# Diversification quotients: Quantifying diversification via risk measures

Xia Han *        Liyuan Lin†        Ruodu Wang‡

June 25, 2024

**Abstract**

We establish the first axiomatic theory for diversification indices using six intuitive axioms: non-negativity, location invariance, scale invariance, rationality, normalization, and continuity. The unique class of indices satisfying these axioms, called the diversification quotients (DQs), are defined based on a parametric family of risk measures. A further axiom of portfolio convexity pins down DQ based on coherent risk measures. DQ has many attractive properties, and it can address several theoretical and practical limitations of existing indices. In particular, for the popular risk measures Value-at-Risk and Expected Shortfall, the corresponding DQ admits simple formulas and it is efficient to optimize in portfolio selection. Moreover, it can properly capture tail heaviness and common shocks, which are neglected by traditional diversification indices. When illustrated with financial data, DQ is intuitive to interpret, and its performance is competitive against other diversification indices.

**Keywords**: Expected Shortfall, axiomatic framework, diversification benefit, portfolios, quasi-convexity

## 1 Introduction

Portfolio diversification refers to investment strategies that spread out among many assets, usually with the hope to reduce the volatility or risk of the resulting portfolio. A mathematical formalization of diversification in a portfolio selection context was made by Markowitz (1952), and some early literature on diversification includes Sharpe (1964), Samuelson (1967), Levy and Sarnat (1970) and Fama and Miller (1972), amongst others.

*School of Mathematical Sciences and LPMC, Nankai University, China. xiahan@nankai.edu.cn

†Department of Statistics and Actuarial Science, University of Waterloo, Canada. l89lin@uwaterloo.ca

‡Department of Statistics and Actuarial Science, University of Waterloo, Canada. wang@uwaterloo.ca

Although diversification is conceptually simple, the question of how to measure diversification *quantitatively* is never well settled. An intuitive, but non-quantitative, approach is to simply count the number of distinct stocks or industries of substantial weight in the portfolio; see e.g., Green and Hollifield (1992), Denis et al. (2002) and DeMiguel et al. (2009) in different contexts. This approach is heuristic as it does not involve statistical or stochastic modeling. The second approach is to compute a quantitative index of the portfolio model, based on e.g., the volatility, variance, an expected utility, or a risk measure; this idea is certainly along the direction of Markowitz (1952). In addition, one may empirically address diversification by combining both approaches; see e.g., Tu and Zhou (2011) for the performance of different diversified portfolio strategies, D'Acunto et al. (2019) in the context of robo-advising, and Berger and Eeckhoudt (2021) from the perspective of risk aversion and ambiguity aversion. Green and Hollifield (1992) studied conditions under which the two approaches are roughly in-line with each other.

In this paper, we take the second approach by assigning a quantifier, called a *diversification index*, to each modeled portfolio. Carrying the idea of Markowitz (1952), we start our journey with a simple index, the diversification ratio (DR) based on the standard deviation (SD). For a random vector $\mathbf{X} = (X_1, \dots, X_n)$ representing future random losses and profits of individual components in a portfolio in one period,[1] DR based on SD is defined as

$$\mathrm{DR}^{\mathrm{SD}}(\mathbf{X}) = \frac{\mathrm{SD}\left(\sum_{i=1}^n X_i\right)}{\sum_{i=1}^n \mathrm{SD}(X_i)}; \tag{1}$$

see Choueifaty and Coignard (2008). One can also replace SD by variance. Intuitively, with a smaller value indicating a stronger diversification, the index $\mathrm{DR}^{\mathrm{SD}}$ quantifies the improvement of the portfolio SD over the sum of SD of its components, and it has several convenient properties. Nevertheless, it is well-known that SD is a coarse, non-monotone and symmetric measurement of risk, making it unsuitable for many risk management applications, especially in the presence of heavy-tailed and skewed loss distributions; see Embrechts et al. (2002) for thorough discussions.

Risk measures, in particular the Value-at-Risk (VaR) and the Expected Shortfall (ES), are more flexible quantitative tools, widely used in both financial institutions' internal risk management and banking and insurance regulatory frameworks, such as Basel III/IV (BCBS (2019)) and Solvency II (EIOPA (2011)). ES has many nice theoretical properties and satisfies the four axioms of coherence (Artzner et al. (1999)), whereas VaR is not subadditive in general, but it enjoys other practically useful properties; see Embrechts et al. (2014, 2018), Emmer et al. (2015) and the references therein for more discussions on the issues of VaR versus ES.

Some indices of diversification based on various risk measures have been proposed in the literature. For a given risk measure $\phi$, an example of a diversification index is DR in (1) with

[1] We focus on the one-period losses to establish the theory. This is consistent with the vast majority of literature on risk measures and decision models.

SD replaced by $\phi$; see Tasche (2007). For a review of diversification indices, see Koumou (2020). We find several demerits of DR built on a general risk measure $\phi$ such as VaR or ES in Section 2. A natural question is whether we can design a suitable index based on risk measures to quantify the magnitude of diversification, which avoids the deficiencies of DR. Answering this and related questions is the main purpose of this paper.

We take an axiomatic approach to find our desirable diversification indices. Axiomatic approaches for risk and decision indices have been prolific in economic and statistical decision theories; see e.g., the recent discussions of Gilboa et al. (2019) and the monographs Gilboa (2009) and Wakker (2010). Closely related to diversification indices, risk measures (Artzner et al. (1999); Frittelli and Rosazza Gianin (2002); Föllmer and Schied (2016)) and acceptability indices (Cherny and Madan (2009)) also admit sound axiomatic foundation; the particular cases of VaR and ES are studied by Chambers (2009) and Wang and Zitikis (2021).

In Section 3, as our main contributions, we establish the first axiomatic foundation of diversification indices.[2] This axiomatic theory leads to the class of diversification quotients (DQs), the main object of this paper, which have an interpretation parallel to DR. Six simple axioms—non-negativity, location invariance, scale invariance, rationality, normalization, and continuity—are introduced and justified for their desirability in quantifying diversification. Their interpretations are self-evident and they describe the basic requirements for a diversification index. In Theorem 1, these six axioms characterize DQ based on monetary and positive homogeneous risk measures. A seventh axiom of portfolio convexity, planting an intuitive ordering over portfolio weights in the index, further pins down DQ based on coherent risk measures in Theorem 2. Further, Proposition 1 gives conditions for which such DQ has the range of a standard interval. Portfolio convexity means that, with a given list of assets, combining a portfolio with a better-diversified one does not lead to worse diversification than the original portfolio, reflecting a fundamental principle in economics (Mas-Colell et al. (1995)). The financial interpretation of DQ is that it quantifies the improvement of a risk-level parameter (such as the parameter in VaR or ES) caused by pooling assets, and this is discussed in Section 3.4.

A detailed analysis of the properties of DQ based on general risk measures is discussed in Section 4, which reveals that DQ has many appealing features, both theoretically and practically. In addition to standard operational properties (Proposition 2), DQ has intuitive behaviour for several benchmark portfolio scenarios (Theorem 3). Moreover, DQ allows for consistency with stochastic dominance (Proposition 3) and a fair comparison across portfolio dimensions (Proposition 4). We proceed to focus on VaR and ES in Section 5. It turns out that DQs based on VaR

[2]A different list of desirable axioms for diversification indices is studied by Koumou and Dionne (2022). Their framework is mathematically different from ours as their diversification indices are mappings of portfolio weights, instead of mappings of portfolio random vectors. They did not provide axiomatic characterization results.

and ES have convenient alternative formulations (Theorem 4) and a natural range of $[0, n]$ and $[0, 1]$, respectively (Proposition 5). Further, they report intuitive comparisons between normal and t-models and it has the nice feature that it can capture heavy tails and common shocks.

In Section 6, efficient algorithms for DQs based on VaR and ES in portfolio optimization based on empirical observations are obtained (Proposition 6). Our new diversification index is applied to financial data in Section 7, where several empirical observations highlight the advantages of DQ. We conclude the paper in Section 8 by discussing a number of implications and promising future directions for DQ. Some additional results, proofs, and some omitted numerical results are relegated to the E-Companion.

**Notation.** Throughout this paper, $(\Omega, \mathcal{F}, \mathbb{P})$ is an atomless probability space, on which almost surely equal random variables are treated as identical. A risk measure $\phi$ is a mapping from $\mathcal{X}$ to $\mathbb{R}$, where $\mathcal{X}$ is a convex cone of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ representing losses faced by a financial institution or an investor (i.e., a sign flip from Artzner et al. (1999)), and $\mathcal{X}$ is assumed to include all constants (i.e., degenerate random variables). For $p \in (0, \infty)$, denote by $L^p = L^p(\Omega, \mathcal{F}, \mathbb{P})$ the set of all random variables $X$ with $\mathbb{E}[|X|^p] < \infty$ where $\mathbb{E}$ is the expectation under $\mathbb{P}$. Furthermore, $L^\infty = L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ is the space of all (essentially) bounded random variables, and $L^0 = L^0(\Omega, \mathcal{F}, \mathbb{P})$ is the space of all random variables. Write $X \sim F$ if the random variable $X$ has the distribution function $F$ under $\mathbb{P}$, and $X \buildrel \mathrm{d} \over = Y$ if two random variables $X$ and $Y$ have the same distribution. Further, denote by $\mathbb{R}_+ = [0, \infty)$ and $\overline{\mathbb{R}} = [-\infty, \infty]$. Terms such as increasing or decreasing functions are in the non-strict sense. For $X \in L^0$, $\text{ess-sup}(X)$ and $\text{ess-inf}(X)$ are the essential supremum and the essential infimum of $X$, respectively. Let $n$ be a fixed positive integer representing the number of assets in a portfolio, and write $[n] = \{1, \ldots, n\}$. It does not hurt to think about $n \geqslant 2$ although our results hold also (trivially) for $n = 1$. The vector $\mathbf{0}$ represents the $n$-vector of zeros, and we always write $\mathbf{X} = (X_1, \ldots, X_n)$ and $\mathbf{Y} = (Y_1, \ldots, Y_n)$.

## 2 Preliminaries and motivation

The main object of the paper, a *diversification index* $D$ is a mapping from $\mathcal{X}^n$ to $\overline{\mathbb{R}}$, which is used to quantify the magnitude of diversification of a risk vector $\mathbf{X} \in \mathcal{X}^n$ representing portfolio losses. Our convention is that a smaller value of $D(\mathbf{X})$ represents a stronger diversification in a sense specified by the design of $D$.

As the evaluation of diversification is closely related to that of risk, diversification indices in the literature are often defined through risk measures. An example of a diversification index is the diversification ratio (DR) mentioned in the Introduction based on measures of variability such as the standard deviation (SD) and variance (var):

$$\mathrm{DR}^{\mathrm{SD}}(\mathbf{X}) = \frac{\mathrm{SD}\left(\sum_{i=1}^n X_i\right)}{\sum_{i=1}^n \mathrm{SD}(X_i)} \quad \text{and} \quad \mathrm{DR}^{\mathrm{var}}(\mathbf{X}) = \frac{\mathrm{var}\left(\sum_{i=1}^n X_i\right)}{\sum_{i=1}^n \mathrm{var}(X_i)},$$

with the convention $0/0 = 0$. We refer to Rockafellar et al. (2006), Furman et al. (2017) and Bellini et al. (2022) for general measures of variability. DRs based on SD and var satisfy the three simple properties below, which can be easily checked.

[+] Non-negativity: $D(\mathbf{X}) \geqslant 0$ for all $\mathbf{X} \in \mathcal{X}^n$.

[LI] Location invariance: $D(\mathbf{X}+\mathbf{c}) = D(\mathbf{X})$ for all $\mathbf{c} = (c_1, \dots, c_n) \in \mathbb{R}^n$ and all $\mathbf{X} \in \mathcal{X}^n$.

[SI] Scale invariance: $D(\lambda\mathbf{X}) = D(\mathbf{X})$ for all $\lambda > 0$ and all $\mathbf{X} \in \mathcal{X}^n$.

The first property, [+], simply means that diversification is measured in non-negative values, where 0 typically represents a fully diversified or hedged portfolio (in some sense). The property [LI] means that injecting constant losses or gains to components of a portfolio, or changing the initial price of assets in the portfolio,[3] does not affect its diversification index. The property [SI] means that rescaling a portfolio does not affect its diversification index. The latter two properties are arguably natural, although they are not satisfied by some diversification indices used in the literature (see (2) below). A diversification index satisfying both [LI] and [SI] is called location-scale invariant.

Next, we define the two popular risk measures in banking and insurance practice. The VaR at level $\alpha \in [0,1)$ is defined as

$$\mathrm{VaR}_\alpha(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leqslant x) \geqslant 1-\alpha\}, \quad X \in L^0,$$

and the ES (also called CVaR, TVaR or AVaR) at level $\alpha \in (0,1)$ is defined as

$$\mathrm{ES}_\alpha(X) = \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_\beta(X)\mathrm{d}\beta, \quad X \in L^1,$$

and $\mathrm{ES}_0(X) = \text{ess-sup}(X) = \mathrm{VaR}_0(X)$, which may be $\infty$. The probability level $\alpha$ above is typically very small, e.g., 0.01 or 0.025 in BCBS (2019); note that we use the "small $\alpha$" convention. Artzner et al. (1999) introduced *coherent* risk measures $\phi : \mathcal{X} \to \mathbb{R}$ as those satisfying the following four properties.

[M] Monotonicity: $\phi(X) \leqslant \phi(Y)$ for all $X, Y \in \mathcal{X}$ with $X \leqslant Y$.[4]

[3] Recall that $X_i$ represents the loss from asset $i$. Suppose that two agents purchased the same portfolio of assets but at different prices of each asset. Denote by $\mathbf{X}$ the portfolio loss vector of agent 1. The portfolio loss vector of agent 2 is $\mathbf{X}+\mathbf{c}$, where $\mathbf{c}$ is the vector of differences between their purchase prices. The two agents should have the same level of diversification regardless of their purchase prices, as they hold the same portfolio.

[4] The inequality $X \leqslant Y$ between two random variables $X$ and $Y$ is pointwise.

[CA] Constant additivity: $\phi(X+c)=\phi(X)+c$ for all $c\in\mathbb{R}$ and $X\in\mathcal{X}$.

[PH] Positive homogeneity: $\phi(\lambda X)=\lambda\phi(X)$ for all $\lambda\in(0,\infty)$ and $X\in\mathcal{X}$.

[SA] Subadditivity: $\phi(X+Y)\leqslant\phi(X)+\phi(Y)$ for all $X,Y\in\mathcal{X}$.

ES satisfies all four properties above, whereas VaR does not satisfy [SA]. We say that a risk measure is *monetary* if it satisfies [CA] and [M], and it is *MCP* if it satisfies [M], [CA] and [PH]. For discussions and interpretations of these properties, we refer to Föllmer and Schied (2016).

Some diversification indices are defined via risk measures, such as DR (e.g., Bürgi et al. (2008), Mainik and Rüschendorf (2010) and Embrechts et al. (2015)) and the diversification benefit (DB, e.g., Embrechts et al. (2009) and McNeil et al. (2015)). For a risk measure $\phi$, DR and DB based on $\phi$ are defined as[5]

$$\mathrm{DR}^{\phi}(\mathbf{X})=\frac{\phi\left(\sum_{i=1}^{n}X_i\right)}{\sum_{i=1}^{n}\phi(X_i)}\quad\text{and}\quad\mathrm{DB}^{\phi}(\mathbf{X})=\sum_{i=1}^{n}\phi(X_i)-\phi\left(\sum_{i=1}^{n}X_i\right).\tag{2}$$

In contrast to DR, a larger value of DB represents a stronger diversification, but this convention can be easily modified by flipping the sign to consider $-\mathrm{DB}^{\phi}$. By definition, DR is the ratio of the pooled risk to the sum of the individual risks, and thus a measurement of how substantially pooling reduces risk; similarly, DB measures the difference instead of the ratio.

DR has a number of deficiencies. First, the value of $\mathrm{DR}^{\phi}$ is not necessarily non-negative, violating [+]. Since the risk measure $\phi$ may take negative values,[6] it would be difficult to interpret the case where either the numerator or denominator in DR is negative, and this makes optimization of DR troublesome. An example is a portfolio of credit default losses, where VaR of individual losses is often 0 or negative but VaR of the portfolio loss is positive; see McNeil et al. (2015, Example 2.25). Second, for common risk measures, DR violates [LI], meaning that adding a risk-free asset changes the value of DR. Third, DR is not necessarily quasi-convex in portfolio weights; this point is more subtle and will be explained later. In addition to the above drawbacks, we also find that DR has wrong incentives for some simple models; for instance, it suggests that an iid portfolio of $t$-distributed risks is less diversified than a portfolio with a common shock and the same marginals; see Section 5.2 for details. Similarly to DR, the index DB satisfies [LI] for $\phi$ satisfying [CA], but it does not satisfy [SI] for common risk measures, and it may take both positive and negative values.

In financial applications, the risk measures VaR and ES are specified in regulatory documents such as BCBS (2019) and EIOPA (2011), and therefore it is beneficial to stick to VaR

[5] If the denominator in the definition of $\mathrm{DR}^{\phi}(\mathbf{X})$ is 0, then we use the convention $0/0=0$ and $1/0=\infty$.

[6] A negative value of a risk measure has a concrete meaning as the amount of capital to be withdrawn from a portfolio position while keeping it acceptable; see Artzner et al. (1999).

or ES as the risk measure when assessing diversification. Both $\mathrm{DR}^{\mathrm{VaR}_\alpha}$ and $\mathrm{DR}^{\mathrm{ES}_\alpha}$ satisfy [SI], but they do not satisfy [+] or [LI].[7] It remains unclear how one can define a diversification index based on VaR or ES satisfying these properties. In the remainder of the paper, we will introduce and study a new index of diversification to bridge this gap.

# 3 Diversification indices: An axiomatic theory

In this section, we fix $\mathcal{X} = L^\infty$ as the standard choice in the literature of axiomatic theory of risk measures. In addition to [+], [LI] and [SI] introduced in Section 2, we propose four new axioms. The first six axioms together characterize a new class of diversification indices, that is, diversification quotients (DQ) based on MCP risk measures. With the seventh axiom of portfolio convexity, we further pin down the class of DQ based on coherent risk measures.

## 3.1 Axioms of rationality, normalization, and continuity

We first present three axioms, which depend on a risk measure $\phi$. These three axioms are standard and weak in the sense that they do not impose a specific functional structure on $D$ other than some forms of monotonicity, normalization, and continuity.

For a risk measure $\phi$, we say that two vectors $\mathbf{X}, \mathbf{Y} \in \mathcal{X}^n$ are *$\phi$-marginally equivalent* if $\phi(X_i) = \phi(Y_i)$ for each $i \in [n]$, and we denote this by $\mathbf{X} \overset{\mathrm{m}}{\simeq} \mathbf{Y}$. In other words, if an agent evaluates risks using the risk measure $\phi$, then she would view the individual components of $\mathbf{X}$ and those of $\mathbf{Y}$ as equally risky. Similarly, denote by $\mathbf{X} \overset{\mathrm{m}}{\succeq} \mathbf{Y}$ if $\phi(X_i) \leqslant \phi(Y_i)$ for each $i \in [n]$, and by $\mathbf{X} \overset{\mathrm{m}}{\succ} \mathbf{Y}$ if $\phi(X_i) < \phi(Y_i)$ for each $i \in [n]$. The other three desirable axioms are presented below, and they are built on a given risk measure $\phi$, such as VaR or ES, typically specified exogenously by financial regulation.

$[\mathrm{R}]_\phi$ Rationality: $D(\mathbf{X}) \leqslant D(\mathbf{Y})$ for $\mathbf{X}, \mathbf{Y} \in \mathcal{X}^n$ satisfying $\mathbf{X} \overset{\mathrm{m}}{\simeq} \mathbf{Y}$ and $\sum_{i=1}^n X_i \leqslant \sum_{i=1}^n Y_i$.

To interpret the axiom $[\mathrm{R}]_\phi$, consider two portfolios $\mathbf{X}$ and $\mathbf{Y}$ satisfying $\mathbf{X} \overset{\mathrm{m}}{\simeq} \mathbf{Y}$. If further $\sum_{i=1}^n X_i \leqslant \sum_{i=1}^n Y_i$ holds, then the total loss from portfolio $\mathbf{X}$ is always less or equal to that from portfolio $\mathbf{Y}$, making the portfolio $\mathbf{X}$ safer than $\mathbf{Y}$. Since the individual components in $\mathbf{X}$ and those in $\mathbf{Y}$ are equally risky, the fact that $\mathbf{X}$ is safer in aggregation is a result of the different diversification effects in $\mathbf{X}$ and $\mathbf{Y}$, leading to the inequality $D(\mathbf{X}) \leqslant D(\mathbf{Y})$. This axiom is called rationality because a rational agent always prefers to have smaller losses.

Next, we formulate our idea about normalizing representative values of the diversification index. First, we assign the zero portfolio $\mathbf{0}$ the value $D(\mathbf{0}) = 0$, as it carries no risk in every

[7] An impossibility result (Proposition EC.1) is presented in Appendix B, which suggests that it is not possible to construct non-trivial diversification indices like DR and DB satisfying [+], [LI] and [SI].

sense.[8] A natural benchmark of a non-diversified portfolio is one in which all components are the same. Such a portfolio $\mathbf{X}^{\mathrm{du}} = (X, \dots, X)$ will be called a *duplicate* portfolio, and we may, ideally, wish to assign the value $D(\mathbf{X}^{\mathrm{du}}) = 1$. However, since the zero portfolio $\mathbf{0}$ is also duplicate but $D(\mathbf{0}) = 0$, we will require the weaker assumption $D(\mathbf{X}^{\mathrm{du}}) \leqslant 1$ for duplicate portfolios.[9] Lastly, we should understand for what portfolios $D(\mathbf{X}) \geqslant 1$ needs to occur. We say that a portfolio $\mathbf{X}^{\mathrm{wd}} = (X_1, \dots, X_n)$ is *worse than duplicate*, if there exists a duplicate portfolio $\mathbf{X}^{\mathrm{du}} = (X, \dots, X)$ such that $\mathbf{X}^{\mathrm{wd}} \overset{\mathrm{m}}{\succ} \mathbf{X}^{\mathrm{du}}$ and $\sum_{i=1}^n X_i \geqslant nX$. Intuitively, this means that each component of $\mathbf{X}^{\mathrm{wd}}$ is strictly less risky than $X$, but putting them together always incurs a larger loss than $nX$; in this case, diversification creates nothing but a penalty to the risk manager, and we assign $D(\mathbf{X}^{\mathrm{wd}}) \geqslant 1$.[10] Existence of worse-than-duplicate portfolios is characterized in Appendix C.1. Putting all of the considerations above, we propose the following normalization axiom.

$[\mathrm{N}]_\phi$ Normalization: $D(\mathbf{0}) = 0$, $D(\mathbf{X}) \leqslant 1$ if $\mathbf{X}$ is duplicate, and $D(\mathbf{X}) \geqslant 1$ if $\mathbf{X}$ is worse than duplicate.

Finally, we propose a continuity axiom which is mainly for technical convenience.

$[\mathrm{C}]_\phi$ Continuity: For $\{\mathbf{Y}^k\}_{k\in\mathbb{N}} \subseteq \mathcal{X}^n$ and $\mathbf{X} \in \mathcal{X}^n$ satisfying $\mathbf{Y}^k \overset{\mathrm{m}}{\simeq} \mathbf{X}$ for each $k$, if $(\sum_{i=1}^n X_i - \sum_{i=1}^n Y_i^k)_+ \xrightarrow{L^\infty} 0$ as $k \to \infty$, then $(D(\mathbf{X}) - D(\mathbf{Y}^k))_+ \to 0$.

The axiom $[\mathrm{C}]_\phi$ is a special form of semi-continuity. To interpret it, consider portfolios $\mathbf{X}$ and $\mathbf{Y}$ that are marginally equivalent. If the sum of components of $\mathbf{X}$ is not much worse than that of $\mathbf{Y}$ in $L^\infty$, then the axiom says that the diversification of $\mathbf{X}$ is not much worse than the diversification of $\mathbf{Y}$. This property allows for a special form of stability or robustness[11] with respect to statistical errors when estimating the distributions of portfolio losses.

One can check that the axioms $[\mathrm{R}]_\phi$, $[\mathrm{N}]_\phi$ and $[\mathrm{C}]_\phi$ are satisfied by $\mathrm{DR}^{\mathrm{VaR}_\alpha}$ and $\mathrm{DR}^{\mathrm{ES}_\alpha}$ if we only consider positive portfolio vectors. The axioms are not satisfied by $\mathrm{DR}^{\mathrm{SD}}$ because SD is not monotone and hence the inequalities $\sum_{i=1}^n X_i \leqslant \sum_{i=1}^n Y_i$ and $\sum_{i=1}^n X_i \geqslant nX$ used in $[\mathrm{R}]_\phi$ and $[\mathrm{N}]_\phi$ are not relevant for SD.

[8] Indeed, the value of $D(\mathbf{0})$ may be rather arbitrary; this is the case for DR where $0/0$ occurs.

[9] Theorem 3 gives some mild conditions that yield $D(\mathbf{X}^{\mathrm{du}}) = 1$ for the class $D$ characterized in this section.

[10] Such situations may be regarded as diversification disasters; see Ibragimov et al. (2011).

[11] In the literature of statistical robustness, often a different metric than the $L^\infty$ metric is used; see Huber and Ronchetti (2009) for a general treatment. Our choice of formulating continuity via the $L^\infty$ metric is standard in the axiomatic theory of risk mappings on $L^\infty$.

## 3.2 Portfolio convexity

The next axiom, different from the three above, imposes a natural form of convexity on the diversification index. Portfolio diversification is intrinsically connected to convexity of ordering relations. Quoting Mas-Colell et al. (1995, p. 44), "Convexity can also be viewed as the formal expression of a basic inclination of economic agents for diversification." For this purpose, we propose an axiom of portfolio convexity in this section.

Let a random vector $\mathbf{X} \in \mathcal{X}^n$ represent losses from $n$ assets and a vector $\mathbf{w} = (w_1, \dots, w_n) \in \Delta_n$ of portfolio weights, where $\Delta_n$ is the standard $n$-simplex, given by

$$\Delta_n = \{\mathbf{x} \in [0,1]^n : x_1 + \dots + x_n = 1\}.$$

The total loss of the portfolio is $\mathbf{w}^\top \mathbf{X}$. We write $\mathbf{w} \odot \mathbf{X} = (w_1 X_1, \dots, w_n X_n)$, which is the portfolio loss vector with the weight $\mathbf{w}$. The *portfolio convexity* axiom is formulated below.

[PC] Portfolio convexity: The set $\{\mathbf{w} \in \Delta_n : D(\mathbf{w} \odot \mathbf{X}) \leqslant d)\}$ is convex for each $\mathbf{X} \in \mathcal{X}^n$ and $d \in \overline{\mathbb{R}}$.

Intuitively, portfolio convexity means that, for a given vector $\mathbf{X}$ of assets, combining a portfolio strategy with a better-diversified one on the same set of assets does not result in a portfolio that is less diversified than the original portfolio. As convexity is the decision-theoretic counterpart of diversification, [PC] is desirable for diversification indices.

