Robustness in risk measurement: the impact of incentives

Alexander Schied
University of Waterloo

Bachelier Finance Society
One World Seminar

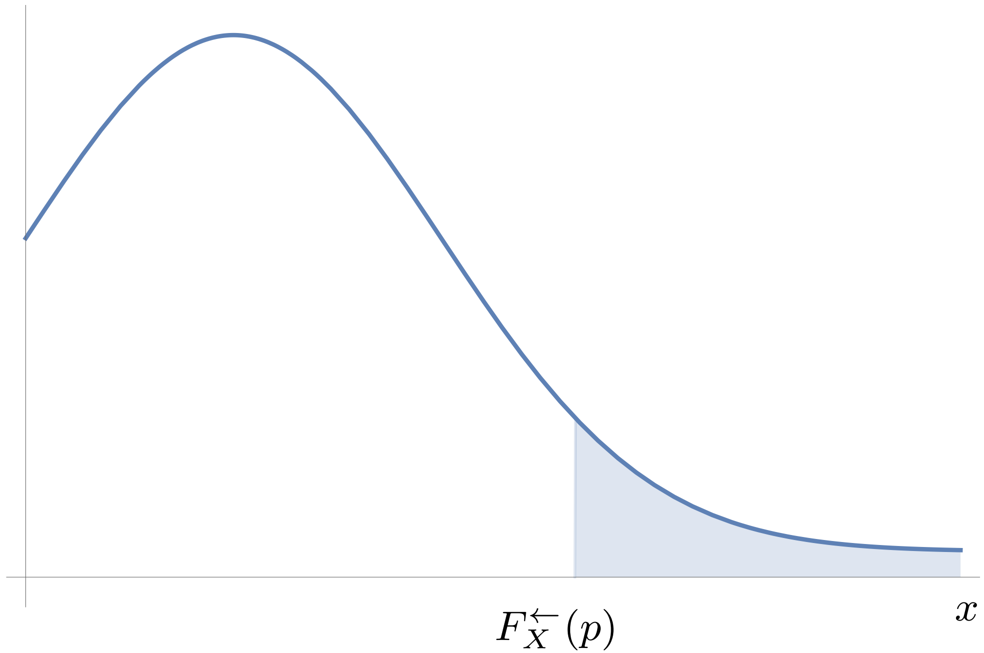
February 11, 2021

Joint work with
Paul Embrechts and Ruodu Wang

1 Introduction

The risk of a portfolio or an entity is often measured by the Value at Risk: If the random variable X describes the future loss and $p \in (0,1)$, then

$$\mathrm{VaR}_p(X) = F_X^{\leftarrow}(p) = \inf\{x \in \mathbb{R} : \mathbb{P}[X \leq x] \geq p\}$$



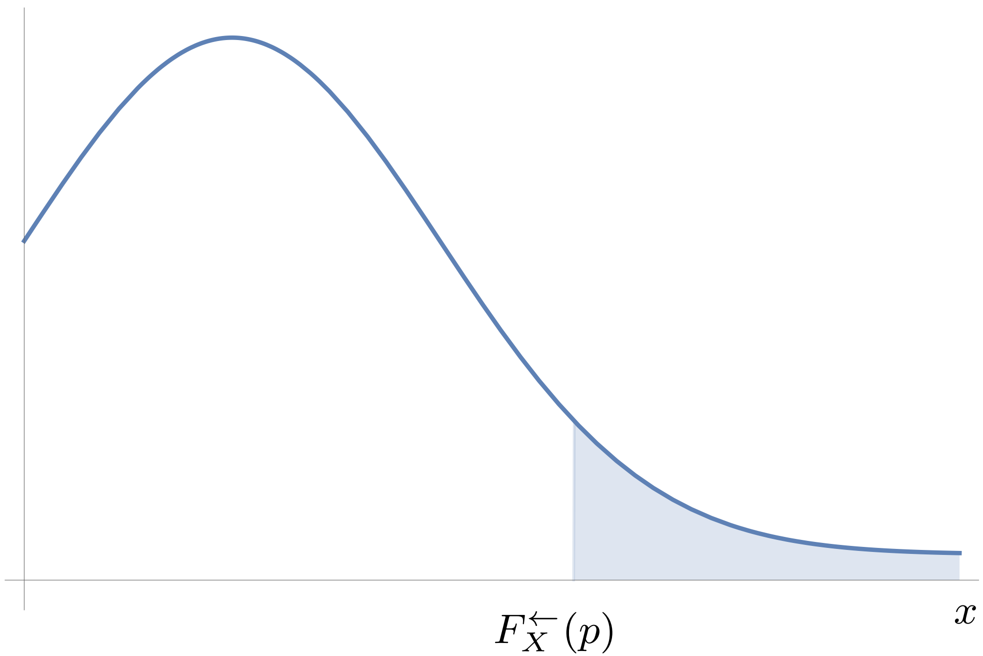
Artzner, Delbaen, Eber & Heath (1999): VaR is problematic, because:

- it discourages diversification
- it encourages concentration of risk on scenarios with small (model) probability

The most popular alternative to VaR is Expected Shortfall (also called AVaR, TailVaR, TVaR, CVaR, ...):

$$\mathrm{ES}_p(X) = \frac{1}{1-p}\int_{1-p}^{1} \mathrm{VaR}_t(X)\,dt = \mathbb{E}[X \mid X \geq \mathrm{VaR}_p(X)]$$

(where the latter identity holds, e.g., if F_X is continuous).



