Statistical Hedging

Motivating the use of Convex Risk Measures for Hedging Portfolios of Derivatives over one Time Step in the presence of General Transaction Cost

A Summary for Derivative Quants

Hans Buehler
QR Equities & Investor Services
JP Morgan*

First version: January 2013; this version: March 2019

Abstract

This note discusses how we may construct a hedge for a portfolio of derivatives in the presence of general transaction cost for one step. This automatically leads to a concept of risk-adjusted *cost* as a measure of risk.

Mathematically, our focus is the intuitive construction of "quadratic" convex risk measures as an extension of the classic Markoviz-style mean-variance approach. Our proposed approach boils down to using

$$\inf_{\boldsymbol{\delta},W} \left\{ \mathrm{cost}(\boldsymbol{\delta}) - \operatorname{ess\,inf} W + \mathrm{risk}(Z+\boldsymbol{\delta}\mathbf{H}-W) \,\middle|\, \mathrm{risk}(Z+\boldsymbol{\delta}\mathbf{H}-W) \leq \mathrm{risklimit} \right\} .$$

Here, $\mathbf{H}$ is a vector of hedging instruments and W is a potential write-off. The "risk" term is meant being quadratic functional such as $X \mapsto \|\min\{X,0\}\|_2$ which leads to our preferred "Quadratic CVaR" convex risk measure. In particular, our method has no need for computing "Greeks". We comment on the use of greeks for hedging is a separate section.

This note contains few genuine contributions, and should be seen as a guide on existing research, in particular of the two articles Ben-Tal/ Teboulle [3] and Filipovic/Kupper [10]. It summarizes results presented at Global Derivatives 2013 [6] and 2014 [7] where we presented results on hedging exotics with the methods discussed here.

It also motivates the use of Convex Risk Measures in our more recent work on the multi-step problem "Deep Hedging" [8].

*Opinions expressed in this paper are those of the authors, and do not necessarily reflect the view of JP Morgan

Contents

<u>Version History</u>

- V0.9: initial release.
- V0.92: fixed mistake for norm-based convex risk measures (i.e. CVaR and Quadratic CVaR are naturally normalized); added remark 4.2 on re-hedging.
- V0.93: added proofs to show that π (corollary 4.1) and ρ are convex risk measures (prop 2.1); added few more comments on arbitrage in section 4.2; removed the need for trading cost to be convex. The function c is now assumed to be non-negative, normalized, increasing on the positive axis, and decreasing on the negative axis. We also added a comment in convexity of trading cost with remark 4.1.

- V0.931: renamed "upper continuity" into the more commonly used "continuity from above".
- V0.932, 0.933: minor stylistic changes.
- V0.933: fixed various sign errors around utilities. A utility u is increasing and concave; the associated risk kernel is $\rho(X) = -\mathbb{E}[\,u(X)\,]$.
- V1.0: added references to our work on "Deep Hedging" [8].

1 Introduction

This note considers the problem of hedging a portfolio of derivatives using a range of hedging instruments under transaction costs and liquidity constraints over one period such as a week. In our more recent work "Deep Hedging" [8] we consider the same problem over several hedging steps; the present work is the theoretical underpinning of our approach there: for example, contrary to common derivative literature, we focus here on the hedging of instruments marked using complex derivative models, but under real life hedging conditions. This is the standard set up for literature on "risk measures" in the sense of "capital", most notably Artzner, Delbaen, Eber and Heath's seminal article [2] and Föllmer and Schied's book [12].

Assume that we are given a portfolio of derivatives with mark-to-model values Z_t. We wish to hedge the return of Z over the period $[0,\tau]$ by investing into a range of hedging instruments with values $\mathbf{H}_t = (H_t^1, \dots, H_t^n)$. We denote by $\mathrm{d}Z$ and $\mathrm{d}\mathbf{H}$ the returns over the period $[0,\tau]$.

The aim of this note is to present an approach akin to the Markoviz paradigm for hedging, i.e. we wish to provide a sensible interpretation of a setup of the form

$$\inf_{\boldsymbol{\delta}} \left\{ \ \mathrm{cost}(\boldsymbol{\delta}) + \mathrm{risk}(Z + \boldsymbol{\delta}\mathbf{H}) \ \middle| \ \mathrm{risk}(Z + \boldsymbol{\delta}\mathbf{H}) \le \mathrm{risklimit} \ \right\} .$$

We will show that "convex risk measures" are a well-motivated, and numerically tractable choice for a concept of "risk" in such a setup. We assume that in addition to being able to hedge with $\mathbf{H}$, we are also allowed "writing off" any part W of our portfolio. Then, our proposal is to use

$$\inf_{\boldsymbol{\delta}, W} \left\{ \ \mathrm{cost}(\boldsymbol{\delta}) - \operatorname{ess\,inf} W + \mathrm{risk}(Z + \boldsymbol{\delta}\mathbf{H} - W) \ \middle| \ \mathrm{risk}(Z + \boldsymbol{\delta}\mathbf{H} - W) \le \mathrm{risklimit} \ \right\} .$$

as a measure for the cost of risk of Z.

Our approach departs from nearly all classic derivative literature in that we do not make use of any "Greeks", and that we see "models" merely as methods of producing model values upon which we design an otherwise non-parametric statistical hedging strategy. Most specifically, we do not assume that the models used for obtaining derivative values are in any way related, or "equivalent" to the real world market.

This note is structured as follows: section 2 introduces our set up and motivates the use of convex risk measures, in particular constructed via the "montone cash-invariant hull by Filipovic/Kupper [10]. We also show, as a minor original contribution, when such measures are "relevant" in the sense that they give finite measures of risk.

In section 3 we discuss examples and explain why our preferred measure of risk is "quadratic CVaR",

$$\eta(Z) := \lambda \| \min\{Z, 0\} \|_2 \ ,$$

Section 4 covers the use of our convex risk measures for determining a hedging strategy and deriving a price for new trades. The main result is that the relative cost of hedging Z is given by the program

$$\mathcal{R}(Z) := \quad \inf_{\boldsymbol{\delta}\in\mathcal{L},\, W\in\mathcal{W}} \Big\{ \underbrace{c(\boldsymbol{\delta}) - \operatorname{ess\,inf} W_\tau}_{\substack{\text{relative}\\\text{implementation}\\\text{cost}}} + \underbrace{\mu\big(\mathrm{d}\Pi\big)}_{\substack{\text{risk}\\\text{penalty}}} \ \Big| \ \underbrace{\nu\big(\mathrm{d}\Pi\big) \le 1}_{\substack{\text{PnL risk}\\\text{limit}}} \\ \mathrm{d}\Pi := \mathrm{d}Z + \boldsymbol{\delta}\mathrm{d}\mathbf{H} - W_\tau \Big\} \ .$$

We also comment on marginal pricing of a derivative P which we are to sell to a client.

In our final section 5 we discuss how our results relate to the standard practise of using "Greeks" to manage risk and why this is an inferior approach.

This note contains few original contributions. It is meant as a motivation to use convex risk measures and provides (we hope) a more accessible discussion than that of Filipovic/Kupper's very comprehensive work [10], which we will quote extensively.

2 Risk

To reflect the practical situation in a typical investment bank, we assume that positions are valued internally "to model" with some classic derivative pricing models. In contrast to most literature on derivative risk management, we do not assume that any of these internal pricing models are particularly related to actual market dynamics: specifically, we do not assume that our models are using a martingale measure equivalent to the statistical measure, or that the market is complete. Rather, we are agnostic towards the pricing models used, and see them merely as maps which provide mark-to-model values for instruments based on prevailing market data. The only constraint we want to impose is that if an instrument is liquidly traded, then the model value falls between bid and ask prices to avoid obvious mark-to-market vs. mark-to-model "arbitrage".

Our discussion covers generic derivatives, securities, cash, currency and other tradable instruments which can be valued either by using market prices or some derivative pricing model. For simplicity it is convenient to just refer to "derivatives" as the most general term. However, all of the below also applies to portfolios of cash equities, in which case the "model" value becomes, say, the volume-weighted mid-price.

Technical Conditions

Notation – we define the usual operators $a \wedge b := \min\{a, b\}$, $a \vee b := \max\{a, b\}$ and $a^+ := \max\{a, 0\}$ as well as $a^- := (-a)^+$.

Measures – we assume we are given a probability space with a "statistical" measure $\mathbb{Q}$, which we have estimated, essentially, with classic statistical methods. See the presentations [6] and [7] as a reference.

This measure $\mathbb{Q}$ also defines all zero sets. That means that quantities such as essential infima or suprema are defined with respect to $\mathbb{Q}$, in the sense $\operatorname{ess\,inf} X := \sup\{x \in \mathbb{R} : \mathbb{Q}[X \geq x] = 1\}$ where $\sup \emptyset = -\infty$. On the envisaged use-case of a finite probability space, it makes sense to assume that $\mathbb{Q}$ has full support, i.e. that no state we observe has zero probability under $\mathbb{Q}$. Furthermore, we assume that all variables of interest such as portfolio value, hedges, and write-offs are square-integrable with respect to $\mathbb{Q}$, c.f. Toussaint and Sircar in [16] on motivating a square-integrable setup. We also denote by $\mathcal{P}$ the set of all probability measures $\mathbb{P}$ which are absolutely continuous with respect to $\mathbb{Q}$. Most specifically, this implies that $\mathbb{E}_{\mathbb{Q}}[X] \geq \operatorname{ess\,inf} X$ for all X.

We use $\mathbb{E}$ to refer to the expectation operator under $\mathbb{Q}$.

Interest – we assume that over the horizon of interest, rates are deterministic. To ease notation, we therefore assume future values are implicitly discounted by the respective discount factor. We may therefore assume for simplicity that rates are just zero.

We refer to El Karoui and Ravanelli's work [9] on the topic of asymmetric funding or stochastic rates in the context of convex risk measures.