*Remark* 1. Axiom [PC] is equivalent to *quasi-convexity* of $\mathbf{w} \mapsto D(\mathbf{w} \odot \mathbf{X})$ for each $\mathbf{X} \in \mathcal{X}^n$; that is, $D((\lambda \mathbf{w} + (1-\lambda)\mathbf{w}') \odot \mathbf{X}) \leqslant D(\mathbf{w} \odot \mathbf{X}) \vee D(\mathbf{w}' \odot \mathbf{X})$ for all $\lambda \in [0,1]$, $\mathbf{w}, \mathbf{w}' \in \Delta_n$ and $\mathbf{X} \in \mathcal{X}^n$.

*Remark* 2. Convexity or quasi-convexity of $\mathbf{X} \mapsto D(\mathbf{X})$ is not natural or desirable. For instance, combining two diversified portfolios $(X, Y)$ and $(Y, X)$ may result in a duplicate portfolio; see Example 3 in Appendix C.2. Convexity of $\mathbf{w} \mapsto D(\mathbf{w} \odot \mathbf{X})$, which is stronger than [PC], is not desirable either; see Example 4 in Appendix C.2.

The four axioms introduced above, together with the three in Section 2, lead to a class of diversification indices, which we define next.

## 3.3 Characterization results

We first formally introduce the diversification index DQ relying on a parametric class of risk measures, which will be characterized in two results below.

**Definition 1.** Let $\rho = (\rho_\alpha)_{\alpha \in I}$ be a class of risk measures indexed by $\alpha \in I = (0, \overline{\alpha})$ with $\overline{\alpha} \in (0, \infty]$ such that $\rho_\alpha$ is decreasing in $\alpha$. For $\mathbf{X} \in \mathcal{X}^n$, the *diversification quotient* based on the class $\rho$ at level $\alpha \in I$ is defined by

$$\mathrm{DQ}^{\rho}_{\alpha}(\mathbf{X}) = \frac{\alpha^*}{\alpha}, \quad \text{where } \alpha^* = \inf\left\{\beta \in I : \rho_\beta\left(\sum_{i=1}^{n} X_i\right) \leqslant \sum_{i=1}^{n} \rho_\alpha(X_i)\right\}, \tag{3}$$

with the convention $\inf(\varnothing) = \overline{\alpha}$.

We first characterize DQ based on MCP risk measures by six axioms without [PC].

**Theorem 1.** *A diversification index $D : \mathcal{X}^n \to \overline{\mathbb{R}}$ satisfies* [+], [LI], [SI], $[\mathrm{R}]_\phi$, $[\mathrm{N}]_\phi$ *and* $[\mathrm{C}]_\phi$ *for some MCP risk measure $\phi$ if and only if $D$ is $\mathrm{DQ}^{\rho}_{\alpha}$ for some $\alpha$ and decreasing class $\rho$ of MCP risk measures. Moreover, in both directions of the above equivalence, it can be required that $\rho_\alpha = \phi$.*

Theorem 1 gives the first axiomatic characterization of diversification indices, to the best of our knowledge. The proof techniques to show the important "only if" statement of Theorem 1 are based on a sophisticated analysis of an auxiliary mapping

$$R : \mathcal{X} \to [0,\infty],\ R(X) = \inf\left\{D(\mathbf{X}) : X \leqslant \sum_{i=1}^{n} X_i,\ \mathbf{X} \overset{\mathrm{m}}{\simeq} \mathbf{0}\right\},$$

and this is explained in Appendix A.

Next, we incorporate portfolio convexity into our axiomatic framework. For this purpose, it is natural to build the diversification indices based on risk measures with convexity. When formulated on monetary risk measures, convexity represents the idea that diversification reduces the risk; see Föllmer and Schied (2016). For risk measures that are not constant additive, Cerreia-Vioglio et al. (2011) argued that quasi-convexity is more suitable than convexity to reflect the consideration of diversification; moreover, convexity and quasi-convexity are equivalent if [CA] holds. A risk measure is *linear* if it satisfies $\phi(aX + bY) = a\phi(X) + b\phi(Y)$ for all $X, Y \in \mathcal{X}$ and $a, b \in \mathbb{R}$. Since linear risk measures correspond to expectations (under monotonicity), which do not reflect diversification, we will focus on non-linear ones. The next theorem characterizes DQ based on coherent risk measures.

**Theorem 2.** *Suppose $n \geqslant 4$ and $\phi$ is a non-linear coherent risk measure. A diversification index $D : \mathcal{X}^n \to \overline{\mathbb{R}}$ satisfies* [+], [LI], [SI], $[\mathrm{R}]_\phi$, $[\mathrm{N}]_\phi$, $[\mathrm{C}]_\phi$ *and* [PC] *if and only if $D$ is $\mathrm{DQ}^{\rho}_{\alpha}$ for some $\alpha$ and decreasing class $\rho$ of coherent risk measures with $\rho_\alpha = \phi$.*

The conditions $n \geqslant 4$ and non-linearity of $\phi$ are essential to the proof of Theorem 2. They are harmless for financial applications since typical portfolios have more than a few components, and common risk measures are not linear.

Although portfolio convexity is crucial for diversification indices, making Theorem 2 a central result, we present Theorem 1 separately for the following reasons. First, Theorem 1

Table 1. Summary of axioms satisfied by diversification indices $\mathrm{DR}^{\phi}$, $\mathrm{DB}^{\phi}$ and $\mathrm{DQ}^{\rho}_{\alpha}$ (with $\phi=\rho_\alpha$), where $\mathcal{X}_+$ is the set of non-negative elements in $\mathcal{X}$ and $\alpha\in(0,1)$

| Index | Domain | [+] | [LI] | [SI] | $[\mathrm{R}]_\phi$ | $[\mathrm{N}]_\phi$ | $[\mathrm{C}]_\phi$ | [PC] |
|---|---|---|---|---|---|---|---|---|
| $\mathrm{DR}^{\mathrm{VaR}_\alpha}$ and $\mathrm{DR}^{\mathrm{ES}_\alpha}$ | $\mathcal{X}^n$ | × | × | √ | × | × | × | × |
| $\mathrm{DR}^{\mathrm{VaR}_\alpha}$ | $\mathcal{X}^n_+$ | √ | × | √ | √ | √ | √ | × |
| $\mathrm{DR}^{\mathrm{ES}_\alpha}$ | $\mathcal{X}^n_+$ | √ | × | √ | √ | √ | √ | √ |
| $\mathrm{DR}^{\mathrm{SD}}$ | $\mathcal{X}^n$ | √ | √ | √ | × | × | × | √ |
| $\mathrm{DR}^{\mathrm{var}}$ | $\mathcal{X}^n$ | √ | √ | √ | × | × | × | × |
| $-\mathrm{DB}^{\mathrm{VaR}_\alpha}$ | $\mathcal{X}^n$ | × | √ | × | √ | × | √ | × |
| $-\mathrm{DB}^{\mathrm{ES}_\alpha}$ | $\mathcal{X}^n$ | × | √ | × | √ | × | √ | √ |
| $\mathrm{DQ}^{\mathrm{VaR}}_\alpha$ | $\mathcal{X}^n$ | √ | √ | √ | √ | √ | √ | × |
| $\mathrm{DQ}^{\mathrm{ES}}_\alpha$ | $\mathcal{X}^n$ | √ | √ | √ | √ | √ | √ | √ |

reveals the fundamental properties needed to pin down the form of DQ and this helps to clarify the role of [PC]. Second, the proof of Theorem 2 is technically built on Theorem 1. Third, the class of DQ characterized by Theorem 1 allows for DQ based on VaR, which is popular in financial regulation.

In the next proposition, we show that for sub-linear risk measures, DQ satisfies [PC] (thus, the "if" direction of Theorem 2 does not need [M] and [CA] for $\rho$), and its range is $[0,1]$ under mild conditions, avoiding non-degeneracy. A risk measure is sub-linear if it satisfies subadditivity and positive homogeneity (equivalently, convexity and positive homogeneity).

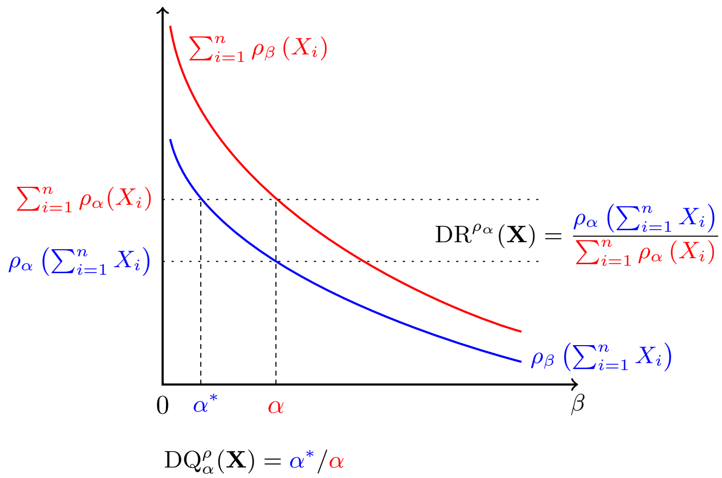
**Proposition 1.** *Let $\rho=(\rho_\beta)_{\beta\in I}$ be a decreasing class of sub-linear risk measures and $\alpha\in I$. Then $\mathrm{DQ}^{\rho}_{\alpha}$ satisfies [PC]. If $n\geqslant 3$, $\rho_\alpha$ is non-linear and there exists $X\in\mathcal{X}$ such that $\beta\mapsto\rho_\beta(X)$ is strictly decreasing, then $\{\mathrm{DQ}^{\rho}_{\alpha}(\mathbf{X}):\mathbf{X}\in\mathcal{X}^n\}=[0,1]$.*

Given a sub-linear risk measure $\rho_\alpha$, the conditions in Proposition 1 for $\{\mathrm{DQ}^{\rho}_{\alpha}(\mathbf{X}):\mathbf{X}\in\mathcal{X}^n\}=[0,1]$ are mild and satisfied by e.g., DQ based on the family of ES. In contrast to DQ, DR based on sub-linear risk measures may not satisfy [PC] since the denominator in (2) may be negative. For a clear comparison, we summarize in Table 1 the axioms satisfied by the diversification indices that appear in the paper.

### 3.4 Interpretation of DQ

DQ based on MCP or coherent risk measures have been characterized axiomatically, but we have not interpreted the meaning of DQ in (3). For an interpretation, consider a decreasing class of risk measures $(\rho_\beta)_{\beta\in I}$. The values of risk measures typically represent the capital requirement

Figure 1. Conceptual symmetry between DQ and DR




of a risky asset, and hence $\beta$ is interpreted as a parameter of risk level (as in $\mathrm{VaR}_\beta$ or $\mathrm{ES}_\beta$), that is, a smaller $\beta$ means a larger capital requirement for the same risk. Notice from (3) that, under mild conditions, $\alpha^*$ is uniquely determined by

$$\rho_{\alpha^*}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \rho_\alpha(X_i).$$

Therefore, $\alpha^*$ is the parameter of risk level achieved by pooling, assuming that the portfolio maintains the same total capital requirement assessed by $\rho_\alpha$ when there is no pooling, that is, $\sum_{i=1}^n \rho_\alpha(X_i)$. As $\mathrm{DQ}_\alpha^\rho(\mathbf{X}) = \alpha^*/\alpha$, DQ is the ratio of the risk-level parameters before and after pooling. To summarize,

*the index DQ quantifies the improvement of the risk-level parameter caused by pooling assets.*

In the most relevant case $\rho_\alpha\left(\sum_{i=1}^n X_i\right) < \sum_{i=1}^n \rho_\alpha(X_i)$, we present in Figure 1 the conceptual symmetry between DQ, which measures the improvement by pooling in the horizontal direction, and DR, which measures an improvement in the vertical direction. In particular, in the case of VaR, DQ measures the probability improvement, whereas DR measures the quantile improvement; see Theorem 4 and (7) below.

*Remark* 3. The idea of improvement of risk level is closely related to acceptability indices, proposed by Cherny and Madan (2009). More precisely, an acceptability index for a loss $X \in \mathcal{X}$ is defined by $\mathrm{AI}^\rho(X) = \sup\{\gamma \in \mathbb{R}_+ : \rho_{1/\gamma}(X) \leqslant 0\}$ for a decreasing class of coherent risk measures $(\rho_\gamma)_{\gamma\in\mathbb{R}_+}$, which has visible similarity to $\alpha^*$ in (3); see Kováčová et al. (2020) for optimization of acceptability indices. If $\rho$ is a class of risk measures satisfying [CA], then

$$\mathrm{DQ}_\alpha^\rho(\mathbf{X}) = \frac{1}{\alpha}\left(\mathrm{AI}^\rho\left(\sum_{i=1}^n (X_i - \rho_\alpha(X_i))\right)\right)^{-1}.$$

Dhaene et al. (2012) studied several methods for capital allocation, among which the quantile allocation principle computes a capital allocation $(C_1, \dots, C_n)$ such that $\sum_{i=1}^n C_i = \mathrm{VaR}_\alpha(\sum_{i=1}^n X_i)$ and $C_i = \mathrm{VaR}_{c\alpha}(X_i)$ for some $c > 0$. The constant $c$ appearing as a nuisance parameter in the above rule has a visible mathematical similarity to $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$. Mafusalov and Uryasev (2018) studied the so-called buffered probability of exceedance, which is the inverse of the ES curve $\beta \mapsto \mathrm{ES}_\beta(X)$ at a specific point $x \in \mathbb{R}$; note that $\alpha^*$ in (3) is obtained by inverting the ES curve $\beta \mapsto \mathrm{ES}_\beta(\sum_{i=1}^n X_i)$ at $\sum_{i=1}^n \mathrm{ES}_\alpha(X_i)$.

We close the section with discussions on the construction of DQ. First, DQ can be constructed from any monotonic parametric family of risk measures. All commonly used risk measures belong to a monotonic family, as this includes VaR, ES, expectiles (e.g., Bellini et al. (2014)), mean-variance (e.g., Markowitz (1952) and Maccheroni et al. (2009)), and entropic risk measures (e.g., Föllmer and Schied (2016)); some choices do not guarantee all axioms to hold. Our results imply that using ES or expectiles guarantees all axioms and non-degeneracy for DQ. In addition, there are ways to construct DQ from a single risk measure; see Appendix C.3. DQ can also be axiomatized using preferences instead of risk measures; see Appendix C.4.

DQ can be used as a normative tool for measuring diversification. In this context, the choice of the parametric family of risk measures is up to the user, and DQ serves as a versatile tool that accommodates various risk attitudes. The choice of risk measures (e.g., VaR, ES) and the determination of the confidence level ($\alpha$) should be aligned with the risk tolerance, objectives, and regulatory requirements of the decision maker. For instance, conservative investors, prioritizing capital preservation, may gravitate towards the family of ES at a high level $\alpha$, which reflects an assessment of downside risk, whereas those with aggressive risk preferences may opt for VaR or ES at a lower level $\alpha$. Most generally, we would recommend the use of DQ based on ES, which has a natural and strong connection to financial regulation and tail risk management, and the parameter $\alpha$ allows for flexibility in the assessment of tail risk.

# 4 Properties of DQ

In this section, we study the properties of DQ defined in Definition 1. For the greatest generality, we do not impose any properties of risk measures in the decreasing family $\rho = (\rho_\alpha)_{\alpha \in I}$, i.e., the family $\rho$ is not limited to MCP or coherent risk measures, so that our results can be applied to more flexible contexts in which some of the seven axioms are relaxed. In this section, $\mathcal{X}$ is not restricted to $L^\infty$.

### 4.1 Basic properties

We first make a few immediate observations by the definition of DQ. From (3), we can see that computing $\mathrm{DQ}^{\rho}_{\alpha}$ is to invert the decreasing function $\beta \mapsto \rho_\beta(\sum_{i=1}^n X_i)$ at $\sum_{i=1}^n \rho_\alpha(X_i)$. For the cases of VaR and ES, $I = (0,1)$, $\alpha^* \in [0,1]$, and DQ has simple formulas; see Theorem 4 in Section 5. For a fixed value of $\sum_{i=1}^n \rho_\alpha(X_i)$, DQ is larger if the curve $\beta \mapsto \rho_\beta(\sum_{i=1}^n X_i)$ is larger, and DQ is smaller if the curve $\beta \mapsto \rho_\beta(\sum_{i=1}^n X_i)$ is smaller. This is consistent with our intuition that a diversification index is large if there is little or no diversification, thus a large value of the portfolio risk, and a diversification index is small if there is strong diversification.

In Theorem 1, we have seen that DQ satisfies [SI] and [LI] if $\rho$ is a class of MCP risk measures. These properties of DQ can be obtained based on a more general version of properties [CA] and [PH] of risk measures, allowing us to include SD and the variance. The results are summarized in Proposition 2, which are straightforward to check by definition.

$[\mathrm{CA}]_m$ Constant additivity with $m \in \mathbb{R}$: $\phi(X + c) = \phi(X) + mc$ for all $c \in \mathbb{R}$ and $X \in \mathcal{X}$.

$[\mathrm{PH}]_\gamma$ Positive homogeneity with $\gamma \in \mathbb{R}$: $\phi(\lambda X) = \lambda^\gamma \phi(X)$ for all $\lambda \in (0, \infty)$ and $X \in \mathcal{X}$.

**Proposition 2.** *Let $\rho = (\rho_\alpha)_{\alpha \in I}$ be a class of risk measures decreasing in $\alpha$. For each $\alpha \in I$,*

*(i) if $\rho_\beta$ satisfies $[\mathrm{PH}]_\gamma$ with the same $\gamma$ across $\beta \in I$, then $\mathrm{DQ}^{\rho}_{\alpha}$ satisfies [SI].*

*(ii) if $\rho_\beta$ satisfies $[\mathrm{CA}]_m$ with the same $m$ across $\beta \in I$, then $\mathrm{DQ}^{\rho}_{\alpha}$ satisfies [LI].*

*(iii) if $\rho_\alpha$ satisfies [SA], then $\mathrm{DQ}^{\rho}_{\alpha}$ takes value in $[0,1]$.*

It is clear that [CA] is $[\mathrm{CA}]_m$ with $m = 1$ and [PH] is $[\mathrm{PH}]_\gamma$ with $\gamma = 1$. More properties of DQs on the important families of VaR and ES will be discussed in Section 5. In particular, we will see that the ranges of $\mathrm{DQ}^{\mathrm{VaR}}_{\alpha}$ and $\mathrm{DQ}^{\mathrm{ES}}_{\alpha}$ are $[0, n]$ and $[0,1]$, respectively.

**Example 1** (Liquidity and temporal consistency)**.** In risk management practice, liquidity and time-horizon of potential losses need to be taken into account; see BCBS (2019, p.89). If liquidity risk is of concern, one may use a risk measure with $[\mathrm{PH}]_\gamma$ with $\gamma > 1$ to penalize large exposures of losses. For such risk measures, $\mathrm{DQ}^{\rho}_{\alpha}$ remains scale invariant, as shown by Proposition 2. On the other hand, if the risk associated to the loss $\mathbf{X}(t)$ at different time spots $t > 0$ is scalable by a function $f > 0$ (usually of the order $f(t) = \sqrt{t}$ in standard models such as the Black-Scholes), then DQ is consistent across different horizons in the sense that $\mathrm{DQ}^{\rho}_{\alpha}(\mathbf{X}(t)) = \mathrm{DQ}^{\rho}_{\alpha}(\mathbf{X}(s))$ for two time spots $s, t > 0$, given that $\rho_\beta(X_i(t)) = f(t)\rho_\beta(X_i(1))$ for $i \in [n]$, $t > 0$ and $\beta \in I$.

Next, we explain that the values taken by DQ are consistent with our usual perceptions of portfolio diversification. For a given risk measure $\phi$ and a portfolio risk vector $\mathbf{X}$, we consider the following three situations which yield intuitive values of DQ.

(i) There is no insolvency risk with pooled individual capital, i.e., $\sum_{i=1}^n X_i \leqslant \sum_{i=1}^n \phi(X_i)$ a.s.;

(ii) diversification benefit exists, i.e., $\phi\left(\sum_{i=1}^n X_i\right) < \sum_{i=1}^n \phi(X_i)$;

(iii) the portfolio relies on a single asset, i.e., $\mathbf{X} = (\lambda_1 X, \dots, \lambda_n X)$ for some $X \in \mathcal{X}$ and $\lambda_1, \dots, \lambda_n \in \mathbb{R}_+$. A duplicate portfolio relies on a single asset.

The above three situations receive special attention because they intuitively correspond to very strong diversification, some diversification, and no diversification, respectively. Naturally, we would expect DQ to be very small for (i), DQ to be smaller than 1 for (ii), and DQ to be 1 for (iii). It turns out that the above intuitions all check out under very weak conditions that are satisfied by commonly used classes of risk measures.

Before presenting this result, we fix some technical terms. For a class $\rho$ of risk measures $\rho_\alpha$ decreasing in $\alpha$, we say that $\rho$ is *non-flat from the left* at $(\alpha, X)$ if $\rho_\beta(X) > \rho_\alpha(X)$ for all $\beta \in (0, \alpha)$, and $\rho$ is *left continuous* at $(\alpha, X)$ if $\alpha \mapsto \rho_\alpha(X)$ is left continuous. A random vector $(X_1, \dots, X_n)$ is *comonotonic* if there exists a random variable $Z$ and increasing functions $f_1, \dots, f_n$ on $\mathbb{R}$ such that $X_i = f_i(Z)$ a.s. for every $i \in [n]$. A risk measure is *comonotonic-additive* if $\phi(X+Y) = \phi(X) + \phi(Y)$ for comonotonic $(X, Y)$. Each of ES and VaR satisfies comonotonic-additivity, as well as any distortion risk measures (Yaari (1987), Kusuoka (2001)) and signed Choquet integrals (Wang et al. (2020)). We denote by $\rho_0 = \lim_{\alpha \downarrow 0} \rho_\alpha$. Note that $\rho_0 = \text{ess-sup}$ for common classes $\rho$ such as VaR, ES, expectiles, and entropic risk measures.

**Theorem 3.** *For given $\mathbf{X} \in \mathcal{X}^n$ and $\alpha \in I$, if $\rho$ is left continuous and non-flat from the left at $(\alpha, \sum_{i=1}^n X_i)$, the following hold.*

*(i) Suppose $\rho_0 \leqslant \text{ess-sup}$. If for $\rho_\alpha$ there is no insolvency risk with pooled individual capital, then $\mathrm{DQ}^\rho_\alpha(\mathbf{X}) = 0$. The converse holds true if $\rho_0 = \text{ess-sup}$.*

*(ii) Diversification benefit exists if and only if $\mathrm{DQ}^\rho_\alpha(\mathbf{X}) < 1$.*

*(iii) If $\rho_\alpha$ satisfies [PH] and $\mathbf{X}$ relies on a single asset, then $\mathrm{DQ}^\rho_\alpha(\mathbf{X}) = 1$.*

*(iv) If $\rho_\alpha$ is comonotonic-additive and $\mathbf{X}$ is comonotonic, then $\mathrm{DQ}^\rho_\alpha(\mathbf{X}) = 1$.*

In (i), we see that if there is no insolvency risk with pooled individual capital, then $\mathrm{DQ}^\rho_\alpha(\mathbf{X}) = 0$. In typical models, such as some elliptical models in Section 5.2, $\sum_{i=1}^n X_i$ is unbounded from above unless it is a constant. Hence, for such models and $\rho$ satisfying $\rho_0 = \text{ess-sup}$, $\mathrm{DQ}^\rho_\alpha(\mathbf{X}) = 0$ if and only if $\sum_{i=1}^n X_i$ is a constant, thus full hedging is achieved. This is also consistent with our intuition of full hedging as the strongest form of diversification. The existence of diversification benefit is the main idea behind coherent risk measures of Artzner et al. (1999). By (ii), DQ and DR agree on whether diversification benefit exists under mild conditions, and this intuition is consistent with Artzner et al. (1999).

*Remark* 4. We require $\rho$ to be left continuous and non-flat from the left to make the inequality in (ii) holds strictly. This requirement excludes, in particular, trivial cases like $\mathbf{X} = \mathbf{c} \in \mathbb{R}^n$ which gives $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = 0$ by definition. In case the conditions fail to hold, $\mathrm{DQ}_\alpha^{\rho}(\mathbf{X}) < 1$ may not guarantee $\rho_\alpha\left(\sum_{i=1}^n X_i\right) < \sum_{i=1}^n \rho_\alpha(X_i)$, but it implies the non-strict inequality $\rho_\alpha\left(\sum_{i=1}^n X_i\right) \leqslant \sum_{i=1}^n \rho_\alpha(X_i)$, and thus the portfolio risk is not worse than the sum of the individual risks.

## 4.2 Stochastic dominance and dependence

In this section, we discuss the consistency of DQ with respect to stochastic dominance, as well as the best and worst cases for DQ among all dependence structures with given marginal distributions of the risk vector.

For a diversification index, monotonicity with respect to stochastic dominance yields consistency with common decision-making criteria such as the expected utility model and the rank-dependent utility model. A random variable $X$ (representing random loss) is dominated by a random variable $Y$ in second-order stochastic dominance (SSD) if $\mathbb{E}[f(X)] \leqslant \mathbb{E}[f(Y)]$ for all decreasing concave functions $f : \mathbb{R} \to \mathbb{R}$ provided that the expectations exist, and we denote this by $X \leqslant_{\mathrm{SSD}} Y$.[12] A risk measure $\phi$ is *SSD-consistent* if $\phi(X) \geqslant \phi(Y)$ for all $X, Y \in \mathcal{X}$ whenever $X \leqslant_{\mathrm{SSD}} Y$. SSD consistency is known as strong risk aversion in the classic decision theory literature (Rothschild and Stiglitz (1970)). SSD-consistent monetary risk measures, which include all law-invariant convex risk measures such as ES, admit an ES-based characterization (Mao and Wang (2020)).

**Proposition 3.** *Assume that $\rho = (\rho_\alpha)_{\alpha \in I}$ is a decreasing class of SSD-consistent risk measures. For $\mathbf{X}, \mathbf{Y} \in \mathcal{X}^n$ and $\alpha \in I$, if $\sum_{i=1}^n \rho_\alpha(X_i) \leqslant \sum_{i=1}^n \rho_\alpha(Y_i)$ and $\sum_{i=1}^n X_i \leqslant_{\mathrm{SSD}} \sum_{i=1}^n Y_i$, then $\mathrm{DQ}_\alpha^{\rho}(\mathbf{X}) \geqslant \mathrm{DQ}_\alpha^{\rho}(\mathbf{Y})$.*

Proposition 3 follows from the simple observation that

$$\left\{\beta \in I : \rho_\beta\left(\sum_{i=1}^n X_i\right) \leqslant \sum_{i=1}^n \rho_\alpha(X_i)\right\} \subseteq \left\{\beta \in I : \rho_\beta\left(\sum_{i=1}^n Y_i\right) \leqslant \sum_{i=1}^n \rho_\alpha(Y_i)\right\},$$

and we omit the proof.

Assume $\rho$ is a class of SSD-consistent risk measures (e.g., law-invariant convex risk measures). Proposition 3 implies that if the sum of marginal risks is the same for $\mathbf{X}$ and $\mathbf{Y}$ (this holds in particular if $\mathbf{X}$ and $\mathbf{Y}$ have the same marginal distributions), then DQ is decreasing in SSD of the total risk. The dependence structures which maximize or minimize DQ for $\mathbf{X}$ with specified marginal distributions are discussed in Appendix D.1. For instance, a comonotonic portfolio has the largest DQ (thus the smallest diversification) among all portfolios with the same marginal distributions; this observation is related to Proposition 2 (iii) and Theorem 3 (iv).

