Relative deviation metrics with applications in finance

Stoyan V. Stoyanov

FinAnalytica, Inc., USA

e-mail: stoyan.stoyanov@finanalytica.com

Svetlozar T. Rachev *

University of Karlsruhe, Germany and

University of California Santa Barbara, USA

e-mail: rachev@pstat.ucsb.edu

Sergio Ortobelli

University of Bergamo, Italy

e-mail: sergio.ortobelli@unibg.it

Frank J. Fabozzi

Yale University, School of Management

e-mail: frank.fabozzi@yale.edu

*Prof Rachev gratefully acknowledges research support by grants from Division of Mathematical, Life and Physical Sciences, College of Letters and Science, University of California, Santa Barbara, the Deutschen Forschungsgemeinschaft and the Deutscher Akademischer Austausch Dienst.

Relative deviation metrics with applications in finance

Abstract

In the paper, we generalize the classic benchmark tracking problem by introducing the class of relative deviation metrics (r.d. metrics). We consider specific families of r.d. metrics and how they interact with one another. Our approach towards their classification is inspired by the theory of probability metrics — we distinguish between compound, primary and simple r.d. metrics and introduce minimal r.d. metrics. Finally we propose several non-trivial examples of simple r.d. metrics.

1 Introduction

In the last years there has been a heated debate on the "best" measures to use in risk management and portfolio theory. The fundamental work of Artzner et al. (1998) has been the starting point to define the properties that a measure has to satisfy in order to price coherently the risk exposure of a financial position. Many other works have developed these basic concepts introducing several different families of risk measures (see, among others, Föllmer and Schied (2002), Frittelli and Gianin (2002), Rockafellar et al. (2006), Szegö (2004) and the reference therein).

In this paper, we deal with the benchmark tracking-error problem which is a type of an optimal portfolio problem and can also be looked at as an approximation problem. We consider it from a very general viewpoint and replace the tracking-error measure by a general functional satisfying a number of axioms. We call this functional a metric of relative deviation. The axioms are introduced and motivated in Section 2. In Section 3 we establish a relationship between the relative deviation metrics and the deviation measures introduced in Rockafellar et al. (2006).

Our approach to the tracking-error problem is based on the universal methods of the theory of probability metrics. In Section 4, we distinguish between compound, simple and primary relative deviation metrics, the properties of which influence the nature of the approximation problem. Next we introduce a minimal functional establishing an important connection between the compound and the simple classes.

Throughout the paper, we focus on the static tracking-error problem; that is, the metric of relative deviation is defined on the space of joint distributions generated by pairs of real-valued random variables. The only reason is our intention to keep the notation simple and not an inherent limitation of the methodology. Exactly the same approach works in a dynamic setting, if the metrics of relative deviation concern random processes.

2 The tracking error problem

The the minimal tracking-error problem has the following form

$$\min_{w \in \mathcal{W}} \sigma(w'r - r^b)$$

where $w = (w_1, \ldots, w_n)$ is a vector of portfolio weights, $\mathcal{W}$ is a set of admissible vectors w, $r = (r_1, \ldots, r_n)$ is a vector of stocks returns, r^b is the return of a benchmark portfolio, $w'r = \sum_{i=1}^{n} w_i r_i$ is the return of the portfolio in which w_i is the weight of the i-th stock with return r_i, and $\sigma(X)$ stands for the standard deviation of the random variable (r.v.) X. The goal is to find a portfolio which is closest to the benchmark in a certain sense, in this case, determined by the standard deviation. Each feasible vector of weights w defines a portfolio with return $w'r$. Therefore, where appropriate, instead of $\mathcal{W}$ we use $\mathcal{X}$ to denote the feasible set of random variables $w'r$.

In essence, in the minimal tracking error problem we minimize the uncertainty of the excess returns of the feasible portfolios relative to the benchmark portfolio. A serious disadvantage of the tracking error is that it penalizes in the same way the positive and the negative deviations from the mean excess return while our attitude towards them is asymmetric. We are inclined to pay more attention to the negative ones since they represent relative loss. This argument leads to the conclusion that a more realistic measure of "closeness" should be asymmetric.

Our aim is to re-state the minimal tracking-error problem in the more general form

$$\min_{w \in \mathcal{W}} \mu(w'r, r^b) \tag{1}$$

where $\mu(X, Y)$ is a measure of the deviation of X relative to Y. Due to this interpretation, we regard μ as a functional which metrizes[1] relative deviation and we call it *a relative deviation metric* or simply, r.d. metric.

In our definition, the r.d. metric involves real-valued r.v.s. Not every real-valued r.v. can be interpreted as the return of a portfolio. Nevertheless, those r.v.s for which this can be done, constitute a set in this space and any r.d. metric defined on the entire space can be adopted as an r.d. metric in the corresponding set.

What are the properties that μ should satisfy? If the portfolio $w'r$ is an exact copy of the benchmark, i.e. it contains exactly the same stocks in the same amounts, then the relative deviation of $w'r$ to r^b should be zero. The converse should also hold but, generally, could be in a somewhat weaker sense. Otherwise problem (1) is not a "tracking" problem. If the deviation of $w'r$ relative to r^b is zero, then the portfolio and the benchmark are indistinguishable but only in the sense implied by μ. They may, or may

[1]We use the word metrize in a broad sense and not in the sense of metrizing a given topological structure.

not, be identical.

Let us assume for now the strongest version of similarity for this situation, that both portfolios are identical. In case the relative deviation of $w'r$ to r^b is non-zero, then μ is positive. These arguments imply the following properties

P1. $\mu(X, Y) \geq 0$ and $\mu(X, Y) = 0$, if $X \overset{as}{=} Y$, or

$\widetilde{\text{P1}}$. $\mu(X, Y) \geq 0$ and $\mu(X, Y) = 0$, if and only if $X \overset{as}{=} Y$,

where $\overset{as}{=}$ denotes equality in "almost sure" sense. Whenever the tilde sign is used in the paper, it is implied that the property is a stronger alternative.

We already mentioned that in general it is meaningful for μ to be asymmetric but the tracking-error is an example of a symmetric μ. For this reason, at times we may assume symmetry,

P2. $\mu(X, Y) = \mu(Y, X)$ for all X, Y

but in general P2 will not hold.

