Financial measures of risk and performance

MATRIX, 20–21 November 2017

Part 1: A review of the theory of risk measures

Mikhail Zhitlukhin
Steklov Mathematical Institute, Moscow

Introduction

We'll consider families of risk measures $F_\lambda(X)$ parametrized by $\lambda \in [0,1]$ (or another set):

$$F_\lambda \colon L^p \to \mathbb{R}, \qquad \lambda \in [0,1],$$

and their inverse functions in λ:

$$G(X,x) = \lambda \text{ such that } F_\lambda(X) = x.$$

Example (Rockafellar, Royset, 2010)

For a random variable $X \in L^1$ with a continuous distribution we can define the conditional value at risk[1] at a level $\lambda \in [0, 1)$ by

$$\mathrm{CVAR}(X, \lambda) = E(X \mid X > Q_X(\lambda)),$$

where $Q_X(\lambda)$ is the λ-quantile of X, i.e. $P(X \leqslant Q_X(\lambda)) = \lambda$.

Define its inverse $\mathbb{P}(X, x)$:

$$\mathbb{P}(X, x) = \lambda \text{ such that } \mathrm{CVAR}(X, \lambda) = x.$$

We'll show that $\mathbb{P}$ is a "good modification" of $P(X > x)$.

[1] Also called Average Value at Risk, Tail Value at Risk, Expected Shortfall.

Questions of interest:

- properties and interpretations;
- how to use them in optimization problems;
- how to compute them without explicitly inverting the functions.

Plan of the minicourse

1. A review of risk measures
2. Inverse functions of CVAR and L^p-CVAR
3. Monotone Sharpe ratios, connection with inverses of L^p-CVAR
4. Performance measures: inverse functions of general risk measures
5. Applications to some non-Markov problems of optimal control

1. VAR and CVAR

Let X be a random variable describing loss[2]. The magnitude of loss can be quantified by probabilities

$$P(X \leqslant x)$$

or, alternatively, by quantiles (Value at Risk)

$$\mathrm{VAR}_\lambda(X) \equiv Q(X, \lambda) = \text{smallest } x \text{ s.t. } P(X \leqslant x) \geqslant \lambda$$

(left inverse of the distribution function).

[2] Small values are good, large values are bad.

Bad properties of $Q(X, \lambda)$ (and $P(X \leqslant x)$)

1. $Q(X, \lambda)$ doesn't show "what happens" above it
   - In finance: no information about losses larger than $Q(X, \lambda)$
2. Not convex in X
   - Non-convexity may discourage diversification
   - Can make difficult minimization of $Q(X, \lambda)$, e.g. portfolio selection
3. Discontinuous in λ and in X for discrete distributions
   - Distributions with atoms may appear in models with jumps or from empirical distributions

Artzner at. al. (1999), Rockafellar and Uryasev (2000) and others systematically studied coherent risk measures.

Basic example – CVAR:

$$\mathrm{CVAR}(X, \lambda) = \alpha Q(X, \lambda) + (1 - \alpha) E(X \mid X > Q(X, \lambda))$$

where $\alpha = \alpha(X, \lambda) \in [0, 1]$ corrects for continuity ($\alpha = 0$ at a point of continuity of the distribution).

Nice properties:

1. Average loss above the quantile
2. Convex in X
3. Continuous[3] in λ and in X

[3]Except possibly $\lambda = 1$. The accurate definition of the continuity in X will be given below.

Alternative representations of CVAR

1. As an optimization problem:

$$\mathrm{CVAR}(X,\lambda) = \min_{c\in\mathbb{R}}\big(\tfrac{1}{1-\lambda}E(X-c)_+ + c\big), \qquad \text{where } x_+ = \max(x,0)$$

2. Dual ("robust") representation:

$$\mathrm{CVAR}(X,\lambda) = \max\big\{E(RX) \mid 0 \leqslant R \leqslant \tfrac{1}{1-\lambda},\ ER = 1\big\}$$

3. Average VAR

$$\mathrm{CVAR}(X,\lambda) = \int_\lambda^1 Q(X,\lambda)d\lambda$$

CVAR as an optimization problem

Rockafellar, Uryasev (2000) obtained the formula:

$$\mathrm{CVAR}(X,\lambda) = \min_{c\in\mathbb{R}}\left(\tfrac{1}{1-\lambda}E(X-c)_+ + c\right)$$

with the minimum attained at $c^* \in \mathrm{VAR}_\lambda(X)$

Advantages: the objective function is convex in c and continuously differentiable (if X has density) or piecewise-linear (if X is discrete).

This makes convenient using CVAR in portfolio selection:

$$\begin{aligned} \text{minimize:} \quad & \tfrac{1}{1-\lambda}E\left(\textstyle\sum_i \alpha_i X_i - c\right)_+ + c \quad \text{over } \alpha_i \text{ and } c \\ \text{subject to:} \quad & \textstyle\sum_i \alpha_i = 1 \\ & \overline{\alpha} \in \mathcal{A} \end{aligned}$$

where X_i are asset returns and $\mathcal{A}$ is a convex set of constraints.

2. Axiomatics of risk measures and their representations

A functional $\rho\colon L^p \to \mathbb{R}$ is called a coherent risk measure if it satisfies the following properties:

1. Monotonicity
   if $X \leqslant Y$ a.s., then $\rho(X) \leqslant \rho(Y)$
2. Positive homogeneity
   $\rho(\alpha X) = \alpha\rho(X)$ for any real $\alpha \geqslant 0$
3. Convexity
   $\rho(\alpha X+(1-\alpha)Y) \leqslant \alpha\rho(X)+(1-\alpha)\rho(Y)$ for any X, Y and $\alpha \in [0,1]$
4. Translation invariance
   $\rho(X+c) = \rho(X) + c$ for any $c \in \mathbb{R}$
5. Continuity
   $p \in [1,\infty)$: $\rho(X)$ is continuous w.r.t. the L^p-norm
   $p = \infty$: the Fatou property (weak lower semicontinuity; see below)

The dual ("robust") representation

A functional $\rho\colon L^p \to \mathbb{R}$ is a coherent risk measure if and only if there exists a set $\mathcal{R} \subset L^q$ such that

$$\rho(X) = \sup_{R \in \mathcal{R}} E(RX) \text{ for any } X \in L^p$$

and $R \geqslant 0$, $ER = 1$ for any $R \in \mathcal{R}$.

Here, $E(RX)$ can be interpreted as expectations w.r.t. different probability measures Q with densities $\frac{dQ}{dP} = R$.

A note about continuity and topologies on L^p

Proposition (follows from the Fenchel-Moreau theorem)
Let $\mathcal{X}$ be a Hausdorff locally convex space. If $f\colon \mathcal{X} \to (-\infty, +\infty]$ is a proper sublinear lower semicontinuous function, then

$$f(x) = \sup_{l \in C} l(x) \text{ for all } x \in \mathcal{X}$$

with some convex closed set $C \in \mathcal{X}'$ of continuous linear functionals $l(x)$.

To have "nice" linear functionals $l(X) = E(RX)$ for $R \in L^q$ we need a "nice" topology on L^p.

- For $p \in [1, \infty)$: the strong (norm) topology is nice.
- For $p = \infty$: we need the weak-* topology (the smallest topology which makes all $l(X) = E(RX)$ continuous).

- For $p \in [1, \infty)$
  On a Banach space, a finite convex l.s.c. functional is continuous
- For $p = \infty$
  Given convexity, l.s.c. is equivalent to the Fatou property:
  $$\text{if } \|X_n\|_\infty \leqslant 1 \text{ and } X_n \xrightarrow{P} X, \text{ then } \rho(X) \leqslant \liminf \rho(X_n)$$

The properties $R \geqslant 0$ and $ER = 1$ for any $R \in \mathcal{R}$ become equivalent to the monotonicity and the translation invariance of a risk measure.

Example: CVAR

For the conditional value at risk:

$$\mathrm{CVAR}(X,\lambda) = \max\left\{E(RX) \ \mid 0 \leqslant R \leqslant \tfrac{1}{1-\lambda},\ ER = 1\right\}$$

The maximum is attained at

$$R^* = \tfrac{1}{1-\lambda}\mathbf{I}(X > q) + \alpha\mathbf{I}(X = q)$$

where $q = \mathrm{Q}(X,\lambda)$ and α is the appropriate constant to make $ER^* = 1$ ($\alpha = 0$ if $P(X = q) = 0$).

Extension: convex risk measures

A functional $\rho\colon L^p \to \mathbb{R}$ is called a convex risk measure if it satisfies the properties:

1. Monotonicity
2. Convexity
3. Translation invariance
4. Continuity

(without positive homogeneity)

A robust representation for convex risk measures

Any convex risk measure can be represented in the form

$$\rho(X) = \sup_{R \in L^q} \left(E(RX) - \alpha(R) \right)$$

with the convex penalty function $\alpha \colon L^q \to (-\infty, +\infty]$

$$\alpha(R) = \sup_{X \in L^p} \left(E(RX) - \rho(X) \right)$$

For example, for a coherent risk measure: $\alpha(R) = 0$ if $R \in \mathcal{R}$ and $\alpha(R) = +\infty$ if $R \notin \mathcal{R}$.

Example: the entropic risk measure

Consider the following penalty function on L^1:

$$\alpha(R) = \beta^{-1} H(R \mid P) \qquad \text{with } H(R \mid P) = E(R \log R),$$

where $\beta > 0$ is constant and $H(R \mid P)$ is the relative entropy of the measure $dQ = R dP$ w.r.t. P (Kullback–Leibler divergence)

It is possible to show that

$$E(RX) - \tfrac{1}{\beta}\alpha(R) \leqslant \tfrac{1}{\beta} \log E(e^{\beta X})$$

and the inequality is attained at $R^* = e^{\beta X}/E(e^{\beta X})$. Then the entropic risk measure

$$\rho(X) = \tfrac{1}{\beta} \log E(e^{\beta X}), \qquad X \in L^\infty$$

has the representation

$$\rho(X) = \sup_{R \in L^1} \left(E(RX) - \alpha(R)\right)$$

3. Construction of new risk measures

3.1. Acceptance sets

Coherent risk measures are in one-to-one correspondence with their acceptance sets

$$A_\rho = \{X : \rho(X) \leqslant 0\}$$

The class of acceptance sets is the class of non-empty convex closed cones, which do not include $\mathbb{R}_+ \setminus \{0\}$ and satisfy the monotonicity condition: if $X \in A_\rho$ and $Y \leqslant X$ a.s., then $Y \in A_\rho$.

Given a set A of this type, the corresponding coherent risk measure is

$$\rho_A(X) = \min\{c \in \mathbb{R} : X - c \in A\}$$

Example: the mean-variance risk measure

For fixed $\lambda > 0$, define

$$\mathring{A} = \{X \in L^2 : E(X) + \lambda\sigma(X) \leqslant 0\}$$

Then $\mathring{A}$ is a (not monotone) closed convex cone in L^2, hence the following is an acceptance set

$$A = \{X \in L^2 : \exists\, Y \geqslant X \text{ s.t. } E(Y) + \lambda\sigma(Y) \leqslant 0\}$$

The corresponding risk measure

$$\rho(X) = \min\{c : \exists\, Y \geqslant X \text{ s.t. } E(Y) + \lambda\sigma(Y) \leqslant c\}$$

3.2. Maximums and integrals of risk measures

If $\{\rho_\lambda\}_{\lambda\in\Lambda}$ is a family of risk measures on L^∞, then the following are also risk measures:

1. $\rho(X) = \sum_i a_i \rho_{\lambda_i}(X)$, where $a_i > 0$ and $\sum_i a_i = 1$
2. More generally, $\rho(X) = \int_\Lambda \rho_\lambda(X)\mu(d\lambda)$ if $\lambda \mapsto \rho_\lambda(X)$ is measurable and μ is a probability measure on Λ (with some σ-algebra on it)
3. $\rho(X) = \max\limits_{\lambda\in\Lambda} \rho_\lambda(X)$, provided it's finite-valued

L^∞ is considered for simplicity, to ensure the above risk measures are finite-valued.

Weighted VAR and law-invariant risk measures

Lemma

For any $X \in L^\infty$, the function $f(\lambda) = \mathrm{CVAR}(X, \lambda)$ is continuous and non-decreasing on $[0, 1]$.

Definition

Let μ be a probability measure on $[0, 1]$. The weighted VAR

$$\mathrm{WVAR}(X, \mu) = \int_{[0,1]} \mathrm{CVAR}_\lambda(X, \lambda)\mu(d\lambda)$$

is a coherent risk measure on L^∞.

