An Axiomatic Foundation of the Expected Shortfall

Ruodu Wang

http://sas.uwaterloo.ca/~wang

Department of Statistics and Actuarial Science
University of Waterloo

INFORMS Annual Meeting 2019
Seattle, USA October 2019
(revised talk slides)

Agenda

1. The main question
2. Economic axioms
3. Tail events and risk concentration
4. Risk aggregation
5. Concluding remarks

Based on joint work with Ričardas Zitikis (Western Ontario)

Risk measures

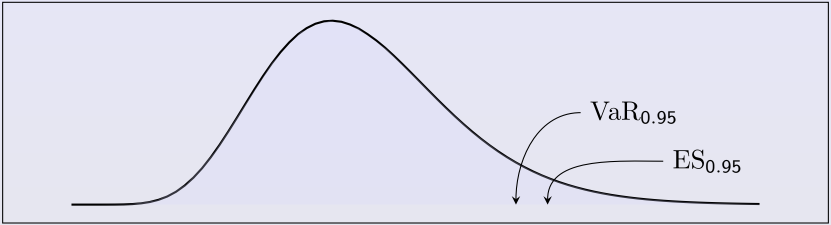
A risk measure $\rho : \mathcal{X} \to \mathbb{R}$ maps a risk (via a model) to a number

- regulatory capital calculation $\leftarrow$ our main focus
- decision making, optimization, portfolio selection, ...
- performance analysis and capital allocation
- pricing

Risks ...

- $\mathcal{X}$ is a set of random losses in one period (e.g. 10d) in an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$
- F_X denotes the cdf of $X \in \mathcal{X}$

VaR and ES



Value-at-Risk (VaR), $p \in (0,1)$

$\mathrm{VaR}_p : L^0 \to \mathbb{R}$,

$$\mathrm{VaR}_p(X) = F_X^{-1}(p) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \le x) \ge p\}.$$

(right-quantile)

Expected Shortfall (ES), $p \in (0,1)$

$\mathrm{ES}_p : L^1 \to \mathbb{R}$,

$$\mathrm{ES}_p(X) = \frac{1}{1-p}\int_p^1 \mathrm{VaR}_q(X)\mathrm{d}q$$

(also: TVaR/CVaR/AVaR)

FRTB

The Basel Committee on Banking Supervision (BCBS)

Fundamental Review of the Trading Book (FRTB), Jan 2016

- $\mathrm{VaR}_{0.99}$ is officially replaced by $\mathrm{ES}_{0.975}$ as the standard risk measure for market risk
- 10-day portfolio loss

Page 1, Executive Summary:

> "*Use of ES will help to ensure a more prudent capture of "tail risk" and capital adequacy ...*"

What is so special about ES?

What is magical about ES?

An ES is

- Coherent (Artzner-Delbaen-Eber-Heath'99, Acerbi-Tasche'02)
- Comonotone-additive (Kusuoka'01) (also VaR)
- Tail-relevant (Liu-W.'18) (also VaR)
- Min-convex expectation (Rockafellar-Uryasev'00)

What is so special about ES?

What is magical about ES?

An ES is

- Coherent (Artzner-Delbaen-Eber-Heath'99, Acerbi-Tasche'02)
- Comonotone-additive (Kusuoka'01) (also VaR)
- Tail-relevant (Liu-W.'18) (also VaR)
- Min-convex expectation (Rockafellar-Uryasev'00)

None of the above, and not even all together, characterizes ES

- e.g. Gini Shortfall (Furman-W.-Zitikis'17)

Axiomatic approach for ES

Target: Find a set of meaningful axioms that uniquely characterizes the family of ES

Theory and Decision
https://doi.org/10.1007/s11238-018-09685-1

What are axiomatizations good for?

