Risk Concentration and the Mean-Expected Shortfall Criterion

Xia Han[*]   Bin Wang[†]   Ruodu Wang[‡]   Qinyu Wu[§]

17th September 2022

Abstract

Expected Shortfall (ES, also known as CVaR) is the most important coherent risk measure in finance, insurance, risk management, and engineering. Recently, Wang and Zitikis (2021) put forward four economic axioms for portfolio risk assessment and provide the first economic axiomatic foundation for the family of ES. In particular, the axiom of no reward for concentration (NRC) is arguably quite strong, which imposes an additive form of the risk measure on portfolios with a certain dependence structure. We move away from the axiom of NRC by introducing the notion of *concentration aversion*, which does not impose any specific form of the risk measure. It turns out that risk measures with concentration aversion are functions of ES and the expectation. Together with the other three standard axioms of monotonicity, translation invariance and lower semicontinuity, concentration aversion uniquely characterizes the family of ES. In addition, we establish an axiomatic foundation for the problem of mean-ES portfolio selection and new explicit formulas for convex and consistent risk measures. Finally, we provide an economic justification for concentration aversion via a few axioms on the attitude of a regulator towards dependence structures.

KEYWORDS: Risk measures, dependence, tail event, concentration aversion, portfolio selection.

1   Introduction

The quantification of market risk for pricing, portfolio selection, and risk management purposes has long been a point of interest to researchers and practitioners in finance. Since the early 1990s,

---

[*]Department of Statistics and Actuarial Science, University of Waterloo, Canada. E-mail: `x235han@uwaterloo.ca`

[†]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China. E-mail: `wangbin@amss.ac.cn`

[‡]Department of Statistics and Actuarial Science, University of Waterloo, Canada. E-mail: `wang@uwaterloo.ca`

[§]Department of Statistics and Finance, University of Science and Technology of China, China. E-mail: `wu051555@mail.ustc.edu.cn`

Value-at-Risk (VaR) has been the leading tool for measuring market risk because of its conceptual simplicity and easy evaluation. It is well known that VaR has been criticized because of its fundamental deficiencies; for instance, it does not account for "tail risk" and its lack of subadditivity or convexity; see e.g., Daníelsson et al. (2001). These limitations have prompted the implementation of an alternative measure of risk, the Expected Shortfall (ES), also known as CVaR, TVaR and AVaR in various contexts.

As the dominating class of risk measures in financial practice, ES has many nice theoretical properties. In particular, ES satisfies the four axioms of coherence (Artzner et al., 1999), and it is also additive for comonotonic risks (Kusuoka, 2001), and thus it is a convex Choquet integral (Yaari, 1987; Schmeidler, 1989). In addition to these theoretic properties, ES admits a nice representation as the minimum of expected losses (Rockafellar and Uryasev, 2002), which allows for convenience in convex optimization. In the recent Fundamental Review of the Trading Book (BCBS, 2016, 2019), the Basel Committee on Banking Supervision proposed a shift from the 99% VaR to the 97.5% ES as the standard risk measure for internal models in market risk assessment. All the above reasons make ES arguably the most important risk measure in banking practice and insurance regulation.

The study of axiomatic characterization of risk measures provides guidelines for choosing among various choices of risk measures. Several sets of axioms have been established to characterize VaR, including those of Chambers (2009), Kou and Peng (2016), He and Peng (2018), and Liu and Wang (2021). Fewer scholars analyze the axiomatic foundation for ES. In some papers, ES is identified based on its joint property with the corresponding VaR; in particular, ES is the smallest law-invariant coherent risk measure dominating VaR (Delbaen, 2002), the only coherent distortion risk measure co-elicitable with VaR (Wang and Wei, 2020), and the only coherent Bayes risk measure with VaR being its Bayes estimator (Embrechts et al., 2021).

Different from the above literature relying on VaR to identify ES, Wang and Zitikis (2021) proposed four axioms, monotonicity, law invariance, prudence and no reward for concentration (NRC), in the context of portfolio risk assessment, which jointly characterize the family of ES. The key axiom [NRC] means that a *concentrated* portfolio, whose components incur large losses simultaneously in a stress event A of regulatory interest, does not receive any capital reduction. This axiom reflects two important common features in portfolio risk assessment. The first is that regulators are concerned with tail events, which are rare events (i.e., have small probabilities) in which risky positions incur large losses, and the second concerns diversification and risk concentration. Mathematically, [NRC] is quite a strong property as it gives the additive form of the risk measure on concentrated portfolios. Hence, [NRC] does not apply in contexts where values of the underlying

risk measures are not meant to be additive, such as risk rating or ranking decisions; nevertheless, ES can be used for rating or ranking credit risks, as in, e.g., Guo et al. (2020).

The main purpose of this paper is the study of an alternative, more natural, property which does not impose any specific functional form and can replace [NRC]. This alternative property will be called *concentration aversion* (CA), whose desirability in regulation can be justified by the arguments of Wang and Zitikis (2021) who extensively discussed issues related to risk concentration and diversification benefit. Although reflecting similar economic considerations, none of [CA] and [NRC] implies the other. As [CA] is free of any particular functional form, it is invariant under any strictly increasing transforms on the risk measure, and this invariance is not shared by [NRC]. In Section 2, some preliminaries about risk measures are collected, and the key property [CA] is formulated. We show that together with law invariance, [CA] is equivalent to a more mathematically tractable property [p-CA] in Proposition 1.

As the first main result of this paper, Theorem 1 in Section 3 says that the risk measures satisfying [p-CA] are precisely functions of ES and expectation. The proof of Theorem 1 is quite different from techniques used in Wang and Zitikis (2021), and it requires some novel mathematical tools including a recent advanced result from Wang and Wu (2020). We proceed to illustrate in Theorem 2 that [CA] characterizes the mean-ES criteria in portfolio selection, thus providing an axiomatic foundation for such optimization problems. The mean-risk portfolio selection problem has a long history since Markowitz (1952); see also Basak and Shapiro (2001), Rockafellar and Uryasev (2002) and the more recent Herdegen and Khan (2022).

In Section 4, we concentrate on monetary risk measures, the most popular type of risk measures; for a comprehensive treatment, see Föllmer and Schied (2016). It turns out that monetary risk measures satisfying [CA] admit a simple representation as a special type of mean-deviation risk measures (Theorem 3), where the deviation is measured by a transformed difference between ES and the mean. Quite surprisingly, if we further impose lower semi-continuity, then such a monetary risk measure has to be an ES (Theorem 4). Compared to the main result of Wang and Zitikis (2021), our new characterization enhances the axiomatic theory for ES as no particular additive form needs to be assumed ex ante. Moreover, we obtain characterizations for coherent, convex, or consistent risk measures (Mao and Wang, 2020) satisfying [CA], giving rise to many new explicit examples of convex and nonconvex consistent risk measures.

In the main part of the paper, the domain of risk measures of interest is chosen as the set of bounded random variables. Generalizations and technical remarks related extending the above results to larger spaces of random variables are discussed in Section 5. In particular, all our main

results can be readily extended to L^q spaces for $q \geqslant 1$ under a continuity assumption.

Finally, in Section 6, we provide an endogenous economic reasoning for concentration aversion via a few axioms on the attitude of a regulator towards bivariate dependence structures. We show in Theorems 9 and 6 that, the four natural requirements of non-diversifiability, dependence monotonicity, convexity and maximality jointly characterize the dependence structures modeling risk concentration in this paper. This result provides a theoretical support to [CA], as well as [NRC] of Wang and Zitikis (2021), in the context of regulatory risk measures. To the best of our knowledge, there is no similar study in the literature on axiomatizing sets of dangerous dependence structures.

2 Risk concentration and concentration aversion

Throughout this paper, we work with an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$. All equalities and inequalities of functionals on $(\Omega, \mathcal{F}, \mathbb{P})$ are under $\mathbb{P}$ almost surely ($\mathbb{P}$-a.s.) sense. A risk measure ρ is a mapping from $\mathcal{X}$ to $(-\infty, \infty]$, where $\mathcal{X}$ is a convex cone of random variables representing losses faced by financial institutions. For $q \in (0, \infty)$, denote by $L^q = L^q(\Omega, \mathcal{F}, \mathbb{P})$ the set of all random variables X with $\mathbb{E}[|X|^q] < \infty$ 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 denoted by the space of all random variables. Positive values of random variables in $\mathcal{X}$ represent one-period losses. We write $X \stackrel{\mathrm{d}}{=} Y$ if two random variables X and Y have the same law. We first collect the key concepts of tail events and risk concentration as in Wang and Zitikis (2021).

Definition 1 (Tail events and risk concentration). Let X be a random variable and $p \in (0, 1)$.

  (i) A *tail event* of X is an event $A \in \mathcal{F}$ with $0 < \mathbb{P}(A) < 1$ such that $X(\omega) \geqslant X(\omega')$ holds for a.s. all $\omega \in A$ and $\omega' \in A^c$, where A^c stands for the complement of A.

 (ii) A *p-tail event* of X is a tail event of X with probability $1 - p$.

(iii) A random vector $(X_1, \ldots, X_n)$ is *p-concentrated* if its components share a common p-tail event.

(iv) A random vector $(X_1, \ldots, X_n)$ is *comonotonic* if there exists a random variable Z and increasing functions $f_1, \ldots, f_n$ on $\mathbb{R}$ such that $X_i = f_i(Z)$ a.s. for every $i = 1, \ldots, n$.

The terminology that a p-tail event has probability $1 - p$ stems from the regulatory language where, for instance, a tail event with probability 1% corresponds to the calculation of a 99% VaR. A random vector $(X_1, \ldots, X_n)$ is p-concentrated for all $p \in (0, 1)$ if and only if it is comonotonic;

4

see Theorem 4 of Wang and Zitikis (2021). Hence, p-concentration can be seen as a weaker notion of positive dependence than comonotonicity, which is a popular notion in the axiomatic characterization of risk functionals and preferences; see e.g., Yaari (1987) and Schmeidler (1989). For more details and a real-data example on p-concentration, see Wang and Zitikis (2021).

Next, we define the two important risk measures in banking and insurance practice. The VaR at level $p \in (0, 1)$ is the functional $\mathrm{VaR}_p : L^0 \to \mathbb{R}$ defined by

$$\mathrm{VaR}_p(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leqslant x) \geqslant p\},$$

which precisely is the left p-quantile of X, and the ES at level $p \in (0, 1)$ is the functional $\mathrm{ES}_p : L^1 \to \mathbb{R}$ defined by

$$\mathrm{ES}_p(X) = \frac{1}{1-p} \int_p^1 \mathrm{VaR}_s(X) \mathrm{d}s.$$

In this paper, terms such as increasing or decreasing functions are in the non-strict sense. A few axioms and properties of a risk measure ρ on $\mathcal{X}$ are collected below, where all random variables are tacitly assumed to be in the space $\mathcal{X}$.

[M] Monotonicity: $\rho(X) \leqslant \rho(Y)$ whenever $X \leqslant Y$ (pointwise).

[TI] Translation invariance: $\rho(X + c) = \rho(X) + c$ for all $c \in \mathbb{R}$.

[LI] Law invariance: $\rho(X) = \rho(Y)$ whenever $X \overset{\mathrm{d}}{=} Y$.

[P] Lower semicontinuity: $\liminf_{n \to \infty} \rho(X_n) \geqslant \rho(X)$ if $X_n \to X$ (pointwise).