Obviously, WVAR is law-invariant in the following sense:

$$\text{if } X \stackrel{d}{=} Y, \text{ then } \mathrm{WVAR}(X, \mu) = \mathrm{WVAR}(Y, \mu)$$

Theorem (Kusuoka, 2001)

Suppose the probability space is atomless (there exists a continuous random variable on it).

Then a risk measure ρ on L^∞ is law invariant if and only if there exists a set $\mathcal{M}$ of probability measures on $[0, 1]$ such that

$$\rho(X) = \sup_{\mu \in \mathcal{M}} \mathrm{WVAR}(X, \mu)$$

Comonotone risk measures

Two random variables X and Y are called comonotone if there is Z and two non-decreasing functions f, g such that that $(X, Y) \stackrel{d}{=} (f(Z), g(Z))$

Equivalently their joint distribution $F(x, y) = \min\{F_X(x), F_Y(y)\}$

A risk measure is called comonotone if for any comonotone X, Y

$$\rho(X + Y) = \rho(X) + \rho(Y)$$

Theorem (Kusuoka, 2001)

Suppose the probability space is atomless. A law-invariant risk measure ρ on L^∞ is comonotone if and only if there is μ such that

$$\rho(X) = \mathrm{WVAR}(X, \mu)$$

Higher order monotonicity

A functional $\rho(X)$ is called

- 1st order monotone if $\rho(X) \geqslant \rho(Y)$ for any X, Y such that
  - $EU(X) \geqslant EU(Y)$ for any convex U (if $EU(X)$, $EU(Y)$ exist);
  - equivalently, $P(X \geqslant t) \geqslant P(Y \geqslant t)$ for any $t \in \mathbb{R}$;
  - equivalently, $Y = X + Z$, where $E(Z \mid X) = 0$

- 2nd order monotone if $\rho(X) \geqslant \rho(Y)$ for any X, Y such that
  - $EU(X) \geqslant EU(Y)$ for any increasing convex U (if $EU(X)$, $EU(Y)$ exists)
  - equivalently, $\int_{-\infty}^{x} (F_Y(t) - F_X(t))dt \geqslant 0$ for any $x \in \mathbb{R}$
  - equivalently, $Y = X + Z$, where $E(Z \mid X) \leqslant 0$

- Dilatation monotone if $\rho(X) \geqslant \rho(E(X \mid \mathcal{G}))$ for any X and any sub-σ-algebra $\mathcal{G}$

Theorem (Cherny, Grigoriev (2007))

Suppose the probability space is atomless. Then for a convex lsc functional $\rho\colon L^p \to \mathbb{R}$, the following are equivalent:

1. ρ is law-invariant
2. ρ is 1st order monotone
3. ρ is dilatation monotone

If additionally ρ is monotone w.r.t. a.s.-order, then 1-3 are equivalent to

4. ρ is 2nd order monotone

Corollary

On an atomless probability spaces, law-invariant risk measures are 1st order, 2nd order and dilatation monotone.

Example: monotonicity of CVAR

It is easy to show that for a law-invariant risk measure dilatation monotonicity implies the 2nd order monotonicity.

Let's show in a simple way that CVAR is dilatation monotone.

$$\begin{aligned}\mathrm{CVAR}(E(X \mid \mathcal{G}), \lambda) &= \min_{c \in \mathbb{R}} \big(\tfrac{1}{1-\lambda} E(E(X - c \mid \mathcal{G})_+) + c \big) \\ &\geqslant \min_{c \in \mathbb{R}} \big(\tfrac{1}{1-\lambda} E(X - c)_+ + c \big) = \mathrm{CVAR}(X, \lambda)\end{aligned}$$

by conditional Jensen's inequality.

3.3. CVAR in L^p and transformed norm risk measures

Recall the representation for CVAR: for $X \in L^1$

$$\mathrm{CVAR}(X, \lambda) = \min_{c \in \mathbb{R}} \left(\tfrac{1}{1-\lambda} E(X - c)_+ + c \right)$$

In a similar way, define CVAR in L^p, $p \in [1, \infty]$:

$$\mathrm{CVAR}_p(X, \lambda) = \min_{c \in \mathbb{R}} \left(\tfrac{1}{1-\lambda} \|(X - c)_+\|_p + c \right) \qquad \text{for } \lambda \in [0, 1)$$

Note that the case $p = \infty$ is not interesting: for any $\lambda \in [0, 1)$ we have $\mathrm{CVAR}_\infty(X, \lambda) = \operatorname{ess\,sup} X$.

Proposition (Krokhmal (2007), Cheridito, Li (2009))

CVAR_p is a law-invariant coherent risk measure on L^p and has the dual representation

$$\mathrm{CVAR}_p(X,\lambda) = \max\big\{E(RX) \mid R \geqslant 0,\ ER = 1,\ \|R\|_q \leqslant 1\big\}$$

The minimum over $c \in \mathbb{R}$ in the definition is attained at c^* such that

$$\frac{\|(X-c^*)_+\|_p}{\|(X-c^*)_+\|_{p-1}} = (1-\lambda)^{-\frac{1}{p-1}}$$

and the maximum in the dual representation is attained at

$$R^* = \frac{(X-c^*)_+^{p-1}}{E\big((X-c^*)_+^{p-1}\big)}$$

A connection between CVAR_p and CVAR_{p-1}

From the above formula for the optimal c^* for CVAR_p:

$$\begin{aligned}\mathrm{CVAR}_p(X,\lambda) &= \tfrac{1}{1-\lambda}\|(X-c^*)_+\|_p + c^* \\ &= \tfrac{1}{(1-\lambda)^q}\|(X-c^*)_+\|_{p-1} + c^* \geqslant \mathrm{CVAR}_{p-1}(X,\widetilde{\lambda})\end{aligned}$$

where $\widetilde{\lambda} = 1-(1-\lambda)^q > \lambda$.

Extension to Orlicz spaces (Cheridito, Li (2009))

Suppose $\Phi\colon \mathbb{R}_+ \to \overline{\mathbb{R}}_+$ is a l.s.c. convex function with $\Phi(0) = 0$ and $\Phi(x) \to \infty$ as $x \to \infty$ (a Young function).

Define the Orlicz space L^Φ and the Orlicz heart $H^\Phi \subseteq L^\Phi$:

$$L^\Phi = \{X : E(\Phi(c|X|)) < \infty \text{ for some } c > 0\}$$
$$H^\Phi = \{X : E(\Phi(c|X|)) < \infty \text{ for all } c > 0\}$$

and the norm $\|X\|_\Phi$ on L^Φ (the Luxembourg norm):

$$\|X\|_\Phi = \inf\{c > 0 : E(\Phi(c^{-1}|X|)) \leqslant 1\}$$

These spaces behave similarly to L^p, for example:

1. L^Φ and H^Φ are Banach spaces w.r.t. $\|\cdot\|_\Phi$
2. Let $\Psi\colon \mathbb{R}_+ \to [0,\infty]$ be the conjugate of Φ (also a Young function):

$$\Psi(y) = \sup_{x \geqslant 0}(xy - \Phi(x))$$

If $\Phi(x) < \infty$ for all $x \geqslant 0$, then the dual space $(H^\Phi)' \simeq L^\Psi$

Examples

1. $\Phi(x) = x^p$, then $\Psi(y) = \frac{p^{1-q}}{q} y^q$, where $p \in [1,\infty)$ and $\frac{1}{p} + \frac{1}{q} = 1$
   $L^\Phi = H^\Phi = L^p$, $\|\cdot\|_\Phi = \|\cdot\|_p$, and $L^\Psi = L^q$

2. $\Phi(x) = e^x - 1$. Then $\Psi(y) = (y(\log y - 1) + 1)\mathbf{I}(y \geqslant 1)$, and
   $L^\infty \subsetneq H^\Phi \subsetneq L^p \subsetneq L^\Psi \subsetneq L^1$ for any $p \in (1,\infty)$

The robust representation

Theorem

Suppose $\Phi(x) < \infty$ for all $x \geqslant 0$. Then a functional $\rho(X)$ is a coherent risk measure on H^Φ if

$$\rho(X) = \max_{R \in \mathcal{R}} E(RX)$$

with the $\|\cdot\|_\Psi$-bounded, convex and $\sigma(L^\Psi, H^\Phi)$-closed set $\mathcal{R} \subset L^\Psi$, explicitly given by

$$\mathcal{R} = \{R \in L^\Psi : R \geqslant 0,\ ER = 1,\ E(RX) \leqslant 0 \text{ for any } X \in A_\rho\}$$

where A_ρ is the acceptance set of ρ.

Example 1

$$\mathrm{CVAR}_\Phi(X,\lambda) = \min_{c\in\mathbb{R}}\big(\tfrac{1}{1-\lambda}\|(X-c)_+\|_\Phi + c\big)$$

This is a coherent risk measure on H^Φ with the robust representation

$$\mathrm{CVAR}_\Phi(X,\lambda) = \max\big\{E(RX) \mid R \geqslant 0,\ ER = 1,\ \|R\|_\Phi^* \leqslant \tfrac{1}{1-\lambda}\big\}$$

where $\|R\|_\Phi^*$ is the Orlicz norm on L^Ψ defined as follows:

$$\|R\|_\Phi^* = \max\big\{E(RX) \mid X \in L^\Phi,\ \|X\|_\Phi \leqslant 1\big\}$$

It is known that $\|\cdot\|_\Psi$ and $\|\cdot\|_\Phi^*$ are equivalent: $\|\cdot\|_\Psi \leqslant \|\cdot\|_\Phi^* \leqslant 2\|\cdot\|_\Psi$
Moreover, the Hölder inequality holds:

$$\|RX\|_1 \leqslant \|R\|_\Phi^* \cdot \|X\|_\Phi$$

Example 2: transformed norm risk measures

$$\rho(X) = \min_{c \in \mathbb{R}} (F(\|H(X-c)\|_G) + c)$$

is a convex risk measure on H^Φ with $\Phi = G \circ H$, where

$F \colon \mathbb{R}_+ \to (-\infty, \infty]$ is convex, increasing and left-continuous;

G is a Young function;

$H \colon \mathbb{R} \to [0, \infty)$ is increasing and convex with $H(x) \to \infty$ as $x \to \infty$;

and F, G, H satisfy some additional conditions on their compositions.

References

1. H. Föllmer, A. Schied (2011). *Stochastic Finance: an introduction in dscrete time*, 3rd ed.
2. P. Artzner, F. Delbaen, J. M. Eber D. Heath (1999). Coherent measures of risk. *Mathematical finance*, 9(3), 203-228.
3. R. T. Rockafellar, S. Uryasev (2000). Optimization of conditional value-at-risk. *Journal of risk*, 2, 21-42.
4. S. Kusuoka (2001). On law invariant coherent risk measures. *Advances in mathematical economics*, pp. 83-95.
5. A. S. Cherny, P. G. Grigoriev (2007). Dilatation monotone risk measures are law invariant. *Finance and Stochastics* 11(2), 291-298.
6. P. A. Krokhmal (2007). Higher moment coherent risk measures. *Quantitative Finance* 7(4), 373-387.
7. P. Cheridito, T. Li (2009). Risk measures on Orlicz hearts. *Mathematical Finance* 19(2), 189-214.
8. R. T. Rockafellar, J. O. Royset (2010). On buffered failure probability in design and optimization of structures. *Reliability Engineering and System Safety* 95(5), 499-510.

Financial measures of risk and performance

MATRIX, 20–21 November 2017

Part 2: Inverse functons of risk measures

Mikhail Zhitlukhin
Steklov Mathematical Institute, Moscow

1. Introduction

In some problems, it may be more convenient to use distribution functions rather than quantile functions, despite they are mathematically equivalent, e.g.

minimize: $P(X > x)$ subject to: $X \in \mathcal{X}$	instead of	maximize: $Q(X, \lambda)$ subject to: $X \in \mathcal{X}$

For example, "minimize the probability that the loss will exceed available reserves" rather than "minimize the loss in the worst 1% of cases".

But $P(X > x)$ have all the same bad properties as quantiles ...

We will mostly use $P(X > x)$, not $P(X \leqslant x)$, simply for convenience.

Bad properties of $P(X > x)$:

1. Doesn't show what happens above x
2. Not convex in X
3. May be discontinuous in X
   For example, $X = \sum_i c_i X_i$, where some X_i has a discontinuous distribution. Then changing c_i slightly may change $P(X > x)$ a lot.

Since CVAR allowed to avoid the same problems of VAR, it is possible to use similar ideas for distribution functions.