[12] If $X$ and $Y$ represent gains instead of losses, then SSD is typically defined via increasing concave functions.

### 4.3 Consistency across dimensions

All properties in the previous sections are discussed under the assumption that the dimension $n \in \mathbb{N}$ is fixed. Letting $n$ vary allows for a comparison of diversification between portfolios with different dimensions. In this section, we slightly generalize our framework by considering a diversification index $D$ as a mapping on $\bigcup_{n\in\mathbb{N}} \mathcal{X}^n$; note that the input vector $\mathbf{X}$ of DQ and DR can naturally have any dimension $n$. We present two more useful properties of DQ in this setting. For $\mathbf{X} \in \mathcal{X}^n$ and $c \in \mathbb{R}$, $(\mathbf{X}, c)$ is the $(n+1)$-dimensional random vector obtained by pasting $\mathbf{X}$ and $c$, and $(\mathbf{X}, \mathbf{X})$ is the $(2n)$-dimensional random vector obtained by pasting $\mathbf{X}$ and $\mathbf{X}$.

[RI] Riskless invariance: $D(\mathbf{X}, c) = D(\mathbf{X})$ for all $n \in \mathbb{N}$, $\mathbf{X} \in \mathcal{X}^n$ and $c \in \mathbb{R}$.

[RC] Replication consistency: $D(\mathbf{X}, \mathbf{X}) = D(\mathbf{X})$ for all $n \in \mathbb{N}$ and $\mathbf{X} \in \mathcal{X}^n$.

Riskless invariance means that adding a risk-free asset into the portfolio $\mathbf{X}$ does not affect its diversification. For instance, the Sharpe ratio of the portfolio does not change under such an operation. Replication consistency means that replicating the same portfolio composition does not affect $D$. Both properties are arguably desirable in most applications due to their natural interpretations.

**Proposition 4.** *Let $\rho = (\rho_\alpha)_{\alpha\in I}$ be a class of risk measures decreasing in $\alpha$. For $\alpha \in I$,*

*(i) If $\rho_\beta$ satisfies* [CA]$_\mathrm{m}$ *with $m \in \mathbb{R}$ for $\beta \in I$ and $\rho_\alpha(0) = 0$ then $\mathrm{DQ}^\rho_\alpha$ satisfies* [RI].

*(ii) If $\rho_\beta$ satisfies* [PH] *for $\beta \in I$, then $\mathrm{DQ}^\rho_\alpha$ satisfies* [RC].

We further show in Proposition EC.7 that if [RI] is assumed, then the only option for DR is to use a non-negative $\phi$ (which is a subclass of DQ). Thus, if [RI] is considered as desirable, then DQ becomes useful compared to DR as it offers more choices, and in particular, it works for any classes $\rho$ of monetary risk measures with $\rho_\alpha(0) = 0$ including VaR and ES. Both DQ and DR satisfy [RC] and [RI] for MCP risk measures.

**Example 2.** Let $\phi$ be a risk measure satisfying [CA], such as $\mathrm{ES}_\alpha$ or $\mathrm{VaR}_\alpha$. Suppose that $\phi(\sum_{i=1}^n X_i) = 1$ and $\sum_{i=1}^n \phi(X_i) = 2$, and thus $\mathrm{DR}^\phi(\mathbf{X}) = 1/2$. If a non-random payoff of $c > 0$ is added to the portfolio, then $\mathrm{DR}^\phi(\mathbf{X}, -c) = (1-c)/(2-c)$, which turns to 0 as $c \uparrow 1$, and it becomes negative as soon as $c > 1$. Hence, $\mathrm{DR}^\phi$ is improved or made negative by including constant payoffs (either as a new asset or added to an existing asset). This creates problematic incentives in optimization. On the other hand, DQ does not suffer from this problem due to [LI] and [RI].

# 5 DQ based on the classes of VaR and ES

Since VaR and ES are the two most common classes of risk measures in practice, we focus on the theoretical properties of $\mathrm{DQ}^{\mathrm{VaR}}_{\alpha}$ and $\mathrm{DQ}^{\mathrm{ES}}_{\alpha}$ in this section. We fix the parameter range $I=(0,1)$, and we choose $\mathcal{X}^n$ to be $(L^0)^n$ when we discuss $\mathrm{DQ}^{\mathrm{VaR}}_{\alpha}$ and $(L^1)^n$ when we discuss $\mathrm{DQ}^{\mathrm{ES}}_{\alpha}$, but all results hold true if we fix $\mathcal{X}=L^1$.

## 5.1 General properties

We first provide alternative formulations of $\mathrm{DQ}^{\mathrm{VaR}}_{\alpha}$ and $\mathrm{DQ}^{\mathrm{ES}}_{\alpha}$. The formulations offer clear interpretations and simple ways to compute the values of DQs. The formula (6) below can be derived from the optimization formulation for the buffered probability of exceedance in Proposition 2.2 of Mafusalov and Uryasev (2018).

**Theorem 4.** *For a given $\alpha\in(0,1)$, $\mathrm{DQ}^{\mathrm{VaR}}_{\alpha}$ and $\mathrm{DQ}^{\mathrm{ES}}_{\alpha}$ have the alternative formulas*

$$\mathrm{DQ}^{\mathrm{VaR}}_{\alpha}(\mathbf{X})=\frac{1}{\alpha}\mathbb{P}\left(\sum_{i=1}^n X_i>\sum_{i=1}^n \mathrm{VaR}_{\alpha}(X_i)\right),\quad \mathbf{X}\in\mathcal{X}^n, \tag{4}$$

*and*

$$\mathrm{DQ}^{\mathrm{ES}}_{\alpha}(\mathbf{X})=\frac{1}{\alpha}\mathbb{P}\left(Y>\sum_{i=1}^n \mathrm{ES}_{\alpha}(X_i)\right),\quad \mathbf{X}\in\mathcal{X}^n, \tag{5}$$

*where $Y=\mathrm{ES}_U(\sum_{i=1}^n X_i)$ and $U\sim\mathrm{U}[0,1]$. Furthermore, if $\mathbb{P}(\sum_{i=1}^n X_i>\sum_{i=1}^n \mathrm{ES}_{\alpha}(X_i))>0$, then*

$$\mathrm{DQ}^{\mathrm{ES}}_{\alpha}(\mathbf{X})=\frac{1}{\alpha}\min_{r\in(0,\infty)}\mathbb{E}\left[\left(r\sum_{i=1}^n (X_i-\mathrm{ES}_{\alpha}(X_i))+1\right)_+\right], \tag{6}$$

*and otherwise $\mathrm{DQ}^{\mathrm{ES}}_{\alpha}(\mathbf{X})=0$.*

As a first observation from Theorem 4, it is straightforward to compute $\mathrm{DQ}^{\mathrm{VaR}}_{\alpha}$ and $\mathrm{DQ}^{\mathrm{ES}}_{\alpha}$ on real or simulated data by applying (4) and (5) to the empirical distribution of the data.

Theorem 4 also gives $\mathrm{DQ}^{\mathrm{VaR}}_{\alpha}$ a clear economic interpretation as the improvement of insolvency probability when risks are pooled, making the discussion in Section 3.4 more concrete. Suppose that $X_1,\dots,X_n$ are continuously distributed and they represent losses from $n$ assets. The total pooled capital is $s_\alpha=\sum_{i=1}^n \mathrm{VaR}_{\alpha}(X_i)$, which is determined by the marginals of $\mathbf{X}$ but not the dependence structure. An agent investing only in asset $X_i$ with capital computed by $\mathrm{VaR}_\alpha$ has an insolvency probability $\alpha=\mathbb{P}(X_i>\mathrm{VaR}_{\alpha}(X_i))$. On the other hand, by Theorem 4, $\alpha^*$ is the probability that the pooled loss $\sum_{i=1}^n X_i$ exceeds the pooled capital $s_\alpha$. The improvement from $\alpha$ to $\alpha^*$, computed by $\alpha^*/\alpha$, is precisely $\mathrm{DQ}^{\mathrm{VaR}}_{\alpha}(\mathbf{X})$. From here, it is also clear that $\mathrm{DQ}^{\mathrm{VaR}}_{\alpha}(\mathbf{X})<1$ is equivalent to $\mathbb{P}\left(\sum_{i=1}^n X_i>s_\alpha\right)<\alpha$.

To compare $\mathrm{DQ}^{\mathrm{VaR}}_{\alpha}$ with $\mathrm{DR}^{\mathrm{VaR}_{\alpha}}$, recall that the two diversification indices can be rewritten as

$$\mathrm{DQ}^{\mathrm{VaR}}_{\alpha}(\mathbf{X}) = \frac{\mathbb{P}\left(\sum_{i=1}^{n} X_i > s_{\alpha}\right)}{\alpha} \quad \text{and} \quad \mathrm{DR}^{\mathrm{VaR}_{\alpha}}(\mathbf{X}) = \frac{\mathrm{VaR}_{\alpha}\left(\sum_{i=1}^{n} X_i\right)}{s_{\alpha}}. \tag{7}$$

From (7), we can see a clear symmetry between DQ, which measures the probability improvement, and DR, which measures the quantile improvement. DQ and DR based on ES have a similar comparison.

The range of DQ based on VaR is different from that based on ES, which is $[0,1]$ by Proposition 1. We summarize them below.

**Proposition 5.** *For $\alpha \in (0,1)$ and $n \geqslant 2$, $\{\mathrm{DQ}^{\mathrm{VaR}}_{\alpha}(\mathbf{X}) : \mathbf{X} \in \mathcal{X}^n\} = [0, \min\{n, 1/\alpha\}]$ and $\{\mathrm{DQ}^{\mathrm{ES}}_{\alpha}(\mathbf{X}) : \mathbf{X} \in \mathcal{X}^n\} = [0,1]$.*

Both $\mathrm{DQ}^{\mathrm{VaR}}_{\alpha}$ and $\mathrm{DQ}^{\mathrm{ES}}_{\alpha}$ take values on a bounded interval. In contrast, the diversification ratio $\mathrm{DR}^{\mathrm{VaR}_{\alpha}}$ is unbounded, and $\mathrm{DR}^{\mathrm{ES}_{\alpha}}$ is bounded above by 1 only when the ES of the total risk is non-negative.

*Remark* 5. It is a coincidence that $\mathrm{DQ}^{\mathrm{VaR}}_{\alpha}$ for $\alpha < 1/n$ and $\mathrm{DR}^{\mathrm{var}}$ both have a maximum value $n$. The latter maximum value is attained by a risk vector $(X/n, \dots, X/n)$ for any $X \in L^2$.

### 5.2 Capturing heavy tails and common shocks

In this section, we analyze three simple normal and t-models to illustrate some features of DQ regarding heavy tails and common shocks in the portfolio models. Here, we only present some key observations. A detailed study of DQs based on VaR and ES for elliptical distributions and multivariate regularly varying models, including explicit formulas to compute DQ for these models, can be found in Han et al. (2023).

Let $\mathbf{Z} = (Z_1, \dots, Z_n)$ be an $n$-dimensional standard normal random vector, and let $\xi^2$ have an inverse gamma distribution independent of $\mathbf{Z}$. Denote by $\mathrm{it}_n(\nu)$ the joint distribution with $n$ independent t-marginals $\mathrm{t}(\nu, 0, 1)$, where the parameter $\nu$ represents the degrees of freedom; see McNeil et al. (2015) for t-models. The model $\mathbf{Y} = (Y_1, \dots, Y_n) \sim \mathrm{it}_n(\nu)$ can be stochastically represented by

$$Y_i = \xi_i Z_i, \quad \text{for } i \in [n], \tag{8}$$

where $\xi_1, \dots, \xi_n$ are iid following the same distribution as $\xi$, and independent of $\mathbf{Z}$. In contrast, a joint t-distributed random vector $\mathbf{Y}' = (Y_1', \dots, Y_n') \sim \mathrm{t}(\nu, \mathbf{0}, I_n)$ has a stochastic representation $\mathbf{Y}' = \xi \mathbf{Z}$, that is,

$$Y_i' = \xi Z_i, \quad \text{for } i \in [n]. \tag{9}$$

In other words, $\mathbf{Y}'$ is a standard normal random vector multiplied by a heavy-tailed common shock $\xi$. All three models $\mathbf{Z}, \mathbf{Y}, \mathbf{Y}'$ have the same correlation matrix, the identity matrix $I_n$.

Table 2. DQs/DRs based on VaR, ES, SD and var, where $\alpha = 0.05$, $n = 10$ and $\nu = 3$; numbers in bold indicate the most diversified among $\mathbf{Z}, \mathbf{Y}, \mathbf{Y}'$ according to the index $D$

| $D$ | $\mathrm{DQ}_{\alpha}^{\mathrm{VaR}}$ | $\mathrm{DQ}_{\alpha}^{\mathrm{ES}}$ | $\mathrm{DR}^{\mathrm{VaR}_{\alpha}}$ | $\mathrm{DR}^{\mathrm{ES}_{\alpha}}$ | $\mathrm{DR}^{\mathrm{SD}}$ | $\mathrm{DR}^{\mathrm{var}}$ |
|---|---|---|---|---|---|---|
| $\mathbf{Z} \sim \mathrm{N}(\mathbf{0}, I_n)$ | $\mathbf{2.0 \times 10^{-6}}$ | $\mathbf{1.9 \times 10^{-9}}$ | **0.3162** | 0.3162 | **0.3162** | **1** |
| $\mathbf{Y} \sim \mathrm{it}_n(3)$ | 0.0235 | 0.0124 | 0.3569 | **0.2903** | **0.3162** | **1** |
| $\mathbf{Y}' \sim \mathrm{t}(3, \mathbf{0}, I_n)$ | 0.0502 | 0.0340 | **0.3162** | 0.3162 | **0.3162** | **1** |
| $D(\mathbf{Z}) < D(\mathbf{Y})$ | Yes | Yes | Yes | No | No | No |
| $D(\mathbf{Y}) < D(\mathbf{Y}')$ | Yes | Yes | No | Yes | No | No |

Table 3. DQs/DRs based on VaR, ES, SD and var, where $\alpha = 0.05$, $n = 10$ and $\nu = 4$; numbers in bold indicate the most diversified among $\mathbf{Z}, \mathbf{Y}, \mathbf{Y}'$ according to the index $D$

| $D$ | $\mathrm{DQ}_{\alpha}^{\mathrm{VaR}}$ | $\mathrm{DQ}_{\alpha}^{\mathrm{ES}}$ | $\mathrm{DR}^{\mathrm{VaR}_{\alpha}}$ | $\mathrm{DR}^{\mathrm{ES}_{\alpha}}$ | $\mathrm{DR}^{\mathrm{SD}}$ | $\mathrm{DR}^{\mathrm{var}}$ |
|---|---|---|---|---|---|---|
| $\mathbf{Z} \sim \mathrm{N}(\mathbf{0}, I_n)$ | $\mathbf{2.0 \times 10^{-6}}$ | $\mathbf{1.9 \times 10^{-9}}$ | **0.3162** | 0.3162 | **0.3162** | **1** |
| $\mathbf{Y} \sim \mathrm{it}_n(4)$ | 0.0050 | 0.0017 | 0.3415 | **0.2828** | **0.3162** | **1** |
| $\mathbf{Y}' \sim \mathrm{t}(4, \mathbf{0}, I_n)$ | 0.0252 | 0.0138 | **0.3162** | 0.3162 | **0.3162** | **1** |
| $D(\mathbf{Z}) < D(\mathbf{Y})$ | Yes | Yes | Yes | No | No | No |
| $D(\mathbf{Y}) < D(\mathbf{Y}')$ | Yes | Yes | No | Yes | No | No |

Because of the common shock $\xi$ in (9), large losses from components of $\mathbf{Y}'$ are more likely to occur simultaneously, compared to $\mathbf{Y}$ in (8), which does not have a common shock. Indeed, $\mathbf{Y}'$ is tail dependent (Example 7.39 of McNeil et al. (2015)) whereas $\mathbf{Y}$ is tail independent. As such, at least intuitively (if not rigorously), diversification for portfolio $\mathbf{Y}'$ should be considered as weaker than $\mathbf{Y}$, although both models are uncorrelated and have the same marginals.[13] By the central limit theorem, for $\nu > 2$, the component-wise average of $\mathbf{Y}$ (scaled by its variance) is asymptotically normal as $n$ increases, whereas the component-wise average of $\mathbf{Y}'$ is always t-distributed. Hence, one may intuitively expect the order $D(\mathbf{Z}) < D(\mathbf{Y}) < D(\mathbf{Y}')$ to hold.

In Tables 2 and 3, we present DQ and DR for a few different models based on $\mathrm{N}(\mathbf{0}, I_n)$, $\mathrm{t}(\nu, \mathbf{0}, I_n)$, and $\mathrm{it}_n(\nu)$. We choose $n = 10$ and $\nu = 3$ or 4,[14] and thus we have five models in total. As we see from Tables 2 and 3, DQs based on both VaR and ES report a lower value for $\mathrm{it}_n(\nu)$ and a larger value for $\mathrm{t}(\nu, \mathbf{0}, I_n)$, meaning that diversification is weaker for the common

[13] On a related note, as discussed by Embrechts et al. (2002), correlation is not a good measure of diversification in the presence of heavy-tailed and skewed distributions.

[14] Most financial asset log-loss data have a tail-index between $[3, 5]$, which corresponds to $\nu \in [3, 5]$; see e.g., Jansen and De Vries (1991).

shock t-model (9) than the iid t-model (8). For the iid normal model, the diversification is the strongest according to DQ. In contrast, DR sometimes reports that the iid t-model has a larger diversification than the common shock t-model, which is counter-intuitive. In the setting of both Tables 2 and 3, a risk manager governed by $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ would prefer the iid portfolio over the common shock portfolio, but the preference is flipped if the risk manager uses $\mathrm{DR}^{\mathrm{VaR}_\alpha}$. A more detailed analysis on this phenomenon for varying $\alpha \in (0, 0.1]$ is presented in Figure EC.1 in Appendix E, and consistent results are observed.

# 6 Portfolio selection with DQ

Next, we focus on the optimal diversification problem

$$\min_{\mathbf{w}\in\Delta_n} \mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{w}\odot\mathbf{X}) \quad \text{and} \quad \min_{\mathbf{w}\in\Delta_n} \mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{w}\odot\mathbf{X}); \tag{10}$$

recall that a smaller value of DQ means better diversification.[15] Recall from Table 1 that the first optimization is not quasi-convex and the second one is quasi-convex (Proposition 1). We do not say that optimizing a diversification index has a decision-theoretic benefit; here we simply illustrate the advantage of DQ in computation and optimization. Whether optimizing diversification is desirable for individual or institutional investors is an open-ended question which goes beyond the current paper; we refer to Van Nieuwerburgh and Veldkamp (2010), Boyle et al. (2012) and Choi et al. (2017) for relevant discussions.

For the portfolio weight $\mathbf{w}$, DQ based on VaR at level $\alpha \in (0,1)$ is given by

$$\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{w}\odot\mathbf{X}) = \frac{1}{\alpha}\inf\left\{\beta\in(0,1) : \mathrm{VaR}_\beta\left(\sum_{i=1}^n w_i X_i\right) \leqslant \sum_{i=1}^n w_i \mathrm{VaR}_\alpha(X_i)\right\},$$

and DQ based on ES is similar. In what follows, we fix $\alpha \in (0,1)$ and $\mathbf{X} = (X_1, \dots, X_n) \in \mathcal{X}^n$, where $\mathcal{X}$ is $L^0$ for VaR and $L^1$ for ES, as in Section 5. Write $\mathbf{0} = (0, \dots, 0) \in \mathbb{R}^n$ and $\mathbf{x}_\alpha^\rho = (\rho_\alpha(X_1), \dots, \rho_\alpha(X_n))$ for a given risk measure $\rho$.

**Proposition 6.** *Fix $\alpha \in (0,1)$ and $\mathbf{X} \in \mathcal{X}^n$. The optimization of $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{w}\odot\mathbf{X})$ in (10) can be solved by*

$$\min_{\mathbf{w}\in\Delta_n} \mathbb{P}\left(\mathbf{w}^\top\left(\mathbf{X} - \mathbf{x}_\alpha^{\mathrm{VaR}}\right) > 0\right). \tag{11}$$

[15] A possible alternative formulation to (10) is to use DQ as a constraint instead of an objective in the optimization. This is mathematically similar to a risk measure constraint (e.g., Basak and Shapiro (2001), Rockafellar and Uryasev (2002) and Mafusalov and Uryasev (2018)), but with a different interpretation, as DQ is not designed to measure or control risk.

*Assuming $\mathbb{P}(X_i > \mathrm{ES}_\alpha(X_i)) > 0$ for each $i \in [n]$, the optimization of $\mathrm{DQ}^{\mathrm{ES}}_\alpha(\mathbf{w} \odot \mathbf{X})$ in (10) can be solved by the convex program*

$$\min_{\mathbf{v} \in \mathbb{R}^n_+ \setminus \{\mathbf{0}\}} \mathbb{E}\left[\left(\mathbf{v}^\top \left(\mathbf{X} - \mathbf{x}^{\mathrm{ES}}_\alpha\right) + 1\right)_+\right], \tag{12}$$

*and the optimal $\mathbf{w}^*$ is given by $\mathbf{v}/\|\mathbf{v}\|_1$.*

Proposition 6 offers efficient algorithms to optimize $\mathrm{DQ}^{\mathrm{VaR}}_\alpha$ and $\mathrm{DQ}^{\mathrm{ES}}_\alpha$ in real-data applications. The values of $\mathbf{x}^{\mathrm{VaR}}_\alpha$ and $\mathbf{x}^{\mathrm{ES}}_\alpha$ can be computed by many existing estimators of the individual losses (see e.g., McNeil et al. (2015)). In particular, a simple way to estimate these risk measures is to use an empirical estimator. More specifically, if we have data $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(N)}$ sampled from $\mathbf{X}$ satisfying some ergodicity condition (being iid would be sufficient), then the empirical version of the problem (11) is

$$\text{minimize} \quad \sum_{j=1}^N \mathbb{1}_{\left\{\mathbf{w}^\top\left(\mathbf{X}^{(j)} - \widehat{\mathbf{x}}^{\mathrm{VaR}}_\alpha\right) > 0\right\}} \quad \text{over } \mathbf{w} \in \Delta_n, \tag{13}$$

where $\widehat{\mathbf{x}}^{\mathrm{VaR}}_\alpha$ is the empirical estimator of $\mathbf{x}^{\mathrm{VaR}}_\alpha$ based on sample $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(N)}$; see McNeil et al. (2015). Write $\mathbf{y}^{(j)} = \mathbf{X}^{(j)} - \widehat{\mathbf{x}}^{\mathrm{VaR}}_\alpha$ and $z_j = \mathbb{1}_{\{\mathbf{w}^\top \mathbf{y}^{(j)} > 0\}}$ for $j \in [n]$. Problem (13) involves a chance constraint (see e.g., Luedtke (2014) and Liu et al. (2016)). By using the big-M method (see e.g., Shen et al. (2010)) via choosing a sufficient large $M$ (e.g., it is sufficient if $M$ is larger than the components of $\mathbf{y}^{(j)}$ for all $j$), (13) can be converted into the following linear integer program:

$$\begin{aligned} &\text{minimize} && \textstyle\sum_{j=1}^N z_j \\ &\text{subject to} && \mathbf{w}^\top \mathbf{y}^{(j)} - M z_j \leqslant 0, \quad \textstyle\sum_{i=1}^n w_i = 1, \\ &&& z_j \in \{0,1\}, \quad w_i \geqslant 0 \quad \text{for all } j \in [N] \text{ and } i \in [n]. \end{aligned} \tag{14}$$

Similarly, the optimization problem (12) for $\mathrm{DQ}^{\mathrm{ES}}_\alpha$ can be solved the empirical version of the problem (12), which is a convex program:

$$\text{minimize} \quad \sum_{j=1}^N \max\left\{\mathbf{v}^\top\left(\mathbf{X}^{(j)} - \widehat{\mathbf{x}}^{\mathrm{ES}}_\alpha\right) + 1, 0\right\} \quad \text{over } \mathbf{v} \in \mathbb{R}_+, \tag{15}$$

where $\widehat{\mathbf{x}}^{\mathrm{ES}}_\alpha$ is the empirical estimator of $\mathbf{x}^{\mathrm{ES}}_\alpha$ based on sample $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(N)}$. Both problems (14) and (15) can be efficiently solved by modern optimization programs, such as CVX programming (see e.g., Matmoura and Penev (2013)).

Additional linear constraints, such as those on budget or expected return, can be easily included in (11)-(15), and the corresponding optimization problems can be solved similarly.

Tie-breaking needs to be addressed when working with (13) since its objective function takes integer values. In dynamic portfolio selection, it is desirable to avoid adjusting positions too drastically or frequently. Therefore, in the real-data analysis in Section 7.3, among tied optimizers, we pick the closest one (in $L^1$-norm $\|\cdot\|_1$ on $\mathbb{R}^n$) to a given benchmark $\mathbf{w}_0$, the portfolio weight of the previous trading period. With this tie-breaking rule, we solve

$$\text{minimize} \quad \|\mathbf{w}-\mathbf{w}_0\|_1 \quad \text{over } \mathbf{w}\in\Delta_n \quad \text{subject to} \quad \sum_{j=1}^{N} \mathbb{1}_{\{\mathbf{w}^\top \mathbf{y}^{(j)}>0\}} \leqslant m^*, \tag{16}$$

where $m^*$ is the optimum of (13). A tie-breaking for (15) may need to be addressed similarly since (15) is not strictly convex.

# 7 Numerical illustrations

To illustrate the performance of DQ, we collect historical asset prices from Yahoo Finance and conduct three sets of numerical experiments based on the data. We use the period from January 3, 2012, to December 31, 2021, with a total of 2518 observations of daily losses and 500 trading days for the initial training. In Section 7.1, we first compare DQs and DRs based on VaR and ES. In Section 7.2, we calculate the values of $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ and $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ under different selections of stocks. Finally, we construct portfolios by minimizing $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$, $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ and $\mathrm{DR}^{\mathrm{SD}}$ and by the mean-variance criterion in Section 7.3.

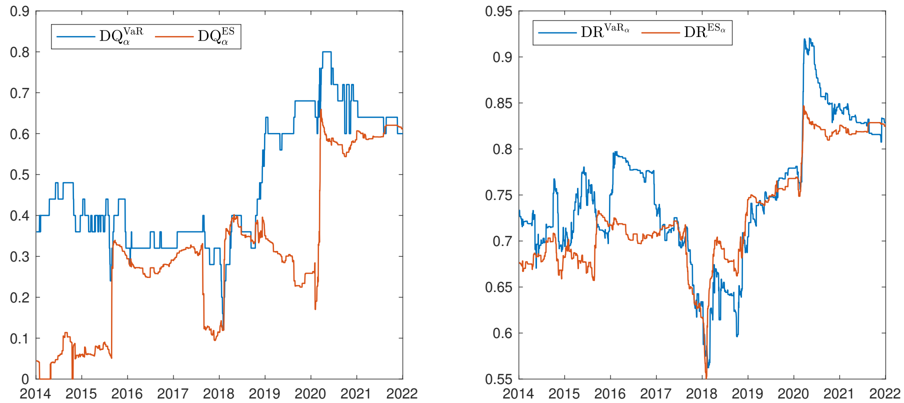
## 7.1 Comparing DQ and DR

We first identify the largest stock in each of the S&P 500 sectors ranked by market cap in 2012. Among these stocks, we select the 5 largest stocks[16] to build our portfolio. We compute $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$, $\mathrm{DQ}_\alpha^{\mathrm{ES}}$, $\mathrm{DR}^{\mathrm{VaR}_\alpha}$ and $\mathrm{DR}^{\mathrm{ES}_\alpha}$ on each day using the empirical distribution in a rolling window of 500 days, where we set $\alpha = 0.05$.