In contrast to VaR, Expected Shortfall encourages diversification and takes the sizes of large losses into account. It is a coherent risk measure in the sense of Artzner et al. (1999).

In **Basel IV**:

$$\mathrm{VaR}_{0.99} \qquad \text{was replaced with} \qquad \mathrm{ES}_{0.975}$$

(with VaR still kept for backtesting), but this and similar moves are also meeting some resistance.

More recently, the following two technical issues have been discussed:

- The possibility to backtest (Gneiting, Ziegel, Acerbi ...)
- Robustness properties for estimating the risk measure from historical data or from Monte Carlo simulations.

  Cont, Deguest & Scandolo (2010):

  - estimating VaR is (essentially) qualitatively robust (in the sense of Hampel (1971))
  - estimating ES is not qualitatively robust
  - more generally, estimating any law-invariant coherent risk measure is not qualitatively robust

Here, we will focus on some aspects of the robustness of VaR and ES.

Qualitative robustness, informally

Let

$$\widehat{\rho}_{n,\mu}$$

denote an estimator of $\rho(X)$ computed from n i.i.d. random variables $X_1, \dots, X_n$ with law μ

The estimator is **qualitatively robust** if

$$\widehat{\rho}_{n,\mu} \qquad \text{and} \qquad \widehat{\rho}_{n,\nu} \qquad \text{are close in law provided that } \mu \text{ and } \nu \text{ are close}$$

E.g., robustness may lead to more stable margin requirements

Cont, Deguest & Scandolo (2010): Estimating VaR_p is (basically) qualitatively robust while estimating any coherent risk measure ρ is not qualitatively robust

Example 1. The median ($= \mathrm{VaR}_{1/2}$) is basically qualitatively robust, whereas estimating the mean (corresponding to $\rho(X) = \mathbb{E}[X]$) is not.

Sounds like a strong argument in favor of VaR and against coherent risk measures and ES in particular. But let's review the argument:

- It is technical and does not address the main objection against VaR: VaR may discourage diversification and encourage concentration of risk on scenarios with perceived small probability

- **Hampel's theorem:** Qualitative robustness of an estimator is basically equivalent to the continuity of that estimator with respect to weak convergence.

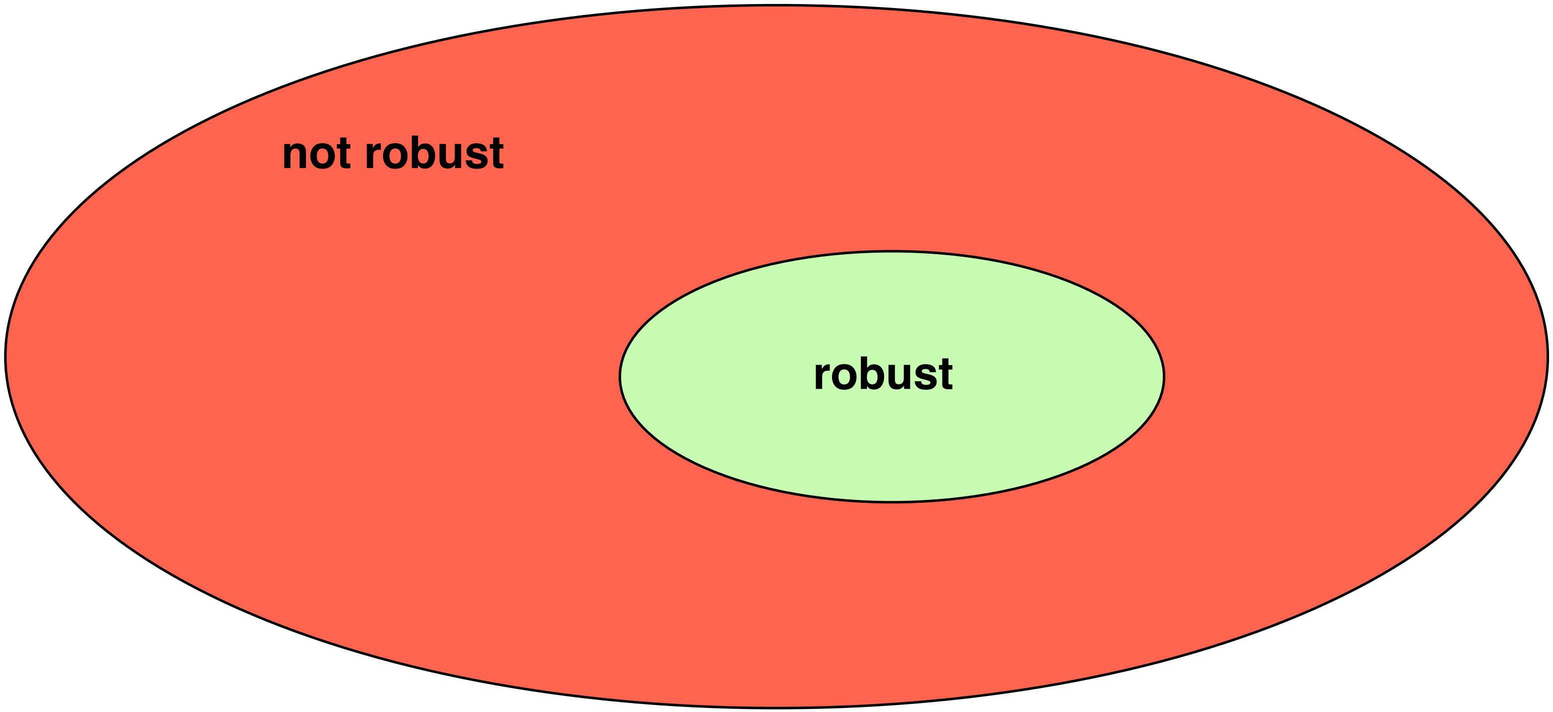
  Since the compactly supported probability measures are dense with respect to weak convergence, Hampel's theorem thus implies that a qualitatively robust risk measure must necessarily be insensitive to tail behavior.

