Goodhart's Law and Risk Optimization

Ruodu Wang

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

Department of Statistics and Actuarial Science
University of Waterloo

Department of Risk, Insurance, and Healthcare Management
Fox School of Business, Temple University

October 14, 2021

Agenda

Based on joint work with Paul Embrechts (Zurich) and Alexander Schied (Waterloo)

Goodhart's law

Goodhart's law (Goodhart'75)

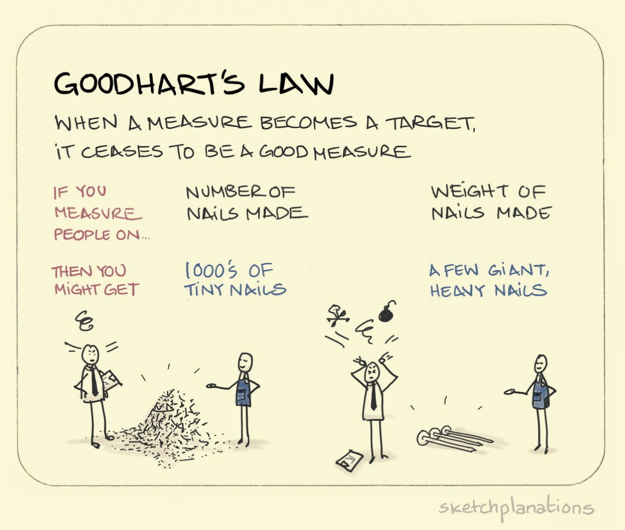
Any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes.

Popular version (Strathern'97)

When a measure becomes a target, it ceases to be a good measure.

- When a feature of the economy is picked as an indicator of the economy, then it inexorably ceases to function as that indicator because people start to game it.
- Monetary policies, scientific impact, economic indices, standardized exams, rankings, ratings, ...

Goodhart's law



Regulatory risk measures

A risk measure ρ maps a risk (via a model) to a number

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

Goodhart's law for risk measures

Goodhart's law for risk measures: *When a risk measure becomes a target, it ceases to be a good risk measure.*

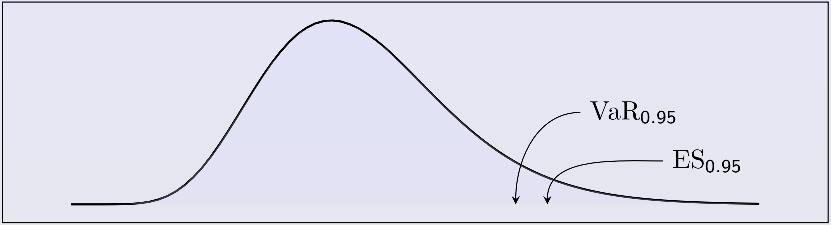
Questions and our work

- Quantitative analysis and explanation
- Comparative results for different risk measures
- Financial consequences and incentives

"Second Goodhart's law" for risk measures

As regulatory target, all risk measures cease to be good, but some risk measures, VaR in particular, are much worse than the others.

VaR and ES



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

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

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

(left-quantile)

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

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

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

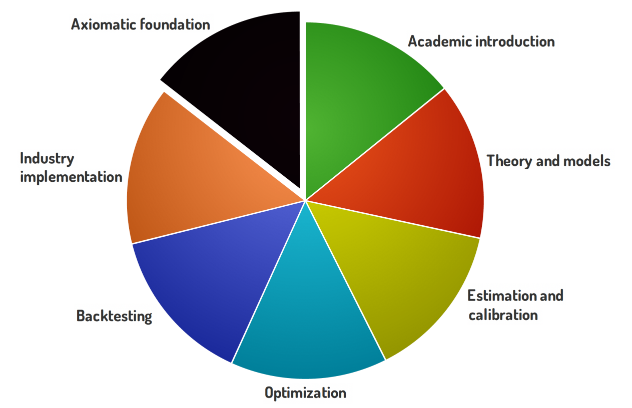
(also: TVaR/CVaR/AVaR)

VaR and ES

The ongoing co-existence of VaR and ES

- Basel IV - ES (with VaR for backtest)
  - $\mathrm{ES}_{0.975}$ replaces $\mathrm{VaR}_{0.99}$
- Solvency II - VaR
- Swiss Solvency Test - ES
- US Solvency Framework (NAIC ORSA) - both