For the next property, we borrow intuition from the tracking error. Suppose that we introduce a third portfolio, the returns of which we denote by r^c. The following inequality holds, after applying the classical Cauchy-Schwartz inequality, $|\text{cov}(X, Y)| \leq \sigma(X)\sigma(Y)$, for the square integrable X and Y,

$$\begin{aligned}\sigma^2(w'r - r^b) &= \sigma^2(w'r - r^c) + \sigma^2(r^c - r^b) + 2\text{cov}(w'r - r^c, r^c - r^b) \\ &\leq (\sigma^2(w'r - r^c) + \sigma^2(r^c - r^b))^2\end{aligned}$$

and therefore $\sigma(w'r - r^b) \leq \sigma(w'r - r^c) + \sigma(r^c - r^b)$ which is readily recognized as the triangle inequality. According to the theory, equality appears only if

$w'r - r^c$ and $r^c - r^b$ are linearly dependent or, equivalently, if r^c is linearly dependent on $w'r$ and r^b. The triangle inequality is fundamental and shows, in this case, the relation between the tracking-error of the excess returns of any three portfolios. We can easily obtain upper and lower bounds for the third tracking-error from the other two.

Due to its fundamental importance, we assume that the triangle inequality holds for μ,

P3. $\mu(X,Y) \leq \mu(X,Z) + \mu(Z,Y)$ for any X, Y, Z

The triangle inequality and P1 are sufficient to prove that μ is continuous. This is easy to infer from the consequence of P3, $|\mu(X,Y) - \mu(X,Z)| \leq \max(\mu(Z,Y), \mu(Y,Z))$.

Now we are in position to define a few basic terms. Let us denote by $\mathfrak{X}$ the space of r.v.s on a given probability space $(\Omega, \mathfrak{A}, P)$ taking values in $\mathbb{R}$. By $\mathcal{L}\mathfrak{X}_2$ we denote the space of all joint distributions $\mathrm{Pr}_{X,Y}$ generated by the pairs $X, Y \in \mathfrak{X}$.

Definition 1. *Suppose that a mapping $\mu(X,Y) := \mu(Pr_{X,Y})$ is defined on $\mathcal{L}\mathfrak{X}_2$ taking values in the extended interval $[0, \infty]$. If it satisfies properties*

a) $\widetilde{P1}$, P2 and P3, then it is called a probability metric on $\mathfrak{X}$.

b) $\widetilde{P1}$ and P3, then it is called a probability quasi-metric on $\mathfrak{X}$.

c) P1, P2 and P3, then it is called a probability semi-metric on $\mathfrak{X}$.

d) P1 and P3, then it is called a probability quasi-semi-metric on $\mathfrak{X}$.

Suppose that we have the freedom to change the composition of both the main portfolio and the benchmark. Let us select one stock with return r^*, which may or may not be in the portfolios, and re-balance them in such a way that in the end we obtain two new portfolios with returns $(1-a)w'r+ar^*$ and $(1-a)r^b + ar^*$ respectively in which a stands for the weight of the new stock. The re-balancing is easy to implement, we have to sell amounts of the stocks we hold proportionally to the corresponding weights until the cash we have gained equals $100a\%$ of the portfolio present value. Then we invest all the cash in the selected stock. The new portfolios have this stock in common and their relative deviation will decrease if a increases because they become more and more alike. Mathematically, this is guaranteed by property P1 because at the limit the returns are identical in almost sure sense.

An interesting question in this example is if there is any relationship between the relative deviation of $(1-a)w'r + ar^*$ to $(1-a)r^b + ar^*$ and that of $(1-a)w'r$ to $(1-a)r^b$. In the second pair, we have only removed the common stock but we hold the returns of the two portfolios scaled down. The answer is trivial to get for the case of the tracking error from the very definition of it — the tracking error does not change by removing the common stock. In the general case, this relationship cannot be derived from the first three axioms and should be postulated. We consider several versions of it.

If, by removing the common stock, the relative deviation always increases, no matter how the new stock is correlated with the other stocks in the portfolios, then we say it is *strongly regular*. In terms of μ, we obtain

P4. (Strong regularity) $\mu(X+Z, Y+Z) \leq \mu(X, Y)$ for all X, Y, Z

If P4 is satisfied as equality for any Z, then the functional μ is said to be *translation invariant*,

$\widetilde{\text{P4}}$. $\mu(X + Z, Y + Z) = \mu(X, Y)$ for all X, Y, Z

As we noted, the tracking error is translation invariant.

A weaker assumption than P4 may turn out to be reasonable as well. We can postulate that P4 holds not for all X, Y, Z but only for those Z which are independent of X and Y. In the example above, it means that the stock return r^* is independent of $w'r$ and the benchmark return r^b, and therefore r^* does not change the existing portfolios profiles. For instance, if $w'r$ consists of groups of stocks which diversify each other, then the extent of the diversification is going to remain unchanged. Intuitively, with the new stock included in both portfolios, there is no reason to expect an increase in the relative deviation, just the opposite because of the new stock being a common factor. If this property holds, the relative deviation is said to be *weakly regular*.

P4*. (Weak regularity) $\mu(X + Z, Y + Z) \leq \mu(X, Y)$, for all Z independent of X, Y

We will assume that any metric of relative deviation is weakly regular. Clearly, if μ is strongly regular, then it is also weekly regular.

Here we may ask the following related question. What happens if we add to the two initial portfolios other assets, such that their returns become $X + c_1$ and $Y + c_2$, where c_1 and c_2 are arbitrary constants? The relative deviation of $X + c_1$ to $Y + c_2$ should be the same as the one of X to Y because only the location of X and Y changes, P5. $\mu(X+c_1, Y+c_2) = \mu(X,Y)$ for all X, Y and constants c_1, c_2

The above implies that X and $X+c$ are indiscernible for any μ and a constant c. This property allows defining μ on the space of zero-mean r.v.s

$$\mathfrak{X}_0 = \{X \in \mathfrak{X} : EX = 0\}, \tag{2}$$

where $\mathfrak{X}$ is the space of all real-valued r.v.s. and E is the mathematical expectation. Thus for arbitrary real-valued r.v.s the relative deviation of X to Y equals $\mu(X - EX, Y - EX)$ and P5 is automatically satisfied.

Suppose that we add a cash amount c to a portfolio. The portfolio return becomes $(PV_{t_1} - PV_{t_0})/(PV_{t_0} + c)$ which equals $a(PV_{t_1} - PV_{t_0})/PV_{t_0}$ where $a = PV_{t_0}/(PV_{t_0}+c)$, PV_{t_0} is the present value of the portfolio at present time and PV_{t_1} is the random portfolio value at a future time t_1. Therefore the portfolio return appears scaled by the constant a. Now suppose that we add cash amounts to both the main portfolio and the benchmark so that their returns appear scaled down by the same constant a. How does the relative deviation change? If it changes by the same factor raised to some power s, then μ is said to be *positive homogeneous of degree s*,

P6. $\mu(aX, aY) = a^s \mu(X,Y)$ for all $X, Y,\ a, s \geq 0$

Let us revisit the example we considered for the strong regularity property. If μ satisfies P4 and P6, then $\mu((1-a)w'r + ar^*, (1-a)r^b + ar^*) \leq \mu((1-a)w'r, (1-a)r^b) = (1-a)^s \mu(w'r, r^b)$ and therefore the degree of homogeneity determines a limit on the convergence rate to zero as the weight of the new stock approaches one. If $s = 0$, then the rate cannot be determined by this method.

