^\lambda)$.

(d) For each sequence $(\mathfrak{m}_n) \subset \mathcal{M}_1^\lambda$ with $d_\lambda(\mathfrak{m}_n, \mathcal{N}_{0,1}) \longrightarrow 0$ there exist constants $C, \beta > 0$ such that for all $n \in \mathbb{N}$

$$|\mathcal{R}_\rho(\mathfrak{m}_n) - \mathcal{R}_\rho(\mathcal{N}_{0,1})| \leq C\, d_\lambda(\mathfrak{m}_n, \mathcal{N}_{0,1})^\beta.$$

For instance, $\rho = \mathrm{V@R}_\alpha$ $(\lambda = 0,\ \beta = 1)$, $\mathrm{AV@R}_\alpha$ $(\lambda > 1,\ \beta = 1)$, $\mathrm{Ept}_\alpha$ $(\lambda > 1,\ \beta = 1)$, $\mathrm{OsM}_p$ $(\lambda > p,\ \beta = 1/p)$, ...

# 3. Consistency, asymptotic normality

**Basic assumption**

Let $\rho : \mathcal{X} \to \mathbb{R}$ be a law-invariant map, $\lambda > 2$, and assume:

(a) $u_n/n \longrightarrow c \in (0, \infty)$.

(b) $\rho$ is cash additive and positively homogeneous, and $\mathcal{M}_1^\lambda \subseteq \mathcal{M}(\mathcal{X})$.

(c) $\mu \in \mathcal{M}(L^\lambda)$.

(d) For each sequence $(\mathfrak{m}_n) \subset \mathcal{M}_1^\lambda$ with $d_\lambda(\mathfrak{m}_n, \mathcal{N}_{0,1}) \longrightarrow 0$ there exist constants $C, \beta > 0$ such that for all $n \in \mathbb{N}$

$$|\mathcal{R}_\rho(\mathfrak{m}_n) - \mathcal{R}_\rho(\mathcal{N}_{0,1})| \leq C\, d_\lambda(\mathfrak{m}_n, \mathcal{N}_{0,1})^\beta.$$

**. . . is not very restrictive!**

**Theorem**

Under the basic assumption we have

(0) $\frac{1}{n}\mathcal{R}_\rho(\mu^{*n}) \;=\; m + \left\{\frac{1}{\sqrt{n}}\mathcal{R}_\rho(\mathcal{N}_{0,1})\right\} s + \mathcal{O}(n^{-1/2-\gamma})$

(i) $\frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_{u_n},\, n\widehat{s}^2_{u_n}}) \;=\; \widehat{m}_{u_n} + \left\{\frac{1}{\sqrt{n}}\mathcal{R}_\rho(\mathcal{N}_{0,1})\right\} \widehat{s}_{u_n}$

(ii) $\frac{1}{n}\,\mathcal{R}_\rho(\widehat{\mu}^{*n}_{u_n}) \;=\; \widehat{m}_{u_n} + \left\{\frac{1}{\sqrt{n}}\mathcal{R}_\rho(\mathcal{N}_{0,1})\right\} \widehat{s}_{u_n} + \mathcal{O}_{\mathbb{P}\text{-a.s.}}(n^{-1/2-\gamma})$

$\gamma := \min\{\lambda - 2, 1\}/2$

$m := \mathrm{mean}(\mu), \quad s := \mathrm{std}(\mu), \quad \widehat{m}_{u_n} := \frac{1}{u_n}\sum_{i=1}^{u_n} Y_i, \quad \widehat{s}_{u_n} := \ldots$

Krätschmer/H. Z. (2011)
Lauer/H. Z. (2015)

**Theorem**

Under the basic assumption we have

(0) $\frac{1}{n}\mathcal{R}_\rho(\mu^{*n}) \;=\; m + \left\{\frac{1}{\sqrt{n}}\mathcal{R}_\rho(\mathcal{N}_{0,1})\right\} s + \mathcal{O}(n^{-1/2-\gamma})$

(i) $\frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_{u_n},\, n\widehat{s}^2_{u_n}}) \;=\; \widehat{m}_{u_n} + \left\{\frac{1}{\sqrt{n}}\mathcal{R}_\rho(\mathcal{N}_{0,1})\right\} \widehat{s}_{u_n}$

(ii) $\frac{1}{n}\mathcal{R}_\rho(\widehat{\mu}^{*n}_{u_n}) \;=\; \widehat{m}_{u_n} + \left\{\frac{1}{\sqrt{n}}\mathcal{R}_\rho(\mathcal{N}_{0,1})\right\} \widehat{s}_{u_n} + \mathcal{O}_{\mathbb{P}\text{-a.s.}}(n^{-1/2-\gamma})$

**Note**

This shows that the premium determined according to $\rho$ is asymptotically equivalent to the premium determined according to the standard deviation principle with safety loading

$$\frac{1}{\sqrt{n}}\mathcal{R}_\rho(\mathcal{N}_{0,1}).$$

The factor $\frac{1}{\sqrt{n}}$ reflects the balancing of risks in a collective of size $n$.