Figure 2 shows that the values of DQ and DR are between 0 and 1. This corresponds to the observation in Theorem 3 that $\mathrm{DQ}_\alpha^{\rho} < 1$ is equivalent to $\mathrm{DR}^{\rho_\alpha} < 1$. DQ has a similar temporal pattern to DR in the above period of time, with a large jump when COVID-19 exploded, which is more visible for DQ than for DR. We remind the reader that DQ and DR are not meant to be compared on the same scale, and hence the fact that DQ has a larger range than DR should be taken lightly. We also note that the values of $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ are in discrete grids. This is because the empirical distribution function takes value in multiples of $1/N$ there $N$ is the sample size (500 in this experiment) and hence $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ takes the values $k/(N\alpha)$ for an integer $k$; see (4). If a smooth curve is preferred, then one can employ a smoothed VaR through linear interpolation. This is a standard technique for handling VaR; see McNeil et al. (2015, Section 9.2.6) and Li and Wang (2022, Remark 8 and Appendix B).

[16]XOM from ENR, AAPL from IT, BRK/B from FINL, WMT from CONS, and GE from INDU.

Figure 2. DQs and DRs based on VaR and ES with $\alpha = 0.05$




## 7.2 DQ for different portfolios

In this section, we fix $\alpha = 0.05$ and calculate the values of $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ and $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ under different portfolio compositions of stocks. We consider portfolios with the following stock compositions:

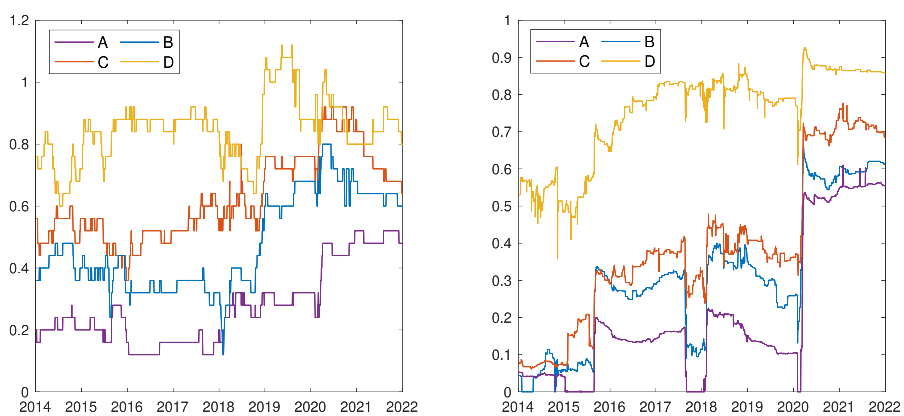
(A) the two largest stocks from each of the 10 different sectors of S&P 500;

(B) the largest stock from each of 5 different sectors of S&P 500 (as in Section 7.1);

(C) the 5 largest stocks, AAPL, MSFT, IBM, GOOGL and ORCL, from the Information Technology (IT) sector;

(D) the 5 largest stocks, BRK/B, WFC, JPM, C and BAC, from the Financials (FINL) sector.

Figure 3. DQs based on VaR (left) and ES (right) with $\alpha = 0.05$



We make a few observations from Figure 3. Both $\mathrm{DQ}^{\mathrm{VaR}}_{\alpha}$ and $\mathrm{DQ}^{\mathrm{ES}}_{\alpha}$ provide similar comparative results. The order (A)$\leqslant$(B)$\leqslant$(C)$\leqslant$(D) is consistent with our intuition.[17] First, portfolio (A) of 20 stocks has the strongest diversification effect among the four compositions. Second, portfolio (B) across 5 sectors has stronger diversification than (C) and (D) within one sector. Third, portfolio (C) of 5 stocks within the IT sector has a stronger diversification than portfolio (D) of 5 stocks within the FINL sector, consistent with the fact that the stocks in the IT sector are less correlated. Moreover, $\mathrm{DQ}^{\mathrm{VaR}}_{\alpha}$ for the FINL sector is larger than 1 during some period of time, which means that there is no diversification benefit if risk is evaluated by VaR. All DQ curves based on ES show a large up-ward jump around the COVID-19 outbreak; such a jump also exists for curves based on VaR but it is less pronounced.

## 7.3 Optimal diversified portfolios

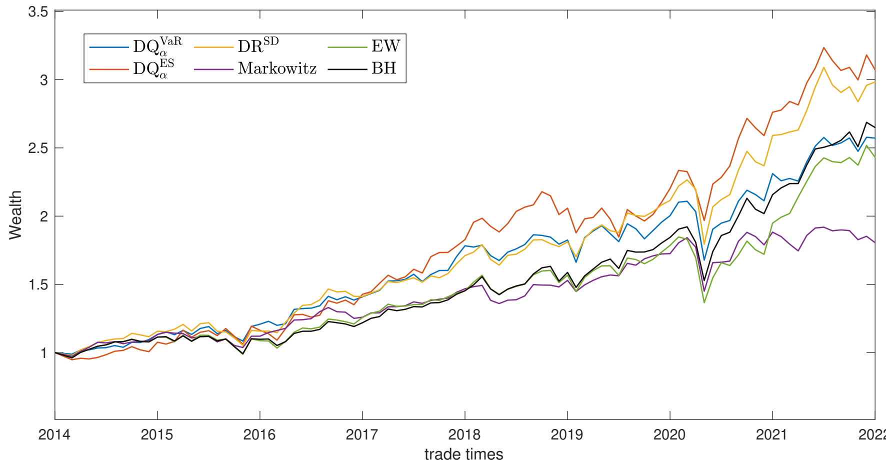
In this section, we fix $\alpha = 0.1$ and build portfolios via $\mathrm{DQ}^{\mathrm{VaR}}_{\alpha}$, $\mathrm{DQ}^{\mathrm{ES}}_{\alpha}$, $\mathrm{DR}^{\mathrm{SD}}$, and the mean-variance criterion in the Markowitz (1952) model.[18] The optimal portfolio problems using $\mathrm{DR}^{\mathrm{SD}}$ and the Markowitz model are well studied in literature; see e.g. Choueifaty and Coignard (2008). We compare these portfolio wealth with the equal weighted (EW) portfolio and the simple buy-and-hold (BH) portfolio. For an analysis on the EW strategy, see DeMiguel et al. (2009).

We apply the algorithms in Proposition 6 to optimize $\mathrm{DQ}^{\mathrm{VaR}}_{\alpha}$ and $\mathrm{DQ}^{\mathrm{ES}}_{\alpha}$, which are extremely fast. A tie-breaking is addressed for each objective as in (16). Minimization of $\mathrm{DR}^{\mathrm{SD}}$ and the Markowitz model can be solved by existing algorithms. The initial wealth is set to 1, and the risk-free rate is $r = 2.84\%$, which is the 10-year yield of the US treasury bill in Jan 2014. The target annual expected return for the Markowitz portfolio is set to 10%. We optimize the portfolio weights in each month with a rolling window of 500 days. That is, in each month, roughly 21 trading days, starting from January 2, 2014, we use the preceding 500 trading days to compute the optimal portfolio weights using the method described above. The portfolio is rebalanced every month. We choose the 4 largest stocks from each of the 10 different sectors of S&P 500 ranked by market cap in 2012 as the portfolio compositions (40 stocks in total). The portfolio performance is reported in Figure 4, and the cumulative distribution of the sorted portfolio weights, averaged over each month, is shown in Figure 5. Summary statistics, including the annualized return (AR), the annualized volatility (AV), the Sharpe ratio (SR), and the average

[17] The observations here are consistent with those from applying $\mathrm{DR}^{\mathrm{SD}}$ (which is also a DQ) in the same setting; see Appendix G.

[18] One may try other portfolio criteria other than mean-variance. For instance, Levy and Levy (2004) found that portfolio strategies based on prospect theory perform similarly to the mean-variance strategies.

Figure 4. Wealth processes for portfolios, 40 stocks, Jan 2014 - Dec 2021




trading proportion (ATP), are reported in Table 4.[19]

From these results, we can see that the portfolio optimization strategies based on minimizing DQ perform quite well, similarly to those based on $DR^{SD}$, and better than the Markowitz strategy. Moreover, ATP and portfolio weight distribution are similar across the strategies based on the three diversification indices and the Markowitz strategy. In contrast, the EW and BH strategies have more uniform portfolio weight distributions and smaller ATP, as anticipated. We remark that it is not our intention to analyze which diversification strategy generates the highest return, which is a challenging question that needs a separate study; also, we do not suggest diversification should or should not be optimized in practice. The empirical results here are presented to illustrate how our proposed diversification indices work in the context of portfolio selection. More empirical results with some other datasets and portfolio strategies are given in Appendix G, and the results show similar patterns.

# 8 Concluding remarks

In this paper, we put forward six axioms to jointly characterize a new class of indices of diversification, and a seventh axiom to specialize this class. The new diversification index DQ has favourable features both theoretically and practically, and it is contrasted with its competitors,

[19]ATP is an approximation of trading costs, and it is computed as the average of $\sum_{t=1}^{T} |w_i^t - w_i^{t-}|$ over $i = 1, \dots, n$, where $T = 96$ is the total number of months, $w_i^t$ is the portfolio weight of asset $i$ at the beginning of month $t$, and $w_i^{t-}$ is the portfolio weight of asset $i$ at the end of month $t-1$, with $w_i^{1-}$ set to $w_i^1$. Note that for BH, ATP is 0 because there is no trading, whereas for EW, ATP is positive, as rebalancing occurs at the end of each month.

Figure 5. Cumulative portfolio weights, 40 stocks, Jan 2014 - Dec 2021

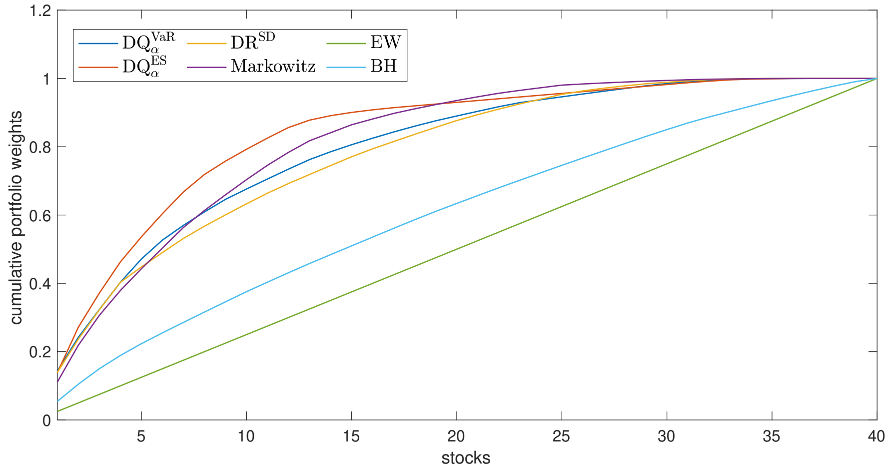



Table 4. Annualized return (AR), annualized volatility (AV), Sharpe ratio (SR), and average trading proportion (ATP) for different portfolio strategies from Jan 2014 to Dec 2021

| % | $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ | $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ | $\mathrm{DR}^{\mathrm{SD}}$ | Markowitz | EW | BH |
|---|---|---|---|---|---|---|
| AR | 12.56 | 14.59 | 14.36 | 7.93 | 11.91 | 12.88 |
| AV | 14.64 | 15.74 | 14.99 | 12.98 | 15.92 | 14.34 |
| SR | 66.40 | 74.66 | 76.85 | 39.22 | 56.95 | 70.02 |
| ATP | 19.29 | 14.75 | 15.61 | 18.79 | 4.43 | 0 |

in particular DR. At a high level, because of the conceptual symmetry in Figure 1 (see also (7)), we expect both DQ and DR to have advantages and disadvantages in different applications, and none should fully dominate the other. Nevertheless, we find many attractive features of DQ through the results in this paper, which suggest that DQ may be a better choice in many situations.

We summarize these features below. Some of these features are shared by DR, but many are not. (i) DQ defined on a class of MCP risk measures can be uniquely characterized by six intuitive axioms (Theorem 1). DQ defined on a class of coherent risk measures can be uniquely characterized by further adding the axiom of portfolio convexity (Theorem 2). These two results lay an axiomatic foundation for using DQ as a diversification index. (ii) DQ further satisfies many properties for common risk measures (Propositions 1-4). These properties are not shared by the corresponding DR. (iii) DQ is intuitive and interpretable with respect to dependence and common perceptions of diversification (Theorem 3). (iv) DQ can be applied to a wide range of risk measures, such as the regulatory risk measures VaR and ES, as well as expectiles. In cases of VaR and ES, DQ has simple formulas and convenient properties (Theorem 4 and Proposition 5). (v) Portfolio optimization of DQs based on VaR and ES can be computed very efficiently (Proposition 6). (vi) DQ can be easily applied to real data and it produces results that are consistent with our usual perception of diversification (Section 7).

Among the class of DQ, for most applications, we generally recommend the use of DQ based on ES for the following reasons: (a) it satisfies all seven axioms of intuitive appeal; (b) it has a simple optimization formula that is very convenient in portfolio optimization; (c) it is closely connected to financial regulation as ES is the standard risk measure of Basel IV; (d) it has a flexible parameter $\alpha$ that allows for reflecting the sensitivity to the tail risk of the decision maker; (e) it is conceptually easy to interpret as the (usually unique) level $\beta$ of the ES family such that $\mathrm{ES}_\beta(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathrm{ES}_\alpha(X_i)$.

We also mention a few interesting questions on DQ, which call for thorough future study. (i) DQ is defined through a class of risk measures. It would be interesting to formulate DQ using expected utility or behavioral decision models, to analyze the decision-theoretic implications of DQ. For instance, DQ based on entropic risk measures can be equivalently formulated using exponential utility functions. Alternatively, one may also build DQ directly from acceptability indices (see Remark 3). (ii) To compute DQ, one needs to invert the decreasing function $\beta \mapsto \rho_\beta(\sum_{i=1}^n X_i)$. In the case of VaR and ES, the formula for this inversion is simple (Theorem 4). For more complicated classes of risk measures, this computation may be complicated and requires detailed analysis. (iii) For general distributions and risk measures other than VaR and ES, finding analytical formulas or efficient algorithms for optimal diversification using either DQ or DR is a challenging task. (iv) Further analysis of DQ without scale-invariance, such as those built on star-shaped risk measures (Castagnoli et al. (2022)), may further generalize the domain of application.

**Acknowledgements.** We thank the Editor, an Associate Editor, and three anonymous referees. The authors also thank Fabio Bellini, Aleš Černý, An Chen, Jan Dhaene, Paul Embrechts, Paul Glasserman, Pablo Koch-Medina, Dimitrios Konstantinides, Henry Lam, Jean-Gabriel Lauzier, Fabio Macherroni, Massimo Marinacci, Cosimo Munari, Alexander Schied, Qihe Tang, Stan Uryasev, Peter Wakker, and Zuoquan Xu for helpful comments and discussions on an earlier version of the paper. Liyuan Lin is supported by the Hickman Scholarship from the Society of Actuaries. Ruodu Wang is supported by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2018-03823, RGPAS-2018-522590) and the Sun Life Research Fellowship.

# References


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

Basak, S. and Shapiro, A. (2001). Value-at-Risk based risk management: Optimal policies and asset prices. *Review of Financial Studies*, **14**(2), 371–405.

BCBS (2019). *Minimum Capital Requirements for Market Risk. February 2019.* Basel Committee on Banking Supervision. Basel: Bank for International Settlements, Document d457.

Bellini, F., Fadina, T., Wang, R. and Wei, Y. (2022). Parametric measures of variability induced by risk measures. *Insurance: Mathematics and Economics*, **106**, 270–284.

Bellini, F., Klar, B., Müller, A. and Rosazza Gianin, E. (2014). Generalized quantiles as risk measures.*Insurance: Mathematics and Economics*, **54**(1), 41–48.

Berger, L. and Eeckhoudt, L. (2021). Risk, ambiguity, and the value of diversification. *Management Science*, **67**(3), 1639–1647.

Bernard, C., Jiang, X. and Wang, R. (2014). Risk aggregation with dependence uncertainty. *Insurance: Mathematics and Economics*, **54**, 93–108.

Boyle, P., Garlappi, L., Uppal, R. and Wang, T. (2012). Keynes meets Markowitz: The trade-off between familiarity and diversification. *Management Science*, **58**(2), 253–272.

Bürgi, R., Dacorogna, M. M. and Iles, R. (2008). Risk aggregation, dependence structure and diversification benefit. In: *Stress Testing for Financial Institutions* (Eds. Rösch, D. and Scheule, H.), 265–306. Riskbooks, Incisive Media, London.

Castagnoli, E., Cattelan, G., Maccheroni, F., Tebaldi, C. and Wang, R. (2022). Star-shaped risk measures. *Operations Research*, **70**(5), 2637–2654.

Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M. and Montrucchio, L. (2011). Risk measures: Rationality and diversification. *Mathematical Finance*, **21**(4), 743–774.

Chambers, C. P. (2009). An axiomatization of quantiles on the domain of distribution functions. *Mathematical Finance*, **19**(2), 335–342.

Cherny, A. and Madan, D. (2009). New measures for performance evaluation. *The Review of Financial Studies*, **22**(7), 2571–2606.

Choi, N., Fedenia, M., Skiba, H. and Sokolyk, T. (2017). Portfolio concentration and performance of institutional investors worldwide. *Journal of Financial Economics*, **123**(1), 189–208.

Choueifaty, Y. and Coignard, Y. (2008). Toward maximum diversification. *The Journal of Portfolio Management*, **35**(1), 40–51.

Corbett, C. J. and Rajaram, K. (2006). A generalization of the inventory pooling effect to non-normal dependent demand. *Manufacturing and Service Operations Management,* **8**(4), 351–358.

D'Acunto, F., Prabhala, N. and Rossi, A. G. (2019). The promises and pitfalls of robo-advising. *The Review of Financial Studies*, **32**(5), 1983–2020.

Debreu, G. (1954). Representation of a preference ordering by a numerical function. *Decision Processes*, **3**, 159–165.

DeMiguel, V., Garlappi, L. and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The Review of Financial Studies*, **22**(5), 1915–1953.

Denis, D. J., Denis, D. K. and Yost, K. (2002). Global diversification, industrial diversification, and firm value. *The Journal of Finance*, **57**(5), 1951–1979.

Denneberg, D. (1990). Premium calculation: Why standard deviation should be replaced by absolute deviation. *ASTIN Bulletin*, **20**, 181–190.

Dhaene, J., Tsanakas, A., Valdez, E. A. and Vanduffel, S. (2012). Optimal capital allocation principles. *Journal of Risk and Insurance*, **79**(1), 1–28.

Drapeau, S. and Kupper, M. (2013). Risk preferences and their robust representation. *Mathematics of Operations Research*, **38**(1), 28–62.

EIOPA (2011). Equivalence assessment of the Swiss supervisory system in relation to articles 172, 227 and 260 of the Solvency II Directive, EIOPA-BoS-11-028.

Embrechts, P., Furrer, H. and Kaufmann, R. (2009). Different kinds of risk. In *Handbook of Financial Time Series* (Eds. Mikosch, T. et al.), 729–751. Springer, Berlin, Heidelberg.

Embrechts, P., Liu, H. and Wang, R. (2018). Quantile-based risk sharing. *Operations Research*, **66**(4), 936–949.

Embrechts, P., McNeil, A. and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In *Risk Management: Value at Risk and Beyond* (Ed. Dempster, M. A. H.), 176–223. Cambridge University Press.

Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R. and Beleraj, A. (2014). An academic response to Basel 3.5. *Risks*, **2**(1), 25-48.

Embrechts, P., Wang, B. and Wang, R. (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics*, **19**(4), 763–790.

Emmer, S., Kratz, M. and Tasche, D. (2015). What is the best risk measure in practice? A comparison of standard measures. *Journal of Risk*, **18**(2), 31–60.

Fama, E. F. and Miller, M. H. (1972). *The Theory of Finance.* Holt, Rinehart, and Winston, New York.

Föllmer, H. and Schied, A. (2016). *Stochastic Finance. An Introduction in Discrete Time.* Fourth Edition. Walter de Gruyter, Berlin.

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

Furman, E., Wang, R. and Zitikis, R. (2017). Gini-type measures of risk and variability: Gini shortfall, capital allocation and heavy-tailed risks. *Journal of Banking and Finance*, **83**, 70–84.

Gilboa, I. (2009). *Theory of Decision under Uncertainty.* Cambridge University Press, Cambridge.

Gilboa, I. Maccheroni, F., Marinacci, M. and Schmeidler, D. (2010). Objective and subjective rationality in a multiple prior model. *Econometrica*, **78**(2), 755–770.

Gilboa, I., Postlewaite, A. and Samuelson, L. (2019). What are axiomatizations good for? *Theory and Decision*, **86**(3–4), 339–359.

Green, R. C. and Hollifield, B. (1992). When will mean-variance efficient portfolios be well diversified? *The Journal of Finance*, **47**(5), 1785–1809.

Guan, Y., Jiao, Z. and Wang, R. (2022). A reverse Expected Shortfall optimization formula. *North American Actuarial Journal*. DOI: 10.1080/10920277.2023.2249524.

Han, X., Lin, L. and Wang, R. (2023). Diversification quotients based on VaR and ES. *Insurance: Mathematics and Economics*, **113**, 185–197.

Huber, P. J. and Ronchetti E. M. (2009). *Robust Statistics*. Second ed., Wiley Series in Probability and Statistics. Wiley, New Jersey. First ed.: Huber, P. (1980).

Ibragimov, R., Jaffee, D. and Walden, J. (2011). Diversification disasters. *Journal of Financial Economics*, **99**(2), 333–348.

Jakobsons, E., Han, X. and Wang, R. (2016). General convex order on risk aggregation. *Scandinavian Actuarial Journal*, **2016**(8), 713-740.

Jansen, D. W. and De Vries, C. G. (1991). On the frequency of large stock returns: Putting booms and busts into perspective. *The Review of Economics and Statistics*, **73**(1), 18–24.

Klibanoff, P., Marinacci, M. and Mukerji, S. (2005). A smooth model of decision making under ambiguity. *Econometrica*, **73**(6), 1849–1892.

Koumou, G. B. (2020). Diversification and portfolio theory: A review. *Financial Markets and Portfolio Management*, **34**(3), 267–312.

Koumou, G. B. and Dionne, G. (2022). Coherent diversification measures in portfolio theory: An axiomatic foundation. *Risks*, **10**(11), 205.

Kováčová, G., Rudloff, B. and Cialenco, I. (2020). Acceptability maximization. *arXiv*: 2012.11972.

Kusuoka, S. (2001). On law invariant coherent risk measures. *Advances in Mathematical Economics*, **3**, 83–95.

Levy, H. and Levy, M. (2004). Prospect theory and mean-variance analysis. *The Review of Financial Studies*, **17**(4), 1015–1041.

Levy, H. and Sarnat, M. (1970). International diversification of investment portfolios. *The American Economic Review*, **60**(4), 668–675.

Li, H. and Wang, R. (2022). PELVE: Probability equivalent level of VaR and ES. *Journal of Econometrics*, published online, doi.org/10.1016/j.jeconom.2021.12.012.

Liu, X., Küçükyavuz, S. and Luedtke, J. (2016). Decomposition algorithms for two-stage chance-constrained programs. *Mathematical Programming*, **157**(1), 219–243.

Luedtke, J. (2014). A branch-and-cut decomposition algorithm for solving chance-constrained mathematical programs with finite support. *Mathematical Programming*, **146**(1), 219–244.

Maccheroni, F., Marinacci, M., Rustichini, A. and Taboga, M. (2009). Portfolio selection with monotone mean-variance preferences. *Mathematical Finance*, **19**(3), 487–521.

Mafusalov A. and Uryasev S. (2018). Buffered probability of exceedance: Mathematical properties and optimization. *SIAM Journal on Optimization*, **28**(2), 1077–1103.

Mainik, G. and Rüschendorf, L (2010). On optimal portfolio diversification with respect to extreme risks. *Finance and Stochastics*, **14**(4), 593–623.

Mao, T. and Wang, R. (2020). Risk aversion in regulatory capital calculation. *SIAM Journal on Financial Mathematics*, **11**(1), 169–200.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, **7**(1), 77–91.

Mas-Colell, A., Whinston, M.D. and Green, J.R. (1995). *Microeconomic Theory.* Oxford University Press.

Matmoura, Y. and Penev, S. (2013). Multistage optimization of option portfolio using higher order coherent risk measures. *European Journal of Operational Research*, **227**(1), 190–198.

McNeil, A. J., Frey, R. and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools.* Revised Edition. Princeton, NJ: Princeton University Press.

Puccetti, G. and Wang R. (2015). Extremal dependence concepts. *Statistical Science*, **30**(4), 485–517.

Rockafellar, R. T. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, **26**(7), 1443–1471.

Rockafellar, R. T., Uryasev, S. and Zabarankin, M. (2006). Generalized deviations in risk analysis. *Finance and Stochastics*, **10**(1), 51–74.

Rothschild, M. and Stiglitz, J. E. (1970). Increasing risk: I. A definition. *Journal of Economic Theory*, **2**(3), 225–243.

Rüschendorf, L. (2013).*Mathematical Risk Analysis. Dependence, Risk Bounds, Optimal Allocations and Portfolios.* Springer, Heidelberg.

Samuelson, P. A. (1967). General proof that diversification pays. *Journal of Financial and Quantitative Analysis*, **2**(1), 1–13.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders.* Springer Series in Statistics.

Sharpe, (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, **19**(3), 425–442.

Shen, S., Smith, J. C. and Ahmed, S. (2010). Expectation and chance-constrained models and algorithms for insuring critical paths. *Management Science*, **56**(10), 1794–1814.

Tasche, D. (2007). Capital allocation to business units and sub-portfolios: The Euler principle. *arXiv*: 0708.2542.

Tu, J. and Zhou, G. (2011). Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics*, **99**(1), 204–215.

Van Nieuwerburgh, S. and Veldkamp, L. (2010). Information acquisition and under-diversification. *The Review of Economic Studies*, **77**(2), 779–805.

Wakker, P.P. (2010). *Prospect Theory: For Risk and Ambiguity.* Cambridge University Press, Cambridge.

Wang, B. and Wang, R. (2016). Joint mixability. *Mathematics of Operations Research*, **41**(3), 808–826.

Wang, R., Wei, Y. and Willmot, G. E. (2020). Characterization, robustness and aggregation of signed Choquet integrals. *Mathematics of Operations Research*, **45**(3), 993–1015.

Wang, R. and Wu, Q. (2020). Dependence and risk attitudes: An equivalence. *SSRN*: 3707709.