The Market

Portfolio and Hedging Instruments – we denote by Z a generic (long) portfolio position of derivatives we want to hedge, each valued with its respective derivative pricing model. We use Z_t to refer to its (mark-to-) model value at t; today is $t = 0$. As it is a total value, it also includes any cash position, e.g. those accrued as part of trading the position Z. In the case of liquid instruments, the mark-to-model value is the official book value, e.g. for equity a volume-weighted mid-price.

Our focus is to hedge the exposure arising from Z after a period τ, Z_τ. The total return[1] over this period is denoted by $\mathrm{d}Z := Z_\tau - Z_0$.[2] Note that we assume

[1] We assume throughout that any cashflows such as coupons or dividends are accrued within the valuation of the respective derivative or security. At expiry, the value of a derivative becomes its settlement value, accrued accordingly when time passes. In this sense, each position is "self-financed": there are no cashflows in our out of the position.

[2] In practical applications, we would usually consider the weekly PnL (Proft-and-Loss) arising from daily delta-hedging within $[0, \tau]$, where the delta used is the respective model

that we have proper return samples, and not just "delta-gamma" approximated returns as often used in Greek-based VaR, e.g. Alexander, Coleman and Li [1]. See our presentations [6] and [7] on how to generate such returns.

To hedge Z, we are given a range of hedging instruments $\mathbf{H}$ with model values $\mathbf{H}_t = (H_t^1, \ldots, H_t^n)'$ and return vector $\mathrm{d}\mathbf{H}$ to τ. This could be any range of instruments: shares, equity options, CDS, swaptions, FX spot, FX butterflies, etc.

We are also simplifying the *cost of carry* in that we assume that there is a uniform carry cost "numeriare" for both long and short cash positions, by which all model values are discounted back to time $t = 0$. Not that this assumption does not preclude that some instruments having a non-zero expected return. It does allow us imposing the following very basic no-arbitrage condition that $Z_0 \geq \operatorname{ess\,inf} Z_\tau$ and $H_0^i \geq \operatorname{ess\,inf} H_\tau^i$, i.e. today's model value is at least the worst-case model value of tomorrow.

Cost – to enter into a position $\boldsymbol{\delta} = (\delta^1, \ldots, \delta^n) \in \mathbb{R}^n$ of $\mathbf{H}$ today, we have to pay the market price $\hat{H}(\boldsymbol{\delta})$, rather than the model value $\boldsymbol{\delta}\mathbf{H}_0$. A negative market price $\hat{H}(\boldsymbol{\delta})$ means that we are getting paid to enter into the position, for example if $\boldsymbol{\delta}$ represents a short position in a hedging instrument with a non-negative payoff. We note that we do not assume that $\hat{H}$ is linear to allow for cross-instrument price effects.

In order to avoid obvious mark-to-model arbitrage we assume that the trading cost function $c(\boldsymbol{\delta}) := \hat{H}(\boldsymbol{\delta}) - \boldsymbol{\delta}\mathbf{H}_0$ is normalized to $c(0) = 0$ and non-negative. The latter means that the price to enter into a position exceeds its model value; in this sense we assume that all of our pricing models are "well-calibrated".

It is mathematically not necessary – but certainly more realistic – to also impose that c is increasing on $\mathbb{R}^n_+$ and decreasing on $\mathbb{R}^n_-$, i.e. to assume that buying more or selling more incurs higher trading cost.

Liquidity and Limits – if we model limited liquidity as infinite trading cost, then the available liquidity in $\mathbf{H}$ can be expressed as the affine set $\{\boldsymbol{\delta} \in \mathbb{R}^n : |\hat{H}(\boldsymbol{\delta})| < \infty\}$. We further assume that our trading is subject to various market risk limits such as greek exposure and stress scenario limits. These are typically linear and affine. We will assume that the respective limits can be reached with the available liquidity, hence that the resulting intersection of allowed trading activity and available liquidity, $\mathcal{L} := \{\boldsymbol{\delta} \in \mathbb{R}^n : |\hat{H}(\boldsymbol{\delta})| < \infty\} \cap \{\text{market risk limits}\}$, is not empty.

Write-Offs – the final component of our discussion is the ability to "write off" arbitrary payoffs of our portfolio over the period τ at a cost of their worst-case outcome: assume we wish to write off W_τ, such that that the new portfolio value becomes $Z_\tau - W_\tau$, then the cost for doing so is $-\operatorname{ess\,inf} W_\tau$. An extreme example is obviously $W_\tau = -Z_\tau$, in which case the entirety of the portfolio is written off at the cost of the worst-case outcome of the overall portfolio.

---

delta.

Also note that if the market price of entering into a hedging instrument H^i is higher than the best-case outcome, then we will not trade in that instrument but prefer to write-off the respective payoff if necessary.

We symbolically refer to write-offs as W, with values $W_0 = 0$ and $W_\tau \in \mathcal{L}^2_\tau$. We define by $\mathcal{W} := \{W_\tau \in \mathcal{L}^2_\tau : \operatorname{ess\,inf} W_\tau > -\infty\}$ the space of reasonable write-offs.

In the subsequent text we will occasionally drop the index "τ" in order to increase readability.

Risk vs. Cost – the task at hand is to find for a given long portfolio Z a "best" hedge $(\boldsymbol{\delta}^*, W^*) \in \mathcal{L} \times \mathcal{W}$ such that the resulting hedged position

$$Z + \boldsymbol{\delta}^* \mathbf{H} - W^*$$

is optimal in the sense that it has the lowest cost to bringing the overall position within acceptable risk limits. We therefore need a notion of "risk" in order to make such an optimal choice.

2.1 Risk as a Cost

In this section we aim to construct intuitively a family of measures of risk which can subsequently be used for practical optimization of portfolios of derivatives. Our notion of risk is in fact much more like a "risk-adjusted cost".[3] We will use the variable X throughout to denote a generic square-integrable random variables representing a long position.

The most natural property of any notion of risk is *diversification* or, in mathematical language, *convexity*: it simply means that the risk of a joint position is less than the sum of the individual risks.

Definition 2.1 (Risk Pre-Kernel) *We call a map $\eta : L^2(\mathbb{Q}) \to \bar{\mathbb{R}}$ a* risk pre-kernel *if it is convex and normalized to $\eta(0) = 0$.*
We also impose the technical condition of continuity from above: assume that $X_n \downarrow X$ in L^2, then $\rho(X_n) \uparrow \rho(X)$, c.f. Toussaint and Sircar in [16] proposition 2.4.

However, our concept of *risk* is to be different than that of raw "uncertainty". We aim to combine the overall expected benefit of a position with the probability of a loss. Most importantly, a notion of "risk" needs to be actionable: it is to dictate an action of to either add add or remove money from a position. This means that we wish to develop a notion of risk which can be interpreted as a (risk-adjusted) cost.

An immediate consequence of this is that *risk* ought to be monotone decreasing: risk (as in cost) must be to be lower for a position which is strictly better than another.

[3] We distinguish between 'cost" as an internal hurdle rate and "prices" as a negotiation statement.

Definition 2.2 (Risk Kernel) *We call a pre-kernel η a* (proper) risk kernel *if it is monotone decreasing, i.e. $\eta(X_1) \leq \eta(X_2)$ whenever $X_1 \geq X_2$. We will generally just refer to η as being "monotone".*

This formulation disqualifies notions of *risk* which are actually notions of "uncertainty" such as plain variance.

A trivial example of a proper risk kernel for X is the "careless" kernel,

$$X \mapsto 0 \ .$$

Under this kernel, nothing has any risk.

Another extreme example is the worst-case risk measure

$$X \mapsto -\operatorname{ess\,inf} X$$

which embodies total risk aversion. It represents an upper bound on any reasonable risk measure as we will show in our section 2.3 on "relevance".

A further well-known group of kernels are the "risk-neutral" expectation operators with respect to a measure $\mathbb{P} \in \mathcal{P}$,

$$X \mapsto -\mathbb{E}_{\mathbb{P}}\left[\, X \,\right] \ .$$

Such measures $\mathbb{P}$ might be chosen to comparatively stress (presumably extreme) scenarios;[4] indeed, in all the following examples, we may use a measure $\mathbb{P}$ other than $\mathbb{Q}$ to define expectations or norms.

Classic non-trivial examples for mere pre-kernels are the Markoviz-style [13] mean-variance and mean-volatility operators

$$X \mapsto -\mathbb{E}\left[\, X \,\right] + \lambda \left\| X - \mathbb{E}X \right\|_2^2 \quad \text{and} \quad X \mapsto -\mathbb{E}\left[\, X \,\right] + \lambda \left\| X - \mathbb{E}X \right\|_2 \ . \tag{1}$$

Both are are convex, c.f. proposition 3.1 on page 14. However, both are well-known for not being monotone.[5] Hence they are merely pre-kernels. From a practical point of view, they are unsatisfactory since they penalize upside gains as much as downside losses – in other words, they are not actually measures of "risk", but measures of "uncertainty".

This non-monotonicity of the Markoviz pre-kernels (1) is rectified by using their downside-risk "semi-quadratic" risk-kernels

$$X \mapsto -\mathbb{E}\left[\, X \,\right] + \lambda \left\| (X - \mathbb{E}X)^- \right\|_2^p \quad p \in \{1, 2\} \tag{2}$$

which are both monotone and convex, c.f. proposition 3.2 on page 15. These are intuitively more appropriate measures of risk since they penalize only losses. (Unfortunately, we will show in section 3.2.1 on page 17 that this class of pre-kernels leads to trivial linear convex risk measures.)

[4]Recall that $\mathbb{P}$ is absolutely continuous w.r.0 $\mathbb{Q}$, which makes sure that $\operatorname{ess\,inf} X \leq \mathbb{E}_{\mathbb{P}} X$.

[5]As an example for $\lambda \uparrow \infty$, consider the payoff zero, compared to a random payoff which pays 0 or 1 randomly.