**Proof**

For instance, the first identity follows from

$$\begin{aligned}\mathcal{R}_\rho(\mu^{*n}) &= \mathcal{R}_\rho(\mathcal{N}_{nm,\,ns^2}) + \big(\mathcal{R}_\rho(\mu^{*n}) - \mathcal{R}_\rho(\mathcal{N}_{nm,\,ns^2})\big)\\ &= \rho(nm + \sqrt{n}sZ) + \big(\rho(\sqrt{n}sZ_n + nm) - \rho(\sqrt{n}sZ + nm)\big)\\ &= nm + \sqrt{n}s\rho(Z) + \sqrt{n}s\big(\rho(Z_n) - \rho(Z)\big)\\ &= nm + \sqrt{n}s\mathcal{R}_\rho(\mathcal{N}_{0,1}) + \sqrt{n}s\big(\mathcal{R}_\rho(\mathrm{law}\{Z_n\}) - \mathcal{R}_\rho(\mathcal{N}_{0,1})\big)\end{aligned}$$

$\big($with $Z_n := \frac{1}{\sqrt{n}s}\sum_{i=1}^n (X_i - m)$ and $Z \sim \mathcal{N}_{0,1}\big)$ and

$$\begin{aligned}&\sqrt{n}s\,\big|\mathcal{R}_\rho(\mathrm{law}\{Z_n\}) - \mathcal{R}_\rho(\mathcal{N}_{0,1})\big|\\ &\quad\le \sqrt{n}s \cdot \mathsf{const}_\rho \cdot \sup_{x\in\mathbb{R}} |F_{Z_n}(x) - \Phi_{0,1}(x)|\,(1+|x|^\lambda)\\ &\quad\le \sqrt{n}s \cdot \mathsf{const}_\rho \cdot \mathsf{const}_\lambda \cdot n^{-\gamma}.\end{aligned}$$

The last step is ensured by Petrov's nonuniform Berry–Esséen inequality.

## Corollary

Under the basic assumption we have

(0) $\frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{nm,ns^2}) - \frac{1}{n}\mathcal{R}_\rho(\mu^{*n}) = \mathcal{O}(n^{-1/2-\gamma})$.

(i) $\frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_{u_n},n\widehat{s}^2_{u_n}}) - \frac{1}{n}\mathcal{R}_\rho(\mu^{*n}) \;=\; (\widehat{m}_{u_n} - m) + o_{\mathbb{P}\text{-a.s.}}(n^{-1/2})$.

(ii) $\frac{1}{n}\mathcal{R}_\rho(\widehat{\mu}^{*n}_{u_n}) - \frac{1}{n}\mathcal{R}_\rho(\mu^{*n}) \;=\; (\widehat{m}_{u_n} - m) + o_{\mathbb{P}\text{-a.s.}}(n^{-1/2})$.

$\gamma := \min\{\lambda - 2, 1\}/2$

$m := \mathrm{mean}(\mu), \quad s := \mathrm{std}(\mu), \quad \widehat{m}_{u_n} := \frac{1}{u_n}\sum_{i=1}^{u_n} Y_i, \quad \widehat{s}_{u_n} := \ldots$

Krätschmer/H. Z. (2011)
Lauer/H. Z. (2015)

## Corollary

Under the basic assumption we have

(0) $\frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{nm,ns^2}) - \frac{1}{n}\mathcal{R}_\rho(\mu^{*n}) = \mathcal{O}(n^{-1/2-\gamma})$.

(i) $\frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_{u_n},n\widehat{s}^2_{u_n}}) - \frac{1}{n}\mathcal{R}_\rho(\mu^{*n}) = (\widehat{m}_{u_n} - m) + o_{\mathbb{P}\text{-a.s.}}(n^{-1/2})$.

(ii) $\frac{1}{n}\mathcal{R}_\rho(\widehat{\mu}^{*n}_{u_n}) - \frac{1}{n}\mathcal{R}_\rho(\mu^{*n}) = (\widehat{m}_{u_n} - m) + o_{\mathbb{P}\text{-a.s.}}(n^{-1/2})$.

## Note

$$
\begin{aligned}
& n^r\big((\widehat{m}_{u_n} - m) + o_{\mathbb{P}\text{-a.s.}}(n^{-1/2})\big) \\
&= \frac{u_n^r}{n^r} \cdot u_n^r(\widehat{m}_{u_n} - m) + \frac{o_{\mathbb{P}\text{-a.s.}}(n^{-1/2})}{n^{-r}} \\
&\xrightarrow{\text{a.s.}} 0 \qquad \text{for all } r < 1/2
\end{aligned}
$$

**Corollary**

Under the basic assumption we have

(0) $\frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{nm,ns^2}) - \frac{1}{n}\mathcal{R}_\rho(\mu^{*n}) = \mathcal{O}(n^{-1/2-\gamma})$.

(i) $\frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_{u_n},n\widehat{s}^2_{u_n}}) - \frac{1}{n}\mathcal{R}_\rho(\mu^{*n}) \;=\; (\widehat{m}_{u_n} - m) + o_{\mathbb{P}\text{-a.s.}}(n^{-1/2})$.

(ii) $\frac{1}{n}\mathcal{R}_\rho(\widehat{\mu}^{*n}_{u_n}) - \frac{1}{n}\mathcal{R}_\rho(\mu^{*n}) \;=\; (\widehat{m}_{u_n} - m) + o_{\mathbb{P}\text{-a.s.}}(n^{-1/2})$.

**Note**

$$\begin{aligned}
&\sqrt{u_n}\big((\widehat{m}_{u_n} - m) \;+\; o_{\mathbb{P}\text{-a.s.}}(n^{-1/2})\big) \\
&= \quad \sqrt{u_n}(\widehat{m}_{u_n} - m) \;+\; \frac{o_{\mathbb{P}\text{-a.s.}}(n^{-1/2})}{n^{-1/2}} \\
&\xrightarrow{\text{d}} \quad Z \;\sim\; \mathcal{N}_{0,s^2}
\end{aligned}$$

**Corollary**

Under the basic assumption we have

(0) $\frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{nm,ns^2}) - \frac{1}{n}\mathcal{R}_\rho(\mu^{*n}) = \mathcal{O}(n^{-1/2-\gamma})$.

