

SENSITIVITY ANALYSIS IN APPLICATIONS WITH DEVIATION, RISK, REGRET, AND ERROR MEASURES*

BOGDAN GRECHUK[†] AND MICHAEL ZABARANKIN[‡]

Abstract. The envelope theorem is derived for optimization problems with positively homogeneous convex functionals defined on a space of random variables. Those problems include linear regression with general error measures and optimal portfolio selection with the objective function being either a general deviation measure or a coherent risk measure subject to a constraint on the expected rate of return. The obtained results are believed to be novel even for Markowitz’s mean-variance portfolio selection, but are far more general and include explicit envelope relationships for the rates of return of portfolios that minimize lower semivariance, mean absolute deviation, deviation measures of \mathcal{L}^p -type and semi- \mathcal{L}^p type, and conditional value-at-risk. In each case, the envelope theorem yields explicit estimates for the absolute value of the difference between deviation/risk of optimal portfolios with the initial and perturbed asset probability distributions in terms of a norm of the perturbation.

Key words. Deviation measure, risk measure, error measure, portfolio optimization, sensitivity analysis, envelope theorem.

AMS subject classifications. 90C31

1. Introduction. Financial markets, being inherently uncertain, require elaborate stochastic models for solving a variety of problems ranging from predicting asset prices, setting interest rates, and selecting appropriate portfolios to making any kind of short-term and long-term decisions. In general, any quantity of interest, e.g., asset price and interest rate, can be modeled as a random variable (r.v.) X . However, uncertainty ingrained in X can be assessed from at least *three* different perspectives, which depend on particular application and decision maker preferences:

- *How random or how non-constant* X is. This aspect is central in estimating the accuracy of an output of a stochastic model. For example, given two weather forecasts “the temperature tomorrow morning will be between 10°C and 30°C ” and “it will be between 19°C and 21°C ,” the latter provides a much shaper estimate—*deviation* in the latter is much smaller than in the former.
- *How non-zero* X is. This aspect is critical when X models the random *error* of an output from a stochastic model. A typical task in stochastic modeling is choosing model parameters that minimize the error, which is an r.v.—one needs a way to measure r.v.’s “nonzeroness” to formulate an appropriate optimization problem.
- *How risky* X is. This aspect arises when X models, for example, an uncertain outcome from a financial investment, or from a business project. In this case, $X > 0$ and $X < 0$ correspond to profit and loss, respectively, and the problem is to assess *hazard* associated with X . Hazard can be split into two types: overall hazard in X to be called *risk* and hazard relative to some benchmark to be called *regret*.

Non-constancy, non-zeroness, overall hazard and relative hazard in X are measured by *deviation measure* $\mathcal{D}(X)$, *error measure* $\mathcal{E}(X)$, *risk measure* $\mathcal{R}(X)$ and *regret measure* $\mathcal{V}(X)$, respectively. Four measures \mathcal{D} , \mathcal{E} , \mathcal{R} and \mathcal{V} can all be defined as *positively homogenous convex functionals* in a unified axiomatic framework—such four measures related by

$$(1) \quad \mathcal{D}(X) = \inf_{C \in \mathbb{R}} \mathcal{E}(X - C), \quad \mathcal{R}(X) = \mathcal{D}(X) - \mathbb{E}[X], \quad \mathcal{V}(X) = \mathcal{E}(X) - \mathbb{E}[X],$$

*Submitted to the editors April 27, 2017.

[†]Department of Mathematics, University of Leicester, LE1 7RH, UK; e-mail: bg83@leicester.ac.uk

[‡]Department of Mathematical Sciences, Stevens Institute of Technology, Hoboken, NJ 07030, USA; e-mail: mzabaran@stevens.edu

form *fundamental quadrangle* [44], where $\mathbb{E}[\cdot]$ denotes the expected value. In (1), \mathcal{D} is called deviation *projected* from error measure \mathcal{E} , $\mathcal{S}(X) = \operatorname{argmin}_{C \in \mathbb{R}} \mathcal{E}(X - C)$, i.e., the set of C for which the minimum in (1) is attained and finite, is called the *statistic* of X associated with \mathcal{E} [50], \mathcal{R} is known under different names: *averse measure of risk*¹ [50], *expected gain-confidence* [3], *safety measure* [35] and *safety-risk measure* [40], and $U(X) = -\mathcal{V}(X)$ is argued to measure *utility* of X [44]. For a detailed discussion of measures \mathcal{D} , \mathcal{E} , \mathcal{R} and \mathcal{V} and statistic \mathcal{S} , see [18, 22, 44, 45, 46, 47, 48, 49, 50].

Virtually any problem in finance deals with at least one of the four: error, deviation, risk, and regret/utility, which can arise in objective functions and constraints. For example, the bi-criteria risk-reward approach chooses a portfolio with the desired level of reward and the minimum level of risk, where risk can be measured by either a risk measure or a deviation measure. It is originated from the classical mean-variance portfolio theory [36, 37] and since then has been extensively used in a vast number of works on portfolio selection, see, e.g. [22, 34, 43, 47, 48, 49]. Alternatively, an investor may minimize risk or regret of a portfolio without a constraint on the expected return. In this case, he/she may measure portfolio risk by coherent risk measures [1, 2, 11, 12, 15, 23, 57]. Optimization of an error measure arises, for example, in linear regression, where a random target is approximated by a linear combination of given r.v.'s (explanatory variables), and regression coefficients are found by minimizing the error measure of the difference between the target and its approximation [50].

In the above discussion, returns of assets in a portfolio problem and the target and the explanatory variables in a generalized linear regression are all assumed to be known r.v.'s. However, this is rarely, if ever, the case in practice. Distributions of those r.v.'s are estimated from data and as such they will change with the arrival of any new data and/or with the choice of different time intervals and a different time window. How are sensitive then optimal values of measures \mathcal{D} , \mathcal{E} , \mathcal{R} and \mathcal{V} in corresponding optimization problems to disturbances in underlying random parameters? Estimation errors in means and variance-covariance matrix of asset returns was already acknowledged by Markowitz [37] and addressed further in subsequent works of Kalymon [30] and Frankfurter et al. [16]. Remarkably, Kallberg & Ziemba [29] showed that the optimal solution of the mean-variance portfolio optimization was considerably more sensitive to errors in mean returns than to errors in the variance-covariance matrix. Also, Best & Grauer [5] showed that a quite small change in the mean return of just one asset can “drive half the securities from the portfolio.” However, they observed that the expected value and variance of the optimal portfolio return were much more stable. Guigues [25] obtained upper bounds for a possible variation of the variance of the optimal portfolio return in terms of variations in the mean returns and variance-covariance matrix. Grechuk & Zabarankin [24] analyzed sensitivity of the mean-variance optimal portfolio return to changes in the weights assigned to historical data.

Much less is known about sensitivity of optimal values of general deviation measures to disturbances in assets' returns in portfolio selection. Assuming assets' rates of return to be normally distributed, Simaan [53] showed that compared to the classical mean-variance model, the mean absolute deviation model is more sensitive to estimation errors in the normal distribution parameters. Under the same assumption of normally distributed rates of return and imposing no constraint on portfolio expected return, Kondor et al. [33] concluded that optimal portfolios that minimize standard deviation, mean absolute deviation, CVaR deviation, and maximum loss deviation are quite sensitive to the noise in the normal distribution parameters. Kandasamy [31] analyzed sensitivity of optimal portfolios that minimize a risk/deviation measure, e.g. CVaR and semivariance, through numerical experiments only.

¹Given a random (nonconstant) payoff X with expected value $\mathbb{E}[X]$, risk aversion is the preference $\mathbb{E}[X] > X$. This yields $\mathcal{R}(\mathbb{E}[X]) < \mathcal{R}(X)$, which with (1) and $\mathcal{R}(\mathbb{E}[X]) = \mathbb{E}[-\mathbb{E}[X]] + \mathcal{D}(\mathbb{E}[X]) \equiv \mathbb{E}[-X]$ simplifies to $\mathcal{D}(X) > 0$.

In the mathematical literature, the study of how the optimal value in an optimization problem changes for perturbed parameters is known as *sensitivity analysis*, *parametric programming*, and *perturbation analysis*, and goes back at least to Hoffman & Jacobs [27], Orchard-Hays [39], and Williams [55], who studied this question for linear programs. One of the standard approaches, which is attributed arguably to Danskin [10], is to derive upper and lower estimates for the directional derivatives of an optimal value function with respect to parameters. For other approaches, e.g. second-order estimates, and for a comprehensive up-to-date survey, see Bonnans & Shapiro [7]. In the economics and finance literature, a formula for the derivative of the optimal value of the objective function in an optimization problem with respect to problem parameters is given by *envelope theorem*. For an expected utility, the corresponding formula is known as *Roy's identity* [54, Ex. 4, p. 123], while in the context of production economy, a related result is known as *Hotelling's lemma* [54, Ex. 4, p. 108]. Milgrom & Segal [38] derived the envelope theorem for arbitrary choice sets.

The contribution and organization of this work are as follows. Section 2 reviews the axiomatic framework for measures of deviation, error, risk and regret related by (1). Section 3 derives the envelope theorem for an optimization problem whose objective and constraints involve only positively homogeneous convex functionals, and also extends sensitivity analysis to the case when perturbations are unknown—robust sensitivity analysis. It presents explicit estimates for the absolute value of the difference of the optimal values of the initial and perturbed optimization problems in terms of a norm of the perturbation. Section 4 specializes the theorems to generalized linear regression and to mean-deviation and mean-risk portfolio problems. Section 5 concludes the work. The appendix presents some technical results.

2. Positively homogeneous convex functionals.

2.1. Error, deviation, risk, and regret. Let $(\Omega, \mathcal{M}, \mathbb{P})$ be a probability space, where Ω is the set of future states ω , \mathcal{M} is a field of subsets of Ω , and \mathbb{P} is a probability measure on (Ω, \mathcal{M}) . Here, a random variable (r.v.) is an element of $\mathcal{L}^2(\Omega) = \mathcal{L}^2(\Omega, \mathcal{M}, \mathbb{P})$. Let $F_X(x)$ and $q_X(\alpha) = \inf\{x | F_X(x) > \alpha\}$ be the cumulative distribution function (CDF) and quantile function of an r.v. X , respectively.

DEFINITION 2.1 (basic functionals). *A basic functional is any functional $\mathcal{F} : \mathcal{L}^2(\Omega) \rightarrow (-\infty, +\infty]$ satisfying (see [21])*

- (F1) $\mathcal{F}(\lambda X) = \lambda \mathcal{F}(X)$ for all X and all $\lambda \geq 0$ (here $0\infty = 0$) (positive homogeneity),
- (F2) $\mathcal{F}(X + Y) \leq \mathcal{F}(X) + \mathcal{F}(Y)$ for all X and Y (subadditivity),
- (F3) $\{X \in \mathcal{L}^2(\Omega) | \mathcal{F}(X) \leq c\}$ is closed in $\mathcal{L}^2(\Omega)$ for all $c < \infty$ (lower semicontinuity).

Measures of error, deviation, risk, and regret, which form a fundamental quadrangle [44], are subclasses of basic functionals. Namely, a basic functional is

- An *error measure* $\mathcal{E} : \mathcal{L}^2(\Omega) \rightarrow [0, +\infty]$ if it satisfies (E4) $\mathcal{E}(X) > 0$ for nonzero X with $\mathcal{E}(C) < \infty$ for constant C [50].
- A *deviation measure* $\mathcal{D} : \mathcal{L}^2(\Omega) \rightarrow [0, +\infty]$ if it satisfies (D4) $\mathcal{D}(X) = 0$ for constant X , and $\mathcal{D}(X) > 0$ otherwise (*nonnegativity*) [45, 46]. F1, F2, and D4 imply that $\mathcal{D}(X + C) = \mathcal{D}(X)$ for all constants C (*constant translation invariance*) [46].
- A *risk measure* $\mathcal{R} : \mathcal{L}^2(\Omega) \rightarrow (-\infty, +\infty]$ if it satisfies (R4) $\mathcal{R}(C) = -C$ for constants C . With F2, R4 is equivalent to $\mathcal{R}(X + C) = \mathcal{R}(X) - C$ for all constants C [46]. \mathcal{R} is called *coherent* [1], if satisfies F1–F3, R4, and (R5): $\mathcal{R}(X) \leq 0$ for all $X \geq 0$. With F2, R5 is equivalent to $\mathcal{R}(X) \leq \mathcal{R}(Y)$ when $X \geq Y$. \mathcal{R} is called *averse* [46], if it satisfies F1–F3, R4, and (R6): $\mathcal{R}(X) > \mathbb{E}[-X]$ for all non-constant X .
- A *regret measure* $\mathcal{V} : \mathcal{L}^2(\Omega) \rightarrow (-\infty, +\infty]$ if it satisfies

(V4) $\mathcal{V}(X) > \mathbb{E}[-X]$ for all non-zero X (see [44]).

V4 is similar to R6, but it holds for every *non-zero* X , not just for *non-constant* one.

A basic functional \mathcal{F} is *law invariant* if $\mathcal{F}(X) = \mathcal{F}(Y)$ whenever X and Y have the same distribution. Every law invariant \mathcal{R} , which can be defined on an *atomless* probability space, i.e. for which there exists an r.v. with a continuous CDF, except only $\mathcal{R}(X) = -\mathbb{E}[X]$, is averse [20, Proposition 8].

Relationships (1) show that \mathcal{E} is a “generator” of a quadrangle. In fact, given an error measure \mathcal{E} , a deviation measure \mathcal{D} can also be obtained *from penalties relative to expectation*:

$$(2) \quad \mathcal{D}(X) = \mathcal{E}(X - \mathbb{E}[X]),$$

and the second and third relationships in (1) still hold with such \mathcal{D} , see [46]. For example, for the \mathcal{L}^p norm $\mathcal{E}(X) = \|X\|_p \equiv \mathbb{E}[|X|^p]^{1/p}$ with $p \geq 1$, (2) and (1) yield $\mathcal{D}(X) = \|X - \mathbb{E}[X]\|_p$, $\mathcal{R}(X) = \|X - \mathbb{E}[X]\|_p - \mathbb{E}[X]$ and $\mathcal{V}(X) = \|X\|_p - \mathbb{E}[X]$. In particular, for $p = 2$, $\mathcal{E}(X) = \|X\|_2$ is mean-square error and $\mathcal{D}(X) = \|X - \mathbb{E}[X]\|_2 = \sigma(X)$ standard deviation, whereas for $p = 1$, $\mathcal{E}(X) = \|X\|_1$ is mean absolute error and $\mathcal{D}(X) = \|X - \mathbb{E}[X]\|_1 = \text{MAD}(X)$ mean absolute deviation. The functionals $\mathcal{R}(X) = \mathbb{E}[-X] + \lambda \sigma(X)$ and $\mathcal{R}(X) = \mathbb{E}[-X] + \lambda \text{MAD}(X)$, popular in finance literature [37, 34], are averse measures of risk for any $\lambda > 0$, and the latter is also coherent if $\lambda \leq 1/2$. However, $\mathcal{E}(X) = \|X\|_p$ and $\mathcal{D}(X) = \|X - \mathbb{E}[X]\|_p$ treat ups and downs of X equally, which is not a desirable property in finance applications.

A non-symmetric generalization of $\mathcal{E}(X) = \|X\|_p$ is given by [50, (2.3)]

$$(3) \quad \mathcal{E}_{a,b,p}(X) = \|a[X]_+ + b[X]_-\|_p \quad \text{with } a > 0, b > 0, \text{ and } p \in [1, \infty],$$

where $[X]_{\pm} = \max\{0, \pm X\}$. For example, $\mathcal{E}_{1,1,p}(X) = \|X\|_p$, whereas $\mathcal{E}_{0,1,p}(X - \mathbb{E}[X])$ and $\mathcal{E}_{1,0,p}(X - \mathbb{E}[X])$ are deviation measures of semi- \mathcal{L}^p type, e.g. $\mathcal{E}_{0,1,2}(X - \mathbb{E}[X]) = \sigma_-(X)$ is the *standard lower semideviation*. Then $\mathcal{R}(X) = \mathbb{E}[-X] + \lambda \mathcal{E}_{a,b,p}(X - \mathbb{E}[X])$ with $\lambda > 0$ is an averse risk measure, which is coherent at least when $a = 0, b = 1$ and $\lambda = 1$ or when $a + b \leq 1, p = 1$ and $\lambda = 1$ [46, Theorem 3]. For any $\alpha \in (0, 1)$, $\mathcal{E}_{1,1/\alpha-1,1}$ is known as the Koenker-Bassett error measure used in quantile regression [32, 44]. In this case, \mathcal{R} and \mathcal{D} determined through (1) are conditional value-at-risk (CVaR) and CVaR deviation, respectively [50, Example 2.4]:

$$(4) \quad \text{CVaR}_{\alpha}(X) \equiv -\frac{1}{\alpha} \int_0^{\alpha} q_X(s) ds, \quad \text{CVaR}_{\alpha}^{\Delta}(X) \equiv \mathbb{E}[X] - \frac{1}{\alpha} \int_0^{\alpha} q_X(s) ds,$$

which can be generalized into mixed CVaR and mixed CVaR-deviation, respectively [50, Example 2.4]:

$$(5) \quad \mathcal{R}(X) = \int_0^1 \text{CVaR}_{\alpha}(X) d\lambda(\alpha), \quad \mathcal{D}(X) = \int_0^1 \text{CVaR}_{\alpha}^{\Delta}(X) d\lambda(\alpha)$$

with some $\lambda(\alpha) \geq 0$ such that $\int_0^1 d\lambda(\alpha) = 1$. Note that mixed CVaR is a coherent risk measure [46]. Mixed CVaR and mixed CVaR-deviation are used in inverse portfolio optimization [22, 23], where $\lambda(\alpha)$ is recovered from investor’s presumably optimal portfolios. For other examples of $\mathcal{E}, \mathcal{D}, \mathcal{R}$, and \mathcal{V} , see [44, 45, 46, 50].