[NRC] No reward for concentration: There exists an event $A \in \mathcal{F}$ such that $\rho(X + Y) = \rho(X) + \rho(Y)$ holds for all risks X and Y sharing the tail event A.

[p-TA] p-tail additivity with $p \in (0, 1)$: $\rho(X + Y) = \rho(X) + \rho(Y)$ for all X and Y sharing a p-tail event.[1]

Wang and Zitikis (2021) proposed [M], [LI], [P] and [NRC] as four axioms, and showed that they together characterize the class of ES up to scaling; see their Theorem 1 and Endnote 14.[2] Axioms [M], [TI], and [LI] are standard in the literature of monetary risk measures. Axiom [P] is motivated by the statistical consideration of robustness (Hampel (1971)). It reflects the idea that if the loss X is statistically modelled using a truthful approximation (e.g., via a consistent distribution

---

[1] The property [p-TA] is called p-additivity by Wang and Zitikis (2021).

[2] Lower semicontinuity is called prudence by Wang and Zitikis (2021) and hence the abbreviation [P].

estimator), then the approximated risk model should not underreport the capital requirement as the approximation error tends to zero. Therefore, Axiom [P] is also a natural requirement for a reasonable risk measure used in practice, as argued by Wang and Zitikis (2021). The inequality in [M] and convergence in [P] are formulated in a pointwise sense, making these axioms weaker and the corresponding characterization results stronger. Nevertheless, as discussed in Remark 1 of Wang and Zitikis (2021), one can replace "$X \leqslant Y$ (pointwise)" in [M] by "$X \leqslant Y$ $\mathbb{P}$-a.s." or "$X \preceq_{\text{st}} Y$[3]", and replace pointwise convergence in [P] by in probability, in distribution, or a.s. convergence.[4] All results in this paper would still hold with the above modified versions.

Since Axiom [P] is relatively new in the literature, we discuss it in some more detail. Cont et al. (2010) showed that Hampel's notion of robustness for law-invariant risk measure is equivalent to continuity with respect to convergence in distribution. For law-invariant risk measures, [P] is equivalent to lower semi-continuity with respect to convergence in distribution, and hence it is half of Hampel's robustness. Moreover, it is the most relevant half of Hampel's robustness: Risk estimation can be asymptotically conservative or accurate, but it cannot be asymptotically anti-conservative. As shown by Wang and Zitikis (2021), ES satisfies lower semicontinuity with respect to convergence in distribution, but not continuity (as already shown by Cont et al. (2010)). Axiom [P] will be essential in Theorem 4 where ES is characterized among a large class of monetary risk measures, and it is not needed in our other results, except for Proposition 6 and Theorem 6 which are extensions of Theorem 4.

As discussed above, [NRC] intuitively means that a concentrated portfolio, whose components incur large losses simultaneously in the stress event A, does not receive any diversification benefit. For a law-invariant risk measure, the property [NRC] is equivalent to [p-TA] for some $p \in (0,1)$; see Proposition 4 of Wang and Zitikis (2021). Thus, it suffices to work with [p-TA] when analyzing the property [NRC] of law-invariant risk measures. As [NRC] imposes an additive form for the risk measure evaluated on concentrated portfolios, it may be seen as a quite strong property mathematically, and it cannot be used in a context, such as rating or ranking risks, where values of risk measures or preference functionals are not interpreted as additive units. Therefore, finding an alternative property, without the additive form, that may replace [NRC] to characterize ES (and

---

[3]The partial order $X \preceq_{\text{st}} Y$ means $\mathbb{P}(X \leqslant t) \geqslant \mathbb{P}(Y \leqslant t)$ for all $t \in \mathbb{R}$.

[4]Continuity with respect to convergence in distribution is an equivalent formulation of robustness as shown by Hampel (1971). A related property in the literature of convex risk measures is the Fatou property, meaning that $\liminf_{n\to\infty} \rho(X_n) \geqslant \rho(X)$ whenever $\{X_n\}_{n\in\mathbb{N}}$ is a bounded sequence in $\mathcal{X}$ converging to X (pointwise). Clearly, [P] is stronger than the Fatou property, and we will see in Theorem 4 that [P] is used to the characterization of ES. The Fatou property is essential to a dual representation of convex risk measures, and we do not assume convexity in most results in our paper.

preferences induced by ES) becomes a natural problem.[5]

To address this problem, we propose the property *concentration aversion* (CA), in a way similar to [NRC] but without imposing additivity. Instead of assuming that the risk measure is additive for concentrated portfolios, the new property of [CA] requires that the risk measure (or decision maker) assigns a larger or equal value for concentrated portfolios, compared to a portfolio that is not concentrated and otherwise identical.

Definition 2. A risk measure ρ satisfies *concentration aversion* if there exists an event $A \in \mathcal{F}$ with $\mathbb{P}(A) \in (0, 1)$ such that $\rho(X + Y) \leqslant \rho(X' + Y')$ if $X \stackrel{\mathrm{d}}{=} X'$, $Y \stackrel{\mathrm{d}}{=} Y'$, and X' and Y' share the tail event A. This property is denoted by [CA].

The event A in [CA] should be interpreted as a stress event of interest to the regulator. We will show in Proposition 1 that the specification of A does not matter in characterization results on law-invariant risk measures, and only the probability of A is relevant. A similar observation can also be found in Wang and Zitikis (2021). The property [CA] has a straightforward preference interpretation; that is, with marginal distributions fixed, the decision maker prefers non-concentrated portfolios over concentrated ones. Similarly to [NRC], the desirability of [CA] for regulatory risk measures depends on whether one agrees that p-concentration represents a dangerous dependence structure of regulatory concern. This issue has been discussed by Wang and Zitikis (2021) in detail; see also BCBS (2019) for evidence and considerations in regulatory practice. In Section 6, we will provide a first axiomatic justification of [CA] from a few natural properties on the set of adverse dependence structures, and thus showing that [CA] (or [NRC]) can be motivated endogenously.

None of [CA] and [NRC] implies the other one, although they are closely related. For instance, $X \mapsto \exp(\mathbb{E}[X])$ satisfies [CA] but not [NRC], whereas $X \mapsto -\mathrm{ES}_p(X)$ satisfies [NRC] but not [CA].

Many characterization axioms in the literature, including [p-TA] and [NRC], compare the value of a risk measure applied to a portfolio with a combined value of the risk measure applied to individual risks. For instance, subadditivity means that a merger does not create extra risk (Artzner et al., 1999), convexity means diversification does not increase risk level (Föllmer and Schied, 2002), and comonotonic additivity means that a comonotonic portfolio does not receive any risk reduction (Kusuoka, 2001; Marinacci and Montrucchio, 2004). In contrast, the property [CA] is defined by comparing two portfolios, not comparing values of the specific risk measure; thus this property is free of the specific functional form. For instance, if ρ satisfies [CA], then so is $f \circ \rho$ for any increasing function f; such a feature is not shared by the above properties in the risk measure

---

[5]We thank Martin Herdegen for raising this question during a seminar at the University of Warwick in October 2020.

literature, although it widely appears in the literature of decision theory.

Similar to the translation between [NRC] and [p-TA], the property [CA] can also be translated to a mathematical property that is easier to analyze. This property, called *p-concentration aversion* [p-CA], will be the central property analyzed in this paper.

Definition 3. Let $p \in (0,1)$. A risk measure ρ satisfies *p-concentration aversion* if $\rho(X + Y) \leqslant \rho(X' + Y')$ for all (X, Y) and p-concentrated (X', Y') satisfying $X \stackrel{\mathrm{d}}{=} X'$ and $Y \stackrel{\mathrm{d}}{=} Y'$. This property is denoted by [p-CA].

We first verify that [CA] can be replaced by [p-CA] for some $p \in (0,1)$ in our subsequent analysis.

Proposition 1. *For a risk measure ρ on $\mathcal{X}$, the following are equivalent.*

(i) ρ satisfies [LI] *and* [CA].

(ii) ρ satisfies [p-CA] *for some $p \in (0,1)$.*

Proof. "(ii) $\Rightarrow$ (i)": We first show [p-CA] implies [LI]. Let $Y = Y' = 0$. Take identical distributed X and X', and note that (X', Y') is p-concentrated since Y' is a constant. Property [p-CA] implies that $\rho(X) = \rho(X + Y) \leqslant \rho(X' + Y') = \rho(X')$, and exchanging the positions of (X, Y) and (X', Y') we also have $\rho(X') = \rho(X' + Y') \leqslant \rho(X + Y) = \rho(X)$. Therefore, ρ is law invariant. To verify [CA], take any event A with probability $1 - p$, and it is straightforward that ρ satisfies [CA] with A being the stress event.

"(i) $\Rightarrow$ (ii)": Suppose that ρ satisfies [CA] with A being the stress event, and let $p = 1 - \mathbb{P}(A)$. Let $X, Y \in \mathcal{X}$ be two random variables which share a tail event B of probability $1 - p$. It suffices to show that for any $\widetilde{X}, \widetilde{Y} \in \mathcal{X}$ with $\widetilde{X} \stackrel{\mathrm{d}}{=} X$ and $\widetilde{Y} \stackrel{\mathrm{d}}{=} Y$, we have $\rho(\widetilde{X} + \widetilde{Y}) \leqslant \rho(X + Y)$. Similar to the proof of Proposition 4 in Wang and Zitikis (2021), we construct two random variables $X', Y' \in \mathcal{X}$ such that X' and Y' share the same tail event A, and (X', Y') and (X, Y) are identically distributed. Using [LI], we have $\rho(X' + Y') = \rho(X + Y)$. It then follows from [CA], $\widetilde{X} \stackrel{\mathrm{d}}{=} X'$ and $\widetilde{Y} \stackrel{\mathrm{d}}{=} Y'$ that

$$\rho(\widetilde{X} + \widetilde{Y}) \leqslant \rho\left(X' + Y'\right) = \rho(X + Y),$$

which completes the proof. $\qquad\square$

It is immediate from Theorem 5 of Wang and Zitikis (2021) that ES_p satisfies [p-CA]. Moreover, the mean $\mathbb{E}$ and convex combinations of $(\mathbb{E}, \mathrm{ES}_p)$ such as $\lambda\mathbb{E} + (1 - \lambda)\mathrm{ES}_p$ for $\lambda \in (0,1)$ also satisfy [p-CA]. For applications in regulatory risk assessment, the value of p should be close to 1, indicating

an emphasis on tail events with large losses that happen with a small probability. In BCBS (2019), the choice of p in ES_p is 0.975.

In the following sections, we will formally study risk measures with the property of [p-CA]; equivalently, they are law invariant risk measures satisfying [CA].

Remark 1. The property [p-CA] is defined for an arbitrary but fixed $p \in (0, 1)$. If we allow p to take value 0, then [p-CA] in Definition 3 degenerates to the property that $\rho(X + Y) = \rho(X' + Y')$ for any $X \stackrel{\mathrm{d}}{=} X'$ and $Y \stackrel{\mathrm{d}}{=} Y'$. Such a property is called *dependence neutrality* by Wang and Wu (2020), who showed that this property is only satisfied by a transformation of the mean.

3   Concentration aversion characterizes mean-ES criteria

In this section, we present our first main result that the property [p-CA] characterizes the class of functionals that are transformations of ES_p and the mean. In this and the next sections, we assume that ρ is a risk measure on $\mathcal{X} = L^\infty$, which is the standard choice in the risk measure literature (see e.g., Föllmer and Schied, 2016). The extension of $\mathcal{X}$ to more general spaces will be discussed in Section 5.