(i) $\frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_{u_n},n\widehat{s}^2_{u_n}}) - \frac{1}{n}\mathcal{R}_\rho(\mu^{*n}) \;=\; (\widehat{m}_{u_n} - m) + o_{\mathbb{P}\text{-a.s.}}(n^{-1/2})$.

(ii) $\frac{1}{n}\mathcal{R}_\rho(\widehat{\mu}^{*n}_{u_n}) - \frac{1}{n}\mathcal{R}_\rho(\mu^{*n}) \;=\; (\widehat{m}_{u_n} - m) + o_{\mathbb{P}\text{-a.s.}}(n^{-1/2})$.

**Note**

. . . in particular,

$$\left[\frac{1}{n}\mathcal{R}_\rho\big(\widehat{\mu^{*n}_{u_n}}\big) - \frac{\widehat{s}_{u_n}}{\sqrt{u_n}}\,\Phi^{-1}\Big(1-\frac{\alpha}{2}\Big)\,,\;\frac{1}{n}\mathcal{R}_\rho\big(\widehat{\mu^{*n}_{u_n}}\big) - \frac{\widehat{s}_{u_n}}{\sqrt{u_n}}\,\Phi^{-1}\Big(\frac{\alpha}{2}\Big)\right]$$

provides an asymptotic confidence interval for the individual premium $\frac{1}{n}\mathcal{R}_\rho(\mu^{*n})$ at level $\alpha$ for both $\widehat{\mu^{*n}_{u_n}} := \mathcal{N}_{n\widehat{m}_{u_n},n\widehat{s}^2_{u_n}}$ and $\widehat{\mu^{*n}_{u_n}} := \widehat{\mu}^{*n}_{u_n}$.

## Numerical example

We let $u_n = n$ and $\rho = \mathrm{V@R}_{0.99}$ and $\mu = (1-p)\delta_0 + p\mathrm{P}_{a,b}$ for the Pareto distribution $\mathrm{P}_{a,b} = f_{a,b}\ell$ with Lebesgue density

$$f_{a,b}(x) := ab^{-1}\big(b^{-1}x+1\big)^{-(a+1)}\,\mathbb{1}_{(0,\infty)}(x) \qquad (a,b>0).$$

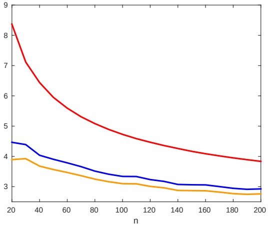
We fixed $p = 0.1$, considered the following fours sets of parameters

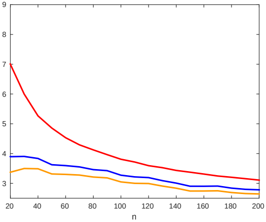
| $a$ | $b$ | mean($\mu$) | std($\mu$) | mean($\mathrm{P}_{a,b}$) | std($\mathrm{P}_{a,b}$) |
|---|---|---|---|---|---|
| 2.1 | 10 | 1 | 14.80 | 10 | 45.83 |
| 3 | 20 | 1 | 6.25 | 10 | 17.32 |
| 6 | 50 | 1 | 4.90 | 10 | 12.25 |
| 10 | 90 | 1 | 4.64 | 10 | 11.18 |

and did Monte Carlo simulations based on 50 repetitions.

(Note that $u_n$ is “small”, $\rho$ is “strict”, and $\mu$ is “risky” **!!!**)

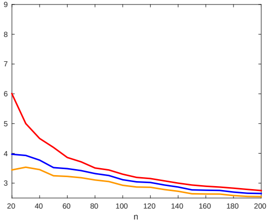
$a$=2.1
$b$=11

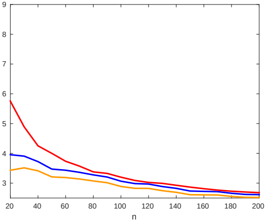



$a$=3
$b$=20

$a$=6
$b$=50

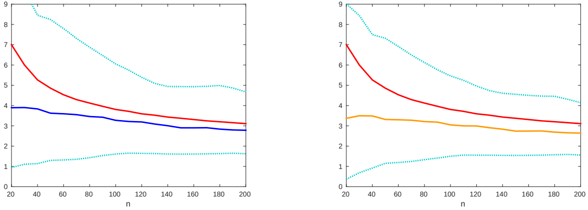


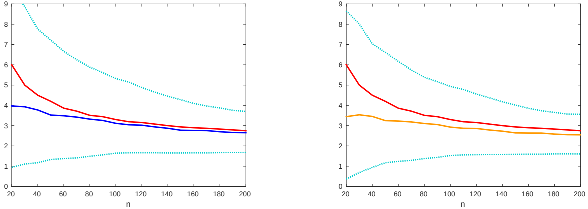

$a$=10
$b$=90

$\frac{1}{n}\mathcal{R}_\rho(\mu^{*n})$ $\quad \widehat{\text{mean}}\ \frac{1}{n}\mathcal{R}_\rho(\widehat{\mu}^{*n}_{u_n})$ $\quad \widehat{\text{mean}}\ \frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_{u_n}, n\widehat{s}^2_{u_n}})$

$a$=3,
$b$=20




$a$=6,
$b$=50




$\frac{1}{n}\mathcal{R}_\rho(\mu^{*n})$ $\widehat{\text{mean}}\ \frac{1}{n}\mathcal{R}_\rho(\widehat{\mu}^{*n}_{u_n})$ $\widehat{\text{mean}}\ \frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_{u_n}, n\widehat{s}^2_{u_n}})$

The light blue lines represent the empirical 5%- and 95%-quantiles.