Development of a new regulatory risk measure



- VaR axiomatization:
  Chambers'09 MF, Kou-Peng'16 OR, He-Peng'18 OR, Liu-W.'21 MOR
- ES axiomatization: W.-Zitikis'21 MS

VaR and ES

Key advantages of ES

- Coherent (Artzner-Delbaen-Eber-Heath'99)
- Capturing the tail risk (Embrechts-Liu-W.'18)
- Proper diversification (Föllmer-Schied'02)
- Convex optimization (Rockafellar-Uryasev'02)

Key advantages of VaR

- Statistical robustness (Cont-Deguest-Scandolo'10)
- Easy to forecast and backtest (Gneiting'11)

Progress

The optimization problem

General setup

- $\mathcal{G}_n = \{$measurable functions from $\mathbb{R}^n$ to $\mathbb{R}\}$
- $X \in (L^0)^n$ is an economic vector, representing all random sources
- $\mathcal{G} \subset \mathcal{G}_n$ is a decision set
- $g(X)$ for $g \in \mathcal{G}$ represents a risky position of an investor
- ρ is an objective functional mapping $\{g(X) : g \in \mathcal{G}\}$ to $\overline{\mathbb{R}}$

"The optimization problem"

to minimize $\rho(g(X))$ over $g \in \mathcal{G}$

The optimization problem

Let

$$\mathcal{G}_X(\rho) = \operatorname*{arg\,min}_{g \in \mathcal{G}} \rho(g(X)).$$

We call

- $g_X \in \mathcal{G}_X(\rho)$ an optimizing function
- $g_X(X)$ an optimized position
- $\rho(g_X(X))$: minimized risk

Uncertainty in optimization

- The optimization problem is subject to model uncertainty
- Let $\mathcal{Z}$ be a set of possible economic vectors including X
  - $\mathcal{Z}$: the set of alternative models
  - e.g. a parametric family of models (parameter uncertainty)
- The true economic vector $Z \in \mathcal{Z}$ is likely different from the perceived economic vector X
  - X: best-of-knowledge model
  - Z: true model (unknowable)
- $g_X \in \mathcal{G}_X(\rho)$ is a best-of-knowledge decision
  - true position $g_X(Z)$
  - perceived position $g_X(X)$

Uncertainty in optimization

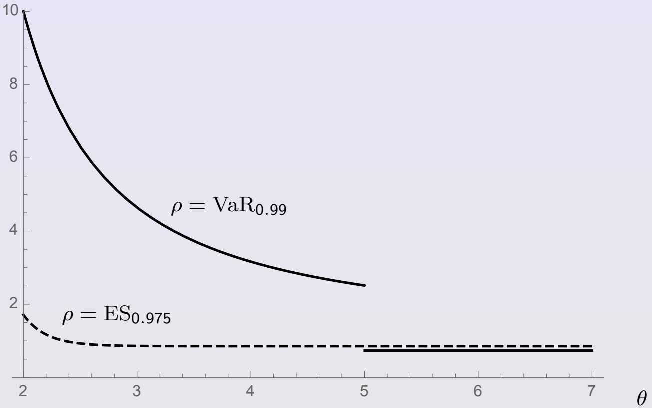


Figure: $\rho(g_X(Z))$ for $Z \sim \mathrm{Pareto}(\theta)$ and $X \sim \mathrm{Pareto}(\widehat{\theta} = 5)$. The function g_X minimizes $\rho(g(X))$ within the class of all measurable functions g satisfying $0 \leq g(x) \leq x$ and $\mathbb{E}[Xg(X)] \geq 1$.

Robust statistics

Statistical robustness addresses the question of "what if the data is compromised with small error?"

- Originally robustness is defined on estimators (estimation procedures)
- Models are at most "approximately correct" $\Rightarrow$ robustness
- (Huber-Hampel's) robustness of a statistical functional typically refers to continuity with respect to some metric

General reference: Huber-Ronchetti'07

Robustness of risk measures

- With respect to weak convergence π^W:
  - VaR_p is continuous at distributions whose quantile is continuous at p. VaR_p is argued as being almost robust.
  - ES_p is not continuous for any $\mathcal{X} \supset L^\infty$
- ES_p is continuous w.r.t. some other (stronger) metric, e.g., the L^q metric π^q, $q \geq 1$ (or the Wasserstein-L^q metric)

Robustness of risk measures

- Classic robustness: VaR and ES are applied to the same financial position.
- The regulatory choice of ρ creates certain incentives, effective before ρ is applied to assess risks.
- Once a specific ρ has been chosen, portfolios will be managed according to ρ (at least to some extend).
- In reality, VaR and ES will not be applied to the same position.