Define the new functional as the inverse of CVAR by analogy with that P is the inverse of Q:

$$\begin{tikzcd}
P(x) \arrow[r, leftrightarrow] \arrow[d, dashed, leftrightarrow] & \mathrm{VAR}(\lambda) \arrow[d, dashed, leftrightarrow] \\
\mathbb{P}(x) \arrow[r, leftrightarrow] & \mathrm{CVAR}(\lambda)
\end{tikzcd}$$

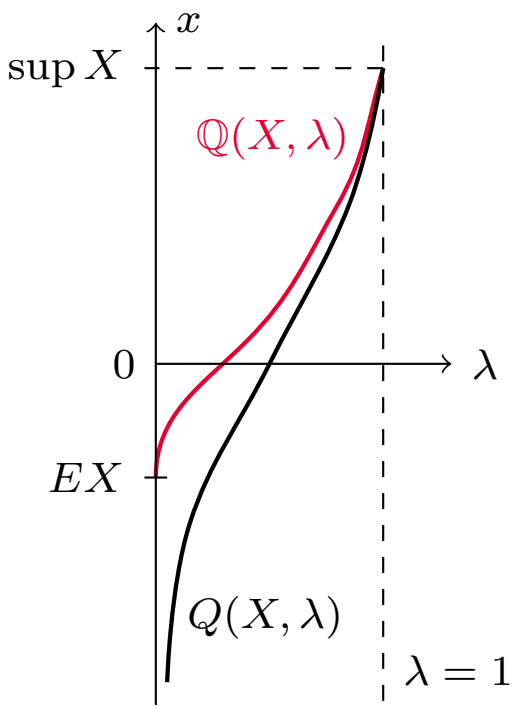
This idea was proposed by Rockafellar and Royset (2010) in relation to engineering applications, and called buffered probability (of failure).

Results below include the results of that paper and also Uryasev, Mafusalov (2014).

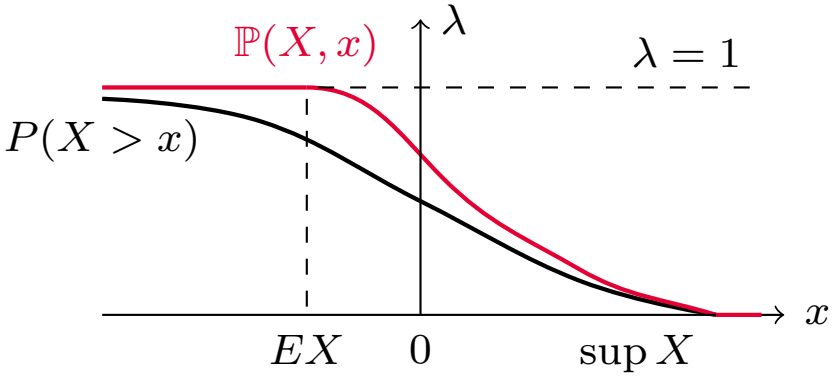
2. The definition

The definition "in pictures"

VAR and CVAR



Inverse functions



From now on, we'll use the notation $\mathbb{Q}(X,\lambda)$ for $\mathrm{CVAR}(X,\lambda)$.

Formal definition

Lemma

For any random variable $X \in L^1$, the function $f(\lambda) = \mathbb{Q}(X, \lambda)$ has the following properties:

1. $f(0) = EX$
2. $f(\lambda) = \sup X$ for $\lambda \geqslant 1 - P(X = \sup X)$, with, possibly, $\sup X = \infty$
3. $f(\lambda)$ is continuous and non-decreasing
4. $f(\lambda)$ is strictly increasing on the set $\{\lambda : f(\lambda) < \sup X\}$

Now define

$$\mathbb{P}(X,x) = \begin{cases} 0, & \text{if } x > \sup X \\ P(X = \sup X), & \text{if } x = \sup X \\ 1 - \mathbb{Q}^{-1}(X,x), & \text{if } EX < x < \sup X \\ 1, & \text{if } x \leqslant EX \end{cases}$$

$\mathbb{P}(X,x)$ was called "buffered probability (that $X > x$)":

$$\mathbb{P}(X,x) = \text{“probability that } X > x\text{”} + \text{“buffer”}$$

3. A representation and properties

Theorem

The following representation holds for any $X \in L^1$:

$$\mathbb{P}(X, x) = \min_{c \geqslant 0} E\big(c(X - x) + 1\big)_+$$

Note the function $c \mapsto E\big(c(X - x) + 1\big)_+$ is convex.

Properties w.r.t. x

Proposition

1. The function $x \mapsto \mathbb{P}(X, x)$ is continuous on $x \in (-\infty, \sup X)$, and is strictly decreasing on $[EX, +\infty)$.
2. The function $x \mapsto (\mathbb{P}(X, x))^{-1}$ is convex in x

Properties w.r.t. X

Proposition

1. Invariance under a linear transformation: $\mathbb{P}(X, x) = \mathbb{P}(aX+b, ax+b)$ for any $a \geqslant 0$ and $b \in \mathbb{R}$.
2. Second order monotonicity: $\mathbb{P}(X, x) \leqslant \mathbb{P}(Y, x)$ for any X, Y such that $EU(X) \leqslant EU(Y)$ for any increasing concave U
3. Continuity in X
4. Quasi-convexity[1]: $\{X : \mathbb{P}(X, x) \geqslant c\}$ is a convex set of random variables for any $x, c \in R$

[1] Quasi-convexity is a weaker property than convexity; a convex function is always quasi-convex, but not vice versa.

Example: $\mathbb{P}(X, x)$ is not convex in X

Take $X \equiv 2$, and $Y \equiv -1$. Then

$$\mathbb{P}((X+Y)/2,\ 0) = 1 \not\leq \tfrac{1}{2} = \tfrac{1}{2}\mathbb{P}(X, 0) + \tfrac{1}{2}\mathbb{P}(Y, 0)$$

Concavity w.r.t. mixtures of distributions

It can be proved[2] that $\mathbb{P}(X, x)$ is concave in X w.r.t. mixtures of distributions, i.e. suppose $X_i \in L^1$ and N is a random variable assuming values $\{1, \ldots, n\}$ with probabilities p_i and is independent of X_i.

Then

$$P(X_N, x) \geqslant \sum_{i=1}^{n} p_i P(X_i, x)$$

Note that $P(X_N > x) = \sum_i p_i P(X_i > x)$.

[2] See Mafusalov, Uryasev (2014).

Usage in optimization problems

Consider the problem with a convex set $\mathcal{X}$:

$$\begin{aligned} \text{minimize: } & \mathbb{P}(X, x) \text{ over } X \in L^1 \\ \text{subject to: } & X \in \mathcal{X} \end{aligned}$$

It reduces to the following optimization problem:

$$\begin{aligned} \text{minimize: } & E(c(X - x) + 1)_+ \\ \text{subject to: } & X \in \mathcal{X},\ c \in \mathbb{R}_+ \end{aligned} \quad = \quad \begin{aligned} \text{minimize: } & E(Y + 1)_+ \\ \text{subject to: } & Y \in \operatorname{cone}(\mathcal{X} - x) \end{aligned}$$

Example: portfolio selection

Suppose there is 1 riskless asset and n risky assets:

– $R_0 = 0$ is the rate of return of the riskless asses;

– $R_1, \ldots, R_n$ are the rates of return of the risky assets.

Consider the problem:

$$\begin{aligned} \text{minimize:} \quad & \mathbb{P}(-a \cdot R,\ 0) \quad \text{over } a \in \mathbb{R}^n \\ \text{subject to:} \quad & E(a \cdot R) = \delta \end{aligned}$$

i.e. find a portfolio which has the given expected rate of return, and the smallest buffered probability of loss.

Notation: $R = (R_1, \ldots, R_n)$, $a = (a_1, \ldots, a_n)$, and $a \cdot R = \sum_i a_i R_i$.

We have:

$$x = 0,$$
$$\mathcal{X} = \{a \cdot R \mid a \in \mathbb{R}^n,\ a \cdot E(R) = \delta\},$$
$$\operatorname{cone}(\mathcal{X}) = \{a \cdot R \mid a \in \mathbb{R}^n,\ a \cdot E(R) \geqslant 0\}.$$

The minimal buffered probability of loss is the value of the problem

$$\begin{aligned} \text{minimize:} \quad & E(1 - a \cdot R)_+ \\ \text{subject to:} \quad & E(a \cdot R) \geqslant 0 \end{aligned}$$

Observe the constraint can be removed here, because $E(1 - a \cdot R)_+ \geqslant 1$ if $E(a \cdot R) \leqslant 0$. Then the problem reduces to

$$\text{minimize:} \quad E(1 - a \cdot R)_+$$

If a'^* is optimal in this problem, the optimal a^* in the original problem can be found as

$$a^* = \frac{\delta a'^*}{E(a'^* \cdot R)}$$

We'll see this problem is related to maximization of the Sharpe ratio.

4. A generalization to L^p

Recall the definition of CVAR in L^p:

$$\mathbb{Q}_p(X,\lambda) = \min_{c\in\mathbb{R}}\big(\tfrac{1}{1-\lambda}\|(X-c)_+\|_p + c\big)$$

and the dual representation

$$\mathbb{Q}_p(X,\lambda) = \max\big\{E(RX) \mid R \geqslant 0,\ ER = 1,\ \|R\|_q \leqslant \tfrac{1}{1-\lambda}\big\}$$

Lemma

For $X \in L^p$, the function $f(\lambda) = \mathbb{Q}_p(X,\lambda)$ has the following properties:

1. $f(0) = EX$
2. $f(\lambda) = \sup X$ for $\lambda \geqslant 1 - (P(X = \sup X))^{\frac{1}{p}}$
3. $f(\lambda)$ is continuous and non-decreasing
4. $f(\lambda)$ is strictly increasing on the set $\{\lambda : f(\lambda) < \sup X\}$

Similarly to the definition for L^1, we define

$$\mathbb{P}_p(X,x) = \begin{cases} 0, & \text{if } x > \sup X \\ (P(X = \sup X))^{\frac{1}{p}}, & \text{if } x = \sup X \\ 1 - \mathbb{Q}_p^{-1}(X,x), & \text{if } EX < x < \sup X \\ 1, & \text{if } x \leqslant EX \end{cases}$$

Representation

For any $X \in L^p$

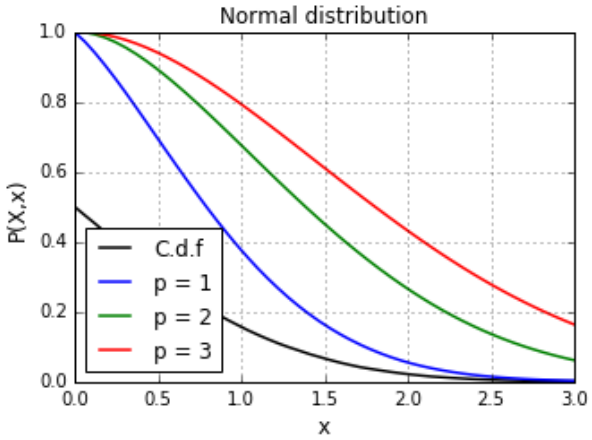
$$\mathbb{P}_p(X, x) = \min_{c \geqslant 0} \|(c(X - x) + 1)_+\|_p$$

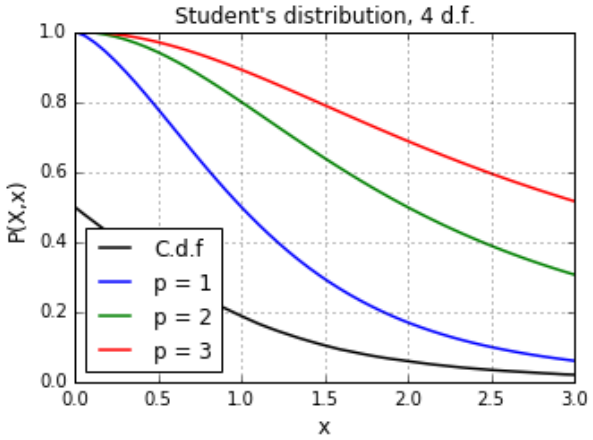
Properties of $\mathbb{P}_p(X, x)$

1. $x \mapsto \mathbb{P}_p(X, x)$ is continuous and strictly decreasing on $[EX, \sup(X))$
2. $X \mapsto \mathbb{P}_p(X, x)$ is quasi-convex, 2nd order monotone, and continuous
3. $X \mapsto \mathbb{P}_p(X, x)$ is concave w.r.t. mixtures of distributions
4. $p \mapsto \mathbb{P}_p(X, x)$ is non-decreasing in p

There results can be proved in a way similar to L^1.