## Conclusion

- The estimators are asymptotically equivalent.
  $\frac{1}{n}\mathcal{R}_\rho(\widehat{\mu}_{u_n}^{*n})$ is better than $\frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_{u_n}, n\widehat{s}^2_{u_n}})$ for finite sample.
- Good applicability to light-tailed $\mu$ and moderate $n$.
  Good applicability to medium-tailed $\mu$ and large $n$.
  Moderate applicability to heavy-tailed $\mu$ and large $n$.
- Both estimators have a negative bias.
  $\Longrightarrow$ Bias correction (see Section 4)

## Outlook

- Comparison with parametric models.

# 4. Bias correction through bootstrap

**Problem**

Both estimators have a negative bias.

**Countermeasure**

Estimate the bias and subtract it from the original estimator.

Use, for instance, a bootstrap-based estimator for the bias.

For simplicity I here restrict to Efron's bootstrap
(i.e. to the multiplier bootstrap with multinomial weights).

We have seen in Section 3 (Corollary) that

$$\mathrm{law}\Big\{\sqrt{u_n}\Big(\tfrac{1}{n}\mathcal{R}_\rho(\widehat{\mu^{*n}_{u_n}}) - \tfrac{1}{n}\mathcal{R}_\rho(\mu^{*n})\Big)\Big\} \xrightarrow{\mathsf{w}} \mathcal{N}_{0,s^2}$$

One can show (Lauer/H. Z. (2015+)) that also

$$\mathrm{law}\Big\{\sqrt{u_n}\Big(\tfrac{1}{n}\mathcal{R}_\rho(\widehat{\mu^{*n}_{u_n}}^{\mathsf{B}}) - \tfrac{1}{n}\mathcal{R}_\rho(\widehat{\mu^{*n}_{u_n}})\Big)\,\Big|\,(Y_1,\ldots,Y_{u_n})\Big\} \xrightarrow{\mathsf{p,w}} \mathcal{N}_{0,s^2}$$

Here

$\widehat{\mu^{*n}_{u_n}}$ is based on $(Y_1,\ldots,Y_{u_n})$,

$\widehat{\mu^{*n}_{u_n}}^{\mathsf{B}}$ is based on $(Y^{\mathsf{B}}_{n,1},\ldots,Y^{\mathsf{B}}_{n,u_n})$,

where, given $(Y_1,\ldots,Y_{u_n})$, the bootstrap sample $(Y^{\mathsf{B}}_{n,1},\ldots,Y^{\mathsf{B}}_{n,u_n})$ is drawn from the "urn" $\{Y_1,\ldots,Y_{u_n}\}$ with replacement.

That is, for large $n$ we have

$$\operatorname{law}\left\{ \tfrac{1}{n}\mathcal{R}_\rho(\widehat{\mu_{u_n}^{*n}}) - \tfrac{1}{n}\mathcal{R}_\rho(\mu^{*n}) \right\} \approx \mathcal{N}_{0,s^2/u_n}$$

and

$$\operatorname{law}\left\{ \tfrac{1}{n}\mathcal{R}_\rho(\widehat{\mu_{u_n}^{*n}}^{\mathsf{B}}) - \tfrac{1}{n}\mathcal{R}_\rho(\widehat{\mu_{u_n}^{*n}}) \,\Big|\, (Y_1,\ldots,Y_{u_n}) \right\} \approx \mathcal{N}_{0,s^2/u_n}.$$

Here

$\widehat{\mu_{u_n}^{*n}}$ is based on $(Y_1,\ldots,Y_{u_n})$,

$\widehat{\mu_{u_n}^{*n}}^{\mathsf{B}}$ is based on $(Y_{n,1}^{\mathsf{B}},\ldots,Y_{n,u_n}^{\mathsf{B}})$,

where, given $(Y_1,\ldots,Y_{u_n})$, the bootstrap sample $(Y_{n,1}^{\mathsf{B}},\ldots,Y_{n,u_n}^{\mathsf{B}})$ is drawn from the "urn" $\{Y_1,\ldots,Y_{u_n}\}$ with replacement.

That is, for large $n$ we have

$$\mathrm{law}\big\{\tfrac{1}{n}\mathcal{R}_\rho(\widehat{\mu^{*n}_{u_n}}) - \tfrac{1}{n}\mathcal{R}_\rho(\mu^{*n})\big\} \approx \mathcal{N}_{0,s^2/u_n} \qquad (\bigstar)$$

and

$$\mathrm{law}\big\{\tfrac{1}{n}\mathcal{R}_\rho(\widehat{\mu^{*n}_{u_n}}^{\mathsf{B}}) - \tfrac{1}{n}\mathcal{R}_\rho(\widehat{\mu^{*n}_{u_n}}) \,\big|\, (Y_1,\ldots,Y_{u_n})\big\} \approx \mathcal{N}_{0,s^2/u_n}.$$

($\bigstar$) pretends that $\mathrm{law}\big\{\tfrac{1}{n}\mathcal{R}_\rho(\widehat{\mu^{*n}_{u_n}}) - \tfrac{1}{n}\mathcal{R}_\rho(\mu^{*n})\big\}$ has mean zero.

However, from the numerical example in Section 3 we known that the mean is strictly negative.

Thus for our purpose $\mathcal{N}_{0,\widehat{s}^2_{u_n}/u_n}$ is **not** a reasonable estimator for the law of the empirical error. A better estimator can be defined as follows ...