3.1   Two technical lemmas

We first collect two lemmas that will become useful tools in the proof of our main result. Denote by F_X the distribution function of a random variable X. Let F_X^{-1} be the left quantile function of X, i.e.,

$$F_X^{-1}(p) = \mathrm{VaR}_p(X) = \inf\{x : F_X(x) \geqslant p\}.$$

Noting that the probability space is atomless, there exists a uniform random variable U on $[0, 1]$ such that $F_X^{-1}(U) = X$ a.s.; see e.g., Lemma A.32 of Föllmer and Schied (2016). Denote by ess-inf$X$ and ess-sup$X$ the essential infimum and essential supremum of a random variable X, respectively. Moreover, define

$$L(F_X) = \text{ess-sup}X - \text{ess-inf}X = F_X^{-1}(1-) - F_X^{-1}(0+),$$

and let $T(F_X)$ be the distribution of $F_X^{-1}(U)/2 + F_X^{-1}(1 - U)/2$ for $U \sim \mathrm{U}(0, 1)$. The first lemma below discusses the relationship between $L(F_X)$ and $L(T(F_X))$. The second lemma of Wang and Wu (2020) is highly nontrivial, which gives the existence of identically distributed random variables whose difference is a pre-specified random variable with mean 0.

Lemma 1. *We have $L(F_X) \geqslant 2L(T(F_X))$ for any random variable $X \in L^\infty$.*

Proof. Write $Y = F_X^{-1}(U)/2 + F_X^{-1}(1-U)/2$. It is easy to verify that

$$\frac{F_X^{-1}(0+) + F_X^{-1}(0.5)}{2} \leqslant Y \leqslant \frac{F_X^{-1}(0.5) + F_X^{-1}(1-)}{2}.$$

Hence,

$$\text{ess-inf}\,X \leqslant \frac{F_X^{-1}(0+) + F_X^{-1}(0.5)}{2} \leqslant \text{ess-inf}\,Y \leqslant \text{ess-sup}\,Y \leqslant \frac{F_X^{-1}(0.5) + F_X^{-1}(1-)}{2} \leqslant \text{ess-sup}\,X.$$

As a consequence, we obtain

$$L(T(F_X)) \leqslant \frac{\text{ess-sup}\,X - \text{ess-inf}\,X}{2} = \frac{L(F_X)}{2},$$

thus showing the lemma. $\qquad\qquad\square$

Lemma 2 (Lemma 1 of Wang and Wu (2020))**.** *For a random variable X with $\mathbb{E}[X] = 0$, there exist identically distributed random variables V and V' such that $V - V' \overset{\mathrm{d}}{=} X$ and $L(V) = L(V') \leqslant L(X)$.*

3.2   The main characterization result

We are now ready to present our main result in this section on the characterization of functionals satisfying [p-CA]. In what follows, we denote by $\mathbb{H}$ the half-space $\{(x, y) \in \mathbb{R}^2 : x \geqslant y\}$. The proof of Theorem 1 requires sophisticated constructions of many random variables, utilizing both Lemmas 1 and 2.

Theorem 1. *Let $p \in (0, 1)$ and $\rho : L^\infty \to (-\infty, \infty]$. The following two statements hold.*

(i) ρ satisfies [p-CA] if and only if it has the form $f(\text{ES}_p, \mathbb{E})$, where $f : \mathbb{H} \to (-\infty, \infty]$ is increasing in its first argument.

(ii) ρ satisfies [M] and [p-CA] if and only if it has the form $f(\text{ES}_p, \mathbb{E})$, where $f : \mathbb{H} \to (-\infty, \infty]$ is increasing in both arguments.

Proof. (i) The sufficiency statement follows from the fact that ES_p takes its largest possible value for a p-concentrated portfolio among all portfolio vectors with given marginal distributions. To be specific, by Theorem 5 of Wang and Zitikis (2021), (X_1, X_2) is p-concentrated if and only if (X_1, X_2) maximizes the ES_p aggregation; that is, for all (X', Y') and p-concentrated (X, Y) satisfying $X \overset{\mathrm{d}}{=} X'$ and $Y \overset{\mathrm{d}}{=} Y'$, one has $\text{ES}_p(X' + Y') \leqslant \text{ES}_p(X + Y)$ and $\mathbb{E}[X' + Y'] = \mathbb{E}[X + Y]$. Hence, if ρ is of the form $f(\text{ES}_p, \mathbb{E})$ and f is increasing in its first argument, we have $\rho(X' + Y') \leqslant \rho(X + Y)$, which implies that ρ satisfies [p-CA].

We now prove the necessity statement. First, it is clear that [p-CA] implies that $\rho(X + Y) = \rho(X_1 + Y_1)$ if (X, Y) and (X_1, Y_1) are both p-concentrated and $X \overset{\mathrm{d}}{=} X_1 \in L^\infty$, $Y \overset{\mathrm{d}}{=} Y_1 \in L^\infty$.

For any $Z \in L^\infty$, denote by $m = \mathrm{VaR}_p(Z) = F_Z^{-1}(p)$, and by a and b two constants such that

$$a = \mathrm{ES}_p(Z) = \frac{1}{1-p} \int_p^1 F_Z^{-1}(t)\mathrm{d}t, \quad b = \mathrm{ES}_p^-(Z) := \frac{1}{p} \int_0^p F_Z^{-1}(t)\mathrm{d}t.$$

Note that $\mathbb{E}[Z] = (1-p)a + pb$. We aim to prove $\rho(Z) = \rho(Z^*)$ where $Z^* \sim (1-p)\delta_a + p\delta_b$, which justifies that $\rho(Z)$ is determined only by values of $\mathrm{ES}_p(Z)$ and $\mathbb{E}[Z]$.

There is nothing to show if $a = b$, which implies that Z is a constant and thus $Z = Z^*$ a.s. We will assume $a > b$ in what follows.

Denote by G, H the distribution functions of $F_Z^{-1}(U_1)$ and $F_Z^{-1}(U_2)$, respectively, where $U_1 \sim \mathrm{U}[p, 1], U_2 \sim \mathrm{U}[0, p]$. Then we write $F_Z = (1-p)G + pH$ with $\mathrm{Support}(G) \subseteq [m, \mathrm{ess\text{-}sup}(F_Z)]$ and $\mathrm{Support}(H) \subseteq [\mathrm{ess\text{-}inf}(F_Z), m]$. Take $U \sim \mathrm{U}[0, 1]$ and define

$$X = Y = X_1 = \frac{F_Z^{-1}(U)}{2} \quad \text{and} \quad Y_1 = \begin{cases} \dfrac{F_Z^{-1}(p - U)}{2} & \text{if } U < p, \\[2mm] \dfrac{F_Z^{-1}(1 + p - U)}{2} & \text{if } U > p. \end{cases}$$

We can verify that (X, Y, X_1, Y_1) is p-concentrated with common p-tail event $\{U > p\}$, $X \overset{\mathrm{d}}{=} X_1$ and $Y \overset{\mathrm{d}}{=} Y_1$. Moreover, by letting $Z_1 = X_1 + Y_1$, we have

$$X + Y = F_Z^{-1}(U) \overset{\mathrm{d}}{=} Z \quad \text{and} \quad F_{Z_1} = (1-p)T(G) + pT(H).$$

Note that $\mathrm{ES}_p(Z_1) = \mathrm{ES}_p(Z)$ and $\mathbb{E}[Z_1] = \mathbb{E}[Z]$. Properties [$p$-CA] and [LI] lead to $\rho(Z) = \rho(Z_1)$. By Lemma 1, we further obtain

$$\mathrm{ess\text{-}sup}(Z_1) - F_{Z_1}^{-1}(p+) = L(T(G)) \leqslant \frac{L(G)}{2} = \frac{1}{2}(\mathrm{ess\text{-}sup}(Z) - F_Z^{-1}(p+)),$$

and

$$F_{Z_1}^{-1}(p) - \mathrm{ess\text{-}inf}(Z_1) = L(T(H)) \leqslant \frac{L(H)}{2} = \frac{1}{2}(F_Z^{-1}(p) - \mathrm{ess\text{-}inf}(Z)).$$

We repeat the above argument to construct Z_2 with Z_1 replacing the position of Z. Take any $\varepsilon \in (0, (a - b)/4)$. For large enough n (more precisely, $n \geqslant \log_2(L(F_Z)/\varepsilon)$), we have

$$\mathrm{ess\text{-}sup}(Z_n) - F_{Z_n}^{-1}(p+) < \varepsilon \quad \text{and} \quad F_{Z_n}^{-1}(p) - \mathrm{ess\text{-}inf}(Z_n) < \varepsilon.$$

Combining with $\mathrm{ES}_p(Z_n) = a$ and $\mathrm{ES}_p^-(Z_n) = b$, it then follows that

$$\mathbb{P}(a - \varepsilon < Z_n < a + \varepsilon) = 1 - p \quad \text{and} \quad \mathbb{P}(b - \varepsilon < Z_n < b + \varepsilon) = p.$$

Note that the above construction preserves the value of ρ, that is,

$$\rho(Z) = \rho(Z_1) = \rho(Z_2) = \cdots = \rho(Z_n).$$

Denote by G_n and H_n the distribution functions of $F_{Z_n}^{-1}(U_1)$ and $F_{Z_n}^{-1}(U_2)$, respectively. It follows that $F_{Z_n} = (1-p)G_n + pH_n$. Moreover, the mean of G_n is a and the mean of H_n is b, and

$$\mathrm{Support}(G_n) \subseteq (a - \varepsilon, a + \varepsilon) \quad \text{and} \quad \mathrm{Support}(H_n) \subseteq (b - \varepsilon, b + \varepsilon).$$

Note that in an atomless probability, there exists a random vector with any specified distribution (e.g., Lemma D.1 of Vovk and Wang (2021)). We take a random vector $(\mathbb{1}_A, V_1, \ldots, V_4)$ such that A is an event independent of (V_1, V_2, V_3, V_4) satisfying $\mathbb{P}(A) = 1 - p$, and $(V_1, \ldots, V_4)$, whose existence is justified by Lemma 2, satisfies

$$V_1 \stackrel{\mathrm{d}}{=} V_2, \ V_1 - V_2 + a \sim G_n \text{ and } \mathrm{Support}(V_1) = \mathrm{Support}(V_2) \subseteq [-\varepsilon, \varepsilon],$$

$$V_3 \stackrel{\mathrm{d}}{=} V_4, \ V_3 - V_4 + b \sim H_n \text{ and } \mathrm{Support}(V_3) = \mathrm{Support}(V_4) \subseteq [-\varepsilon, \varepsilon].$$

Define

$$X = \mathbb{1}_A \left(V_1 + \frac{a}{2}\right) + \mathbb{1}_{A^c} \left(V_3 + \frac{b}{2}\right), \quad Y = \mathbb{1}_A \left(-V_2 + \frac{a}{2}\right) + \mathbb{1}_{A^c} \left(-V_4 + \frac{b}{2}\right),$$

$$X^* = \mathbb{1}_A \left(V_1 + \frac{a}{2}\right) + \mathbb{1}_{A^c} \left(V_3 + \frac{b}{2}\right), \quad Y^* = \mathbb{1}_A \left(-V_1 + \frac{a}{2}\right) + \mathbb{1}_{A^c} \left(-V_3 + \frac{b}{2}\right).$$

Since $|V_1|, |V_2|, |V_3|, |V_4| \leqslant \varepsilon < (b-a)/4$, for any $\omega \in A$ and $\omega' \in A^c$, we have

$$X(\omega) = V_1(\omega) + \frac{a}{2} > \frac{a}{2} - \varepsilon > \frac{b}{2} + \varepsilon > V_3(\omega') + \frac{b}{2} = X(\omega').$$

Similarly, we have $Y(\omega) > Y(\omega')$, $Y^*(\omega) > Y^*(\omega')$ and $X^*(\omega) > X^*(\omega')$. Hence, (X, Y, X^*, Y^*) is p-concentrated with common p-tail event A, and $X = X^*$, $Y \stackrel{\mathrm{d}}{=} Y^*$, $X + Y \stackrel{\mathrm{d}}{=} Z_n$. Therefore,

$$\rho(Z) = \rho(Z_n) = \rho(X + Y) = \rho(X^* + Y^*) = \rho(\mathbb{1}_A \times a + \mathbb{1}_{A^c} \times b) = \rho(Z^*),$$

thus showing the desirable statement that the value of $\rho(X)$ only depends on $\mathrm{ES}_p(X)$ and $\mathbb{E}[X]$, that is, ρ has the form $f(\mathrm{ES}_p, \mathbb{E})$.