Wang, R. and Zitikis, R. (2021). An axiomatic foundation for the Expected Shortfall. *Management Science*, **67**(3), 1413–1429.

Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica*, **55**(1), 95–115.

# Technical appendices

## Outline of the appendices

We organize the technical appendices as follows. The proofs of the main results, Theorems 1–4, are presented in Appendix A. Additional results, discussions, and proofs of propositions are presented in Appendices B (for Section 2), C (for Section 3), D (for Section 4), E (for Section 5), and F (for Section 6). Finally, in Appendix G, we present other examples for the optimal portfolio problem that complement the empirical studies in Section 7.3.

# A Proofs of Theorems 1–4

*Proof of Theorem 1.* For $\mathbf{X} \in \mathcal{X}^n$ and a risk measure $\phi : \mathcal{X} \to \mathbb{R}$, denote by $S(\mathbf{X}) = \sum_{i=1}^n X_i$ and $\mathbf{X}_\phi = (X_1 - \phi(X_1), \dots, X_n - \phi(X_n))$.

We first verify the "if" statement. Using the definition of DQ and properties of MCP risk measures, it is straightforward to verify [+], [LI], [SI]. Below we check the other three axioms.

To show $[\mathrm{R}]_\phi$, for $\mathbf{X}, \mathbf{Y} \in \mathcal{X}^n$ such that $\mathbf{X} \overset{\mathrm{m}}{\simeq} \mathbf{Y}$ and $\sum_{i=1}^n X_i \leqslant \sum_{i=1}^n Y_i$, we have $\sum_{i=1}^n \rho_\alpha(X_i) = \sum_{i=1}^n \rho_\alpha(Y_i)$ and $\rho_\beta(\sum_{i=1}^n X_i) \leqslant \rho_\beta(\sum_{i=1}^n Y_i)$ for all $\beta \in I$. Hence,

$$\begin{aligned}\mathrm{DQ}_\alpha^\rho(\mathbf{X}) &= \frac{1}{\alpha} \inf \left\{ \beta \in I : \rho_\beta \left( \sum_{i=1}^n X_i \right) \leqslant \sum_{i=1}^n \rho_\alpha(X_i) \right\} \\ &\leqslant \frac{1}{\alpha} \inf \left\{ \beta \in I : \rho_\beta \left( \sum_{i=1}^n Y_i \right) \leqslant \sum_{i=1}^n \rho_\alpha(Y_i) \right\} = \mathrm{DQ}_\alpha^\rho(\mathbf{Y}).\end{aligned}$$

To show $[\mathrm{N}]_\phi$, it is straightforward that $\mathrm{DQ}_\alpha^\rho(\mathbf{0}) = 0$. Let $\mathbf{X} = (X, \dots, X)$ for any $X \in \mathcal{X}$. We have

$$\mathrm{DQ}_\alpha^\rho(\mathbf{X}) = \frac{1}{\alpha} \inf\{\beta \in I : \rho_\beta(nX) \leqslant n\rho_\alpha(X)\} \leqslant \frac{\alpha}{\alpha} = 1.$$

If $\mathbf{Y} \overset{\mathrm{m}}{\succ} (X, \dots, X)$ and $\sum_{i=1}^n Y_i \geqslant nX$, then $\sum_{i=1}^n \rho_\alpha(Y_i) < n\rho_\alpha(X) \leqslant \rho_\alpha(\sum_{i=1}^n Y_i)$. Hence, $\mathrm{DQ}_\alpha^\rho(\mathbf{Y}) \geqslant 1$.

To show $[\mathrm{C}]_\phi$, for $\mathbf{X} \in \mathcal{X}^n$, we have $a_{\mathbf{X}}^* = \inf\{\beta \in I : \rho_\beta(S(\mathbf{X}_{\rho_\alpha})) \leqslant 0\}$. If $a_{\mathbf{X}}^* = 0$, it is clear that $\mathrm{DQ}_\alpha^\rho(\mathbf{X}) = 0$ and $[\mathrm{C}]_\phi$ holds as $\mathrm{DQ}_\alpha^\rho(\mathbf{Y}^k) \geqslant 0$ for any $\mathbf{Y}^k \in \mathcal{X}^n$. Now, we assume $a_{\mathbf{X}}^* > 0$. For any $0 \leqslant \beta < a_{\mathbf{X}}^*$, we have $\rho_\beta(S(\mathbf{X}_{\rho_\alpha})) > 0$. Since $\mathbf{Y}^k \overset{\mathrm{m}}{\simeq} \mathbf{X}$ for each $k$ and $(S(\mathbf{X}) - S(\mathbf{Y}^k))_+ \xrightarrow{L^\infty} 0$ as $k \to \infty$, for any $\varepsilon > 0$, there exists $K$ such that $S(\mathbf{X}_{\rho_\alpha}) - S(\mathbf{Y}_{\rho_\alpha}^k) \leqslant \varepsilon$ for all $k \geqslant K$. For any $0 < \delta < a_{\mathbf{X}}^*$, let $0 < \varepsilon < \rho_{a_{\mathbf{X}}^* - \delta}(S(\mathbf{X}_{\rho_\alpha}))$. It is clear that $0 < \varepsilon < \rho_\beta(S(\mathbf{X}_{\rho_\alpha}))$ for all $0 < \beta < a_{\mathbf{X}}^* - \delta$. Hence, for all $0 < \beta < a_{\mathbf{X}}^* - \delta$, there exists $K$ such that $0 < \rho_\beta(S(\mathbf{X}_{\rho_\alpha})) - \varepsilon \leqslant \rho_\beta(S(\mathbf{Y}_{\rho_\alpha}^k))$ for all $k \geqslant K$, which implies $a_{\mathbf{Y}^k}^* \geqslant a_{\mathbf{X}}^* - \delta$. Therefore, $(\mathrm{DQ}_\alpha^\rho(\mathbf{X}) - \mathrm{DQ}_\alpha^\rho(\mathbf{Y}^k))_+ \to 0$.

Next, we show the "only if" statement. Assume that $D:\mathcal{X}^n\to\overline{\mathbb{R}}$ satisfies [+], [LI], [SI], $[\mathrm{R}]_\phi$, $[\mathrm{N}]_\phi$ and $[\mathrm{C}]_\phi$. Note that $\mathbf{X}_\phi \overset{\mathrm{m}}{\simeq} \mathbf{Y}_\phi$ for all $\mathbf{X},\mathbf{Y}\in\mathcal{X}^n$ since $\phi(X-\phi(X))=0$ for any $X\in\mathcal{X}$. Hence, by using $[\mathrm{R}]_\phi$, we know that $S(\mathbf{X}_\phi)=S(\mathbf{Y}_\phi)$ implies $D(\mathbf{X}_\phi)=D(\mathbf{Y}_\phi)$. Further, by [LI], we have $D(\mathbf{X})=D(\mathbf{Y})$ if $S(\mathbf{X}_\phi)=S(\mathbf{Y}_\phi)$. This means that $D(\mathbf{X})$ is determined by $S(\mathbf{X}_\phi)$. Define the mapping

$$R:\mathcal{X}\to[0,\infty],\ R(X)=\inf\{D(\mathbf{X}):X\leqslant S(\mathbf{X}_\phi),\ \mathbf{X}\in\mathcal{X}^n\},\tag{EC.1}$$

with the convention $\inf\varnothing=\infty$. Next, we will verify several properties of $R$.

(a) $R(S(\mathbf{X}_\phi))=D(\mathbf{X})$ for $\mathbf{X}\in\mathcal{X}^n$. The inequality $R(S(\mathbf{X}_\phi))\leqslant D(\mathbf{X})$ follows directly from (EC.1). To see the opposite direction of the inequality, suppose $R(S(\mathbf{X}_\phi))<D(\mathbf{X})$. By (EC.1), there exists $\mathbf{Y}\in\mathcal{X}^n$ such that $D(\mathbf{Y})<D(\mathbf{X})$ and $S(\mathbf{X}_\phi)\leqslant S(\mathbf{Y}_\phi)$. This contradicts $[\mathrm{R}]_\phi$ of $D$.

(b) $R(\lambda X)=R(X)$ for all $\lambda>0$ and $X\in\mathcal{X}$. This follows directly from (EC.1), [SI] of $D$ and positive homogeneity of $\phi$ which gives $(\lambda\mathbf{X})_\phi=\lambda\mathbf{X}_\phi$.

(c) $R(X)\leqslant R(Y)$ for all $X,Y\in\mathcal{X}$ with $X\leqslant Y$. This follows directly from (EC.1).

(d) $R(0)=0$. This follows directly from (EC.1) and $D(\mathbf{0})=0$ in $[\mathrm{N}]_\phi$.

(e) $\lim_{c\downarrow 0}R(S(\mathbf{X}_\phi)-c)=R(S(\mathbf{X}_\phi))$ for $\mathbf{X}\in\mathcal{X}^n$. Let $X=S(\mathbf{X}_\phi)$. By (c), we have $\lim_{c\downarrow 0}R(X-c)\leqslant R(X)$. Assume $\lim_{c\downarrow 0}R(X-c)<R(X)$; that is, there exists $\delta>0$ such that $R(X-c)<R(X)-\delta$ for all $c>0$. Let $c_k=1/k$ for $k\in\mathbb{N}$. By (EC.1), there exists a sequence $\{\mathbf{Y}^k\}_{k\in\mathbb{N}}$ such that $X-c_k\leqslant S(\mathbf{Y}^k_\phi)$ and $D(\mathbf{Y}^k_\phi)<D(\mathbf{X}_\phi)-\delta$. For $\{\mathbf{Y}^k_\phi\}_{k\in\mathbb{N}}$, we have $0\leqslant(S(\mathbf{X}_\phi)-S(\mathbf{Y}^k_\phi))_+\leqslant c_k$, which implies $(S(\mathbf{X}_\phi)-S(\mathbf{Y}^k_\phi))_+\xrightarrow{L^\infty}0$ as $k\to\infty$. By $[\mathrm{C}]_\phi$, we have $(D(\mathbf{X}_\phi)-D(\mathbf{Y}^k_\phi))_+\to 0$; that is, for any $\delta>0$, there exists $K\in\mathbb{N}$ such that $D(\mathbf{X}_\phi)-\delta\leqslant D(\mathbf{Y}^k_\phi)$ for all $k>K$, which is a contradiction. Therefore, we have $\lim_{c\downarrow 0}R(S(\mathbf{X}_\phi)-c)=R(S(\mathbf{X}_\phi))$.

Let $I=(0,\infty)$. For each $\beta\in(0,\infty)$, let $\mathcal{A}_\beta=\{X\in\mathcal{X}:R(X)\leqslant\beta\}$. Since $R$ is monotone, $\mathcal{A}_\beta$ is a decreasing set; i.e., $X\in\mathcal{A}_\beta$ implies $Y\in\mathcal{A}_\beta$ for all $Y\leqslant X$. Moreover, $\mathcal{A}_\beta$ is conic; i.e., $X\in\mathcal{A}_\beta$ implies $\lambda X\in\mathcal{A}_\beta$ for all $\lambda>0$. Moreover, we have $\mathcal{A}_\beta\subseteq\mathcal{A}_\gamma$ for $\beta\leqslant\gamma$, and $\mathcal{A}_\beta\neq\varnothing$ since $0\in\mathcal{A}_0$.

Let $\rho_\beta(X)=\inf\{m\in\mathbb{R}:X-m\in\mathcal{A}_\beta\}$ for $\beta\in I$. Since $\rho_\beta$ is defined via a conic acceptance set, $(\rho_\beta)_{\beta\in I}$ is a class of MCP risk measures; see Föllmer and Schied (2016). It is also clear that $\rho_\beta$ is decreasing in $\beta$. Note that $X\in\mathcal{A}_\beta$ implies $\rho_\beta(X)\leqslant 0$. Hence,

$$R(X)=\inf\{\beta\in I:R(X)\leqslant\beta\}=\inf\{\beta\in I:X\in\mathcal{A}_\beta\}\geqslant\inf\{\beta\in I:\rho_\beta(X)\leqslant 0\}.$$

For $X \in \{S(\mathbf{X}_\phi) : \mathbf{X} \in \mathcal{X}^n\}$, using (e), we have $R(X-m) \leqslant \beta$ for all $m > 0$ implies $R(X) \leqslant \beta$. Then we have

$$\inf\{\beta \in I : \rho_\beta(X) \leqslant 0\} = \inf\{\beta \in I : R(X-m) \leqslant \beta \text{ for all } m > 0\} \geqslant \inf\{\beta \in I : R(X) \leqslant \beta\}.$$

Therefore, $\inf\{\beta \in I : \rho_\beta(S(\mathbf{X}_\phi)) \leqslant 0\} = \inf\{\beta \in I : R(S(\mathbf{X}_\phi)) \leqslant \beta\} = R(S(\mathbf{X}_\phi))$ for all $\mathbf{X} \in \mathcal{X}^n$. Using (a), we get, for all $\mathbf{X} \in \mathcal{X}^n$,

$$D(\mathbf{X}) = R(S(\mathbf{X}_\phi)) = \inf\left\{\beta \in I : \rho_\beta\left(\sum_{i=1}^n X_i\right) \leqslant \sum_{i=1}^n \phi(X_i)\right\}.$$

Take a duplicate portfolio $\mathbf{X} = (X, \dots, X)$. Together with [PH] of $\rho_\beta$, $D(\mathbf{X}) \leqslant 1$ implies

$$D(X, \dots, X) = \inf\{\beta \in I : \rho_\beta(X) \leqslant \phi(X)\} \leqslant 1,$$

which is equivalent to $\rho_\beta(X) \leqslant \phi(X)$ for $\beta > 1$. For $m < \phi(X)$, take any $\mathbf{Y} = (Y_1, \dots, Y_n)$ satisfying $S(\mathbf{Y}_\phi) \geqslant X - m$; such $\mathbf{Y}$ may not exist. Let $\mathbf{Z} = (Y_1 - \phi(Y_1) + m/n, \dots, Y_n - \phi(Y_n) + m/n)$, yielding $\sum_{i=1}^n Z_i = \sum_{i=1}^n Y_i - \sum_{i=1}^n \phi(Y_i) + m \geqslant X$ and $(Z_1 \dots, Z_n) \overset{\mathrm{m}}{\succ} (X/n, \dots, X/n)$. Hence, $\mathbf{Z}$ is worse than duplicate. By [LI] and $[\mathrm{N}]_\phi$, we have $D(\mathbf{Y}) = D(\mathbf{Z}) \geqslant 1$. Since $D(\mathbf{Y}) \geqslant 1$ for all such $\mathbf{Y}$, by (EC.1), we have $R(X-m) \geqslant 1$. The above observation implies $\rho_{1-\varepsilon}(X) \geqslant \phi(X)$ for any $\varepsilon > 0$ since $\rho_{1-\varepsilon}(X) = \inf\{m \in \mathbb{R} : R(X-m) \leqslant 1-\varepsilon\}$. Therefore, we get $\rho_{1-\varepsilon} \geqslant \phi \geqslant \rho_{1+\varepsilon}$ for all $\varepsilon > 0$. Let $\tilde{\rho}_1 = \phi$, and $\tilde{\rho}_\beta = \rho_\beta$ for $\beta \neq 1$. The class $\tilde{\rho} = (\tilde{\rho}_\beta)_{\beta \in I}$ of MCP risk measures is decreasing in $\beta$ by the above argument. Moreover, for any $X \in \mathcal{X}$, since the two decreasing curves $\beta \mapsto \rho_\beta(X)$ and $\beta \mapsto \tilde{\rho}_\beta(X)$ differ at only one point, their (left) inverses coincide, and we have, for all $\mathbf{X} \in \mathcal{X}^n$,

$$\inf\left\{\beta \in I : \rho_\beta\left(\sum_{i=1}^n X_i\right) \leqslant \sum_{i=1}^n \phi(X_i)\right\} = \inf\left\{\beta \in I : \tilde{\rho}_\beta\left(\sum_{i=1}^n X_i\right) \leqslant \sum_{i=1}^n \tilde{\rho}_1(X_i)\right\},$$

which implies $D = \mathrm{DQ}_1^{\tilde{\rho}}$ on $\mathcal{X}^n$. A reparametrization via $\hat{\rho}_\beta = \tilde{\rho}_{\beta/\alpha}$ leads to $D = \mathrm{DQ}_\alpha^{\hat{\rho}}$ and $\hat{\rho}_\alpha = \phi$. □

The next two remarks are useful in the proof of Theorem 2.

*Remark* EC.1. In the proof of Theorem 1, the constructed class of risk measures $(\rho_\beta)_{\beta \in I}$ exhibits right continuity in $\beta$. This is established based on the condition $\bigcap_{\beta > \beta^*} \mathcal{A}_\beta = \mathcal{A}_{\beta^*}$.

*Remark* EC.2. For a non-linear coherent risk measure $\phi$, there exists $Y \in \mathcal{X}$ such that $\phi(Y) + \phi(-Y) > 0$. Suppose otherwise. Since $\phi$ is coherent risk measure, we have $\phi(Y) + \phi(-Y) \geqslant 0$, and this implies $\phi(Y) + \phi(-Y) = 0$ for all $Y \in \mathcal{X}$. We obtain $\phi(Y) \leqslant \phi(X+Y) + \phi(-X) = \phi(X+Y) - \phi(X)$ and $\phi(X+Y) \leqslant \phi(X) + \phi(Y)$ for any $X, Y \in \mathcal{X}$. This implies $\phi(X+Y) = \phi(X) + \phi(Y)$ for any $X, Y \in \mathcal{X}$, contradicting the non-linearity of $\phi$.

*Proof of Theorem 2.* For the"if" statement, since a coherent risk measure is also MCP, it follows that $\mathrm{DQ}^{\rho}_{\alpha}$ satisfies [+], [LI], [SI], $[\mathrm{R}]_{\phi}$ and $[\mathrm{N}]_{\phi}$, $[\mathrm{C}]_{\phi}$ by Theorem 1. Next, we show that $\mathrm{DQ}^{\rho}_{\alpha}$ satisfies [PC].

For any $\mathbf{X} \in \mathcal{X}^n$, let $r^{\mathbf{X}}_{\beta} : \Delta_n \to \mathbb{R}$ be given by

$$r^{\mathbf{X}}_{\beta}(\mathbf{w}) = \rho_{\beta}\left(\sum_{i=1}^{n} w_i X_i\right) - \sum_{i=1}^{n} \rho_{\alpha}\left(w_i X_i\right)$$

for $\beta \in I$. From [PH] of $\rho_{\alpha}$, we have $r^{\mathbf{X}}_{\beta}(\mathbf{w}) = \rho_{\beta}\left(\sum_{i=1}^{n} w_i X_i\right) - \mathbf{w}^{\top}\mathbf{x}^{\rho}_{\alpha}$. Convexity of $\rho_{\beta}$ implies convexity of $\mathbf{w} \mapsto r^{\mathbf{X}}_{\beta}(\mathbf{w})$. Hence, for the portfolio weight $\lambda\mathbf{w} + (1-\lambda)\mathbf{v} \in \Delta_n$, DQ based on $\rho$ at level $\alpha \in (0,1)$ is given by

$$\begin{aligned}
\mathrm{DQ}^{\rho}_{\alpha}((\lambda\mathbf{w} + (1-\lambda)\mathbf{v}) \odot \mathbf{X}) &= \frac{1}{\alpha}\inf\left\{\beta \in I : r^{\mathbf{X}}_{\beta}(\lambda\mathbf{w} + (1-\lambda)\mathbf{v}) \leqslant 0\right\} \\
&\leqslant \frac{1}{\alpha}\inf\left\{\beta \in I : \lambda r^{\mathbf{X}}_{\beta}(\mathbf{w}) + (1-\lambda) r^{\mathbf{X}}_{\beta}(\mathbf{v}) \leqslant 0\right\} \\
&\leqslant \frac{1}{\alpha}\max\left\{\inf\{\beta \in I : r^{\mathbf{X}}_{\beta}(\mathbf{w}) \leqslant 0\}, \inf\{\beta \in I : r^{\mathbf{X}}_{\beta}(\mathbf{v}) \leqslant 0\}\right\} \\
&= \max\left\{\mathrm{DQ}^{\rho}_{\alpha}(\mathbf{w} \odot \mathbf{X}), \mathrm{DQ}^{\rho}_{\alpha}(\mathbf{v} \odot \mathbf{X})\right\},
\end{aligned}$$

which gives us quasi-convexity of $\mathbf{w} \mapsto \mathrm{DQ}^{\rho}_{\alpha}(\mathbf{w} \odot \mathbf{X})$. By Remark 1, we have that $\mathrm{DQ}^{\rho}_{\alpha}$ satisfies [PC].

For the "only if" statement, we have constructed a class of MCP risk measures $\rho = (\rho_{\beta})_{\beta \in (0,\infty)}$ with $\rho_{\alpha} = \phi$ in the proof of Theorem 1. We will further show that $\rho$ is a class of convex risk measures using [PC].

For any $\lambda \in [0,1]$, $\mathbf{w}, \mathbf{v} \in \Delta_n$, $\mathbf{X} \in \mathcal{X}^n$ and $\beta \in (0,\infty)$, if $r^{\mathbf{X}}_{\beta}(\mathbf{w}) \leqslant 0$ and $r^{\mathbf{X}}_{\beta}(\mathbf{v}) \leqslant 0$, then we have $\mathrm{DQ}^{\rho}_{\alpha}(\mathbf{w} \odot \mathbf{X}) \leqslant \beta$ and $\mathrm{DQ}^{\rho}_{\alpha}(\mathbf{v} \odot \mathbf{X}) \leqslant \beta$. By [PC], we have $\mathrm{DQ}^{\rho}_{\alpha}((\lambda\mathbf{w} + (1-\lambda)\mathbf{v}) \odot \mathbf{X}) \leqslant \beta$. As discussed in Remark EC.1, $\rho_{\beta}$ is right-continuous for any $X \in \mathcal{X}$. Hence, we have $r^{\mathbf{X}}_{\beta}(\lambda\mathbf{w} + (1-\lambda)\mathbf{v}) \leqslant 0$. That is, the set $\{\mathbf{w} \in \Delta_n : r^{\mathbf{X}}_{\beta}(\mathbf{w}) \leqslant 0\}$ is convex for any $\mathbf{X} \in \mathcal{X}^n$ and $\beta \in I$.

Let $\mathrm{Conv}\{X - \rho_{\alpha}(X) : X \in \mathcal{X}\}$ be the convex hull of $\{X - \rho_{\alpha}(X) : X \in \mathcal{X}\}$. Next, we show that $\mathrm{Conv}\{X - \rho_{\alpha}(X) : X \in \mathcal{X}\} = \{X \in \mathcal{X} : \rho_{\alpha}(X) \leqslant 0\}$. For any $Z \in \mathrm{Conv}\{X - \rho_{\alpha}(X) : X \in \mathcal{X}\}$, there exist $(\lambda_1, \dots, \lambda_n) \in \Delta_n$ and $X_1, \dots, X_n \in \mathcal{X}$ such that $Z = \sum_{i=1}^{n} \lambda_i (X_i - \rho_{\alpha}(X_i))$. Since $\rho_{\alpha}$ is convex, we have

$$\rho_{\alpha}(Z) = \rho_{\alpha}\left(\sum_{i=1}^{n} \lambda_i (X_i - \rho_{\alpha}(X_i))\right) \leqslant \sum_{i=1}^{n} \lambda_i \rho_{\alpha}(X_i - \rho_{\alpha}(X_i)) = 0.$$

Hence, $\mathrm{Conv}\{X - \rho_{\alpha}(X) : X \in \mathcal{X}\} \subseteq \{X \in \mathcal{X} : \rho_{\alpha}(X) \leqslant 0\}$.

On the other hand, since $\rho_{\alpha}$ is a non-linear coherent risk measure, as noted in Remark EC.2, there exists $Y_0$ such that $\rho_{\alpha}(Y_0) + \rho_{\alpha}(-Y_0) > 0$. Let $Y_1 = 1 - Y_0$. For any $Z \in \mathcal{X}$ with $\rho_{\alpha}(Z) \leqslant 0$, we can find $\theta > 0$ such that

$$Z = (1 - 2\lambda)(X - \rho_{\alpha}(X)) + \lambda(\theta Y_0 - \rho_{\alpha}(\theta Y_0)) + \lambda(\theta Y_1 - \rho_{\alpha}(\theta Y_1))$$

with $\lambda = -\rho_\alpha(Z)/(\theta\rho_\alpha(Y_0) + \theta\rho_\alpha(-Y_0))$ and $X = 1/(1-2\lambda)Z$. It is clear that $\lambda \in [0, 1/2]$ holds if $\theta$ is sufficiently large. Hence, $Z \in \mathrm{Conv}\{X - \rho_\alpha(X) : X \in \mathcal{X}\}$. This implies $\mathrm{Conv}\{X - \rho_\alpha(X) : X \in \mathcal{X}\} \supseteq \{X \in \mathcal{X} : \rho_\alpha(X) \leqslant 0\}$.

Therefore, for any $X, Y \in \mathcal{X}$ with $\rho_\alpha(X) \leqslant 0$ and $\rho_\alpha(Y) \leqslant 0$, we can find $\mathbf{w}, \mathbf{v} \in \Delta^n$ and $\mathbf{X} \in \mathcal{X}^n$ with $n \geqslant 4$ such that $X = \mathbf{w}^\top(\mathbf{X} - \mathbf{x}^\rho_\alpha)$ and $Y = \mathbf{v}^\top(\mathbf{X} - \mathbf{x}^\rho_\alpha)$. Since $r_\beta(\mathbf{w}) = \rho_\beta\left(\mathbf{w}^\top(\mathbf{X} - \mathbf{x}^\rho_\alpha)\right)$, we have $\rho_\beta(X) = r^{\mathbf{X}}_\beta(\mathbf{w})$, $\rho_\beta(Y) = r^{\mathbf{X}}_\beta(\mathbf{v})$ and $\rho_\beta(\lambda X + (1-\lambda)Y) = r^{\mathbf{X}}_\beta(\lambda\mathbf{w} + (1-\lambda)\mathbf{v})$. If $\rho_\beta(X) \leqslant 0$ and $\rho_\beta(Y) \leqslant 0$, we have $r^{\mathbf{X}}_\beta(\lambda\mathbf{w} + (1-\lambda)\mathbf{v}) \leqslant 0$ for any $\lambda \in [0, 1]$; that is $\rho_\beta(\lambda X + (1-\lambda)Y) \leqslant 0$. Hence, $\rho_\beta$ is quasi-convex. Since $\rho_\beta$ is MCP, we further have $\rho_\beta$ is coherent. $\square$

*Proof of Theorem 3.* (i) As $\sum_{i=1}^n X_i \leqslant \sum_{i=i}^n \rho_\alpha(X_i)$ a.s. and $\rho_0 \leqslant$ ess-sup, it is clear that $\alpha^* = 0$ in (3), which implies $\mathrm{DQ}^\rho_\alpha(\mathbf{X}) = 0$. Conversely, if $\mathrm{DQ}^\rho_\alpha(\mathbf{X}) = 0$, then $\alpha^* = 0$. By definition of $\rho_0$ and $\mathrm{DQ}^\rho_\alpha$, this implies $\rho_0(\sum_{i=1}^n X_i) \leqslant \sum_{i=1}^n \rho_\alpha(X_i)$, and hence $\sum_{i=1}^n X_i \leqslant \sum_{i=1}^n \rho_\alpha(X_i)$ a.s.