One cannot decouple the technical properties of a risk measure from the incentives it creates.

Uncertainty in optimization

We are interested in the insolvency gap

$$\underbrace{\rho(g_X(Z))}_{\text{true risk}} - \underbrace{\rho(g_X(X))}_{\text{perceived risk}}$$

not the optimality gap

$$\underbrace{\rho(g_Z(Z))}_{\text{true optimum}} - \underbrace{\rho(g_X(Z))}_{\text{true risk}}$$

or the optimality shift

$$\underbrace{\rho(g_Z(Z))}_{\text{true optimum}} - \underbrace{\rho(g_X(X))}_{\text{perceived optimum}}$$

Uncertainty in optimization

- If the modeling has good quality, $Z \approx X$ according to some metric π
- $\rho(g_X(Z)) \approx \rho(g_X(X))$ to make sense of the optimizing function g_X $\Rightarrow$ some continuity of the mapping $Z \mapsto \rho(g_X(Z))$ at $Z = X$
- We call $(\mathcal{G}, \mathcal{Z}, \pi)$ an uncertainty triplet if $\mathcal{G} \subset \mathcal{G}_n$ and $(\mathcal{Z}, \pi)$ is a pseudo-metric space of n-random vectors.
- Assume that ρ is compatible: $\rho(g(Y)) = \rho(g(Z))$ for all $g \in \mathcal{G}$ and $Y, Z \in \mathcal{Z}$ with $\pi(Y, Z) = 0$.

Robustness in optimization

Definition 1

An objective functional ρ is robust against optimization at $X \in \mathcal{Z}$ for an uncertainty triplet $(\mathcal{G}, \mathcal{Z}, \pi)$ if there exists $g_X \in \mathcal{G}_X(\rho)$ such that the function $Y \mapsto \rho(g_X(Y))$ is π-continuous at $Y = X$.

- Robustness is a joint property of the tuple $(\rho, X, \mathcal{G}, \mathcal{Z}, \pi)$
- Only a π-neighbourhood of X in $\mathcal{Z}$ matters

Robustness in optimization

Remarks.

- If ρ is robust against optimization at X for $(\mathcal{G}, \mathcal{Z}, \pi)$, then it also holds
  - for $(\mathcal{G}, \mathcal{Z}', \pi)$ if $X \in \mathcal{Z}' \subset \mathcal{Z}$;
  - for $(\mathcal{G}, \mathcal{Z}, \pi')$ if π' is stronger than π
- If $\mathcal{G}_X(\rho) = \varnothing$, then ρ is not robust at X
- Alternatives
  - One can use topologies instead of metrics
  - One can consider uncertainty on the set of probability measures instead of on the set of random vectors
  - One can require the continuity for all $g \in \mathcal{G}_X(\rho)$ instead of that for some g.

Progress

Functional optimization problems

Setup

- An n-dimensional random vector X
- Two measurable functions $v, w : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty\}$
- A measurable price density $\gamma : \mathbb{R}^n \to (0, \infty)$
- A constant $x_0 \in \mathbb{R}$

Risk minimization under budget constraint

$$\min: \ \rho(g(X)) \text{ subject to } v \leq g \leq w, \ \mathbb{E}[\gamma(X)g(X)] \geq x_0.$$

Examples

Optimal investment

- $S = (S_t)_{t\in[0,T]}$ is a d-dimensional price process with a martingale measure $\mathbb{Q}$ and price density $\gamma = \mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}$ on $\mathcal{F}_T^S$
- $\mathbb{Q}$ is unique $\Leftrightarrow$ completeness of the market
- X with $\sigma(X) = \mathcal{F}_T^S$ represents market randomness
- An investor has budget v_0 and an obligation $f(X)$ at time T
- The investor minimizes $\rho(f(X) - V_T)$, where $V_T := V_T(X)$ is the time-T value of a self-financing trading strategy V satisfying $V_0 = \mathbb{E}[\gamma(X)V_T] \le v_0$ and $v(X) \le V_T \le w(X)$
- A special case of our setting with $g(x) = f(x) - V_T(x)$