Example: $\mathbb{P}_p(X, x)$ for two distributions and different p



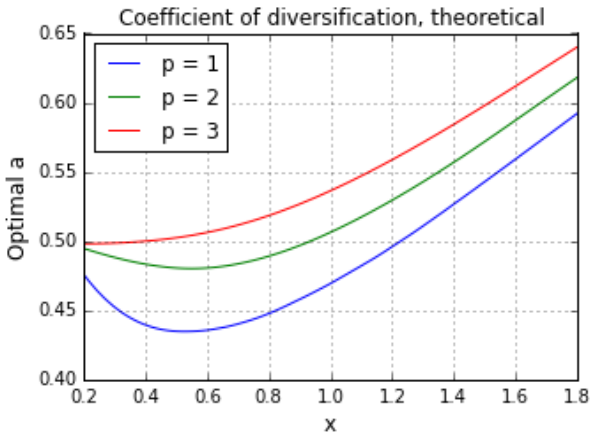


Example: portfolio diversification with different p

Suppose $X_1 \sim N$ and $X_2 \sim t(4)$ are independent and scaled so that $\operatorname{Var} X_1 = \operatorname{Var} X_2 = 1$.

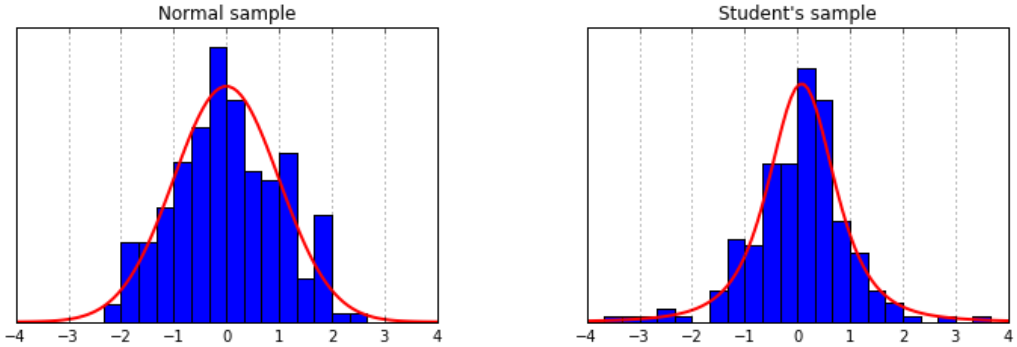
For $x > 0$, choose $a \in [0, 1]$ minimizing $\mathbb{P}_p(aX_1 + (1-a)X_2,\ x)$.

The following picture shows the results for the theoretical distributions.

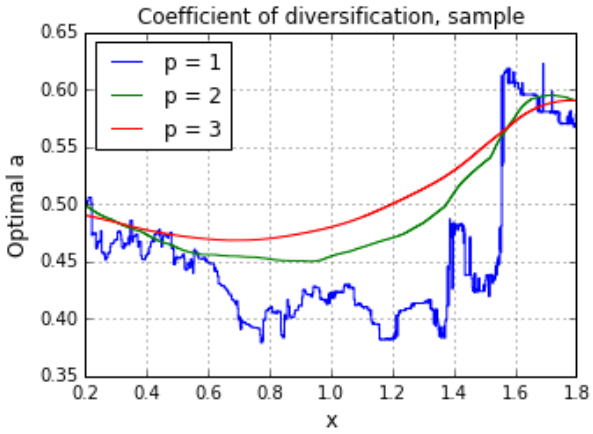


However, it turns out that the diversification coefficient for $\mathbb{P}_1(X, x)$ is "unstable" when applied to a sample distribution.

Consider two samples from the above distributions, $n = 200$:



We get the following coefficient a of the portfolio:



Financial measures of risk and performance

MATRIX, 20–21 November 2017

Part 3: Risk measures and the monotone Sharpe ratio

Mikhail Zhitlukhin
Steklov Mathematical Institute, Moscow

1. Introduction

We'll see that $\mathbb{P}_p(X, x)$ for $p = 1, 2$ is equivalent to some monotone modification of the Sharpe ratio $S(X)$ for $X \in L^2$ and a similar functional for $X \in L^1$, i.e.

$$S(X) = \frac{E(X)}{\sigma(X)}$$

In particular, for the portfolio selection problem we'll see that

> maximization of the (monotone) Sharpe ratio is equivalent to minimization of the buffered probability of loss.

2. Review: the mean-variance portfolio selection problem

Consider a market consisting of $n+1$ assets with rates of return R_i, so 1\$ invested at $t=0$ turns into $(1+R_i)$\$ at $t=1$:

- one asset is riskless with constant return $R_0 = r \geqslant 0$,
- the other assets are risky with random returns $R_i \in L^2$.

We'll denote the vector of returns by $R = (R_0 \ldots, R_n)$.

A portfolio of assets is identified with a vector $x \in \mathbb{R}^{n+1}$, such that $\sum_i x_i = 1$, where x_i is the proportion of capital invested into asset i.

The return of a portfolio x is $R_x = x \cdot R$.

The Markowitz portfolio selection problem:

$$\text{maximize } E(R_x) \text{ and minimize } \mathrm{Var}(R_x).$$

Namely, for given $\delta \geqslant r$ consider the optimization problem

$$\begin{aligned} \text{minimize:} \quad & \mathrm{Var}(x \cdot R) \quad \text{over } x \in \mathbb{R}^{n+1} \\ \text{subject to:} \quad & x \cdot e = 1 \\ & E(x \cdot R) \geqslant \delta \end{aligned}$$

where $e = (1, 1, \ldots, 1)$.

Via Lagrange multipliers, this is equivalent to solving for $\tau \geqslant 0$

$$\begin{aligned} \text{minimize:} \quad & \mathrm{Var}(x \cdot R) - 2\tau E(x \cdot R) \\ \text{subject to:} \quad & x \cdot e = 1 \end{aligned}$$

Let $\mu \in \mathbb{R}^n$ denote the vector of means, i.e. $\mu = (ER_1, \ldots, ER_n)$, Σ denote the covariance matrix $\Sigma = \mathrm{cov}(R_i, R_j)_{i,j=1}^n$.

Assumptions:

- $\det(\Sigma) > 0$
- $\mu_i \neq \mu_j$ for some $i, j \geqslant 1$

Then the solution of the above problem for any $\tau \geqslant 0$ is given by

$$x_\tau^* = x_0^* + \tau y^*$$

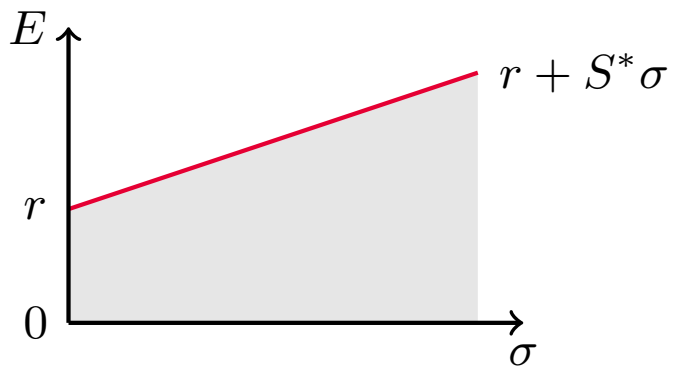
where

$$x_0^* = (1, 0, \ldots, 0), \qquad y^* = (y_0^*,\ \widehat{y}^*)$$

and

$$\widehat{y}^* = \Sigma^{-1}(\mu - re), \qquad y_0^* = -\widehat{y}_0^* \cdot e$$

The points $(\sigma(R_{x_\tau^*}),\ E(R_{x_\tau^*}))$ are located on a straight line in the (σ, E)-plane – the efficient frontier; all other portfolios are not optimal and are below the efficient frontier:



Any optimal portfolio x_τ^* can be obtained through diversification between the riskless portfolio x_0^* and some optimal portfolio, say x_1^* (Tobin's mutual fund theorem).

All optimal portfolios with $\tau > 0$ have the same Sharpe ratio $S(R_{x^*})$,

$$S^* = S(R_{x^*}) = \frac{E(R_{x^*} - r)}{\sigma(R_{x^*})}$$

which is the slope of the efficient frontier.

The portfolio selection problem is equivalent to finding a certain portfolio with the maximal Sharpe ratio, for example

$$\begin{aligned} \text{maximize:} \quad & S(R_x) \\ \text{subject to:} \quad & x \cdot e = 1 \\ & x \cdot \mu = r + \delta \end{aligned}$$

3. A monotone modification of the Sharpe ratio

Non-monotonicity of the Sharpe ratio

The Sharpe ratio is not monotone: for a random variable X there may exist a random variable Y such that

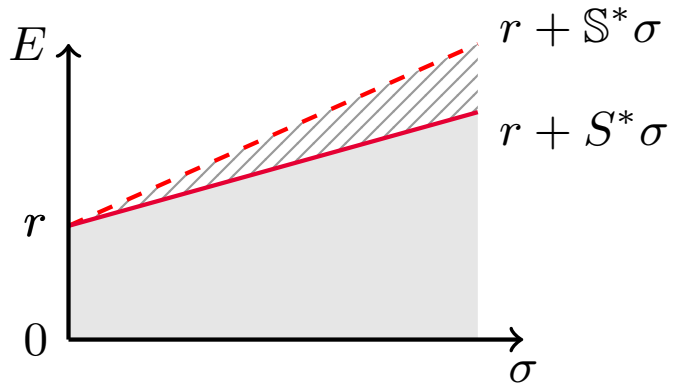
$$Y \leqslant X \text{ a.s., but } S(Y) > S(X),$$

where $C = X - Y$ represents the disposed profit (or consumption).

For example,
$X \sim N(1,1)$, $Y = \min(X, 2) \implies S(X) = 1$, but $S(Y) \approx 1.08$.

Actually, for any random variable X such that $EX > 0$, $P(X < 0) > 0$ and $\operatorname{ess\,sup} X = +\infty$ it is possible to find such Y.

For the portfolio selection problem, this means that there might exist portfolios with consumption which are above the efficient frontier:



The slope of the new efficient frontier is

$$\mathbb{S}^* = \max_{\substack{x \cdot e = 1 \\ C \geqslant 0}} \frac{E(R_x - C - r)}{\sigma(R_x - C)}$$

Making the Sharpe ratio monotone

Define the monotone Sharpe ratio

$$\mathbb{S}(X) = \sup_{Y \leqslant X} \frac{E(Y)}{\sigma(Y)}$$

where $\sup$[1] is over random variables Y such that $P(Y \leqslant X) = 1$.

$\mathbb{S}(X)$ is monotone in X, and the slope of the new efficient frontier $\mathbb{S}^*$ is the maximal monotone Sharpe ratio.

[1] The supremum is attained, except some degenerate cases.

The following formula is true for $X \in L^2$ (Z., 2015; see also below):

$$\mathbb{S}(X) = \left(\frac{1}{\min\limits_{c \geqslant 0} E(1 - cX)_+^2} - 1 \right)^{\frac{1}{2}}$$

where $\min$ is over real numbers $c \geqslant 0$.

This can be rewritten as follows:

$$\frac{1}{1 + (\mathbb{S}(X))^2} = \min_{c \geqslant 0} E(1 - cX)_+^2$$

The right-hand side is $(\mathbb{P}_2(-X, 0))^2$! Then, maximization of $\mathbb{S}(X)$ is equivalent to minimization of $\mathbb{P}_2(-X, 0)$.

We are going to show that:

- A similar relation between $\mathbb{P}_1$ and $\mathbb{S}_1$ holds, where $\mathbb{S}_1$ is an analogue of $\mathbb{S}$ for L^1, which we'll define later.
- For $p \neq 1, 2$, we'll introduce $\mathbb{S}_p$ in a similar way and obtain a good representation for it, but it'll be slightly different from $\mathbb{P}_p$.

4. A generalization to L^p

Recall the definition of the monotone Sharpe ratio in L^2:

$$\mathbb{S}(X) = \sup_{Y \leqslant X} \frac{E(Y)}{\sigma(Y)}, \qquad X, Y \in L^2.$$

We want to replace $\sigma(X)$ by a similar functional in L^p, call it $\sigma_p(X)$.

On one hand, we want $\sigma_p(X)$ to be similar to $\sigma(X)$. On the other hand, we want it to yield a convenient representation for $\mathbb{S}$.