Itzhak Gilboa[1,2] · Andrew Postlewaite[3] · Larry Samuelson[4] · David Schmeidler[2]

Axiomatic approaches for VaR

Axiomatic characterizations of VaR (quantile):

- Chambers'09: ordinal-covariance + monotonicity + law-invariance
- Kou-Peng'16: elicitability + comonotonic-additivity + monotonicity
- He-Peng'18: surplus-invariance + law-invariance + positive homogeneity
- Liu-W.'18: elicitability + tail-relevance + positive homogeneity

all + some form of continuity

Axiomatic approach for ES

If the set of economic axioms for ES:

- correctly reflects the regulators' practical intentions
  $\Rightarrow$ justify and support the use of ES in regulation
- contradicts the regulators' intentions
  $\Rightarrow$ discuss whether ES is still the best risk measure to use

Axiomatic approach for risk functionals

Decision theory

- Expected utility: von Neumann-Morgenstern'44, Savage'54
- Dual utility: Yaari'87
- Variational preferences: Gilboa-Schmeidler'89, Schmeidler'89, Maccheroni-Marinacci-Rustichini'06

Banking and insurance

- Coherent risk measures: Artzner-Delbaen-Eber-Heath'99
- Convex risk measures: Föllmer-Schied'02, Fritteli-Rosazza Gianin'02
- Insurance pricing: Wang-Young-Panjer'97
- Systemic risk measures: Chen-Iyengar-Moallemi'13

Progress

Axioms **M** and **LI**

A risk measure $\rho : \mathcal{X} \to \mathbb{R}$

- $\rho(X)$ is the amount of regulatory capital for a particular risk model X
- e.g. $\mathcal{X} = L^0,\ L^1,\ L^\infty,\ \ldots$

Axioms **M** and **LI**

A risk measure $\rho : \mathcal{X} \to \mathbb{R}$

- $\rho(X)$ is the amount of regulatory capital for a particular risk model X
- e.g. $\mathcal{X} = L^0, L^1, L^\infty, \ldots$

Two intuitive axioms

M. (Monotonicity) A surely larger or equal loss leads to a larger or equal risk value, that is, $\rho(X) \le \rho(Y)$ whenever $X \le Y$.

Axioms **M** and **LI**

A risk measure $\rho : \mathcal{X} \to \mathbb{R}$

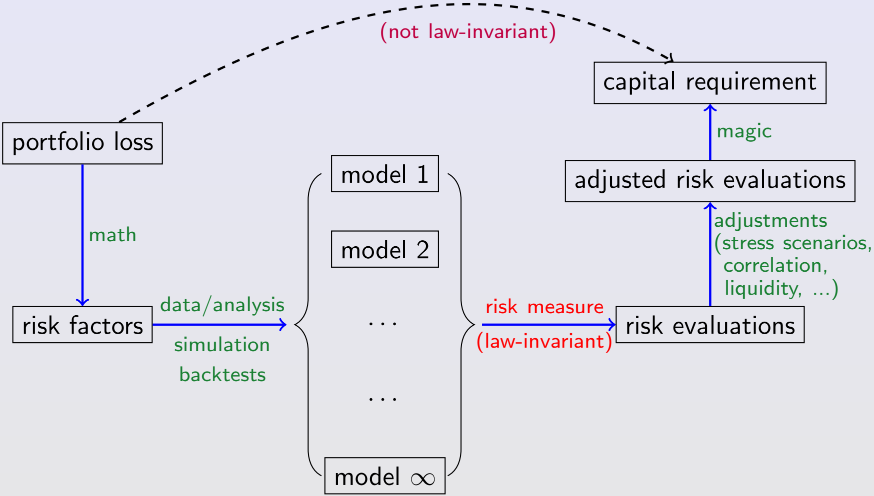
- $\rho(X)$ is the amount of regulatory capital for a particular risk model X
- e.g. $\mathcal{X} = L^0,\ L^1,\ L^\infty,\ \ldots$

Two intuitive axioms

M. (Monotonicity) A surely larger or equal loss leads to a larger or equal risk value, that is, $\rho(X) \le \rho(Y)$ whenever $X \le Y$.

LI. (Law-invariance) The risk value depends on the loss via its distribution, that is, $\rho(X) = \rho(Y)$ whenever $X \stackrel{d}{=} Y$.

The risk assessment process



Axiom P

A third intuitive axiom

P. (Prudence) The risk value is not underestimated by approximations, that is, $\lim_n \rho(\xi_n) \geq \rho(X)$ whenever $\xi_n \to X$ point-wise and $\lim_n \rho(\xi_n)$ exists.