Examples

Insurance design

- Let $X \geq 0$ represent a random future loss to an insured
- f is an insurance indemnity function
- $\gamma \geq 1$ and $\gamma \mathbb{E}[f(X)]$ is the price of contract f
- y_0 is the budget of the insured
- Standard optimal insurance problem with risk measure ρ:

$$\min: \ \rho(X - f(X)) \text{ subject to } 0 \leq f(X) \leq X, \ \gamma \mathbb{E}[f(X)] \leq y_0$$

- A special case of our setting with $g(x) = x - f(x)$

Assumptions

$$\mathcal{G} = \left\{ g \in \mathcal{G}_n : v \le g \le w \text{ and } \mathbb{E}[\gamma(X)g(X)] \ge x_0 \right\}$$

Assumption G

$\mathbb{E}[\gamma(X)] < \infty$ *and* $\mathcal{G} \neq \varnothing$; *the distribution measure* μ_X *of* X *has a positive density on its support, which is a convex subset of* $\mathbb{R}^n$, *and* $(\mathcal{Z}, \pi)$ *is* (L^0_n, π^W_n) *or* (L^q_n, π^q_n), $q \in [1, \infty]$.

A special case (1-d)

$$\text{min: } \rho(g(X)) \text{ subject to } 0 \le g(X) \le X, \ \mathbb{E}[\gamma(X)g(X)] \ge x_0. \quad \text{(S)}$$

Robustness of VaR

$$\rho(X;\mathcal{G}) = \inf\{\rho(g(X)) : g \in \mathcal{G}\}$$

Assumption V

$\text{ess-sup}(v) < \rho(X;\mathcal{G}) < \rho(w(X))$ *and* γ *is bounded from above.*

Assumption V is quite general and weak.

- the lower bound v is not too large
- the optimization problem is not solved by $g = w$.
- boundedness of γ can be relaxed

Robustness of VaR

Theorem 1

For $p \in (0,1)$, under Assumptions G and V, $\rho = \mathrm{VaR}_p$ is not robust against optimization at X for $(\mathcal{G}, \mathcal{Z}, \pi)$.

- VaR_p is not robust for all commonly used metrics and a general continuously distributed X
- VaR_p has the poorest possible robustness in our setup
- Any optimizing function g_X always has a jump at the p-quantile of $g_X(X)$

Robustness of VaR

Problem (S): $0 \le g(X) \le X$

$$q = \mathrm{VaR}_p(X; \mathcal{G}) \quad \text{and} \quad Y = (X - q)\gamma(X)$$

Proposition 2

Suppose that Assumptions G and V hold, $p \in (0,1)$, $\rho = \mathrm{VaR}_p$, $\mathbb{E}[\gamma(X)X] < \infty$ and $\mathbb{P}(Y \le \mathrm{VaR}_p(Y)) = p$. Problem (S) admits a μ_X-a.s. unique solution of the form

$$g_X(x) = x\mathbb{1}_{\{(x-q)\gamma(x)>c\}} + (x \wedge q)\mathbb{1}_{\{(x-q)\gamma(x)\le c\}},$$

where $c = \mathrm{VaR}_p(Y)$. Moreover, $p\mathrm{ES}_{1-p}(-Y_+) = x_0 - \mathbb{E}[\gamma(X)X]$.

Robustness of convex risk measures

Assumption P

The functions γ, v and w are μ_X-a.e. continuous and $\gamma(X)$ has a continuous density. Moreover, $-\infty \le \mathbb{E}[\gamma(X)v(X)] \le x_0 \le \mathbb{E}[\gamma(X)w(X)] \le \mathbb{E}[|\gamma(X)w(X)|] < \infty$.