  More generally, there is a tradeoff between robustness and sensitivity. In risk management, regulators may prefer a risk measure that is **sensitive** to outliers in the form of crash scenarios.

  But if, e.g., $\mathbb{P}[A] < 1 - p$, then

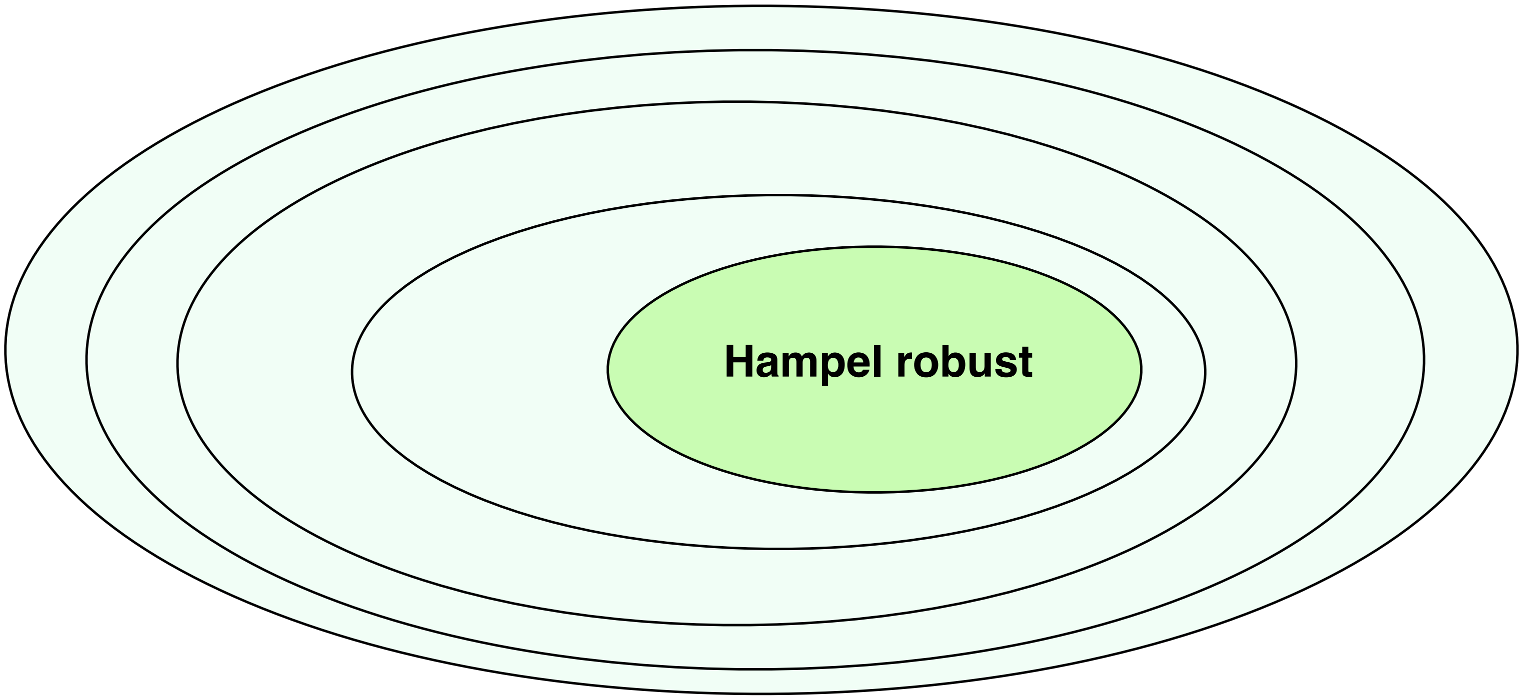
$$\mathrm{VaR}_p(c\mathbb{1}_A) = 0 \qquad \text{for all } c.$$

- Hampel's terminology of qualitative robustness generates a sharp division of estimators into those that are called "robust" and others that are called "not robust".