We define σ_p as the distance in L^p-norm to the set of constants:

$$\sigma_p(X) = \min_{c \in \mathbb{R}} \|X - c\|_p$$

In particular,

- for $p = 1$: $\sigma_1(X) = E|X - \mathrm{med}(X)|$
- for $p = 2$: $\sigma_2(X) = \|X - EX\|_2 = \sigma(X)$

It is easy to see that

- $\sigma_p(X) = 0$ if and only if $X = const$
- $\sigma_p(aX + b) = |a|^p \sigma(X)$ for any $a, b \in \mathbb{R}$

Lemma

For any $X \in L^p$ the following representation holds:

$$\sigma_p(X) = \max\{E(RX) \mid R \in L^q,\ ER = 0, \|R\|_q \leqslant 1\}$$

where $\frac{1}{p} + \frac{1}{q} = 1$.

Define the monotone Sharpe ratio in L^p for $X \not\equiv 0$

$$\mathbb{S}_p(X) = \sup_{Y \leqslant X} \frac{E(Y)}{\sigma_p(Y)}.$$

The main case of interest will be when $EX > 0$ and $\inf X < 0$, where $\mathbb{S}_p(X) \in (0, \infty)$ and $\sup$ in the definition is attained.

Otherwise

- if $EX \leqslant 0$, then $\mathbb{S}_p(X) = 0$;
- if $\operatorname{ess\,inf} X > 0$, then $\mathbb{S}_p(X) = +\infty$.

A representation theorem for $\mathbb{S}_p$

If $E(X) > 0$ and $\operatorname{ess\,inf} X < 0$, then the following representation is true:

$$(\mathbb{S}_p(X))^q = \max_{a,b\in\mathbb{R}} \Big\{ b - E\Big(\tfrac{q-1}{q^p} \big|(aX+b)_+ - q\big|^p + (aX+b)_+ \Big) \Big\}$$

For $p = 1, 2$ this further simplifies to one-dimensional optimization:

$$\frac{1}{1 + (\mathbb{S}_p(X, r))^p} = \min_{c \geqslant 0} E(1 - cX)_+^p$$

In particular, the right-hand side is $(\mathbb{P}_p(-X, 0))^p$.

Why there's no simple connection between $\mathbb{S}_p$ and $\mathbb{P}_p$ for $p \neq 1, 2$

$\mathbb{P}_p$ and $\mathbb{S}_p$ reduce to slightly different optimization problems:

For $\mathbb{P}_p$		For $\mathbb{S}_p$	
minimize:	$\|R\|_q$	minimize:	$\|R-1\|_q$
subject to:	$\mathbb{R} \geqslant 0$ a.s.	subject to:	$\mathbb{R} \geqslant 0$ a.s.
	$ER = 1$		$ER = 1$
	$E(RX) = 0$		$E(RX) = 0$

But for $p = 1, 2$ they turn out to be equivalent.

If instead of $\sigma_p(X)$ we consider $\widetilde{\sigma}_p(X) = \|X - EX\|_p$, then

$$\widetilde{\sigma}_p(X) = \sup_{R \in L^q} \{E(RX) \mid ER = 0,\ \sigma_q(R) \leqslant 1\}$$

The optimization problem becomes

$$\begin{aligned} \text{minimize:} \quad & \|R - c\|_q \quad \text{over } R \in L^q,\ c \in \mathbb{R} \\ \text{subject to:} \quad & \mathbb{R} \geqslant 0 \text{ a.s.} \\ & ER = 1 \\ & E(RX) = 0 \end{aligned}$$

5. Properties of $\mathbb{S}_p$

Theorem

For any $p \in [1, \infty)$, the following properties are satisfied by $\mathbb{S}_p$:

1. Quasi-concavity
   $\{X : \mathbb{S}_p(X) \geqslant C\}$ is a convex set in L^p for any $C \in \mathbb{R}$
2. Scaling invariance
   $\mathbb{S}_p(\lambda X) = \mathbb{S}_p(X)$ for any real $\lambda > 0$
3. Law invariance
   If $X \overset{d}{=} Y$, then $\mathbb{S}_p(X) = \mathbb{S}_p(Y)$
4. Monotonicity
   $\mathbb{S}_p(X)$ is 2nd order monotone w.r.t. X
5. Continuity
   $\mathbb{S}_p(X)$ is continuous in L^p-norm at X s.t. $EX > 0$ and $\operatorname{ess\,inf} X < 0$

1. Quasi-concavity

The quasi-concavity property can be rewritten as follows:

$$\begin{aligned}&\text{if } \mathbb{S}_p(X) \geqslant c \text{ and } \mathbb{S}_p(Y) \geqslant c,\\ &\text{then } \mathbb{S}_p(\lambda X + (1-\lambda)Y) \geqslant c \text{ for any } \lambda \in [0,1],\end{aligned}$$

i.e. $\mathbb{S}_p$ of a diversified portfolio is not worse than any of its components.

Proof follows from that the Sharpe ratio $S_p = \frac{EX}{\sigma_p(X)}$ is quasi-concave: if $S_p(X) \geqslant a$ and $S_p(Y) \geqslant a$, then

$$S_p(\lambda X + (1-\lambda)Y) \geqslant \frac{\lambda EX + (1-\lambda)EY}{\lambda\sigma_p(X) + (1-\lambda)\sigma_p(Y)} \geqslant a$$

and then $\mathbb{S}_p$ is the maximum of $f_Z(X) = S_p(X - Z)$ over $Z \geqslant 0$.

2. Scaling invariance

Suppose X is the r-adjusted return of a portfolio $x = (x_0, \ldots, x_n)$

$$X = x \cdot (1 + R) - (1 + r) = \sum_{i=1}^{n} x_i (R_i - r)$$

where R_i are the returns of the risky assets, r is the risk-free return.

Then λX is the return of the leveraged portfolio $x' = (x'_0 \ldots, x'_n)$

$$x'_i = \lambda x_i, \qquad x'_0 = 1 - \lambda \sum_i x_i$$

Scaling invariance implies that leverage doesn't change $\mathbb{S}_p$.

3. Law invariance

Law invariance can be obtained from the fact that $\sup$ over $Y \leqslant X$ in the definition of $\mathbb{S}_p(X)$ can be taken over Y that are $\sigma(X)$-measurable.

Indeed, if $Y' = E(Y \mid X)$, then $EY' = EY$, and

$$\sigma_p(Y') = \min_c \|E(Y - c \mid X)\|_p \leqslant \min_c \|Y - c\|_p = \sigma_p(Y)$$

by Jensen's inequality.

4. Monotonicity

Recall that the 2nd order monotonicity means that if $EU(X) \leqslant EU(Y)$ for any increasing concave $U(\cdot)$ such that $EU(X)$ and $EU(Y)$ exist, then $\mathbb{S}_p(X) \leqslant \mathbb{S}_p(Y)$.

Equivalently: $X \stackrel{d}{=} Y + Z$, where $E(Z \mid Y) \leqslant 0$.

A simple proof

Suppose $X = Y + Z$ on the same Ω and $\mathbb{S}_p(X) = \frac{EX'}{\sigma_p(X')}$ for $X' \leqslant X$.

Take $Y' = E(X' \mid Y)$.
Then $Y' \leqslant Y$, $EY' = EX'$, and $\sigma_p(Y') \leqslant \sigma_p(X')$.

6. Normalization and an application to data

Problem: how to compare values of $\mathbb{S}$? For example, if $\mathbb{S}(X_1) = 1$ and $\mathbb{S}(X_2) = 1.1$, do X_1 and X_2 have significantly different performance?

We can take some "reference" distribution, for example $Z \sim N(0,1)$.

Define

$$f(x) = \mathbb{S}(Z + x), \qquad x > 0$$

and the "normalized" Sharpe ratio

$$\bar{\mathbb{S}}(X) = f^{-1}(\mathbb{S}(X))$$

"S.R. of a normal r.v. with the same $\mathbb{S}$", $\qquad \bar{\mathbb{S}}(N(\mu, \sigma^2)) = \dfrac{\mu}{\sigma}$

$\bar{\mathbb{S}}(X) > \frac{EX}{\sigma(X)} \implies$ "better" than the normal distribution

$\bar{\mathbb{S}}(X) < \frac{EX}{\sigma(X)} \implies$ "worse" than the normal distribution

In a similar way, for any $p \geqslant 1$ we can define

$$f_p(x) = \mathbb{S}_p(Z + \sigma_p x), \qquad \bar{\mathbb{S}}_p(X) = f_p^{-1}(\mathbb{S}_p(X))$$

where $\sigma_p = \sigma_p(Z) = \|Z\|_p$ (for example, $\sigma_1 = \sqrt{2/\pi}$, $\sigma_2 = 1$, etc.)

Then $\bar{\mathbb{S}}_p(x + Z) = S_p(x + Z) = \frac{x}{\sigma_p}$.

$\bar{\mathbb{S}}_p(X) > S_p(X)$ $(\ldots < \ldots)$ $\implies$ "better" ("worse") than normal

Computations for $p = 1, 2$

$$\mathbb{S}_p(Z + \sigma_p x) = ((\mathbb{P}_p(Z, \sigma_p x))^{-p} - 1)^{\frac{1}{p}}$$

Since $\mathbb{P}_p$ and $\mathbb{Q}_p$ are inverse, we can find that if $\mathbb{S}_p(X) \in (0, \infty)$, then

$$\bar{\mathbb{S}}_p(X) = \frac{1}{\sigma_p}\mathbb{Q}_p(Z,\ \lambda) \text{ with } \lambda = 1 - (1 + \mathbb{S}_p(X))^{-\frac{1}{p}}$$

Moreover, recall that for $p = 1$, CVAR can be computed explicitly:

$$\mathbb{Q}_1(Z, \lambda) = \frac{1}{(1-\lambda)\sqrt{2\pi}} e^{-\frac{q(\lambda)^2}{2}}$$

where $q(\lambda)$ is the standard normal quantile function. Then

$$\bar{\mathbb{S}}_1(X) = \frac{1+\mathbb{S}_1(X)}{2} e^{-\frac{q(\lambda)^2}{2}} \text{ with } \lambda = \frac{\mathbb{S}_1(X)}{1+\mathbb{S}_1(X)}$$

Funds from the Morningstar Database

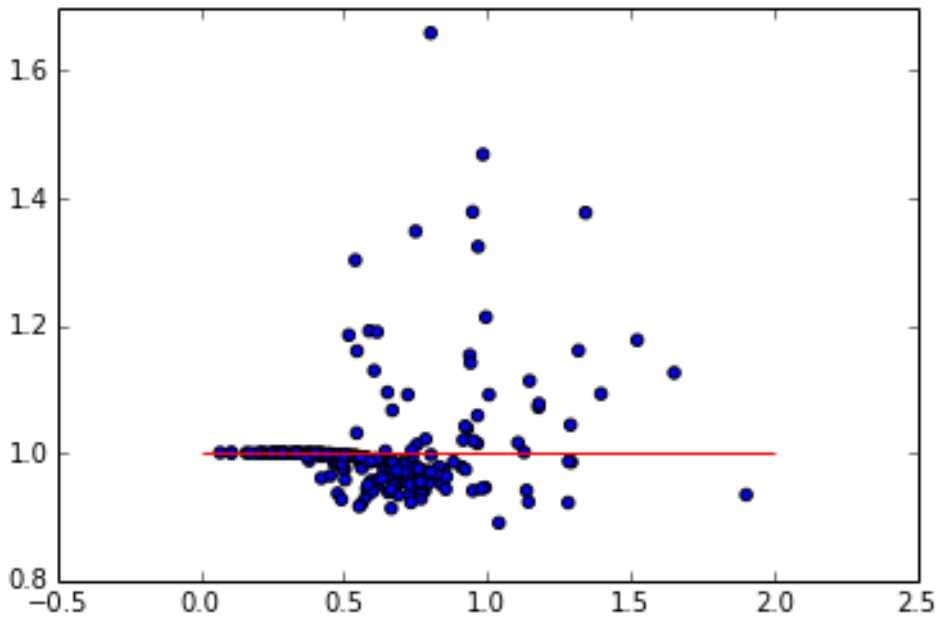
The Morningstar CISDM[2] Database:

- The oldest hedge fund and CTA database in the market
- More than 6000 funds since 1994
- Includes live as well as dead funds

We'll look at funds active in 1995-2009 (262 funds) and plot their Sharpe ratio $S_2(X)$ and compare with the normalized Sharpe ratio $\bar{\mathbb{S}}_2(X)$.