In the benchmark-tracking problem (1), the benchmark is fixed; the weights of the main portfolio are varied until we find the composition with least deviation from the benchmark. If the portfolio contains stocks which are included in the benchmark, then a larger s may indicate higher sensitivity of μ towards the shared stocks; that is, a μ with larger s may identify the common stocks more easily.

Now we are ready to define the r.d. metrics.

Definition 2. *Any quasi-metric μ which satisfies P4*, P5 and P6 is said to be a metric of relative deviation.*

3 Examples of relative deviation metrics

In this section, we deal with translation invariant r.d. metrics and show that there is a one-to-one correspondence with the r.d. metrics generated by the class of deviation measures. The deviation measures are introduced in Rockafellar et al. (2006) and are defined below.

Definition 3. *A deviation measure is any functional $D : \mathfrak{X} \longrightarrow [0, \infty]$ satisfying*

(D1) $D(X + C) = D(X)$ *for all* X *and constants* C*,*

(D2) $D(0) = 0$ *and* $D(\lambda X) = \lambda D(X)$ *for all* X *and all* $\lambda > 0$*,*

(D3) $D(X + Y) \leq D(X) + D(Y)$ *for all* X *and* Y*,*

(D4) $D(X) \geq 0$ *for all* X*, with* $D(X) > 0$ *for non-constant* X*.*

Proposition 1. *The functional $\mu(X,Y) = D(X-Y)$, $X,Y \in \mathfrak{X}_0$, satisfies $\widetilde{P1}$, P3, $\widetilde{P4}$, P5, P6 with $s = 1$ where D is a deviation measure in the sense of Definition 3. Therefore μ is a translation invariant r.d. metric.*

Proof. $\widetilde{\mathrm{P1}}$ follows from D1 and D4. $\widetilde{\mathrm{P4}}$ is trivial. P5 follows from D1 and P6 with $s = 1$ from D2. P3 is easy to show

$$\begin{aligned}\mu(X,Y) = D(X-Y) &= D(X-Z+(Z-Y)) \\ &\leq D(X-Z) + D(Z-Y) = \mu(X,Z) + \mu(Z,Y)\end{aligned}$$

□

In the reverse direction, we can also show that a translation invariant r.d. metric generates a deviation measure.

Proposition 2. *The functional $D(X) = \mu(X - EX, 0)$, $X \in \mathfrak{X}$, is a deviation measure in the sense of Definition 3 if μ is a translation invariant r.d. metric, positively homegeneous of degree one.*

Proof. D1 follows trivially from P5. $D(0) = \mu(0,0) = 0$ and P6 with $s = 1$ guarantees D2. We can easily show that D3 holds making use of the triangle inequality,

$$
\begin{aligned}
D(X+Y) = \mu(X - EX + Y - EY, 0) &= \mu(X - EX, -Y + EY) \\
&\leq \mu(X - EX, 0) + \mu(0, -Y + EY) \\
&= \mu(X - EX, 0) + \mu(Y - EY, 0) \\
&= D(X) + D(Y)
\end{aligned}
$$

Finally, D4 holds due to P1. □

Therefore, we can conclude that all deviation measures in the sense of Definition 3 arise from translation invariant r.d. metrics. Another very simple to establish property is that any translation invariant r.d. metric is convex in any of the two arguments.

Proposition 3. *Any translation invariant r.d. metric μ, for which P6 holds with $s = 1$, satisfies*

$$
\begin{aligned}
\mu(\alpha X + (1-\alpha)Y, Z) &\leq \alpha\mu(X, Z) + (1-\alpha)\mu(Y, Z) \\
\mu(X, \alpha Y + (1-\alpha)Z) &\leq \alpha\mu(X, Y) + (1-\alpha)\mu(X, Z)
\end{aligned} \tag{3}
$$

where $\alpha \in [0, 1]$ and X, Y, Z are any real-valued random variables.

Proof. We demonstrate how to obtain the first inequality, the second follows by a similar argument.

$$
\begin{aligned}
\mu(\alpha X + (1-\alpha)Y, Z) &= \mu(\alpha(X-Z) + (1-\alpha)(Y-Z), 0) \\
&= \mu(\alpha(X-Z), -(1-\alpha)(Y-Z)) \\
&\leq \mu(\alpha(X-Z), 0) + \mu(0, -(1-\alpha)(Y-Z)) \\
&= \mu(\alpha(X-Z), 0) + \mu((1-\alpha)(Y-Z), 0) \\
&= \alpha\mu(X-Z, 0) + (1-\alpha)\mu(Y-Z, 0) \\
&= \alpha\mu(X, Z) + (1-\alpha)\mu(Y, Z)
\end{aligned}
$$

The first two equalities hold because of $\widetilde{\mathrm{P4}}$. The inequality follows from P3 and then we make use of $\widetilde{\mathrm{P4}}$ and P6. $\square$

The convexity property (3) guarantees that problem (1) is a convex optimization problem if μ satisfies the assumptions in the proposition and the feasible set $\mathcal{X}$ is a convex set in the sense that if $X, Y \in \mathcal{X}$, then $\alpha X + (1-\alpha)Y \in \mathcal{X}$, $\alpha \in [0,1]$. Both inequalities in (3) are due to the translation invariance property. Therefore, we have to impose the convexity properties if we relax $\widetilde{\mathrm{P4}}$ to P4 or P4* in order for μ to be convex.

Definition 4. *Any r.d. metric which satisfies any of the inequalities (3) is called a convex r.d. metric.*

At this point we have shown that the translation invariant r.d. metrics are equivalent to the metrics spawned by the deviation measures and are well-positioned for the benchmark tracking problem (1). Beside this class, there are other examples of r.d. metrics suitable for (1). The next section is devoted to them.

In the remaining part of this section, we provide several examples of r.d. metrics generated from deviation measures. We consider firstly the semi-standard deviation and, secondly, the value-at-risk and the conditional vlaue-at-risk which are, generally, risk measures but if restricted to $\mathfrak{X}_0$ they become deviation measures.

Semi-standard deviation

Suppose that D is the semi-standard deviation

$$D(X) = \sigma_-(X) = (E(X - EX)_-^2)^{1/2}$$

where $X_- = \max(0, -X)$. Thus $\mu(w'r, r^b) = (E(w'r_0 - r_0^b)_-^2)^{1/2}$ in which $w'r_0$ and r_0^b denote centered returns is the corresponding r.d. metric. In terms of the benchmark tracking problem, we minimize the negative deviations of $w'r_0$ relative to r_0^b.