- The loss X is modelled truthfully (e.g. consistent estimators) $\Rightarrow$ estimated risk $\geq$ true risk asymptotically

Axiom P

A third intuitive axiom

P. (Prudence) The risk value is not underestimated by approximations, that is, $\lim_n \rho(\xi_n) \geq \rho(X)$ whenever $\xi_n \to X$ point-wise and $\lim_n \rho(\xi_n)$ exists.

- The loss X is modelled truthfully (e.g. consistent estimators) $\Rightarrow$ estimated risk $\geq$ true risk asymptotically

Proposition

For $p \in (0,1)$, both ES_p and VaR_p on $\mathcal{X} = L^1$ satisfy Axioms **M**, **LI** *and* **P**.

Toward the fourth axiom: step 0

Practitioners' intuitions: BCBS (Feb 2019)

10.22 Diversification: the reduction in risk at a portfolio level due to holding risk positions in different instruments that are not perfectly correlated with one another.

22.4 No diversification benefit is recognised between the DRC requirements for: (1) non-securitisations; (2) securitisations (non-CTP); and (3) securitisations (CTP).

30.17(3b) [...] with sufficient consideration given to ensuring: [...] that the models reflect concentration risk that may arise in an undiversified portfolio.

30.20 Banks' stress scenarios must cover a range of factors that (i) can create extraordinary losses or gains in trading portfolios, or (ii) make the control of risk in those portfolios very difficult. These factors include low-probability events in all major types of risk, [...]

Toward the fourth axiom: step 1

For a portfolio vector $(X_1, \ldots, X_n)$, there is diversification benefit if

$$\rho\left(\sum_{i=1}^{n} X_i\right) < \sum_{i=1}^{n} \rho(X_i).$$

Toward the fourth axiom: step 1

For a portfolio vector $(X_1, \ldots, X_n)$, there is diversification benefit if

$$\rho\left(\sum_{i=1}^{n} X_i\right) < \sum_{i=1}^{n} \rho(X_i).$$

Three features of portfolio regulatory capital:

- rewards diversification: $\rho\left(\sum_{i=1}^{n} X_i\right) < \sum_{i=1}^{n} \rho(X_i)$ if the portfolio is properly diversified

Toward the fourth axiom: step 1

For a portfolio vector $(X_1, \ldots, X_n)$, there is diversification benefit if

$$\rho\left(\sum_{i=1}^{n} X_i\right) < \sum_{i=1}^{n} \rho(X_i).$$

Three features of portfolio regulatory capital:

- rewards diversification: $\rho\left(\sum_{i=1}^{n} X_i\right) < \sum_{i=1}^{n} \rho(X_i)$ if the portfolio is properly diversified
- penalizes risk concentration: $\rho\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \rho(X_i)$ if the portfolio is concentrated/non-diversified

Toward the fourth axiom: step 1

For a portfolio vector $(X_1, \ldots, X_n)$, there is diversification benefit if

$$\rho\left(\sum_{i=1}^{n} X_i\right) < \sum_{i=1}^{n} \rho(X_i).$$

Three features of portfolio regulatory capital:

- rewards diversification: $\rho\left(\sum_{i=1}^{n} X_i\right) < \sum_{i=1}^{n} \rho(X_i)$ if the portfolio is properly diversified
- penalizes risk concentration: $\rho\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \rho(X_i)$ if the portfolio is concentrated/non-diversified
- tail events: a focus on events of small probability that the most severe loss occurs

Toward the fourth axiom: step 2

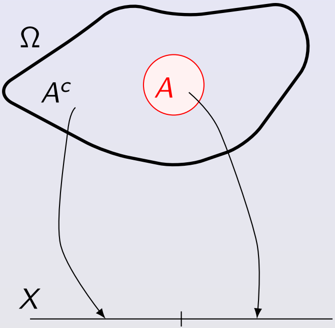
Definition (Tail events)

A tail event of X is $A \in \mathcal{F}$ such that

a) $0 < \mathbb{P}(A) < 1$

b) $X(\omega) \geq X(\omega')$

for a.s. all $\omega \in A$ and $\omega' \in A^c$



Remark.