2.2. Dual representation and differentiability. Every basic functional $\mathcal{F}(X)$ can be represented by [21, Proposition 3.1]

$$(6) \quad \mathcal{F}(X) = \sup_{Y \in \mathcal{Y}} \mathbb{E}[XY], \quad \mathcal{Y} \subset \mathcal{L}^2(\Omega),$$

where \mathcal{Y} will be called the *dual set* of \mathcal{F} and can be recovered by

$$(7) \quad \mathcal{Y} = \{ Y \in \mathcal{L}^2(\Omega) \mid \mathbb{E}[XY] \leq \mathcal{F}(X) \text{ for all } X \in \mathcal{L}^2(\Omega) \}.$$

Relations (6) and (7) establish a one-to-one correspondence between basic functionals $\mathcal{F} : \mathcal{L}^2(\Omega) \rightarrow (-\infty, +\infty]$ and nonempty closed convex sets in $\mathcal{L}^2(\Omega)$.

This work is confined to basic functionals, which are *continuous* everywhere on $\mathcal{L}^2(\Omega)$. A basic functional \mathcal{F} is continuous if and only if it is finite: $\mathcal{F}(X) < \infty$ for all $X \in \mathcal{L}^2(\Omega)$ [19, Proposition 3.12], and also it is continuous if and only if its dual set \mathcal{Y} is bounded [46, Theorem 1]². Since $\mathcal{L}^2(\Omega)$ is a reflexive Banach space, this implies that \mathcal{Y} is weakly compact by Kakutani's theorem [13, Theorem 3.31], which means that the supremum in (6) is always attained [28]. Consequently, for any $X \in \mathcal{L}^2(\Omega)$, the subdifferential $\partial\mathcal{F}(X) = \arg \max_{Y \in \mathcal{Y}} \mathbb{E}[XY]$ is nonempty closed convex set, see also [48] and [52, Corollary 3.1].

The representation (6) determines

- An *error measure*, \mathcal{E} , if and only if \mathcal{Y} is a nonempty closed convex set such that (i) $|\mathbb{E}[Y]| \leq c$ for some $c < \infty$ and every $Y \in \mathcal{Y}$ and (ii) $\mathbb{E}[XY] > 0$ for some $Y \in \mathcal{Y}$ and every nonzero $X \in \mathcal{L}^2(\Omega)$.³
- A *deviation measure*, \mathcal{D} , if and only if \mathcal{Y} is a non-empty closed convex set such that (i) $\mathbb{E}[Y] = 0$ for all $Y \in \mathcal{Y}$ and (ii) $\mathbb{E}[XY] > 0$ for some $Y \in \mathcal{Y}$ and every nonconstant $X \in \mathcal{L}^2(\Omega)$.
- A *risk measure*, \mathcal{R} , if and only if \mathcal{Y} is a nonempty closed convex set such that (i) $\mathbb{E}[Y] = -1$ for all $Y \in \mathcal{Y}$. \mathcal{R} is *averse* if \mathcal{Y} satisfies (i) and $\mathbb{E}[XY] > \mathbb{E}[-X]$ for some $Y \in \mathcal{Y}$ and every nonconstant $X \in \mathcal{L}^2(\Omega)$, and \mathcal{R} is *coherent* if \mathcal{Y} satisfies (i) and $Y \leq 0$ for all $Y \in \mathcal{Y}$.
- A *regret measure*, \mathcal{V} , if and only if \mathcal{Y} is a nonempty closed convex set such that $\mathbb{E}[XY] > \mathbb{E}[-X]$ for some $Y \in \mathcal{Y}$ and every nonzero $X \in \mathcal{L}^2(\Omega)$.

For \mathcal{D} and \mathcal{R} , $\mathcal{Q}_{\mathcal{D}} = 1 - \mathcal{Y}$ and $\mathcal{Q}_{\mathcal{R}} = -\mathcal{Y}$ are called *risk envelopes* of \mathcal{D} and \mathcal{R} , respectively, whereas $\mathcal{Q}_{\mathcal{D}}(X) = 1 - \partial\mathcal{D}(X)$ and $\mathcal{Q}_{\mathcal{R}}(X) = -\partial\mathcal{R}(X)$ are called *risk identifier sets* of X with respect to \mathcal{D} and \mathcal{R} , respectively, and $Q \in \mathcal{Q}_{\mathcal{D}}(X)$ and $Q \in \mathcal{Q}_{\mathcal{R}}(X)$ are called risk identifiers of X for \mathcal{D} and \mathcal{R} [1, 46, 48]. If \mathcal{D} and \mathcal{R} are related by (1), then $\mathcal{Q}_{\mathcal{D}} = \mathcal{Q}_{\mathcal{R}}$.

For σ , σ_- , MAD and $\text{CVaR}_{\alpha}^{\Delta}$ and for any non-constant X , risk identifier sets are given by [48, (9), (11), (16), (31)], which yield the subdifferentials $\partial\mathcal{D}(X) = 1 - \mathcal{Q}_{\mathcal{D}}(X)$:

$$(8a) \quad \sigma(X) = \|X - \mathbb{E}[X]\|_2, \quad \partial\sigma(X) = \{ \sigma(X)^{-1}(X - \mathbb{E}[X]) \},$$

$$(8b) \quad \sigma_-(X) = \|Z\|_2, \quad Z = [X - \mathbb{E}[X]]_-, \quad \partial\sigma_-(X) = \{ -\sigma_-(X)^{-1}(Z - \mathbb{E}[Z]) \},$$

$$(8c) \quad \text{MAD}(X) = \mathbb{E}[|X - \mathbb{E}[X]|], \quad \partial\text{MAD}(X) = \{ Z - \mathbb{E}[Z] \mid Z \in \text{sign}[X - \mathbb{E}[X]] \},$$

where $\text{sign}Y(\omega) = \pm 1$ when $Y(\omega) \gtrless 0$ and $\text{sign}Y(\omega) \in [-1, 1]$ when $Y(\omega) = 0$,

(9)

$$\mathcal{D}(X) = \text{CVaR}_{\alpha}^{\Delta}(X), \quad Y \in \partial\text{CVaR}_{\alpha}^{\Delta}(X) \iff \begin{cases} Y \in [1 - \alpha^{-1}, 1], & \mathbb{E}[Y] = 0, \\ Y(\omega) = 1 & \text{when } X(\omega) > q_X(\alpha), \\ Y(\omega) = 1 - \alpha^{-1} & \text{when } X(\omega) < q_X(\alpha). \end{cases}$$

For deviation measures $\mathcal{D}_p(X) = \|X - \mathbb{E}[X]\|_p$ and $\mathcal{D}_-(X) = \|[X - \mathbb{E}[X]]_-\|_p$ with $p \in (1, \infty)$ and for any non-constant X with $\mathcal{D}_p(X) < \infty$ and $\mathcal{D}_-(X) < \infty$, the corresponding risk identifier

²[19, Proposition 3.12] and [46, Theorem 1] are formulated for deviation measures, but property (D4) is not used in the proofs.

³(i) $\iff \mathcal{F}(C) < \infty$ for constants C , and (ii) $\iff \mathcal{E}(X) > 0$ for nonzero X .

sets $\mathcal{D}_p(X)$ and $\mathcal{D}_-(X)$ are singletons⁴ (see [14]), and so are the subdifferentials:

$$(10a) \quad \partial \mathcal{D}_p(X) = -\mathcal{D}_p(X)^{1-p} (\mathbb{E}[Z] - Z), \quad Z = |X - \mathbb{E}[X]|^{p-1} \text{sign}[X - \mathbb{E}[X]],$$

$$(10b) \quad \partial \mathcal{D}_-(X) = \mathcal{D}_-(X)^{1-p} (\mathbb{E}[Z] - Z), \quad Z = [X - \mathbb{E}[X]]_-^{p-1}.$$

For mixed CVaR-deviation in (5) and for a continuous r.v. X , the subdifferential is a singleton $\{Y\}$ such that $q_Y(\alpha) = 1 - \int_{(\alpha,1)} p^{-1} d\lambda(p)$, $\alpha \in (0, 1)$, and Y is comonotone⁵ with X . For an arbitrary r.v. X , it is given by Proposition 6.2 in [8].

A basic functional \mathcal{F} , being convex, has one-sided derivatives in any direction, and

$$(11) \quad \frac{\partial^+}{\partial t} \mathcal{F}(X + tZ) = \sup_{Y \in \partial \mathcal{F}(X+tZ)} \mathbb{E}[YZ],$$

for every $X, Z \in \mathcal{L}^2(\Omega)$, see Theorem 2.1 in [9], which is proved for risk measures, but its proof uses only F1–F3. There exists a *two-sided* derivative in (11) for $t = 0$ in any direction if and only if $\partial \mathcal{F}(X)$ is a singleton [9].

3. Envelope theorem.

3.1. Ordinary envelope theorem. The ordinary envelope theorem [54, Theorem 3.10.4] deals with the optimization problem

$$(12) \quad v(\boldsymbol{\xi}) = \max f(\mathbf{x}, \boldsymbol{\xi}) \quad \text{subject to} \quad \mathbf{g}(\mathbf{x}, \boldsymbol{\xi}) \leq \mathbf{0},$$

where $\mathbf{x} = (x_1, \dots, x_n)^\top$, $\boldsymbol{\xi} = (\xi_1, \dots, \xi_t)^\top$ is the vector of real parameters, $\mathbf{g} = (g_1, \dots, g_m)^\top$, and $v(\boldsymbol{\xi})$ is a value function. It states that if

- (a) For given $\boldsymbol{\xi}$, problem (12) has a unique solution $\mathbf{x}^* = \mathbf{x}^*(\boldsymbol{\xi})$;
- (b) There exists an open ball $\mathcal{B}_\varepsilon(\boldsymbol{\xi})$ and $\delta > 0$ such that for every $\widehat{\boldsymbol{\xi}} \in \mathcal{B}_\varepsilon(\boldsymbol{\xi})$, problem (12) has at least one solution $\widehat{\mathbf{x}} \in \mathcal{B}_\delta(\mathbf{x}^*)$;
- (c) f, g_1, \dots, g_m are continuously differentiable functions in some open ball around $(\mathbf{x}^*(\boldsymbol{\xi}), \boldsymbol{\xi})$; and
- (d) The gradients $\nabla_{\mathbf{x}} g_j(\mathbf{x}^*(\boldsymbol{\xi}), \boldsymbol{\xi})$ corresponding to those constraints that are active for given $\boldsymbol{\xi}$ are linearly independent (*constraint qualification condition*),

then $v(\boldsymbol{\xi})$ is differentiable at $\boldsymbol{\xi}$, and

$$(13) \quad \nabla v(\boldsymbol{\xi}) = \nabla_{\boldsymbol{\xi}} \mathcal{L}(\mathbf{x}, \boldsymbol{\xi}) \Big|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\xi}), \boldsymbol{\lambda}=\boldsymbol{\lambda}^*(\boldsymbol{\xi})},$$

where $\mathcal{L}(\mathbf{x}, \boldsymbol{\xi}) = f(\mathbf{x}, \boldsymbol{\xi}) - \boldsymbol{\lambda}^\top \mathbf{g}(\mathbf{x}, \boldsymbol{\xi})$ is the Lagrangian of (12), $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)^\top$ is the vector of Lagrange multipliers, and $\boldsymbol{\lambda}^*(\boldsymbol{\xi})$ is the value of $\boldsymbol{\lambda}$ at optimality.

If $\boldsymbol{\xi}$ is perturbed by $t\mathbf{h}$ with $\|\mathbf{h}\|_{\mathbb{E}} \equiv \sqrt{\sum_{i=1}^n \mathbb{E}[h_i^2]} = 1$ and some small $t \in \mathbb{R}$ then $v(\boldsymbol{\xi} + t\mathbf{h})$ can be estimated by

$$v(\boldsymbol{\xi} + t\mathbf{h}) = v(\boldsymbol{\xi}) + t\mathbf{h}^\top \nabla v(\boldsymbol{\xi}) + o(t) = v(\boldsymbol{\xi}) + t\mathbf{h}^\top \nabla_{\boldsymbol{\xi}} \mathcal{L}(\mathbf{x}, \boldsymbol{\xi}) \Big|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\xi}), \boldsymbol{\lambda}=\boldsymbol{\lambda}^*(\boldsymbol{\xi})} + o(t).$$

3.2. Envelope theorem for optimization problems with basic functionals. Central to this work is the optimization problem

$$(14) \quad v(\boldsymbol{\xi}) = \min_{\mathbf{x} \in X_0} \mathcal{F}_0(\boldsymbol{\xi}^\top \mathbf{x}) \quad \text{subject to} \quad \mathcal{F}_j(\boldsymbol{\xi}^\top \mathbf{x}) \leq C_j, \quad j = 1, \dots, k,$$

⁴For any choice of $Z \in \text{sign}[X - \mathbb{E}[X]]$, the r.v. $Z|X - \mathbb{E}[X]|^{p-1}$ is the same.

⁵Two r.v.'s $X : \Omega \rightarrow \mathbb{R}$ and $Y : \Omega \rightarrow \mathbb{R}$ are said to be *comonotone*, if there exists a set $A \subset \Omega$ such that $P[A] = 1$ and $(X(\omega_1) - X(\omega_2))(Y(\omega_1) - Y(\omega_2)) \geq 0$ for all $\omega_1, \omega_2 \in A$.

and its perturbed version

$$(15) \quad \phi(t) = \min_{\mathbf{x} \in X_0} \mathcal{F}_0((\boldsymbol{\xi} + t\mathbf{h})^\top \mathbf{x}) \quad \text{subject to} \quad \mathcal{F}_j((\boldsymbol{\xi} + t\mathbf{h})^\top \mathbf{x}) \leq C_j, \quad j = 1, \dots, k,$$

where $\mathbf{x} = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$ is the vector of decision variables, $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)^\top \in [\mathcal{L}^2(\Omega)]^n$ is the vector of random parameters, $\mathbf{h} = (h_1, \dots, h_n)^\top \in [\mathcal{L}^2(\Omega)]^n$ is the vector of random disturbances of $\boldsymbol{\xi}$, $\mathcal{F}_0, \dots, \mathcal{F}_k$ are basic functionals, $(C_1, \dots, C_k) \in \mathbb{R}^k$, $X_0 \subseteq \mathbb{R}^n$ is a closed convex set, and $t \in [-t_0, t_0]$ for some $t_0 > 0$.

Problem (15) will be studied under the following four assumptions:

(A1) No-redundancy: there is no $t \in [-t_0, t_0]$ and $\mathbf{x} \neq 0$ such that $(\boldsymbol{\xi} + t\mathbf{h})^\top \mathbf{x} = 0$, i.e., there is no $i \in \{1, \dots, n\}$ such that $\xi_i + th_i$ is a linear combination of $\xi_j + th_j$ for $j \neq i$.

(A2) Slater's constraint qualification condition: there exists $\widehat{\mathbf{x}} \in X_0$ such that

$$\mathcal{F}_j((\boldsymbol{\xi} + t\mathbf{h})^\top \widehat{\mathbf{x}}) < C_j, \quad j = 1, \dots, k, \quad t \in [-t_0, t_0].$$

(A3) $\mathcal{F}_0(X) < \infty, \dots, \mathcal{F}_k(X) < \infty$ for all $X \in \mathcal{L}^2(\Omega)$. This implies that $\mathcal{F}_j : \mathcal{L}^2(\Omega) \rightarrow \mathbb{R}$, $j = 0, \dots, k$, are continuous functionals [42, Corollary 8B].

(A4) For every $t \in [-t_0, t_0]$, there is a constant $C_t \in \mathbb{R}$ such that the set

$$(16) \quad X_t = \{x \in X_0 \mid \mathcal{F}_0((\boldsymbol{\xi} + t\mathbf{h})^\top \mathbf{x}) \leq C_t, \mathcal{F}_j((\boldsymbol{\xi} + t\mathbf{h})^\top \mathbf{x}) \leq C_j, j = 1, \dots, k\},$$

is non-empty and bounded. If X_t is empty for all $C_t \in \mathbb{R}$, then the feasible set in (15) is empty. The condition that X_t is bounded for some C_t guarantees that it is a compact subset of \mathbb{R}^n , and minimum in (15) exists by Weierstrass extreme value theorem [6].

Assumptions A1–A4 imply the following four properties of (15).

(P1) A1 implies that $\xi_i + th_i \neq 0$ for all i and all $t \in [-t_0, t_0]$ and that there could be at most one i such that $\xi_i + th_i \in \mathbb{R}$. For such i (if exists), it is assumed that $\xi_i \in \mathbb{R}$ and $h_i \in \mathbb{R}$, i.e., variation of a constant is a constant.

(P2) A3 implies that for every $j = 0, \dots, k$ and every fixed $\boldsymbol{\xi}, \mathbf{h}$ and t , $\mathcal{F}_j((\boldsymbol{\xi} + t\mathbf{h})^\top \mathbf{x})$ is a convex function of \mathbf{x} , is finite everywhere on \mathbb{R}^n , and, therefore, is continuous, see [41, Theorem 10.1]. By the same argument, $\mathcal{F}_j((\boldsymbol{\xi} + t\mathbf{h})^\top \mathbf{x})$ is continuous in t for every fixed \mathbf{x} .

(P3) For every fixed $t \in [-t_0, t_0]$, (15) can be rewritten as $\min_{\mathbf{x} \in X_t} \mathcal{F}_0((\boldsymbol{\xi} + t\mathbf{h})^\top \mathbf{x})$ with X_t given by (16). Since $\mathcal{F}_0(\cdot)$ is convex and continuous by A3, and X_t is a closed convex set, which is also non-empty and bounded by A4, an optimal solution of (15) always exists, and

$$(17) \quad \mathcal{X}^*(t) = \arg \min_{\mathbf{x} \in X_t} \mathcal{F}_0((\boldsymbol{\xi} + t\mathbf{h})^\top \mathbf{x})$$

is a non-empty, closed, convex, and bounded set.