It remains to prove that the function $x \mapsto f(x, y)$ is increasing for each fixed $y \in \mathbb{R}$. Suppose $0 < p \leqslant 1/2$, and let $X \sim p\delta_{-(1-p)a} + (1-p)\delta_{pa}$ and $Y \sim (1-p)\delta_{-pb+y} + p\delta_{(1-p)b+y}$ with $0 \leqslant b \leqslant a$. For $U \sim \mathrm{U}(0,1)$, take $X_1 = X_2 = F_X^{-1}(U)$, $Y_1 = F_Y^{-1}(U)$ and $Y_2 = F_Y^{-1}(1-U)$. By straightforward calculation, we obtain

$$\mathbb{E}[X_1 + Y_1] = \mathbb{E}[X_2 + Y_2] = y, \quad \mathrm{ES}_p(X_1 + Y_1) = p(a-b) + \frac{pb}{1-p} + y, \quad \mathrm{ES}_p(X_2 + Y_2) = p(a-b) + y.$$

Note that (X_1, Y_1) is p-concentrated and $X_1 \overset{\mathrm{d}}{=} X_2$, $Y_1 \overset{\mathrm{d}}{=} Y_2$. Hence, by using [$p$-CA], we obtain

$$f\left(p(a-b) + \frac{pb}{1-p} + y, y\right) = \rho(X_1 + Y_1) \geqslant \rho(X_2 + Y_2) = f\left(p(a-b) + y, y\right).$$

Since $a - b \geqslant 0$ and $b \geqslant 0$ can be arbitrarily chosen, we have that $x \mapsto f(x, y)$ is increasing for each y. Using similar arguments, increasing monotonicity of $x \mapsto f(x, y)$ also holds for $1/2 < p < 1$. Hence, we complete the proof of (i).

(ii) The sufficiency statement is straightforward. To show the necessity statement, based on the result in (i), it remains to show that [M] implies the increasing monotonicity of the function $y \mapsto f(x, y)$. Take $A \in \mathcal{F}$ with probability $1 - p$. Define two random variables X and Y such that $X(\omega) = Y(\omega) = x$ for $\omega \in A$, and $X(\omega) = x_1$, $Y(\omega) = x_2$ for $\omega \in A^c$, where $x_1 \leqslant x_2 \leqslant x$. Obviously, we have $X \leqslant Y$, and it follows that

$$\begin{aligned} f(x, (1-p)x + px_1) &= f(\mathrm{ES}_p(X), \mathbb{E}[X]) \\ &= \rho(X) \leqslant \rho(Y) = f(\mathrm{ES}_p(Y), \mathbb{E}[Y]) = f(x, (1-p)x + px_2). \end{aligned}$$

Hence, we obtain the increasing monotonicity of $y \mapsto f(x, y)$ since $x_1 \leqslant x_2 \leqslant x$ can be arbitrarily chosen. $\qquad\square$

Remark 2. The functional ES_p^- is used in the proof of Theorem 1, but not in its statement. There is a linear relationship between ES_p, ES_p^- and $\mathbb{E}$, that is,

$$p\mathrm{ES}_p^-(X) + (1-p)\mathrm{ES}_p(X) = \mathbb{E}[X].$$

Therefore, the form $f(\mathrm{ES}_p, \mathbb{E})$ of the risk measure in Theorem 1 can also be represented as $f_1(\mathrm{ES}_p^-, \mathbb{E})$ or $f_2(\mathrm{ES}_p, \mathrm{ES}_p^-)$ with different conditions on f_1 and f_2.

3.3 Mean-ES portfolio selection

There is a large literature on mean-risk portfolio selection since Markowitz (1952) who measured risk by using variance. In the more recent literature, risk is often measured by a risk measure, such as VaR (Basak and Shapiro, 2001; Gaivoronski and Pflug, 2005), ES (Rockafellar and Uryasev, 2000, 2002; Embrechts et al., 2022), or expectiles (Bellini et al., 2014; Lin et al., 2021). For a recent work on mean-ρ optimization where ρ is a coherent risk measure, see Herdegen and Khan (2022).

Remarkably, Theorem 1 gives rise to an axiomatic foundation for the mean-ES portfolio selection. Consider a classical optimization problem

$$\min_{\mathbf{a} \in A} \mathcal{V}(g(\mathbf{X}, \mathbf{a})) \tag{1}$$

where A is a set of possible actions, $\mathcal{V} : \mathcal{X} \to (-\infty, \infty]$ is an objective functional, $\mathbf{X}$ is the underlying d-dimensional risk vector, and $g : \mathbb{R}^d \times A \to \mathbb{R}$ is a function representing the portfolio value. Constraints on the optimization problem can be incorporated into either A or $\mathcal{V}$. For instance, one may set $\mathcal{V}$ to be ∞ for positions that violate certain constraints, as we will see below.

We say that the optimization problem (1) is *a mean-ρ optimization* for some risk measure ρ, if $\mathcal{V}$ is determined by $\mathbb{E}$ and ρ and increasing in both. There are two classic versions of mean-ρ optimization problems:

(a) Maximizing expected return with a target risk $r \in \mathbb{R}$, that is

$$\max_{\mathbf{a} \in A} \mathbb{E}[-\mathbf{a}^{\mathrm{T}}\mathbf{X}] \quad \text{subject to } \rho(\mathbf{a}^{\mathrm{T}}\mathbf{X}) \leqslant r, \tag{2}$$

where $\mathbf{X}$ is the vector of losses (negative returns) from individual assets and A is a subset of $\mathbb{R}^d$; recall that $-\mathbf{a}^{\mathrm{T}}\mathbf{X}$ represents the future portfolio wealth. By choosing

$$\mathcal{V}(X) = \mathbb{E}[X]\mathbb{1}_{\{\rho(X) \leqslant r\}} + \infty\mathbb{1}_{\{\rho(X) > r\}} \quad \text{and} \quad g(\mathbf{X}, \mathbf{a}) = \mathbf{a}^{\mathrm{T}}\mathbf{X}$$

with the convention $\infty \times 0 = 0$, (2) becomes (1), which is clearly a mean-ρ optimization.

(b) Minimizing risk with a target expected return $u \in \mathbb{R}$, that is,

$$\min_{\mathbf{a} \in A} \rho(\mathbf{a}^{\mathrm{T}}\mathbf{X}) \quad \text{subject to } \mathbb{E}[-\mathbf{a}^{\mathrm{T}}\mathbf{X}] \geqslant u. \tag{3}$$

This time, by choosing

$$\mathcal{V}(X) = \rho(X)\mathbb{1}_{\{\mathbb{E}[X] \leqslant -u\}} + \infty\mathbb{1}_{\{\mathbb{E}[X] > -u\}} \text{ and } g(\mathbf{X}, \mathbf{a}) = \mathbf{a}^{\mathrm{T}}\mathbf{X},$$

we arrive again at (1).

Using Theorem 1, we obtain a characterization of mean-ES (i.e., mean-ES$_p$ for some $p \in (0, 1)$) optimization problems, which include the classical problems (2) and (3) with $\rho = \mathrm{ES}_p$.

Theorem 2. *An optimization problem* (1) *is a mean-ES optimization if and only if its objective* $\mathcal{V}$ *satisfies* [M] *and* [p-CA] *for some* $p \in (0, 1)$.

Theorem 2 illustrates that a preference for dependence (i.e., [p-CA]) can help to pin down the particular form of optimization problems, in addition to characterizing risk measures. In the next section, we continue to explore the relationship between concentration aversion and characterizing risk measures.

4 Monetary risk measures satisfying [CA]

In this section, we again assume that ρ is a risk measure on $\mathcal{X} = L^\infty$, and further investigate *monetary risk measures* satisfying [p-CA]. A *monetary risk measure* is a risk measure satisfying [M] and [TI]; see Föllmer and Schied (2016). It is well known that monetary risk measures are one-to-one corresponding to acceptance sets. An *acceptance set* $\mathcal{A}$ is a subset of $\mathcal{X}$ which is generated by some monetary risk measure ρ via $\mathcal{A} = \{X \in \mathcal{X} : \rho(X) \leqslant 0\}$. Also note that a monetary risk measure ρ is finite on L^∞ as long as it is finite at some $X \in L^\infty$. Therefore, we can safely assume $\rho : L^\infty \to \mathbb{R}$ in this section.

4.1 Concentration-averse monetary risk measures

Let us first recall the definition of *second-order stochastic dominance* (SSD). We say that X is second-order stochastically dominated by Y, denoted by $X \preceq_{\mathrm{SSD}} Y$, if $\mathbb{E}[u(X)] \leqslant \mathbb{E}[u(Y)]$ for all increasing convex functions u.[6] We collect two properties from Mao and Wang (2020).

[SC] SSD-consistency: $\rho(X) \leqslant \rho(Y)$ whenever $X \preceq_{\mathrm{SSD}} Y$.

[DC] Diversification consistency: $\rho(X + Y) \leqslant \rho(X^c + Y^c)$ whenever $X \overset{\mathrm{d}}{=} X^c$, $Y \overset{\mathrm{d}}{=} Y^c$ and (X^c, Y^c) is comonotonic.

---

[6]SSD is also known as increasing convex order in probability theory and stop-loss order in actuarial science.

The property [SC] is often called *strong risk aversion* for a preference functional (Hadar and Russell, 1969; Rothschild and Stiglitz, 1970), while [DC] is called *comovement aversion* (Wang and Wu, 2020). By using a risk measure satisfying [SC], a financial institution makes decisions that are consistent with the common notion of risk aversion and, in particular, favours a risk with small variability over one with a large variability. Mao and Wang (2020) showed that, for a monetary risk measure, [SC] and [DC] are equivalent, and they called monetary risk measures satisfying [SC] *consistent risk measure*, which have a representation based on ES; see their Theorem 3.1. Since p-concentration is weaker than comonotonicity, [p-CA] implies [DC], and hence a monetary risk measure satisfying [p-CA] is automatically a consistent risk measure. In the following theorem, a representation of such a risk measure is established. This result leads to a class of risk measures (5) that is new to the literature. In what follows, we say that a real-valued function g satisfies the 1-Lipschitz condition if

$$|g(x) - g(y)| \leqslant |x - y| \quad \text{for } x, y \text{ in the domain of } g. \tag{4}$$

Theorem 3. *Let $p \in (0, 1)$ and ρ be a risk measure on L^∞. Then, ρ satisfies [M], [TI], [p-CA] and $\rho(0) = 0$ if and only if it has the form*

$$\rho(X) = g(\mathrm{ES}_p(X) - \mathbb{E}[X]) + \mathbb{E}[X], \quad X \in L^\infty, \tag{5}$$

for some increasing function $g : [0, \infty) \to \mathbb{R}$ with $g(0) = 0$ satisfying the 1-Lipschitz condition. In particular, such ρ is a consistent risk measure.