But, e.g., estimating the expected value should be "more robust" than estimating variance.

In Krätschmer, A.S. & Zähle (2012, 2014, 2017), we developed a **refined** notion of robustness by replacing weak convergence with stronger concepts, such as convergence in the L^p-Wasserstein distance. This gives a picture like the following one:



Thus, robustness is not lost entirely but only to some degree when Value at Risk is replaced by a coherent risk measure. It also captures to some degree the natural tradeoff between robustness and tail-sensitivity. The concept is not limited to risk measures.

- When comparing the robustness of VaR and ES, we estimate their values on the same loss profile X.

  But this is not what is happening in reality:

  - The regulatory choice of one risk measure creates certain **incentives.**
  - These incentives become effective even before the risk measure has ever been applied in measuring the risk.

    That is, once a specific risk measure has been chosen, positions will be optimized with respect to that risk measure (at least to some extent). For instance:

    * maxing out leverage
    * slicing the senior tranche of an ABS so that it qualifies for a AAA rating
  - Thus, in reality, VaR and ES will not be applied to the same position.
  - *Therefore, one cannot decouple the technical properties of a risk measure from the incentives it creates.*

These observations are the starting point for our talk. It will lead us to a re-evaluation of the robustness properties of VaR and ES.

The basic problem formulation

Our ingredients:

- a risk factor X is obtained from a best-of-knowledge model
- a regulatory risk measure ρ is imposed
- an agent takes the position $g_X(X)$ obtained by minimizing the risk $\rho(g(X))$ under certain constraints on g.
- the true model Z, however, is **unknowable**

Robustness

Is the true risk

$$\rho(g_X(Z))$$

close to the perceived risk

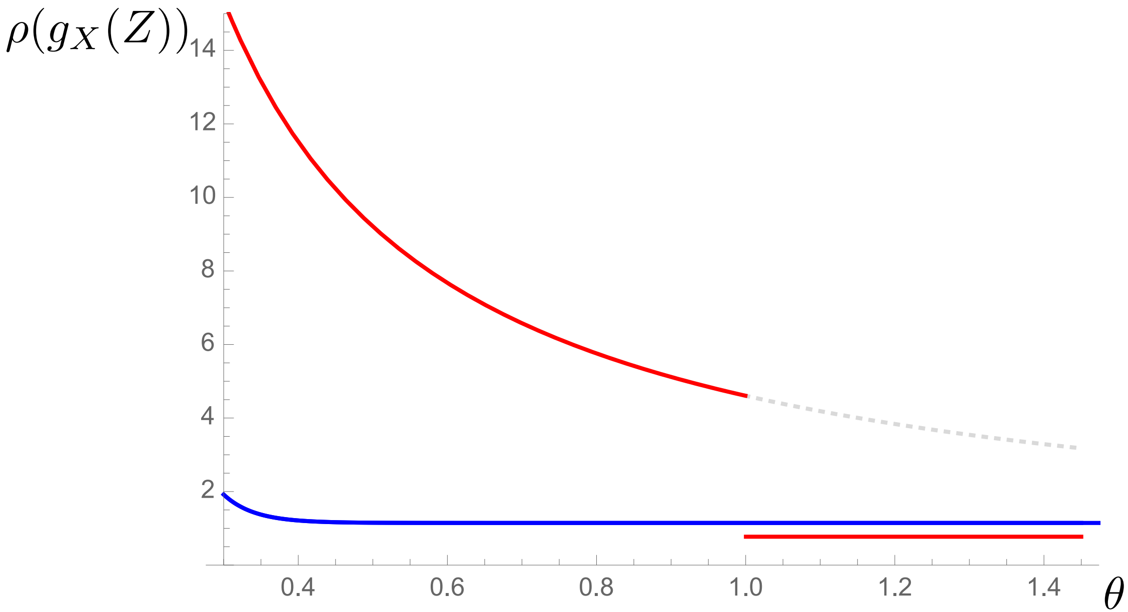
$$\rho(g_X(X))$$

if Z is close to X?

A toy model

Suppose that X has an exponential distribution with estimated parameter $\widehat{\theta} = 1$

The agent finds g_X by minimizing $\rho(g(X))$ under the budget constraint $\mathbb{E}[\gamma(X)g(X)] \geq 1$ and the additional constraint $0 \leq g(x) \leq x$ for the 'price density' $\gamma(x) = x$



True risk $\rho(g_X(Z))$ for $\rho = \mathrm{ES}_{0.975}$ and $\rho = \mathrm{VaR}_{0.99}$ when $Z \sim \mathrm{Exp}(\theta)$