(P4) Since $\mathcal{F}_j((\boldsymbol{\xi} + t\mathbf{h})^\top \mathbf{x})$, $j = 0, \dots, k$, are continuous convex functions of \mathbf{x} , the subdifferentials $\partial_{\mathbf{x}} \mathcal{F}_j$, $j = 0, \dots, k$, with respect to \mathbf{x} are non-empty convex compact sets. A2 implies that $\mathbf{x}^* \in X_0$ is a solution to (15) if and only if there exists $\lambda^* \in (\mathbb{R}^+)^k$ such that

$$(18) \quad 0 \in \partial_{\mathbf{x}} \mathcal{F}_0((\boldsymbol{\xi} + t\mathbf{h})^\top \mathbf{x}^*) + \sum_{j=1}^k \lambda_j^* \partial_{\mathbf{x}} \mathcal{F}_j((\boldsymbol{\xi} + t\mathbf{h})^\top \mathbf{x}^*) + N_{X_0}(\mathbf{x}^*)$$

and

$$(19) \quad \lambda_j^* (\mathcal{F}_j((\boldsymbol{\xi} + t\mathbf{h})^\top \mathbf{x}^*) - C_j) = 0, \quad j = 1, \dots, k,$$

where $N_{X_0}(\mathbf{x}^*) = \{\mathbf{y} \in \mathbb{R}^n \mid \mathbf{y}^\top (\mathbf{x} - \mathbf{x}^*) \leq 0 \forall \mathbf{x} \in X_0\}$ is the normal cone to X_0 at \mathbf{x}^* , see [51, Theorem 3.34]. With the chain rule for subdifferentials [26, Corollary 16]

$$\partial_{\mathbf{x}} \mathcal{F}_j((\boldsymbol{\xi} + t\mathbf{h})^\top \mathbf{x}^*) = \{\mathbb{E}[Y_j(\boldsymbol{\xi} + t\mathbf{h})] \mid Y_j \in \partial \mathcal{F}_j((\boldsymbol{\xi} + t\mathbf{h})^\top \mathbf{x}^*)\}, \quad j = 0, \dots, k,$$

so that condition (18) is equivalent to the existence of $Y_j^* \in \partial \mathcal{F}_j((\boldsymbol{\xi} + t\mathbf{h})^\top \mathbf{x}^*)$, $j = 0, \dots, k$, and of Lagrange multipliers $\lambda^* \in (\mathbb{R}^+)^k$ such that

$$(20) \quad 0 \in \mathbb{E} \left[\left(Y_0^* + \sum_{j=1}^k \lambda_j^* Y_j^* \right) (\boldsymbol{\xi} + t\mathbf{h}) \right] + N_{X_0}(\mathbf{x}^*).$$

Let

$$(21) \quad \mathcal{Y}^*(t) = \left\{ Y_0^* + \sum_{j=1}^k \lambda_j^* Y_j^* \mid \lambda_j^* \text{ and } Y_j^* \text{ satisfy (20) and (19)} \right\}.$$

Let $\mathcal{X}^* = \mathcal{X}^*(0)$ and $\mathcal{Y}^* = \mathcal{Y}^*(0)$ be the set of optimal solutions (17) and corresponding subdifferentials (21), respectively, for the unperturbed problem (14).

THEOREM 3.1. *Under A1–A4, ϕ in (15) is continuous and left and right differentiable at 0 with*

$$(22) \quad \begin{aligned} \phi'(0+) &= \min_{\mathbf{x} \in \mathcal{X}^*} \max_{Y \in \mathcal{Y}^*} \mathbb{E} [Y \mathbf{h}^\top \mathbf{x}] = \max_{Y \in \mathcal{Y}^*} \min_{\mathbf{x} \in \mathcal{X}^*} \mathbb{E} [Y \mathbf{h}^\top \mathbf{x}], \\ \phi'(0-) &= \max_{\mathbf{x} \in \mathcal{X}^*} \min_{Y \in \mathcal{Y}^*} \mathbb{E} [Y \mathbf{h}^\top \mathbf{x}] = \min_{Y \in \mathcal{Y}^*} \max_{\mathbf{x} \in \mathcal{X}^*} \mathbb{E} [Y \mathbf{h}^\top \mathbf{x}]. \end{aligned}$$

Proof. By [51, Theorem 4.7], (18) and (19) hold if and only if $(\mathbf{x}^*, \lambda^*)$ is a saddle point of the Lagrangian

$$(23) \quad \mathcal{L}(\mathbf{x}, \lambda, t) = \mathcal{F}_0((\boldsymbol{\xi} + t\mathbf{h})^\top \mathbf{x}) + \sum_{j=1}^k \lambda_j (\mathcal{F}_j((\boldsymbol{\xi} + t\mathbf{h})^\top \mathbf{x}) - C_j),$$

i.e., $\mathcal{L}(\mathbf{x}^*, \lambda, t) \leq \mathcal{L}(\mathbf{x}^*, \lambda^*, t) \leq \mathcal{L}(\mathbf{x}, \lambda^*, t)$ for all $\mathbf{x} \in X_0$, $\lambda \in (\mathbb{R}^+)^k$, and $t \in [-t_0, t_0]$. By [51, Theorem 4.8], problem (15) can then be recast in the form

$$(24) \quad \phi(t) = \max_{\lambda \in (\mathbb{R}^+)^k} \inf_{\mathbf{x} \in X_0} \mathcal{L}(\mathbf{x}, \lambda, t) = \min_{\mathbf{x} \in X_0} \sup_{\lambda \in (\mathbb{R}^+)^k} \mathcal{L}(\mathbf{x}, \lambda, t).$$

The dual characterization (6) of $\mathcal{F}_0, \dots, \mathcal{F}_k$ through the corresponding dual sets $\mathcal{Y}_0, \dots, \mathcal{Y}_k$ yields

$$(25) \quad \phi(t) = \min_{\mathbf{x} \in X_0} \sup_{Y_j \in \mathcal{Y}_j, \lambda \in (\mathbb{R}^+)^k} \underbrace{\left(\mathbb{E} \left[\left(Y_0 + \sum_{j=1}^k \lambda_j Y_j \right) ((\boldsymbol{\xi} + t\mathbf{h})^\top \mathbf{x}) \right] - \sum_{j=1}^k \lambda_j C_j \right)}_{=f(\mathbf{x}, Y_0, \dots, Y_k, \lambda, t)}.$$

Since f in (25) and

$$(26) \quad \frac{\partial f}{\partial t} = \mathbb{E} \left[\left(Y_0 + \sum_{j=1}^k \lambda_j Y_j \right) \mathbf{h}^\top \mathbf{x} \right]$$

are continuous functions of all their arguments, (22) follows⁶ from Theorem 5 in [38]. \square

COROLLARY 3.2. *If in Theorem 3.1, the sets $\mathcal{X}^* = \{\mathbf{x}^*\}$ and $\mathcal{Y}^* = \{Y^*\}$ are singletons, then $\phi'(0+) = \phi'(0-)$, and (22) simplifies to*

$$(27) \quad \phi'(0) = \left. \frac{\partial}{\partial t} \mathcal{L}(\mathbf{x}, \lambda, t) \right|_{\mathbf{x}=\mathbf{x}^*, \lambda=\lambda^*, t=0} = \mathbb{E} [Y^* \mathbf{h}^\top \mathbf{x}^*],$$

where $\mathcal{L}(\mathbf{x}, \lambda, t)$ is defined by (23). Formula (27) is a random-parameter version of the ordinary envelope relationship (13).

⁶Theorem 5 in [38] requires $\mathcal{X}^*(t) \times \mathcal{Y}^*(t) \neq \emptyset$ for $t \in [0, 1]$ and feasible sets for \mathbf{x} , Y and λ to be compact: these technicalities are addressed in Appendix A.

COROLLARY 3.3. *The problem*

$$(28) \quad v(\boldsymbol{\xi}) = \min_{\mathbf{x} \in X_0} \mathcal{F}_0(\boldsymbol{\xi}^\top \mathbf{x}), \quad X_0 = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} \geq \mathbf{b}\},$$

where A and \mathbf{b} are given matrix and vector, respectively, is a particular case of (14) and is frequently encountered in various applications, e.g. in portfolio selection [47, 48]. Let \mathbf{h} be defined as in Theorem 3.1, then $\phi(t) = v(\boldsymbol{\xi} + t\mathbf{h})$ is a particular case of (15), so that $\phi'(0\pm)$ is determined by (22) with \mathcal{X}^* and \mathcal{Y}^* given by

$$(29) \quad (\mathcal{X}^*, \mathcal{Y}^*) = \left\{ (\mathbf{x}^*, Y) \mid Y \in \partial \mathcal{F}_0(\boldsymbol{\xi}^\top \mathbf{x}^*), \mathbb{E}[Y\boldsymbol{\xi}] = A^\top \mathbf{z}, \mathbf{z} \geq 0, \mathbf{z}^\top (\mathbf{A}\mathbf{x}^* - \mathbf{b}) = 0 \right\}.$$

For every $Y \in \mathcal{Y}^*$, there exists \mathbf{z} such that $\mathbb{E}[Y\boldsymbol{\xi}] = A^\top \mathbf{z}$, $\mathbf{z} \geq 0$ and $\mathbf{z}^\top (\mathbf{A}\mathbf{x}^* - \mathbf{b}) = 0$. This implies that $\mathbb{E}[Y\boldsymbol{\xi}^\top \mathbf{x}^*] = \mathbf{z}^\top \mathbf{A}\mathbf{x}^*$, or, equivalently, that $\mathcal{F}_0(\boldsymbol{\xi}^\top \mathbf{x}^*) = \mathbf{b}^\top \mathbf{z}$. In particular, when A is a row \mathbf{a}^\top and \mathbf{b} is a scalar $b \neq 0$, \mathbf{z} is the scalar $z = b^{-1} \mathcal{F}_0(\boldsymbol{\xi}^\top \mathbf{x}^*)$ provided that $z > 0$, and in this case,

$$(30) \quad (\mathcal{X}^*, \mathcal{Y}^*) = \left\{ (\mathbf{x}^*, Y) \mid Y \in \partial \mathcal{F}_0(\boldsymbol{\xi}^\top \mathbf{x}^*), \mathbb{E}[Y\boldsymbol{\xi}] = b^{-1} \mathcal{F}_0(\boldsymbol{\xi}^\top \mathbf{x}^*) \mathbf{a} \right\}.$$

3.3. Robust optimization with unknown variation. In applications, the variation \mathbf{h} in (15) is rarely known. Let $\tilde{\boldsymbol{\xi}}$ be perturbation of $\boldsymbol{\xi}$, and suppose that only the magnitude of $\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi}$:

$$\delta = \|\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi}\|_{\mathbb{E}} \equiv \sqrt{\sum_{i=1}^n \mathbb{E}[(\tilde{\xi}_i - \xi_i)^2]}$$

can be reliably estimated. The next theorem estimates then $v(\tilde{\boldsymbol{\xi}}) - v(\boldsymbol{\xi})$ in terms of δ .

THEOREM 3.4. *Let $v = v(\boldsymbol{\xi})$ and $\tilde{v} = v(\tilde{\boldsymbol{\xi}})$ with v defined in (14). Then under assumptions A1–A4,*

$$(31) \quad \begin{aligned} \tilde{v} - v &\leq \min_{\mathbf{x} \in \mathcal{X}^*} \|\mathbf{x}\|_{\mathbb{R}^n} \max_{Y \in \mathcal{Y}^*} \sqrt{\mathbb{E}[Y^2]} \cdot \delta + o(\delta), \\ v - \tilde{v} &\leq \max_{\mathbf{x} \in \mathcal{X}^*} \|\mathbf{x}\|_{\mathbb{R}^n} \min_{Y \in \mathcal{Y}^*} \sqrt{\mathbb{E}[Y^2]} \cdot \delta + o(\delta), \end{aligned}$$

where \mathcal{X}^* and \mathcal{Y}^* are defined in Theorem 3.1 for (14) with $\boldsymbol{\xi}$. In particular, if $\mathcal{X}^* = \{\mathbf{x}^*\}$ and $\mathcal{Y}^* = \{Y^*\}$ are singletons, then

$$(32) \quad |\tilde{v} - v| \leq \|\mathbf{x}^*\|_{\mathbb{R}^n} \sqrt{\mathbb{E}[(Y^*)^2]} \cdot \delta + o(\delta).$$

Proof. Let $\mathbf{h} = (\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi})/\delta$, so that $\|\mathbf{h}\|_{\mathbb{E}} = \sqrt{\sum_{i=1}^n \mathbb{E}[h_i^2]} = 1$. With $\|\mathbf{x}\|_{\mathbb{R}^n} = \sqrt{\sum_{i=1}^n x_i^2}$, the Cauchy-Schwarz inequality implies that

$$\mathbb{E}[Y\mathbf{h}^\top \mathbf{x}] \leq \sum_{i=1}^n \sqrt{\mathbb{E}[Y^2 x_i^2]} \sqrt{\mathbb{E}[h_i^2]} \leq \sqrt{\mathbb{E}[Y^2] \sum_{i=1}^n x_i^2} \|\mathbf{h}\|_{\mathbb{E}} = \sqrt{\mathbb{E}[Y^2]} \|\mathbf{x}\|_{\mathbb{R}^n},$$

and Theorem 3.1 yields

$$\begin{aligned} \phi'(0+) &= \min_{\mathbf{x} \in \mathcal{X}^*} \max_{Y \in \mathcal{Y}^*} \mathbb{E}[Y\mathbf{h}^\top \mathbf{x}] \leq \min_{\mathbf{x} \in \mathcal{X}^*} \|\mathbf{x}\|_{\mathbb{R}^n} \max_{Y \in \mathcal{Y}^*} \sqrt{\mathbb{E}[Y^2]}, \\ \phi'(0+) &= - \max_{\mathbf{x} \in \mathcal{X}^*} \min_{Y \in \mathcal{Y}^*} -\mathbb{E}[Y\mathbf{h}^\top \mathbf{x}] \geq - \max_{\mathbf{x} \in \mathcal{X}^*} \|\mathbf{x}\|_{\mathbb{R}^n} \min_{Y \in \mathcal{Y}^*} \sqrt{\mathbb{E}[Y^2]}. \end{aligned}$$

Consequently, for $t > 0$,

$$\begin{aligned}\phi(t) - \phi(0) &= t\phi'(0+) + o(t) \leq \min_{\mathbf{x} \in \mathcal{X}^*} \|\mathbf{x}\|_{\mathbb{R}^n} \max_{Y \in \mathcal{Y}^*} \sqrt{\mathbb{E}[Y^2]} |t| + o(t), \\ \phi(0) - \phi(t) &= -t\phi'(0+) + o(t) \leq \max_{\mathbf{x} \in \mathcal{X}^*} \|\mathbf{x}\|_{\mathbb{R}^n} \min_{Y \in \mathcal{Y}^*} \sqrt{\mathbb{E}[Y^2]} |t| + o(t),\end{aligned}$$

which with $t = \delta$ yields (31). \square

3.4. Reinterpretation of sensitivity. When $\boldsymbol{\xi}$ is perturbed/updated, the true question is not, in fact, about the change in the optimal value of \mathcal{F}_0 in (15), i.e. about the difference $\phi(t) - \phi(0)$, but whether the original solution \mathbf{x}^* of (14) should be recomputed. In other words, how far is \mathbf{x}^* from the optimality in (15)? This can be assessed by the difference

$$(33) \quad \psi(t) = \mathcal{F}_0((\boldsymbol{\xi} + t\mathbf{h})^\top \mathbf{x}^*) - \phi(t) + \max_{\lambda^* \in \Lambda(\mathbf{x}^*)} \sum_{j=1}^k \lambda_j^* (\mathcal{F}_j((\boldsymbol{\xi} + t\mathbf{h})^\top \mathbf{x}^*) - C_j),$$

where $\mathcal{F}_0((\boldsymbol{\xi} + t\mathbf{h})^\top \mathbf{x}^*) - \phi(t)$ and $\lambda_j^* (\mathcal{F}_j((\boldsymbol{\xi} + t\mathbf{h})^\top \mathbf{x}^*) - C_j)$ are penalties for the objective function and constraint violation, respectively, and $\Lambda(\mathbf{x}^*)$ is the set of Lagrange multipliers λ^* satisfying the optimality conditions (19) and (20) for $t = 0$ and $\mathbf{x} = \mathbf{x}^*$. In this case, (11) and (22) imply that

$$(34) \quad \psi'(0+) = \max_{\lambda^* \in \Lambda(\mathbf{x}^*)} \max_{Y_j \in \partial \mathcal{F}_j(\boldsymbol{\xi}^\top \mathbf{x}^*)} \mathbb{E}[(Y_0 + \sum_{j=1}^k \lambda_j^* Y_j) \mathbf{h}^\top \mathbf{x}^*] - \min_{\mathbf{x} \in \mathcal{X}^*} \max_{Y \in \mathcal{Y}^*} \mathbb{E}[Y \mathbf{h}^\top \mathbf{x}].$$

If the sets $\mathcal{X}^* = \{\mathbf{x}^*\}$ and $\partial \mathcal{F}_j(\boldsymbol{\xi}^\top \mathbf{x}^*) = \{Y_j^*\}$, $j = 0, 1, \dots, k$, are all singletons, then (34) yields $\psi'(0) = 0$, which also follows from (27). For twice differentiable $\psi(t)$, $\psi(t) = O(t^2)$, which indicates that for small t , \mathbf{x}^* can be kept as a near optimal solution for (15).

3.5. Extensions and generalizations. Theorem 3.1 relies on assumptions A1, A2, and A4 which do not allow problem (15) to de-generate in certain senses. It can be extended in several ways: (i) relaxing A3, i.e., establishing the envelope theorem for not necessarily finite basic functionals on $\mathcal{L}^2(\Omega)$, (ii) replacing $\mathcal{L}^2(\Omega)$ by $\mathcal{L}^p(\Omega)$, $1 \leq p \leq \infty$, and (iii) relaxing F1, i.e., establishing the envelope theorem for arbitrary convex functionals.

In (i), the envelope theorem cannot be established for all such functionals. For example, $\mathcal{F}_0(X) = -\inf X$ is a basic functional on $\mathcal{L}^2(\Omega)$, which is a coherent risk measure. Its dual set (7) is given by $\mathcal{Y} = \{Y \in \mathcal{L}^2(\Omega) \mid \mathbb{E}[Y] = -1\}$. This set is unbounded, and, for any bounded continuous r.v. X , e.g. for X being uniformly distributed on $(0, 1)$, the supremum in (6) is not attained, so that $\partial \mathcal{F}_0(X) = \emptyset$, and (22) does not hold. Of course, failure of (22) does not imply that the directional derivatives $\phi'(0\pm)$ do not exist. However, Appendix B presents an optimization problem with $\mathcal{F}_0(X) = -\inf X$, where $\phi'(0+)$ does not even exist.