Value-at-Risk

The Value-at-Risk (VaR) is defined as $VaR_\alpha(X) = -\inf\{x \in \mathbb{R} : P(X \leq x) \geq \alpha\}$ and α is the tail probability. If considered on on $\mathfrak{X}_0$, it turns into a deviation measure, see Rockafellar et al. (2006). Therefore the functional $\mu(w'r, r^b) = VaR_\alpha(w'r_0 - r_0^b)$ is the corresponding r.d. metric. It measures a given quantile of the centered excess return. By minimizing it, we would like to find a portfolio closest to the benchmark with an emphasis on the corresponding quantile.

Conditional Value-at-Risk

The conditional value-at-risk (CVaR) is defined as,

$$CVaR_\alpha(X) = \frac{1}{\alpha}\int_0^\alpha VaR_p(X)dp$$

where $VaR_p(X)$ is the Value-at-Risk. The CVaR is a coherent risk measure but if considered on $\mathfrak{X}_0$, it becomes a deviation measure, see Rockafellar et al. (2006). The corresponding r.d. metric is $\mu(w'r, r^b) = CVaR_\alpha(w'r_0 - r_0^b)$ and measures the average VaR of the centered excess return smaller than $VaR_\alpha(X)$.

4 Relation to probability metrics — representation theorems

In this section we examine the relationship between r.d. metrics and probability metrics (p. metrics). We have already defined the term *probability metric* in Definition 1. The basic differences are that r.d. metrics are asymmetric (P2 does not hold) and property P5. Therefore if we consider a p. metric defined on $\mathfrak{X}_0$, the asymmetry remains the only difference. Any quasi-semi-metric generates a p. semi-metric via symmetrization. If μ is a quasi-metric, then $\overline{\mu}(X,Y) = (\mu(X,Y)+\mu(Y,X))/2$ is symmetric and also satisfies P1 and P3. Hence $\overline{\mu}(X,Y)$ is a p. metric. As a result, any r.d. metric generates a p. metric via symmetrization. Due to the close connection, all classifications of p. metrics remain valid for r.d. metrics.

In the theory of p. metrics, we distinguish between *primary*, *simple*, and *compound* metrics (for more information about p. metrics, see Rachev (1991)). We obtain the same classes of r.d. metrics by defining primary, simple, and compound quasi-semi-metrics. In order to introduce primary p. quasi-semi-metrics, we need some additional notation.

Let h be a mapping defined on $\mathfrak{X}$ with values in $\mathbb{R}^J$, that is we associate a vector of numbers with a random variable. The vector of numbers could be interpreted as a set of some characteristics of the random variable. An example of such a mapping is: $X \to (EX, \sigma(X))$ where the first element is the mathematical expectation and the second is the standard deviation. Furthermore, the mapping h induces a partition of $\mathfrak{X}$ into classes of equivalence. That is, two random variables X and Y are regarded as equivalent, $X \sim Y$, if their corresponding characteristics agree,

$$X \sim Y \quad \Longleftrightarrow \quad h(X) = h(Y)$$

Since the p. metric is defined on the space of pairs of r.v.s $\mathcal{L}\mathfrak{X}_2$, we have to translate the equivalence into the case of pairs of r.v.s. Two sets of pairs (X_1, Y_1) and (X_2, Y_2) are said to be equivalent if there is equivalence on an element-by-element basis, i.e. $h(X_1) = h(X_2)$ and $h(Y_1) = h(Y_2)$.

Definition 5. *Let μ be a probability quasi-semi-metric such that μ is constant on the equivalence classes induced by the mapping h:*

$$(X_1, Y_1) \sim (X_2, Y_2) \quad \Longrightarrow \quad \mu(X_1, Y_1) = \mu(X_2, Y_2)$$

Then μ is called primary probability quasi-semi-metric. If the converse implication ($\Longleftarrow$) also holds, then μ is said to be a primary quasi-metric.

Here is an illustration. Assume that h maps the r.v.s to their first absolute moments, $h(X) = E|X|$. Thus $(X_1, Y_1) \sim (X_2, Y_2)$ means that X_1 and X_2 on one hand, and Y_1 and Y_2 on the other, have equal first absolute moments, i.e. $(E|X_1|, E|Y_1|) = (E|X_2|, E|Y_2|)$. In this situation, a primary quasi-semi-metric would measure the two distances, $\mu(X_1, Y_1)$ and $\mu(X_2, Y_2)$, as equal. The word *quasi* reminds one that $\mu(X_1, Y_1)$ may not equal $\mu(Y_1, X_1)$. Nevertheless $\mu(Y_1, X_1) = \mu(Y_2, X_2)$. Moreover, if $h(X_1) = h(Y_1)$, then necessarily $\mu(X_1, Y_1) = 0$. The word *semi* signifies that $\mu(X_1, Y_1) = 0$ may not induce equality among the corresponding characteristics.

These considerations show that a primary quasi-semi-metric in $\mathfrak{X}$ generates a quasi-semi-metric $\mu_1(h(X), h(Y)) = \mu(X, Y)$ in the space of the corresponding characteristics $h(\mathfrak{X}) = \{h(X) \in \mathbb{R}^J,\ X \in \mathfrak{X}\} \subseteq \mathbb{R}^J$. For example, consider our portfolio with return $w'r$ and the benchmark with return r^b and choose two characteristics — the expected return and the standard deviation. The functional

$$\mu_1(w'r, r^b) = |Ew'r - Er^b| + |\sigma(w'r) - \sigma(r^b)|$$

is a primary probability metric. If $\mu_1(w'r, r^b) = 0$, then $Ew'r = Er^b$ and $\sigma(w'r) = \sigma(r^b)$ but it is not clear if the two portfolios have the same distribution or if they are indistinguishable in almost sure sense. Now let us consider the functional

$$\mu_2(w'r, r^b) = (Ew'r - Er^b)_- + (\sigma(w'r) - \sigma(r^b))_+$$

where $(x)_- = \max(-x, 0)$ and $(x)_+ = \max(x, 0)$. Assume that the two characteristics of the portfolio returns agree, then $\mu_2(w'r, r^b) = 0$. In contrast to μ_1, the condition $\mu_2(w'r, r^b) = 0$ implies $Ew'r \geq Er^b$ and $\sigma(w'r) \leq \sigma(r^b)$, which is an order relation in the space of characteristics.

In similar vein, we can define primary metrics involving any number of moments or other characteristics such as quantiles, conditional expectations, etc. We proceed with the definition of the simple metrics.