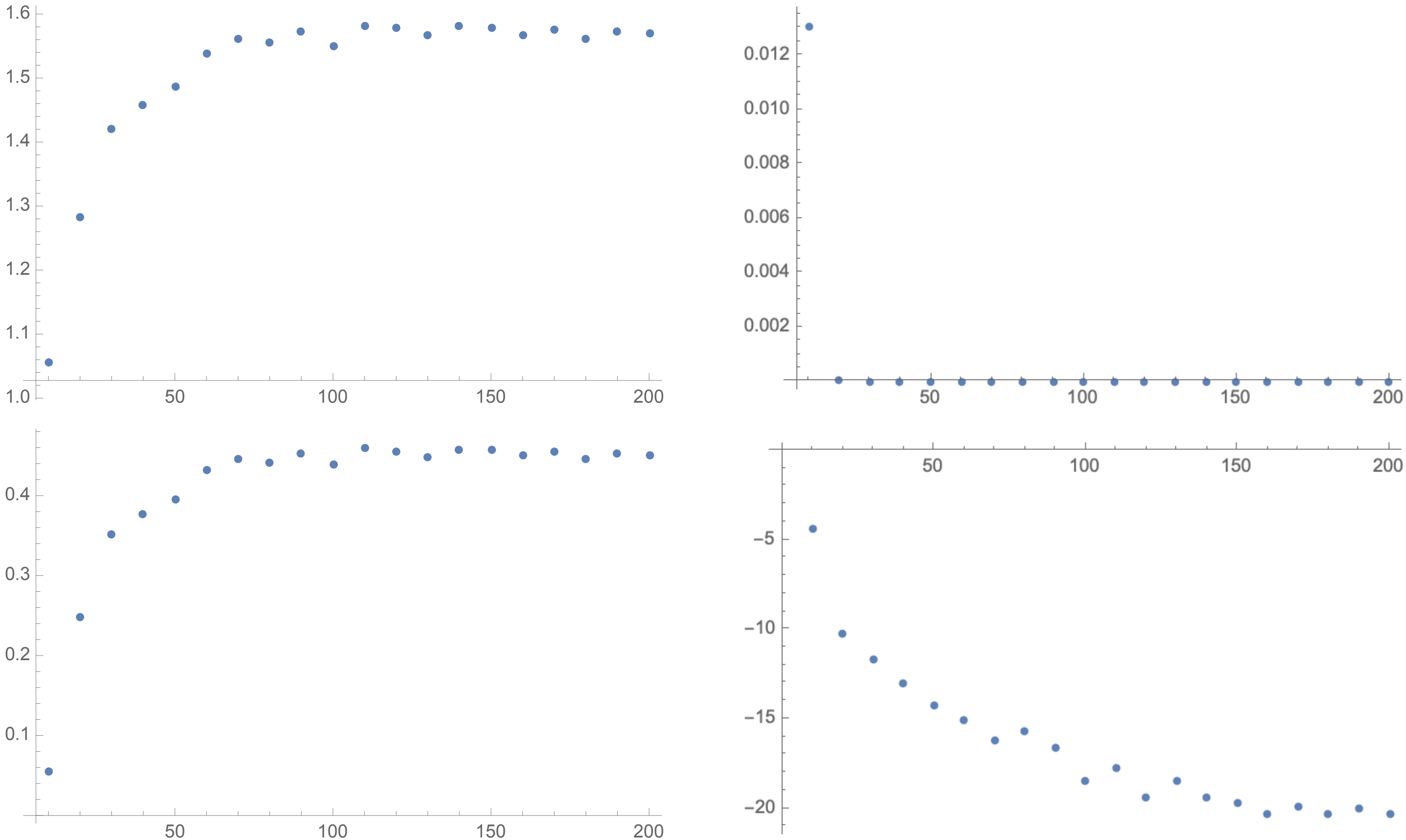
- For VaR, a 0.1% error in estimating θ leads to a 400% risk increase
- For ES, the true risk $\mathrm{ES}_{0.975}(g_X(Z))$ deviates from the model value $\mathrm{ES}_{0.975}(g_X(X))$ by only 0.5% as long as θ is within $\pm 50\%$ of the estimated value $\widehat{\theta} = 1$

The same issue viewed from a different angle

- Suppose that Z has a Pareto distribution with parameter $\theta = 5$
- The agent, unaware of the true value of θ, uses n i.i.d. copies $Z_1, \dots, Z_n$ of Z for a maximum likelihood estimator $\widehat{\theta}$
- $X \sim \text{Pareto}(\widehat{\theta})$ will then serve as model for the true risk factor
- The agent then determines g_X as minimizer of $\rho(g_X(X))$ subject to the constraints $0 \leq g(x) \leq x$ and $\mathbb{E}[\gamma(X)g(X)] \geq 1$ for $\gamma(x) = x$
- Then we compare the true risk value $\rho(g_X(Z))$ to the perceived risk value $\rho(g_X(X))$
- For each n, we repeat this procedure 10,000 times and compute the mean-squared error, i.e., the average of

$$|\rho(g_X(Z)) - \rho(g_X(X))|^2$$

  of all 10,000 sample points
- As the number n of i.i.d. realizations of Z increases, the estimate $\widehat{\theta}$ becomes ever more accurate, and we may expect the mean-squared error of the risk differences to decrease



Empirical mean-squared error as a function of n for $\mathrm{VaR}_{0.99}$ (left) and $\mathrm{ES}_{0.975}$ (right). Original values top, log-values bottom

2 Formal setup and mathematical results

In the sequel, $X = (X_1, \dots, X_n)$ will denote an n-dimensional random vector describing all random sources in an economic model (e.g., all prices of risky assets in a discrete-time market model).