The final, related, class of examples are the CVaR-style risk kernels of the form

$$X \mapsto \lambda \left\| X^- \right\|_p^q \tag{3}$$

for $\lambda \geq 1$. The most famous case is CVaR itself, given by $p = q = 1$, where λ is typically given as $\lambda := 1/(1-\alpha)$ for a confidence level $\alpha \in [0, 1)$.

The quadratic "Quadratic CVaR" case $p = 2$ and $q = 1$ is our preferred risk model as it is numerically efficient to implement and provides a standard interpretation of risk.

2.2 Convex Risk Measures

Given a pre-kernel η, and a long position X, we may now ask what is the optimal write-off in order to achieve the minimum stated risk. This translates into the program

$$\rho(X) := \inf_{W \in \mathcal{W}} \Big\{ \underbrace{-\operatorname{ess\,inf} W_\tau}_{\text{cost}} + \underbrace{\eta(X - W)}_{\text{risk}} \Big\} \ . \tag{4}$$

We may write any $W \in \mathcal{W}$ as $W = W^+ - w$ where $w \in \mathbb{R}$ and $W^+ \geq 0$. Hence, we may write

$$\rho(X) = \inf_{w \in \mathbb{R}, W^+ \geq 0} \Big\{ w + \eta(X + w - W^+) \Big\} \ . \tag{5}$$

Moreoever, if η is a proper risk measure, monotonicity implies that (4) simplifies to

$$\rho(X) = \inf_{w \in \mathbb{R}} \Big\{ w + \eta(X + w) \Big\} \ . \tag{6}$$

The monotone, convex functional ρ is the "monotone cash-invariant hull" of μ as discussed in Ben-Tal/Teboulle's work [3] or in Filipovic/Kupper's rather technical article [10]. (Also note that the "inf" here is properly defined over L^2 due to the assumption of upper continuity; c.f. Toussaint and Sircar's proposition 2.4 in [16].) In other words, ρ is a *convex risk measure* a'la Föllmer and Schied [12]:

Definition 2.3 (Convex Risk Measure) *A functional ρ is called a* convex risk measure *if it has the following properties:*

1. ***Monotonicity**: it is cheaper to hedge a more favourable position. For two portfolios $X^1 \geq X^2$ pricing is monotone decreasing, i.e. $\rho(X^1) \leq \rho(X^2)$.*

2. ***Convexity (Diversification)**: hedging parts of the same portfolio separately raises the cost of doing so. Formally: for any two portfolios X^1 and X^2 and any $\alpha \in [0, 1]$, $\alpha' := 1 - \alpha$ we have $\rho(\alpha X^1 + \alpha' X^2) \leq \alpha\rho(X^1) + \alpha'\rho(X^2)$.*

3. ***Cash-Invariance**: if cash is added to a position, then the Cost of Risk drops by that amount: $\rho(X + c) = \rho(X) - c$ for all $c \in \mathbb{R}$.*

On $L^2(\mathbb{Q})$ we also impose the technical condition

4. ***Continuity from Above****: For any $X_n \downarrow X$ in L^2, we have $\rho(X_n) \uparrow \rho(X)$.*

Note that if η is already cash-invariant, then

$$\rho(X) = \inf_{W^+ \geq 0} \Big\{ \eta(X - W^+) \Big\} \ . \tag{7}$$

Cash-Invariance implies in particular that

$$\rho(X + \rho(X)) = 0 \ ,$$

which means $\rho(X)$ has a clear interpretation as the required cash amount to add to our position such that the new position has zero risk. At first sight, one may argue that maintaining *zero* risk is economically imprudent, and that we should allow ourselves some capacity to warehouse risk. However, this is a misunderstanding of the term "risk". For us, "risk" here is really a "cost", and having zero cost for a non-zero position is not desirable. Further information can be found in Föllmer and Schied's book [12].

We now extend our definition (4) to the case where we use a pre-kernel both for risk-adjustment and as a constraint. This captures the common case where a portfolio optimization problem is solved under a total risk constraint, for example on the risk defined by a set of stylized risk-factors such as industry sectors.

Proposition 2.1 *Assume that μ and ν are pre-kernels. Then,*

$$\varrho(X) := \inf_{W \in \mathcal{W}} \Big\{ -\operatorname{ess\,inf} W_\tau + \mu(X - W) \ \Big| \ \nu(X - W) \leq 1 \ \Big\} \tag{8}$$

is a convex risk measure.
We may also write

$$\varrho(X) = \inf_{w \in \mathbb{R}, W^+ \geq 0} \Big\{ -w + \mu(X + w - W^+) \ \Big| \ \nu(X + w - W^+) \leq 1 \ \Big\} \ .$$

Moreover

1. *If both μ and ν are normalized, then ϱ is normalized.*
2. *If both μ and ν are monotone, then (8) simplifies to*

$$\varrho(X) = \inf_{w \in \mathbb{R}} \Big\{ w + \mu(X + w) \ \Big| \ \nu(X + w) \leq 1 \ \Big\} \tag{9}$$

Proof – We first show that ϱ is monotone: assume that $X \geq Y$. For any W, $\eta(X - W) \leq \eta(X - W)$ for $\eta \in \{\mu, \nu\}$, hence monotonicity follows.

To see convexity, assume $\alpha \in (0,1)$ and set $\alpha' := 1-\alpha$. Then,

$$
\begin{aligned}
&\alpha\varrho(X)+\alpha'\varrho(X)\\
&= \alpha \inf_{U\in\mathcal{W}} \left\{ -\operatorname{ess\,inf} U_\tau + \mu(X-U) \;\middle|\; \nu(X-U)\le 1 \right\}\\
&\quad +\alpha' \inf_{V\in\mathcal{W}} \left\{ -\operatorname{ess\,inf} V_\tau + \mu(Y-V) \;\middle|\; \nu(Y-V)\le 1 \right\}\\
&= \inf_{U\in\mathcal{W},V\in\mathcal{W}} \Big\{ -\operatorname{ess\,inf} \alpha U_\tau - \operatorname{ess\,inf} \alpha' V_\tau + \alpha\mu(X-U) + \alpha'\mu(Y-V) \;\Big|\\
&\qquad\qquad \alpha\nu(X-U)\le\alpha\,,\ \alpha'\nu(Y-V)\le\alpha' \Big\}\\
&\ge \inf_{U\in\mathcal{W},V\in\mathcal{W}} \Big\{ -\operatorname{ess\,inf} \alpha U_\tau - \operatorname{ess\,inf} \alpha' V_\tau + \alpha\mu(X-U) + \alpha'\mu(Y-V) \;\Big|\\
&\qquad\qquad \alpha\nu(X-U)+\alpha'\nu(Y-V)\le 1 \Big\}\\
&\ge \inf_{U\in\mathcal{W},V\in\mathcal{W}} \Big\{ -\operatorname{ess\,inf} \alpha U_\tau - \operatorname{ess\,inf} \alpha' V_\tau + \mu(\alpha X+\alpha' Y-(\alpha U+\alpha' V)) \;\Big|\\
&\qquad\qquad \nu(\alpha X+\alpha' Y-(\alpha U+\alpha' V))\le 1 \Big\}\\
&\ge \inf_{U\in\mathcal{W},V\in\mathcal{W}} \Big\{ -\operatorname{ess\,inf}\, (\alpha U_\tau + \alpha' V_\tau) + \mu(\alpha X+\alpha' Y-(\alpha U+\alpha' V)) \;\Big|\\
&\qquad\qquad \nu(\alpha X+\alpha' Y-(\alpha U+\alpha' V))\le 1 \Big\}\\
&= \inf_{W=\alpha U+\alpha' V:\, U\in\mathcal{W},V\in\mathcal{W}} \Big\{ -\operatorname{ess\,inf} W_\tau + \mu(\alpha X+\alpha' Y-W) \;\Big|\\
&\qquad\qquad\qquad \nu(\alpha X+\alpha' Y-W)\le 1 \Big\}\,.
\end{aligned}
$$

To show cash-invariance, let $c\in\mathbb{R}$. Then,

$$
\begin{aligned}
\varrho(X+c) &= \inf_{W\in\mathcal{W}} \left\{ -\operatorname{ess\,inf} W_\tau + \mu(X+c-W) \;\middle|\; \nu(X+c-W)\le 1 \right\}\\
&= \inf_{W':=W-c\in\mathcal{W}} \left\{ -\operatorname{ess\,inf}\,(W'_\tau + c) + \mu(X-W') \;\middle|\; \nu(X-W')\le 1 \right\}\\
&= \varrho(X)-c\,.
\end{aligned}
$$

If μ and ν are both normalized, then clearly $\varrho(0)\le 0$. For any W we have $\mu(-W)\ge \operatorname{ess\,inf} W_\tau$, hence $-\operatorname{ess\,inf} W_\tau + \mu(-W)\ge 0$. Therefore, $\varrho(0)=0$.

□

2.3 Relevance

Our construction of convex risk measures is, as we hope, well motivated and easy to implement. However, it has one potential deficiency: consider the pre-kernel

$$\eta(X) := \lambda\|X_\tau^-\|_p \tag{10}$$

with $\lambda\in(0,1)$. The value for $\rho(0)$ is given by $\inf_w\{w+\lambda|w^-|\}$, whose infimum for $\lambda<1$ is attained in $w\downarrow-\infty$. In other words,

$$\rho(0) = -\infty\ ,$$

which means we derive infinite benefits of holding nothing.

This occurs because the benefit of taking out what amounts to a loan of size w today outweighs the perceived risk of having to pay it back tomorrow. This is clearly not a desirable outcome.

DEFINITION 2.4 (Relevance) *We call a convex risk measure ρ* relevant *if it is finite for any bounded payoff, i.e.*

$$\rho(X) > -\infty \ .$$

for all X with $\operatorname{ess\,sup} X < \infty$*.*

The following improves upon the results in both Ben-Tal/Teboulle [3] and Filipovic/Kupper [10] since it only requires to look at the finiteness of ρ at zero:

PROPOSITION 2.2 *A convex risk measure is relevant if and only if $\rho(0) > -\infty$.*

Note that a *normalized* convex risk measure, i.e. one with $\rho(0) = 0$ is by definition relevant. For example, will show below in our example section that for $\lambda > 1$, (10) is relevant.