In (ii), Theorem 3.1 does not hold all finite basic functionals \mathcal{F} on $\mathcal{L}^p(\Omega)$, $p \in [1, \infty]$. By A3, \mathcal{F} is finite and hence continuous [42, Corollary 8B], whereas continuity of \mathcal{F} implies that the dual set \mathcal{Y} in (6) is bounded [46, Theorem 1]. For $p = 2$, $\mathcal{L}^p(\Omega)$ is a reflexive Banach space, in which case, bounded \mathcal{Y} is weakly compact [13, Theorem 3.31], so that the supremum in (6) is always attained [28] and $\partial \mathcal{F}(X) \neq \emptyset$. For $p \neq 2$, $\mathcal{L}^p(\Omega)$ is not reflexive, and the above argument does not hold. In fact, $\mathcal{F}_0(X) = -\inf X$, defined on $\mathcal{L}^\infty(\Omega)$, is a finite and continuous basic functional that can be represented by (6) with $\mathcal{Y} = \{Y \in \mathcal{L}^1(\Omega) \mid \mathbb{E}[Y] = -1\}$, but the supremum in (6) is not attained. Thus, at least in the case of $p = \infty$, \mathcal{F} requires additional conditions, such as ‘‘continuity from below’’ [15, Corollary 4.37], to guarantee that the supremum in (6) is attained.

For convex functionals on $\mathcal{L}^2(\Omega)$, which are finite and continuous but not positively homogeneous, the dual characterization (6) does not hold and Theorem 5 in [38] does not seem to be directly applicable. Extending Theorem 3.1 for this case calls for a separate work.

4. Optimization with error, deviation, risk, and regret measures.

4.1. Generalized linear regression. The generalized linear regression introduced in [50] seeks to approximate an r.v. η by a linear combination of given r.v.'s ξ_1, \dots, ξ_n . It minimizes an error measure $\mathcal{E} : \mathcal{L}^2(\Omega) \rightarrow [0, \infty]$ of the approximation error $\eta - c_0 - \sum_{k=1}^n c_k \xi_k$ with respect to regression coefficients c_0, c_1, \dots, c_n :

$$(35) \quad \min_{c_0 \in \mathbb{R}, \mathbf{c} \in \mathbb{R}^n} \mathcal{E}(\eta - c_0 - \boldsymbol{\xi}^\top \mathbf{c}),$$

where $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)^\top$ and $\mathbf{c} = (c_1, \dots, c_n)^\top$.

The *error decomposition theorem* in [50] states that (35) is equivalent to minimizing the deviation measure \mathcal{D} projected from \mathcal{E} with respect to \mathbf{c} :

$$(36) \quad v_{\mathcal{E}}(\boldsymbol{\xi}, \eta) = \min_{\mathbf{c} \in \mathbb{R}^n} \mathcal{D}(\eta - \boldsymbol{\xi}^\top \mathbf{c}),$$

and then setting optimal \mathbf{c}^* to the statistic \mathcal{S} of $\eta - \boldsymbol{\xi}^\top \mathbf{c}^*$ associated with \mathcal{E} , i.e.

$$\mathbf{c}^* \in \mathcal{S}(\eta - \boldsymbol{\xi}^\top \mathbf{c}^*) \equiv \arg \min_{c_0 \in \mathbb{R}} \mathcal{E}(\eta - c_0 - \boldsymbol{\xi}^\top \mathbf{c}^*).$$

By [50, Theorem 4.1], \mathbf{c}^* solves (36) if and only if there exists a $Y \in \partial \mathcal{D}(\eta - \boldsymbol{\xi}^\top \mathbf{c}^*)$ such that $\mathbb{E}[Y \boldsymbol{\xi}] = 0$, i.e. the optimal solution set \mathcal{C}^* and corresponding subdifferential \mathcal{Y}^* are given by

$$(37) \quad (\mathcal{C}^*, \mathcal{Y}^*) = \left\{ (\mathbf{c}^*, Y) \mid Y \in \partial \mathcal{D}(\eta - \boldsymbol{\xi}^\top \mathbf{c}^*), \mathbb{E}[Y \boldsymbol{\xi}] = 0 \right\}.$$

Let $\boldsymbol{\xi} + t \mathbf{h} = (\xi_1 + t h_1, \dots, \xi_n + t h_n)$ and $\eta + t h_0$ be perturbed $\boldsymbol{\xi}$ and η , respectively, where $h_0 \in \mathcal{L}^2(\Omega)$, $\mathbf{h} \in (\mathcal{L}^2(\Omega))^n$, and $t \in [-t_0, t_0]$ for some $t_0 > 0$, and let

$$(38) \quad \phi_{\mathcal{E}}(t) = v_{\mathcal{E}}(\boldsymbol{\xi} + t \mathbf{h}, \eta + t h_0).$$

Then Theorem 3.1 yields

$$(39) \quad \phi'_{\mathcal{E}}(0+) = \min_{\mathbf{c} \in \mathcal{C}^*} \max_{Y \in \mathcal{Y}^*} \mathbb{E}[Y(h_0 - \mathbf{h}^\top \mathbf{c})] = \max_{Y \in \mathcal{Y}^*} \min_{\mathbf{c} \in \mathcal{C}^*} \mathbb{E}[Y(h_0 - \mathbf{h}^\top \mathbf{c})].$$

EXAMPLE 4.1 (quadratic error). *Problem (35) with $\mathcal{E}(X) = \|X\|_2$ is equivalent to (36) with $\mathcal{D} = \sigma$, and (39) with (8a) implies that*

$$\phi'_{\mathcal{E}}(0+) = \mathbb{E} \left[\frac{\eta - \boldsymbol{\xi}^\top \mathbf{c}^* - \mathbb{E}[\eta - \boldsymbol{\xi}^\top \mathbf{c}^*]}{\sigma(\eta - \boldsymbol{\xi}^\top \mathbf{c}^*)} (h_0 - \mathbf{h}^\top \mathbf{c}^*) \right] = \frac{\text{cov}(\eta - \boldsymbol{\xi}^\top \mathbf{c}^*, h_0 - \mathbf{h}^\top \mathbf{c}^*)}{\sigma(\eta - \boldsymbol{\xi}^\top \mathbf{c}^*)},$$

where \mathbf{c}^* is the (unique) optimal solution to (36) and where cov denotes covariance.

EXAMPLE 4.2 (Koenker-Bassett error measure [32]). *Problem (35) with $\mathcal{E} = \mathcal{E}_{1,1/\alpha-1,1}$, $\alpha \in (0, 1)$, is equivalent to (36) with $\mathcal{D} = \text{CVaR}_{\alpha}^{\Delta}$. If (36) has a unique solution \mathbf{c}^* such that $\rho^* = \eta - \boldsymbol{\xi}^\top \mathbf{c}^*$ satisfies $\mathbb{P}[\rho^* = q_{\rho^*}(\alpha)] = 0$, then (39) with (9) implies that*

$$(40) \quad \phi'_{\mathcal{E}}(0+) = \mathbb{E}[h_0 - \mathbf{h}^\top \mathbf{c}^*] - \mathbb{E}[h_0 - \mathbf{h}^\top \mathbf{c}^* \mid \rho^* < q_{\rho^*}(\alpha)].$$

For (35) with $\mathcal{E} = \|X\|_1 \equiv \mathcal{E}_{1,1,1}$ (median regression), $\phi'_{\mathcal{E}}(0+)$ is given by (40) with $\alpha = 1/2$.

Function (33) shows how far the solution \mathbf{c}^* of (36) is from the optimality in (38):

$$\psi_{\mathcal{E}}(t) = \mathcal{D}((\eta + t h_0) - (\boldsymbol{\xi} + t \mathbf{h})^\top \mathbf{c}^*) - \phi_{\mathcal{E}}(t),$$

and (34) implies that

$$(41) \quad \psi'_{\mathcal{E}}(0+) = \max_{Y \in \partial \mathcal{D}(\eta - \boldsymbol{\xi}^\top \mathbf{c}^*)} \mathbb{E}[Y(h_0 - \mathbf{h}^\top \mathbf{c}^*)] - \min_{\mathbf{c} \in \mathcal{C}^*} \max_{Y \in \mathcal{Y}^*} \mathbb{E}[Y(h_0 - \mathbf{h}^\top \mathbf{c})].$$

For example, for $\mathcal{D} = \sigma$, $\psi'_{\mathcal{E}}(0+) = \psi'_{\mathcal{E}}(0-) = 0$.

EXAMPLE 4.3 (Koenker-Bassett error measure: non-singleton case). In (38) with $\mathcal{D} = \text{CVaR}_{1/3}^\Delta$, let $\Omega = \{\omega_1, \omega_2, \omega_3\}$, $\mathbb{P}[\omega_i] = 1/3$, $i = 1, 2, 3$, $\eta = (\eta(\omega_1), \eta(\omega_2), \eta(\omega_3)) = (0, 1, 3)$, $n = 1$, $\xi_1 = (-1, 0, 1)$, $h_0 = (0, 0, 0)$, $h_1 = (0, -1, 1)$. Then by evaluating directly, $\phi_{\mathcal{D}}(t) = \frac{1}{3} - \frac{2t}{1+2t}$ for $t \in [0, 3/4]$, so that $\phi'_{\mathcal{D}}(0+) = -2$. For (36), $\mathcal{C}^* = [1, 2]$ and $\mathcal{Y}^* = \{Y^*\} = \{(1, -2, 1)\}$, and (39) reduces to $\min_{c_1 \in [1, 2]}(-c_1) = -2$. Now, let $c_1^* = 1$. Then by evaluating directly, $\psi_{\mathcal{D}}(t) = \frac{1}{3} - \phi_{\mathcal{D}}(t) = \frac{2t}{1+2t}$, and $\psi'_{\mathcal{D}}(0+) = 2$. Also, $\partial \mathcal{D}(\eta - \xi^\top \mathbf{c}^*) = \{Y | Y = (y, -1 - y, 1), y \in [-2, 1]\}$, and (41) reduces to $\psi'_{\mathcal{D}}(0+) = \max_{y \in [-2, 1]} \frac{-y-2}{3} - (-2) = 2$.

In application to (36), Theorem 3.4 yields the following corollary.

COROLLARY 4.1 (robust linear regression). Let $v = v_{\mathcal{D}}(\xi, \eta)$ and $\tilde{v} = v_{\mathcal{D}}(\tilde{\xi}, \tilde{\eta})$, where $v_{\mathcal{D}}$ is defined in (36) with a continuous deviation measure \mathcal{D} . Then (31) holds with $\mathcal{X}^* = \{\mathbf{x} = (1, \mathbf{c}), \mathbf{c} \in \mathcal{C}^*\}$, where \mathcal{C}^* and \mathcal{Y}^* are given by (37), $\delta = \sqrt{\mathbb{E}[(\tilde{\eta} - \eta)^2] + \|\tilde{\xi} - \xi\|_{\mathbb{E}}^2}$, and where $\mathbb{E}[Y^2]$ simplifies to $\sigma(Y)$. A less precise but easier to compute estimate is given by

$$(42) \quad |\tilde{v} - v| \leq \max_{\mathbf{c} \in \mathcal{C}^*} \sqrt{1 + \sum_{k=1}^n (c_k^*)^2} \max_{Y \in \partial \mathcal{D}(\eta - \xi^\top \mathbf{c})} \sigma(Y) \cdot \delta + o(\delta).$$

EXAMPLE 4.4 (quadratic error: robust regression). For (36) with $\mathcal{D} = \sigma$, $\mathcal{C}^* = \{\mathbf{c}^*\}$ is a singleton, $\sigma(Y) = 1$, and (42) simplifies to $|\tilde{v} - v| \leq \sqrt{1 + \sum_{k=1}^n (c_k^*)^2} \cdot \delta + o(\delta)$.

EXAMPLE 4.5 (Koenker-Bassett error measure: robust regression). For (36) with $\mathcal{D} = \text{CVaR}_{\alpha}^\Delta$, and any $\mathbf{c} \in \mathcal{C}^*$, $\max_{Y \in \partial \mathcal{D}(\eta - \xi^\top \mathbf{c})} \sigma(Y)$ in (42) simplifies to $\alpha^{-1} - 1$.

Detail. Let $Y^* \in \partial \text{CVaR}_{\alpha}^\Delta(\rho^*)$, $\rho^* = \eta - \xi^\top \mathbf{c}$, be such that $\mathbb{P}[Y^* = \alpha^{-1}] = \alpha$ and $\mathbb{P}[Y^* = 0] = 1 - \alpha$, with $\sigma^2(Y^*) = \alpha \cdot \alpha^{-2} + (1 - \alpha) \cdot 0 - 1^2 = \alpha^{-1} - 1$. Then (9) implies that Y^* is dominated by all other elements of $\partial \text{CVaR}_{\alpha}^\Delta(\rho^*)$ with respect to concave ordering⁷, so that $\sigma^2(Y^*) \geq \sigma^2(Y)$ for $Y \in \partial \text{CVaR}_{\alpha}^\Delta(\rho^*)$, see [17]. \square

4.2. Minimization of deviation and regret. Suppose there are n risky instruments with rates of return r_1, \dots, r_n . Let x_1, \dots, x_n be proportions of the initial capital (portfolio weights) invested in those instruments. Then x_1, \dots, x_n satisfy the budget constraint $\mathbf{e}^\top \mathbf{x} = 1$, where $\mathbf{e} = (1, \dots, 1)^\top$, and the portfolio rate of return is determined by $r_p = \mathbf{r}^\top \mathbf{x}$, where $\mathbf{r} = (r_1, \dots, r_n)^\top$ and $\mathbf{x} = (x_1, \dots, x_n)^\top$. An investor may wish then to minimize either a deviation measure \mathcal{D} or a regret measure \mathcal{V} of r_p :

$$(43) \quad v_{\mathcal{D}}(\mathbf{r}) = \min_{\mathbf{x} \in X_0} \mathcal{D}(\mathbf{r}^\top \mathbf{x}), \quad v_{\mathcal{V}}(\mathbf{r}) = \min_{\mathbf{x} \in X_0} \mathcal{V}(\mathbf{r}^\top \mathbf{x}),$$

where $X_0 = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{e}^\top \mathbf{x} = 1\}$. Let $\mathbf{r} + t \mathbf{h}$ be disturbance of \mathbf{r} , where $\mathbf{h} \in (\mathcal{L}^2(\Omega))^n$, and $t \in [-t_0, t_0]$ for some $t_0 > 0$, and let

$$(44) \quad \phi_{\mathcal{D}}(t) = v_{\mathcal{D}}(\mathbf{r} + t \mathbf{h}), \quad \phi_{\mathcal{V}}(t) = v_{\mathcal{V}}(\mathbf{r} + t \mathbf{h}).$$

With assumptions A1–A4, there is no asset redundancy and \mathcal{D} and \mathcal{V} are continuous, and both problems in (43) are particular cases of (28) with $\xi = \mathbf{r}$, $A = \mathbf{e}^\top$ and with \mathbf{b} being scalar of 1. Then Corollary 3.3 implies that $\phi'_{\mathcal{D}}(0\pm)$ and $\phi'_{\mathcal{V}}(0\pm)$ are given by (22), in which $\mathcal{X}_{\mathcal{D}}^*(0)$, $\mathcal{Y}_{\mathcal{D}}^*(0)$, $\mathcal{X}_{\mathcal{V}}^*(0)$ and $\mathcal{Y}_{\mathcal{V}}^*(0)$ are determined by (30) with $\mathcal{F}_0 = \mathcal{D}$ and $\mathcal{F}_0 = \mathcal{V}$, respectively:

$$(\mathcal{X}_{\mathcal{F}_0}^*, \mathcal{Y}_{\mathcal{F}_0}^*) = \left\{ (\mathbf{x}^*, Y) \mid Y \in \partial \mathcal{F}_0(\xi^\top \mathbf{x}^*), \mathbb{E}[Y \xi] = \mathcal{F}_0(\xi^\top \mathbf{x}^*) \mathbf{e} \right\}, \quad \mathcal{F}_0 = \mathcal{D}, \mathcal{V}.$$

⁷An r.v. X dominates an r.v. Y with respect to concave ordering, or $X \succeq_c Y$, if $\mathbb{E}[X] = \mathbb{E}[Y]$, and $\int_{-\infty}^x F_X(t) dt \leq \int_{-\infty}^x F_Y(t) dt$ for all $t \in \mathbb{R}$, see [17].

If $\mathcal{X}_{\mathcal{D}}^*$, $\mathcal{Y}_{\mathcal{D}}^*$, $\mathcal{X}_{\mathcal{V}}^*$ and $\mathcal{Y}_{\mathcal{V}}^*$ are all singletons, then $\phi'_{\mathcal{D}}(0+) = \phi'_{\mathcal{D}}(0-)$ and $\phi'_{\mathcal{V}}(0+) = \phi'_{\mathcal{V}}(0-)$, and $\phi'_{\mathcal{D}}(0)$ and $\phi'_{\mathcal{V}}(0)$ are given by (27). For brevity, let $h^* = \mathbf{h}^\top \mathbf{x}^*$.

EXAMPLE 4.6. For σ , σ_- , MAD, and $\text{CVaR}_{\alpha}^{\Delta}$, the subdifferentials are given by (8a)–(8c), and (9), respectively, and for $v_{\mathcal{D}}(\mathbf{r})$ in (43) with these deviation measures, (27) yields

$$\phi'_{\mathcal{D}}(0) = \begin{cases} \mathbb{E}[\partial\sigma(\rho^*)h^*], & \mathcal{D} = \sigma, \\ \mathbb{E}[\partial\sigma_-(\rho^*)h^*], & \mathcal{D} = \sigma_-, \\ 2(\mathbb{E}[h^* | \rho^* > \mathbb{E}[\rho^*]] - \mathbb{E}[h^*])\mathbb{P}[\rho^* > \mathbb{E}[\rho^*]], & \mathcal{D} = \text{MAD}, \\ \mathbb{E}[h^*] - \mathbb{E}[h^* | \rho^* < q_{\rho^*}(\alpha)], & \mathcal{D} = \text{CVaR}_{\alpha}^{\Delta}, \end{cases}$$

where for $\mathcal{D} = \text{MAD}$ and $\mathcal{D} = \text{CVaR}_{\alpha}^{\Delta}$, it is assumed that $\arg \min_{\mathbf{x} \in X_0} \mathcal{D}(\mathbf{r}^\top \mathbf{x})$ is a singleton $\{\mathbf{x}^*\}$ such that $\rho^* = \mathbf{r}^\top \mathbf{x}^*$ satisfies $\mathbb{P}[\rho^* = \mathbb{E}[\rho^*]] = 0$ and $\mathbb{P}[\rho^* = q_{\rho^*}(\alpha)] = 0$, respectively.