(ii) We first show the "only if" statement. As $\rho$ is left continuous and non-flat from the left at $(\alpha, \sum_{i=1}^n X_i)$ and $\sum_{i=1}^n \rho_\alpha(X_i) - \rho_\alpha\left(\sum_{i=1}^n X_i\right) > 0$, there exists $\delta > 0$ such that

$$\rho_\beta\left(\sum_{i=1}^n X_i\right) - \rho_\alpha\left(\sum_{i=1}^n X_i\right) < \sum_{i=1}^n \rho_\alpha(X_i) - \rho_\alpha\left(\sum_{i=1}^n X_i\right)$$

for all $\beta \in (\alpha - \delta, \alpha)$. Hence, we have $\alpha^* \leqslant \alpha - \delta < \alpha$, which leads to $\mathrm{DQ}^\rho_\alpha(\mathbf{X}) < 1$.

Next, we show the "if" statement. As $\mathrm{DQ}^\rho_\alpha(\mathbf{X}) < 1$, we have $\alpha > \alpha^*$. By (3), there exists $\beta \in (\alpha^*, \alpha)$ such that

$$\sum_{i=1}^n \rho_\alpha(X_i) \geqslant \rho_\beta\left(\sum_{i=1}^n X_i\right).$$

Because $\rho$ is non-flat from the left at $(\alpha, \sum_{i=1}^n X_i)$, we have

$$\sum_{i=1}^n \rho_\alpha(X_i) \geqslant \rho_\beta\left(\sum_{i=1}^n X_i\right) > \rho_\alpha\left(\sum_{i=1}^n X_i\right).$$

(iii) If $\rho_\alpha$ satisfies [PH], for $\mathbf{X} = (\lambda_1 X, \dots, \lambda_n X)$ where $\lambda_1, \dots, \lambda_n \geqslant 0$, we have

$$\alpha^* = \inf\left\{\beta \in I : \rho_\beta\left(\sum_{i=1}^n \lambda_i X\right) \leqslant \sum_{i=1}^n \lambda_i \rho_\alpha(X)\right\}.$$

It is clear that $\rho_\alpha\left(\sum_{i=1}^n \lambda_i X\right) = (\sum_{i=1}^n \lambda_i)\rho_\alpha(X)$. Together with the non-flat condition and $\rho_\beta\left(\sum_{i=1}^n \lambda_i X\right) > \sum_{i=1}^n \lambda_i \rho_\alpha(X)$ for all $\beta < \alpha$, we have $\alpha^* = \alpha$, and thus $\mathrm{DQ}^\rho_\alpha(\mathbf{X}) = 1$.

(iv) If $\rho_\alpha$ is comonotonic-additive and $\mathbf{X}$ is comonotonic, then

$$\alpha^* = \inf\left\{\beta \in I : \rho_\beta\left(\sum_{i=1}^n X_i\right) \leqslant \sum_{i=1}^n \rho_\alpha(X_i) = \rho_\alpha\left(\sum_{i=1}^n X_i\right)\right\},$$

which, together with the non-flat condition, implies that $\alpha^* = \alpha$, and thus $\mathrm{DQ}^\rho_\alpha(\mathbf{X}) = 1$. $\square$

*Proof of Theorem 4.* We first show (4). For any $X \in \mathcal{X}$, $t \in \mathbb{R}$ and $\alpha \in (0,1)$, by Lemma 1 of Guan et al. (2022), $\mathbb{P}(X > t) \leqslant \alpha$ if and only if $\mathrm{VaR}_\alpha(X) \leqslant t$. Hence,

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^n X_i > \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i)\right) &= \inf\left\{\beta \in (0,1) : \mathbb{P}\left(\sum_{i=1}^n X_i > \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i)\right) \leqslant \beta\right\}\\ &= \inf\left\{\beta \in (0,1) : \mathrm{VaR}_\beta\left(\sum_{i=1}^n X_i\right) \leqslant \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i)\right\},\end{aligned}$$

and (4) follows. The formula (5) for $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ follows from a similar argument to (4) by noting that $Y$ is a random variable with $\mathrm{VaR}_\alpha(Y) = \mathrm{ES}_\alpha(\sum_{i=1}^n X_i)$.

Next, we show the last statement of the theorem. If $\mathbb{P}(\sum_{i=1}^n X_i > \sum_{i=1}^n \mathrm{ES}_\alpha(X_i)) = 0$, then $\mathrm{DQ}_\alpha^{\mathrm{ES}}(X) = 0$ by Theorem 3 (i).

Below, we assume $\mathbb{P}(\sum_{i=1}^n X_i > \sum_{i=1}^n \mathrm{ES}_\alpha(X_i)) > 0$. The formula (6) is very similar to Proposition 2.2 of Mafusalov and Uryasev (2018), where we additionally show that the minimizer to (6) is not 0. Here we present a self-contained proof based on the well-known formula of ES (Rockafellar and Uryasev (2002)),

$$\mathrm{ES}_\beta(X) = \min_{t\in\mathbb{R}}\left\{t + \frac{1}{\beta}\mathbb{E}\left[(X-t)_+\right]\right\}, \quad \text{for } X \in \mathcal{X} \text{ and } \beta \in (0,1).$$

Using this formula, we obtain, by writing $X_i' = X_i - \mathrm{ES}_\alpha(X_i)$ for $i \in [n]$,

$$\begin{aligned}\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X}) &= \frac{1}{\alpha}\inf\left\{\beta \in (0,1) : \mathrm{ES}_\beta\left(\sum_{i=1}^n X_i\right) - \sum_{i=1}^n \mathrm{ES}_\alpha(X_i) \leqslant 0\right\}\\ &= \frac{1}{\alpha}\inf\left\{\beta \in (0,1) : \min_{t\in\mathbb{R}}\left\{t + \frac{1}{\beta}\mathbb{E}\left[\left(\sum_{i=1}^n X_i' - t\right)_+\right]\right\} \leqslant 0\right\}\\ &= \frac{1}{\alpha}\inf\left\{\beta \in (0,1) : \frac{1}{\beta}\mathbb{E}\left[\left(\sum_{i=1}^n X_i' - t\right)_+\right] \leqslant -t \text{ for some } t \in \mathbb{R}\right\}\\ &= \frac{1}{\alpha}\inf\left\{\beta \in (0,1) : \mathbb{E}\left[\left(r\sum_{i=1}^n X_i' + 1\right)_+\right] \leqslant \beta \text{ for some } r \in (0,\infty)\right\}\\ &= \frac{1}{\alpha}\inf_{r\in(0,\infty)}\mathbb{E}\left[\left(r\sum_{i=1}^n X_i' + 1\right)_+\right].\end{aligned}$$

Let $f : [0,\infty) \to [0,\infty)$, $r \mapsto \mathbb{E}[(r\sum_{i=1}^n X_i' + 1)_+]$. It is clear that $f(0) = 1$. Moreover,

$$f(r) \geqslant r\mathbb{E}\left[(X_i')_+\right] \to \infty \quad \text{as } r \to \infty.$$

By Theorem 1 (iii), we have $\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X}) \leqslant 1$, and hence $\inf_{r\in(0,\infty)} f(r) \leqslant \alpha < 1$. The continuity of $f$ yields $\inf_{r\in(0,\infty)} f(r) = \min_{r\in(0,\infty)} f(r)$, and thus (6) holds. □

# B Additional results for Section 2

In this appendix, we present an impossibility result showing a conflicting nature of the three natural properties [+], [LI] and [SI] for some diversification indices defined via risk measures. As mentioned in Section 2, the most commonly used diversification indices depend on $\mathbf{X}$ through its values assessed by some risk measure $\phi$. That is, given a risk measures $\phi$ and a portfolio $\mathbf{X}$, the diversification index can be written as

$$D(\mathbf{X}) = R\left(\phi\left(\sum_{i=1}^{n} X_i\right), \phi(X_i), \dots, \phi(X_n)\right) \text{ for some function } R : \mathbb{R}^{n+1} \to \overline{\mathbb{R}}. \tag{EC.2}$$

We will say that $D$ is *$\phi$-determined* if (EC.2) holds. Often, one may further choose $R$ so that $D(\mathbf{X})$ decreases in $\phi(\sum_{i=1}^{n} X_i)$ and increases in $\phi(X_i)$ for each $i \in [n]$, for a proper interpretation of measuring diversification.

We show that a diversification index based on an MCP risk measure, such as VaR or ES satisfying all three properties [+], [LI] and [SI] can take at most 3 different values. In this case, we will say that the diversification index $D$ is *degenerate*. In fact, this result can be extended to more general properties $[\mathrm{PH}]_\gamma$ and $[\mathrm{CA}]_m$ with $\gamma \in \mathbb{R}$ and $m \in \mathbb{R}$ of the risk measure $\phi$, with definitions given at the beginning of Section 4.

**Proposition EC.1.** *Fix $n \geqslant 1$. Suppose that a risk measure $\phi$ satisfies $[\mathrm{PH}]_\gamma$ and $[\mathrm{CA}]_\mathrm{m}$ with $\gamma \in \mathbb{R}$ and $m \neq 0$. A diversification index $D$ is $\phi$-determined and satisfies [+], [LI] and [SI] if and only if for all $\mathbf{X} \in \mathcal{X}^n$,*

$$D(\mathbf{X}) = C_1 \mathbb{1}_{\{d<0\}} + C_2 \mathbb{1}_{\{d=0\}} + C_3 \mathbb{1}_{\{d>0\}}, \tag{EC.3}$$

*where $d = \mathrm{DB}^\phi(\mathbf{X}) = \sum_{i=1}^{n} \phi(X_i) - \phi\left(\sum_{i=1}^{n} X_i\right)$ for some $C_1, C_2, C_3 \in \mathbb{R}_+ \cup \{\infty\}$.*

We first present a lemma to prepare for the proof of Proposition EC.1.

**Lemma EC.1.** *A function $R : \mathbb{R}^{n+1} \to \overline{\mathbb{R}}$ satisfies, for all $x_0 \in \mathbb{R}$, $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$, $\mathbf{c} = (c_1, \dots, c_n) \in \mathbb{R}^n$ and $\lambda > 0$, (i) $R\left(x_0 + \sum_{i=1}^{n} c_i, \mathbf{x} + \mathbf{c}\right) = R(x_0, \mathbf{x})$ and (ii) $R(\lambda x_0, \lambda \mathbf{x}) = R(x_0, \mathbf{x})$, if and only if there exist $C_1, C_2, C_3 \in \overline{\mathbb{R}}$ such that*

$$R(x_0, \mathbf{x}) = C_1 \mathbb{1}_{\{r<0\}} + C_2 \mathbb{1}_{\{r=0\}} + C_3 \mathbb{1}_{\{r>0\}}, \tag{EC.4}$$

*where $r = \sum_{i=1}^{n} x_i - x_0$, for all $x_0 \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$.*

*Proof.* First, we show that $R$ in (EC.4) satisfies (i) and (ii). Assume $r < 0$. For any $\mathbf{c} \in \mathbb{R}^n$ and $\lambda > 0$, it is clear that $x_0 + \sum_{i=1}^{n} c_i < \sum_{i=1}^{n}(x_i + c_i)$ and $\lambda x_0 < \sum_{i=1}^{n} \lambda x_i$. Therefore, (i) and (ii) are satisfied. The cases of $r = 0$ and $r > 0$ follow by the same argument.

Next, we verify the "only if" part. Given $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$ satisfying $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$, let $\mathbf{c} = \mathbf{y} - \mathbf{x}$. For any $x_0 \in \mathbb{R}$, we have $\sum_{i=1}^{n} c_i = \sum_{i=1}^{n}(y_i - x_i) = 0$. Therefore,

$$R(x_0, \mathbf{x}) = R\left(x_0 + \sum_{i=1}^{n} c_i, \mathbf{x} + \mathbf{c}\right) = R(x_0, \mathbf{y}).$$

Thus, the value of $R(x_0, \mathbf{x})$ only depends on $x_0$ and $\sum_{i=1}^n x_i$. Let $\tilde{R}: \mathbb{R}^2 \to \overline{\mathbb{R}}$ be a function such that $\tilde{R}\left(x_0, \sum_{i=1}^n x_i\right) = R(x_0, \mathbf{x})$. From the properties of $R$, $\tilde{R}$ satisfies $\tilde{R}(a+c, b+c) = \tilde{R}(a,b)$ for any $c \in \mathbb{R}$, and $\tilde{R}(\lambda a, \lambda b) = R(a,b)$ for any $\lambda > 0$. Hence, we have

$$\tilde{R}(a,b) = \tilde{R}(a-b,0) = \tilde{R}(1,0) \quad \text{for} \quad a > b,$$

$$\tilde{R}(a,b) = \tilde{R}(0,b-a) = \tilde{R}(0,1) \quad \text{for} \quad a < b,$$

and

$$\tilde{R}(a,b) = \tilde{R}(a-a,b-a) = \tilde{R}(0,0) \quad \text{for} \quad a = b.$$

Let $C_1 = \tilde{R}(1,0)$, $C_2 = \tilde{R}(0,0)$ and $C_3 = \tilde{R}(0,1)$. We have $R(x_0, \mathbf{x}) = \tilde{R}(x_0, \sum_{i=1}^n x_i)$, which has the form in (EC.4). $\square$

*Proof of Proposition EC.1.* Let us first prove sufficiency. By definition, $D$ satisfies [+] and $D$ is $\phi$-determined. Next, we prove $D$ satisfies [LI] and [SI]. Similarly to Lemma EC.1, we only prove the case $d < 0$. It is straightforward that

$$\phi\left(\sum_{i=1}^n \lambda X_i\right) = \lambda^\gamma \phi\left(\sum_{i=1}^n X_i\right) < \lambda^\gamma \sum_{i=1}^n \phi(X_i) = \sum_{i=1}^n \phi(\lambda X_i),$$

and

$$\phi\left(\sum_{i=1}^n (X_i + c_i)\right) = \phi\left(\sum_{i=1}^n X_i\right) + m\sum_{i=1}^n c_i < \sum_{i=1}^n \left(\phi(X_i) + mc_i\right) = \sum_{i=1}^n \phi(X_i + c_i).$$

Thus, we have $D(\lambda \mathbf{X}) = C_1$ and $D(\mathbf{X} + \mathbf{c}) = C_1$, which completes the proof of sufficiency.

Next, we show the necessity. Define the set

$$\mathcal{A} = \left\{ \left( \phi\left(\sum_{i=1}^n X_i\right), \phi(X_1), \dots, \phi(X_n) \right) : (X_1, \dots, X_n) \in \mathcal{X}^n \right\}.$$

Note that $\phi$ satisfies $[\mathrm{PH}]_\gamma$ with $\gamma \neq 0$ since $[\mathrm{CA}]_m$ for $m \neq 0$ implies $\rho(2) \neq \rho(1)$, which in turn implies $\gamma \neq 0$. We always write $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{c} = (c_1, \dots, c_n)$. Consider the two operations $(x_0, \mathbf{x}) \mapsto (x_0 + \sum_{i=1}^n c_i, \mathbf{x} + \mathbf{c})$ for some $\mathbf{c} \in \mathbb{R}^n$ and $(x_0, \mathbf{x}) \mapsto (\lambda x_0, \lambda \mathbf{x})$ for some $\lambda > 0$. Let $r(x_0, \mathbf{x}) = \sum_{i=1}^n x_i - x_0$. By using $[\mathrm{CA}]_m$ and $[\mathrm{PH}]_\gamma$ of $\phi$, we have that (see also the proof of Lemma EC.1) the regions $\mathcal{A}_+ := \{(x_0, \mathbf{x}) : r(x_0, \mathbf{x}) > 0\}$, $\mathcal{A}_0 := \{(x_0, \mathbf{x}) : r(x_0, \mathbf{x}) = 0\}$ and $\mathcal{A}_- := \{(x_0, \mathbf{x}) : r(x_0, \mathbf{x}) < 0\}$ are closed under the above two operations, and each of them is connected via the above two operations. Therefore, $\mathcal{A}$ is the union of some of $\mathcal{A}_+$, $\mathcal{A}_0$ and $\mathcal{A}_-$.

We define a function $R: \mathbb{R}^{n+1} \to \overline{\mathbb{R}}$. For $(x_0, \mathbf{x}) \in \mathcal{A}$, let $R(x_0, \mathbf{x}) = D(X_1, \dots, X_n)$, where $(X_1, \dots, X_n)$ is any random vector such that $x_0 = \phi(\sum_{i=1}^n X_i)$ and $\mathbf{x} = (\phi(X_1), \dots, \phi(X_n))$. The choice of $(X_1, \dots, X_n)$ is irrelevant since $D$ is $\phi$-determined. For $(x_0, \mathbf{x}) \in \mathbb{R}^{n+1} \setminus \mathcal{A}$, let $R(x_0, \mathbf{x}) = 0$. We will verify that $R$ satisfies conditions (i) and (ii) in Lemma EC.1.

For $(x_0, \mathbf{x}) \in \mathcal{A}$, there exists $\mathbf{X} = (X_1, \dots, X_n) \in \mathcal{X}^n$ such that $x_0 = \phi(\sum_{i=1}^n X_i)$ and $\mathbf{x} = (\phi(X_1), \dots, \phi(X_n))$. For any $\mathbf{c} \in \mathbb{R}^n$, using $[\mathrm{CA}]_m$ with $m \neq 0$ of $\phi$ and [LI] of $D$, we obtain

$$\begin{aligned} R\left(x_0 + \sum_{i=1}^n c_i, \mathbf{x} + \mathbf{c}\right) &= R\left(\phi\left(\sum_{i=1}^n X_i\right) + \sum_{i=1}^n c_i, \phi(X_1) + c_1, \dots, \phi(X_n) + c_n\right) \\ &= R\left(\phi\left(\sum_{i=1}^n \left(X_i + \frac{c_i}{m}\right)\right), \phi\left(X_1 + \frac{c_1}{m}\right), \dots, \phi\left(X_n + \frac{c_n}{m}\right)\right) \\ &= D\left(\mathbf{X} + \mathbf{c}/m\right) = D(\mathbf{X}) = R\left(x_0, \mathbf{x}\right). \end{aligned}$$

Using $[\mathrm{PH}]_\gamma$ with $\gamma \neq 0$ of $\phi$ and [SI] of $D$, for any $\lambda > 0$, we obtain

$$\begin{aligned} R(\lambda x_0, \lambda \mathbf{x}) &= R\left(\lambda\phi\left(\sum_{i=1}^n X_i\right), \lambda\phi(X_1), \dots, \lambda\phi(X_n)\right) \\ &= R\left(\phi\left(\sum_{i=1}^n \lambda^{1/\gamma} X_i\right), \phi(\lambda^{1/\gamma} X_1), \dots, \phi(\lambda^{1/\gamma} X_n)\right) \\ &= D(\lambda^{1/\gamma}\mathbf{X}) = D(\mathbf{X}) = R\left(x_0, \mathbf{x}\right). \end{aligned}$$

Hence, $R$ satisfies (i) and (ii) in Lemma EC.1 on $\mathcal{A}$. By definition, $R$ satisfies (i) and (ii) also on $\mathbb{R}^{n+1} \setminus \mathcal{A}$. Since $\mathcal{A}$ and $\mathbb{R}^{n+1} \setminus \mathcal{A}$ are both closed under the two operations, we know that $R$ satisfies (i) and (ii) on $\mathbb{R}^{n+1}$.

Using Lemma EC.1, we have $R$ has the representation (EC.4), which gives

$$D(\mathbf{X}) = C_1 \mathbb{1}_{\{d<0\}} + C_2 \mathbb{1}_{\{d=0\}} + C_3 \mathbb{1}_{\{d>0\}}$$

with $d = \sum_{i=1}^n \phi(X_i) - \phi\left(\sum_{i=1}^n X_i\right)$ and $C_1, C_2, C_3 \in \overline{\mathbb{R}}$ for all $\mathbf{X} \in X^n$. As $D$ satisfying [+], we have $C_1, C_2, C_3 \in \mathbb{R}_+ \cup \{\infty\}$. □

# C Additional results and proofs for Section 3

## C.1 Existence of worse-than-duplicate portfolios

We discuss the existence of worse-than-duplicate portfolios. First, note that if a vector $\mathbf{X}^{\mathrm{wd}} = (X_1^{\mathrm{wd}}, \dots, X_n^{\mathrm{wd}})$ is worse than a duplicate portfolio $\mathbf{X}^{\mathrm{du}} = (X, \dots, X)$ under a given MCP risk measure $\phi$, then we have

$$\phi\left(\sum_{i=1}^n X_i^{\mathrm{wd}}\right) \geqslant \phi(nX) = n\phi(X) > \sum_{i=1}^n \phi(X_i^{\mathrm{wd}})$$

and thus $\phi$ violates subadditivity with $\mathbf{X}^{\mathrm{wd}}$. Therefore, a necessary condition for the existence of a vector that is worse than a duplicate under a MCP risk measure $\phi$ is that $\phi$ violates subadditivity.

We further provide a necessary and sufficient condition for the existence or non-existence of duplicate portfolios.

**Lemma EC.2.** *For a monotone risk measure $\phi$, there exists a worse-than-duplicate portfolio if and only if there exist $X, X_1, \dots, X_n \in \mathcal{X}$ with $X_1 + \cdots + X_n = nX$ such that $\phi(X) > \phi(X_i)$ for $i \in [n]$.*

*Proof.* This follows directly by monotonicity. ☐

A risk measure $\phi : \mathcal{X} \to \mathbb{R}$ is *scale-continuous* if the mapping $\lambda \mapsto \phi(\lambda X)$ on $(0,1)$ is continuous for every $X$. This condition is very weak; for instance it is weaker than continuity on any $L^p$-space $\mathcal{X}$.

**Proposition EC.2.** *For a monotone risk measure $\phi$ scale-continuous on $\mathcal{X} = L^\infty$, there exists no worse-than-duplicate portfolio if and only if $\phi$ is quasi-convex.*

*Proof.* If $\phi$ is quasi-convex, then for any $X_1, \dots, X_n$,

$$\phi\left(\frac{X_1}{n} + \cdots + \frac{X_n}{n}\right) \leqslant \max\{\phi(X_1), \dots, \phi(X_n)\}. \tag{EC.5}$$

By Lemma EC.2, there exists no worse-than-duplicate portfolio. Conversely, if exists no worse-than-duplicate portfolio, then (EC.5) holds for all $X_1, \dots, X_n$. It suffices to verify that this implies quasi-convexity of $\phi$. That is, we need to show that for $\lambda \in (0,1)$ and $X_1, X_2$,

$$\phi(\lambda X_1 + (1-\lambda)X_2) \leqslant \max\{\phi(X_1), \phi(X_2)\}. \tag{EC.6}$$

First, suppose that $\lambda = p/n^q$ where $p, q \in \mathbb{N}$. Repeatedly applying (EC.5) $q$ times, we get, for all $Y_1, \dots, Y_m$ where $m = n^q$,

$$\phi\left(\sum_{i=1}^{m} \frac{Y_i}{m}\right) \leqslant \max\{\phi(Y_1), \dots, \phi(Y_m)\}. \tag{EC.7}$$

Letting $Y_i = X_1$ for $i \leqslant p$ and $Y_i = X_2$ for $i > p$ in (EC.7), we get (EC.6). Next, consider a general $\lambda \in (0,1)$. Let $X_1' = \lambda X_1/t$ and $X_2' = (1-\lambda)X_2/(1-t)$, where $t = p/n^q \in (0,1)$ for some $p, q \in \mathbb{N}$. Using (EC.7), we get

$$\phi(\lambda X_1 + (1-\lambda)X_2) = \phi(tX_1' + (1-t)X_2') \leqslant \max\{\phi(X_1'), \phi(X_2')\}. \tag{EC.8}$$

Sending $t \to \lambda$ and using continuity we obtain the desired result. ☐

## C.2 Examples and proofs related to portfolio convexity

In the first example, we show that convexity or quasi-convexity of $\mathbf{X} \mapsto D(\mathbf{X})$ should not hold for a diversification index $D$.

**Example 3** (Quasi-convexity on $\mathcal{X}^n$ is not desirable)**.** Let $(X, Y) \in \mathcal{X}^2$ represent any diversified portfolio (e.g., with iid normal components), and assume that $Z := (X+Y)/2$ is not a constant.

Since the portfolio $(Z, Z)$ relies only on one asset and has no diversification benefit, for a good diversification index $D$ we naturally want $D(Z, Z)$ to be larger than both $D(X, Y)$ and $D(Y, X)$; recall that $D(Z, Z) = 1$ in the setting of Theorem 3 (iii). This argument shows that it is unnatural to require $D$ to be convex or quasi-convex on $\mathcal{X}^2$; the case of $\mathcal{X}^n$ is similar. Indeed, if a real-valued $D$ satisfies [SI] and convexity on $\mathcal{X}^n$, then it is a constant; this is shown in the proposition below.

**Proposition EC.3.** *A mapping $D : \mathcal{X}^n \to \mathbb{R}$ satisfies* [SI] *and convexity if and only if $D(\mathbf{X}) = c$ for all $\mathbf{X} \in \mathcal{X}$ and some constant $c \in \mathbb{R}$.*

*Proof.* If $D$ is a constant for all $\mathbf{X} \in \mathcal{X}^n$, it is clear that $D$ satisfies [SI] and convexity. Next we will show the "only if" part. Let $d_0 = D(\mathbf{0}) \in \mathbb{R}$.

(i) If $d_0 \geqslant D(\mathbf{X})$ for all $\mathbf{X} \in \mathcal{X}^n$ and there exists $\mathbf{X}_0$ such that $D(\mathbf{X}_0) < d_0$, then

$$D\left(\frac{1}{2}\mathbf{X}_0 + \frac{1}{2}(-\mathbf{X}_0)\right) = D(\mathbf{0}) > \frac{1}{2}D(\mathbf{X}_0) + \frac{1}{2}D(-\mathbf{X}_0),$$

which contradicts the convexity of $D$.

(ii) If there exists $\mathbf{X}_0$ such that $d_0 < D(\mathbf{X}_0)$, then, by [SI] of $D$,

$$D\left(\frac{1}{2}\mathbf{0} + \frac{1}{2}\mathbf{X}_0\right) = D(\mathbf{X}_0) > \frac{1}{2}D(\mathbf{0}) + \frac{1}{2}D(\mathbf{X}_0),$$

which contradicts the convexity of $D$.

By (i) and (ii), we can conclude that $D$ only takes the value $d_0$. □

From the proof of Proposition EC.3, we see that the conflict between convexity and [SI] holds for real-valued mappings on any closed convex cone, not necessarily on $\mathcal{X}^n$.

In the second example, we see that convexity of $\mathbf{w} \mapsto D(\mathbf{w} \odot \mathbf{X})$ is not desirable either for a good diversification index.