Proof of proposition 2.2– We only have to show one direction. Assume that $\rho(0) > -\infty$ and that X is bounded from below. Then $\rho(X) \geq \rho(\operatorname{ess\,sup} X) = \rho(0) - \operatorname{ess\,sup} X > -\infty$. □

If ρ is relevant then we may w.l.g. assume that we are given the normalized convex risk measure

$$X \longmapsto \rho(X) - \rho(0) = \rho\big(X + \rho(0)\big) \ .$$

COROLLARY 2.1 *Let ρ be a normalized convex risk measure. Then it satisfies*

- ***Boundness from Below****: $\rho(X) \geq -\operatorname{ess\,sup} X$: the Cost of Risk of a position can never be better (be more negative) than the best-case outcome.*
- ***Boundness from Above****: $\rho(X) \leq -\operatorname{ess\,inf} X$: the Cost of Risk of a position can never be worse (be more positive) than the worst-case loss.*
- ***Cash-Equivalence****: $\rho(c) = -c$ for all $c \in \mathbb{R}$.*

If η is normalized, then the upper bound is attained with the choice $W = -X$, which means that we may restrict our search space to $W \in \mathcal{W} : -W \leq X$ or, in the formulation of (5) to (w, W^+) such that $w \leq -\operatorname{ess\,inf} X$ and $0 \leq W^+ \leq X - \operatorname{ess\,inf} X$.

2.3.1 Linearity

The opposite concern to "relevance" is the question on whether a particular pre-kernel accidentally yields a too simple risk measure. Consider, for example, the rather natural risk-kernel

$$\eta(X) := -\mathbb{E}X + \lambda \|X^-\|_p^q$$

for $p, q \geq 1$ and $\lambda \geq 0$ (c.f. proposition 3.2 for convexity and monotonicity). The associated convex risk measure is

$$\rho(X) = \inf_{w \in \mathbb{R}} \Big\{ -\mathbb{E}[X] + \lambda \|(X+w)^-\|_p^q \Big\}$$

which reduces by taking $w \uparrow \infty$ to the trivial

$$\rho(X) \equiv -\mathbb{E}X \ .$$

The issue at hand is that η becomes simply linear on the positive upside. That means we are tempted to write off any negative part of X, such that $X + w \geq 0$, $w > 0$. We pay $|w|$ to do so, but this is compensated by the linear contribution w from η.

We note the following first result:

Lemma 2.1 *Assume that ρ is normalized and that*

$$\rho(X) \leq -\mathbb{E}_{\mathbb{P}}X \ , \tag{11}$$

for any $\mathbb{P} \in \mathcal{P}$. Then ρ is actually linear,

$$\rho(X) = -\mathbb{E}_{\mathbb{P}}X \ .$$

Proof – We show that $\rho(X) = -\mathbb{E}_{\mathbb{P}}X$: first, $\rho(0) = -\mathbb{E}_{\mathbb{P}}0 = 0$. Assume Y is such that $\rho(Y) < -\mathbb{E}_{\mathbb{P}}Y$. Write $0 = 0.5Y + 0.5(-Y)$, hence $-\rho(-Y) \leq \rho(Y) < -\mathbb{E}_{\mathbb{P}}X$ which yields $\rho(-Y) > -\mathbb{E}_{\mathbb{P}}(-Y)$, a contradiction to (11). □

Definition 2.5 *We call a risk measure ρ* law-invariant *under $\mathbb{P} \in \mathcal{P}$ if $\rho(X) = \rho(Y)$ for all X and Y which have the same law under $\mathbb{P}$.*

Föllmer and Knispel show in [11] in section 5.4 the intuitive result:

Theorem 2.1 *Assume that all variables are L^∞. Every law-invariant convex risk measure ρ under $\mathbb{P}$ is* linearly bounded from below *in the sense*

$$\rho(X) \geq -\mathbb{E}_{\mathbb{P}}[X] \ .$$

We therefore may write ρ in terms of the non-negative $\gamma(Y) := -\rho(Y - \mathbb{E}_{\mathbb{P}}[Y]) = \rho(Y) + \mathbb{E}_{\mathbb{P}}[Y]$ as

$$\rho(X) \equiv -\mathbb{E}_{\mathbb{P}}[X] + \gamma(X) \ .$$

As shown in [11] in section 5.4, γ is then a measure of dispersion in the sense that

- $\gamma(X) \geq 0$ *with* $\gamma(X) > 0$ *if* X *is not a constant* $\mathbb{P}$*-a.s.,*
- $\gamma(X + c) = \gamma(X)$ *for all* $c \in \mathbb{R}$.
- γ *is convex.*
- $\gamma(\lambda X) = \lambda\gamma(X)$

2.4 Related Work

We claim no novelty in the above construction of convex risk measures. We refer to Ben-Tal and Teboulle's work [3] who, to our knowledge, first used the concept of a cash-invariant hull to generate convex risk measures.

They called their approach the *Optimized Certainty Equivalent*, or *OCE*; the authors work with utility functions; recall that u is a utility function if it is normalized, monotone increasing and concave. Ben-Tal and Teboulle also comment on relevance, see their proposition 2.1 and remark 2.1. As a minor contribution from our side, we generalize those results into proposition 2.2.

Their results translate into our setup by using the risk kernel

$$\eta(X) := -\mathbb{E}[u(X_\tau)] \ . \tag{12}$$

Assuming smoothness of u, this yields the condition

$$w(X) \ : \ 1 \stackrel{!}{=} \mathbb{E}[\partial_w u(X_\tau - w)] \tag{13}$$

for the optimal w in (9). *Relevance* is then implied by the existence of a finite $w(0)$ such that $1 = u'(-w(0))$, c.f. definition 2.1 in [3].

If η is indeed relevant, then we may set

$$\rho(X) := -w(X) - \mathbb{E}[u(X_\tau - w(X))] \ , \tag{14}$$

as before.

The extension of the OCE to non-monotonic functions was provided by Filipovic and Kupper in [10] who introduced $\mathcal{P}$-monotone hulls i.e. our "prekernels". Our minor contribution is the interpretation of this being the result of allowing writing off parts of the payoff, and subsequently the introduction of hedging instruments.

3 Examples

We follow up with a list of examples for practical consideration. Nearly of these examples were discussed already in Filipovic/Kupper [10]. The only minor addition to the material there is our coverage of the mean-semi norm case.

We start with a few general comments on convexity and monotonicity:

Proposition 3.1 *The map* $X \mapsto \|X\|_p^q$ *is convex for* $p, q \geq 1$ *and all measures* $\mathbb{P} \in \mathcal{P}$.

Proof – Any norm is convex.[6] For $p > 1$, note that the function $x \mapsto x^q$ for $q \geq 1$ over $\mathbb{R}_+$ is increasing and convex, hence[7] $\|\cdot\|_p^q$ is convex for all $p, q \geq 1$. $\square$

Proposition 3.2 *The map $X \mapsto \|X^-\|_p^q$ is convex and monotone decreasing for $p, q \geq 1$.*

Proof – Following the argument for the previous proposition it is sufficient to show the claim for $q = 1$.

Let therefore $g(X) := \lambda \|\min\{0, X\}\|_p$.

We show first monotonicty: the function g is scale-invariant, i.e. $g(aX) = ag(X)$ for all $a \geq 0$, and therefore sub-additive: $\frac{1}{2}g(X+Y) = g(\frac{1}{2}X + \frac{1}{2}Y) \leq \frac{1}{2}g(X) + \frac{1}{2}g(Y)$. To show that it is decreasing, assume $X^1 \geq X^2$, such that $X^1 = X^2 + Y$ where $Y \geq 0$. Then $g(X^1) = g(X^2 + Y) \leq g(X^2) + 0$.

To show convexity, let $\alpha \in [0,1]$ and $\alpha' := 1 - \alpha$. We note that $\min\{0, \alpha X_1 + \alpha' X_2\} \geq \alpha \min\{0, X_1\} + \alpha' \min\{0, X_2\}$ because of concavity of $\min\{0, \cdot\}$. Monotonicity yields $g(\alpha X_1 + \alpha' X_2) = g(\min\{0, \alpha X_1 + \alpha' X_2\}) \leq g(\alpha \min\{0, X_1\} + \alpha' \min\{0, X_2\}) = \|\alpha \min\{0, X_1\} + \alpha' \min\{0, X_2\}\|_p$. Applying now the previously shown convexity of $\|\cdot\|_p$ shows that g is convex, too. $\square$

3.1 Generalized CVaR: Semi-Norm

We may generate a wide range of convex risk measures using an extension of the CVaR-approach. To this end, consider the monotone and convex kernel

$$\eta(X) := \lambda \|X^-\|_p^q \tag{15}$$

with $p, q \geq 1$ and $\lambda > 1$, where the norm might be defined with respect to any measure $\mathbb{P} \in \mathcal{P}$. Relevance follows from noting that the map

$$w \longmapsto -w + \lambda |w|^q$$

is minimized in $w = 0$ for $q = 1$ and $w = (q\lambda)^{1/(q-1)}$ for $q > 1$.[8] For the case $q = 2$ specifically, we get $w = 2\lambda$. See also Filipovic/Kupper [10], section 5.4.

3.1.1 CVaR ("Short Fall")

The classic example of risk measures along the lines of (15) is CVaR with $p = q = 1$ and $\lambda = 1/(1-\alpha)$ for a confidence level $\alpha \in [0,1]$, for example $\alpha = 95\%$. Note that $\lambda \in [1, \infty]$. Minimizing the unconditional objective

$$w \longmapsto -w + \lambda \mathbb{E}\left[(X - w)^-\right]$$

[6] Let $\alpha \in [0,1]$ and $\alpha' := 1 - \alpha$, then $\|\alpha X_1 + \alpha' X_2\|_p \leq \|\alpha X_1\|_p + \|\alpha' X_2\|_p$ by sub-additivity. Since $\|\alpha X_1\|_p = \alpha \|X_1\|_p$, convexity follows.