Definition 6. *A probability quasi-semi-metric is said to be simple if for each* $(X, Y) \in \mathcal{LX}_2$*,*

$$P(X \leq t) = P(Y \leq t), \ \forall t \in \mathbb{R} \quad \Longrightarrow \quad \mu(X, Y) = 0$$

If the converse implication ($\Longleftarrow$) *also holds, then* μ *is said to be a simple quasi-metric.*

Examples of simple quasi-metrics include the celebrated Kolmogorov metric

$$\rho(X, Y) = \sup_{t \in \mathbb{R}} |F_X(t) - F_Y(t)| \tag{4}$$

and the Kantorovich metric

$$\kappa(X, Y) = \int_{-\infty}^{\infty} |F_X(t) - F_Y(t)| dt. \tag{5}$$

Clearly, if the two returns of the two portfolios in the example above have equal distributions, then, for instance, $\rho(w'r, r^b) = 0$. Conversely, if $\rho(w'r, r^b) = 0$, then $w'r$ and r^b are equal in distribution, which also means that their characteristics, such as moments, quantiles, etc. agree if they are finite.

Equality in distribution does not indicate if the random variables coincide almost surely, in case they are defined on a common probability space. We proceed with the definition of the most general notion of compound metrics.

Definition 7. *Any probability quasi-semi-metric in the sense of Definition 1 is said to be compound.*

Consider, for instance, the average compound metric

$$\mathcal{L}_p(X,Y) = (E(|X-Y|)^p)^{1/p} \tag{6}$$

where $p \in [1, \infty)$ and the the Birnbaum-Orlicz compound metric

$$\Theta_p(X,Y) = \left[\int_{-\infty}^{\infty} (\tau(t;X,Y))^p dt\right]^{1/p} \tag{7}$$

where $p \geq 1$, $\tau(t;X,Y) = P(X \leq t < Y) + P(Y \leq t < X)$. If the two random variables are equal in almost sure sense, then both (6) and (7) become equal to zero. The converse is also true. The limit case $\mathcal{L}_\infty(X,Y) = \inf\{\epsilon > 0 : P|X-Y| > \epsilon) < \epsilon\}$ is known as the Ky-Fan p. metric.

In summary, $\mu(X,Y) = 0$ implies that (i) $X \overset{as}{=} Y$ if μ is a compound quasi-metric, (ii) $F_X(t) = F_Y(t)$, $\forall t \in \mathbb{R}$ if μ is a simple quasi-metric, and (iii) $h(X) = h(Y)$ if μ is a primary quasi-metric.

In the discussion of P1, we noted that we would assume the strong property that $\mu(X,Y) = 0$ implies $X \overset{as}{=} Y$. This may be too demanding to measure relative deviation. From a practical viewpoint, it may be more appropriate to consider a r.d. metric such that $\mu(X,Y) = 0$ implies only equality in distribution, or equality of some characteristics of X and Y, without the requirement that X and Y should coincide almost surely. For such cases, simple and primary r.d. metrics are better positioned.

In order to obtain an r.d. metric, we have to check the additional properties of weak regularity, P4*, and positive homogeneity, P5. Weak regularity is most difficult to verify directly. Therefore, we can take advantage of the notion of the ideal p. metric.

Definition 8. *A p. metric is called ideal of order* $p \in \mathbb{R}$ *if for any r.v.s* $X, Y, Z \in \mathfrak{X}$ *and any non-zero constant* c *the next two properties are satisfied*

(a) *Regularity (strong or weak):* $\mu(X + Z, Y + Z) \leq \mu(X, Y)$

(b) *Homogeneity:* $\mu(cX, cY) \leq |c|^s \mu(X, Y)$

If μ *is a simple metric, then* Z *is assumed to be independent of* X *and* Y*, that is, weak regularity holds. In this case,* μ *is said to be weakly-perfect of order* s*.*

Due to (a) and (b), any ideal metric defined on $\mathfrak{X}_0$ turns into a symmetric r.d. metric.

We can obtain a quasi-metric from a p. metric by breaking the symmetry axiom while keeping the triangle inequality. Let us take as an example the average compound metric defined in (6). Then one way to break the symmetry in this case is to consider

$$\mathcal{L}_p^*(X, Y) = (E(\max(X - Y, 0))^p)^{1/p} \tag{8}$$

A special limit case here is $\mathcal{L}_\infty^*(X, Y) = \inf\{\epsilon > 0 : P(\max(X - Y, 0) > \epsilon) = 0\}$ which is the quasi-metric corresponding to the limit case of the Ky-Fan

p. metric $\mathcal{L}_\infty(X,Y) = \inf\{\epsilon > 0 : P|X-Y| > \epsilon) = 0\}$. Still there is a deviation measure behind $\mathcal{L}^*_p(X,Y)$ because it is translation invariant. Our main goal will be to develop examples of r.d. metrics which are asymmetric and are not generated from a deviation measure.

Proposition 4. *The functional defined in equation (8) is a compound quasi-semi-metric in $\mathfrak{X}$ and a compound quasi-metric in $\mathfrak{X}_0$. It is translation invariant and homogeneous of degree $s = 1$.*

Proof. Let us first consider (8) in $\mathfrak{X}$. P1 is trivial. The triangle inequality follows from the sub-additivity of the max function and Minkowski's inequality.

Next we show that (8) is a quasi-metric in $\mathfrak{X}_0$. Suppose that $\mathcal{L}^*_p(X,Y)$ is defined on $\mathfrak{X}_0$ and, additionally, that $\mathcal{L}^*_p(X,Y) = 0$. From the definition, we easily observe that the last assumption implies $Y \geq X$. Strict inequality is easy to rule out because if $Y > X$, then $EY > EX$ and this is impossible by construction since $EX = EY = 0$. Therefore $Y \overset{as}{=} X$ and hence $\mathcal{L}^*_p$ is a quasi-metric in $\mathfrak{X}_0$, i.e. $\widetilde{\text{P1}}$ holds.

Both translation invariance and the homogeneity degree are obvious. $\square$

As a second example, let us examine the Birnbaum-Orlicz compound metric defined in (7). In this case, it is easier to obtain a quasi-metric. Consider the functional

$$\Theta^*_p(X,Y) = \left[\int_\infty^\infty (\tau^*(t;X,Y))^p dt\right]^{1/p} \tag{9}$$

where $\tau^*(t;X,Y) = P(Y \leq t < X)$ and $p \geq 1$.

Proposition 5. *The functional defined in equation (9) is a compound quasi-semi-metric in $\mathfrak{X}$ and a compound quasi-metric in $\mathfrak{X}_0$. It satisfies the weak regularity property P4* and is homogeneous of degree $s = 1/p$.*

Proof. In order to prove that (9) is a quasi-semi-metric in $\mathfrak{X}$, we need to verify P1 and P3.

P1. It is trivial since $P(X \leq t < X) = 0$ for all $t \in \mathbb{R}$.