$\mathscr{G}$ will be a class of admissible functions $g : \mathbb{R}^n \to \mathbb{R}$. E.g., all hedged positions of X_n satisfying a certain budget constraint.

ρ will be a given risk measure

The goal of the agent is to

(1) $$\text{minimize } \rho(g(X)) \text{ over } g \in \mathscr{G}.$$

The optimization problem

$$\text{minimize } \rho(g(X)) \text{ over } g \in \mathscr{G} \tag{1}$$

may be subject to significant model uncertainty resulting from the assumptions made on X.

Let $\mathscr{Z}$ be a class of possible models, including X.

- X best-of-knowledge model
- $Z \in \mathscr{Z}$ true model (unknown and typically $\neq X$)

Then

$$g_X(X)$$

will be the perceived position, but

$$g_X(Z)$$

will be the real position

If the model is good, X and Z are close to each other with respect to some distance d.

In this case, the risk $\rho(g_X(Z))$ of the real position should be close to the risk $\rho(g_X(X))$ of the perceived position.

We therefore desire continuity of the map $Z \to \rho(g_X(Z))$ at $Z = X$ in a (semi-)metric d

Definition 1. ρ is robust against optimization at $X \in \mathscr{X}$ with respect to d if there exists $g_X \in \mathscr{G}$ such that g_X solves (1) and $Z \to \rho(g_X(Z))$ is continuous at $Z = X$.

Note that we are interested in the solvency gap

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

not the *optimality gap*

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

nor the *optimality shift*

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

In the sequel, we will make the following specific assumptions:

- $\mathscr{X} = L^q$ with $q \in \{0\} \cup [1, \infty]$ and $d = \|\cdot\|_q$ for $q \geq 1$ and convergence in law for $d = 0$
- $X = (X_1, \ldots, X_n)$ and $\mathbb{P} \circ X^{-1}$ has a convex support and admits a density
- There are measurable functions
  - $v : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty\}$
  - $w : \mathbb{R}^n \to \mathbb{R}$
  - $\gamma : \mathbb{R}^n \to (0, \infty)$ with $\mathbb{E}[\gamma(X)] < \infty$
- such that the set $\mathscr{G}$ is of the form

$$\mathscr{G} = \Big\{ g : v \leq g \leq w \text{ and } \mathbb{E}[g(X)\gamma(X)] \geq x_0 \Big\} \neq \emptyset$$

Thus, our optimization problem is

$$\text{(2)} \qquad \text{minimize: } \rho(g(X)) \qquad \text{subject to } v \leq g \leq w, \quad \mathbb{E}[\gamma(X)g(X)] \geq x_0.$$

We assume that it is not trivial in the sense that

$$\operatorname{ess\,sup}(v) < \inf_{g \in \mathscr{G}} \rho(g(X)) < \rho(w(X))$$

2.1 Value at Risk

Theorem 1. *Suppose that γ is bounded from above. Then* VaR_p *is* **not robust against optimization** *at X for every $p \in (0,1)$ and neither for convergence in law nor for L^q-convergence with $q \in [1,\infty]$.*

The preceding result can be seen as an example of Goodhart's law (Goodhart 1975) as paraphrased by Strathern (1997):

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

But we are also going to see that some other risk measures do much better than VaR.

2.2 Two classes of convex risk measures

We make the following standing assumption:

Assumption 1. The functions, γ, v and w are $\mathbb{P} \circ X^{-1}$-a.e. continuous. Moreover,

$$-\infty \leq \mathbb{E}[\gamma(X)v(X)] \leq x_0 \leq \mathbb{E}[\gamma(X)w(X)] \leq \mathbb{E}[|\gamma(X)w(X)|] < \infty$$

2.2.1 Divergence risk measures

Let $\varphi : \mathbb{R} \to [0, +\infty]$ be a proper closed convex function whose effective domain is an interval with endpoints $a < b$. We assume moreover that $a < 1 < b$ and that $0 = \varphi(1) = \min_x \varphi(x)$. Then the φ-divergence of a probability measure $\mathbb{Q}$ with respect to $\mathbb{P}$ is

(3)
$$I_\varphi(\mathbb{Q}|\mathbb{P}) := \begin{cases} \int \varphi\Big(\frac{d\mathbb{Q}}{d\mathbb{P}}\Big) d\mathbb{P} & \text{if } \mathbb{Q} \ll \mathbb{P}, \\ +\infty & \text{otherwise.} \end{cases}$$

The corresponding divergence risk measure is defined as

(4)
$$\rho(Y) := \sup_{\mathbb{Q} \ll \mathbb{P}} \big(\mathbb{E}_\mathbb{Q}[Y] - I_\varphi(\mathbb{Q}|\mathbb{P})\big), \qquad Y \in L^\infty.$$

Important special cases:

- The entropic risk measure for $\varphi(x) = x \log x - x + 1$
- The Expected Shortfall, ES_p, for $\varphi = \infty \cdot \mathbb{1}_{[1/(1-p),\infty)}$

Theorem 2. *We assume in addition that v and w are bounded. Then the divergence risk measure ρ is robust against optimization at $X \in L^0$ for convergence in law.*

Theorem 2 relies on a duality formula from Ben-Tal & Teboulle (1987, 2007), which has only been established on L^∞. This is one of the reasons for assuming the boundedness of v and w. It is possible, however, to relax their boundedness by imposing a suitable growth condition, given that the aforementioned duality formula extends to L^p. An important special case for which this is possible is Expected Shortfall, ES_p.