EXAMPLE 4.7. Let $\mathcal{V}(X) = 2\mathbb{E}[X]_-$, and let $\mathcal{X}_{\mathcal{V}}^* = \{\mathbf{x}^*\}$ be a singleton such that $\rho^* = \mathbf{r}^\top \mathbf{x}^*$ satisfies $\mathbb{P}[\rho^* = 0] = 0$, then for $v_{\mathcal{V}}(\mathbf{r})$ in (43), (27) yields

$$\phi'_{\mathcal{V}}(0) = -2\mathbb{E}[h^* | \rho^* < 0]\mathbb{P}[\rho^* < 0].$$

4.3. Mean-deviation portfolio optimization. This time suppose there are $n + 1$ instruments available for investment: a risk-free instrument with the rate of return r_0 and n risky instruments with rates of return r_1, \dots, r_n . Let x_0 and x_1, \dots, x_n be proportions of the initial capital (portfolio weights) invested in the risk-free instrument and in n risky instruments, respectively. Then the portfolio rate of return is determined by $r_p = r_0 x_0 + \mathbf{r}^\top \mathbf{x}$, where $\mathbf{r} = (r_1, \dots, r_n)^\top$ and $\mathbf{x} = (x_1, \dots, x_n)^\top$. The portfolio problem is to minimize a deviation measure \mathcal{D} of r_p subject to the budget constraint $x_0 + \mathbf{e}^\top \mathbf{x} = 1$, where $\mathbf{e} = (1, \dots, 1)^\top$, and subject to a constraint on the portfolio expected rate of return to be greater than or equal to $r_0 + \Delta$ with $\Delta > 0$ being a premium over r_0 , i.e.

$$(45) \quad v_{\mathcal{D}}^{\Delta}(r_0, \mathbf{r}) = \min_{(x_0, \mathbf{x}) \in X_0} \mathcal{D}(\mathbf{r}^\top \mathbf{x}) \quad \text{subject to} \quad \mathbb{E}[r_0 x_0 + \mathbf{r}^\top \mathbf{x}] \geq r_0 + \Delta,$$

where $X_0 = \{(x_0, \mathbf{x}) \mid x_0 + \mathbf{e}^\top \mathbf{x} = 1\}$. It generalizes Markowitz's mean-variance portfolio selection that corresponds to $\mathcal{D} = \sigma$. Positive homogeneity of \mathcal{D} implies that (x_0, \mathbf{x}) solves (45) if and only if

$$\mathbf{x} = \Delta \cdot \mathbf{y}, \quad x_0 = 1 - \Delta \cdot \mathbf{e}^\top \mathbf{y},$$

where \mathbf{y} , called *basic fund* [47], is a solution to the optimization problem

$$(46) \quad v_{\mathcal{D}}(r_0, \mathbf{r}) = \min_{\mathbf{y} \in \mathbb{R}^n} \mathcal{D}(\mathbf{r}^\top \mathbf{y}) \quad \text{subject to} \quad \mathbb{E}[(\mathbf{r} - r_0 \mathbf{e})^\top \mathbf{y}] \geq 1,$$

and the optimal values in (45) and (46) are related by

$$(47) \quad v_{\mathcal{D}}^{\Delta}(r_0, \mathbf{r}) = \Delta \cdot v_{\mathcal{D}}(r_0, \mathbf{r}).$$

Let $r_0 + t h_0$ and $\mathbf{r} + t \mathbf{h}$ be variations of r_0 ⁸ and \mathbf{r} , respectively, where $h_0 \in \mathbb{R}$, $\mathbf{h} = (h_1, \dots, h_n)$ with $h_i \in \mathcal{L}^2$, $i = 1, \dots, n$, and $t \in [-t_0, t_0]$ for some $t_0 > 0$. Then let

$$(48) \quad \phi_{\mathcal{D}}(t) = v_{\mathcal{D}}(r_0 + t h_0, \mathbf{r} + t \mathbf{h}).$$

Under A1–A4, (48) is a special case of (15) with parameters $\xi = \mathbf{r} - r_0 \mathbf{e}$, disturbances $\mathbf{h} - h_0 \mathbf{e}$, $X_0 = \mathbb{R}^n$, $j = 1$, $\mathcal{F}_0 = \mathcal{D}$, and $\mathcal{F}_1 = -\mathbb{E}$. In this case, A2 is equivalent to the condition that

⁸Different banks offer different risk-free rates and change them based on market conditions.

for every $t \in [-t_0, t_0]$, $\mathbb{E}[(r_i + th_i)]$, $i = 1, \dots, n$ are not all the same, A3 implies that \mathcal{D} is continuous, and A4 holds automatically, see [47, Proposition 4]. Then Theorem 4 in [48] implies that for (46), $Y_0 \in \partial \mathcal{D}(\mathbf{r}^\top \mathbf{y}^*)$, $\lambda_1^* = \mathcal{D}(\mathbf{r}^\top \mathbf{y}^*)$ and $Y_1^* = -1$ in (21), so that

$$(49) \quad (\mathcal{X}^*, \mathcal{Y}^*) = \{(\mathbf{y}^*, Y) \mid Y + \mathcal{D}(\mathbf{r}^\top \mathbf{y}^*) \in \partial \mathcal{D}(\mathbf{r}^\top \mathbf{y}^*), \mathbb{E}[Y\mathbf{r}] = -\mathcal{D}(\mathbf{r}^\top \mathbf{y}^*)r_0\mathbf{e}\},$$

and Theorem 3.1 yields

$$(50) \quad \begin{aligned} \phi'_{\mathcal{D}}(0+) &= \min_{\mathbf{y} \in \mathcal{X}^*} \max_{Y \in \mathcal{Y}^*} \mathbb{E}[Y(\mathbf{h} - h_0\mathbf{e})^\top \mathbf{y}] = \max_{Y \in \mathcal{Y}^*} \min_{\mathbf{y} \in \mathcal{X}^*} \mathbb{E}[Y(\mathbf{h} - h_0\mathbf{e})^\top \mathbf{y}], \\ \phi'_{\mathcal{D}}(0-) &= \max_{\mathbf{y} \in \mathcal{X}^*} \min_{Y \in \mathcal{Y}^*} \mathbb{E}[Y(\mathbf{h} - h_0\mathbf{e})^\top \mathbf{y}] = \min_{Y \in \mathcal{Y}^*} \max_{\mathbf{y} \in \mathcal{X}^*} \mathbb{E}[Y(\mathbf{h} - h_0\mathbf{e})^\top \mathbf{y}]. \end{aligned}$$

If $\mathcal{X}^* = \{\mathbf{y}^*\}$ and $\partial \mathcal{D}(\mathbf{r}^\top \mathbf{y}^*) = \{Y^*\}$ are singletons, then $\phi'_{\mathcal{D}}(0+) = \phi'_{\mathcal{D}}(0-)$ and with $h^* = \mathbf{h}^\top \mathbf{y}^*$, (50) simplifies to

$$(51) \quad \phi'_{\mathcal{D}}(0) \equiv \left. \frac{\partial}{\partial t} v_{\mathcal{D}}(r_0 + th_0, \mathbf{r} + t\mathbf{h}) \right|_{t=0} = \mathbb{E}[Y^* h^*] - v_{\mathcal{D}}(r_0, \mathbf{r}) (\mathbb{E}[h^*] - h_0\mathbf{e}^\top \mathbf{y}^*).$$

COROLLARY 4.2. For (45), let $\phi_{\Delta}(t) = v_{\mathcal{D}}^{\Delta}(r_0 + th_0, \mathbf{r} + t\mathbf{h})$. Then $\phi'_{\Delta}(0\pm) = \Delta \cdot \phi'_{\mathcal{D}}(0\pm)$, where $\phi'_{\mathcal{D}}(0\pm)$ are given by (50).

COROLLARY 4.3. The investor, who demands $\mathbb{E}[r_p]$ to be μ , solves

$$(52) \quad v_{\mathcal{D}}^{\mu}(r_0, \mathbf{r}) = \min_{(x_0, \mathbf{x}) \in X_0} \mathcal{D}(\mathbf{r}^\top \mathbf{x}) \quad \text{subject to} \quad \mathbb{E}[r_0 x_0 + \mathbf{r}^\top \mathbf{x}] \geq \mu,$$

where $X_0 = \{(x_0, \mathbf{x}) \mid x_0 + \mathbf{e}^\top \mathbf{x} = 1\}$. Then $\phi_{\mu}(t) = v_{\mathcal{D}}^{\mu}(r_0 + th_0, \mathbf{r} + t\mathbf{h}) = (\mu - r_0 - th_0)\phi_{\mathcal{D}}(t)$, and

$$(53) \quad \phi'_{\mu}(0\pm) = (\mu - r_0)\phi'_{\mathcal{D}}(0\pm) - h_0\phi_{\mathcal{D}}(0) = (\mu - r_0)\phi'_{\mathcal{D}}(0\pm) - \frac{h_0}{\mu - r_0}\phi_{\mu}(0),$$

where $\phi'_{\mathcal{D}}(0\pm)$ is given by (50).

In the following examples, let \mathbf{y}^* be the vector of optimal portfolio weights in (46), and for brevity, let $\rho^* = \mathbf{r}^\top \mathbf{y}^*$ and $h^* = \mathbf{h}^\top \mathbf{y}^*$.

EXAMPLE 4.8 (standard deviation). Let Λ be the covariance matrix for the random rates of return r_1, \dots, r_n and let Λ^{-1} be its inverse, then for (46) with $\mathcal{D} = \sigma$,

$$\mathbf{y}^* = \frac{\Lambda^{-1}\boldsymbol{\zeta}}{\boldsymbol{\zeta}^\top \Lambda^{-1}\boldsymbol{\zeta}}, \quad \sigma(\rho^*) = \frac{1}{\sqrt{\boldsymbol{\zeta}^\top \Lambda^{-1}\boldsymbol{\zeta}}}, \quad \boldsymbol{\zeta} = \mathbb{E}[\mathbf{r}] - r_0\mathbf{e},$$

see [56, §8.2.1], and with (8a), (51) yields

$$(54) \quad \phi'_{\sigma}(0) = \frac{\mathbb{E}[\left(\frac{(\mathbf{r} - \mathbb{E}[\mathbf{r}])^\top \Lambda^{-1}\boldsymbol{\zeta} - 1}{\boldsymbol{\zeta}^\top \Lambda^{-1}\boldsymbol{\zeta}}\right) \mathbf{h}^\top \Lambda^{-1}\boldsymbol{\zeta}]}{(\boldsymbol{\zeta}^\top \Lambda^{-1}\boldsymbol{\zeta})^{3/2}} + h_0 \frac{\mathbf{e}^\top \Lambda^{-1}\boldsymbol{\zeta}}{(\boldsymbol{\zeta}^\top \Lambda^{-1}\boldsymbol{\zeta})^{3/2}}.$$

In the case of $h_0 = 0$, Appendix C verifies (54) directly.

EXAMPLE 4.9. For σ_- , MAD, $\text{CVaR}_{\alpha}^{\Delta}$, and deviations of \mathcal{L}^p -type and semi- \mathcal{L}^p type, i.e. \mathcal{D}_p and \mathcal{D}_- , the subdifferentials are given by (8b), (8c), (9), (10a), and (10b),⁹ respec-

⁹For $p > 2$, $\mathcal{D}_p(X)$ is not continuous on $\mathcal{L}^2(\Omega)$, and derivation of (50) does not work as stated. However, under the assumption that $r_i \in \mathcal{L}^p(\Omega)$ and $h_i \in \mathcal{L}^p(\Omega)$, $i = 1, \dots, n$, $\mathcal{D}_p(X)$ can be defined on $\mathcal{L}^p(\Omega)$, and in this case, it is continuously differentiable, and (51) follows from (27). The same remark concerns deviation measures of semi- \mathcal{L}^p type.

tively, and for (46) with these deviation measures, (51) yields

$$\phi'_{\mathcal{D}}(0) = \begin{cases} \mathbb{E}[\partial \sigma_{-}(\rho^*)h^*] - \sigma_{-}(\rho^*)(\mathbb{E}[h^*] - h_0 \mathbf{e}^{\top} \mathbf{y}^*), \\ 2(\mathbb{E}[h^* | \rho^* > \mathbb{E}[\rho^*]] - \mathbb{E}[h^*]) \mathbb{P}[\rho^* > \mathbb{E}[\rho^*]] - \text{MAD}(\rho^*)(\mathbb{E}[h^*] - h_0 \mathbf{e}^{\top} \mathbf{y}^*), \\ \mathbb{E}[h^*] - \mathbb{E}[h^* | \rho^* < q_{\rho^*}(\alpha)] - \text{CVaR}_{\alpha}^{\Delta}(\rho^*)(\mathbb{E}[h^*] - h_0 \mathbf{e}^{\top} \mathbf{y}^*), \\ \mathbb{E}[\partial \mathcal{D}_p(\rho^*)h^*] - \mathcal{D}_p(\rho^*)(\mathbb{E}[h^*] - h_0 \mathbf{e}^{\top} \mathbf{y}^*), \\ \mathbb{E}[\partial \mathcal{D}_{-}(\rho^*)h^*] - \mathcal{D}_{-}(\rho^*)(\mathbb{E}[h^*] - h_0 \mathbf{e}^{\top} \mathbf{y}^*), \end{cases}$$

respectively, where for $\mathcal{D} = \text{MAD}$ and $\mathcal{D} = \text{CVaR}_{\alpha}^{\Delta}$, it is assumed that (45) has unique solutions \mathbf{y}^* such that $\rho^* = \mathbf{r}^{\top} \mathbf{y}^*$ satisfies $\mathbb{P}[\rho^* = \mathbb{E}[\rho^*]] = 0$ and $\mathbb{P}[\rho^* = q_{\rho^*}(\alpha)] = 0$, respectively.

EXAMPLE 4.10 ($\mathcal{D} = \text{CVaR}_{\alpha}^{\Delta}$: non-singleton case). In (48) with $\mathcal{D} = \text{CVaR}_{1/3}^{\Delta}$, let Ω be defined as in Example 4.3, and let $n = 2$, $r_1 = \eta$, $r_2 = -\xi_1$, $r_0 = h_0 = 0$, $h_1 = (0, 0, 0)$, $h_2 = (0, 1, -1)$ with η and ξ_1 from Example 4.3. Then by evaluating directly, $\phi_{\mathcal{D}}(t) = \frac{1}{4} - \frac{3t}{2+4t}$ for $t \in [0, 3/4]$, so that $\phi'_{\mathcal{D}}(0+) = -3/2$. For (46), $\mathcal{X}^* = \{(y_1, y_2) | y_1 = 3/4, y_2 \in [3/4, 3/2]\}$ and $\mathcal{Y}^* = \{Y^*\} = \{(3/4, -9/4, 3/4)\}$, and (50) reduces to $\min_{y_2 \in [3/4, 3/2]} (-y_2) = -3/2$.

EXAMPLE 4.11 ($\mathcal{D} = \sigma$: numerical illustration). Let a portfolio be formed out of 99 instruments from the FTSE 100 index,¹⁰ and let $r_0 = 0.01\%$ and $\Delta = 0.02\%$. Means and variance-covariance matrix of instruments' daily rates are estimated from the data for the period 2015/04/01–2016/04/01 ($T = 251$ days). The disturbance vector \mathbf{h} is selected randomly: for each instrument $i = 1, \dots, n$, and scenario $j = 1, \dots, T$, h_{ij} comes from a standard normal distribution. Figure 1 shows the value function $v(t) = \Delta \cdot \phi_{\sigma}(t)$ in (48) with $\mathcal{D} = \sigma$ and also shows its linear approximation $v(0) + v'(0)t$ with $v'(0) = \Delta \cdot \phi'_{\sigma}(0)$ given by (54).

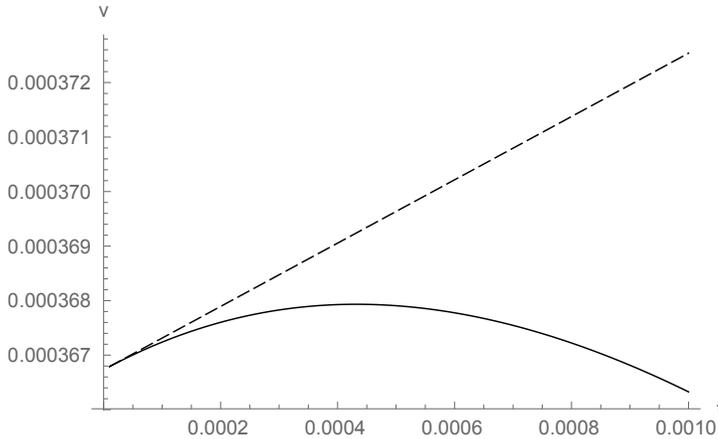


FIG. 1. The value function $v(t) = \Delta \cdot \phi_{\sigma}(t)$ in (48) with standard deviation and its linear approximation $v(0) + v'(0)t$ with $v'(0) = \Delta \cdot \phi'_{\sigma}(0)$ given by (54): solid and dashed curves, respectively.

For r , r_0 , Δ and h from Example 4.11, Figure 2 depicts the value function $v(t) = \Delta \cdot \phi_{\mathcal{D}}(t)$ in (48) with $\mathcal{D} = \sigma_{-}$ and also shows its linear approximation $v(0) + v'(0)t$ with $v'(0) = \Delta \cdot \phi'_{\mathcal{D}}(0)$ given in Example 4.9 for $\mathcal{D} = \sigma_{-}$.

When r_0 and \mathbf{r} are perturbed/updated, the investor may be concerned how far the original optimal solution \mathbf{y}^* in (46) deviates from the optimality in terms of penalty function (33),

¹⁰The index contains 101 instruments, two of which are excluded due to the lack of data.