P3. We start by decomposing the function $P(Y \leq t < X)$, t is fixed.

$$
\begin{aligned}
P(Y \leq t < X) &= P\left(\{Y \leq t\} \cap \{X > t\}\right) \\
&= P\left(\{Y \leq t\} \cap \{X > t\} \cap \{\{Z \leq t\} \cup \{Z > t\}\}\right) \\
&= P\left(\{Y \leq t\} \cap \{X > t\} \cap \{Z \leq t\}\right) \\
&\qquad + P\left(\{Y \leq t\} \cap \{X > t\} \cap \{Z > t\}\right) \\
&\leq P\left(\{X > t\} \cap \{Z \leq t\}\right) + P\left(\{Y \leq t\} \cap \{Z > t\}\right) \\
&= P(Z \leq t < X) + P(Y \leq t < Z)
\end{aligned}
$$

The third equality holds because the corresponding events have empty intersection. The inequality appears because we ignore events which, generally, have probability less than one. In effect, by Minkowski's inequality,

$$
\begin{aligned}
\Theta_p^*(X,Y) &= \left[\int_{\infty}^{\infty}(P(Y \le t < X))^p dt\right]^{1/p} \\
&\le \left[\int_{\infty}^{\infty}(P(Z \le t < X) + P(Y \le t < Z))^p dt\right]^{1/p} \\
&= \left[\int_{\infty}^{\infty}(P(Z \le t < X))^p dt\right]^{1/p} + \left[\int_{\infty}^{\infty}(P(Y \le t < Z))^p dt\right]^{1/p} \\
&= \Theta_p^*(X,Z) + \Theta_p^*(Z,Y)
\end{aligned}
$$

and we receive the triangle inequality.

It is simple to verify that $\widetilde{\text{P1}}$ holds in $\mathfrak{X}_0$. Similarly to Proposition 4, suppose that Θ_p^* is defined on $\mathfrak{X}_0$. If $\Theta_p^*(X,Y) = 0$, then $\tau^*(t; X, Y) = 0$, $\forall t \in \mathbb{R}$. Since $P(Y \le t < X) = 0$ for any t implies $Y \ge X$, by the same argument, as Proposition 4, we obtain that Θ_p^* is a quasi-metric in $\mathfrak{X}_0$ and a quasi-semi-metric in $\mathfrak{X}$. The weak regularity property P4* is checked by applying Young's convolution inequality. The homogeneity order is verified directly by change of variables.

□

The functional in (9) is an example of a r.d. metric which is not generated from a deviation measure. A special limit example of (5) is $\Theta_\infty^*(X,Y) = \sup_{t\in\mathbb{R}} P(Y \le t < X)$ which is an asymmetric version of the compound p. metric $\Theta_\infty(X,Y) = \sup_{t\in\mathbb{R}} \tau(t; X, Y)$ generating as a minimal metric the celebrated Kolmogorov metric in the space of distribution functions.

One might be tempted to surmise that a structured approach towards the generation of classes of r.d. metrics is through asymmetrization of classes of ideal p. metrics, to make them *quasi*-metrics, and using them on the sub-space $\mathfrak{X}_0$. The functionals in (8) and (9) support such a generalization. For instance, asymmetrization of (8) with the max function turns the p. metric into a p. quasi-semi-metric on $\mathfrak{X}$ and considering it on the sub-space $\mathfrak{X}_0$ turns it into a quasi-metric. In spite of these two examples, in the general case, such an approach would be incorrect because it would sometimes lead to a quasi-semi-metric even on $\mathfrak{X}_0$.

4.1 Minimal quasi-semi-metrics

The minimal quasi-metric is a type of simple quasi-metric obtained from a compound quasi-metric in a special way. Thus certain properties valid for the compound quasi-metric are inherited by the generated minimal quasi-metric.

In the theory of probability metrics, the p. minimal metrics are defined in the following way.

Definition 9. *For a given p. metric μ on $\mathcal{L}\mathfrak{X}_2$, the functional $\widehat{\mu}$ defined by the equality*

$$\widehat{\mu}(X, Y) = \inf\{\mu(\widetilde{X}, \widetilde{Y}) : \widetilde{X} \stackrel{d}{=} X, \widetilde{Y} \stackrel{d}{=} Y\} \tag{10}$$

where $\stackrel{d}{=}$ means equality in distribution, is said to be minimal p. metric (w.r.t. μ).

Many simple p. metrics arise as minimal p. metrics with respect to a compound p. metric. For example, the Kolmogorov p. metric (4) is the minimal metric w.r.t. the Θ_∞ compound metric defined in (7) and the Kantorovich p. metric (5) is minimal w.r.t. $\mathcal{L}_1$ defined in (6). See Rachev (1991) for other examples.

We apply (10) to obtain functionals minimal w.r.t. quasi-semi-metrics because our main goal is to derive classes simple r.d. metrics from compound r.d. metrics. Suppose that we use a simple r.d. metric in the benchmark tracking problem (1). Then the optimization can be equivalently re-stated as

$$\min_{w \in \mathcal{W}} \mu(F_{w'r}, F_{r^b}) \tag{11}$$

in which $F_{w'r}$ denotes the c.d.f. of the portfolio return $w'r$ and F_{r^b} is the c.d.f. of the benchmark returns because the simple r.d. metric measures the relative deviation of the c.d.f. of the portfolio return to the benchmark return. That is, by varying the portfolio weights w, we aim at getting the composition of stocks such that the c.d.f. of the resulting portfolio return least deviates from the c.d.f. of the benchmark return. To this end, we have to be able to compute $F_{w'r}$ for any feasible vector of weights w. Since the portfolio return is a linear combination of the equity returns it is composed of, we can calculate $F_{w'r}$ for any w if we have a hypothesis not only for the individual c.d.f.s of the equity returns but also for the dependence between them. Therefore, we need a hypothesis for the multivariate c.d.f. of the entire vector of stock returns r.

Note that this is not an extraordinary requirement. For example, in the classical tracking-error problem, we need the covariance matrix. This is certainly not an assumption for a multivariate c.d.f. but we know that the tracking-error problem is consistent only with the multivariate normal distribution and therefore even if there is not an explicit assumption, implicitly it exists.

Throughout the paper, we regard the joint distribution of the stock returns in r as available. We will also assume that the following non-restrictive continuity property holds.

CP. Assume that the pairs $(X_n, Y_n) \in \mathcal{LX}_2,\ n \in \mathbb{N}$ converge in distribution to $(X, Y) \in \mathcal{LX}_2$ and that $\mu(X_n, Y_n) \to 0$ as $n \to \infty$. Then $\mu(X, Y) = 0$.

All examples of μ that we will consider satisfy this property. We state it only for technical reasons. In the next lemma, we establish the properties that the minimal functional in (10) satisfies.

Lemma 1. *Suppose that μ is a compound quasi-semi-metric. Then $\widehat{\mu}$ defined in (10) is a simple quasi-semi-metric. Moreover, if μ is a compound quasi-metric satisfying the CP condition, then $\widehat{\mu}$ is a simple quasi-metric. Finally, if μ satisfies any of the inequalities*

$$\mu(\alpha X + (1-\alpha)Y, Z) \leq \alpha\mu(X, Z) + (1-\alpha)\mu(Y, Z)$$
$$\mu(X, \alpha Y + (1-\alpha)Z) \leq \alpha\mu(X, Y) + (1-\alpha)\mu(X, Z)$$

then so does $\widehat{\mu}$ on condition that the bivariate law in the convex combination, either (X, Y) or (Y, Z), is known.