That is, for large $n$ we have

$$\mathrm{law}\left\{\tfrac{1}{n}\mathcal{R}_\rho(\widehat{\mu^{*n}_{u_n}}) - \tfrac{1}{n}\mathcal{R}_\rho(\mu^{*n})\right\} \approx \mathcal{N}_{0,s^2/n}$$

and

$$\mathrm{law}\left\{\tfrac{1}{n}\mathcal{R}_\rho(\widehat{\mu^{*n}_{u_n}}^{\mathsf{B}}) - \tfrac{1}{n}\mathcal{R}_\rho(\widehat{\mu^{*n}_{u_n}}) \,\Big|\, (Y_1,\ldots,Y_{u_n})\right\} \approx \mathcal{N}_{0,s^2/n}.$$

Let $L \gg n$ and $\widehat{\mu^{*n}_{u_n}}^{\mathsf{B},1},\ldots,\widehat{\mu^{*n}_{u_n}}^{\mathsf{B},L}$ be based on $L$ independent bootstrap samples $(Y^{\mathsf{B},\ell}_{n,1},\ldots,Y^{\mathsf{B},\ell}_{n,n})$, $\ell = 1,\ldots,L$. Then

$$\tfrac{1}{L}\sum_{\ell=1}^{L}\delta_{\left(\frac{1}{n}\mathcal{R}_\rho(\widehat{\mu^{*n}_{u_n}}^{\mathsf{B},\ell}) - \frac{1}{n}\mathcal{R}_\rho(\widehat{\mu^{*n}_{u_n}})\right)}$$

provides the bootstrap estimator for

$$\mathrm{law}\left\{\tfrac{1}{n}\mathcal{R}_\rho(\widehat{\mu^{*n}_{u_n}}) - \tfrac{1}{n}\mathcal{R}_\rho(\mu^{*n})\right\}$$

... and one can use its mean

$$\widehat{\mathrm{Bias}}^{\mathsf{B}}_n := \frac{1}{L}\sum_{\ell=1}^{L}\left(\frac{1}{n}\mathcal{R}_\rho(\widehat{\mu^{*n}_{u_n}}^{\mathsf{B},\ell}) - \frac{1}{n}\mathcal{R}_\rho(\widehat{\mu^{*n}_{u_n}})\right)$$

as an estimator for

$$\mathrm{Bias}(\widehat{T}_n;\mu) := \mathbb{E}^\mu\left[\frac{1}{n}\mathcal{R}_\rho(\widehat{\mu^{*n}_{u_n}}) - \frac{1}{n}\mathcal{R}_\rho(\mu^{*n})\right].$$

In particular,

$$\widehat{T_n}^{\mathsf{BSC}} := \widehat{T}_n - \widehat{\mathrm{Bias}}^{\mathsf{B}}_n$$

provides an estimator for $T_n(\mu) := \frac{1}{n}\mathcal{R}_\rho(\mu^{*n})$ with smaller bias than

$$\widehat{T}_n := \frac{1}{n}\mathcal{R}_\rho(\widehat{\mu^{*n}_{u_n}}).$$

## Numerical example

We let $u_n = n = 100$ and $\rho = \mathrm{V@R}_{0.99}$ and $\mu = (1-p)\delta_0 + p\mathrm{P}_{a,b}$ for the Pareto distribution $\mathrm{P}_{a,b} = f_{a,b}\ell$ with Lebesgue density

$$f_{a,b}(x) := ab^{-1}\big(b^{-1}x+1\big)^{-(a+1)}\,\mathbb{1}_{(0,\infty)}(x) \qquad (a,b>0).$$

We fixed $p = 0.1$ and the following set of parameters

| $a$ | $b$ | mean($\mu$) | std($\mu$) | mean($\mathrm{P}_{a,b}$) | std($\mathrm{P}_{a,b}$) |
|---|---|---|---|---|---|
| 6 | 50 | 1 | 4.90 | 10 | 12.25 |

We did a Monte Carlo simulation based on $500$ repetitions (where the bootstrap estimator for the bias was based on $L = 1.000$ repetitions) and obtained

| $\frac{1}{n}\mathcal{R}_\rho(\mu^{*n})$ | $\widehat{\text{mean}}\ \frac{1}{n}\mathcal{R}_\rho(\widehat{\mu}_n^{*n})$ | $\widehat{\text{mean}}\ \frac{1}{n}\mathcal{R}_\rho(\widehat{\mu}_n^{*n})^{\text{BSC}}$ |
|---|---|---|
| 3.3 | 3.06 | 3.13 |

## Numerical example

We let $u_n = n = 100$ and $\rho = \mathrm{V@R}_{0.99}$ and $\mu = (1-p)\delta_0 + p\mathrm{P}_{a,b}$ for the Pareto distribution $\mathrm{P}_{a,b} = f_{a,b}\ell$ with Lebesgue density

$$f_{a,b}(x) := ab^{-1}\big(b^{-1}x+1\big)^{-(a+1)}\,\mathbb{1}_{(0,\infty)}(x) \qquad (a,b>0).$$

We fixed $p = 0.1$ and the following set of parameters

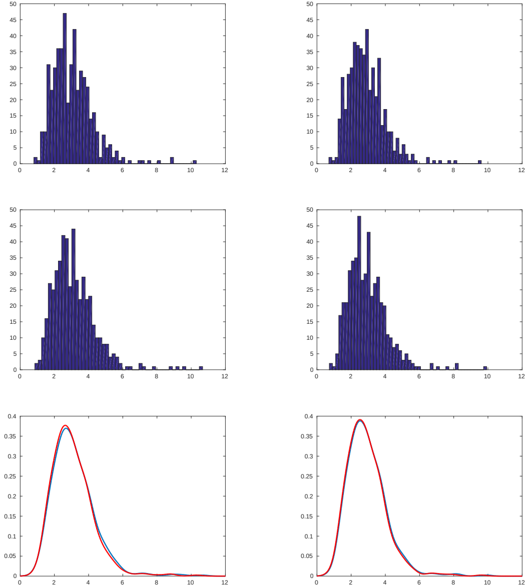
| $a$ | $b$ | mean($\mu$) | std($\mu$) | mean($\mathrm{P}_{a,b}$) | std($\mathrm{P}_{a,b}$) |
|---|---|---|---|---|---|
| 6 | 50 | 1 | 4.90 | 10 | 12.25 |