[2] Center for International Securities & Derivatives Markets

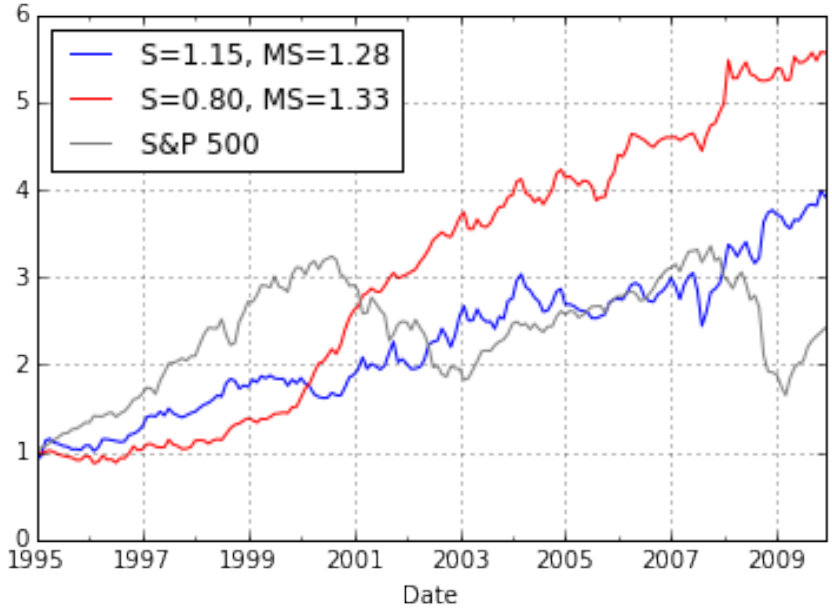
$S(X)$ and $\bar{\mathbb{S}}(X)$ for different funds in 1995–2009.

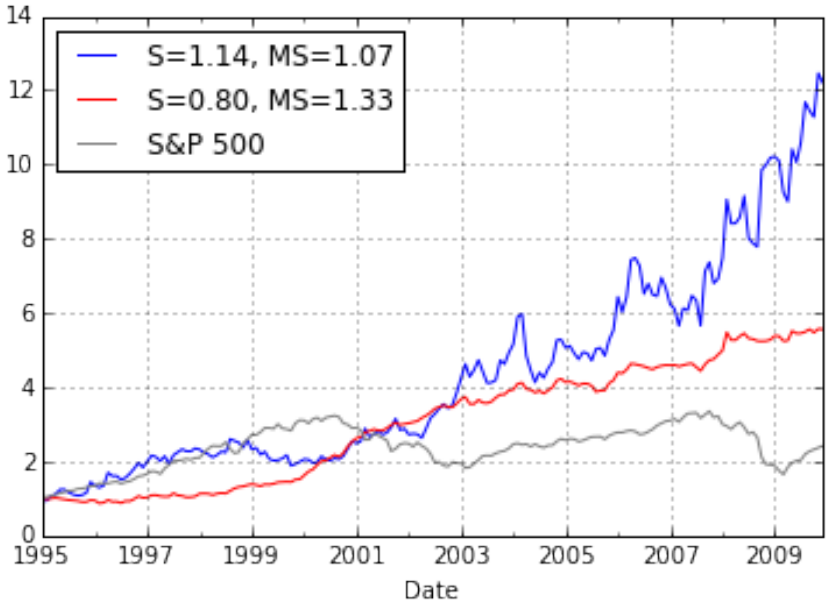


x-axis: $S(X)$, y-axis: $\bar{\mathbb{S}}(X)/S(X)$

Below the red line: "worse" than the normal distribution.

A fund with high $\bar{\mathbb{S}}/S$ (red) compared with two other funds (blue):





7. Compensation for consumption

We can extend the definition of $\mathbb{S}_p$ to allow to trade the disposed amount $X - Y$ in the future for a cash equivalent $v(X - Y)$ now:

$$\mathbb{S}_{p,v}(X) = \sup_{Y \leqslant X} \frac{E(Y + v(X - Y))}{\sigma_p(Y)}$$

Here $Z = X - Y$ is the disposed part of the profit (consumption), which is a considered as a derivative security sold for $v(Z)$ at time $t = 0$.

We consider only the simplest case when the premium $v(Z)$ is invested into the risk-free asset.

We'll assume that

$$v(Z) = E(RZ),$$

where $R \in L^q$, $R \geqslant 0$ and $E(RX) = 0$.

The main example: the density of an equivalent martingale measure, maybe multiplied by a discount factor.

It is assumed that $R \not\equiv 1$, i.e. the actual measure and the martingale measure are different (otherwise $\mathbb{S}_{p,v}(X) = 0$).

A representation theorem for $\mathbb{S}_{p,R}$

The following representation holds:

$$
\begin{aligned}
&(\mathbb{S}_{p,R}(X))^q = \\
&\qquad = \max_{a,b\in\mathbb{R}} \Big\{ E\Big(|\widetilde{R}|^q - q(b-aX)\widetilde{R} + (q-1)|b-aX|^p \Big) \times \\
&\qquad\qquad \times \mathbf{I}\big\{ b - aX < \operatorname{sgn}(\widetilde{R})|\widetilde{R}|^{\frac{p}{q}} \big\} - (q-1)|b-aX|^p - aq)
\end{aligned}
$$

where $\widetilde{R} = R - 1$.

Note that the function in $\max\{\ldots\}$ is convex in a, b.

The proof is similar to the case when $R = 0$, but reduces to the optimization problem

$$\begin{aligned} \text{minimize:} \quad & \|R' - 1\|_q \\ \text{subject to:} \quad & R' \geqslant R \text{ a.s.} \\ & ER' = 1 \\ & E(R'X) = 0 \end{aligned}$$

Financial measures of risk and performance

MATRIX, 20–21 November 2017

Part 4: Acceptability indices and performance measures

Mikhail Zhitlukhin
Steklov Mathematical Institute, Moscow

Introduction

Recall the properties of $\mathbb{S}_p(X)$:

1. Quasi-concavity
2. Scaling invariance
3. Law invariance
4. Monotonicity
5. Continuity

We'll consider general functionals satisfying these properties.

Following Cherny and Madan (2009), these functionals will be called acceptability indices or performance measures.

1. Axiomatics of acceptability indices

Definition

$\alpha\colon L^\infty \to [a,b]$ is called an acceptability index (performance measure) if

1. Quasi-concavity
   $\alpha(\lambda X + (1-\lambda)Y) \geqslant \min(\alpha(X), \alpha(Y))$
2. Monotonicity (w.r.t. 2nd order dominance)
   if $EU(X) \geqslant EU(Y)$ for any incr. concave U, then $\alpha(X) \geqslant \alpha(Y)$
3. Scale invariance
   $\alpha(\lambda X) = \alpha(X)$ for any $\lambda > 0$
4. Fatou property (weak-$*$ l.s.c.)
   if $|X_n| \leqslant 1$, $\alpha(X_n) \geqslant x$ and $X_n \xrightarrow{P} X$, then $\alpha(X) \geqslant x$
5. Law invariance
   if $X \stackrel{d}{=} Y$, then $\alpha(X) = \alpha(Y)$

Remarks

- We consider L^∞ in order not to worry about integrability, but all the interesting results can be extended to L^p
- Without loss of generality $a = 0$, $b = +\infty$, as in Cherny-Madan. Otherwise, apply a monotone function $[a, b] \to [0, +\infty]$
- Cherny and Madan require only 1-4 (without the law invariance), but all interesting examples are law-invariant
- Under the law invariance, the second-order monotonicity is equivalent to a.s.-monotonicity

Examples

1. $\mathbb{S}_p$ – the monotone Sharpe ratio on L^p

2. $1 - \mathbb{P}(-X, 0)$ is an acceptability index with values in $[0, 1]$

   Note: $1 - \mathbb{P}(-X, 0) \not\equiv \mathbb{P}(X, 0)$

3. "Gain–loss ratio" is an acceptability index with values in $[0, \infty]$:

$$GLR(X) = \begin{cases} \frac{EX}{EX_-}, & \text{if } EX > 0 \\ 0, & \text{otherwise} \end{cases}$$

Relation to risk measures

Theorem

A functional α is an acceptability index if and only if there exists a non-decreasing[1] family of risk measures $\rho_x(X)$, $x \in \mathbb{R}_+$ such that

$$\alpha(X) = \sup\{x \in \mathbb{R}_+ : \rho_x(-X) \leqslant 0\}$$

"The largest level x at which the risk of loss is still acceptable"

For example (recall that $\mathbb{Q}_p(\cdot, \lambda)$ is non-decreasing in λ):

$$\text{for } \mathbb{S}_1: \ \rho_x(-X) = \mathbb{Q}_1\Big(-X, \ \frac{x}{1+x}\Big)$$

$$\text{for } \mathbb{S}_2: \ \rho_x(-X) = \mathbb{Q}_2\Big(-X, \ 1 - \tfrac{1}{\sqrt{1+x^2}}\Big)$$

[1] $x \mapsto \rho_x(X)$ is non-decreasing for any X

More examples

1. Since $\mathbb{Q}_p(\cdot, \lambda)$ is non-decreasing in λ, for any non-decreasing function $f\colon [0, \infty] \to [0, 1]$, an acceptability index can be defined by

$$\rho_x(X) = \mathbb{Q}_p(-X, f(x)).$$

   Using that $1 - \mathbb{P}(-X, 0) = \sup\{\lambda : \mathbb{Q}(-X, \lambda) \leqslant 0\}$, such an acceptability index is

$$\alpha(X) = f^{-1}(1 - \mathbb{P}_p(-X, 0))$$

2. Cherny and Madan defined $AIT(X) =$ "Tail VAR acceptability index" as $\mathbb{S}_1$ (but with different notation)

2. Extreme measures. Comparison of acceptability indices

Let $\mathcal{P} = \{R \in L^1 : R \geqslant 0, ER = 1\}$ be the set of densities of equivalent probability measures.

For any x, let $\mathcal{R}_x$ be the maximal subset of $\mathcal{P}$ such that

$$\rho_x(X) = \sup_{R \in \mathcal{R}_x} E(RX)$$

Then $\mathcal{R}_x$ is convex and closed in L^1 and explicitly given by

$$\mathcal{R}_x = \{R \in \mathcal{P} : E(RX) \leqslant \rho_x(X) \text{ for any } X\}$$

From now on, we'll $\sup$ in ρ_x is attained.

For all $X \in L^\infty$, we'll denote by $R_x(X)$ extreme measures, i.e. such that

$$\rho_x(X) = E(R_x(X)X)$$

(may be not unique, but in many examples unique and $\sigma(X)$-measurable) and denote

$$R^*(X) = R_{\alpha(X)}(X)$$

Proposition (Cherny, 2007)

If $R_x(X)$ exists and is unique, then for any $Y \in L^\infty$

$$\lim_{\varepsilon \downarrow 0} \frac{\rho_x(X + \varepsilon Y) - \rho(X)}{\varepsilon} = E(R_x(X)Y)$$

Interpretation

If X is the loss of a portfolio, and εY is the loss of an additional marginal trade, then

- if $E(R_x(X)Y) > 0$, then the trade εY decreases the acceptability of the whole portfolio
- if $E(R_x(X)Y) < 0$, then the trade εY increases the acceptability of the whole portfolio

Reasonable properties of extreme measures

If $R_x(X)$ is $\sigma(X)$-measurable, the following properties of the function $r(z) = R_x(z)$ are desirable:

1. $r(z)$ is non-decreasing and strictly increasing at least for $z \geqslant z_0$: "better to lose additional money Y when X is small rather than large";
2. $r(z) \to \infty$ as $z \to +\infty$
3. $r(z) \to 0$ as $z \to -\infty$

Compare with dis-utility (convex) functions:

$$\lim_{\varepsilon \downarrow 0} \frac{EU(X + \varepsilon Y) - EU(X)}{\varepsilon} = E(U'(X)Y)$$

For example, $U(x) = e^{ax}$, where $a > 0$, satisfies the above properties for $r(z) = U'(z)$.

Example 1: acceptability indices based on $\mathbb{P}_p$

Consider $\mathbb{Q}_p(X,\lambda) = \min_{c\in\mathbb{R}}(\frac{1}{1-\lambda}\|(X-c)_+\|_p + c)$

If for some X, the minimum is attained at $c^* = c^*(X,\lambda)$, then

$$R(X) = \frac{(X-c^*)_+^{p-1}}{E\big((X-c^*)_+^{p-1}\big)}$$

In particular, the function

$$z \mapsto \frac{(z-c^*)_+^{p-1}}{E\big((X-c^*)_+^{p-1}\big)}$$

satisfies the above properties 1–3 if $p > 1$, but not if $p = 1$.