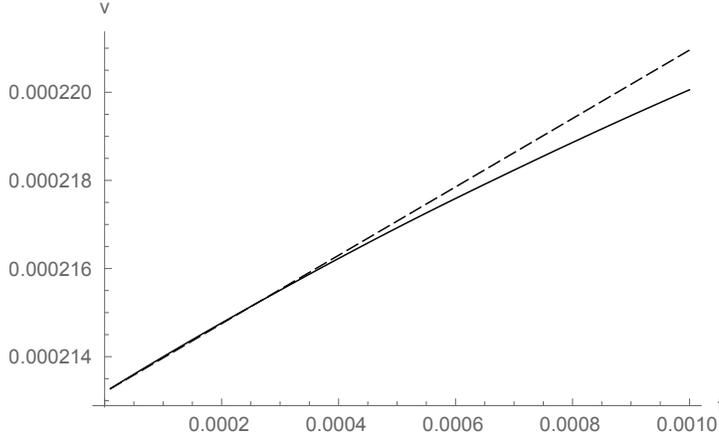


FIG. 2. The value function $v(t) = \Delta \cdot \phi_{\sigma_-}(t)$ in (48) with standard lower semideviation and its linear approximation $v(0) + v'(0)t$ with $v'(0) = \Delta \cdot \phi'_{\sigma_-}(0)$ given in Example 4.9 for $\mathcal{D} = \sigma_-$: solid and dashed curves, respectively.

which simplifies to

$$\psi_{\mathcal{D}}(t) = \mathcal{D}((\mathbf{r} + t\mathbf{h})^\top \mathbf{y}^*) - \phi_{\mathcal{D}}(t) + \mathcal{D}(\mathbf{r}^\top \mathbf{y}^*) (1 - \mathbb{E}[\langle (\mathbf{r} + t\mathbf{h}) - (r_0 + t h_0)\mathbf{e}, \mathbf{y}^* \rangle]).$$

In this case, $\psi'_{\mathcal{D}}(0+)$ is given by (34), which with (11) and (50) takes the form

$$(55) \quad \psi'_{\mathcal{D}}(0+) = \max_{Y_0 \in \partial \mathcal{D}(\mathbf{r}^\top \mathbf{y}^*)} \mathbb{E}[(Y_0 - \mathcal{D}(\mathbf{r}^\top \mathbf{y}^*)) (\mathbf{h} - h_0 \mathbf{e})^\top \mathbf{y}^*] - \min_{\mathbf{y} \in \mathcal{X}^*} \max_{Y \in \mathcal{Y}^*} \mathbb{E}[Y (\mathbf{h} - h_0 \mathbf{e})^\top \mathbf{y}].$$

If \mathbf{y}^* is the only solution to (46) and if $\partial \mathcal{D}(\mathbf{r}^\top \mathbf{y}^*) = \{Y^*\}$ is a singleton, then $\psi'_{\mathcal{D}}(0+) = 0$.

EXAMPLE 4.12 (CVaR $_{\alpha}^{\Delta}$). In Example 4.10, let $\mathbf{y}^* = (3/4, 3/4)$. Then by evaluating directly, $\psi_{\mathcal{D}}(t) = \frac{3t}{2+4t}$, and $\psi'_{\mathcal{D}}(0+) = 3/2$. Also, $\partial \mathcal{D}(\mathbf{r}^\top \mathbf{y}^*) = \{Y | Y = (y, -1 - y, 1), y \in [-2, 1]\}$, and (55) reduces to $\psi'_{\mathcal{D}}(0+) = \max_{y \in [-2, 1]} \frac{-y-2}{4} - (-3/2) = 3/2$.

4.3.1. Robust portfolio optimization: application of Theorem 3.4 to problem (46).

COROLLARY 4.4. Let $v = v_{\mathcal{D}}(r_0, \mathbf{r})$ and $\tilde{v} = v_{\mathcal{D}}(\tilde{r}_0, \tilde{\mathbf{r}})$ be optimal values in (46) with the rates or return (r_0, \mathbf{r}) and $(\tilde{r}_0, \tilde{\mathbf{r}})$, respectively, and let \mathcal{D} be a continuous deviation measure. Then (31) holds with $\delta = \sqrt{\sum_{i=1}^n \mathbb{E}[(\tilde{r}_i - \tilde{r}_0) - (r_i - r_0)]^2}$ and with \mathcal{X}^* and \mathcal{Y}^* determined in (49). A less precise but easier to compute estimate is given by

$$(56) \quad |\tilde{v} - v| \leq \max_{\mathbf{y} \in \mathcal{X}^*} \|\mathbf{y}\|_{\mathbb{R}^n} \max_{Y \in \partial \mathcal{D}(\mathbf{r}^\top \mathbf{y})} \sqrt{\sigma^2(Y) + v^2} \cdot \delta + o(\delta).$$

In particular, if $\mathcal{X}^* = \{\mathbf{y}^*\}$ and $\partial \mathcal{D}(\mathbf{r}^\top \mathbf{y}^*) = \{Y^*\}$ are singletons, then (32) simplifies to

$$(57) \quad |\tilde{v} - v| \leq \|\mathbf{y}^*\|_{\mathbb{R}^n} \sqrt{\sigma^2(Y^*) + v^2} \cdot \delta + o(\delta).$$

EXAMPLE 4.13 (standard deviation). For (46) with $\mathcal{D} = \sigma$, the solution \mathbf{y}^* is given in Example 4.8, and $v = \sigma(\mathbf{r}^\top \mathbf{y}^*)$. Then since $\sigma^2(Y^*) = 1$, (57) takes the form

$$|\tilde{v} - v| \leq \frac{\sqrt{\boldsymbol{\zeta}^\top \Lambda^{-2} \boldsymbol{\zeta}} \sqrt{1 + \boldsymbol{\zeta}^\top \Lambda^{-1} \boldsymbol{\zeta}}}{(\boldsymbol{\zeta}^\top \Lambda^{-1} \boldsymbol{\zeta})^{3/2}} \cdot \delta + o(\delta).$$

EXAMPLE 4.14 (standard lower semideviation). For $\mathcal{D} = \sigma_-$, (57) holds with

$$\sigma^2(Y^*) + v^2 = 1 - \frac{(\mathbb{E}[\rho^* - \mathbb{E}[\rho^*]])_-^2}{\sigma_-^2(\rho^*)} + \sigma_-^2(\rho^*), \quad \rho^* = \mathbf{r}^\top \mathbf{y}^*.$$

EXAMPLE 4.15 (mean absolute deviation and CVaR deviation). For $\mathcal{D} = \text{MAD}$ and $\mathcal{D} = \text{CVaR}_\alpha^\Delta$, and any $\mathbf{y} \in \mathcal{X}^*$, $\max_{Y \in \partial \mathcal{D}(\mathbf{r}^\top \mathbf{y})} \sigma^2(Y)$ in (56) takes the form

$$(58) \quad \max_{Y \in \partial \mathcal{D}(\mathbf{r}^\top \mathbf{y})} \sigma^2(Y) = \begin{cases} 4p^*(1-p^*), & \mathcal{D} = \text{MAD}, \\ \alpha^{-1} - 1, & \mathcal{D} = \text{CVaR}_\alpha^\Delta, \end{cases}$$

where $p^* = \max\{\mathbb{P}[\rho^* > \mathbb{E}[\rho^*]], \mathbb{P}[\rho^* < \mathbb{E}[\rho^*]], 1/2\}$ and $\rho^* = \mathbf{r}^\top \mathbf{y}$.

Detail. If $Y \in \partial \text{MAD}(\rho^*)$, then $Y = Z - \mathbb{E}[Z]$ for some $Z \in \text{sign}[\rho^* - \mathbb{E}[\rho^*]]$, and $\sigma^2(Y) = \sigma^2(Z)$. Let $Z^* \in \text{sign}[\rho^* - \mathbb{E}[\rho^*]]$ be such that $\mathbb{E}[Z^*] = \mathbb{E}[Z]$ and let Z^* take values -1 and 1 only: $\mathbb{P}[Z^* = 1] = p$ and $\mathbb{P}[Z^* = -1] = 1 - p$, where $p \in [\mathbb{P}[\rho^* > \mathbb{E}[\rho^*]], 1 - \mathbb{P}[\rho^* < \mathbb{E}[\rho^*]]]$. Then Z dominates Z^* with respect to concave ordering, which implies that $\sigma^2(Z^*) \geq \sigma^2(Z)$, see [17], and $\sigma^2(Z^*) = 4p(1-p)$ attains maximum at $p = p^*$. This proves (58) for $\mathcal{D} = \text{MAD}$, whereas (58) for $\mathcal{D} = \text{CVaR}_\alpha^\Delta$ follows from Example 4.5. \square

4.3.2. Comparison of robust properties for different deviation measures. With (56), the relative error $|\tilde{v} - v|/v$ is estimated by

$$(59) \quad \frac{|\tilde{v} - v|}{v} \leq \max_{\mathbf{y} \in \mathcal{X}^*} \|\mathbf{y}\|_{\mathbb{R}^n} \sqrt{\frac{1}{v^2} \max_{Y \in \partial \mathcal{D}(\mathbf{r}^\top \mathbf{y})} \sigma^2(Y) + 1} \cdot \delta + \frac{o(\delta)}{v}.$$

For any continuous deviation measure \mathcal{D} and any r.v. X , the smaller the coefficient

$$(60) \quad \gamma(\mathcal{D}, X) = \frac{1}{\mathcal{D}(X)^2} \max_{Y \in \partial \mathcal{D}(X)} \sigma^2(Y),$$

the better estimate (59) for $X = \mathbf{r}^\top \mathbf{y}$.

PROPOSITION 4.5. For any deviation measure \mathcal{D} and any r.v. X , $\gamma(\mathcal{D}, X) \geq \gamma(\sigma, X)$.

Proof. For any $Y \in \partial \mathcal{D}(X)$,

$$\gamma(\mathcal{D}, X) \geq \frac{\sigma^2(Y)}{\mathcal{D}(X)^2} = \frac{\sigma^2(Y)}{\text{cov}(Y, X)^2} \geq \frac{\sigma^2(Y)}{\sigma^2(Y)\sigma^2(X)} = \frac{1}{\sigma^2(X)} = \gamma(\sigma, X),$$

where the last equality follows from (60) and Example 1 in [48]. \square

Proposition 4.5 suggests that *optimization with standard deviation is more robust than that with any other deviation measure* in the sense that estimate (59) is the tightest for standard deviation. In the view of Proposition 4.5, the ratio

$$\frac{\gamma(\mathcal{D}, X)}{\gamma(\sigma, X)} = \frac{\sigma^2(X)}{\mathcal{D}(X)^2} \max_{Y \in \partial \mathcal{D}(X)} \sigma^2(Y)$$

may indicate “loss in robustness” when \mathcal{D} is used in place of σ .

EXAMPLE 4.16 (normal distribution). Let X be normally distributed. Then

$$\frac{\gamma(\sigma_-, X)}{\gamma(\sigma, X)} = 2 - \frac{2}{\pi} \approx 1.36, \quad \frac{\gamma(\text{MAD}, X)}{\gamma(\sigma, X)} = \frac{\pi}{2} \approx 1.57, \quad \frac{\gamma(\text{CVaR}_\alpha^\Delta, X)}{\gamma(\sigma, X)} = \frac{\alpha(1-\alpha)}{\phi^2(\Phi^{-1}(\alpha))},$$

where ϕ and Φ are probability density function and CDF of a standard normal distribution, respectively. In particular,

$$\frac{\gamma(\text{CVaR}_{0.5}^\Delta, X)}{\gamma(\sigma, X)} = \frac{\pi}{2}, \quad \frac{\gamma(\text{CVaR}_{0.05}^\Delta, X)}{\gamma(\sigma, X)} \approx 4.47, \quad \frac{\gamma(\text{CVaR}_{0.01}^\Delta, X)}{\gamma(\sigma, X)} \approx 13.94.$$

Detail. Since the ratio $\gamma(\mathcal{D}, X)/\gamma(\sigma, X)$ is invariant under a linear transformation of X , X can be assumed to have a standard normal distribution. Then, the elements Y of the subdifferential for the r.v. X with respect to σ_- , MAD and $\text{CVaR}_\alpha^\Delta$ are given by (8b), (8c), and (9), respectively, for which

$$\sigma^2(Y) = \frac{\mathbb{E}[[X]_-]^2 - (\mathbb{E}[[X]_-])^2}{\sigma_-(X)^2} = 1 - \frac{1}{\pi}, \quad \sigma^2(Y) = 1, \quad \sigma^2(Y) = \alpha^{-1} - 1,$$

respectively. With $\sigma_-(X) = 1/\sqrt{2}$, $\text{MAD}(X) = \sqrt{2/\pi}$ and $\text{CVaR}_\alpha^\Delta(X) = \alpha^{-1}\phi(\Phi^{-1}(\alpha))$, the statement follows. \square

It is known that if the rates of return \mathbf{r} are jointly normally distributed, then optimization with an arbitrary deviation measure yields the same optimal solution \mathbf{x}^* , and in this case, all deviation measures are equivalent. However, perturbed rates of return $\tilde{\mathbf{r}}$ are not assumed to be normally distributed, and the deviation measures are *not* equivalent from the robustness perspective. While standard deviations $\sigma(\mathbf{r}^\top \mathbf{x})$ and $\sigma(\tilde{\mathbf{r}}^\top \mathbf{x})$ of the original and perturbed portfolios are relatively close to each other, CVaR deviations of these portfolios can differ significantly, especially for small α . Example 4.16 is in agreement with the well-known fact that CVaR is difficult to estimate for small α [4] and with the observation that “estimation risk” for mean absolute deviation is greater than that for standard deviation [53].

4.3.3. One fund theorem and capital asset pricing model.

Theorem 3 in [47] generalizes the one-fund theorem for optimal mean-variance portfolio: for any $\Delta > 0$, optimal portfolio in (45) is a combination of a risk-free asset and a basic fund with zero weight in the risk-free asset. Depending on the value of the risk-free rate of return, a basic fund may have either positive, negative or zero price: a basic fund with price of ± 1 is called *master fund of positive/negative types*, whereas a basic fund with zero price is called *master fund of threshold type* (see [48, Definition 2]). The classical one-fund theorem (for $\mathcal{D} = \sigma$) covers only the case of positive master fund, which is called *market portfolio*.

Suppose that r_M is the rate of return of a master fund of positive type for given r_0 and \mathbf{r} and suppose r_p is the rate of return of an optimal portfolio in (45), then the generalized one-fund theorem [47, 48] implies that

$$(61a) \quad x_0^*(r_0, \mathbf{r}) = \frac{\mathbb{E}[r_M] - \mathbb{E}[r_p]}{\mathbb{E}[r_M] - r_0},$$

$$(61b) \quad \mathcal{D}(r_p) = \mathcal{D} \left(\frac{\mathbb{E}[r_M] - \mathbb{E}[r_p]}{\mathbb{E}[r_M] - r_0} r_0 + \frac{\mathbb{E}[r_p] - r_0}{\mathbb{E}[r_M] - r_0} r_M \right) \equiv \frac{\mathbb{E}[r_p] - r_0}{\mathbb{E}[r_M] - r_0} \mathcal{D}(r_M),$$

where $1 - x_0^*(r_0, \mathbf{r})$ is the weight of the master fund in an optimal portfolio and (61b) is called *efficient frontier* (*security market line* for $\mathcal{D} = \sigma$), see Figure 3.

In fact, if \mathbf{y}^* is a basic fund, i.e. a solution to (46), such that $\mathbf{e}^\top \mathbf{y}^* > 0$, then $\mathbf{x}^* = \mathbf{y}^*/\mathbf{e}^\top \mathbf{y}^*$ defines a master fund of positive type [47, Definition 2] with the rate of return $r_M = \mathbf{r}^\top \mathbf{x}^*$, and

$$(62) \quad \frac{\mathcal{D}(r_M)}{\mathbb{E}[r_M] - r_0} = \mathcal{D}(\mathbf{r}^\top \mathbf{y}^*) \equiv v_{\mathcal{D}}(r_0, \mathbf{r}),$$

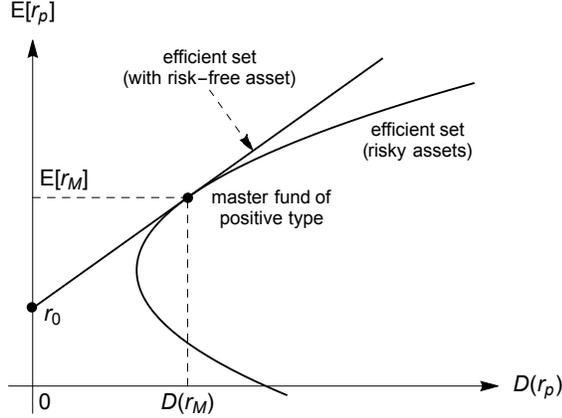


FIG. 3. Efficient sets (frontiers) for portfolio problem (45): expected rate of return versus deviation of optimal mean-deviation portfolios when the risk-free asset is available (security market line for $\mathcal{D} = \sigma$) and when it is not.

where $v_{\mathcal{D}}(r_0, \mathbf{r})$ is defined in (46). The ratio

$$S(r_0, \mathbf{r}) = \frac{\mathbb{E}[r_M] - r_0}{\mathcal{D}(r_M)} \equiv \frac{1}{v_{\mathcal{D}}(r_0, \mathbf{r})}$$

is the slope of the efficient frontier and is also called the *Sharpe ratio* for $\mathcal{D} = \sigma$. In this case,

$$(63) \quad \phi_S(t) = S(r_0 + t h_0, \mathbf{r} + t \mathbf{h}) \equiv \frac{1}{\phi_{\mathcal{D}}(t)} \quad \text{and} \quad \phi'_S(0_{\pm}) = -\frac{\phi'_{\mathcal{D}}(0_{\pm})}{(\phi_{\mathcal{D}}(0))^2},$$

where $\phi_{\mathcal{D}}(t)$ and $\phi'_{\mathcal{D}}(0_{\pm})$ are given by (48) and (50), respectively.

In fact, with (62), (61b) takes the form

$$(64) \quad \mathcal{D}(r_p) = v_{\mathcal{D}}(r_0, \mathbf{r}) (\mathbb{E}[r_p] - r_0),$$

which also follows from (47).