We did a Monte Carlo simulation based on $500$ repetitions (where the bootstrap estimator for the bias was based on $L = 1.000$ repetitions) and obtained

| $\frac{1}{n}\mathcal{R}_\rho(\mu^{*n})$ | $\widehat{\text{mean}}\ \frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_n, n\widehat{s}_n^2})$ | $\widehat{\text{mean}}\ \frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_n, n\widehat{s}_n^2})^{\text{BSC}}$ |
|---|---|---|
| 3.3 | 2.89 | 2.94 |

$a=6, b=50$
$n=100$

$\widehat{\text{law}}\ \frac{1}{n}\mathcal{R}_\rho(\widehat{\mu}_n^{*n})$ $\widehat{\text{law}}\ \frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_n, n\widehat{s}_n^2})$



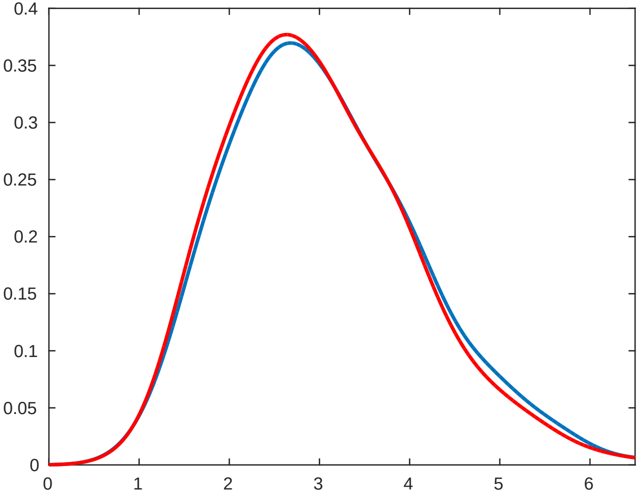

without BSC

with BSC

comparison

$a{=}6, b{=}50$
$n{=}100$

$\widehat{\text{law}}\ \frac{1}{n}\mathcal{R}_\rho(\widehat{\mu}_n^{*n})$




comparison

# 5. Qualitative robustness

As before we let

$$\Omega := \mathbb{R}^{\mathbb{N}}, \qquad \mathcal{F} := \mathcal{B}(\mathbb{R})^{\otimes \mathbb{N}}, \qquad \mathbb{P}^{\mu} := \mu^{\otimes \mathbb{N}}$$

and consider the statistical model

$$\big(\Omega, \mathcal{F}, \{\mathbb{P}^{\mu} : \mu \in \mathcal{M}(\mathcal{X})\}\big).$$

We assume that $\frac{1}{n}\sum_{i=1}^{n} \delta_{y_i} \in \mathcal{M}(\mathcal{X})$ for all $n \in \mathbb{N}$ and $y_i \in \mathbb{R}$.

The aspect of interest is

$$T_n(\mu) \ := \ \frac{1}{n}\mathcal{R}_{\rho}(\mu^{*n}),$$

and we use the estimator

$$\widehat{T}_n \ := \ \frac{1}{n}\mathcal{R}_{\rho}\big(\widehat{\mu}_{u_n}^{*n}\big).$$

## Definition

The sequence of estimators $(\widehat{T}_n)$ is said to be **robust on** $M \subseteq \mathcal{M}(\mathcal{X})$ if for every $\mu \in M$ and $\varepsilon > 0$ there exist a $\delta > 0$ such that for all $n \in \mathbb{N}$

$$\nu \in M, \quad d_{\text{weak}}(\mu, \nu) \leq \delta \quad \Longrightarrow \quad d_{\text{Lévy}}\left(\mathbb{P}^{\mu}_{\widehat{T}_n}, \mathbb{P}^{\nu}_{\widehat{T}_n}\right) \leq \varepsilon.$$

The definition was proposed by Hampel (1971) for $M = \mathcal{M}(\mathcal{X})$, where

$d_{\text{weak}}$ = any metric generating the weak topology, e. g. $d_{\text{weak}} = d_{\text{Lévy}}$

$d_{\text{Lévy}}(\mu, \nu) := \inf\{\varepsilon > 0 : \; F_\mu(x-\varepsilon)-\varepsilon \leq F_\nu(x) \leq F_\mu(x+\varepsilon)+\varepsilon \; \forall\, x \in \mathbb{R}\}$

**Definition**

The sequence of estimators $(\widehat{T}_n)$ is said to be **robust on** $M \subseteq \mathcal{M}(\mathcal{X})$ if for every $\mu \in M$ and $\varepsilon > 0$ there exist a $\delta > 0$ such that for all $n \in \mathbb{N}$

$$\nu \in M, \quad d_{\text{weak}}(\mu, \nu) \leq \delta \qquad \Longrightarrow \qquad d_{\text{Lévy}}\big(\mathbb{P}^{\mu}_{\widehat{T}_n}, \mathbb{P}^{\nu}_{\widehat{T}_n}\big) \leq \varepsilon.$$

**Corollary to Hampel's theorem**

If $\{T_n : n \in \mathbb{N}\}$ is equicontinuous for the weak topology, then $(\widehat{T}_n)$ is robust on $M = \mathcal{M}(\mathcal{X})$.