[7] We have $f(g(x))'' = [f'(g(x)) * g'(x)]' = f''(g(x))g'(x)^2 + f'(g(x))g''(x)$. Hence $f \circ g$ is convex if $'f \geq 0$ which is the case for $f(x) = x^q$ over $\mathbb{R}_+$.

[8] Assume $q > 1$, then the minimal w solves $1 = \lambda q w^{q-1}$ for $w \geq 0$.

yields $1 = \lambda\mathbb{P}[X \le w]$; in other words, the optimal point w^* satisfies

$$\mathbb{P}[X \le w^*] = 1 - \alpha \ .$$

This representation of CVaR as essentially a linear proramming problem is due to Rockafellar/Uryasev [15].

CVaR is not only a convex risk measure, but also coherent and the prototype example of the theory discussed by Artzner et.al. [2]. From a theoratical point of view, this makes it an attractive approach.

The drawback of using CVaR is that for commercially reasonable confidence levels of α, it focuses entirely on the $(1-\alpha)$'th percentile of the distribution of a portfolio. If we consider a sample space of, say, 255 data points, then using CVaR@95% means that we only care about the value of the worst 13 sample points in the distribution of a trial portfolio for a given level of average return. We are practically insensitive to the behaviour of the portfolio under "normal" circumstances. Of course, for regular distributions controlling the 95% usually results in controlling "normal" behaviour, too, but in our view the latter should rather be a genuine property of the risk measure used.

3.1.2 Quadratic CVaR

This concern is addressed by using the quadratic version of (15) with $p = 2$, i.e.

$$\eta(X) := \lambda\| \min\{0, X\}\|_2$$

for $\lambda \ge 1$. In this setup, the respective numerical optimization problem (22) remains numerically efficient since it can be represented as a cone problem.

As in the CVaR case, consider the unconditional objective

$$w \longmapsto -w + \lambda\|(X - w)^-\|_2$$

The optimal point w^* satisfies

$$\frac{1}{\lambda} = \frac{\|(w^* - X)^+\|_1}{\|(w^* - X)^+\|_2} \ ,$$

which does not have a simple interpretation as in the CVaR case. However, we may parameterize λ as $\lambda := 1 + \delta$ where δ might be interpreted as classic risk aversion in terms of standard deviation. (We saw above with the example in the beginning of section 2.3 that a value of $\lambda < 1$ does not create a relevant risk measure, because in this case the "risk" associated to entering into a positive position is smaller than the "gain" of writing it off.)

This is the favourite measure of the authors, mainly because of its efficiency of implementation and closeness to CVaR.

3.1.3 Symmetric Generalized CVaR: Norm

We shall also consider the symmetric family of pre-kernels

$$\eta(X) := \lambda\|X\|_p^q \tag{16}$$

where as before $p, q \geq 1$ and $\lambda \geq 1$. The norm might be defined w.r.t. any $\mathbb{P} \in \mathcal{P}$. These pre-kernels penalize losses as much as gains, which is not a particularly natural assumption. However, as we will see below, they are closely related to some forms of mean-risk kernels, hence we note here that (16) is a indeed a family of relevant normalized pre-kernels: in order to show relevance, assume by contradiction that there is a W such that

$$0 > -\operatorname{ess\,inf} W + \lambda \|W\|_p^q =: m(W)$$

In this case $\operatorname{ess\,inf} W > 0$, i.e. $W > 0$, but this means $\|W\| \geq \operatorname{ess\,inf} W$, hence $m(W)$ cannot be negative. This contradicts our assumption $m(0) = 0$, hence we have shown that (16) is indeed relevant.

See also Filipovic/Kupper [10], section 5.3.

3.2 Mean-Risk Kernels

In this section we investigate the relevance of a range of classic mean-risk kernels in the spirit of Markoviz' work [13]. In all cases, we combine an expected "average" return with a risk penalty term which we apply to either the outright position or its performance vs. its expected value. We consider both the symmetric and the CVaR "downside" risk case.

As a general comment, we notice that if η is a [pre-] kernel, then

$$\eta'(X) := -\mathbb{E}_{\mathbb{P}} X + \eta(X)$$

is also a [pre-] kernel for any $\mathbb{P} \in \mathcal{P}$.

3.2.1 Mean-Semi Norm

The downside mean-risk kernel is given by

$$\eta(X) := -\mathbb{E}[X] + \lambda \|X^-\|_p^q \tag{17}$$

for $p, q \geq 1$ and $\lambda \geq 0$. It is decreasing. As shown in section 2.3.1, the associated convex risk measure is actually linear

$$\rho(X) \equiv -\mathbb{E}[X] \ .$$

This disqualifies the otherwise very natural kernel (17).

3.2.2 Mean-Norm

The symmetric centralized mean-risk pre-kernel is given by

$$\eta(X) := -\mathbb{E}[X] + \lambda \|X\|_p^q$$

for $p, q \geq 1$ and $\lambda \geq 0$; c.f. Filipovic/Kupper [10], section 5.3. The function

$$W \longmapsto -\operatorname{ess\,inf} W + \mathbb{E} W + \lambda \|W\|^q$$

has a minimum in $W = 0$ following similar arguments to those for showing (16). Therefore, η is relevant and normalized. This setup allows those who dislike the downside formulation (17) still using quadratic mean-risk. Here, λ is truly the sought-after standard deviation.

3.2.3 Mean-Semi Deviation

The downside centralized case is given by

$$\eta(X) := -\mathbb{E}[X] + \lambda \|(X - \mathbb{E}X)^-\|_p^q$$

for $p, q \geq 1$ and $\lambda \geq 0$. It is convex but not decreasing for $\lambda > 1$, c.f. Filipovic/Kupper [10], section 5.2. The case $\lambda \in [0, 1]$ is too restrictive to be of practical use hence we focus on the general non-monotone case: we see that

$$W \longmapsto -\operatorname{ess\,inf} W + \mathbb{E}W + \lambda \|W - \mathbb{E}W\|_p^q$$

attains its minimum in $W = 0$, simply because the norm only ever contributes positively to the right hand side. Hence, η is normalized and therefore relevant. While it is not monotone, it is cash-invariant, which means that computing ρ only requires the reduced form (7).

3.2.4 Mean-Deviation

The symmetric centralized case is given by

$$\eta(X) := -\mathbb{E}[X] + \lambda \|X - \mathbb{E}X\|_p^q$$

for $p, q \geq 1$ and $\lambda \geq 0$. Again, this is a normalized relevant cash-invariant pre-kernel. See also Filipovic/Kupper [10], section 5.1. Its program for computing ρ is reduced to (7).

In the special case where X is normal, the mean-deviation measure is equal to the certainty equivalent of the exponential utility, and therefore *per se* a convex risk measure.

3.3 Certainty Equivalent of the Exponential Utility

The exponential utility is given as

$$u(x) := \frac{1}{\lambda}\left(1 - e^{-\lambda x}\right) \tag{18}$$

for a risk-aversion parameter $\lambda > 0$.

It is normalized, increasing and concave, hence we may define the associated risk kernel as in (12) as $\eta(X) := -\mathbb{E}[u(X_\tau)]$. The respective condition (13) for w is

$$1 = \mathbb{E}\left[-\partial_w \frac{1}{\lambda} e^{-\lambda(X_\tau - w)}\right] = e^{\lambda w}\mathbb{E}\left[e^{-\lambda X_\tau}\right]$$

which yields $w = -\frac{1}{\lambda} \log \mathbb{E}\left[e^{-\lambda X_\tau} \right]$. Plugging this into (14) gives, as expected, the (standard) *certainty equivalent of the expected utility*,

$$\rho(X) = \frac{1}{\lambda} \log \mathbb{E}\left[e^{-\lambda X_\tau} \right] \ . \tag{19}$$

If considered for a normal X with mean m and volatility σ, i.e. $X = m + \sigma N$ with N standard normal, we obtain

$$\rho(X) = \frac{1}{2}\lambda\sigma^2 - m \ .$$

This gives the above an interpretation of a mean-"variance" based risk measure a'la Markoviz [13].

We note that (19) is simply the classic utility uncertainty equivalent of the exponential utility:

DEFINITION 3.1 *Recall that u is called a* utility function *if it is increasing and concave. The* certainty equivalent *of u is defined as*

$$U(X) := -u^{-1}\mathbb{E}[u(X)] \ .$$

The exponential utility (18) has inverse $u^{-1}(y) = -\frac{1}{\lambda}\log(1-\lambda y)$. Its certainty equivalent is therefore indeed (19).

REMARK 3.1 *U is also not always convex, c.f. Ben-Tal and Teboulle [4], theorem 5.1. An example of a non-trivial utility which yields a convex certainty equivalent is*

$$u(x) := (\alpha x + b)^{\frac{\alpha-1}{\alpha}}$$

REMARK 3.2 *U is not generally cash-invariant; this is here a consequence of the exponential utility having "constant absolute risk aversion".*

In fact, the only utility for which the uncertainty equivalent is a convex risk measure is the exponential utility with the linear limit case included, c.f. Ben-Tal and Teboulle [4], corollary 5.1 and the commentary thereafter.

4 Hedging

In the previous section we have learned how to derive convex risk measures from (hopefully) natural assumptions by just using the ability to write off parts of our portfolio returns. In this section, we go back to our original question on how to use a range of hedging instruments to hedge a long position Z using a range of hedging instruments $\mathbf{H}$.