Proof. We make use of the ideas behind the proof that the minimal functional (10) is a p. semi-metric if μ is a semi-metric, see Rachev (1991). Repeating the arguments there we see that P1, $(\widetilde{\text{P1}})$, P3 hold for $\widehat{\mu}$.

The verification of convexity is very similar to checking P3. We prove that $\widehat{\mu}$ is convex in the first argument; the same reasoning can be used to prove the other inequality. Let $X_1, X_2, Z \in \mathfrak{X}_0$ and $\alpha \in [0,1]$ and assume that we know the bivariate law (X_1, X_2); this way we know the r.v. $X_\alpha := \alpha X_1 + (1-\alpha)X_2,\ X_\alpha \in \mathfrak{X}_0$. We also assume the technical condition that the underlying probability space $(\Omega, \mathfrak{A}, P)$ is rich enough. Hence, for any $\epsilon > 0$, we can choose bivariate laws $(\widetilde{X}_1, \widetilde{Z})$, $(\widetilde{X}_2, \widetilde{Z})$ and $(\widetilde{X}_\alpha, \widetilde{Z})$ such that $\widetilde{X}_1 \stackrel{d}{=} X_1$, $\widetilde{X}_2 \stackrel{d}{=} X_2$, $\widetilde{Z} \stackrel{d}{=} Z$, $\widetilde{X}_\alpha \stackrel{as}{=} \alpha\widetilde{X}_1 + (1-\alpha)\widetilde{X}_2 \stackrel{d}{=} X_\alpha$, $\widehat{\mu}(X_1, Z) + \epsilon \geq \mu(\widetilde{X}_1, \widetilde{Z})$, and $\widehat{\mu}(X_2, Z) + \epsilon \geq \mu(\widetilde{X}_2, \widetilde{Z})$. By assumption μ is convex and therefore $\mu(\widetilde{X}_\alpha, \widetilde{Z}) \leq \alpha\mu(\widetilde{X}_1, \widetilde{Z}) + (1-\alpha)\mu(\widetilde{X}_2, \widetilde{Z})$ holds. Hence

$$\begin{aligned}
\widehat{\mu}(X_\alpha, Z) &\leq \mu(\widetilde{X}_\alpha, \widetilde{Z}) \leq \alpha\mu(\widetilde{X}_1, \widetilde{Z}) + (1-\alpha)\mu(\widetilde{X}_2, \widetilde{Z}) \\
&\leq \alpha(\widehat{\mu}(X_1, Z) + \epsilon) + (1-\alpha)(\widehat{\mu}(X_2, Z) + \epsilon) \\
&= \alpha\widehat{\mu}(X_1, Z) + (1-\alpha)\widehat{\mu}(X_2, Z) + \epsilon
\end{aligned}$$

Since we minimize over all possible ways to couple X_α and Z, we implicitly assume that the distribution of X_α is known and we hold it fixed in this calculation.

The chain of inequalities above is true for any $\epsilon > 0$, therefore letting $\epsilon \to 0$ we prove the convexity of $\widehat{\mu}$. □

The result contained in Lemma 1 will be used to derive classes of r.d. metrics consistent with the notion of convergence in distribution. First we have to check whether it is true that if μ is a r.d. metric, then the minimal quasi-metric is also a r.d. metric.

Corollary 1. *If μ is a compound r.d. metric, then $\widehat{\mu}$ defined in (10) is a simple r.d. metric.*

Proof. By assumption, μ satisfies P1, P3, P4 or P4*, P5, P6. From Lemma 1, we know that $\widehat{\mu}$ satisfies P1 and P3, provided that the continuity condition holds. It remains to check P4-P6. P4 or P4* follows from the argument which we used to prove convexity, P5 holds trivially since μ (and therefore $\widehat{\mu}$) is defined in $\mathfrak{X}_0$. P6 is easy to check.

$$\begin{aligned}
\widehat{\mu}(aX, aY) &= \inf\{\mu(\widetilde{X}, \widetilde{Y}) : \widetilde{X} \stackrel{d}{=} aX, \widetilde{Y} \stackrel{d}{=} aY\} \\
&= \inf\{a^s \mu(\widetilde{X}/a, \widetilde{Y}/a) : \widetilde{X}/a \stackrel{d}{=} X, \widetilde{Y}/a \stackrel{d}{=} Y\} \\
&= a^s \inf\{\mu(\widetilde{X}/a, \widetilde{Y}/a) : \widetilde{X}/a \stackrel{d}{=} X, \widetilde{Y}/a \stackrel{d}{=} Y\} \\
&= a^s \widehat{\mu}(X, Y)
\end{aligned}$$

□

We observe the following relationships between the introduced classes of r.d. metrics:

a) *compound translation invariant r.d. metrics $\subset$ compound convex r.d. metrics $\subset$ compound r.d. metrics $\subset$ compound quasi-metrics $\subset$ compound quasi-semi-metrics.*

b) *simple convex r.d. metrics $\subset$ compound convex r.d. metrics*

c) *simple translation invariant r.d. metrics $\not\subset$ compound translation invariant r.d. metrics*

d) *ideal p. metrics* $\subset$ *r.d. metrics*

e) *primary quasi-semi-metrics* $\subset$ *simple quasi-semi-metrics* $\subset$ *compound quasi-semi-metrics*

Item c) appears because generally $\widetilde{\mathrm{P4}}$ may fail to hold for the minimal metric. Item e) is due to the general relationship between the corresponding p. metrics.

Apart from providing an interesting link between the classes of compound and simple quasi-semi-metrics, the minimal metrics can be used to construct simple quasi-semi-metrics with suitable properties. For example, it is easier to establish the regularity property, or the convexity property, for a compound metric using the method of one probability space. These properties are inherited by the corresponding simple minimal metric. In this fashion, taking advantage of Corollary 1, we can construct simple r.d. metrics.

Under some conditions, $\widehat{\mu}$ can be explicitly computed. We can obtain explicit representations through the Cambanis-Simons-Stout theorem, see Cambanis et al. (1976). The basic results are contained in the next theorem. We adopt the notation $\mu_\phi(X, Y) := E\phi(X, Y)$ where ϕ is a function.