**Problem**

$\{T_n : n \in \mathbb{N}\}$ is **not** equicontinuous for the weak topology for any law-invariant coherent risk measure $\rho$ (if $\mathcal{X}$ is the "natural" domain of $\rho$).

**Definition**

The sequence of estimators $(\widehat{T}_n)$ is said to be **robust on** $M \subseteq \mathcal{M}(\mathcal{X})$ if for every $\mu \in M$ and $\varepsilon > 0$ there exist a $\delta > 0$ such that for all $n \in \mathbb{N}$

$$\nu \in M, \quad d_{\text{weak}}(\mu, \nu) \leq \delta \quad \Longrightarrow \quad d_{\text{Lévy}}\big(\mathbb{P}^{\mu}_{\widehat{T}_n}, \mathbb{P}^{\nu}_{\widehat{T}_n}\big) \leq \varepsilon.$$

**Corollary to Hampel's theorem**

If $\{T_n : n \in \mathbb{N}\}$ is equicontinuous for the weak topology,
then $(\widehat{T}_n)$ is robust on $M = \mathcal{M}(\mathcal{X})$.

**To do**

Find a suitable generalization of Hampel's theorem!

## The $p$-weak topology

Let $p \in [0, \infty]$. On (any subset of)

$$\mathcal{M}(L^p) := \left\{ \mu \in \mathcal{M}_1(\mathbb{R}) \, : \, \int |x|^p \, \mu(dx) < \infty \right\}$$

we may impose the $p$-weak topology, that is, the coarsest topology for which all mappings $\mu \longmapsto \int f \, d\mu$, $f \in C_p(\mathbb{R})$, are continuous, where

$$C_p(\mathbb{R}) := \left\{ f \in C(\mathbb{R}) \, : \, |f(x)| \leq c|x|^p \text{ for some } c \in (0, \infty) \right\}.$$

Note that

$$\begin{aligned} & \mu_n \to \mu \quad p\text{-weakly} \\ \iff \quad & \int f \, d\mu_n \to \int f \, d\mu \text{ for all } f \in C_p(\mathbb{R}) \\ \iff \quad & \mu_n \to \mu \text{ weakly and } \int |x|^p \, \mu_n(dx) \to \int |x|^p \, \mu(dx) \end{aligned}$$

Let $\mathcal{X} = L^p$, i.e. the domain of $\mathcal{R}_\rho$ is $\mathcal{M}(L^p)$.

**Definition**

The sequence $(\widehat{T}_n)$ is said to be $p$-**robust on** $M \subseteq \mathcal{M}(L^p)$ if for every $\mu \in \mathcal{M}(L^p)$ and $\varepsilon > 0$ there exist a $\delta > 0$ such that for all $n \in \mathbb{N}$

$$\nu \in \mathcal{M}(L^p), \quad d_{p\text{-weak}}(\mu, \nu) \leq \delta \quad \Longrightarrow \quad d_{\text{Lévy}}\left(\mathbb{P}^{\mu}_{\widehat{T}_n}, \mathbb{P}^{\nu}_{\widehat{T}_n}\right) \leq \varepsilon.$$

**Theorem**

If $\{T_n : n \in \mathbb{N}\}$ is equicontinuous for the $p$-weak topology, then $(\widehat{T}_n)$ is $p$-robust on every locally uniformly $p$-integrating set $M \subseteq \mathcal{M}(L^p)$.

H. Z. (2016)
Krätschmer/Schied/H. Z. (2012, 2014)

Let $\mathcal{X} = L^p$, i.e. the domain of $\mathcal{R}_\rho$ is $\mathcal{M}(L^p)$.

**Definition**

The sequence $(\widehat{T}_n)$ is said to be **$p$-robust on** $M \subseteq \mathcal{M}(L^p)$ if for every $\mu \in \mathcal{M}(L^p)$ and $\varepsilon > 0$ there exist a $\delta > 0$ such that for all $n \in \mathbb{N}$

$$\nu \in \mathcal{M}(L^p), \quad d_{p\text{-weak}}(\mu, \nu) \leq \delta \quad \Longrightarrow \quad d_{\text{Lévy}}\left(\mathbb{P}^{\mu}_{\widehat{T}_n}, \mathbb{P}^{\nu}_{\widehat{T}_n}\right) \leq \varepsilon.$$

**Theorem**

If $\{T_n : n \in \mathbb{N}\}$ is equicontinuous for the $p$-weak topology, then $(\widehat{T}_n)$ is $p$-robust on every locally uniformly $p$-integrating set $M \subseteq \mathcal{M}(L^p)$.

*Here $M \subseteq \mathcal{M}_1$ is said to be locally uniformly $p$-integrating if for every $\mu \in M$ and $\varepsilon > 0$ there exist $\delta > 0$ and $a > 0$ such that*

$$\nu \in M, \quad d_{\text{weak}}(\mu, \nu) \leq \delta \quad \Longrightarrow \quad \int |x|^p \mathbb{1}_{\{|x|^p \geq a\}}\, \nu(dx) \leq \varepsilon.$$

Let $\mathcal{X} = L^p$, i.e. the domain of $\mathcal{R}_\rho$ is $\mathcal{M}(L^p)$.

**Definition**

The sequence $(\widehat{T}_n)$ is said to be **$p$-robust on** $M \subseteq \mathcal{M}(L^p)$ if for every $\mu \in \mathcal{M}(L^p)$ and $\varepsilon > 0$ there exist a $\delta > 0$ such that for all $n \in \mathbb{N}$

$$\nu \in \mathcal{M}(L^p), \quad d_{p\text{-weak}}(\mu, \nu) \leq \delta \quad \Longrightarrow \quad d_{\text{Lévy}}\big(\mathbb{P}^{\mu}_{\widehat{T}_n}, \mathbb{P}^{\nu}_{\widehat{T}_n}\big) \leq \varepsilon.$$

**Theorem**

If $\{T_n : n \in \mathbb{N}\}$ is equicontinuous for the $p$-weak topology, then $(\widehat{T}_n)$ is $p$-robust on every locally uniformly $p$-integrating set $M \subseteq \mathcal{M}(L^p)$.