Definition 4.1 (Cost of Risk) *Assume that ν and μ are two pre-kernels. The* Cost of Risk *for a position Z is then given by the program*

$$\pi(Z) \quad := \quad \inf_{\boldsymbol{\delta}\in\mathcal{L},\, W\in\mathcal{W}} \Big\{ \underbrace{\hat{H}(\boldsymbol{\delta}) - \operatorname{ess\,inf} W_\tau}_{\substack{\textit{implementation}\\ \textit{cost}}} + \underbrace{\mu(\Pi_\tau)}_{\substack{\textit{risk}\\ \textit{penalty}}} \ \Big|\ \underbrace{\nu(\Pi_\tau) \le 1}_{\substack{\textit{risk}\\ \textit{constraint}}} \quad \underbrace{\Pi := Z + \boldsymbol{\delta}\mathbf{H} - W}_{\textit{hedged position}} \Big\} \tag{20}$$

We call any $(\boldsymbol{\delta}^*, W^*)$ which minimizes the Cost of Risk an optimal hedge *for Z.*

Note in particular that the ability to "write off" parts – and therefore potentially all – of the portfolio Z ensures that the risk constraint $\nu \le 1$ can always be achieved.

Note that we may write

$$\pi(Z) = \inf_{\boldsymbol{\delta}\in\mathcal{L}} \Big\{ \hat{H}(\boldsymbol{\delta}) + \varrho\big(Z + \boldsymbol{\delta}\mathbf{H}\big) \Big\} = \inf_{\boldsymbol{\delta}\in\mathcal{L}} \Big\{ c(\boldsymbol{\delta}) + \varrho\big(Z + \boldsymbol{\delta}\mathrm{d}\mathbf{H}\big) \Big\} \ , \tag{21}$$

where ϱ was defined in (8) as

$$\varrho(Z) := \inf_{W\in\mathcal{W}} \Big\{ -\operatorname{ess\,inf} W_\tau + \mu(Z - W) \ \Big|\ \nu(Z - W) \le 1 \Big\} \ .$$

Corollary 4.1 *π is a convex risk measure.*

Proof – Cash-Invariance and monotonicity follow directly by construction from the cash-invariance and monotonicity of ϱ. To show convexity, let $\alpha \in (0,1)$ and set $\alpha' := 1 - \alpha$. Then,

$$\begin{aligned} \pi(\alpha X + \alpha' Y) &= \inf_{\boldsymbol{\delta}\in\mathcal{L}} \Big\{ \hat{H}(\boldsymbol{\delta}) + \varrho(\alpha X + \alpha' Y + \boldsymbol{\delta}\mathbf{H}) \Big\} \\ &\le \inf_{\boldsymbol{\delta}\in\mathcal{L}} \Big\{ \alpha(\hat{H}(\boldsymbol{\delta}) + \varrho(X + \boldsymbol{\delta}\mathbf{H})) + \alpha'(\hat{H}(\boldsymbol{\delta}) + \varrho(Y + \boldsymbol{\delta}\mathbf{H})) \Big\} \\ &\le \alpha\pi(X) + \alpha'\pi(Y) \ , \end{aligned}$$

hence ϱ is a convex risk measure. □

If μ and ν are normalized, then $\pi(0) \le 0$, but per se we cannot assume that $\pi(0) = 0$. In fact, when using statistical data to estimate the returns of hedging instruments vs. current market prices, it is rather likely that one finds a case of *statistical arbitrage*. We will comment on this further in section 4.2 below.

Remark 4.1 (Convexity of Trading Cost) *We had only assumed that our trading cost c is non-negative and normalized. That means that the function*

$$\boldsymbol{\delta} \mapsto \hat{H}(\boldsymbol{\delta}) + \varrho(X + \boldsymbol{\delta}\mathbf{H})$$

may be rather irregular. However, if c is itself convex, then so is above function,[9] *hence our program (21) is a convex optimization problem.*

However, convex trading cost functions exclude the case of fixed transaction fees, e.g. trading costs of the form

$$\hat{H}(\boldsymbol{\delta}) = \sum_{i=1}^{n} k_0^i 1_{\delta^i \neq 0} + H_0^i \left(\, k_b^i \max\{0, \delta^i\} + k_s^i \max\{0, -\delta^i\} \, \right) .$$

To come back to π: in case the two kernels μ and ν are monotone decreasing, we obtain the simplified result:

PROPOSITION 4.1 *If μ and ν are proper risk kernels, then the infimum over random variables W in (20) is attained in a real number w, and we may write the Cost or Risk as*

$$\pi(Z) \quad := \quad \inf_{\boldsymbol{\delta}\in\mathcal{L},\, w\in\mathbb{R}} \Big\{ \underbrace{\hat{H}(\boldsymbol{\delta}) + w}_{\substack{\textit{implementation}\\ \textit{cost}}} + \underbrace{\mu\big(\Pi_\tau\big)}_{\substack{\textit{risk}\\ \textit{penalty}}} \; \Big| \; \underbrace{\nu\big(\Pi_\tau\big) \leq 1}_{\substack{\textit{risk}\\ \textit{constraint}}} \quad \underbrace{\Pi := Z + \boldsymbol{\delta}\mathbf{H} + w}_{\textit{hedged position}} \Big\} \tag{22}$$

Problem (22) is computationally much more efficient than (20) as it does not require optimization over the entire sample space via W in the way (20) does.

INTERPRETATION 4.1 *The Cost of Risk for Z represents the* additional external *cash requirement (a negative numbers implies a possible withdrawal) for holding the position Z after an optimal hedge $(\boldsymbol{\delta}^*, W^*)$ has been put in place. Note that this covers the cost of implementing the hedge itself.*

In particular, the relationship

$$\pi(Z + \pi(Z)) = 0$$

means we may interpret $\pi(Z)$ as the all-in cost of Z which maintains a nominally riskless position.[10]

REMARK 4.2 *Note that the use of (truly) convex trading cost implies typically that*

$$\pi\big(\pi(Z) + Z + \boldsymbol{\delta}^*\mathbf{H} - \hat{H}(\boldsymbol{\delta}^*) - W^* + \operatorname{ess\,inf} W^*\big) < 0 \ .$$

Basically, that means that if we have found an optimal hedge, implemented it, and then re-optimize, we may find an even better hedge. The reason is technical: in our setup, trading cost and therefore liqudity and limits are set up around a

[9] $\hat{H}(\alpha\boldsymbol{\delta} + \alpha'\boldsymbol{\delta}') + \varrho\big(\alpha(X + \boldsymbol{\delta}\mathbf{H}) + \alpha'(X + \boldsymbol{\delta}'\mathbf{H})\big) \leq \hat{H}(\boldsymbol{\delta}) + \varrho(X + \boldsymbol{\delta}\mathbf{H})$

[10] Classic literature on convex risk measures refers to the set $\{X : \rho(X) \leq 0\}$ as the *acceptance set* of ρ; c.f. Föllmer and Schied [12]; we go one step further and provide a monetary gain for any X such that $\rho(X) < 0$.

"zero" hedge. For example, assume that we are only allowed to trade one unit of H^1. However, by running our optimization twice, we are now able to purchase twice H^1.

To alleviate this, the second optimization needs to base the new hedge around the first one, in which case re-optimization will not yield a better result.

Formally, that means if $\boldsymbol{\delta}_0$ is the result of the first optimization, then the second optimization uses $c'(\boldsymbol{\delta}') := c(\boldsymbol{\delta}' + \boldsymbol{\delta}_0)$ for $\boldsymbol{\delta}' \in \mathcal{L}' := \{\boldsymbol{\delta}' : \boldsymbol{\delta}' + \boldsymbol{\delta}_0 \in \mathcal{L}\}$ accordingly.

4.1 Marginal Pricing

Let us assume we have an existing portfolio Z of derivatives, and we are asked to sell a derivative P to a client. The marginal Cost of Risk is given by

$$m_0(P) := \pi(Z - P) - \pi(Z) \ . \tag{23}$$

It is the minimum price we should charge for selling P in the presence of an existing long position Z.

Representation (23) uses via (20) absolute price levels of both the position Z and the hedges. This is mathematically correct in the sense that it assumes that Z encompasses the entire current portfolio including current cash holdings.

Assume now that we wish to operate rather in return space, i.e. $\mathrm{d}Z$ and $\mathrm{d}\mathbf{H}$. This is a much more natural situation in a financial institution since capital requirements and the like are necessarily held properly only at higher level where the overall non-linear calculation can be performed. If we are anyway rather interested in marginal pricing, then such a relative representation is rather useful, as we will see shortly.

Notation 1 *We call*

$$\mathcal{R}(Z) := \pi(\mathrm{d}Z)$$

the Relative Cost of Risk *for Z. It has the following intuitive representation:*

$$\mathcal{R}(Z) = \inf_{\boldsymbol{\delta}\in\mathcal{L},\, W\in\mathcal{W}} \Big\{ \underbrace{c(\boldsymbol{\delta}) - \operatorname{ess\,inf} W_\tau}_{\substack{\text{relative}\\ \text{implementation}\\ \text{cost}}} + \underbrace{\mu\big(\mathrm{d}\Pi\big)}_{\substack{\text{risk}\\ \text{penalty}}} \ \Big| \ \underbrace{\nu\big(\mathrm{d}\Pi\big) \le 1}_{\substack{\text{PnL risk}\\ \text{limit}}} \quad \mathrm{d}\Pi := \mathrm{d}Z + \boldsymbol{\delta}\mathrm{d}\mathbf{H} - W \Big\} \ . \tag{24}$$

As before, if μ and ν are monotone, then

$$\mathcal{R}(Z) = \inf_{\boldsymbol{\delta}\in\mathcal{L},\, w\in\mathbb{R}} \Big\{ c(\boldsymbol{\delta}) + w + \mu\big(\mathrm{d}\Pi\big) \ \Big| \ \nu\big(\mathrm{d}\Pi\big) \le 1 \quad \mathrm{d}\Pi := \mathrm{d}Z + \boldsymbol{\delta}\mathrm{d}\mathbf{H} + w \Big\} \ . \tag{25}$$

If μ and ν are convex risk measures, then

$$\mathcal{R}(Z) = \inf_{\boldsymbol{\delta}\in\mathcal{L}} \Big\{ c(\boldsymbol{\delta}) + \mu\big(\mathrm{d}\Pi\big) \ \Big| \ \nu\big(\mathrm{d}\Pi\big) \le 1; \ \mathrm{d}\Pi := \mathrm{d}Z + \boldsymbol{\delta}\mathrm{d}\mathbf{H} \Big\} \ . \tag{26}$$

Proposition 4.2 *We have the trivial relationship $\mathcal{R}(Z) - Z_0 = \pi(Z)$, hence the marginal minimum price for selling a position P is given in terms of Relative Costs of Risk as*

$$m_0(P) := \underbrace{P_0}_{\substack{\textit{model}\\ \textit{price}}} + \underbrace{\mathcal{R}(Z-P) - \mathcal{R}(Z)}_{\substack{\textit{risk}\\ \textit{adjustment}}} \ .$$

Most importantly, we do not need to know the total value of Z_0, only its return distribution $\mathrm{d}Z$.