From this point of view, $\mathbb{Q}_p$, $p > 1$, are better than $\mathbb{Q}_1$.

Example 2: gain-loss ratio

Recall the definition of the gain-loss ratio:

$$GLR(X) = \begin{cases} \frac{EX}{EX_-}, & \text{if } EX > 0 \\ 0, & \text{otherwise} \end{cases}$$

Proposition

The gain-loss ratio is defined by

$$\rho_x(X) = \sup_{R \in \mathcal{R}_x} E(RX), \qquad x > 0$$

where

$$\mathcal{R}_x = \{R = c(1+Y) \mid c > 0,\ 0 \leqslant Y \leqslant x,\ E(R) = 1\}$$

The extreme measures for a continuous r.v. X are of the form

$$R_x^* = \frac{1 + x\mathbf{I}(X \geqslant a)}{1 + xP(X \geqslant a)}$$

for an appropriate choice of a. Doesn't satisfy properties 1–3.

3. Acceptability indices based on probability distortion

Let $\Psi_x \colon [0,1] \to [0,1]$ be a family of probability distortion functions parametrized by $x \in [0,\infty)$ and decreasing in x:

1. $y \mapsto \Psi_x(y)$ is convex, increasing and continuous
2. $\Psi_x(0) = 0$, $\Psi_x(1) = 1$
3. $x \mapsto \Psi_x(y)$ decreases for any $y \in (0,1)$

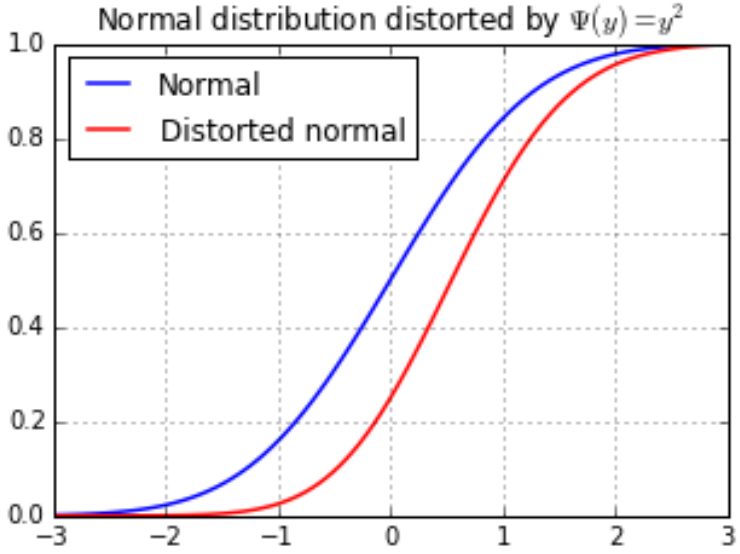
We will apply $\Psi(y)$ to distribution functions of random variables: if $F(x)$ is a distribution function, then $\Psi(F(x))$ is again a distribution function.

Example

$\Psi(y) = y^2$ is a distortion function.

If $F(x)$ is a distribution function then $\Psi(F(x)) = P(\max(X_1, X_2) \leqslant x)$, where $X_1, X_2 \sim F$ i.i.d.

For example, for the standard normal distribution:



Coherent risk measures based on distortion

Let $\Psi_x(y)$ be a decreasing family of distortion functions. Then

$$\rho_x(X) = \int_{\mathbb{R}} y d\Psi_x(F_X(y))$$

defines an increasing family of law-invariant coherent risk measures.

Proof

By a change of variables,

$$\rho_x(X) = \int_0^1 \mathbb{Q}_1(X, \lambda)\mu_x(d\lambda), \quad \text{where} \quad (\Psi_x(y))' = \int_{[0,y)} \frac{1}{1-\lambda}\mu(d\lambda)$$

Hence $\rho_x(X)$ is a comonotone law-invariant risk measure.

Remark: the converse is also true, a comonotone risk measure can be expressed through a distortion function.

Extreme measures for ρ_x

For X with a continuous distribution

$$R_x^*(X) = (\Psi_x)'(F_X(X))$$

(Ψ_x' is defined except maximum a countable number of point; this is enough to define $R^*(X)$ a.s., since $F_X(X)$ has the uniform distribution).

See Föllmer, Schied (2004), Corollary 4.74.

Properties of extreme measures we'd like to have:

1. $z \mapsto R_x^*(z)$ is non-decreasing: follows from the convexity of Ψ
2. $R_x^*(z) \to \infty$ as $z \to \infty$: need to check $(\Psi_x)'(y) \to \infty$ as $y \to 1$
3. $R_x^*(z) \to 0$ as $z \to -\infty$: need to check $(\Psi_x)'(y) \to 0$ as $y \to 0$

Let's construct examples of acceptability indices from families of distortion risk measures:

$$\alpha(X) = \sup\Big\{x : \int_{\mathbb{R}} y d\Psi_x(F_X(y)) \leqslant 0\Big\}$$

Example 1

Let $\Psi_x(y) = y^{x+1}$. In particular, if x is integer, then

$$\rho_x(X) = EY, \qquad \text{where } Y \stackrel{d}{=} \max\{X_1, \ldots, X_{x+1}\}$$

where X_i are independent copies of X.

In this case

$$\Psi'_x(y) = (x+1)y^x$$

Problem: $\Psi_x(y) \to (x+1) < +\infty$ as $y \to 1$.

Example 2

Let $\Psi_x(y) = 1 - (1-y)^{\frac{1}{x+1}}$. If x is integer, then:

$$\rho_x(X) = EY, \qquad \text{where } \min\{Y_1, \ldots, Y_{x+1}\} \stackrel{d}{=} X$$

where Y_i are independent copies of y.

In this case

$$\Psi'_x(y) = \frac{1}{x+1} \cdot \frac{1}{(1-y)^{\frac{x}{x+1}}}$$

Problem: $\Psi_x(y) \to \frac{1}{x+1} > 0$ as $y \to 0$.

Example 3

Let $\Psi_x(y) = 1 - (1 - y^{x+1})^{\frac{1}{x+1}}$. If x is integer, then

$$\rho_x(X) = EY, \qquad \text{where } \min\{Y_1, \ldots, Y_{x+1}\} \stackrel{d}{=} \max\{X_1, \ldots, X_{x+1}\}$$

where X_i are independent copies of X; Y_i are independent copies of Y.

We have

$$\Psi'_x(y) = \frac{(x+1)y^x}{(1 - y^{x+1})^{\frac{x}{x+1}}}$$

This function satisfies properties 1–3.

Financial measures of risk and performance

MATRIX, 20–21 November 2017

Part 5: Dynamic maximization of the Sharpe ratio

Mikhail Zhitlukhin
Steklov Mathematical Institute, Moscow

1. Dynamic portfolios that maximize the Sharpe ratio

The market model

- One riskless asset with price $B_t \equiv 1$
- One risky asset, a geometric Brownian motion

$$dS_t = S_t(\mu dt + \sigma dW_t), \qquad S_0 = 1$$

- A trading strategy is identified with two predictable control processes v_t, u_t, the amount of money invested in the risky and riskless assets. The capital $X_t^{u,v} = v_t + u_t$ of a self-financing strategy controlled by v_t, u_t satisfies the SDE

$$dX_t^{u,v} = v_t \frac{dB_t}{B_t} + u_t \frac{dS_t}{S_t}, \qquad X_0^{u,v} = x_0$$

In a similar way, $B_t = e^{rt}$ can be considered.

– Since $B_t \equiv 1$, everything can be expressed only in terms of u_t (and then $v_t = X_t - u_t$):

$$dX_t^u = \mu u_t dt + \sigma u_t dW_t, \qquad X_0^u = x_0$$

– The class of admissible control processes $\mathcal{U} = \{u_t : E \int_0^T u_t^2 dt\} < \infty$.

We consider the problem[1]

$$\text{maximize } S(u) = \frac{EX_T^u - x_0}{\sigma(X_T^u - x_0)} \quad \text{over } u \in \mathcal{U}.$$

Without loss of generality, below $x_0 = 0$ and $\mu > 0$, $\sigma > 0$.

[1] Note: the HJB approach cannot be applied directly here.

A brief literature review

Typically, the problem is first reduced to the constraint optimization problem

$$\begin{aligned} \text{minimize:} \quad & E(X_T^u)^2 \\ \text{subject to:} \quad & EX_T^u = \delta \end{aligned}$$

1. Richardson (1989) solved this problem by martingale methods:

First, find the optimal terminal capital X_T^*.

Then, find the process X_t^*.

Finally, find the optimal u_t^*.

Let M be the density of $dQ = M dP$, so that S is a Q-martingale: from Girsanov's theorem $M = \exp(-\frac{\mu}{\sigma} W_T - \frac{\mu^2}{2\sigma^2} T)$.

Since $x_0 = 0$, the set of terminal values X_T that can be obtained

$$H = \{X \in L^2 : E^Q X = \langle X, M \rangle = 0\}$$

where $\langle \cdot, \cdot \rangle$ is the scalar product in L^2 w.r.t. P, i.e. $\langle X, Y \rangle = E(XY)$.

Then the problem becomes

$$\begin{aligned} \text{minimize:} \quad & \langle X, X \rangle \text{ over } X \in L^2 \\ \text{subject to:} \quad & \langle X, M \rangle = 0 \\ & \langle X, 1 \rangle = \delta \end{aligned}$$

which is solved by standard methods.

2. Pedersen, Peskir (2013) solved the problem by applying HJB to the Lagrangian:

$$\text{minimize } E(X_T^u)^2 - \lambda E X_T^u \qquad (*)$$

where the Lagrange multiplier $\lambda > 0$ is a parameter has to be determined.

The HJB equation:

$$\inf_{u \in \mathbb{R}} \left\{ V_t' + \mu u V_x' + \tfrac{\sigma^2}{2} u^2 V_{xx}'' \right\} = 0$$
$$V(T, x) = x^2 - \lambda x$$

where $V(t, x)$ is the minimal value of $(*)$ under $X_t = x$.

Minimizing over $u \in \mathbb{R}$, obtain

$$V_t' = \frac{\mu^2}{2\sigma^2} \cdot \frac{(V_x')^2}{V_{xx}''} \qquad \text{and} \qquad u(t, x) = -\frac{\mu}{\sigma^2} \cdot \frac{V_x'(t, x)}{V_{xx}''(t, x)}$$

The key step is to look for the solution in the form

$$V(t,x) = a(t)x^2 + b(t)x + c(t)$$

where the coefficients $a(t)$, $b(t)$ and $c(t)$ can be found explicitly.

Then apply a verification theorem to $V(t,x)$.

3. For other results related to this problem, see the overview in Pedersen, Peskir (2013).

Remark: the Hamilton–Jacobi–Bellman equation

For a controlled diffusion process

$$dX_t^u = \mu(t, X_t^u, u_t)dt + \sigma(t, X_t^u, u_t)dW_t$$

consider the optimization problem

$$V(t,x) = \min_{u \in \mathcal{U}} E_{t,x} F(X_T^u)$$

where $E_{t,x}(\cdot) = E(\,\cdot \mid X_t^u = x)$.

The HJB equation:

$$\inf_{u \in \mathbb{R}} \mathcal{L}V = 0, \qquad \text{where } \ (\mathcal{L}V)(t,x,u) = V_t' + \mu V_x' + \tfrac{1}{2}\sigma^2 V_{xx}''$$

"Derivation": apply Itô's formula to

$$V(t,x) = \inf_{u \in \mathcal{U}} E_{t,x} V(t+\varepsilon, X_{t+\varepsilon}^u)$$

2. A solution through the monotone Sharpe ratios

Step 1: the monotone Sharpe ratio as the optimality criterion

We'll first solve the following problem for $p > 1$:

$$V = \min_{u \in \mathcal{U}} \mathbb{P}_p(-X_T^u, 0)$$

From the representation of $\mathbb{P}_p(X, x)$:

$$V = \min_{c \geqslant 0} \min_{u \in \mathcal{U}} \|(1 - cX_T^u)_+\|_p = \min_{u \in \mathcal{U}} \|(1 - X_T^u)_+\|_p$$

where in the second equality we used that

$$dX_t^u = \mu u_t dt + \sigma u_t dW_t, \qquad X_0^u = 0,$$

so the constant c can be included in the control.

Then

$$V^p = \min_{u \in \mathcal{U}} E|\widetilde{X}_T^u|^p$$

for

$$d\widetilde{X}_t^u = -\mu u_t dt - \sigma u_t dW_t, \qquad \widetilde{X}_0^u = 1.$$

Note that $(\cdot)_+$ was removed since it's never optimal to go below zero.