In particular, in this case $(\widehat{T}_n)$ is robust on every locally uniformly $p$-integrating set $M \subseteq \mathcal{M}(L^p)$ on which the weak topology and the $p$-weak topology coincide.

**Theorem**

For $p \in [0, 1]$, $M \subseteq \mathcal{M}(L^p)$ the following conditions are equivalent:

(a) The weak topology and the $p$-weak topologies on $M$ coincide.

(b) $M$ is locally uniformly $p$-integrating.

(c) Every weakly compact subset of $M$ is uniformly $p$-integrating.

(d) Every weakly convergent sequence in $M$ is uniformly $p$-integrating.

(e) For every sequence $(\mu_n) \subseteq M$ for which $\mu_n$ converges weakly to $\mu_0$ the convergence $\int |x|^p \, \mu_n(dx) \to \int |x|^p \, \mu_0(dx)$ holds.

H. Z. (2016)
Krätschmer/Schied/H. Z. (2015+)

**Theorem**

For $p \in [0,1]$, $M \subseteq \mathcal{M}(L^p)$ the following conditions are equivalent:

(a) The weak topology and the $p$-weak topologies on $M$ coincide.

(b) $M$ is locally uniformly $p$-integrating.

(c) Every weakly compact subset of $M$ is uniformly $p$-integrating.

(d) Every weakly convergent sequence in $M$ is uniformly $p$-integrating.

(e) For every sequence $(\mu_n) \subseteq M$ for which $\mu_n$ converges weakly to $\mu_0$ the convergence $\int |x|^p \, \mu_n(dx) \to \int |x|^p \, \mu_0(dx)$ holds.

**Definition**

A set $M \subseteq \mathcal{M}(L^p)$ satisfying condition (a) is called **w-set in** $\mathcal{M}(L^p)$.

*(Note: The smaller $p$, the larger one can make such w-sets).*

Let $\mathcal{X} = L^p$, i.e. the domain of $\mathcal{R}_\rho$ is $\mathcal{M}(L^p)$.

**Corollary**

If $\{T_n : n \in \mathbb{N}\}$ is equicontinuous for the $p$-weak topology, then $(\widehat{T}_n)$ is robust on every w-set $M$ in $\mathcal{M}(L^p)$.

**Examples**

$\{T_n : n \in \mathbb{N}\}$ is equicontinuous for the $p$-weak topology if

- $\rho = \mathrm{AV@R}_\alpha$ $(p = 1)$, $\quad \rho = \mathrm{Ept}_\alpha$ $(p = 1)$, $\quad \rho = \mathrm{OsM}_{p,a}$, $\quad \ldots$

W-sets in $\mathcal{M}(L^p)$ are fairly large. Examples are

- the set of all normal distributions, the set of all Gamma distributions, the set of all Pareto distributions with tail-index $a \geq a_0 > p$, ...

# Our theory is taken into account in the IAIS *Risk-based Global Insurance Capital Standard*:

IAIS
INTERNATIONAL ASSOCIATION OF INSURANCE SUPERVISORS

Public

**Risk-based Global Insurance Capital Standard**

**17 December 2014**

**Public Consultation Document**
**Comments due by 16 February 2015**



**Table 3. Main features of VaR and Tail-VaR**

| Features/Risk measure | VaR | Tail-VaR |
|---|---|---|
| Frequency captured? | Yes | Yes |
| Severity captured? | No | Yes |
| Sub-additive? | Not always | Always |
| Diversification captured? | Issues | Yes |
| Back-testing? | Straight-forward | Issues |
| Estimation? | Feasible | Issues with data limitation |
| Model uncertainty? | Sensitive to aggregation | Sensitive to tail modelling |
| Robustness I (with respect to "Lévy metric[33]")? | Almost, only minor issues | No |
| Robustness II (with respect to "Wasserstein metric[34]")? | Yes | Yes |

[33] The Lévy metric is a metric on the space of cumulative distribution functions of one-dimensional random variables. It is a special case of the Lévy–Prokhorov metric.

[34] The Wasserstein (or Vasershtein) metric is a distance function defined between probability distributions on a given metric space, the metric is also known for its optimal transport properties.

# Thank you!

# References


- Krätschmer, V. and Zähle, H. (2011).
  Sensitivity of risk measures w.r.t. the normal approximation of total claim distributions.
  *Insurance: Mathematics and Economics*, 49, 335–344.
- Krätschmer, V., Schied, A., and Zähle, H. (2012).
  Qualitative and infinitesimal robustness of tail-dependent statistical functionals.
  *Journal of Multivariate Analysis* 2012, 103, 35–47.
- Krätschmer, V., Schied, A., and Zähle, H. (2014).
  Comparative and qualitative robustness of law-invariant risk measures.
  *Finance and Stochastics*, 18, 271–295.
- Krätschmer, V., Schied, A., and Zähle, H. (2015+).
  Domains of weak continuity of statistical functionals with a view on robust statistics.
  *Preprint*.
- Lauer, A. and Zähle, H. (2015).
  Nonparametric estimation of risk measures of collective risks.
  *Statistics and Risk Modeling*, accepted for publication.
- Zähle, H. (2016).
  A definition of qualitative robustness for general point estimators, and examples.
  *Journal of Multivariate Analysis*, 143, 12–31.