4.2 Relevance under Hedging: Absence of Arbitrage

The introduction of hedging instruments causes the same concern as the construction of convex risk measures in the first place: how to we make sure that the cost and return structure of hedging instruments does not create real or statistical arbitrage?

With "real" arbitrage we refer to cases such as $H_0^1 = 0$ and $H_\tau^1 \geq 0$ with a non-zero $\mathbb{Q}$-probability that H_τ^1 is positive. Note that this does not imply that such opportunity is exploited by program (20): if we use the worst-case measure inf, then only sure, not just likely positive returns will be exploited. Moreoever, any other measure will still balance cost vs. opportunity; we say $\boldsymbol{\delta} \in \mathcal{L}$ is a *static arbitrage opportunity* if

$$\boldsymbol{\delta}\mathbf{H}_\tau - \hat{H}(\boldsymbol{\delta}) \geq 0 \quad \text{and} \quad \mathbb{Q}\big[\ \boldsymbol{\delta}\mathbf{H}_\tau - \hat{H}(\boldsymbol{\delta}) > 0\ \big] > 0\ . \tag{27}$$

It is straight forward to avoid static arbitrage opportunities:

Proposition 4.3 *If $H_0^i > \operatorname{ess\,inf} H_\tau^i$ for all i; then there is exist no static arbitrage opportunity.*

Statistical Arbitrage

The more interesting case, however, is that of a *statistical arbitrage opportunity* which depends on the measure ϱ. It is defined as a $\boldsymbol{\delta} \in \mathcal{L}$ with the property

$$\hat{H}(\boldsymbol{\delta}) + \varrho(\boldsymbol{\delta}\mathbf{H}_\tau) < 0\ . \tag{28}$$

For example, set $\hat{H}(\boldsymbol{\delta}) = \boldsymbol{\delta}H_0$ and $\varrho(X) := -\mathbb{E}_\mathbb{P}[X]$ for a measure $\mathbb{P} \in \mathcal{P}$ such that $\mathbb{E}_\mathbb{P}[H_\tau^1] > H_0^1$. Then $\boldsymbol{\delta} = (1, 0, \dots, 0)$ is such an arbitrage direction.

In fact, given hedges $\mathbf{H}$, we can *always* find a kernel which creates statistical arbitrage as long as the returns of the hedging instruments are not pathological: assume that H_τ^1 is such that the set $A = \{\delta^1 H_\tau^1 > \hat{H}(\delta^1)\}$ has non-zero $\mathbb{Q}$-probability for some $\delta^1 \neq 0$. Define the measure $\mathbb{P}[B] := \mathbb{Q}[A \cap B]/\mathbb{Q}[A]$ which is absolutely continuous with respect to $\mathbb{Q}$. Then $\varrho(X) := -\mathbb{E}_\mathbb{P}[X]$ has the statistical arbitrage opportunity $(\delta^1, 0, \dots, 0)$.

Indeed, if the aim of the practitioner is using statistical tools to detect "alpha", then we would indeed expect to have statistical arbitrage opportunities, presumably up to a size at which point transaction cost would make an investment unwise given our risk measure ϱ. From this point of view, the only real restriction should be that such an arbitrage opportunity cannot be infinite. We recall that $\varrho(Z) \geq \varrho(0) - \operatorname{ess\,sup} Z$ i.e. statistical arbitrage might well be acceptable as long as π is *relevant* as defined before.

Either way, we want to conclude this section with a simple method of ensuring that $\mathbf{H}$ admits no statistical arbitrage.

PROPOSITION 4.4 *Let ϱ be a normalized law-invariant convex risk measure w.r.t. $\mathbb{P}$, and assume that*

$$H_0^i \geq \mathbb{E}_{\mathbb{P}}[H_\tau^i]$$

for $i = 1, \ldots, n$.

Then, π is normalized and therefore does not admit statistical arbitrage.

Proof – Recall that we had assumed that the cost function $c(\boldsymbol{\delta}) := \hat{H}(\boldsymbol{\delta}) - \boldsymbol{\delta}\mathbf{H}_0$ is normalized and non-negative. Theorem 2.1 yields $\varrho(X) \geq -\mathbb{E}_{\mathbb{P}}[X]$. Recall from (21) that $\pi(Z) = \inf_{\boldsymbol{\delta}}\{\hat{H}(\boldsymbol{\delta}) + \varrho(\boldsymbol{\delta}\mathbf{H})\}$. Hence

$$\hat{H}(\boldsymbol{\delta}) + \varrho\big(\boldsymbol{\delta}\mathbf{H}\big) \geq \hat{H}(\boldsymbol{\delta}) - \mathbb{E}_{\mathbb{P}}[\boldsymbol{\delta}\mathbf{H}_\tau] \geq \hat{H}(\boldsymbol{\delta}) - \boldsymbol{\delta}\mathbf{H}_0 = c(\boldsymbol{\delta}) \geq 0 \ ,$$

which proves $\pi(0) = 0$ as claimed. □

5 What Happened to my Greeks?

Our framework goes against much of the common practise of using "greeks" to manage risk: the classic approach in derivative finance for hedging derivatives is to use the valuation models to compute "greeks" with respect to relevant market parameters. In our approach, though, we have not used any greeks. How did that happen?

Assume that relevant market parameters which are inputs into our models are given by a vector $\mathbf{x}_0 = (x_0^1, \ldots, x_0^m)'$. This vector contains, for example, spot, volatilities, reference yield curve rates, etc. The value of our portfolio and our hedges are therefore understood to be functions $Z_0 \equiv Z_0(\mathbf{x}_0)$ and $H_0^j \equiv H_0^j(\mathbf{x}_0)$, respectively.

We denote by $\Delta_i := \partial_{x_i}$ the derivative with respect to the ith market parameter and set $\boldsymbol{\Delta} := (\Delta_1 \ \ldots, \Delta_m)$. We stress again that our $\boldsymbol{\Delta}$ contains not just "Deltas", but all relevant first order risks such as Vega, Rho etc.[11]

[11] For the avoidance of doubt: $\boldsymbol{\Delta}$ includes greeks with respects to parameters which are not stochastic in the underlying model, e.g. Vega risk for Black & Scholes pricers.

Greek Hedging

The standard hedging methodology in derivative finance is based on the assumption of a complete market. In such a situation, a position with zero first order greek sensitivity has no "risk" and perfectly replicates the payoff. In our notation, that means computing the matrix $\mathbf{\Delta H}_0$ and the row vector $\mathbf{\Delta} Z_0$, and then finding a $\boldsymbol{\delta}$ which "flattens the greeks", for example

$$\|\mathbf{\Delta} Z_0 + \boldsymbol{\delta} \cdot \mathbf{\Delta H}_0\|_2 \ . \tag{29}$$

If we wanted to minimize our greek exposure of the overall position irrespective of cost, we may use the Pseudo-Inverse to compute the projection

$$\boldsymbol{\delta}_G := -\left(\mathbf{\Delta H}_0 \mathbf{\Delta H}_0'\right)^{-1} \left(\mathbf{\Delta} Z_0 \mathbf{\Delta H}_0'\right) \ . \tag{30}$$

In practise, a range of heuristics are used to implement risk management based on this approach while taking into account transaction cost; see our own discussion in section 5.2 of [5].

Unfortunately, there is no obvious methodologically sound approach to balance between cost and residual risk since a given greek exposure has no dimensionality which can be compared to the cost of entering into it. Formally, this is addressed by using "Cash Greeks" $\mathbf{\Delta}^{\$} := \mathbf{x} \cdot \mathbf{\Delta}$.[12] We may therefore solve the program

$$\inf_{\boldsymbol{\delta}} \left\{ \hat{H}(\boldsymbol{\delta}) + \|\mathbf{\Delta} Z_0 + \boldsymbol{\delta} \cdot \mathbf{\Delta}^{\$} \mathbf{H}_0\|_2 \right\} \ .$$

In this approach, each parameter is treated according to their "cash greek" value. The actual statistical sensitivity of the driving risk factor plays no role.

Parameter Hedging

The immediate next step to improve upon this approach is to recognize that the return vector $\mathrm{d}\mathbf{x}_0$ has uncertainty, which should drive which greek to hedge and how much: if the Delta exposures for two risk factors are nominally the same, and hedging them cost the same, then we should hedge more the one whose risk factor has higher volatility. Let us therefore define the statistical covariance matrix Σ of $\mathrm{d}\mathbf{x}_0$. Methodologically, we use Taylor to approximate any $F \in \{Z, H^1, \ldots, H^n\}$ with its second-order representation

$$dF_0 = \mathbf{\Delta} F_0 \, \mathrm{d}\mathbf{x}_0 + \mu F \, \mathrm{d}t \ , \quad \mu F := \frac{1}{2} \mathbf{\Gamma} F_0 \, \Sigma \, \frac{\mathrm{d}x^2}{\mathrm{d}t} + \Theta F_0 \ . \tag{31}$$

Here, $\mathbf{\Gamma}$ denotes the second order derivative operator in $\mathbf{x}$, and Θ denotes the derivative operator in time. The drift μ is called the "Gamma-Theta carry" of the position F.