Proof. Let us first prove sufficiency. Obviously, ρ of the form (5) satisfies [TI] and $\rho(0) = 0$. By Theorem 1, we obtain that ρ satisfies [p-CA]. So it remains to verify that ρ is monotone. Suppose $X \leqslant Y$, and define $a_1 = \mathrm{ES}_p(X)$, $b_1 = \mathbb{E}[X]$, $a_2 = \mathrm{ES}_p(Y)$ and $b_2 = \mathbb{E}[Y]$. Obviously, we have $a_1 \leqslant a_2$, $b_1 \leqslant b_2$ and

$$\rho(X) = g(a_1 - b_1) + b_1, \quad \rho(Y) = g(a_2 - b_2) + b_2.$$

If $a_2 - a_1 \geqslant b_2 - b_1$, we have

$$\begin{aligned}
\rho(X) &\leqslant \rho(X) + (b_2 - b_1) \\
&= g(a_1 - b_1) + b_2 \\
&= g((a_1 + b_2 - b_1) - b_2) + b_2 \leqslant g(a_2 - b_2) + b_2 = \rho(Y),
\end{aligned}$$

16

where the second inequality follows from the increasing monotonicity of g. If $a_2 - a_1 < b_2 - b_1$, we have

$$\rho(X) \leqslant \rho(X) + (a_2 - a_1)$$
$$= g(a_2 - (b_1 + a_2 - a_1)) + (b_1 + a_2 - a_1) \leqslant \rho(a_2 - b_2) + b_2 = \rho(Y),$$

where the second inequality follows from the 1-Lipschitz condition of g. Hence, we complete the proof of sufficiency. For the other direction, it follows from the results in Theorem 1 that ρ has the form $f(\mathrm{ES}_p, \mathbb{E})$ for some bivariate function f. Define a function $g : [0, \infty) \to \mathbb{R}$ such that $g(x) = f(x, 0)$ for $x \geqslant 0$. It is clear that $g(0) = f(0, 0) = \rho(0) = 0$. Note that $f(\cdot, y) : [y, \infty) \to \mathbb{R}$ is increasing for all $y \in \mathbb{R}$ (see Theorem 1). It follows that g is increasing. Using [TI], we obtain

$$\rho(X) = \rho(X - \mathbb{E}[X]) + \mathbb{E}[X]$$
$$= f(\mathrm{ES}_p(X) - \mathbb{E}[X], 0) + \mathbb{E}[X] = g(\mathrm{ES}_p(X) - \mathbb{E}[X]) + \mathbb{E}[X].$$

Finally, applying Theorem 1 (ii), we know that the function $f(x, \cdot) : (-\infty, x] \to \mathbb{R}$ is increasing for all $x \in \mathbb{R}$. Hence, we have

$$g(x - y) + y = f(x, y) \leqslant f(x, y') = g(x - y') + y' \quad \text{for all } y < y' \leqslant x,$$

which implies that g is 1-Lipschitz. Hence, we complete the proof. $\qquad \square$

The risk measure ρ with form (5) is the sum of the mean and $g(\mathrm{ES}_p - \mathbb{E})$. Note that $\mathrm{ES}_p - \mathbb{E}$ is both a generalized deviation measure according to Rockafellar et al. (2006) and a coherent measure of variability according to Furman et al. (2017). Hence, $g(\mathrm{ES}_p - \mathbb{E})$ is a transformed deviation or variability measure, and a monetary risk measure satisfying [p-CA] can be seen as a mean-deviation functional.

We continue to characterize the classes of convex,[7] coherent,[8] and consistent risk measures that satisfy [p-CA]. These three classes of risk measures are all monetary risk measure, and thus they can be represented as the form in Theorem 3. Note that for ρ satisfying [p-CA], there is a one-to-one correspondence between ρ and g in (5), and hence the above classes can be identified based on properties of g. The gap between convex risk measure and consistent risk measure is established

---

[7]A convex risk measure is a monetary risk measure which also satisfies *convexity*: $\rho(\lambda X + (1 - \lambda)Y) \leqslant \lambda\rho(X) + (1 - \lambda)\rho(Y)$ for all $\lambda \in [0, 1]$.

[8]A coherent risk measure is a convex risk measure which also satisfies *positive homogeneity*: $\rho(\lambda X) = \lambda\rho(X)$ for all $\lambda \in (0, \infty)$ and $X \in \mathcal{X}$.

clearly in the following proposition. In particular, convexity of g is equivalent to convexity of ρ.

Proposition 2. *Let $p \in (0,1)$ and ρ be a risk measure on L^∞ satisfying [p-CA] and $\rho(0) = 0$.*

(i) *ρ is a consistent risk measure if and only if $\rho = g(\mathrm{ES}_p - \mathbb{E}) + \mathbb{E}$ for some increasing and 1-Lipschitz function $g : [0, \infty) \to \mathbb{R}$ with $g(0) = 0$.*

(ii) *ρ is a convex risk measure if and only if $\rho = g(\mathrm{ES}_p - \mathbb{E}) + \mathbb{E}$ for some increasing, convex and 1-Lipschitz function $g : [0, \infty) \to \mathbb{R}$ with $g(0) = 0$.*

(iii) *ρ is a coherent risk measure if and only if $\rho = \alpha \mathrm{ES}_p + (1 - \alpha)\mathbb{E}$ for some $\alpha \in [0, 1]$.*

Proof. (i) is implied by Theorem 3. To see (ii), applying Theorem 3, it is sufficient to prove that convexity of the function g in (5) is equivalent to convexity of ρ. Note that g is an increasing function. If g is convex, then ρ is a convex risk measure because expectation is linear and ES_p is a convex risk measure. If g is nonconvex, then there exist $0 \leqslant x < y$ and $\lambda \in (0,1)$ such that $g(\lambda x + (1 - \lambda)y) > \lambda g(x) + (1 - \lambda)g(y)$. Suppose (X, Y) is p-concentrated, and satisfies $\mathrm{ES}_p(X) - \mathbb{E}[X] = x$ and $\mathrm{ES}_p(Y) - \mathbb{E}[Y] = y$. Thus, we have

$$
\begin{aligned}
\rho(\lambda X + (1 - \lambda)Y) &= g(\mathrm{ES}_p(\lambda X + (1 - \lambda)Y) - \mathbb{E}[\lambda X + (1 - \lambda)Y]) + \mathbb{E}[\lambda X + (1 - \lambda)Y] \\
&= g(\lambda(\mathrm{ES}_p(X) - \mathbb{E}[X]) + (1 - \lambda)(\mathrm{ES}_p(Y) - \mathbb{E}[Y])) + \lambda\mathbb{E}[X] + (1 - \lambda)\mathbb{E}[Y] \\
&= g(\lambda x + (1 - \lambda)y) + \lambda\mathbb{E}[X] + (1 - \lambda)\mathbb{E}[Y] \\
&> \lambda g(x) + (1 - \lambda)g(y) + \lambda\mathbb{E}[X] + (1 - \lambda)\mathbb{E}[Y] \\
&= \lambda(g(\mathrm{ES}_p(X) - \mathbb{E}[X]) + \mathbb{E}[X]) + (1 - \lambda)(g(\mathrm{ES}_p(Y) - \mathbb{E}[Y]) + \mathbb{E}[Y]) \\
&= \lambda\rho(X) + (1 - \lambda)\rho(Y),
\end{aligned}
$$

which implies ρ is nonconvex. (iii) Sufficiency is straightforward. To show necessity, let X be such that $\mathbb{E}[X] = 0$ and $\mathrm{ES}_p(X) = x > 0$. By Theorem 3, coherence of ρ implies that for all $\lambda > 0$,

$$
g(\lambda x) = \rho(\lambda X) = \lambda\rho(X) = \lambda g(x).
$$

This means that g is linear on $(0, \infty)$. Noting that g is 1-Lipschitz with $g(0) = 0$, we have $g(x) = \alpha x$ for some $\alpha \in [0, 1]$. Hence, we complete the proof of (iii). $\qquad\square$

Since SSD-consistency is strictly weaker than convexity for a law-invariant risk measure, the class of consistent risk measures generalizes that of law-invariant convex risk measures. However, explicit formulas for nonconvex consistent risk measures are rare in the literature; indeed, all examples in Mao and Wang (2020) involve taking an infimum over convex risk measures. Proposition

2 leads to many examples of consistent risk measures with explicit formulas which are outside the classic framework of convex risk measures.

Example 1. We give two examples of new risk measures from Proposition 2. First, take the convex and 1-Lipschitz function $g : x \mapsto x - \log(1 + x)$. The corresponding risk measure is given by $\rho = \mathrm{ES}_p - \log(1 + \mathrm{ES}_p - \mathbb{E})$, which is a convex risk measure by Proposition 2. Second, take the concave and 1-Lipschitz function $g : x \mapsto \log(1 + x)$. The corresponding risk measure is given by $\rho = \mathbb{E} + \log(1 + \mathrm{ES}_p - \mathbb{E})$, which is a consistent risk measure but not convex by Proposition 2.

4.2 A new characterization of the Expected Shortfall

Next, we add lower semicontinuity [P] to the requirements in Theorem 3 and obtain a new characterization of ES. Remarkably, although Theorem 3 allows for many choices of risk measures satisfying [p-CA], lower semicontinuity is enough to force the function g in (5) to collapse to the identity. Hence, for this characterization of ES, we do not need to assume coherence or convexity.

Theorem 4. *Let $p \in (0, 1)$ and ρ be a risk measure on L^∞. Then ρ satisfies* [M], [TI], [P], [p-CA] *and $\rho(0) = 0$ if and only if it is* ES_p.

Proof. Sufficiency follows from Proposition 1 and Theorem 5 of Wang and Zitikis (2021). To see necessity, we first apply the result in Theorem 1 that ρ has the form $f(\mathrm{ES}_p, \mathbb{E})$, and the function $y \mapsto f(x, y)$ is increasing on $(-\infty, x]$ for all $x \in \mathbb{R}$. Next, we will verify that the value of f is independent of its second argument. On the one hand, we have $f(x, x) \geqslant f(x, y)$ for all $x \geqslant y$. On the other hand, define a sequence of random variables $\{X_n\}_{n \in \mathbb{N}}$ such that $\mathbb{P}(X_n = x) = 1 - 1/n$, $\mathbb{P}(X_n = x - n(x - y)) = 1/n$ and $X_n \to x$ a.s.. By the property [P], we have

$$f(x, y) = \liminf_{n \to \infty} f(\mathrm{ES}_p(X_n), \mathbb{E}[X_n]) \geqslant f(x, x).$$

Therefore, we conclude that $f(x, x) = f(x, y)$ for all $x \geqslant y$, and this means $\rho(X) = g(\mathrm{ES}_p(X))$ for some function g. Finally, using [TI] and $\rho(0) = 0$, one can conclude that g is the identity. $\square$

We can equivalently express Theorem 4 in terms of the acceptance set as in the next proposition. A proof is straightforward from the definition of an acceptance set.