**Example 4** (Convexity in $\mathbf{w}$ is not desirable)**.** Let $Z$ be standard normal and $\varepsilon > 0$ be a small constant. Consider a portfolio vector $\mathbf{X} = ((1-\varepsilon)Z, -\varepsilon Z)$. Let $\mathbf{w} = (1, 0)$ and $\mathbf{v} = (\varepsilon, 1-\varepsilon)$. Note that $\mathbf{w} \odot \mathbf{X} = (1-\varepsilon)(Z, 0)$ and $\mathbf{v} \odot \mathbf{X} = (\varepsilon - \varepsilon^2)(Z, -Z)$. The portfolio $\mathbf{w} \odot \mathbf{X}$ is not diversified since it has only one non-zero component, and the portfolio $\mathbf{v} \odot \mathbf{X}$ is perfectly hedged since the sum of its components is 0. Hence, for a good diversification index $D$, it should hold that $D(\mathbf{w} \odot \mathbf{X}) = 1$ and $D(\mathbf{v} \odot \mathbf{X}) = 0$; Theorem 3 confirms this. On the other hand, the portfolio

$$\left(\frac{1}{2}\mathbf{w} + \frac{1}{2}\mathbf{v}\right) \odot \mathbf{X} = \frac{1}{2}\left((1-\varepsilon^2)Z, -(\varepsilon - \varepsilon^2)Z\right)$$

is not well diversified since its second component is very small compared to its first component. Intuitively, for $\varepsilon \approx 0$, we expect $D((\mathbf{w}/2 + \mathbf{v}/2) \odot \mathbf{X}) \approx 1 > D(\mathbf{w} \odot \mathbf{X})/2 + D(\mathbf{v} \odot \mathbf{X})/2$. This shows that $\mathbf{w} \mapsto D(\mathbf{w} \odot \mathbf{X})$ is not convex. One can verify that this is indeed true if $D$ is DQ or DR based on commonly used risk measures such as SD, VaR ($\alpha < 1/2$) and ES.

*Proof of Proposition 1.* Since the proof of Theorem 2 solely relies on convexity and positive homogeneity of $\rho_\alpha$ to show [PC], it is clear that $\mathrm{DQ}_\alpha^\rho$ satisfies [PC].

The subadditivity of $\rho_\alpha$ implies that $\mathrm{DQ}_\alpha^\rho$ takes value in $[0,1]$. Consequently, $\{\mathrm{DQ}_\alpha^\rho(\mathbf{X}) : \mathbf{X} \in \mathcal{X}^n\} \subseteq [0,1]$. We only need to show $[0,1] \subseteq \{\mathrm{DQ}_\alpha^\rho(\mathbf{X}) : \mathbf{X} \in \mathcal{X}^n\}$.

Since $\rho_\alpha$ is non-linear and sub-linear, there exists $Y$ such that $\rho_\alpha(Y)+\rho_\alpha(-Y) > 0$ following the argument of Remark EC.2. Consider $\mathbf{X}^\theta = (X, \theta Y, -\theta Y, 0, \dots, 0)$ with $\theta \geqslant 0$. We have

$$\mathrm{DQ}_\alpha^\rho(\mathbf{X}^\theta) = \frac{1}{\alpha} \inf\{\beta \in I : \rho_\beta(X) \leqslant \rho_\alpha(X) + \theta\rho_\alpha(Y) + \theta\rho_\alpha(-Y)\}.$$

It is clear that $\mathrm{DQ}_\alpha^\rho(\mathbf{X}^0) = 1$, and there exists $\tilde{\theta}$ such that $\mathrm{DQ}_\alpha^\rho(\mathbf{X}^{\tilde{\theta}}) = 0$ with $\rho_\alpha(X)+\tilde{\theta}\rho_\alpha(Y)+\tilde{\theta}\rho_\alpha(-Y) > \rho_0(X)$. Since $\beta \mapsto \rho_\beta(X)$ is strictly decreasing, its generalized inverse is continuous and we can conclude $\{\mathrm{DQ}_\alpha^\rho(\mathbf{X}^\beta) : \beta \in [0, \tilde{\theta}]\} = [0,1]$. Hence, $[0,1] \subseteq \{\mathrm{DQ}_\alpha^\rho(\mathbf{X}) : \mathbf{X} \in \mathcal{X}^n\}$. $\square$

### C.3 Constructing DQ from a single risk measure

In this section, we discuss how to construct DQ from only a single risk measure $\phi$. For commonly used risk measures like VaR and ES, a natural family $\rho$ with $\rho_\alpha = \phi$ exists. If in some applications one needs to use a different $\phi$ which does not belong to an existing family, we will need to construct a family of risk measures for $\phi$.

First, suppose that $\phi$ is MCP. A simple approach is to take $\rho_\alpha = (1-\alpha)\text{ess-sup} + \alpha\phi$ for $\alpha \in (0,1)$. Clearly, $\rho_1 = \phi$. As $\phi$ is MCP, we have $\phi(X) \leqslant \phi(\text{ess-sup}(X)) = \text{ess-sup}(X)$ for all $X \in L^\infty$. Hence, $\rho$ is a decreasing class of MCP risk measures. Therefore, $\mathrm{DQ}_1^\rho$ satisfies the six axioms in Theorem 1. Moreover, by checking the definition, this DQ has an explicit formula

$$\mathrm{DQ}_1^\rho(\mathbf{X}) = \left(\frac{\text{ess-sup}\left(\sum_{i=1}^n X_i\right) - \sum_{i=1}^n \phi(X_i)}{\text{ess-sup}\left(\sum_{i=1}^n X_i\right) - \phi\left(\sum_{i=1}^n X_i\right)}\right)_+ .$$

If $\sum_{i=1}^n X_i \leqslant \sum_{i=1}^n \phi(X_i)$, we have $\text{ess-sup}(\sum_{i=1}^n X_i) \leqslant \sum_{i=1}^n \phi(X_i)$ and $\mathrm{DQ}_1^\rho(\mathbf{X}) = 0$; this is also reflected by Theorem 3 when $\text{ess-sup}(\sum_{i=1}^n X_i) > \phi(\sum_{i=1}^n X_i)$.

For any arbitrary risk measure $\phi$, we can always define the decreasing family $\{\phi_+/\alpha : \alpha \in I\}$ for constructing DQ; here $\phi_+$ is the positive part of $\phi$. This approach leads to DQs that are also DRs.

**Proposition EC.4.** *For a given $\phi : \mathcal{X} \to \mathbb{R}_+$, let $\rho = (\phi/\alpha)_{\alpha \in (0,\infty)}$. For $\alpha \in (0,\infty)$, we have $\mathrm{DQ}_\alpha^\rho = \mathrm{DR}^\phi$. The same holds if $\rho = (b\mathbb{E} + c\phi/\alpha)_{\alpha \in (0,\infty)}$ for some $b \in \mathbb{R}$ and $c > 0$ and $\mathcal{X} = L^1$.*

*Proof.* First, we compute $\alpha^*$ by the definition of $\mathrm{DQ}^\rho_\alpha$. For any $\mathbf{X} \in (L^1)^n$,

$$\begin{aligned}\alpha^* &= \inf\left\{\beta \in (0,\infty) : b\mathbb{E}\left[\sum_{i=1}^n X_i\right] + \frac{c}{\beta}\phi\left(\sum_{i=1}^n X_i\right) \leqslant b\sum_{i=1}^n \mathbb{E}\left[X_i\right] + \sum_{i=1}^n \frac{c}{\alpha}\phi(X_i)\right\}\\ &= \inf\left\{\beta \in (0,\infty) : \frac{\phi\left(\sum_{i=1}^n X_i\right)}{\beta} \leqslant \frac{\sum_{i=1}^n \phi(X_i)}{\alpha}\right\}.\end{aligned}$$

If $\phi(\sum_{i=1}^n X_i) = 0$ and $\sum_{i=1}^n \phi(X_i) = 0$, then $\alpha^* = 0$. If $\phi(\sum_{i=1}^n X_i) > 0$ and $\sum_{i=1}^n \phi(X_i) = 0$, then $\alpha^* = \infty$ because the set on which the infimum is taken is empty. If $\phi(\sum_{i=1}^n X_i) > 0$ and $\sum_{i=1}^n \phi(X_i) > 0$, then $\alpha^* = \alpha\phi(\sum_{i=1}^n X_i)/\sum_{i=1}^n \phi(X_i)$. Hence, $\mathrm{DQ}^\rho_\alpha(\mathbf{X}) = \mathrm{DR}^\phi(\mathbf{X})$ holds for all $\mathbf{X} \in (L^1)^n$. By the same argument, for $\rho = (\phi/\alpha)_{\alpha\in(0,\infty)}$, we get $\mathrm{DQ}^\rho_\alpha(\mathbf{X}) = \mathrm{DR}^\phi(\mathbf{X})$ for all $\mathbf{X} \in \mathcal{X}^n$. $\square$

As a result of Proposition EC.4, DQ built on the family $\rho$ of the mean-SD functions given by $\rho_\alpha(X) = \mathbb{E}[X] + \mathrm{SD}(X)/\alpha$ is precisely $\mathrm{DR}^{\mathrm{SD}}$.

## C.4 Axiomatization of DQ using preferences

The axioms $[\mathrm{R}]_\phi$, $[\mathrm{N}]_\phi$ and $[\mathrm{C}]_\phi$ are formulated based on an exogenously specified risk measure $\phi$, usually by financial regulation. This choice can also be endogenized in the context of internal decision making. In this section, we provide an axiomatization of DQ as in Theorem 1 without specifying a risk measure $\phi$. We first define the preference of a decision maker over risks. A preference relation $\succeq$ is defined by a non-trivial total preorder[20] on $\mathcal{X}$. As usual, $\succ$ and $\simeq$ correspond to the antisymmetric and equivalence relations, respectively. On the preference $\succeq$ of risk, the relation $X \succeq Y$ means the agent prefers $X$ to $Y$ for any $X, Y \in \mathcal{X}$. We will use the following axioms.

[A1] $X \leqslant Y \implies X \succeq Y$.

[A2] $X \succeq Y \implies X + c \succeq Y + c$ for any $c \in \mathbb{R}$.

[A3] $X \succeq Y \implies \lambda X \succeq \lambda Y$ for any $\lambda > 0$.

[A4] For any $X \in \mathcal{X}$, there exists $c \in \mathbb{R}$ such that $X \simeq c$.

The four axioms are rather standard and we only briefly explain them. The axiom [A1] means that the agent always prefers a smaller loss. The axioms [A2] and [A3] mean that if the agent prefers one random loss over another, then this is preserved under any strictly increasing linear

[20] A preorder is a binary relation on $\mathcal{X}$, which is reflexive and transitive. A binary relation $\succeq$ is reflexive if $X \succeq X$ for all $X \in \mathcal{X}$, and transitive if $X \succeq Y$ and $Y \succeq Z$ imply $X \succeq Z$. A non-trivial total preorder is a preorder that in addition is complete, that is, $X \succeq Y$ or $Y \succeq X$ for all $X, Y \in \mathcal{X}$, and there exist at least two alternatives $X$, $Y$ such that $X$ is preferred over $Y$ strictly.

transformations. The axiom [A4] implies that any random losses can be equally favourable as a constant loss which is commonly referred to as a certainty equivalence.

A numerical representation of a preference $\succeq$ is a mapping $\phi : \mathcal{X} \to \mathbb{R}$, such that $X \succeq Y \iff \phi(X) \leqslant \phi(Y)$ for all $X, Y \in \mathcal{X}$. In other words, $\succeq$ is the preference of an agent favouring less risk evaluated via $\phi$. There is a simple relationship between preferences satisfying [A1]-[A4] and MCP risk measures.

**Lemma EC.3.** *A preference satisfies* [A1]–[A4] *if and only if it can be represented by an MCP risk measure* $\phi$.

*Proof.* The "if" statement is straightforward to check, and we will show the "only if" statement. The preference $\succeq$ can be represented by a risk measure $\phi$ through $X \succeq Y \iff \phi(X) \leqslant \phi(Y)$ for all $X, Y \in \mathcal{X}$ since $\succeq$ is separable by [A1] and [A4]; see Debreu (1954) and Drapeau and Kupper (2013). If $\phi(0) = \phi(1)$, then by using [A1]–[A3], the preference $\succeq$ is trivial, contradicting our assumption on $\succeq$. Hence, using [A1], $\phi(0) < \phi(1)$, we can further let $\phi(0) = 0$ and $\phi(1) = 1$. It is then straightforward to verify that $\phi$ is MCP from [A1]-[A3]. $\square$

Similarly to Section 3, but with the preference $\succeq$ replacing the risk measure $\phi$, we denote by $\mathbf{X} \overset{\mathrm{m}}{\simeq} \mathbf{Y}$ if $X_i \simeq Y_i$ for each $i \in [n]$, by $\mathbf{X} \overset{\mathrm{m}}{\succeq} \mathbf{Y}$ if $X_i \succeq Y_i$ for each $i \in [n]$, and by $\mathbf{X} \overset{\mathrm{m}}{\succ} \mathbf{Y}$ if $X_i \succ Y_i$ for each $i \in [n]$. With this new formulation and everything else unchanged, the axioms of rationality, normalization and continuity are now denoted by $[\mathrm{R}]_{\succeq}$, $[\mathrm{N}]_{\succeq}$ and $[\mathrm{C}]_{\succeq}$.

**Proposition EC.5.** *A diversification index* $D : \mathcal{X}^n \to \overline{\mathbb{R}}$ *satisfies* [+], [LI], [SI], $[\mathrm{R}]_{\succeq}$, $[\mathrm{N}]_{\succeq}$ *and* $[\mathrm{C}]_{\succeq}$ *for some preference* $\succeq$ *satisfying* [A1]–[A4] *if and only if* $D$ *is* $\mathrm{DQ}^{\rho}_{\alpha}$ *for some decreasing families* $\rho$ *of MCP risk measures. Moreover, in both directions of the above equivalence, it can be required that* $\rho_\alpha$ *represents* $\succeq$.

*Proof.* The proof follows from Theorem 1 by noting that Lemma EC.3 allows us to convert between a preference $\succeq$ satisfying [A1]-[A4] and an MCP risk measure $\phi$. $\square$

Theorem 2 also admits a formulation via preferences similar to Proposition EC.5.

## C.5 Uniqueness of the risk measure family representing a DQ

Proposition EC.6 below shows that the choice of the risk measure family is unique up to strictly increasing transformation of the parameter if the ordering structure on portfolio diversification is specified by a given ordering relation $\succeq$ on $\mathcal{X}^n$ that can be numerically represented by a DQ.

**Proposition EC.6.** *Let* $n \geqslant 3$, $I = [0, 1]$, $\alpha \in I$ *and* $\phi$ *is a positively homogeneous risk measure with* $\phi(Y) + \phi(-Y) > 0$ *for some* $Y \in \mathcal{X}$. *Suppose that a weak order* $\succeq$ *is numerically represented* *by both* $\mathrm{DQ}^{\rho}_{\alpha}$ *and* $\mathrm{DQ}^{\tau}_{\alpha}$ *such that* $\{\mathrm{DQ}^{\rho}_{\alpha}(\mathbf{X}) : \mathbf{X} \in \mathcal{X}^n\} = \{\mathrm{DQ}^{\tau}_{\alpha}(\mathbf{X}) : \mathbf{X} \in \mathcal{X}^n\} = [0,1]$*, where* $\rho = (\rho_\beta)_{\beta\in I}$ *and* $\tau = (\tau_\beta)_{\beta\in I}$ *are continuous decreasing families of risk measures satisfying* $\phi = \rho_\alpha = \tau_\alpha$*. Then, there exists a strictly increasing* $f : (0,\alpha) \to (0,\alpha)$ *such that* $\tau_\beta = \rho_{f(\beta)}$ *for all* $\beta \in (0,\alpha)$*.*

*Proof.* Since $\succeq$ is represented by both $\mathrm{DQ}^{\rho}_{\alpha}$ and $\mathrm{DQ}^{\tau}_{\alpha}$, there exists a strictly increasing function $g : [0,1] \to [0,1]$ such that $\mathrm{DQ}^{\rho}_{\alpha} = g(\mathrm{DQ}^{\tau}_{\alpha})$. Let $f(\beta) = g(\beta/\alpha)$ for $\beta \in (0,\alpha)$.

Assume that there exists $\beta^* \in (0,\alpha)$ such that $\tau_{\beta^*}(X) > \rho_{f(\beta^*)}(X)$. By positive homogeneity of $\phi$ and $\phi(Y)+\phi(-Y) > 0$, there exists $\varepsilon > 0$ such that $\tau_{\beta^*}(X) > \phi(X)+\phi(\varepsilon Y)+\phi(-\varepsilon Y) > \rho_{f(\beta^*)}(X)$. Let $\mathbf{X} = (X, \varepsilon Y, -\varepsilon Y, 0, \dots, 0)$. Since $\beta \mapsto \rho_\beta(X)$ and $\beta \mapsto \tau_\beta(X)$ are continuous, we have

$$g(\mathrm{DQ}^{\tau}_{\alpha}(\mathbf{X})) = f\left(\inf\{\beta \in I : \tau_\beta(X) \leqslant \phi(X) + \phi(\varepsilon Y) + \phi(-\varepsilon Y)\}\right) > f(\beta^*)$$

and

$$\mathrm{DQ}^{\rho}_{\alpha}(\mathbf{X}) = \inf\{\beta \in I : \rho_\beta(X) \leqslant \phi(X) + \phi(\varepsilon Y) + \phi(-\varepsilon Y)\} \leqslant f(\beta^*),$$

which contradicts $\mathrm{DQ}^{\rho}_{\alpha}(\mathbf{X}) = g(\mathrm{DQ}^{\tau}_{\alpha}(\mathbf{X}))$. □

# D Additional results and proofs for Section 4

In this section, we present additional results, proofs, and discussions supplementing Sections 4.2 and 4.3.

## D.1 Worst-case and best-case dependence for DQ (Section 4.2)

We assume that two random vectors $\mathbf{X}$ and $\mathbf{Y}$ have the same marginal distributions, and we study the effect of the dependence structure. We will assume that a tuple of distributions $\mathbf{F} = (F_1, \dots, F_n)$ is given and each component has a finite mean. Let

$$\mathcal{Y}_{\mathbf{F}} = \{(X_1, \dots, X_n) : X_i \sim F_i \text{ for each } i = 1, \dots, n\}.$$

For $\mathbf{X}, \mathbf{Y} \in \mathcal{Y}_{\mathbf{F}}$, we say that $\mathbf{X}$ is smaller than $\mathbf{Y}$ in sum-convex order, denoted by $\mathbf{X} \leqslant_{\mathrm{scx}} \mathbf{Y}$, if $\sum_{i=1}^n X_i \geqslant_{\mathrm{SSD}} \sum_{i=1}^n Y_i$; see Corbett and Rajaram (2006). We refer to Shaked and Shanthikumar (2007) for a general treatment of multivariate stochastic orders. With arbitrary dependence structures, the best-case value and worst-case value of $\mathrm{DQ}^{\rho}_{\alpha}$ are given by

$$\inf_{\mathbf{X}\in\mathcal{Y}_{\mathbf{F}}} \mathrm{DQ}^{\rho}_{\alpha}(\mathbf{X}) \quad \text{and} \quad \sup_{\mathbf{X}\in\mathcal{Y}_{\mathbf{F}}} \mathrm{DQ}^{\rho}_{\alpha}(\mathbf{X}).$$

For some mapping on $\mathcal{X}^n$, finding the best-case and worst-case values and structures over $\mathcal{Y}_{\mathbf{F}}$ is known as a problem of risk aggregation under dependence uncertainty; see Bernard et al. (2014) and Embrechts et al. (2015).

If $\rho = (\rho_\alpha)_{\alpha \in I}$ is a class of SSD-consistent risk measures such as ES, then, by Proposition 3, $\mathrm{DQ}_\alpha^\rho$ is consistent with the sum-convex order on $\mathcal{Y}_{\mathbf{F}}$. This leads to the following observations on the corresponding dependence structures.

(i) It is well-known (e.g., Rüschendorf (2013)) that the $\leqslant_{\mathrm{scx}}$-largest element of $\mathcal{Y}_{\mathbf{F}}$ is comonotonic, and thus a comonotonic random vector has the largest $\mathrm{DQ}_\alpha^\rho$ in this case. Note that such $\rho$ does not include VaR. Indeed, as we have seen from Proposition 5, $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = 1$ for comonotonic $\mathbf{X}$ under mild conditions, which is not equal to its largest value $n$.

(ii) In case $n = 2$, the $\leqslant_{\mathrm{scx}}$-smallest element of $\mathcal{Y}_{\mathbf{F}}$ is counter-comonotonic, and thus a comonotonic random vector has the smallest $\mathrm{DQ}_\alpha^\rho$.

(iii) For $n \geqslant 3$, the $\leqslant_{\mathrm{scx}}$-smallest elements of $\mathcal{Y}_{\mathbf{F}}$ are generally hard to obtain. If each pair $(X_i, X_j)$ is counter-monotonic for $i \neq j$, then $\mathbf{X}$ is a $\leqslant_{\mathrm{scx}}$-smallest element of $\mathcal{Y}_{\mathbf{F}}$. Pairwise counter-monotonicity puts very strong restrictions on the marginal distributions. For instance, it rules out all continuous marginal distributions; see Puccetti and Wang (2015).

(iv) If a joint mix, i.e., a random vector with a constant component-wise sum, exists in $\mathcal{Y}_{\mathbf{F}}$, then any joint mix is a $\leqslant_{\mathrm{scx}}$-smallest element of $\mathcal{Y}_{\mathbf{F}}$ by Jensen's inequality. See Puccetti and Wang (2015) and Wang and Wang (2016) for results on the existence of joint mixes. In case a joint mix does not exist, the $\leqslant_{\mathrm{scx}}$-smallest elements are obtained by Bernard et al. (2014) and Jakobsons et al. (2016) under some conditions on the marginal distributions such as monotonic densities.

In optimization problems over dependence structures (see e.g., Rüschendorf (2013) and Embrechts et al. (2015)), the above observations yield guidelines on where to look for the optimizing structures.

## D.2 Proofs and related discussions on RI and RC (Section 4.3)

Here we present the proof of Proposition 4 and an additional result (Proposition EC.7) on the properties RI and RC.

*Proof of Proposition 4.* (i) For any $n \in \mathbb{N}$, $\mathbf{X} \in (L^p)^n$ and $c \in \mathbb{R}$, by $[\mathrm{CA}]_m$ of $(\rho_\alpha)_{\alpha \in I}$,

$$\begin{aligned}\mathrm{DQ}_\alpha^\rho(\mathbf{X}, c) &= \frac{1}{\alpha} \inf \left\{ \beta \in I : \rho_\beta \left( \sum_{i=1}^n X_i + c \right) \leqslant \sum_{i=1}^n \rho_\alpha(X_i) + \rho_\alpha(c) \right\} \\ &= \frac{1}{\alpha} \inf \left\{ \beta \in I : \rho_\beta \left( \sum_{i=1}^n X_i \right) + mc \leqslant \sum_{i=1}^n \rho_\alpha(X_i) + mc \right\} = \mathrm{DQ}_\alpha^\rho(\mathbf{X}),\end{aligned}$$

and hence $\mathrm{DQ}_\alpha^\rho$ satisfies [RI].

(ii) For any $n \in \mathbb{N}$ and $\mathbf{X} \in (L^p)^n$, by [PH] of $(\rho_\alpha)_{\alpha \in I}$,

$$\mathrm{DQ}_\alpha^\rho(\mathbf{X}, \mathbf{X}) = \frac{1}{\alpha} \inf \left\{ \beta \in I : \rho_\beta \left( 2 \sum_{i=1}^n X_i \right) \leqslant 2 \sum_{i=1}^n \rho_\alpha(X_i) \right\} = \mathrm{DQ}_\alpha^\rho(\mathbf{X}),$$

and hence $\mathrm{DQ}_\alpha^\rho$ satisfies [RC]. □

**Proposition EC.7.** *Let $\phi : L^p \to \mathbb{R}$ be a continuous and law-invariant risk measure.*

*(i) Suppose that* $\mathrm{DR}^\phi$ *is not degenerate for some input dimension. Then* $\mathrm{DR}^\phi$ *satisfies* [RI] *and* [+] *if and only if $\phi$ satisfies* $[\mathrm{CA}]_0$, $[\pm]$ *and $\phi(0) = 0$.*

*(ii) If $\phi$ satisfies* [PH]*, then* $\mathrm{DR}^\phi$ *satisfies* [RC].

*Proof.* (i) We first show the "if" part. If $\phi$ satisfies $[\mathrm{CA}]_0$ and $\phi(0) = 0$, then $\phi(c) = \phi(0) = 0$ for all $c \in \mathbb{R}$. For any $n \in \mathbb{N}$, $\mathbf{X} \in (L^p)^n$ and $c \in \mathbb{R}$,

$$\mathrm{DR}^\phi(\mathbf{X}, c) = \frac{\phi\left(\sum_{i=1}^n X_i + c\right)}{\sum_{i=1}^n \phi(X_i) + \phi(c)} = \frac{\phi\left(\sum_{i=1}^n X_i\right)}{\sum_{i=1}^n \phi(X_i)} = \mathrm{DR}^\phi(\mathbf{X}).$$

Thus, $\mathrm{DR}^\phi$ satisfies [RI].

For the "only if" part, we first assume $\phi(0) \neq 0$. Since $\mathrm{DR}^\phi$ satisfies [RI], for all $n \in \mathbb{N}$, $c \in \mathbb{R}$ and $\mathbf{X} = \mathbf{0} \in \mathbb{R}^n$, we have

$$\mathrm{DR}^\phi(\mathbf{X}, c) = \frac{\phi(c)}{n\phi(0) + \phi(c)} = \mathrm{DR}^\phi(\mathbf{X}) = \frac{\phi(0)}{n\phi(0)} = \frac{1}{n}.$$

The above equality means that $\phi(c) = n\phi(0)/(n-1)$ holds for any $n \in \mathbb{N}$ and $c \in \mathbb{R}$, and thus we have $\phi(0) = 0$, which violates the assumption $\phi(0) \neq 0$. Hence, $\phi(0) = 0$.

If there exists $c_1 \in \mathbb{R}$ such that $\phi(c_1) \neq 0$, then by [RI] and $\phi(0) = 0$, we have

$$\mathrm{DR}^\phi(c_1, 0, 0, \dots, 0, c) = \frac{\phi(c_1 + c)}{\phi(c_1) + \phi(c)} = \mathrm{DR}^\phi(c_1, 0, 0, \dots, 0) = \frac{\phi(c_1)}{\phi(c_1)} = 1,$$

and thus $\phi(c_1 + c) = \phi(c_1) + \phi(c)$ as long as $\phi(c_1)$ or $\phi(c)$ is not zero. If both of $\phi(c_1)$ and $\phi(c)$ are 0, then $\phi(c_1 + c) = 0$. To sum up, $\phi$ is additive on $\mathbb{R}$. Since $\phi$ is also continuous on $\mathbb{R}$, we know that $\phi$ is linear, that is, $\phi(c) = \beta c$ for some $\beta \in \mathbb{R}$.

Suppose that there exists $X$ such that $\phi(X) \neq 0$; otherwise there is nothing to show. Using [RI] and $\phi(0) = 0$, we have, for $c \in \mathbb{R}$,

$$\mathrm{DR}^\phi(X, 0, 0, \dots, 0, c) = \frac{\phi(X + c)}{\phi(X) + \phi(c)} = \mathrm{DR}^\phi(X, 0, \dots, 0) = 1,$$

which implies $\phi(X + c) = \phi(X) + \phi(c) = \phi(X) + \beta c$.