- tail event $\Longrightarrow$ most severe loss

Toward the fourth axiom: step 3

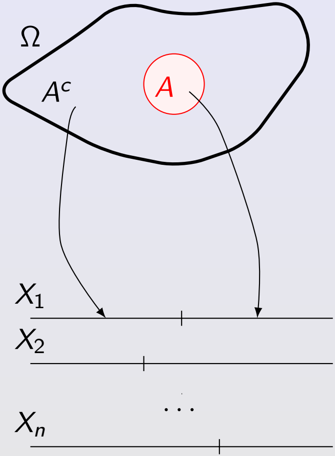
Main idea

concentrated portfolio $\Longleftrightarrow$

severe losses occur simultaneously

on a stress event

- A: a stress event specified by the regulator



The fourth axiom

The fourth key axiom

NRC. (No reward for concentration) There exists an event $A \in \mathcal{F}$ such that $\rho(X+Y) = \rho(X) + \rho(Y)$ holds for all risks X and Y sharing the tail event A.

Remark.

- Axiom **NRC** may be equivalently formulated via: for all $n \geq 2$, $\rho\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \rho(X_i)$ whenever $X_1, \ldots, X_n$ share a tail event A
- Axioms **M**, **P** and **NRC** are model-free (independent of $\mathbb{P}$)

Axiomatic characterization of ES

Theorem

A functional $\rho : L^1 \to \mathbb{R}$ with $\rho(1) = 1$ satisfies Axioms **M**, **LI**, **P** *and* **NRC** *if and only if $\rho = \mathrm{ES}_p$ for some $p \in (0, 1)$.*

Remarks.

- In the forward direction, the value of $p = \mathbb{P}(A)$ specified in Axiom **NRC**
- $\rho(1) = 1$ is normalizing

Axiomatic characterization of ES

None of the axioms rely on integrability.

Is the domain $\mathcal{X} = L^1$ natural?

Axiomatic characterization of ES

None of the axioms rely on integrability.

Is the domain $\mathcal{X} = L^1$ natural?

Theorem

For any $q \in [0,1)$, a functional $\rho : L^q \to \mathbb{R}$ satisfies Axioms **M**, **LI**, **P** *and* **NRC** *if and only if $\rho = 0$ on L^q.*

- No meaningful risk measure satisfying **M**, **LI**, **P** and **NRC** is defined beyond L^1
- For L^q, $q \in [1, \infty]$, the previous ES characterization holds

Independence of the axioms

Axioms **M**, **LI**, **P** and **NRC** are independent on $\mathcal{X} = L^1$:

- **M** + **LI** + **P** − **NRC**: VaR_p $\quad p \in (0,1)$
- **M** + **LI** + **NRC** − **P**: $\mathbb{E}$
- **M** + **P** + **NRC** − **LI**: $X \mapsto X(\omega)$ $\quad \omega \in \Omega$
- **LI** + **P** + **NRC** − **M**: $X \mapsto \mathrm{ES}_p(-X)$ $\quad p \in (0,1)$

Progress

Tail events and risk concentration

For $p \in (0,1)$ and a random vector $(X_1, \dots, X_n)$:

- p-tail event: a tail event of probability $1 - p$
- $(X_1, \dots, X_n)$ is p-concentrated: $X_1, \dots, X_n$ share a p-tail event

Tail events

A p-tail event of X

- always exists
- is a.s. unique if X is continuously distributed
- is invariant under strictly increasing marginal transformations
- A is a p-tail event of X

$$\iff \mathbb{P}(A) = 1 - p \text{ and } \{X > x\} \subset A \subset \{X \geq x\} \text{ a.s.}$$

where $x = \mathrm{VaR}_p(X)$.

Remark.

- The case for discrete random variables is more complicated, but crucial for our theory

Risk concentration

p-concentration as a dependence concept

- A notion of positive dependence

Risk concentration

p-concentration as a dependence concept

- A notion of positive dependence

Theorem

A random vector is p-concentrated for all $p \in (0,1)$ if and only if it is comonotonic.