We will say that a function $f : \mathbb{R}^k \to \mathbb{R}$ has growth index $q \in [0, \infty]$, if f is locally bounded for $q = \infty$ and if for $q < \infty$ there exists a constant c such that $|f(x)| \le c(1 + |x|^q)$ for $x \in \mathbb{R}^k$.

Corollary 1. *We assume in addition that both v and w have growth index $q \in [1, \infty]$. Then Expected Shortfall, ES_p, with $p \in (0, 1)$, is robust against optimization at $X \in L^q$ for L^q-convergence.*

Idea of proof of Theorem 2 and Corollary 1:

- Let

$$\ell(x) := \sup_{y \geq 0}(xy - \varphi(y))$$

For instance,

$$\ell(x) = \frac{x^+}{1-p} \qquad \text{for } \rho = \mathrm{ES}_p$$

- Ben-Tal & Teboulle (1987, 2007):

$$\rho(Y) = \inf_{z \in \mathbb{R}} \big(\mathbb{E}[\ell(Y+z)] - z\big), \qquad Y \in L^\infty. \tag{5}$$

- Show that the infimum in (5) is actually attained.

- Let z^* be such that

$$\rho(g_X(X)) = \mathbb{E}[\ell(g_X(X) + z^*)] - z^*$$

Then,

$$\inf_{g \in \mathscr{G}} \rho(g(X)) \leq \inf_{g \in \mathscr{G}} \big(\mathbb{E}[\ell(g(X) + z^*)] - z^*\big) \leq \mathbb{E}[\ell(g_X(X) + z^*)] - z^* = \inf_{g \in \mathscr{G}} \rho(g(X)).$$

- Hence, g_X minimizes $\mathbb{E}[\ell(g(X) + z^*)]$ over $g \in \mathscr{G}$.

- We'll see below that the solution to the latter problem is continuous $\mathbb{P} \circ X^{-1}$-a.e. □

2.2.2 Utility-based shortfall risk

Let $\ell : \mathbb{R} \to \mathbb{R}$ be a nonconstant, increasing, and convex loss function and x_0 be an interior point in the range of ℓ. The corresponding utility-based shortfall risk measure is given by

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

Theorem 3. *We assume in addition that v and w are bounded. Then the utility-based shortfall risk measure ρ is robust against optimization at $X \in L^0$ for convergence in law.*

A notable special case of a utility-based shortfall risk measure is the expectile of $Y \in L^1$ at level $\tau \in (1/2, 1]$, defined as the unique solution to the equation

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

as it corresponds to the loss function $\ell(x) = \tau x_+ - (1 - \tau)x_-$

Corollary 2. *We assume in addition 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 $X \in L^q$ for L^q-convergence.*

Idea of proof of Theorem 3:

- $z^* := \rho(g_X(X))$ is the unique solution to the equation $\mathbb{E}[\ell(g_X(X) - z)] = x_0$.
- Hence, g_X minimizes $\mathbb{E}[\ell(g(X) - z^*)]$ over $g \in \mathscr{G}$. Indeed, suppose by way of contradiction that there is $g_0 \in \mathscr{G}$ for which

$$\mathbb{E}[\ell(g_0(X) - z^*)] < \mathbb{E}[\ell(g_X(X) - z^*)]$$

  Then the solution, $z_0 = \rho(g_0(X))$, of the equation $\mathbb{E}[\ell(g_0(X) - z)] = x_0$ will be strictly smaller than z^*, a contradiction to the optimality of g_X.
- As before, we'll see in the next section that the solution to the latter problem is continuous $\mathbb{P} \circ X^{-1}$-a.e. $\square$

2.3 Expected utility or loss

We have seen in the proof sketches that the optimizers g_X can also be obtained as a minimizer of the expected loss defined through

$$\rho_\ell(Y) := \mathbb{E}[\ell(Y)],$$

where $\ell : \mathbb{R} \to \mathbb{R}$ is a nonconstant, nondecreasing, and convex function. Up to a sign change and a constant shift, minimizing an expected loss is equivalent to maximizing an expected utility.

This is a classical problem if ℓ is smooth and strictly convex. However, recall that

$$\ell(x) = \frac{x^+}{1-p} \qquad \text{for } \rho = \mathrm{ES}_p$$

which is neither smooth nor strictly convex.