Proposition 3. *Let $p \in (0, 1)$. An acceptance set $\mathcal{A}$ satisfies*

(i) (X, Y) is p-concentrated and $X + Y \in \mathcal{A} \implies X' + Y' \in \mathcal{A}$ for all X', Y' with $X' \overset{\mathrm{d}}{=} X$, $Y' \overset{\mathrm{d}}{=} Y$,

(ii) $X_n \in \mathcal{A}$ for each $n = 1, 2, \ldots$ and $X_n \to X$ pointwise $\implies X \in \mathcal{A}$, and

(iii) $\sup\{c \in \mathbb{R} : c \in \mathcal{A}\} = 0$,

if and only if $\mathcal{A}$ is the acceptance set of ES_p.

5 Generalization to larger spaces

In this section, we generalize the characterization results in Sections 3 and 4 to larger L^q spaces than L^∞. The risk measure $\rho : L^q \to \mathbb{R}$ will be assumed to take real values.

5.1 Generalization to L^q for $q \geqslant 1$

We endow the natural norm on L^q, $q \in [1, \infty)$, i.e., $\|X\|_q = (\mathbb{E}[|X|^q])^{1/q}$ for $X \in L^q$, and continuity is defined with respect to $\|\cdot\|_q$. Furthermore, we recall the notation $\mathbb{H}$ as the half-space $\{(x, y) \in \mathbb{R}^2 : x \geqslant y\}$.

Proposition 4. *Let $p \in (0, 1)$, $q \geqslant 1$ and $\rho : L^q \to \mathbb{R}$ be a continuous risk measure. Then,*

(i) ρ satisfies [p-CA] if and only if it has the form $f(\mathrm{ES}_p, \mathbb{E})$, where $f : \mathbb{H} \to \mathbb{R}$ is a continuous bivariate function which is increasing in its first argument.

(ii) ρ satisfies [M] and [p-CA] if and only if it has the form $f(\mathrm{ES}_p, \mathbb{E})$, where $f : \mathbb{H} \to \mathbb{R}$ is a continuous bivariate function which is increasing in both arguments.

Proof. Sufficiency in both (i) and (ii) is trivial. To see necessity, noting that for any $X \in L^q$, there exists a sequence $\{X_n\}_{n \in \mathbb{N}} \subseteq L^\infty$ converges to X with respect to the norm $\|\cdot\|_q$. By the continuity of ρ, the statements in Theorem 1 are all valid. Thus, it remains to prove that f is continuous on $\mathbb{H}$. For $(x_0, y_0) \in \mathbb{H}$, let $\{(x_n, y_n)\}_{n \in \mathbb{N}} \subseteq \mathbb{H}$ be a sequence converges to (x_0, y_0). Let $A \in \mathcal{F}$ such that $\mathbb{P}(A) = p$. Define a sequence of random variables

$$X_n(\omega) = \frac{y_n - (1 - p)x_n}{p} \text{ for } \omega \in A, \quad \text{and} \quad X_n(\omega) = x_n \text{ for } \omega \in A^c,$$

and let

$$X(\omega) = \frac{y_0 - (1 - p)x_0}{p} \text{ for } \omega \in A, \quad \text{and} \quad X(\omega) = x_0 \text{ for } \omega \in A^c.$$

Obviously, $\mathrm{ES}_p(X_n) = x_n$, $\mathbb{E}[X_n] = y_n$ and $X_n \to X$ in L^q with $\mathrm{ES}_p(X) = x_0$, $\mathbb{E}[X] = y_0$. Hence, we have

$$f(x_n, y_n) = f(\mathrm{ES}_p(X_n), \mathbb{E}[X_n]) = \rho(X_n) \to \rho(X) = f(\mathrm{ES}_p(X), \mathbb{E}[X]) = f(x_0, y_0).$$

This completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Similarly, Theorems 3 and 4 can be generalized to L^q for $q \geqslant 1$.

Proposition 5. *Let $p \in (0,1)$, $q \geqslant 1$ and $\rho : L^q \to \mathbb{R}$ be a continuous risk measure. Then ρ satisfies [M], [TI], [p-CA] and $\rho(0) = 0$ if and only if it has the form $\rho = g(\mathrm{ES}_p - \mathbb{E}) + \mathbb{E}$ for some increasing and 1-Lipschitz function $g : [0, \infty) \to \mathbb{R}$ with $g(0) = 0$. In particular, such ρ is a consistent risk measure.*

Proposition 6. *Let $p \in (0,1)$, $q \geqslant 1$ and $\rho : L^q \to \mathbb{R}$ be a continuous risk measure. Then ρ satisfies [M], [TI], [P], [p-CA] and $\rho(0) = 0$ if and only if it is ES_p.*

5.2 Impossibility results on L^q for $q \in [0,1)$

In this section, we let $q \in [0,1)$ and consider the larger spaces $L^q \supset L^1$ as the domain of the risk measure ρ. It is shown in Theorem 2 of Wang and Zitikis (2021) that the only real-valued risk measure on L^q satisfying [M], [LI], [P] and [NRC] is the constant risk measure $\rho = 0$. A natural question arises: Is there a nonconstant risk measure $\rho : L^q \to \mathbb{R}$ satisfying [p-CA]? We shall first see in the following example that [p-CA] on L^q does not necessarily lead to a constant risk measure.

Example 2. Let $f(x, y)$ be a bounded real function on $\mathbb{H} = \{(x, y) \in \mathbb{R}^2 : x \geqslant y\}$ which is increasing in both x, y and $M > 0$ be such that $|f| \leqslant M$. Define

$$
\rho(X) = \begin{cases} f(\mathrm{ES}_p(X), \mathbb{E}[X]), & X \in L^1, \\ -M, & \mathbb{E}[X_-] = \infty, \ \mathbb{E}[X_+] < \infty, \\ M, & \mathbb{E}[X_+] = \infty, \end{cases}
$$

where $X_+ = \max\{X, 0\}$ and $X_- = \max\{-X, 0\}$. One can verify that ρ satisfies [M] and [p-CA].

As illustrated by Example 2, in contrast to [NRC], we can construct a class of nontrivial risk measures bounded on L^0 that satisfies [p-CA]. Nevertheless, the following proposition illustrates that it is pointless to consider monotone risk measures $\rho : L^q \to \mathbb{R}$ satisfying [p-CA] if ρ is unbounded on the set of constants. As a consequence, we conclude that the domain L^1 is the most natural, and essentially the largest, choice for any real-valued risk measures satisfying [M], [TI] and [p-CA].

Proposition 7. *Let $p \in (0,1)$ and $q \in [0,1)$. There is no such $\rho : L^q \to \mathbb{R}$ that satisfies [M], [p-CA] and $\lim_{c \to \infty} \rho(c) = \infty$.*

Proof. Assume that such ρ exists. Take a nonnegative $X \in L^q \setminus L^1$, and let $X_n = \min\{X, n\} \in L^\infty$ for $n \in \mathbb{N}$. Obviously, we have $X_n \uparrow X$. By Theorem 1, ρ has the form $f(\mathrm{ES}_p, \mathbb{E})$ on L^∞. It

then follows from [M] and the condition $\lim_{c\to\infty} \rho(c) = \infty$ that $\lim_{y\to\infty} f(x,y) = \infty$. Note that $\mathbb{E}[X_n] \to \infty$. Thus, we obtain

$$\rho(X) \geqslant \liminf \rho(X_n) = \liminf f(\mathrm{ES}_p(X_n), \mathbb{E}[X_n]) = \infty,$$

a contradiction. □

Since a monetary risk measure ρ necessarily satisfies $\lim_{c\to\infty} \rho(c) = \infty$, we conclude from Proposition 7 that for $q \in [0, 1)$, there is no monetary risk measure $\rho : L^q \to \mathbb{R}$ that satisfies [p-CA].

6   An economic reasoning for concentration aversion

For the key concept of concentration aversion in this paper, it is assumed in Definition 2 that there exists a tail event A of regulatory concern. Such a tail event A is exogenous to the property [CA]; similarly, the structure of p-concentration is exogenous to the property [p-CA]. For a solid economic foundation of using [CA], it would be more compelling to justify the structure of p-concentration from endogenous reasoning.[9] Addressing this issue is the objective of this section. We will show that, if a regulator is concerned about dangerous dependence structures satisfying a few axioms, then [CA] must hold for the regulator's risk measure.

Assume $\mathcal{X} = L^\infty$ in this section, and denote by $L_c^\infty \subseteq L^\infty$ the set of all continuously distributed random variables in L^∞. We focus on continuous distributions because we will work with dependence structures, which will be modelled by copulas. An n-copula is a joint distribution function on $\mathbb{R}^n$ with standard uniform marginals. Sklar's theorem implies that the joint distribution F of any random vector $\mathbf{X}$ can be expressed by a copula C of $\mathbf{X}$ through $F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n))$ where $F_1, \ldots, F_n$ are the marginals of F. The copula C is unique if $F_1, \ldots, F_n$ are continuous. We denote by $C_{\mathbf{X}}$ the copula of $\mathbf{X}$ if it is unique, and $\mathcal{C}_n$ the set of n-copulas. We refer to Joe (2014) for a general treatment of copulas.

Suppose that a regulator is concerned about random losses that are dependent in an adverse (dangerous) way. The interpretation of dangerousness of a dependence structure is modelled by a set $\mathcal{D} \subseteq \mathcal{C}_2$. We will specify a suitable $\mathcal{D}$ later, but a primary example is

$$\mathcal{D}_p = \{C \in \mathcal{C}_2 : C(p, p) = p\}, \quad p \in (0, 1). \tag{6}$$

By Theorem 3 of Wang and Zitikis (2021), a copula of (X, Y) is in $\mathcal{D}_p$ if and only if (X, Y) is

---

[9]We thank an anonymous referee for bringing this question up.

p-concentrated; hence, $\mathcal{D}_p$ is the set of bivariate copulas for p-concentrated random vectors. Since an adverse dependence structure bears more risk, the regulatory risk measure $\rho : L^\infty \to \mathbb{R}$ should satisfy $\mathcal{D}$-*aversion*, that is, $\rho(X+Y) \leqslant \rho(Z+W)$ for all $X, Y, Z, W \in L^\infty$ satisfying $X \stackrel{\mathrm{d}}{=} Z$, $Y \stackrel{\mathrm{d}}{=} W$, and a copula of (Z, W) is in $\mathcal{D}$. For the special case of $\mathcal{D} = \mathcal{D}_p$ for some $p \in (0, 1)$, $\mathcal{D}_p$-aversion is precisely [p-CA].[10]

In what follows, we discuss reasonable choices of $\mathcal{D}$ for the regulator. A common idea of diversification originates from the Law of Large Numbers (LLN), or its refined versions, the Central Limit Theorems. A dependence structure of risks is arguably quite dangerous if there is no effect of LLN; that is, the average risk does not vanish even if the number of risks in the pool tends to infinity. Inspired by this observation, we define non-diversifiability via violation of LLN. For a copula $C \in \mathcal{C}_2$, we say that a sequence $(X_n)_{n \in \mathbb{N}}$ is *sequentially C-coupled* if $X_n \stackrel{\mathrm{d}}{=} X_{n+1}$ and C is the copula of (X_n, X_{n+1}) for each $n \in \mathbb{N}$. Note that the dependence of (X_n, X_k) for $|n - k| > 1$ is unspecified and it typically has some flexibility. We say that $C \in \mathcal{C}_2$ is *non-diversifiable* if each sequentially C-coupled sequence $(X_n)_{n \in \mathbb{N}}$ in L_c^∞ breaks LLN, that is,

$$\frac{1}{n} \sum_{i=1}^n X_i - \mu \stackrel{\mathrm{p}}{\not\to} 0 \quad \text{as } n \to \infty, \tag{7}$$

where μ is the mean of X_1; otherwise C is *diversifiable*. A simple example of a non-diversifiable copula is the comonotonic copula C^+, defined via $C^+(u, v) = \min(u, v)$. For any sequentially C^+-coupled sequence $(X_n)_{n \in \mathbb{N}}$ in L_c^∞, due to comotonicity, we have $X_n = X_{n+1}$ for $n \in \mathbb{N}$, and

$$\frac{1}{n} \sum_{i=1}^n X_i - \mu = X_1 - \mu \stackrel{\mathrm{p}}{\not\to} 0 \quad \text{as } n \to \infty. \tag{8}$$

On the other hand, the independent copula $C^\perp$, defined via $C^\perp(u, v) = uv$, is clearly diversifiable due to LLN.