Given (31), it is very tempting to approximate $\mathrm{d}\mathbf{x}_0$ itself as a normal with mean $\mathbb{E}[\mathrm{d}\mathbf{x}_0]$ and variance Σ, which means both $\mathrm{d}Z$ and $\mathrm{d}\mathbf{H}$ are via (31) approximated as normals. If only quadratic risk is minimized, then a quadratically

[12] The vector is given by $\mathbf{\Delta}^{\$}{}_i F(\mathbf{x}) := \partial_\varepsilon F(\mathbf{x}(1 + \varepsilon \mathbf{e}_i))|_{\varepsilon=0}$.

optimal hedge is given by the projection

$$\boldsymbol{\delta}_P := -\left(\boldsymbol{\Delta}\mathbf{H}_0\,\Sigma\,\boldsymbol{\Delta}\mathbf{H}_0'\right)^{-1}\left(\boldsymbol{\Delta}Z_0\,\Sigma\,\boldsymbol{\Delta}\mathbf{H}_0'\right)\ . \tag{32}$$

In order to introduce cost, we may apply the classic Markoviz mean-volatility/variance framework and define the program

$$\mathcal{R}(Z) \equiv \inf_{\boldsymbol{\delta}}\left\{ \underbrace{\hat{H}(\boldsymbol{\delta})}_{\substack{\text{implementation}\\ \text{cost}}} - \underbrace{\mathbb{E}[\mathrm{d}Z + \boldsymbol{\delta}\mathrm{d}\mathbf{H}]}_{\substack{\text{Gamma-Theta}\\ \text{carry and drift}}} + \underbrace{\lambda\|(\mathrm{d}Z + \boldsymbol{\delta}\mathrm{d}\mathbf{H}) - \mathbb{E}[\mathrm{d}Z + \boldsymbol{\delta}\mathrm{d}\mathbf{H}]\|_2^p}_{\text{Risk}} \right\} \tag{33}$$

for $p \in \{0, 2\}$. Note that due to the assumed normality of the returns of $\mathrm{d}Z$ and $\mathrm{d}\mathbf{H}$, the objective in this problem is monotone and convex, and obviously also cash-invariant. The expectation term naturally introduces via (31) the Gamma-Theta carry μ of the position into the optimization.

This framework is attractive and is in effect widely used in sensitivity-based VaR calculations. It works well for benign situations where second order effects and discontinuities of the derivative portfolio are comparatively minor. However, the implicit assumption that all derivatives can be modelled over $[0, \tau]$ as normals breaks down in situations with high gamma values, for examples when working with barriers products or short-term options. See Alexander, Coleman and Li's [1] for an early discussion on these issues (they refer to (31) as "delta-gamma approximation").

Statistical Hedging

The next logical step towards hedging a portfolio of derivatives is to use, as we did in the main body of this note, the actual returns of Z and $\mathbf{H}$. A simplistic approach of using Markoviz-style mean-variance optimization was suggested in [6]. We remark that in the absence of cost, and if the drift is not taken into account, and if we only wish to minimize quadratic risk, then an optimal delta is given by

$$\boldsymbol{\delta}_Q := -\left(\,\mathbb{E}[\mathrm{d}\mathbf{H}\mathrm{d}\mathbf{H}']\,\right)^{-1}\left(\,\mathbb{E}[\mathrm{d}Z\mathrm{d}\mathbf{H}']\,\right)\ . \tag{34}$$

However, we reiterate the points made in the previous sections that the associated

$$\mathcal{R}(Z) \equiv \inf_{\boldsymbol{\delta}}\left\{ \underbrace{\hat{H}(\boldsymbol{\delta})}_{\text{cost}} - \underbrace{\mathbb{E}[\mathrm{d}Z + \boldsymbol{\delta}\mathrm{d}\mathbf{H}]}_{\substack{\text{Gamma-Theta}\\ \text{carry and drift}}} + \underbrace{\lambda\|(\mathrm{d}Z + \boldsymbol{\delta}\mathrm{d}\mathbf{H}) - \mathbb{E}[\mathrm{d}Z + \boldsymbol{\delta}\mathrm{d}\mathbf{H}]\|_2^p}_{\text{Risk}} \right\} \tag{35}$$

for $p \in \{1, 2\}$ is not a well-posed problem since its objective is not monotone which may result into hedging positions which appear numerically optimal, but are factually strictly worse than our original position. This is rectified by using the approach discussed above, i.e. to use

$$\inf_{\boldsymbol{\delta},W^+\geq 0}\Bigg\{\ \underbrace{\hat{H}(\boldsymbol{\delta})-\operatorname{ess\,inf} W^+_\tau}_{\text{cost and write-off}}-\underbrace{\mathbb{E}[\mathrm{d}Z+\boldsymbol{\delta}\mathrm{d}\mathbf{H}-W^+_\tau]}_{\substack{\text{Gamma-Theta}\\ \text{carry and drift}}} \tag{36}$$
$$+\underbrace{\lambda\|(\mathrm{d}Z+\boldsymbol{\delta}\mathrm{d}\mathbf{H}-W^+_\tau)-\mathbb{E}[\mathrm{d}Z+\boldsymbol{\delta}\mathrm{d}\mathbf{H}-W^+_\tau]\|_2^p}_{\text{Risk}}\ \Bigg\}$$

in order to find an optimal hedge $\boldsymbol{\delta}^*$ and write-off W^*.

6 Conclusion

We have motivated an alternative approach to mean-variance optimization which is monotone and convex – properties which are natural to any such approach and become crucial when considering portfolios of derivatives. We have also seen how the introduction of being able to write off parts of our portfolios leads naturally to convex risk measures.

Our favourite measure is the "quadratic CVaR" measure

$$\rho(X) := \inf_w \{w + \lambda\|\min\{0, X+w\}\|_2\}$$

for $\lambda \geq 1$.

In our data-driven approach we cease having to use greeks for the calculation of hedging position, and rely entierly on statistical properties of the return distribution.

Our work on "Deep Hedging" [8] extends the material here to multiple steps.

7 Disclaimer

References

[1] *Alexander, Coleman, Li:*
"Minimizing CVaR and VaR for a portfolio of derivatives", Journal of Banking & Finance, 2006
`https://cs.uwaterloo.ca/~yuying/papers/cvarJBF.pdf`

[2] *Artzner, Delbaen, Eber, Heath:*
"Coherent Measures of Risk", Mathematical Finance, Vol. 9, No. 3 (July 1999), 203;228
`http://personal.fmipa.itb.ac.id/khreshna/files/2011/02/artzner1999.pdf`

[3] *Ben-Tal, Teboulle:*
"An old-new concept of convex risk measures: The optimized certainty equivalent", Mathematical Finance, Issue 3, 17 (2007), 449-476
`http://emotion.technion.ac.il/Labs/Opt/opt/Pap/oce-risk.pdf`

[4] *Ben-Tal, Teboulle:*
"Expected utility, penalty functions and duality in stochastic nonlinear programming", Management Science, Vol. 32, No. 11 (Nov., 1986), pp. 1445-1466

[5] *Buehler:*
Volatility Markets: Consistent Modelling, Hedging and Practical Implementation, PhD dissertation, 2006, published in print VMD Verlag 2009
`http://ssrn.com/abstract=1118245`

[6] *Buehler:*
"Statistical Hedging: Application to Stochastic Local Volatility Models", presentation, Global Derivatives Trading & Risk Management Conference, Amsterdam, May 2013

[7] *Buehler:*
"Statistical Hedging - Cost, Carry, Risk", presentation, Global Derivatives Trading & Risk Management Conference, Amsterdam, May 2014

[8] *Buehler, Gonon, Teichmann, Wood:*
"Deep Hedging", Quantitative Finance, vol 0, num 0, 2019, pages 1-21
`https://www.ssrn.com/abstract=3120710`

[9] *El Karoui, Ravanelli:*
"Cash Sub-additive Risk Measures and Interest Rate Ambiguity", Mathematical Finance, Volume 19, Issue 4 October 2009, Pages 561590

[10] *Filipovic, Kupper:*
"Monotone and Cash-Invariant Convex Functions and Hulls", Insurance Mathematics Economics, 41(1): 16, 2007
`http://www.math.lmu.de/~filipo/PAPERS/constconvfinal.pdf`

[11] *Föllmer, Knispel*:
"Convex Risk Measures: Basic Facts, Law-invariance and beyond, Asymptotics for Large Portfolios ", Advances in Mathematical Economics In: Handbook of the Fundamentals of Financial Decision Making:In 2 Parts (World Scientific Handbook in Financial Economics Series), 10 May 2013
`https://www.math.hu-berlin.de/~foellmer/papers/RiskMeasures_FoellmerKnispel.pdf`

[12] *Föllmer, Schied*:
Stochastic Finance: An Introduction in Discrete Time, Walter de Gruyter, 2002

[13] *Markowitz*:
"Portfolio Selection: Efficient Diversification of Investments", New York: John Wiley & Sons. (reprinted by Yale University Press, 1970, ISBN 978-0-300-01372-6; 2nd ed. Basil Blackwell, 1991, ISBN 978-1-55786-108-5

[14] *Meucci*:
Risk and Asset Allocation, Springer, 2005

[15] *Rockafellar, Uryasev:*
"Optimization of conditional value-at-risk", Journal of Risk, 2, 21-41, 2000
`http://www.ise.ufl.edu/uryasev/files/2011/11/CVaR1_JOR.pdf`

[16] *Toussaint, Sircar:*
"A Framework for Dynamic Hedging under Convex Risk Measures", WP, November 2008; revised August 2009
`https://www.princeton.edu/~sircar/Public/ARTICLES/ascona_proc.pdf`