Let $h_0 = 1$ and $\mathbf{h} \equiv 0$, then with (61a) and (61b), (50) yields

$$(65) \quad \frac{\partial^+}{\partial r_0} v_{\mathcal{D}}(r_0, \mathbf{r}) \leq \frac{\mathcal{D}(r_M)}{(\mathbb{E}[r_M] - r_0)^2} \leq \frac{\partial^-}{\partial r_0} v_{\mathcal{D}}(r_0, \mathbf{r}),$$

where $\partial^{\pm}/\partial r_0$ are the right/left partial derivatives. In particular, $v_{\mathcal{D}}(r_0, \mathbf{r})$ is an increasing function on r_0 . With (65), (63) implies that

$$\frac{\partial^+}{\partial r_0} S(r_0, \mathbf{r}) \geq -\frac{1}{\mathcal{D}(r_M)} \geq \frac{\partial^-}{\partial r_0} S(r_0, \mathbf{r}).$$

With (53) and (62), (65) yields

$$(66) \quad \frac{\partial^+}{\partial r_0} v_{\mathcal{D}}^{\mu}(r_0, \mathbf{r}) \leq -\frac{\mathbb{E}[r_M] - \mu}{(\mathbb{E}[r_M] - r_0)^2} \mathcal{D}(r_M) \leq \frac{\partial^-}{\partial r_0} v_{\mathcal{D}}^{\mu}(r_0, \mathbf{r}),$$

where μ and $v_{\mathcal{D}}^{\mu}(r_0, \mathbf{r})$ are defined in (52). In particular, for a master fund, (65) implies that

$$(67) \quad \mu = \mathbb{E}[r_M] \iff \frac{\partial^+}{\partial r_0} v_{\mathcal{D}}^{\mu}(r_0, \mathbf{r}) \leq 0 \leq \frac{\partial^-}{\partial r_0} v_{\mathcal{D}}^{\mu}(r_0, \mathbf{r}),$$

which can be used as a condition for finding $\mathbb{E}[r_M]$.

EXAMPLE 4.17 (standard deviation). In Example 4.8, condition (67) with (53) and (54) for $\mathbf{h} \equiv \mathbf{0}$ and $h_0 = 1$ determines the expected rate of return of a master fund of positive type:

$$\frac{\partial v_{\sigma}^{\mu}(r_0, \mathbf{r})}{\partial r_0} = \frac{(\boldsymbol{\mu} \mathbf{e} - \mathbb{E}[\mathbf{r}])^{\top} \Lambda^{-1} \boldsymbol{\zeta}}{(\boldsymbol{\zeta}^{\top} \Lambda^{-1} \boldsymbol{\zeta})^{3/2}} = 0 \quad \implies \quad \mu_M = \mathbb{E}[r_M] = \frac{\mathbb{E}[\mathbf{r}]^{\top} \Lambda^{-1} \boldsymbol{\zeta}}{\mathbf{e}^{\top} \Lambda^{-1} \boldsymbol{\zeta}}.$$

Let r_M be the rate of return of a master fund of positive type for a given $r_0 = r_0^*$. If r_0 changes by some small ε , can the same master fund be used in the one-fund theorem? This question is similar to that in §3.4, but the difference is that here only the rate of return r_M of the master fund and not the optimal portfolio \mathbf{x}^* is assumed to be fixed, so that weights of the risk-free asset and the master fund should be re-balanced to satisfy the constraint in (48). Let

$$\psi_M(r_0 + \varepsilon) = \frac{\mathcal{D}(r_M)}{\mathbb{E}[r_M] - (r_0 + \varepsilon)} - v_{\mathcal{D}}(r_0 + \varepsilon, \mathbf{r})$$

be the change in the slope of the efficient frontier—note that $\psi_M(r_0) = 0$ by (62). Then, if $\frac{\partial}{\partial r_0} v_{\mathcal{D}}(r_0, \mathbf{r})$ exists at $r_0 = r_0^*$, (65) yields

$$\psi'_M(r_0^*) = \frac{\mathcal{D}(r_M)}{(\mathbb{E}[r_M] - r_0^*)^2} - \frac{\mathcal{D}(r_M)}{(\mathbb{E}[r_M] - r_0^*)^2} = 0,$$

whereas if $\psi_M(r_0)$ is twice differentiable at $r_0 = r_0^*$, we obtain

$$\psi_M(r_0^* + \varepsilon) = \psi_M(r_0^*) + \psi'_M(r_0^*)\varepsilon + \psi''_M(r_0^*)\varepsilon^2 + O(\varepsilon^3) = \psi''_M(r_0^*)\varepsilon^2 + O(\varepsilon^3).$$

In other words, if r_0 is perturbed by ε , the master fund remains optimal up to $O(\varepsilon^2)$. If, however, $\frac{\partial}{\partial r_0} v_{\mathcal{D}}(r_0, \mathbf{r})$ does not exist, then the use of the same master fund in the one-fund theorem may not be appropriate.

EXAMPLE 4.18 (standard deviation). For $\mathcal{D} = \sigma$, $\psi_M(\cdot)$ is twice differentiable and

$$\psi_M(r_0^* + \varepsilon) = \frac{\mathbf{e}^{\top} \Lambda^{-1} (\boldsymbol{\zeta} \mathbf{e}^{\top} - \mathbf{e} \boldsymbol{\zeta}^{\top}) \Lambda^{-1} \boldsymbol{\zeta}}{(\boldsymbol{\zeta}^{\top} \Lambda^{-1} \boldsymbol{\zeta})^{5/2}} \varepsilon^2 + O(\varepsilon^3),$$

where $\boldsymbol{\zeta} = \mathbb{E}[\mathbf{r}] - r_0^* \mathbf{e}$, and Λ is the covariance matrix defined in Example 4.8.

4.3.4. Maximizing expectation with a constraint on deviation. Being still concerned about only the expected value $\mathbb{E}[r_p]$ and a deviation $\mathcal{D}(r_p)$ of portfolio rate of return $r_p = r_0 x_0 + \mathbf{r}^{\top} \mathbf{x}$, the investor may wish to maximize $\mathbb{E}[r_p]$ subject to a constraint on $\mathcal{D}(r_p)$:

$$(68) \quad v_{\mathbb{E}}(r_0, \mathbf{r}) = \max_{(x_0, \mathbf{x}) \in X_0} \mathbb{E}[r_0 x_0 + \mathbf{r}^{\top} \mathbf{x}] \quad \text{subject to} \quad \mathcal{D}(\mathbf{r}^{\top} \mathbf{x}) \leq d,$$

where $X_0 = \{(x_0, \mathbf{x}) \mid x_0 + \mathbf{e}^{\top} \mathbf{x} = 1\}$ and $d > 0$. Let

$$(69) \quad \phi_{\mathbb{E}}(t) = v_{\mathbb{E}}(r_0 + t h_0, \mathbf{r} + t \mathbf{h}).$$

Of course, $\phi'_{\mathbb{E}}(0 \pm)$ can be obtained similarly to $\phi'_{\mathcal{D}}(0 \pm)$ in (50). However, it is instructive to address it through problem (45). Indeed, (64) (one-fund theorem) implies that problems (45) and (68) have the same master funds and the same efficient frontiers¹¹ and

$$v_{\mathbb{E}}(r_0, \mathbf{r}) = \Delta \frac{d}{v_{\mathcal{D}}^{\Delta}(r_0, \mathbf{r})} + r_0 = \frac{d}{v_{\mathcal{D}}(r_0, \mathbf{r})} + r_0,$$

¹¹ $(\mathcal{D}(r_p), \mathbb{E}[r_p])$ is in the efficient frontier of (45) if and only if there is no feasible r'_p such that $\mathbb{E}[r'_p] \geq \mathbb{E}[r_p]$ and $\mathcal{D}(r'_p) \leq \mathcal{D}(r_p)$ with at least one inequality being strict, which implies that $(\mathcal{D}(r_p), \mathbb{E}[r_p])$ is in the efficient frontier of (68).

where the second equality follows from (47). Consequently,

$$(70) \quad \phi_{\mathbb{E}}(t) = \frac{d}{\phi_{\mathcal{D}}(t)} + r_0 + t h_0 \quad \text{and} \quad \phi'_{\mathbb{E}}(0\pm) = -\frac{d}{(\phi_{\mathcal{D}}(0))^2} \phi'_{\mathcal{D}}(0\pm) + h_0,$$

where $\phi_{\mathcal{D}}(t)$ and $\phi'_{\mathcal{D}}(0\pm)$ are given by (48) and (50), respectively.

For σ and σ_- and for MAD and $\text{CVaR}_{\alpha}^{\Delta}$ under the assumption that (68) has unique solutions \mathbf{x}^* such that $\rho^* = \mathbf{r}^{\top} \mathbf{x}^*$ satisfies $\mathbb{P}[\rho^* = \mathbb{E}[\rho^*]] = 0$ and $\mathbb{P}[\rho^* = q_{\rho^*}(\alpha)] = 0$, respectively, $\phi'_{\mathcal{D}}(0+) = \phi'_{\mathcal{D}}(0-)$, and the derivative $\phi'_{\mathcal{D}}(0)$ in (70) is given in Examples 4.8 and 4.9.

4.4. Minimization of risk. The constraint in (45) is always active, and (45) can be equivalently recast in terms of the averse measure of risk $\mathcal{R}(X) = \mathcal{D}(X) - \mathbb{E}[X]$, see (1), as

$$(71) \quad v_{\mathcal{R}}^{\Delta}(r_0, \mathbf{r}) = \min_{(x_0, \mathbf{x}) \in X_0} \mathcal{R}(r_0 x_0 + \mathbf{r}^{\top} \mathbf{x}) \quad \text{subject to} \quad \mathbb{E}[r_0 x_0 + \mathbf{r}^{\top} \mathbf{x}] = r_0 + \Delta$$

with $X_0 = \{(x_0, \mathbf{x}) \mid x_0 + \mathbf{e}^{\top} \mathbf{x} = 1\}$. In this case,

$$(72) \quad v_{\mathcal{R}}^{\Delta}(r_0, \mathbf{r}) = v_{\mathcal{D}}^{\Delta}(r_0, \mathbf{r}) - (r_0 + \Delta) = \Delta \cdot v_{\mathcal{D}}(r_0, \mathbf{r}) - (r_0 + \Delta),$$

where $v_{\mathcal{D}}^{\Delta}(r_0, \mathbf{r})$ is defined in (45). Let

$$\phi_{\mathcal{R}}(t) = v_{\mathcal{R}}^{\Delta}(r_0 + t h_0, \mathbf{r} + t \mathbf{h}).$$

In view of (72), $\phi_{\mathcal{R}}(t) = \Delta \cdot \phi_{\mathcal{D}}(t) - (r_0 + t h_0 + \Delta)$, where $\phi_{\mathcal{D}}(t)$ is defined by (48), and consequently,

$$(73) \quad \phi'_{\mathcal{R}}(0\pm) = \Delta \cdot \phi'_{\mathcal{D}}(0\pm) - h_0$$

with $\phi'_{\mathcal{D}}(0\pm)$ determined by (50).

EXAMPLE 4.19 (CVaR). Let (71) with $\mathcal{R} = \text{CVaR}_{\alpha}$, $\alpha \in (0, 1)$, have a unique solution (x_0^*, \mathbf{x}^*) such that $\rho^* = \mathbf{r}^{\top} \mathbf{x}^*$ satisfies $\mathbb{P}[\rho^* = q_{\rho^*}(\alpha)] = 0$, then (73) and Example 4.9 yield

$$\phi'_{\mathcal{R}}(0+) = \mathbb{E}[h^*] - \mathbb{E}[h^* \mid \rho^* < q_{\rho^*}(\alpha)] - \Delta^{-1} \text{CVaR}_{\alpha}^{\Delta}(\rho^*)(\mathbb{E}[h^*] - h_0 \mathbf{e}^{\top} \mathbf{x}^*) - h_0, \quad h^* = \mathbf{h}^{\top} \mathbf{x}^*.$$

5. Conclusions. The envelope relationship (Theorem 3.1) is obtained for optimization problems whose objectives and constraints involve arbitrary positively homogeneous convex functionals defined on the space of random variables. When problem solution sets and corresponding subdifferentials are singletons, it simplifies to random-parameter version (27) of the ordinary envelope relationship. The function $\psi(t)$, introduced by (33), then indicates how far an optimal solution \mathbf{x}^* of an unperturbed problem is from optimality in the corresponding perturbed problem. When (27) is applicable, $\psi(0) = \psi'(0) = 0$, which implies that \mathbf{x}^* can be kept as a nearly optimal solution in the case of small perturbations. Theorem 3.4 estimates perturbation of the value function of optimization problems in terms of perturbations of problem parameters. Theorems 3.1 and 3.4 are then specialized for optimization problems with error, regret, deviation and risk measures such as generalized linear regression and optimal portfolio selection. The results are believed to be novel even for Markowitz's mean-variance portfolio selection (Examples 4.8, 4.11, 4.13, 4.17, and 4.18). Remarkably, portfolio selection with standard deviation is "more robust" than that with any other deviation measure (§4.3.2).

Acknowledgment. We are grateful to the referees for the comments and suggestions, which helped to improve the quality of the paper. The first author thanks the University of Leicester for granting him the academic study leave to do this research.

Appendix A. Technical details of the proof of Theorem 3.1.

Theorem 5 in [38] requires $\mathcal{X}^*(t) \times \mathcal{Y}^*(t) \neq \emptyset$ for $t \in [0, 1]$ and the feasible sets for \mathbf{x} , \mathcal{Y} and λ to be compact.

Obviously, Theorem 5 in [38] can be applied to $t \in [-t_0, t_0]$ instead of $t \in [0, 1]$, and $\mathcal{X}^*(t) \times \mathcal{Y}^*(t) \neq \emptyset$ for $t \in [-t_0, t_0]$ follows from assumptions A3 and A4.

For every $t \in [-t_0, t_0]$ and for any optimal Lagrange multiplier $\lambda_1^*, \dots, \lambda_k^*$,

$$\phi(t) \leq \mathcal{L}(\widehat{\mathbf{x}}, \lambda^*, t) \leq \mathcal{F}_0((\boldsymbol{\xi} + t\mathbf{h})^\top \widehat{\mathbf{x}}) + \lambda_j^*(\mathcal{F}_j((\boldsymbol{\xi} + t\mathbf{h})^\top \widehat{\mathbf{x}}) - C_j), \quad j = 1, \dots, k,$$

where $\widehat{\mathbf{x}}$ is such that assumption A2 holds, and the first and second inequalities follow from the definition of a saddle point and from non-negativity of Lagrange multipliers, respectively. This implies that

$$0 \leq \lambda_j^* \leq C = \max_{j \in \{1, \dots, k\}} \sup_{t \in [-t_0, t_0]} \frac{\mathcal{F}_0((\boldsymbol{\xi} + t\mathbf{h})^\top \widehat{\mathbf{x}}) - \phi(t)}{C_j - \mathcal{F}_j((\boldsymbol{\xi} + t\mathbf{h})^\top \widehat{\mathbf{x}})} < \infty,$$

where the last inequality follows from assumption A2. Consequently, $(\mathbb{R}^+)^k$ in (24) and (25) can be replaced by the compact set $[0, C]^k$.

Since $\mathcal{F}_0, \mathcal{F}_1, \dots, \mathcal{F}_k$ are finite and continuous by assumption A3, the corresponding $\mathcal{Y}_1, \dots, \mathcal{Y}_k$ are bounded, convex and closed in $\mathcal{L}^2(\Omega)$, so is the set \mathcal{Y} of all r.v.'s $Y_0^* + \sum_{j=1}^k \lambda_j^* Y_j^*$ with $\lambda_j^* \in [0, C]$ and $Y_j^* \in \mathcal{Y}_j$, $j = 0, \dots, k$. Since $\mathcal{L}^2(\Omega)$ is a reflexive Banach space, this implies that \mathcal{Y} is weakly compact by Kakutani's theorem [13, Theorem 3.31], and since f in (25) and $\partial f / \partial t$ in (26) are continuous functions in the weak topology, this is sufficient for application of Theorem 5 in [38].

The remaining issue is a possible non-compactness of $X_0 \subset \mathbb{R}^n$. Assumptions A3 and A4 imply that for every fixed t , $\mathcal{X}^*(t)$ is a non-empty convex compact set. It remains to show that $\mathcal{X}^*(t)$ is bounded uniformly in t .

PROPOSITION A.1. *There exists $t'_0 > 0$ and $C > 0$ such that $\|\mathbf{x}\|_{\mathbb{R}^n} < C$ for every $t \in [-t'_0, t'_0]$ and $\mathbf{x} \in \mathcal{X}^*(t)$.*

With Proposition A.1, we can apply [38, Theorem 5] to $t \in [-t'_0, t'_0]$ and compact set $\mathcal{X} = \{\mathbf{x} \in X_0 \mid \|\mathbf{x}\|_{\mathbb{R}^n} \leq C\}$. The proof of Proposition A.1 relies on the optimization problem

$$(74) \quad v(t) = \min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, t),$$

where $\mathcal{X} \subseteq \mathbb{R}^n$ and t is a real-valued parameter.