- Concentration is a weaker notion than comonotonicity
- Comonotonicity may be too strong a requirement for a "non-diversified portfolio"
- Additional flexibility: $p \in (0,1)$ is specified by the regulator
- Axiom **NRC** implies comonotone-additivity

Properties of risk concentration

$(X_1, \ldots, X_n)$ is p-concentrated $\Rightarrow$ so is each pair (X_i, X_j)

- The converse is true if some X_i is continuously distributed
- The converse is generally not true in sharp contrast to the case of comonotonicity

Example (Pair-wise concentration does not imply concentration)

- A_1, A_2, A_3 are three disjoint, each of probability $p = 1/3$
- $X_i = \mathbb{1}_{A_i}$ for $i = 1, 2, 3$
- (X_i, X_j) has a common p-tail event $A_i \cup A_j$
- (X_1, X_2, X_3) does not have a common p-tail event

Properties of risk concentration

Theorem

For every $p \in (0,1)$ *and every random vector* $(X_1, \ldots, X_n)$, *writing* $S = X_1 + \cdots + X_n$, *equivalent are:*

1. $(X_1, \ldots, X_n)$ *is p-concentrated;*
2. $(X_1, \ldots, X_n, S)$ *is p-concentrated;*
3. $(X_i, S - X_i)$ *is p-concentrated for every* $i = 1, \ldots, n$*;*
4. $(f_1(X_1), \ldots, f_n(X_n))$ *is p-concentrated for all increasing functions* $f_1, \ldots, f_n$*;*
5. *a copula* C *of* $(X_1, \ldots, X_n)$ *satisfies* $C(p, \ldots, p) = p$.

Progress

Risk aggregation

Given any $p \in (0,1)$, the random vector $(X_1, \dots, X_n) \in (L^1)^n$ is said to maximize the ES_p aggregation if

$$\mathrm{ES}_p\left(\sum_{i=1}^n X_i\right) = \max\left\{\mathrm{ES}_p\left(\sum_{i=1}^n X_i'\right) : X_i' \overset{d}{=} X_i,\ i = 1, \dots, n\right\}.$$

Risk aggregation

Given any $p \in (0,1)$, the random vector $(X_1, \dots, X_n) \in (L^1)^n$ is said to maximize the ES_p aggregation if

$$\mathrm{ES}_p\left(\sum_{i=1}^n X_i\right) = \max\left\{\mathrm{ES}_p\left(\sum_{i=1}^n X_i'\right) : X_i' \overset{d}{=} X_i,\ i = 1, \dots, n\right\}.$$

Known: Comonotonicity maximizes ES_p aggregation

Q: Is comonotonicity necessary?

Hint: Comonotonicity $\iff$ p-concentration for all $p \in (0,1)$

Risk aggregation

Theorem

For $p \in (0,1)$ and $(X_1, \dots, X_n) \in (L^1)^n$, equivalent are:

1. $(X_1, \dots, X_n)$ *is p-concentrated;*
2. $(X_1, \dots, X_n)$ *maximizes the* ES_p *aggregation;*
3. $\mathrm{ES}_p\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{ES}_p(X_i)$.

Risk aggregation

Theorem

For $p \in (0,1)$ and $(X_1, \dots, X_n) \in (L^1)^n$, equivalent are:

1. $(X_1, \dots, X_n)$ *is p-concentrated;*
2. $(X_1, \dots, X_n)$ *maximizes the* ES_p *aggregation;*
3. $\mathrm{ES}_p\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{ES}_p(X_i)$.

Remarks.

- Comonotonicity is not necessary for max ES_p aggregation
- ES_p is additive for and only for a p-concentrated portfolio
- ES_p satisfies Axiom **NRC**

Risk aggregation

Proof of (i)⇔(iii). Note the dual representation of ES_p:

$$\mathrm{ES}_p(X) = \sup_{\mathbb{P}(A)=1-p} \mathbb{E}[X|A], \quad X \in L^1.$$

- Lemma. For $p \in (0,1)$, $X \in L^1$ and $\mathbb{P}(A) = 1-p$,

$$\mathrm{ES}_p(X) = \mathbb{E}[X|A] \Leftrightarrow A \text{ is a } p\text{-tail event of } X.$$