Let $v_t = -u_t/\widetilde{X}_t$ and assume $E\exp\big(\frac{\sigma^2 p^2}{2}\int_0^T v_t^s dt\big) < \infty$ for optimal u. Then

$$E|\widetilde{X}_T^u|^p = E\Big\{Z\exp\Big(\int_0^T \big(\mu p v_s + \tfrac{1}{2}\sigma^2(p^2-p)v_s^2\big)ds\Big)\Big\}$$

where Z is the stochastic exponent $Z = \mathcal{E}(\sigma p v_t)$ and $Z > 0$, $EZ = 1$ by Novikov's condition.

By changing the measure to $dQ = ZdP$ we obtain

$$E(\widetilde{X}_T^u)^p = E^Q\left\{\exp\left(\int_0^T \left(\mu p v_s + \tfrac{1}{2}\sigma^2(p^2 - p)v_s^2\right)ds\right)\right\}$$

which is minimized by minimizing the integrand for each t:

$$v^* \equiv -\frac{\mu}{\sigma^2(p-1)}$$

Then an optimal control in the original problem

$$u_t = \frac{\mu}{\sigma^2(p-1)}(1 - X_t^u)$$

Since any control $\widetilde{u}_t = Cu_t$ will be optimal as well, the family of optimal control functions is

$$u(t,x) = \frac{\mu}{\sigma^2(p-1)}(C-x), \qquad C > 0$$

Then the capital processes of strategies with minimal $\mathbb{P}$ satisfy the SDE

$$dX_t = \frac{\mu^2(C-X_t)}{\sigma^2(p-1)}dt + \frac{\mu(C-X_t)}{\sigma(p-1)}dW_t, \qquad X_0 = 0$$

or $X_t = C - Y_t$, where Y_t is the geometric Brownian motion

$$\frac{dY_t}{Y_t} = -\frac{\mu^2}{\sigma^2(p-1)}dt - \frac{\mu}{\sigma(p-1)}dW_t, \qquad Y_0 = C.$$

Verification of the assumption that an optimal strategy satisfies the condition $E\exp(\frac{\sigma^2 p^2}{2}\int_0^T v_t^s dt) < \infty$ can be done by localizing by $\tau = \inf\{t : \widetilde{X}_t^u = \varepsilon\}$.

Step 2: why $\mathbb{S}$-optimality is the same as S-optimality

Formally, the previous step provides the solution to the problem:

$$\text{maximize } \mathbb{S}(X_T^u) = \sup_{Y \leqslant X_T^u} \frac{EY}{\sigma(Y)} \quad \text{over } u \in \mathcal{U}.$$

Why it won't be optimal to dispose of some profit?

Suppose for optimal u^* there is $Y \leqslant X_T^{u^*}$ such that $S(Y) > S(X_T^{u^*})$. Since the BS-market is complete, there is $y < 0$ and a control u_t so that $X_0^u = y_0 < 0$ and $X_T^u = Y$.

Then the capital process $\widetilde{X}_t = y_0 + X_t^u$ has a higher Sharpe ratio than Y and a higher monotone Sharpe ratio than X^{u*} – a contradiction.

3. The dynamic mean-variance problem. Time inconsistency

Consider the problem

$$\begin{aligned} \text{minimize:} \quad & E(X_T^u)^2 \\ \text{subject to:} \quad & EX_T^u = \delta, \; X_0^u = x_0 \end{aligned}$$

where $\delta > 0$ and $x_0 < \delta$ are any constants.

In the "Markov setting[2]":

$$V(t,x) = \min_{u \in \mathcal{U}_t} \{ E_{t,x}(X_T^u)^2 \mid E_{t,x} X_T^u = \delta \}$$

where $E_{t,x}(\,\cdot\,) = E(\,\cdot \mid X_t = x)$ and $\mathcal{U}_t$ is the set of controls u_s admissible for $s \geqslant t$.

[2] Not true Markov: the optimal control will be different for different (t,x).

Suppose for (t, x) the minimum is provided by the control

$$u^{t,x} = u^{t,x}(s, y), \qquad s \geqslant t,\ y \in \mathbb{R}.$$

In a standard Markov problem we would have

$$u^{t_1,x_1}(s, y) = u^{t_2,x_2}(s, y) \qquad \text{for any } t_1 \leqslant t_2 \leqslant s \text{ and } x_1, x_2, y \in \mathbb{R}.$$

We'll solve the dynamic mean-variance problem and see it doesn't have this property.

Solution

The optimal control $u(t,x)$ can be found from the problem:

$$\min_{u\in\mathcal{U}}\{E(\widetilde{X}^u_{T-t})^2 \mid E\widetilde{X}^u_{T-t} = \delta - x,\ \widetilde{X}^u_0 = 0\}$$

If u^* is optimal here, then $u^{t,x}(s,y) = u^*(s-t, y-x)$.

From the previous section:

$$u^*(r,\widetilde{x}) = u^*(\widetilde{x}) = \frac{\mu}{\sigma^2}(C-\widetilde{x}), \qquad r\in[0,T-t],\ \widetilde{x}\in\mathbb{R},$$

where the constant $C = C(t,x)$ is found from the condition

$$E\widetilde{X}^{u^*}_{T-t} = \delta - x$$

We found that

$$\widetilde{X}_r^{u^*} = C - Y_r \qquad \text{with } \frac{dY_r}{Y_r} = -\frac{\mu^2}{\sigma^2}dt - \frac{\mu}{\sigma}dB_r, \;\; Y_0 = C.$$

This gives

$$C = (\delta - x)\Big(1 - e^{\frac{\mu^2}{\sigma^2}(t-T)}\Big)^{-1}$$

and the optimal control for $V(t,x)$

$$u^{t,x}(s,y) = u^{t,x}(y) = \frac{\mu}{\sigma^2}\Big\{(\delta - x)\Big(1 - e^{\frac{\mu^2}{\sigma^2}(t-T)}\Big)^{-1} + x - y\Big\}$$

We obtain

$$u^{t_1,x_1}(y) \neq u^{t_2,x_2}(y)$$

4. The mean-variance optimal selling problem

As before, one risky asset:

$$dS_t = S_t(\mu dt + \sigma dW_t), \qquad S_0 = 1.$$

For a parameter $x \geqslant 1$, we'll consider the problem

$$\text{maximize } \frac{ES_\tau - x}{\sigma(S_\tau)} \text{ over Markov times } \tau$$

(for $x = 1$: "buy and hold" problem)

The interesting case is only when $\mu \in (0, \frac{\sigma^2}{2})$:

- $\mu \leqslant 0$: the optimal $\tau = 0$
- $\mu \geqslant \frac{\sigma^2}{2}$: the process $S_t = \exp(\sigma B_t + (\mu - \frac{\sigma^2}{2})t)$ reaches any level A; take $\tau = \tau_A$ and $A \to +\infty$

The paper by Pedersen, Peskir (2012)

Pedersen and Peskir considered an equivalent problem

$$\text{maximize } ES_\tau - c\,\mathrm{Var}\,S_\tau$$

Their solution briefly:

1. Equivalence to a constrained problem: $\min_\tau\{ES_\tau^2 \mid ES_\tau = \delta\}$
2. Lagrange multipliers: $\min_\tau\{ES_\tau^2 - \lambda ES_\tau\}$
3. Markov formulation: the value function $V(s) = \min_\tau E_s(S_\tau^2 - \lambda S_\tau)$, where $E_s(\cdot) = E(\,\cdot \mid S_0 = s)$
4. Introduce the stopping set $D = \{s : V(s) = s^2 - \lambda s\}$, and the continuation set $C = \mathbb{R}_+ \setminus D = \{s : V(s) < s^2 - \lambda s\}$. From the general theory, the optimal stopping time $\tau = \inf\{t \geqslant 0 : S_\tau \in D\}$

How to find C and D?

A "guess": $D = \{s : s \geqslant b\}$, where b is an unknown stopping boundary.

Solve the free-boundary problem:

$$\begin{aligned}
&\mathcal{L}V(s) = 0 \quad \text{for } s \in C \\
&V(0+) = 0 \\
&V(b) = b^2 - \lambda b \quad \textit{(continuous fit)} \\
&V'(b) = 2b - \lambda \quad \textit{(smooth fit)}
\end{aligned}$$

where $\mathcal{L}V = \frac{1}{2}\sigma^2 s^2 V'' + \mu s V'$.

Three unknown parameters (b and constants C_1, C_2 from the ODE) and three conditions $\implies$ solution.

Then apply a verification theorem.

A solution through the monotone Sharpe ratio

First, let's maximize the monotone Sharpe ratio, or, equivalently

$$\text{minimize } \mathbb{P}_2(-(S_\tau - x), 0) \text{ over Markov times } \tau.$$

From the representation of $\mathbb{P}_2(X, 0)$:

$$V = \min_{c \geqslant 0} \min_{\tau} E(1 - c(S_\tau - x))_+^2$$

The inner problem $\min_\tau E(\ldots)$ is Markovian, and by a simple argument it can be shown that the solution is

$$\tau_c = \inf\{t \geqslant 0 : S_t \geqslant b_c\}$$

where $\tau_c = \infty$ if $S_t < b_c$ for all t.

Hence, for the optimal τ^* the distribution of S_{τ^*} is binomial:

$$S_{\tau^*} = \begin{cases} b^*, & \text{if } \tau^* < \infty \\ 0, & \text{if } \tau^* = \infty \end{cases}$$

Next observe that if τ^* maximizes $\mathbb{S}(S_\tau)$, then

$$\mathbb{S}(S_\tau) := \sup_{Y \leqslant S_{\tau^*}} \frac{EY - x}{\sigma(Y)} = \frac{ES_{\tau^*} - x}{\sigma(S_{\tau^*})}$$

i.e. $Y = S_{\tau^*}$, which follows from that only $\mathscr{F}^{S_{\tau^*}}$-measurable Y can be considered.

Hence, the same τ^* maximizes both $\mathbb{S}$ and S.

The optimal level b can be found as follows. Let $p_b = P(\tau_b < \infty)$:

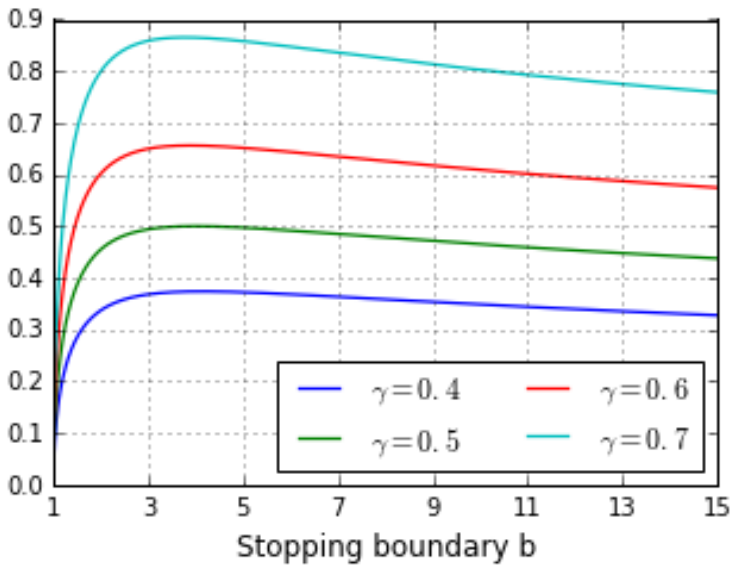
$$\begin{aligned} p_b &= P(S_t \geqslant b \text{ for some } t < \infty) \\ &= P(\sigma W_t + (\mu - \tfrac{\sigma^2}{2})t \geqslant \log b \text{ for some } t < \infty) \\ &= b^{\gamma-1}, \qquad \text{where } \gamma = \tfrac{2\mu}{\sigma^2} < 1 \end{aligned}$$

Then

$$\frac{ES_{\tau_b} - x}{\sigma(S_{\tau_b})} = \frac{bp_b - x}{b\sqrt{p_b(1-p_b)}} = \frac{b^\gamma - x}{b^{\frac{\gamma+1}{2}}(1-b^{\gamma-1})^{\frac{1}{2}}}$$

Choose $b > 1$ which maximizes the right-hand side (no explicit formula for general γ and x).

The right-hand side $\mathrm{RHS}_\gamma(b)$ for different b and γ, and $x = 1$.



Observe that since $\gamma < 1$

$$\mathrm{RHS}_\gamma(b) \to 0 \text{ if } b \to 0 \text{ or } b \to \infty$$

so the maximum exists.