Another important consideration is that positive dependence is more dangerous than negative dependence. Recall that for two bivariate copulas C and C', the point-wise order $C' \geqslant C$, called the concordance order (see e.g., Müller and Stoyan (2002)), compares the level of positive dependence. In particular, if $X \stackrel{\mathrm{d}}{=} Z$ and $Y \stackrel{\mathrm{d}}{=} W$ and $C_{X,Y} \geqslant C_{Z,W}$, then $Z + W \preceq_{\mathrm{SSD}} X + Y$ (see e.g., Wang and Wu (2020)), and thus $X + Y$ bears more risk than $Z + W$ in a commonly agreed sense of riskiness. The comonotonic copula C^+ attains the maximum in concordance order.

---

[10]If we insist using copulas for continuously distributed random variables, we may alternatively require $\rho(X+Y) \leqslant \rho(Z + W)$ to hold only for $X, Y, Z, W \in L_c^\infty$ and (Z, W) with a unique copula in $\mathcal{D}_p$. This property is slightly weaker than [p-CA], but they are equivalent if ρ is monotone and lower semicontinuous with respect to a.s. convergence.

Finally, a combination of dangerous scenarios, in the form of a probability mixture, is still dangerous, because such a mixture represents randomly picking a dangerous scenario.

Translating the above considerations into properties of $\mathcal{D}$, we define a *bivariate concentration class* which is a subset $\mathcal{D}$ of $\mathcal{C}_2$ satisfying the following three properties [ND], [DM] and [Cx].

[ND] Non-diversifiability: Each $C \in \mathcal{D}$ is non-diversifiable.

[DM] Dependence monotonicity: If $C \in \mathcal{D}$ and $C \leqslant C' \in \mathcal{C}_2$, then $C' \in \mathcal{D}$.

[Cx] Convexity: If $C_n \in \mathcal{D}$ for $n \in \mathbb{N}$, then $\sum_{n \in \mathbb{N}} \lambda_n C_n \in \mathcal{D}$ for any non-negative numbers λ_n, $n \in \mathbb{N}$ with $\sum_{n \in \mathbb{N}} \lambda_n = 1$.

The first property, [ND], simply means that each dependence structure in $\mathcal{D}$ breaks LLN. The second property, [DM], says that if C is considered dangerous and C' is more positively dependent than C, then C' is also considered as dangerous. Convexity [Cx] means combining dangerous scenarios leads to a dangerous scenario. The three properties are arguably quite natural for a concept of concentration of interest to a regulator.

We first verify a few important examples of bivariate concentration classes.

Proposition 8. *The sets $\mathcal{D}_p$ for $p \in (0,1)$ and the singleton $\{C^+\}$ are bivariate concentration classes.*

Proof. We first verify the statement for $\{C^+\}$. A singleton is obviously convex, and thus [Cx] holds. By (8), $\{C^+\}$ satisfies [ND]. Since C^+ is the maximum in concordance order, $\{C^+\}$ satisfies [DM]. Next, we show that $\mathcal{D}_p$ for $p \in (0,1)$ satisfies [Cx], [ND] and [DM]. The property [Cx] follows directly from (6). Note that $C'(p,p) \leqslant C^+(p,p) = p$ for all $C' \in \mathcal{C}_2$. Using (6), $C' \geqslant C \in \mathcal{D}_p$ implies $C'(p,p) = C(p,p) = p$, and thus $C' \in \mathcal{D}_p$. This shows that $\mathcal{D}_p$ satisfies [DM]. To show [ND], take $C \in \mathcal{D}_p$ and construct any sequentially C-coupled sequence $(X_n)_{n \in \mathbb{N}}$. Since (X_n, X_{n+1}) is p-concentrated, by Corollary A.1 of Wang and Zitikis (2021), X_n and X_{n+1} share the same a.s. unique p-tail event $A = \{X_n > x_p\} = \{X_{n+1} > x_p\}$ where $x_p = \mathrm{VaR}_p(X_1)$. Applying this argument to $n \in \mathbb{N}$, we know that $X_1, X_2, \ldots$ share the same tail event A which does not depend on n. Write $\widetilde{X}_n = n^{-1} \sum_{i=1}^n X_i$. We note that $\widetilde{X}_n \mathbb{1}_A$ does not converge to $\mathbb{E}[X_1] \mathbb{1}_A$ since

$$\mathbb{E}\left[\widetilde{X}_n \mid A\right] = \mathbb{E}[X_1 \mid A] = \mathrm{ES}_p(X_1) > \mathbb{E}[X_1]$$

and $(\widetilde{X}_n)_{n \in \mathbb{N}}$ is uniformly integrable. Therefore,

$$\widetilde{X}_n = \widetilde{X}_n \mathbb{1}_A + \widetilde{X}_n \mathbb{1}_{A^c} \overset{\mathrm{p}}{\nrightarrow} \mathbb{E}[X_1],$$

24

and thus $(X_n)_{n\in\mathbb{N}}$ is non-diversifiable. This shows that $\mathcal{D}_p$ satisfies [ND]. Therefore, $\mathcal{D}_p$ is a bivariate concentration class. $\square$

If both $\mathcal{D}$ and $\mathcal{D}'$ are bivariate concentration classes, then so is $\mathcal{D}\cap\mathcal{D}'$. Using this relation, we can construct bivariate concentration classes other than the ones in Proposition 8. The next result, which is the most technical result in this section, shows that a bivariate concentration class may not contain anything more than those in $\mathcal{D}_p$ for some $p\in(0,1)$.

Proposition 9. *A copula $C \in \mathcal{C}_2$ is in a bivariate concentration class if and only if $C \in \mathcal{D}_p$ for some $p \in (0,1)$.*

Proof. The "if" statement follows directly from Proposition 8. Below we will show the "only if" statement. We first present some technical preparations. For any copula C and $p \in (0,1)$, denote by $t_C(p)$ the essential infimum of the distribution of V given $U \geqslant p$ where $(U,V) \sim C$, and by $s_C(p)$ the essential supremum of the distribution of V given $U \leqslant p$. Since $\mathbb{P}(U \geqslant p) = 1 - p$, the essential infimum of the distribution of V given $U \geqslant p$ is at most p, which is the case when $\{U \geqslant p\}$ is a p-tail event of V. Therefore, $t_C(p) \leqslant p$ and similarly, $s_C(p) \geqslant p$ for $p \in (0,1)$. Moreover, both t_C and s_C are increasing curves on $(0,1)$. See Figure 1a for an illustration, where the grey area B_C is the area between the curves t_C and s_C.

To proceed, we need the following lemma, which may be of interest in dependence theory by its own right. Below, by saying that a copula C has positive density on a region, we meant that the Lebesgue measure is absolutely continuous with respect to the measure associated with C on this region. Such C has a positive absolutely continuous part on the region, and may also include a non-absolutely continuous part.

Lemma 3. *For any $C \in \mathcal{C}_2$, there exists a copula $\widetilde{C} \geqslant C$ such that $\widetilde{C}$ has positive density on the region $B_C := \{(u,v) \in [0,1]^2 : t_C(u) < v < s_C(u)\}$.*

The proof of Lemma 3 requires some delicate constructions of copulas, and it is put in Appendix A. Here, we briefly explain the intuition behind the proof. For any given copula C supported in a subset of B_C possibly with no density (see Figure 1a), we first mix it with the comonotonic copula C^+, so that the resulting copula $C' \geqslant C$ has a support that includes the diagonal line in $[0,1]$ (see Figure 1b). Second, we run a continuum of concordance-increasing (CI) transfers of Tchen (1980) (see Figure 1b-c) on C' to obtain another copula $\widehat{C} \geqslant C'$ which has positive density on a subset of B_C. Finally, we run another continuum of CI transfers on $\widehat{C}$ to arrive at a copula $\widetilde{C} \geqslant \widehat{C}$ which has positive density on B_C (see Figure 1c-d).

(a) Black lines: support of C; grey area: B_C

(b) Black lines: support of C'; arrows: a CI transfer

(c) Blue area: density area of $\widehat{C}$; arrows: a CI transfer

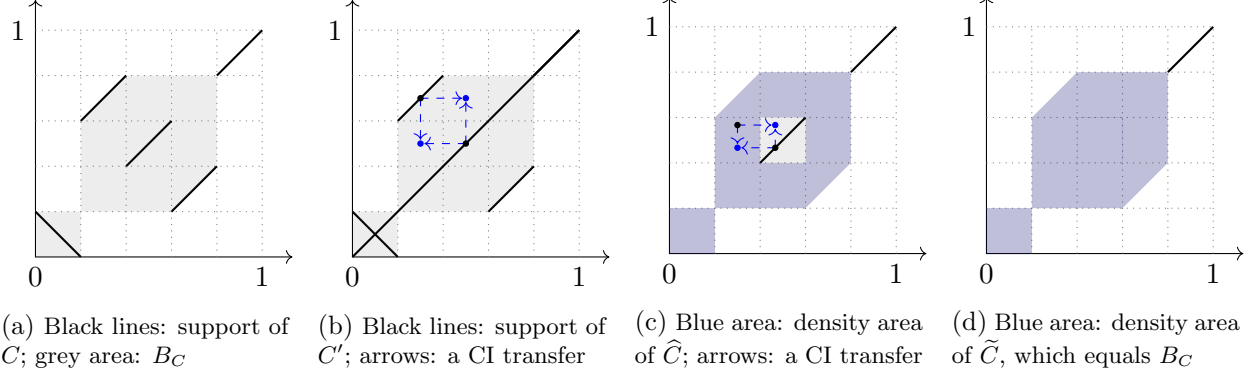(d) Blue area: density area of $\widetilde{C}$, which equals B_C

Figure 1: Intuition behind Lemma 3; here, C is a 5×5 checkerboard copula

We continue to prove Proposition 9. Let $\mathcal{D}$ be a bivariate concentration class and take $C \in \mathcal{D}$. Suppose for the purpose of contradiction that there does not exist $p \in (0,1)$ such that $C(p,p) = p$. Note that if $s_C(p) = p$, then $C(p,p) = \mathbb{P}(U \leqslant p, V \leqslant p) = \mathbb{P}(U \leqslant p) = p$. Similarly, if $t_C(p) = p$, then $C(p,p) = p$. Hence, our assumption on C implies that $t_C(p) < p < s_C(p)$ for all $p \in (0,1)$.