- (i) $\underset{\text{(thm)}}{\Leftrightarrow}$ $\exists$ a common p-tail event A of $X_1, \dots, X_n, S$ $\Rightarrow$

$$\sum_{i=1}^n \mathrm{ES}_p(X_i) \underset{\text{(lemma)}}{=} \sum_{i=1}^n \mathbb{E}[X_i|A] = \mathbb{E}[S|A] \underset{\text{(lemma)}}{=} \mathrm{ES}_p(S) \Rightarrow \text{(iii)}.$$

- (iii) $\Rightarrow$ for a p-tail event A of S,

$$\sum_{i=1}^n \mathrm{ES}_p(X_i) = \mathrm{ES}_p(S) \underset{\text{(lemma)}}{=} \mathbb{E}[S|A] = \sum_{i=1}^n \mathbb{E}[X_i|A] \underset{\text{(lemma)}}{\Rightarrow} \text{(i)}.$$

Risk aggregation

Define the right p-quantile

$$\mathrm{VaR}_p^+(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \le x) > p\}, \quad X \in L^0, \ p \in (0,1).$$

Theorem

For every $p \in (0,1)$ and every p-concentrated vector $(X_1, \dots, X_n)$, writing $S = X_1 + \cdots + X_n$, we have

$$\mathrm{VaR}_p(S) \le \sum_{i=1}^n \mathrm{VaR}_p(X_i) \le \sum_{i=1}^n \mathrm{VaR}_p^+(X_i) \le \mathrm{VaR}_p^+(S).$$

If the quantile function of S is continuous at p, then all inequalities above are equalities.

Risk aggregation

Remarks on VaR_p and VaR_p^+

- They are both additive for any comonotonic portfolio
  - Generally not additive for a p-concentrated portfolio
  - Fail to satisfy Axiom **NRC**
- VaR_p is subadditive for any p-concentrated portfolio
- VaR_p^+ is superadditive for any p-concentrated portfolio
- $\mathrm{VaR}_p(S) < \mathrm{VaR}_p^+(S) \Leftrightarrow$ the quantile of S has a jump at p
  - Such a jump is not strange as p-concentration already imposes some degeneracy

Risk aggregation

Example (VaR_p does not satisfy Axiom NRC)

- $U \sim \mathrm{U}[0,1]$ and $p \in (0,1)$
- A is an event with $\mathbb{P}(A) = p$ independent of U
- $X = U\mathbb{1}_A + \mathbb{1}_{A^c}$ and $Y = (1-U)\mathbb{1}_A + \mathbb{1}_{A^c}$
- A^c is a common p-tail event of X and Y
- $\mathrm{VaR}_p(X) = \mathrm{VaR}_p(Y) = 1$
- $\mathrm{VaR}_p(X+Y) = \mathrm{VaR}_p(\mathbb{1}_A + 2\mathbb{1}_{A^c}) = 1$
- $\Rightarrow \mathrm{VaR}_p(X+Y) < \mathrm{VaR}_p(X) + \mathrm{VaR}_p(Y)$

Progress

1. The main question
2. Economic axioms
3. Tail events and risk concentration
4. Risk aggregation
5. Concluding remarks

Concluding remarks

Main contributions

- Four axioms, **M**, **LI**, **P** and **NRC**, uniquely identify ES
- Mathematical concepts and results
  - Tail events and risk concentration
  - Risk aggregation for ES and VaR
  - Characterization theorems

Discussions

- Are the axioms consistent with regulator's intentions?
- How special is ES?
- Are there other ways to characterize ES?

VaR versus ES: Summary

	Value-at-Risk	Expected Shortfall
Domain	always exists	needs first moment
Capturing	only frequency	frequency and severity
Diversification	non-coherent/non-NRC	coherent/NRC
Optimization	non-convex/non-robust	convex/robust
Backtesting	straightforward	complicated
Estimation	comparably difficult	comparably difficult
Allocation	difficult to estimate	straightforward (Euler)
Robustness	weak topology	L-metrics
Elicitation	complexity $= 1$	complexity $= 2$
Numéraire invariance	yes	no
Surplus invariance	yes	no

Thank you

Thank you for your kind attention

The manuscript is available at SSRN:3423042

Comments are welcome