PROPOSITION A.2. *Let $v(t)$ in (74) exist for every t , and let f be continuous in t for every \mathbf{x} . Then $v(t)$ is continuous.*

Proof. For any t_1, t_2 , let $\mathbf{x}_1, \mathbf{x}_2$ be the corresponding solutions of (74). Then $v(t_1) = f(\mathbf{x}_1, t_1) \leq f(\mathbf{x}_2, t_1)$ and $v(t_2) = f(\mathbf{x}_2, t_2) \leq f(\mathbf{x}_1, t_2)$. Consequently, $f(\mathbf{x}_1, t_1) - f(\mathbf{x}_1, t_2) \leq v(t_1) - v(t_2) \leq f(\mathbf{x}_2, t_1) - f(\mathbf{x}_2, t_2)$ and $0 \leq \lim_{t_2 \rightarrow t_1} (f(\mathbf{x}_1, t_1) - f(\mathbf{x}_1, t_2)) \leq \lim_{t_2 \rightarrow t_1} (v(t_1) - v(t_2)) \leq \lim_{t_2 \rightarrow t_1} (f(\mathbf{x}_2, t_1) - f(\mathbf{x}_2, t_2)) = 0$. \square

PROPOSITION A.3. *Let $v(t)$ in (74) with $\mathcal{X} = \mathbb{R}^n$ exist for every t , and let $\mathcal{X}(t)$ be a set of corresponding solutions. Let also $\mathcal{X}(0)$ be bounded, f convex and continuous in \mathbf{x} for every t , and continuous in t for every \mathbf{x} . Then*

$$(75) \quad \lim_{t \rightarrow 0} \left(\sup_{\mathbf{x}(t) \in \mathcal{X}(t)} \inf_{\mathbf{x}(0) \in \mathcal{X}(0)} \|\mathbf{x}(t) - \mathbf{x}(0)\| \right) = 0.$$

Proof. Fix any $\mathbf{x}^* \in \mathcal{X}(0)$. For any $\varepsilon > 0$, the set $\mathcal{X}_\varepsilon(0) = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{x}(0)\| \leq \varepsilon \text{ for some } \mathbf{x}(0) \in \mathcal{X}(0)\}$ is closed, bounded (hence compact) and convex, so that its boundary $\partial \mathcal{X}_\varepsilon(0)$ is also a compact set. Consequently, the minimum in

$$u(t) = \min_{\mathbf{x} \in \partial \mathcal{X}_\varepsilon(0)} f(\mathbf{x}, t)$$

is attained for every t , and $u(t)$ is continuous by Proposition A.2. Since $\partial \mathcal{X}_\varepsilon(0) \cap \mathcal{X}(0) = \emptyset$, $u(0) > v(0) = f(\mathbf{x}^*, 0)$. Consequently, there exists a $t(\varepsilon) > 0$ such that $u(t) > f(\mathbf{x}^*, t)$ for every $t \in (-t(\varepsilon), t(\varepsilon))$. For every $\mathbf{x} \notin \mathcal{X}_\varepsilon(0)$, there exists an $\alpha \in (0, 1)$ such that $\alpha \mathbf{x} + (1 - \alpha)\mathbf{x}^* \in \partial \mathcal{X}_\varepsilon(0)$. Consequently,

$$\alpha(f(\mathbf{x}, t) - f(\mathbf{x}^*, t)) \geq f(\alpha \mathbf{x} + (1 - \alpha)\mathbf{x}^*, t) - f(\mathbf{x}^*, t) \geq u(t) - f(\mathbf{x}^*, t) > 0, \quad t \in (-t(\varepsilon), t(\varepsilon)),$$

so that $\mathbf{x} \notin \mathcal{X}(t)$. Since $\mathbf{x} \notin \mathcal{X}_\varepsilon(0)$ is arbitrary, this implies that $\mathcal{X}(t) \subseteq \mathcal{X}_\varepsilon(0)$, $|t| \leq t(\varepsilon)$, and (75) follows. \square

Now we can prove Proposition A.1. $\mathcal{X}(t)$ is the solution set in

$$(76) \quad \min_{\mathbf{x} \in X_0} \sup_{\lambda \in (\mathbb{R}^+)^k} \mathcal{L}(\mathbf{x}, \lambda, t) = \min_{\mathbf{x} \in X_0} \max_{\lambda \in [0, C]^k} \mathcal{L}(\mathbf{x}, \lambda, t).$$

Since $\mathcal{L}(\mathbf{x}, \lambda, \cdot)$ is continuous, so is $\max_{\lambda \in [0, C]^k} \mathcal{L}(\mathbf{x}, \lambda, \cdot)$. In application to (76), Proposition A.3 yields $\lim_{t \rightarrow 0} \left(\sup_{\mathbf{x}(t) \in \mathcal{X}^*(t)} \inf_{\mathbf{x}(0) \in \mathcal{X}^*(0)} \|\mathbf{x}(t) - \mathbf{x}(0)\| \right) = 0$, and Proposition A.1 follows.

Appendix B. A counterexample.

The following example demonstrates that assumption A3 in Theorem 3.1 is critical.

EXAMPLE B.1. Let $\phi(t)$ be defined by (15) with $k = 0$, $n = 2$, $\mathcal{F}_0(X) = -\inf X$, $X_0 = \{(x_1, x_2) \mid x_1 + x_2 = 1\}$, and with ξ_1, ξ_2, h_1 , and h_2 defined as follows. Let $\Omega = \{\omega_0, \omega_1, \dots\}$ be countable with $\mathbb{P}[\omega_0] = 3/4$ and $\mathbb{P}[\omega_n] = 5^{-n}$, $n \geq 1$. Then

$$\begin{aligned} \xi_1(\omega_0) &= 0, & \xi_1(\omega_n) &= 2^{-n}, & n \geq 1, \\ \xi_2(\omega_0) &= 1, & \xi_2(\omega_n) &= 2^{-n} - 1, & n \geq 1, \end{aligned}$$

$h_1 = 0$, and

$$h_2(\omega_0) = 0, \quad h_2(\omega_n) = 2^n, \quad n \geq 1.$$

In this case, \mathcal{F}_0 is not a continuous basic functional, and $\phi'(0+)$ does not exist.

Detail. The constraint $x_1 + x_2 = 1$ implies that $\mathbf{x} = (x_1, x_2) = (1 - \alpha, \alpha)$, $\alpha \in \mathbb{R}$, and $(\boldsymbol{\xi} + t\mathbf{h})^\top \mathbf{x} = X_\alpha = (1 - \alpha)\xi_1 + \alpha(\xi_2 + th_2)$. Then $-\inf X_\alpha = -\min\{\alpha, \inf_n(\alpha(2^n - 1) + 2^{-n})\}$. For $t = 0$, $\phi(0) = \min_\alpha(-\inf X_\alpha) = 0$, and for $t \in (0, 1/2]$, optimal α is given by $\alpha = \alpha_t^* \equiv \min_{n \leq \log_2 1/t} 2^{-n}/(2 - 2^n t)$. Then $\phi(t) = -\inf(X_{\alpha_t^*}) = -\alpha_t^*$. However, $\lim_{t \rightarrow 0^+} \alpha_t^*$ does not exist: $\limsup_{t \rightarrow 0^+} \alpha_t^* = 9/8$ and $\liminf_{t \rightarrow 0^+} \alpha_t^* = 1$, so that $\phi'(0+)$ does not exist. \square

Appendix C. Direct verification of (54).

$$\begin{aligned} \left. \frac{\partial v(r_0, \mathbf{r} + t\mathbf{h})}{\partial t} \right|_{t=0} &= \left. \frac{\partial}{\partial t} \frac{1}{\sqrt{(\mathbb{E}[\mathbf{r} + t\mathbf{h}] - r_0 \mathbf{e})^\top \Lambda (\mathbf{r} + t\mathbf{h})^{-1} (\mathbb{E}[\mathbf{r} + t\mathbf{h}] - r_0 \mathbf{e})}} \right|_{t=0} \\ &\quad - \frac{1}{2c^{3/2}} \left(\mathbb{E}[\mathbf{h}]^\top \Lambda^{-1} \boldsymbol{\zeta} + \boldsymbol{\zeta}^\top \frac{\partial}{\partial t} \Lambda (\mathbf{r} + t\mathbf{h})^{-1} \right) \Big|_{t=0} \boldsymbol{\zeta} + \boldsymbol{\zeta}^\top \Lambda^{-1} \mathbb{E}[\mathbf{h}], \end{aligned}$$

where $\boldsymbol{\zeta} = \mathbb{E}[\mathbf{r}] - r_0 \mathbf{e}$ and $c = \boldsymbol{\zeta}^\top \Lambda^{-1} \boldsymbol{\zeta}$. Then

$$\begin{aligned} \left. \frac{\partial}{\partial t} \Lambda(\mathbf{r} + t\mathbf{h})^{-1} \right|_{t=0} &= -\Lambda^{-1} \left. \frac{\partial}{\partial t} \Lambda(\mathbf{r} + t\mathbf{h}) \right|_{t=0} \Lambda^{-1} \\ &= -\Lambda^{-1} (\mathbb{E}[\mathbf{r}\mathbf{h}^\top] + \mathbb{E}[\mathbf{h}\mathbf{r}^\top] - \mathbb{E}[\mathbf{r}]\mathbb{E}[\mathbf{h}^\top] - \mathbb{E}[\mathbf{h}]\mathbb{E}[\mathbf{r}^\top]) \Lambda^{-1}. \end{aligned}$$

For any vector \mathbf{a} and matrix B , $\mathbf{a}^\top B \mathbf{a} = (\mathbf{a}^\top B \mathbf{a})^\top = \mathbf{a}^\top B^\top \mathbf{a}$. With $\mathbf{a} = \Lambda^{-1} \boldsymbol{\zeta}$ and $B = \mathbb{E}[\mathbf{r}\mathbf{h}^\top]$ and with $\mathbf{a} = \Lambda^{-1} \boldsymbol{\zeta}$ and $B = \mathbb{E}[\mathbf{r}]\mathbb{E}[\mathbf{h}^\top]$, this identity yields

$$\boldsymbol{\zeta}^\top \left. \frac{\partial}{\partial t} \Lambda(\mathbf{r} + t\mathbf{h})^{-1} \right|_{t=0} \boldsymbol{\zeta} = -2\boldsymbol{\zeta}^\top \Lambda^{-1} (\mathbb{E}[\mathbf{r}\mathbf{h}^\top] - \mathbb{E}[\mathbf{r}]\mathbb{E}[\mathbf{h}^\top]) \Lambda^{-1} \boldsymbol{\zeta}.$$

Since $\mathbb{E}[\mathbf{h}]^\top \Lambda^{-1} \boldsymbol{\zeta} = \boldsymbol{\zeta}^\top \Lambda^{-1} \mathbb{E}[\mathbf{h}]$,

$$\left. \frac{\partial v(r_0, \mathbf{r} + t\mathbf{h})}{\partial t} \right|_{t=0} = -\frac{1}{2c^{3/2}} \left(2\mathbb{E}[\mathbf{h}]^\top \Lambda^{-1} \boldsymbol{\zeta} - 2\boldsymbol{\zeta}^\top \Lambda^{-1} (\mathbb{E}[\mathbf{r}\mathbf{h}^\top] - \mathbb{E}[\mathbf{r}]\mathbb{E}[\mathbf{h}^\top]) \Lambda^{-1} \boldsymbol{\zeta} \right),$$

which simplifies to (54) with $h_0 = 0$.

REFERENCES

- [1] P. ARTZNER, F. DELBAEN, J.-M. EBER, AND D. HEATH, *Coherent measures of risk*, Math. Financ., 9 (1999), pp. 203–227.
- [2] P. ARTZNER, F. DELBAEN, J.-M. EBER, D. HEATH, AND H. KU, *Coherent multiperiod risk adjusted values and bellmans principle*, Ann. Oper. Res., 152 (2007), pp. 5–22.
- [3] W. BAUMOL, *An expected gain-confidence limit criterion for portfolio selection*, Manage. Sci., 10 (1963), pp. 174–182.
- [4] D. BERTSIMAS, D. BROWN, AND C. CARAMANIS, *Theory and applications of robust optimization*, SIAM review, 53 (2011), pp. 464–501.
- [5] M. J. BEST AND R. GRAUER, *On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results*, Review of Financial Studies, 4 (1991), pp. 315–342.
- [6] J. BIAŁAS AND Y. NAKAMURA, *The theorem of weierstrass*, J. Formalized Math., 5 (1996), pp. 353–359.
- [7] F. BONNANS AND A. SHAPIRO, *Perturbation analysis of optimization problems*, Springer Science & Business Media, 2013.
- [8] A. CHERNY, *Weighted $v@r$ and its properties*, Financ. Stoch., 10 (2006), pp. 367–393.
- [9] A. CHERNY AND D. ORLOV, *On two approaches to coherent risk contribution*, Math. Financ., 21 (2011), pp. 557–571.
- [10] J. M. DANSKIN, *The theory of max-min and its application to weapons allocation problems*, Springer-Verlag, New York, 1967.
- [11] F. DELBAEN, *Coherent risk measures on general probability spaces*, in Advances in Financ. Stoch., Springer, 2002, pp. 1–37.
- [12] S. EMMER, M. KRATZ, AND D. TASCHE, *What is the best risk measure in practice? a comparison of standard measures*, J. Risk, 18 (2015), pp. 31–60.
- [13] M. FABIAN, P. HABALA, P. HÁJEK, V. S. J. PELANT, AND V. ZIZLER, *Functional analysis and infinite-dimensional geometry*, Springer Science & Business Media, 2013.
- [14] D. FILIPOVIĆ AND M. KUPPER, *Monotone and cash-invariant convex functions and hulls*, Insurance: Mathematics and Economics, 41 (2007), pp. 1–16.
- [15] H. FÖLLMER AND A. SCHIED, *Stochastic Finance*, de Gruyter, Berlin New York, 3 ed., 2011.
- [16] G. FRANKFURTER, H. PHILLIPS, AND J. SEAGLE, *Portfolio selection: the effects of uncertain means, variances, and covariances*, J. Financ. Quant. Anal., 6 (1971), pp. 1251–1262.
- [17] B. GRECHUK, A. MOLYBOHA, AND M. ZABARANKIN, *Chebyshev inequalities with law-invariant deviation measures*, Probab. Eng. Inform. Sc., 24 (2010), pp. 145–170.
- [18] ———, *Mean-deviation analysis in the theory of choice*, Risk Anal., 32 (2012), pp. 1277–1292.
- [19] ———, *Cooperative games with general deviation measures*, Math. Financ., 23 (2013), pp. 339–365.
- [20] B. GRECHUK AND M. ZABARANKIN, *Optimal risk sharing with general deviation measures*, Ann. Oper. Res., 200 (2012), pp. 9–21.
- [21] ———, *Schur convex functionals: Fatou property and representation*, Math. Financ., 22 (2012), pp. 411–418.
- [22] ———, *Inverse portfolio problem with mean-deviation model*, Eur. J. Oper. Res., 234 (2014), pp. 481–490.

- [23] ———, *Inverse portfolio problem with coherent risk measures*, Eur. J. Oper. Res., 249 (2016), pp. 740–750.
- [24] ———, *Direct data-based decision making under uncertainty*, Eur. J. Oper. Res., submitted (2017).
- [25] V. GUIGUES, *Sensitivity analysis and calibration of the covariance matrix for stable portfolio selection*, Comput. Optim. Appl., 48 (2011), pp. 553–579.
- [26] A. HANTOUTE, M. LÓPEZ, AND C. ZĂLINESCU, *Subdifferential calculus rules in convex analysis: a unifying approach via pointwise supremum functions*, SIAM J. on Optim., 19 (2008), pp. 863–882.
- [27] A. J. HOFFMAN AND W. JACOBS, *Smooth patterns of production*, Manage. Sci., 1 (1954), pp. 86–91.
- [28] R. JAMES, *Weakly compact sets*, Trans. of Amer. Math. Soc., 113 (1964), pp. 129–140.
- [29] J. G. KALLBERG AND W. T. ZIEMBA, *Mis-specifications in portfolio selection problems*, in Risk and capital, Springer, 1984, pp. 74–87.
- [30] B. KALYMON, *Estimation risk in the portfolio selection model*, J. Financ. Quant. Anal., 6 (1971), pp. 559–582.
- [31] H. KANDASAMY, *Portfolio Selection Under Various Risk Measures*, ProQuest, 2008.
- [32] R. KOENKER AND G. B. JR, *Regression quantiles*, Econometrica, (1978), pp. 33–50.
- [33] I. KONDOR, S. PAFKA, AND G. NAGY, *Noise sensitivity of portfolio selection under various risk measures*, J. Bank. Financ., 31 (2007), pp. 1545–1573.
- [34] H. KONNO AND H. YAMAZAKI, *Mean-absolute deviation portfolio optimization model and its applications to tokyo stock market*, Manage. Sci., 37 (1991), pp. 519–531.
- [35] R. MANSINI AND W. O. G. SPERANZA, *Lp solvable models for portfolio optimization: A classification and computational comparison*, IMA J. Manag. Math., 14 (2003), pp. 187–220.
- [36] H. M. MARKOWITZ, *Portfolio selection*, J. Financ., 7 (1952), pp. 77–91.
- [37] ———, *Portfolio selection: Efficient diversification of investments*, Cowles Foundation monograph no. 16, New York: John Wiley & Sons, Inc, 1959.
- [38] P. MILGROM AND I. SEGAL, *Envelope theorems for arbitrary choice sets*, Econometrica, 70 (2002), pp. 583–601.
- [39] W. ORCHARD-HAYS, *Notes on linear programming (part 6): the rand code for the simplex method (sx4)*, Tech. Report 10, RAND Corporation, 1955.
- [40] S. ORTOBELLI, S. RACHEV, S. STOYANOV, F. FABOZZI, AND A. BIGLOVA, *The proper use of risk measures in portfolio theory*, IJTAF, 8 (2005), pp. 1107–1133.
- [41] R. T. ROCKAFELLAR, *Convex Analysis*, Princeton university press, Princeton, NJ, 1970.
- [42] ———, *Conjugate duality and optimization*, SIAM, 1974.
- [43] R. T. ROCKAFELLAR AND S. URYASEV, *Optimization of conditional value-at-risk*, J. Risk, 2 (2000), pp. 21–42.
- [44] ———, *The fundamental risk quadrangle in risk management, optimization and statistical estimation*, Surveys in Oper. Res. and Manage. Sci., 18 (2013), pp. 33–53.
- [45] R. T. ROCKAFELLAR, S. URYASEV, AND M. ZABARANKIN, *Deviation measures in risk analysis and optimization*, Tech. Report 2002-7, ISE Department, University of Florida, 2002.
- [46] ———, *Generalized deviations in risk analysis*, Financ. Stoch., 10 (2006), pp. 51–74.
- [47] ———, *Master funds in portfolio analysis with general deviation measures*, J. Bank. Financ., 30 (2006), pp. 743–778.
- [48] ———, *Optimality conditions in portfolio analysis with general deviation measures*, Math. Program., 108 (2006), pp. 515–540.
- [49] ———, *Equilibrium with investors using a diversity of deviation measures*, J. Bank. Financ., 31 (2007), pp. 3251–3268.
- [50] ———, *Risk tuning with generalized linear regression*, Math. Oper. Res., 33 (2008), pp. 712–729.
- [51] A. RUSZCZYŃSKI, *Nonlinear optimization*, vol. 13, Princeton university press, 2006.
- [52] A. RUSZCZYŃSKI AND A. SHAPIRO, *Optimization of convex risk functions*, Math. Oper. Res., 31 (2006), pp. 433–452.
- [53] Y. SIMAAN, *Estimation risk in portfolio selection: the mean variance model versus the mean absolute deviation model*, Manage. Sci., 43 (1997), pp. 1437–1446.
- [54] K. SYDSAETER, P. HAMMOND, AND A. SEIERSTAD, *Further mathematics for economic analysis*, Pearson Education, 2008.
- [55] A. WILLIAMS, *Marginal values in linear programming*, SIAM J. Appl. Math., 11 (1963), pp. 82–94.
- [56] M. ZABARANKIN AND S. URYASEV, *Statistical Decision Problems: Selected Concepts and Portfolio Safe-guard Case Studies*, Springer, Berlin, 2014.
- [57] J. ZIEGEL, *Coherence and elicibility*, Math. Financ., 26 (2016), pp. 901–918.