Using the fact that $\mathrm{DR}^\phi$ is not degenerate for some dimension $n$, there exists $\mathbf{X} = (X_1, \dots, X_n)$ such that $\mathrm{DR}^\phi(\mathbf{X}) \in \mathbb{R} \setminus \{0, 1\}$. Note that $\phi(\sum_{i=1}^n X_i) \neq 0$ and $\sum_{i=1}^n \phi(X_i) \neq 0$. Hence,

$$\mathrm{DR}^\phi(\mathbf{X}, 1) = \frac{\phi(\sum_{i=1}^n X_i + 1)}{\sum_{i=1}^n \phi(X_i) + \phi(1)} = \frac{\phi(\sum_{i=1}^n X_i) + \beta}{\sum_{i=1}^n \phi(X_i) + \beta} = \mathrm{DR}^\phi(\mathbf{X}) = \frac{\phi(\sum_{i=1}^n X_i)}{\sum_{i=1}^n \phi(X_i)}.$$

This implies $\beta = 0$, $\phi(c) = 0$ for all $c \in \mathbb{R}$ and $\phi(X + c) = \phi(X)$ for all $X \in L^p$ such that $\phi(X) \neq 0$. For any $X \in L^p$ such that $\phi(X) = 0$ and $c \in \mathbb{R}$, we have

$$\mathrm{DR}^{\phi}(X, 0, \ldots, 0, c) = \frac{\phi(X + c)}{\phi(X) + \phi(c)} = \frac{\phi(X + c)}{\phi(X)} = \frac{\phi(X)}{\phi(X)} = \mathrm{DR}^{\phi}(X, 0, \ldots, 0),$$

which implies $\phi(X + c) = 0 = \phi(X)$. Therefore, $\phi$ satisfies $[\mathrm{CA}]_0$.

Finally, we show $\phi$ is either non-negative or non-positive by considering the following three cases.

(a) Assume that there exists $X \in L^p$ such that $\phi(X) + \phi(-X) > 0$. If there exists $Y \in L^p$ such that $\phi(Y) < 0$, then by continuity of $\phi$ and $\phi(0) = 0$, there exists $m > 0$ such that $0 < -\phi(mY) < \phi(X) + \phi(-X)$. We have

$$\mathrm{DR}^{\phi}(mY, X, -X, 0, \ldots, 0) = \frac{\phi(mY)}{\phi(mY) + (\phi(X) + \phi(-X))} < 0,$$

which contradicts the fact that $\mathrm{DR}^{\phi}$ is non-negative. Hence, $\phi(Y) \geqslant 0$ for all $Y \in L^{\infty}$.

(b) By the same argument, if there exists $X \in L^p$ such that $\phi(X) + \phi(-X) < 0$, then $\phi(Y) \leqslant 0$ for all $Y \in L^{\infty}$.

(c) Assume $\phi(X) + \phi(-X) = 0$ for all $X \in L^{\infty}$. Suppose that there exists $Y \in L^{\infty}$ such that $\phi(Y) < 0$. Using Lemma 1 of Wang and Wu (2020) again, there exist $Z, Z' \in L^{\infty}$ satisfying $Z \stackrel{\mathrm{d}}{=} Z'$ and $Z - Z' \stackrel{\mathrm{d}}{=} Y - \mathbb{E}[Y]$. For $\mathbf{Z} = (Z, -Z', 0, \ldots, 0)$, using the law invariance of $\phi$, we have

$$\mathrm{DR}^{\phi}(\mathbf{Z}) = \frac{\phi\left(Z - Z'\right)}{\phi(Z) + \phi(-Z')} = \frac{\phi\left(Y - \mathbb{E}[Y]\right)}{\phi(Z) + \phi(-Z')} = \frac{\phi\left(Y\right)}{\phi(Z) + \phi(-Z)} = \frac{\phi(Y)}{0} = -\infty,$$

which contradicts $\mathrm{DR}^{\phi}(\mathbf{Z}) \geqslant 0$. Hence, $\phi(X) \geqslant 0$ for all $X \in L^{\infty}$. Together with $\phi(X) + \phi(-X) = 0$, we get $\phi(X) = 0$. To extend this to $L^p$, we simply use continuity. For $X \in L^p$, let $Y_M = (X \wedge M) \vee (-M)$. Hence, $Y_M \in L^{\infty}$ and $Y_M \xrightarrow{L^p} X$ as $M \to \infty$. As a result, we have $\phi(X) = \lim_{M \to \infty} \phi(Y_M) = 0$.

In conclusion, we have $\phi(Y) \geqslant 0$ or $\phi(Y) \leqslant 0$ for all $X \in L^p$. Case (c) is not possible because it contradicts that $\mathrm{DR}^{\phi}$ is not degenerate. Cases (a) and (b) are possible, corresponding to, for instance, (a) $\phi = \mathrm{SD}$; (b) $\phi = -\mathrm{SD}$.

(ii) If $\phi$ satisfies [PH], then for any $n \in \mathbb{N}$ and $\mathbf{X} \in (L^p)^n$,

$$\mathrm{DR}^{\phi}(\mathbf{X}, \mathbf{X}) = \frac{\phi\left(2 \sum_{i=1}^n X_i\right)}{2 \sum_{i=1}^n \phi(X_i)} = \frac{\phi\left(\sum_{i=1}^n X_i\right)}{\sum_{i=1}^n \phi(X_i)} = \mathrm{DR}^{\phi}(\mathbf{X}).$$

Hence, $\mathrm{DR}^{\phi}$ satisfies [RC]. □

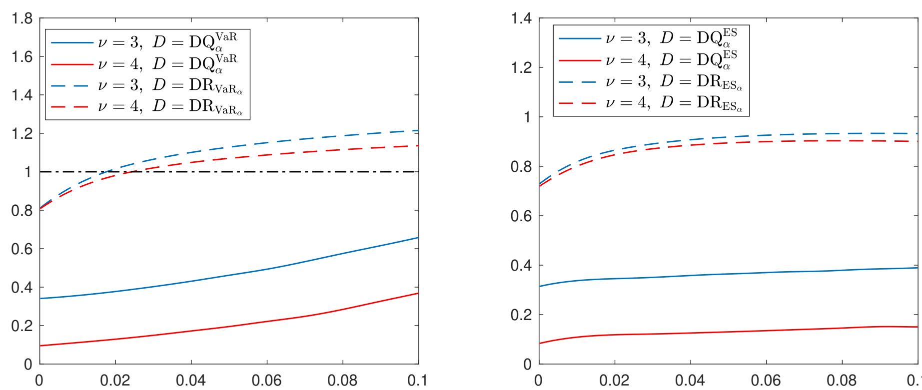
In Proposition EC.7, we show that if [RI] is assumed, then the only option for DR is to use a non-negative $\phi$ (we can use $-\phi$ if $\phi$ is non-positive) such as var or SD. By Proposition EC.4, all such DRs belong to the class of DQs.

# E Additional results and proofs for Section 5

In this section, we present the proof for Proposition 5 and an additional numerical result to complement those in Section 5.2.

*Proof of Proposition 5.* This statement on ES follows from Proposition 1; for the one on VaR, see Theorem 1 (i) of Han et al. (2023). $\square$

Figure EC.1. $D(\mathbf{Y}')/D(\mathbf{Y})$ based on VaR and ES for $\alpha \in (0, 0.1]$ with fixed $n = 10$




We look at the models $\mathbf{Y}'$ and $\mathbf{Y}$ in the setting of Tables 2 and 3. In Figure EC.1, we observe that the values of $D(\mathbf{Y}')/D(\mathbf{Y})$ for $D = \mathrm{DQ}_\alpha^{\mathrm{VaR}}$ or $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ are always smaller than 1 for $\alpha \in (0, 0.1]$, while the values of $D(\mathbf{Y}')/D(\mathbf{Y})$ for $D = \mathrm{DR}^{\mathrm{VaR}_\alpha}$ are only smaller than 1 when $\alpha$ is relatively small. We always observe that, if the desired relation $D(\mathbf{Y}')/D(\mathbf{Y}) < 1$ holds for $D = \mathrm{DR}^{\mathrm{VaR}_\alpha}$ or $\mathrm{DR}^{\mathrm{ES}_\alpha}$ then it holds for $D = \mathrm{DQ}_\alpha^{\mathrm{VaR}}$ or $\mathrm{DQ}_\alpha^{\mathrm{ES}}$, but the converse does not hold. This means that if the iid model is preferred to the common shock model by DR, then it is also preferred by DQ, but in many situations, it is only preferred by DQ not by DR. Similarly to Tables 2 and 3, the iid normal model shows a stronger diversification according to DQ, and this is not the case for DR.

# F Proofs for Section 6

*Proof of Proposition 6.* For the case of $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X})$, (4) in Theorem 4 gives that to minimize $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X})$ is equivalent to minimize

$$\mathbb{P}\left(\mathbf{w}^\top \mathbf{X} > \mathbf{w}^\top \mathbf{x}_\alpha^{\mathrm{VaR}}\right) = \mathbb{P}\left(\mathbf{w}^\top \left(\mathbf{X} - \mathbf{x}_\alpha^{\mathrm{VaR}}\right) > 0\right) \quad \text{over } \mathbf{w} \in \Delta_n.$$

Next, we discuss the case of $\mathrm{DQ}^{\mathrm{ES}}_\alpha(\mathbf{X})$. Let $f(\mathbf{v}) = \mathbb{E}[(\mathbf{v}^\top(\mathbf{X}-\mathbf{x}^{\mathrm{ES}}_\alpha)+1)_+]$ for $\mathbf{v} \in \mathbb{R}^n_+$. It is clear that $f$ is convex. Furthermore, for any $i \in [n]$, we have, for almost every $\mathbf{v} \in \mathbb{R}^n_+$,

$$\begin{aligned}\frac{\partial f}{\partial v_i}(\mathbf{v}) &= \mathbb{E}\left[(X_i - \mathrm{ES}_\alpha(X_i))\mathbb{1}_{\{\mathbf{v}^\top(\mathbf{X}-\mathbf{x}^{\mathrm{ES}}_\alpha)+1>0\}}\right]\\ &= \mathbb{E}\left[(X_i - \mathrm{ES}_\alpha(X_i))\mathbb{1}_{\{\{\mathbf{v}^\top(\mathbf{X}-\mathbf{x}^{\mathrm{ES}}_\alpha)+1>0\}\cap\{X_i-\mathrm{ES}_\alpha(X_i)>0\}\}}\right]\\ &\quad+ \mathbb{E}\left[(X_i - \mathrm{ES}_\alpha(X_i))\mathbb{1}_{\{\{\mathbf{v}^\top(\mathbf{X}-\mathbf{x}^{\mathrm{ES}}_\alpha)+1>0\}\cap\{X_i-\mathrm{ES}_\alpha(X_i)<0\}\}}\right].\end{aligned}$$

The set $\{(\mathbf{v}^\top\mathbf{X}-\mathbf{x}^{\mathrm{ES}}_\alpha)+1>0\}\cap\{X_i-\mathrm{ES}_\alpha(X_i)>0\}$ increases in $v_i$ and the set $\{(\mathbf{v}^\top\mathbf{X}-\mathbf{x}^{\mathrm{ES}}_\alpha)+1>0\}\cap\{X_i-\mathrm{ES}_\alpha(X_i)<0\}$ decreases in $v_i$. Hence, $v_i \mapsto \partial f/\partial v_i(\mathbf{v})$ is increasing. Furthermore, $\partial f/\partial v_i(\mathbf{v}) \to \mathbb{E}[(X_i-\mathrm{ES}_\alpha(X_i))\mathbb{1}_{\{X_i-\mathrm{ES}_\alpha(X_i)>0\}}] > 0$ as $v_i \to \infty$. Also, $\partial f/\partial v_i(\mathbf{v}) \to \mathbb{E}[X_i - \mathrm{ES}_\alpha(X_i)] < 0$ as $\mathbf{v} \downarrow \mathbf{0}$ component-wise. Hence, there exists a minimizer $\mathbf{v}^*$ of the problem $\min_{\mathbf{v}\in\mathbb{R}^n_+\setminus\{\mathbf{0}\}} \mathbb{E}[(\mathbf{v}^\top(\mathbf{X}-\mathbf{x}^{\mathrm{ES}}_\alpha)+1)_+]$.

Let $A = \{\mathbf{v}\in\mathbb{R}^n_+\setminus\{\mathbf{0}\} : \mathbb{P}(\mathbf{v}(\mathbf{X}-\mathbf{x}^{\mathrm{ES}}_\alpha)>0)>0\}$ and $B = \{\mathbf{v}\in\mathbb{R}^n_+\setminus\{\mathbf{0}\} : \mathbb{P}(\mathbf{v}(\mathbf{X}-\mathbf{x}^{\mathrm{ES}}_\alpha)>0)=0\}$. If $B$ is empty, it is clear that $\min_{\mathbf{w}\in\Delta_n} \mathrm{DQ}^{\mathrm{ES}}_\alpha(\mathbf{w}\odot\mathbf{X}) = \min_{\mathbf{v}\in\mathbb{R}^n_+\setminus\{\mathbf{0}\}} \mathbb{E}[(\mathbf{v}^\top(\mathbf{X}-\mathbf{x}^{\mathrm{ES}}_\alpha)+1)_+]$ by Theorem 4.

If $B$ is not empty, assume $\mathbf{v}^* \in A$. For any $\mathbf{v}_A \in A$, $\mathbf{v}_B \in B$ and $k>0$, we have

$$\mathbb{E}\left[\left((\mathbf{v}_A + k\mathbf{v}_B)^\top(\mathbf{X}-\mathbf{x}^{\mathrm{ES}}_\alpha)+1\right)_+\right] \leqslant \mathbb{E}\left[\left(\mathbf{v}_A^\top(\mathbf{X}-\mathbf{x}^{\mathrm{ES}}_\alpha)+1\right)_+\right].$$

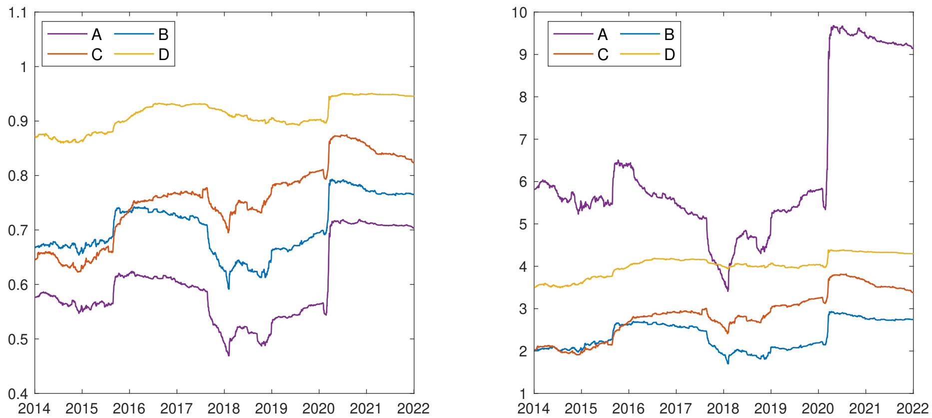
This implies $f(\mathbf{v}^*+k\mathbf{v}_B) = f(\mathbf{v}^*)$ for all $k>0$, which contradicts $\partial f/\partial v_i(\mathbf{v}) > 0$ as $v_i \to \infty$. Hence, we have $\mathbf{v}^* \in B$. For $\mathbf{w}^* = \mathbf{v}^*/\|\mathbf{v}^*\|$, we have $\mathbb{P}((\mathbf{w}^*)^\top(\mathbf{X}-\mathbf{x}^{\mathrm{ES}}_\alpha)>0) = 0$ and $\mathrm{DQ}^{\mathrm{ES}}_\alpha(\mathbf{w}^*\odot\mathbf{X}) = 0$ by Theorem 4, which means that $\mathbf{w}^*$ is the minimizer of the problem $\min_{\mathbf{w}\in\Delta_n} \mathrm{DQ}^{\mathrm{ES}}_\alpha(\mathbf{w}\odot\mathbf{X})$. □

# G Additional empirical results for Section 7

In this section, we present some omitted empirical results to complement those in Sections 7.2 and 7.3. In Section 7.2, the values of DQs based on VaR and ES are reported under different portfolio compositions of stocks during the period from 2014 to 2022. Using the same stock compositions in (A)-(D), we calculate the values of DRs based on SD and var (recall that they are also DQs), to see how they perform. The results are reported in Figure EC.2.

We can see that the same intuitive order (A)⩽(B)⩽(C)⩽(D) as in Figure 3 in Section 7.2 holds for $\mathrm{DR}^{\mathrm{SD}}$, showing some consistency between DQs based on VaR and ES and $\mathrm{DR}^{\mathrm{SD}}$. The values of $\mathrm{DR}^{\mathrm{SD}}$ are between 0 and 1. On the other hand, the values of $\mathrm{DR}^{\mathrm{var}}$ are all larger than 1, and portfolio (A) of 20 stocks has the weakest diversification effect according to $\mathrm{DR}^{\mathrm{var}}$ among the four compositions. This is not in line with our intuition, but is to be expected since variance has a different scaling effect than SD, and more correlated stocks lead to a larger value

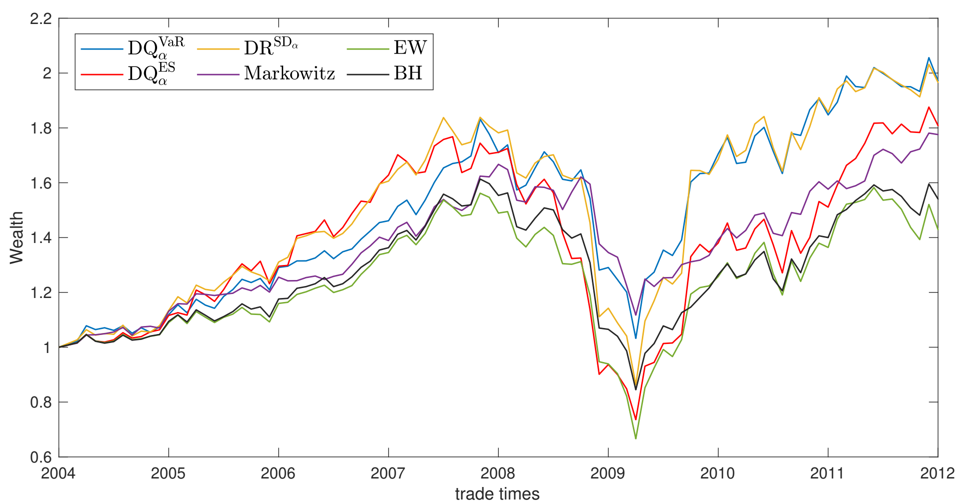
Figure EC.2. DRs based on SD (left) and var (right)




of $\mathrm{DR}^{\mathrm{var}}$ in general. For example, $\mathrm{DR}^{\mathrm{var}}$ equals 1 even for an iid normal model of arbitrarily large dimension (which is often considered as quite well-diversified), and $\mathrm{DR}^{\mathrm{var}}$ equals $n$ if the portfolio has one single asset. These observations show that $\mathrm{DR}^{\mathrm{var}}$ is difficult to interpret if it is used to measure diversification across dimensions.

In Section 7.3, we used the period from January 3, 2012, to December 31, 2021, to build up the portfolios. Next, we consider two different datasets from Section 7.3, first using the period 2002-2011 and second using 20 instead of 40 stocks, to see how the results vary.

Figure EC.3. Wealth processes for portfolios, 40 stocks, Jan 2004 - Dec 2011




For the first experiment, we choose the four largest stocks from each of the 10 different sectors of S&P 500 ranked by market cap in 2002 as the portfolio compositions and use the

Figure EC.4. Cumulative portfolio weights, 40 stocks, Jan 2004 - Dec 2011

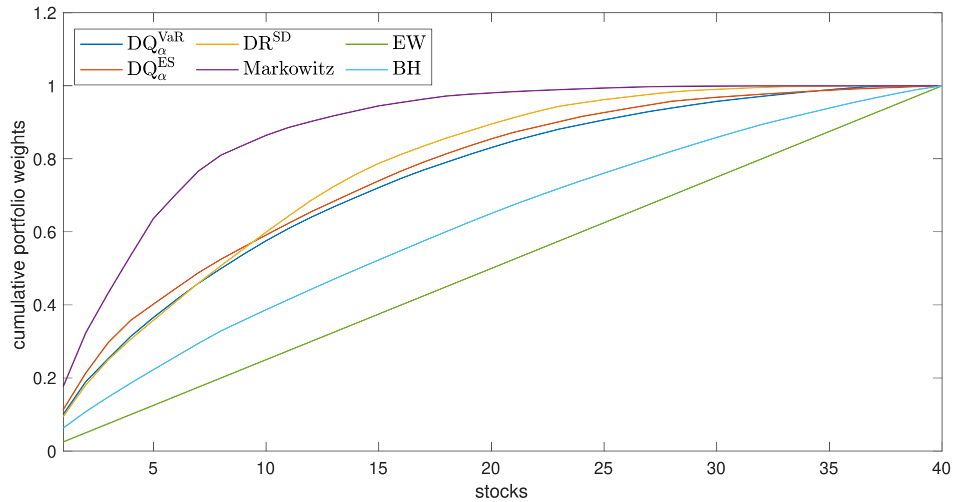



Table EC.1. Annualized return (AR), annualized volatility (AV), Sharpe ratio (SR) and average trading proportion (ATP) for different portfolio strategies from Jan 2004 to Dec 2011

| % | $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ | $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ | $\mathrm{DR}^{\mathrm{SD}}$ | Markowitz | EW | BH |
|---|---|---|---|---|---|---|
| AR | 9.46 | 8.13 | 9.10 | 7.98 | 5.30 | 6.23 |
| AV | 16.65 | 21.45 | 20.92 | 11.98 | 20.15 | 15.53 |
| SR | 30.48 | 17.47 | 22.58 | 30.06 | 4.57 | 11.94 |
| ATP | 37.23 | 28.59 | 20.24 | 24.56 | 5.04 | 0 |

period from January 3, 2002, to December 31, 2011, to build up the portfolio. The risk-free rate $r = 4.38\%$, and the target annual expected return for the Markowitz portfolio is set to 5% due to infeasibility of setting 10%. The results are reported in Figures EC.3, EC.4 and Table EC.1.

Table EC.2. Annualized return (AR), annualized volatility (AV), Sharpe ratio (SR) and average trading proportion (ATP) for different portfolio strategies from Jan 2014 to Dec 2021

| % | $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ | $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ | $\mathrm{DR}^{\mathrm{VaR}_\alpha}$ | $\mathrm{DR}^{\mathrm{ES}_\alpha}$ | $\mathrm{DR}^{\mathrm{SD}}$ | Markowitz | EW | BH |
|---|---|---|---|---|---|---|---|---|
| AR | 13.54 | 14.79 | 12.77 | 13.85 | 14.37 | 8.59 | 12.74 | 14.22 |
| AV | 13.43 | 15.90 | 14.41 | 14.53 | 14.29 | 12.74 | 14.68 | 13.96 |
| SR | 79.69 | 75.17 | 68.89 | 75.79 | 80.67 | 45.14 | 67.40 | 81.54 |
| ATP | 16.07 | 19.24 | 64.77 | 57.56 | 11.81 | 15.19 | 4.45 | 0 |

Figure EC.5. Wealth processes for portfolios, 20 stocks, Jan 2014 - Dec 2021

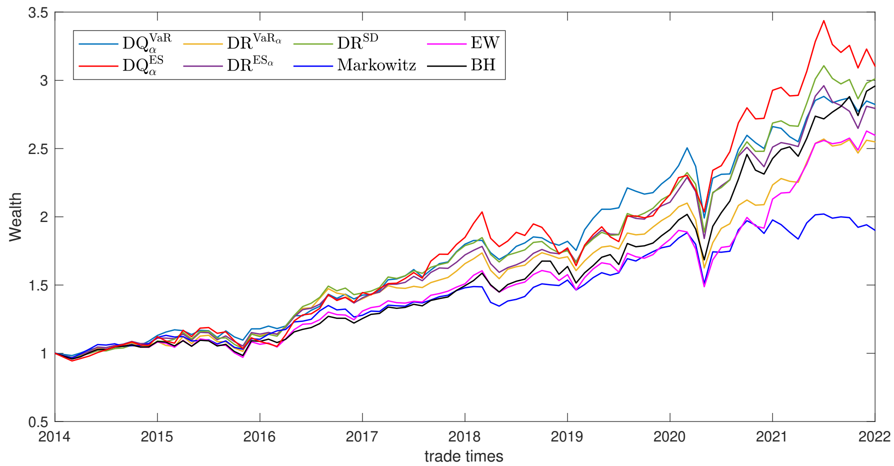
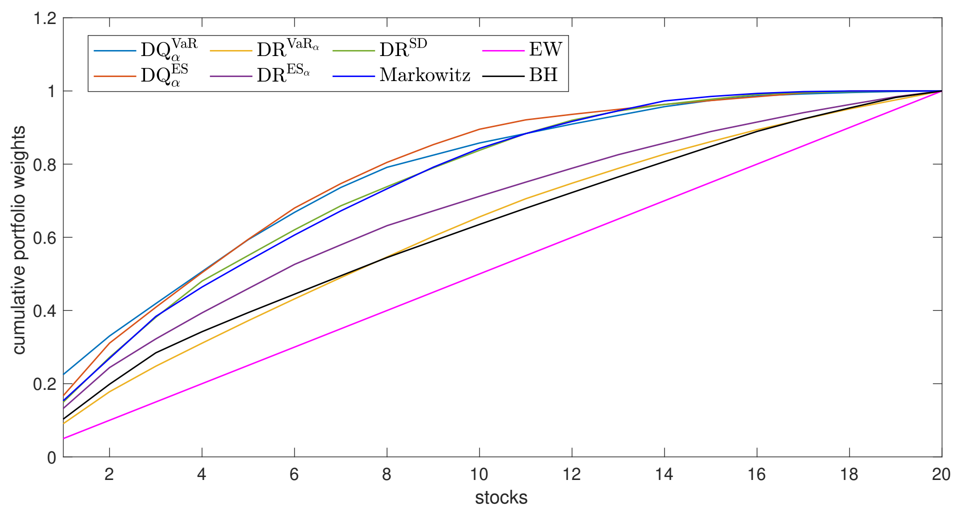



Figure EC.6. Cumulative portfolio weights, 20 stocks, Jan 2014 - Dec 2021




For the second experiment, we choose the top two stocks from each sector to build the portfolios, and all other parameters are the same as in Section 7.3. The results, including two other portfolios built by $\mathrm{DR}^{\mathrm{VaR}_\alpha}$ and $\mathrm{DR}^{\mathrm{ES}_\alpha}$, are reported in Figures EC.5 and EC.6 and Table EC.2. Since we do not find an efficient algorithm for computing $\mathrm{DR}^{\mathrm{VaR}_\alpha}$ and $\mathrm{DR}^{\mathrm{ES}_\alpha}$, we use the preceding 500 trading days to compute the optimal portfolio weights using the random sampling method, which is relatively slow and not very stable. (If the previous month has an optimal weight $\mathbf{w}_{t-1}^*$, then $10^5$ new weights are sampled from $\lambda\mathbf{w}_{t-1}^*+(1-\lambda)\Delta_n$, where $\lambda$ is chosen as 0.9. Tie-breaking is done by picking the one that is closest to $\mathbf{w}_{t-1}^*$. We set $\mathbf{w}_0^* = (1/n, \dots, 1/n)$.) The results show similar observations to those in Section 7.3. The additional observation from

Table EC.2 is that DR strategies have much larger ATP than the others, but this may be partially caused by our random sampling algorithms to optimize DR.