Proposition 1. *Let ℓ^+ have growth index $q^+ \in [1, \infty]$ and suppose that w has growth index $p \in [1, \infty]$. If, moreover, ℓ is not bounded from below, let ℓ^- have growth index $q^- \in [0, 1]$ and suppose that the growth index r of v satisfies $r < \infty$ if $q^- = 0$ and $r \leq pq^+/q^-$ otherwise. Let*

$$I(z) := \inf\{y : \ell'_-(y) > z\} = \sup\{y : \ell'_-(y) \leq z\}$$

denote the right-continuous generalized inverse function of ℓ'_-. Then the problem of minimizing $\rho_\ell(g(X))$ over $g \in \mathscr{G}$ has a solution that takes one of the following forms:

$$g_X(x) = \begin{cases} v(x) \vee I(c\gamma(x)) \wedge w(x) & \textit{for some } c, \textit{ or} \\ v(x) \vee I(a) \wedge w(x) & \textit{for } a := \inf_y \ell'_-(y) \geq 0, \textit{ or} \\ (v \vee I(a) \wedge w)\mathbb{1}_{\{\gamma > c_0\}} + v\mathbb{1}_{\{\gamma \leq c_0\}} & \textit{for some constant } c_0. \end{cases} \tag{6}$$

By our assumptions, all possibilities in (6) are $\mathbb{P} \circ X^{-1}$-a.e. continuous functions. This provides the essential step still missing in the proofs of Theorems 2 and 3 and also yields the following result:

Theorem 4. *Under the assumptions of Proposition 1, the expected loss ρ_ℓ is robust against optimization at $X \in L^{pq^+}$ for L^{pq^+}-convergence.*

As corollaries, we obtain concrete solutions to the problem of minimizing $\rho(g(X))$ over $g \in \mathscr{G}$.

- For constant constraint functions, v and w, and a coherent risk measure ρ, this problem can be formulated as a composite hypothesis testing problem, and there exists a significant amount of corresponding literature.

- Much fewer results were obtained for the case of nonconstant constraint functions, and those that are available often lack some concreteness. It is therefore worth pointing out that our proofs also provide structure results for the solutions of our optimization problems.

- Specifically, in the context of Corollary 1 (robustness of ES), our proof yields that 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:

$$g_X(x) = \begin{cases} (v(x) \vee z^* \wedge w(x))\mathbb{1}_{\{0 < c\gamma(x) < 1\}} & \text{or} \\ (v(x) \vee z^* \wedge w(x))\mathbb{1}_{\{c\gamma(x) > 1\}} + v(x)\mathbb{1}_{\{c\gamma(x) \leq 1\}} \end{cases}$$

  This significantly extending previous results, such as those by Sekine (2004).

3 Conclusion

- Applying the classical notion of classical robustness yields that VaR is (essentially) robust, whereas ES and all other coherent risk measures are not robust.
- A refined notion of robustness shows that the picture is not all black & white but that there are varying degrees of robustness.

  The more robust a risk measure is, the less sensitive it becomes toward extreme risks and crash scenarios.
- However, if one takes the incentives created by a regulatory risk measure into account, the picture can change completely, and VaR becomes much less robust than ES.

  The reason is that portfolios optimized w.r.t. VaR are "too greedy": since VaR ignores tail risk,
- In conclusion, one can say that VaR has no overall advantage over ES with respect to robustness properties.

Thank you

References

Artzner, P., Delbaen, F., Eber, J.-M. & Heath, D. (1999), 'Coherent measures of risk', *Math. Finance* **9**(3), 203–228.

Ben-Tal, A. & Teboulle, M. (1987), 'Penalty functions and duality in stochastic programming via ϕ-divergence functionals', *Math. Oper. Res.* **12**(2), 224–240.

Ben-Tal, A. & Teboulle, M. (2007), 'An old-new concept of convex risk measures: the optimized certainty equivalent', *Math. Finance* **17**(3), 449–476.

Cont, R., Deguest, R. & Scandolo, G. (2010), 'Robustness and sensitivity analysis of risk measurement procedures', *Quant. Finance* **10**(6), 593–606.

Embrechts, P., Schied, A. & Wang, R. (2018), 'Robustness in the optimization of risk measures', *arXiv:1809.09268*.

Goodhart, C. A. (1975), 'Problems of monetary management: the uk experience', *Papers in Monetary Economics.* **1**.

Hampel, F. R. (1971), 'A general qualitative definition of robustness', *Ann. Math. Statist.* **42**, 1887–1896.

Krätschmer, V., Schied, A. & Zähle, H. (2012), 'Qualitative and infinitesimal robustness of tail-dependent statistical functionals', *J. Multivariate Anal.* **103**, 35–47.

Krätschmer, V., Schied, A. & Zähle, H. (2014), 'Comparative and qualitative robustness for law-invariant risk measures', *Finance and Stochastics* **18**, 271–295.

Krätschmer, V., Schied, A. & Zähle, H. (2017), 'Domains of weak continuity of statistical functionals with a view toward robust statistics', *Journal of Multivariate Analysis* **158**, 1–19.

Sekine, J. (2004), 'Dynamic minimization of worst conditional expectation of shortfall', *Math. Finance* **14**(4), 605–618.

Strathern, M. (1997), '"Improving ratings': audit in the British university system', *European review* **5**(3), 305–321.