Convex risk measures (Follmer-Schied'02)

- monotone
- cash invariant
- convex

Robustness of convex risk measures

A divergence risk measure is defined as

$$\rho(Y) := \sup_{\mathbb{Q} \ll \mathbb{P}} \left(\mathbb{E}_{\mathbb{Q}}[Y] - I_{\varphi}(\mathbb{Q}|\mathbb{P}) \right), \qquad Y \in L^{\infty}, \tag{1}$$

where

$$I_{\varphi}(\mathbb{Q}|\mathbb{P}) = \int \varphi\left(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right) \mathrm{d}\mathbb{P}, \tag{2}$$

is the φ-divergence of $\mathbb{Q}$ to $\mathbb{P}$, for a proper closed convex function $\varphi : \mathbb{R} \to [0, +\infty]$ with $0 = \varphi(1) = \min_x \varphi(x)$.

- $\varphi(x) = x \log x - x + 1$: I_{φ} is the relative entropy and ρ is an entropic risk measure
- $\varphi = \infty \cdot \mathbb{1}_{[1/(1-p),\infty)}$: $\rho = \mathrm{ES}_p$

Robustness of convex risk measures

Theorem 3

In addition to Assumptions G and P we assume that v and w are bounded. Then the divergence risk measure ρ is robust against optimization at $X \in L^0_n$ for $(\mathcal{G}, L^0_n, \pi^W_n)$.

- f has growth index q: $|f(x)| \le c(1+|x|^q)$ for some $c > 0$

Corollary 4

In addition to Assumptions G and P, we assume that both v and w have growth index $q \in [1, \infty]$. Then ES_p for $p \in (0,1)$ is robust against optimization at $X \in L^q_n$ for $(\mathcal{G}, L^q_n, \pi^q_n)$.

- Sharp contrast between VaR and ES

Robustness of convex risk measures

For ES, there exists a minimizer g_X that has one of the following two forms, where $z^* \in \mathbb{R}$ and $c > 0$ are suitable constants:

$$g_X(x) = (v(x) \vee z^* \wedge w(x))\mathbb{1}_{\{0<c\gamma(x)<1\}}$$

or

$$g_X(x) = (v(x) \vee z^* \wedge w(x))\mathbb{1}_{\{c\gamma(x)>1\}}.$$

Robustness of convex risk measures

A utility-based shortfall risk measure is given by

$$\rho(Y) = \inf\left\{m \in \mathbb{R} : \mathbb{E}[\ell(Y-m)] \le x_0\right\}, \qquad Y \in L^\infty$$

where $\ell : \mathbb{R} \to \mathbb{R}$ is increasing and convex and $x_0 \in \operatorname{int} \ell(\mathbb{R})$.

Theorem 5

In addition to Assumptions G and P we assume that v and w are bounded. Then the utility-based shortfall risk measure ρ is robust against optimization at $X \in L^0_n$ for $(\mathcal{G}, L^0_n, \pi^W_n)$.

- Similar results are obtained for expected utility maximization

Robustness of convex risk measures

The expectile of $Y \in L^1$ at level $\tau \in [0,1]$ is the unique solution to the equation

$$\tau\mathbb{E}[(Y-z)_+] = (1-\tau)\mathbb{E}[(Y-z)_-],$$

which is a shortfall risk measure with $\ell(x) = \tau x_+ - (1-\tau)x_-$.

Corollary 6

In addition to Assumptions G and P, we assume that both v and w have growth index $q \in [1,\infty]$. Then the expectile at level $\tau \in (1/2, 1]$ is robust against optimization at $X \in L^q_n$ for $(\mathcal{G}, L^q_n, \pi^q_n)$.

Robustness of convex risk measures

Problem (S): $0 \le g(X) \le X$

Proposition 7

Let $p \in (0,1)$ and $\rho = \mathrm{ES}_p$. Suppose that γ is μ_X-a.e. continuous, $\gamma(X)$ has a continuous density, and $0 \le x_0 < \mathbb{E}[\gamma(X)X]$. There exist constants $d > 0$ and $r \ge 0$ such that the function

$$g_X(x) = x \mathbb{1}_{\{\gamma(x) > d\}} + (x \wedge r)\mathbb{1}_{\{\gamma(x) \le d\}}, \quad x \in \mathbb{R}, \tag{3}$$

solves Problem (S). Moreover, r is a p-quantile of $g_X(X)$.