Theorem 1. *Given $X, Y \in \mathfrak{X}$ with finite moment $\int_{\mathbb{R}} \phi(x, a)dF(x)$, $a \in \mathbb{R}$ where $\phi(x, y)$ is a quasi-antitone function, i.e.*

$$\phi(x, y) + \phi(x', y') \leq \phi(x', y) + \phi(x, y')$$

for any $x' > x$ and $y' > y$, then

$$\widehat{\mu}_\phi(X,Y) = \int_0^1 \phi(F_X^{-1}(t), F_Y^{-1}(t))dt$$

where $F_X^{-1}(t) = \inf\{x : F_X(x) \geq t\}$ *is the generalized inverse of the c.d.f.* $F_X(x)$ *and also* $\widehat{\mu}_\phi(X,Y) = \mu_\phi(F_X^{-1}(U), F_Y^{-1}(U))$ *where* U *is a uniformly distributed r.v. on* $(0,1)$.

By virtue of Corollary 1, $\widehat{\mu}_\phi$ is a r.d. metric if μ_ϕ is a r.d. metric and clearly it depends only on the distribution functions of X and Y. The function ϕ should be such that $\phi(x,x) = 0$ but generally it may not be symmetric, $\phi(x,y) \neq \phi(y,x)$. Examples of ϕ include $f(x-y)$ where f is a non-negative convex function in $\mathbb{R}$. In particular, one might choose

$$\phi(x,y) = H_1(\max(x-y,0)) + H_2(\max(y-x,0))$$

where $H_1, H_2 : [0,\infty) \to [0,\infty)$ are convex, non-decreasing functions. If $H_1(t) = t$ and $H_2(t) = 0$, $\phi^*(x,y) = \max(x-y,0)$, then

$$\widehat{\mu}_{\phi^*}(X,Y) = \int_0^1 \max(F_X^{-1}(t) - F_Y^{-1}(t), 0)dt \tag{12}$$

We can see that (12) is the minimal quasi-semi-metric of (8) with $p=1$. If (12) is defined on the space of zero-mean random variables, $X,\ Y \in \mathfrak{X}_0$, then $\mu_{\phi^*}(X,Y)$ is a quasi-metric and, therefore, according to Corollary 1, (12) is a r.d. metric, see Proposition 4. Without this restriction, if $X,\ Y \in \mathfrak{X}$, then $\widehat{\mu}_{\phi^*}(X,Y)$ is a quasi-semi-metric.

Similarly, we obtain the minimal metrics of $\mathcal{L}_p^*$ defined in (8),

$$l_p^*(X,Y) = \widehat{\mathcal{L}}_p^*(X,Y) = \left[\int_0^1 (\max(F_X^{-1}(t) - F_Y^{-1}(t), 0))^p dt\right]^{1/p}.$$

Furthermore, Proposition 4 shows that $\mathcal{L}_p^*$ is a translation invariant r.d. metric and therefore it is convex. As a result, according to Lemma 1, l_p^* is a simple convex r.d. metric.

Other explicit forms can be computed also for the family Θ_p^* defined in (9) through the Frechet-Hoeffding inequality.

Lemma 2. *Suppose that $X, Y \in \mathfrak{X}$ and $H : [0,\infty] \to [0,\infty]$ is non-decreasing. If*

$$\Theta_H^*(X,Y) = \int_{-\infty}^{\infty} H(P(Y \le t < X))dt$$

then

$$\widehat{\Theta}_H^*(X,Y) = \int_{-\infty}^{\infty} H(\max(F_Y(t) - F_X(t), 0))dt$$

Proof. The demonstration is straightforward.

$$\begin{aligned}
\mu(X,Y) &= \int_{-\infty}^{\infty} H(P(Y \le t < X))dt \\
&= \int_{-\infty}^{\infty} H(P(Y \le t) - P(Y \le t, X \le t))dt \\
&\ge \int_{-\infty}^{\infty} H(F_Y(t) - \min(F_X(t), F_Y(t)))dt \\
&= \int_{-\infty}^{\infty} H(\max(F_Y(t) - F_X(t), 0))dt
\end{aligned}$$

The inequality is application of the celebrated Frechet-Hoeffding upper bound $\min(F_X(x), F_Y(y)) \geq P(X \leq x, Y \leq y)$. See Rachev (1991) p. 153, 154 for more general results. $\square$

Corollary 2. *Choosing $H = t^p$, $p \geq 1$, we receive*

$$\theta^*_p(X,Y) = \widehat{\Theta}^*_p(X,Y) = \left[\int_{-\infty}^{\infty} (\max(F_Y(t) - F_X(t), 0))^p dt\right]^{1/p}$$

and $\theta^*_\infty(X,Y) = \widehat{\Theta}^*_\infty(X,Y) = \sup_{t\in\mathbb{R}}[\max(F_Y(t) - F_X(t), 0)]$.

With respect to the homogeneity property, the examples (8), (9) and their minimal counterparts show a wide range of degrees. The family $\mathcal{L}^*_p$ and the minimal metrics l^*_p it generates are homogeneous of degree one. In contrast, the families Θ^*_p and θ^*_p are homogeneous of degree $s = 1/p$. At the limit, Θ^*_∞ and θ^*_∞ are both homogeneous of degree zero.

5 Conclusion

This paper discusses the connections between probability metrics theory and benchmark tracking-error problems. We define a new class of functionals, metrizing the relative deviation of a portfolio to a benchmark. Firstly, we observe that the class of deviation measures introduced by Rockafellar et al. (2006) is generated by the class of translation invariant r.d. metrics. Secondly, we divide all r.d. metrics into three categories — primary, simple and compound functionals and introduce minimal r.d. metrics. The three classes of r.d. metrics assign different properties to the optimal portfolio problem and the minimal functional can be used in a constructive way to obtain simple r.d. metrics. The classification and the methods are inspired by the theory of probability metrics. Further on, we show that under certain conditions, the minimal quasi-semi-metric admits an integral representation.

Even though in the paper we consider a static problem, the generality of the suggested approach allows for extensions in a dynamic setting by studying quasi-semi-metrics not in the space of random variables but in the space of random processes.

References

Artzner, P., F. Delbaen, J.-M. Eber and D. Heath (1998), 'Coherent measures of risk', *Math. Fin.* **6**, 203–228.

Cambanis, S., G. Simons and W. Stout (1976), 'Inequalities for ek(x,y) when the marginals are fixed', *Z. Wahrsch. Verw. Geb.* **36**, 285–294.

Föllmer, H. and A. Schied (2002), 'Convex measures of risk and trading constraints', *Finance and Stochastics* pp. 429–447.

Frittelli, M. and E. Rosazza Gianin (2002), 'Putting order in risk measures', *Journal of Banking and Finance* **26**, 1473–1486.

Rachev, S. T. (1991), *Probability Metrics and the Stability of Stochastic Models*, Wiley, Chichester, U.K.

Rockafellar, R. T., S. Uryasev and M. Zabarankin (2006), 'Generalized deviations in risk analysis', *Finance and Stochastics* **10**, 51–74.

Szegö, G. (2004), *Risk measures for the* 21^{st} *century*, Wiley & Son Chichester.