Progress

Simulation results

- Problem (S)
- $Z \sim \text{Expo}(\theta)$ or $Z \sim \text{Pareto}(\theta)$ with unknown parameter $\theta > 0$
- Estimate $\widehat{\theta}$ for θ
- $X \sim \text{Expo}(\widehat{\theta})$ or $X \sim \text{Pareto}(\widehat{\theta})$
- Minimizes $\rho(g(X))$ in Problem (S)
- $\rho = \mathrm{VaR}_{0.99}$ and $\rho = \mathrm{ES}_{0.975}$ (Basel III)
- $\gamma(x) = x$

Simulation results

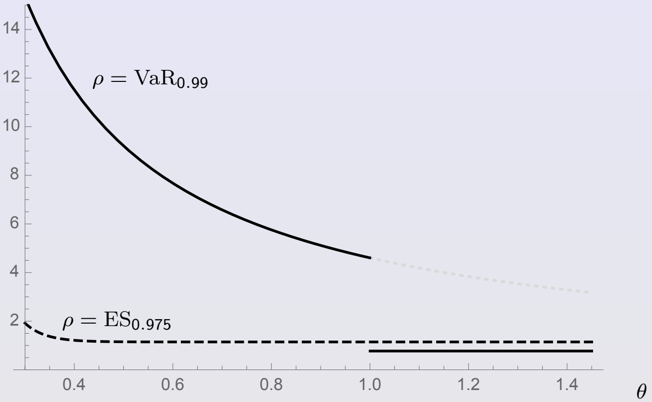


Figure: $\rho(g_X(Z))$ for $Z \sim \mathrm{Expo}(\theta)$ and $X \sim \mathrm{Expo}(\widehat{\theta} = 1)$. The dotted grey curve corresponds to the VaR of the unoptimized position, $\mathrm{VaR}_{0.99}(Z) \approx \mathrm{ES}_{0.975}(Z)$.

Simulation results

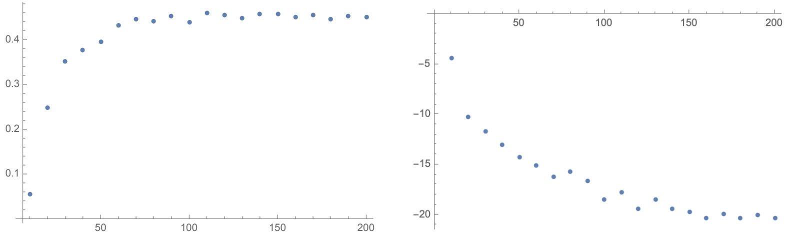


Figure: Mean-squared errors $|\rho(g_X(Z)) - \rho(g_X(X))|^2$ of 10,000 independent sample points of $\rho(g_X(Z))$ and $\rho(g_X(X))$, each with a maximum likelihood estimator $\widehat{\theta}$ computed from n iid realizations of the Pareto(5)-distributed risk factor Z. The horizontal axis shows the number n. The case $\rho = \mathrm{VaR}_{0.99}$ can be found on the left, $\rho = \mathrm{ES}_{0.975}$ is on the right, both in log scale.

Simulation results

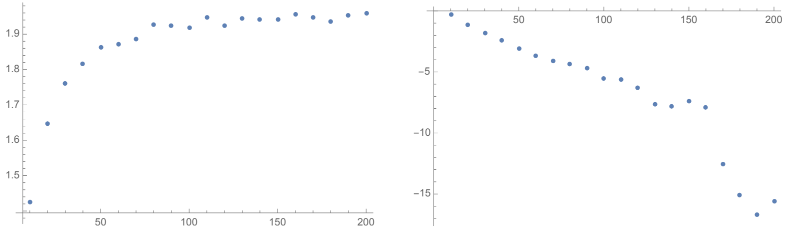


Figure: Mean-squared errors $|\rho(g_X(Z)) - \rho(g_X(X))|^2$ of 10,000 independent sample points of $\rho(g_X(Z))$ and $\rho(g_X(X))$, each with a maximum-likelihood estimator $\widehat{\theta}$ computed from n iid realizations of the Exp(1)-distributed risk factor Z. The horizontal axis shows the number n. The case $\rho = \mathrm{VaR}_{0.99}$ can be found on the left, $\rho = \mathrm{ES}_{0.975}$ is on the right, both in log scale.

Progress

Conclusion

On robustness in optimization:

$$\text{VaR} \lll \text{ES}$$

VaR optimized position for Problem (S)

$$g_X(X) = X\mathbb{1}_{\{(X-q)\gamma(X)>c\}} + (X \wedge q)\mathbb{1}_{\{(X-q)\gamma(X)\le c\}}$$

Observations.

- The discontinuity in $Z \mapsto g_X(Z)$ comes from the fact that optimizing VaR is "too greedy": always ignores tail risk, and hopes that the probability of the tail risk is correctly modelled.

Conclusion

Is risk positions of type g_X realistic?

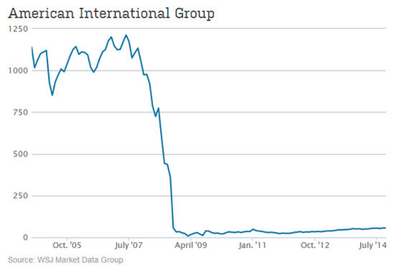
> *"Starting in 2006, the CDO group at UBS noticed that their risk-management systems treated AAA securities as essentially riskless even though they yielded a premium (the proverbial free lunch). So they decided to hold onto them rather than sell them."*

- From Feb 06 to Sep 07, UBS increased investment in AAA-rated CDOs by more than 10 times; many large banks did the same.
  - Take a risk of big loss with small probability
  - Treat it as free money - profit
  - Model uncertainty?

---

quoted from Acharya-Cooley-Richardson-Walter'10

AIG



CEO of AIG Financial Products, August 2007:

"*It is hard for us, without being flippant, to even see a scenario within any kind of realm of reason that would see us losing one dollar in any of those transactions.*"

- AIGFP sold protection on super-senior tranches of CDOs
- US Financial Crisis Inquiry Commission'11: due to unhedged CDS positions

Other questions

Many other questions ...

- other risk measures
- other optimization settings
- connection to distributionally robust optimization
- risk measures as constraints instead of objectives

Thank you

Paul Embrechts
(ETH Zurich)

Alexander Schied
(Waterloo)

- Embrechts-Schied-W., Robustness in the optimization of risk measures
  Operations Research, 2021. SSRN: 3254587

Connection to distributionally robust optimization

Distributionally robust optimization, for $\epsilon > 0$:

$$\text{to minimize:} \quad \sup_{\pi(Y,X)\le\epsilon} \rho(g(Y)) \quad \text{subject to } g \in \mathcal{G}.$$

- $\mathcal{G}_X(\rho,\epsilon)$: the set of functions $g \in \mathcal{G}$ solving this problem
- $\mathcal{G}_X(\rho,0) = \mathcal{G}_X(\rho)$, the original setting
- ρ is robust for the ϵ-problem if there exists $g_X \in \mathcal{G}_X(\rho,\epsilon)$ such that $Z \mapsto \rho(g_X(Z))$ is π-continuous at $Z = X$

e.g., Natarajan-Pachamanova-Sim'08, Zhu-Fukushima'09,

Connection to distributionally robust optimization

Problem: to minimize (1-d)

$$\sup_{\pi^\infty(Y,X)\le\epsilon} \mathrm{VaR}_p(g(Y)) \text{ subject to } g \in \mathcal{G},$$

where $\mathcal{G} = \{g \in \mathcal{G}_1 : \mathbb{E}[\gamma(X)g(X)] \ge x_0,\ 0 \le g \le m\}$. Let

$$q_\epsilon = \inf\left\{\sup_{\pi^\infty(Y,X)\le\epsilon} \mathrm{VaR}_p(g(Y)) : g \in \mathcal{G}\right\}.$$

Assumption D

$q > 0$, $1/2 \le p < 1$, X *has a decreasing density on its support and* γ *is an increasing function of* X.

Connection to distributionally robust optimization

Proposition 8

Under Assumption D, the above problem admits a solution of the form

$$g_X(x) = m\mathbb{1}_{\{x>c+\epsilon\}} + q_\epsilon \mathbb{1}_{\{x\le c+\epsilon\}}, \quad x \in \mathbb{R}, \quad \textit{where } c = \mathrm{VaR}_p(X).$$

- $Z \mapsto \mathrm{VaR}_p(g_X(Z))$ is π^∞-continuous at $Z = X$
- VaR_p is robust for the ϵ-problem
- The ϵ-modification improves the robustness of VaR
- We still get the big-loss-small-probability type of optimizer