Take $\widetilde{C}$ as the one in Lemma 3. Since $\mathcal{D}$ is a bivariate concentration class and $\widetilde{C} \geqslant C$, we have $\widetilde{C} \in \mathcal{D}$. Let

$$\widetilde{C}_{2|1}(v|u) = \frac{\partial \widetilde{C}}{\partial u}(u,v), \quad \text{if the partial derivative exists.} \tag{9}$$

It is known that $\widetilde{C}_{2|1}$ is a conditional distribution of V given U, where $(U,V) \sim \widetilde{C}$; see Joe (2014, Section 2.1.3). As a consequence, $\widetilde{C}_{2|1}$ exists almost everywhere and takes value in $[0,1]$. Let $\widetilde{C}_{2|1}^{-1}(v|u)$ be the corresponding conditional v-quantile of V given $U = u$ for $u,v \in (0,1)$; we omit "almost everywhere".

Take a sequence $(U_n)_{n \in \mathbb{N}}$ of iid random variables uniformly distributed on $[0,1]$. We will construct a Markov process $(X_n)_{n \in \mathbb{N}}$ as follows. Let

$$X_1 = U_1 \quad \text{and} \quad X_{n+1} = \widetilde{C}_{2|1}^{-1}(U_{n+1}|X_n) \text{ for } n \geqslant 1. \tag{10}$$

By construction, (X_n, X_{n+1}) has the distribution $\widetilde{C}$ for each n; see Joe (2014, Section 6.9). Moreover, $(X_n)_{n \in \mathbb{N}}$ is obviously Markov and stationary. Since $t_C(u) < u < s_C(u)$ for all $u \in (0,1)$ and $\widetilde{C}$ has positive density on B_C, we know that each state in $(0,1)$ is reachable from any state in $(0,1)$. This means that the Markov process $(X_n)_{n \in \mathbb{N}}$ is irreducible (Definition 4.1 of Tierney (1996)). Since an irreducible and stationary Markov process satisfies LLN (see Theorem 4.3 of Tierney (1996)), we have $n^{-1} \sum_{i=1}^{n} X_n \xrightarrow{\text{p}} \mathbb{E}[X_1]$. This shows that $\widetilde{C}$ is diversifiable, contradicting $\widetilde{C} \in \mathcal{D}$. Therefore, we conclude that $C(p,p) = p$ for some $p \in (0,1)$, leading to the desired "only if" statement. $\square$

Remark 3. We comment on two technical points. First, convexity [Cx] is not needed in the proof

of the "only if" direction in Proposition 9. Therefore, a copula C is in any set satisfying [ND] and [DM] if and only if $C \in \mathcal{D}_p$ for some $p \in (0, 1)$. Second, to show that $\widetilde{C}$ is diversifiable in the above proof, we constructed the Markov process $(X_n)_{n \in \mathbb{N}}$ in (10) with uniform marginal distributions. A strictly increasing transform on $(X_n)_{n \in \mathbb{N}}$ does not matter as the resulting Markov process is always irreducible. Hence, we may alternatively define non-diversifiability of a copula by requiring (7) to hold only for the sequentially $\widetilde{C}$-coupled process with Markov dependence (10) and a uniform marginal distribution (or another continuous marginal distribution), and our results in Propositions 8 and 9 remain valid.

Proposition 9 leads to a characterization of $\mathcal{D}_p$ as an important subclass of bivariate concentration classes. A bivariate concentration class $\mathcal{D}$ is *maximal* if there does not exist another bivariate concentration class $\mathcal{D}' \neq \mathcal{D}$ containing $\mathcal{D}$.

Theorem 5. *The set $\mathcal{D} \subseteq \mathcal{C}_2$ is a maximal bivariate concentration class if and only if $\mathcal{D} = \mathcal{D}_p$ for some $p \in (0, 1)$.*

Proof. We first show the "if" statement. To show that $\mathcal{D}_p$ is maximal, suppose that a bivariate concentration class $\mathcal{D}$ satisfies $\mathcal{D}_p \subseteq \mathcal{D}$ and $\mathcal{D}$ contains a copula $C \notin \mathcal{D}_p$. This means $C(p, p) < p$ since $C(p, p) \leqslant p$ is satisfied by any copula. Take another copula $C' \in \mathcal{D}_p$ satisfying $C'(u, u) < u$ for all $u \in (0, 1) \setminus \{p\}$. Such a copula can be obtained by, for instance, mixing a Lebesgue measure on $[0, p]^2$ and a Lebesgue measure on $[p, 1]^2$. Let $C^* = C/2 + C'/2$ which is in $\mathcal{D}$ since $\mathcal{D}$ is convex. We have $C^*(u, u) < u$ for all $u \in (0, 1)$, and hence C^* is not in any $\mathcal{D}_u$. By Proposition 9, C^* is not in any bivariate concentration class, a contradiction to $C^* \in \mathcal{D}$. Therefore, $\mathcal{D}_p$ is maximal.

Below, we show the "only if" statement. Let $\mathcal{D}$ be a bivariate concentration class. If there exists $p \in (0, 1)$ such that $C(p, p) = p$ for all $C \in \mathcal{D}$, then we have $\mathcal{D} \subseteq \mathcal{D}_p$, and the maximality of $\mathcal{D}$ implies $\mathcal{D} = \mathcal{D}_p$.

Next, we suppose that there does not exist $p \in (0, 1)$ such that $C(p, p) = p$ for all $C \in \mathcal{D}$. We will show that this case is not possible by contradiction. Take any $\varepsilon \in (0, 1/2)$. For each $p \in [\varepsilon, 1 - \varepsilon]$, our assumption implies that there exists a copula $C^p \in \mathcal{D}$ such that $C^p(p, p) < p$. Note that $C^p(u, u)$ is continuous in u since all copula functions are Lipchitz continuous. As a consequence, there exists an open interval I_p with $p \in I_p$ such that $C^p(u, u) < u$ for all $u \in I_p$. Clearly, $\{I_p : p \in [\varepsilon, 1 - \varepsilon]\}$ is an open cover of $[\varepsilon, 1 - \varepsilon]$. Since $[\varepsilon, 1 - \varepsilon]$ is compact, there exists a finite subcover, denoted by $\{I_{p_i} : i = 1, \ldots, n\}$, which satisfies $\bigcup_{i=1}^n I_{p_i} \supseteq [\varepsilon, 1 - \varepsilon]$. Write $C^{[\varepsilon]} = n^{-1} \sum_{i=1}^n C^{p_i}$. We have $C^{[\varepsilon]} \in \mathcal{D}$ since $\mathcal{D}$ is convex. Moreover, $C^{[\varepsilon]}$ satisfies $C^{[\varepsilon]}(u, u) < u$ for all $u \in [\varepsilon, 1 - \varepsilon]$.

Define $C^* = \sum_{k=1}^\infty 2^{-k} C^{[3^{-k}]}$. Convexity of $\mathcal{D}$ for countable sums implies $C^* \in \mathcal{D}$. On the other hand, $C^*(u, u) < u$ for all $u \in (0, 1)$ by construction. Using Proposition 9, we know that C^*

is not in a bivariate concentration class. This yields a contradiction. □

Remark 4. In our formulation of [Cx], convexity of $\mathcal{D}$ is required to hold for countable sums. This property is used in the last step of the proof of Theorem 5 to yield that C^* is in $\mathcal{D}$. The current proof techniques do not work if we require convexity of $\mathcal{D}$ only for finite sums.

To conclude the paper, we put Theorems 4 and 5 together to arrive at a complete endogenous reasoning for a regulator to use ES.

Theorem 6. *A monetary risk measure ρ on L^∞ satisfies lower semicontinuity, $\rho(0) = 0$, and $\mathcal{D}$-aversion for a maximal bivariate concentration class $\mathcal{D}$ if and only if $\rho = \mathrm{ES}_p$ for some $p \in (0,1)$.*

Proof. This result follows from combining Theorems 4 and 5, and noting that $\mathcal{D}_p$-aversion is equivalent to [p-CA]. □

To interpret Theorem 6, we make the following economic assumptions on the regulator's preference towards dependence structures in risk aggregation. First, the regulator believes that breaking LLN is dangerous ([ND]); second, the regulator believes that more positive dependence is more dangerous ([DM]); third, the regulator believes that a mixture of dangerous structures is dangerous ([Cx]); fourth, the regulator chooses to use a largest possible set to model such dangerous structures (maximality). If all four assumptions are met, then, by Theorem 5, the regulator needs to use a risk measure that satisfies concentration aversion. With some other standard properties in Theorem 4, we further arrive at the class of ES. Certainly, the desirability of the four assumptions on the regulator's dependence preference can be debated, and, based on the main results of this paper, such debates can directly translate to critical arguments for or against the use of ES in financial regulation.

Remark 5. If maximality is removed from the consideration of the regulator, then we can allow for other bivariate concentration classes. The simplest such example is the singleton $\mathcal{D}^+ := \{C^+\}$, which is clearly also the smallest bivariate concentration class. As shown by Mao and Wang (2020), a monetary risk measure is $\mathcal{D}^+$-averse if and only if it is SSD-consistent. Therefore, by Theorem 6, the maximality of the bivariate concentration class $\mathcal{D}$ pins down the class of ES among all lower-semicontinuous consistent risk measures. This shows that, among a general class of risk measures, ES has the largest spectrum of dangerous dependence. In other words, if maximality of $\mathcal{D}$ is desirable, then ES is the only suitable class; if maximality is relaxed to somewhere between the largest and the smallest, then the regulator has more choices of SSD-consistent risk measures. A larger set of dangerous dependence narrows down the corresponding choices of regulatory risk measures, from all SSD-consistent ones to ES.

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. *The Review of Financial Studies*, **14**(2), 371–405.

BCBS (2016). *Minimum Capital Requirements for Market Risk. January 2016.* Basel Committee on Banking Supervision. Basel: Bank for International Settlements. `https://www.bis.org/bcbs/publ/d352.htm`

BCBS (2019). *Minimum Capital Requirements for Market Risk. February 2019.* Basel Committee on Banking Supervision. Basel: Bank for International Settlements. `https://www.bis.org/bcbs/publ/d457.htm`

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

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

Cont, R., Deguest, R. and Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance*, **10**(6), 593–606.

Daníelsson, J., Embrechts, P., Goodhart, C., Keating, C., Muennich, F., Renault, O. and Shin, H. S. (2001). An academic response to Basel II. *LSE Special Paper Series May 2001.*

Delbaen, F. (2002). Coherent risk measures on general probability spaces. In *Advances in Finance and Stochastics* (pp. 1–37). Springer, Berlin, Heidelberg.

Embrechts, P., Mao, T., Wang, Q. and Wang, R. (2021). Bayes risk, elicitability, and the Expected Shortfall. *Mathematical Finance*, **31**, 1190–1217.

Embrechts, P., Schied, A. and Wang, R. (2022). Robustness in the optimization of risk measures. *Operations Research*, **70**(1), 95–110.

Föllmer, H. and Schied, A. (2002). Convex measures of risk and trading constraints. *Finance and Stochastics*, **6**, 429–447.

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

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.

Gaivoronski, A. and Pflug, G. (2005). Value-at-Risk in portfolio optimization: Properties and computational approach. *Journal of Risk*, **7**(2), 1–31.

Guo, N., Kou, S., Wang, B. and Wang, R. (2020). Self-consistency, subjective pricing, and a theory of credit rating. *SSRN*: 3504065.

Hadar, J. and Russell, W. (1969). Rules for ordering uncertain prospects. *American Economic Review*, **59**(1), 25–34.

Hampel, F. (1971). A general qualitative definition of robustness. *Annals of Mathematical Statistics*, **42**(6), 1887–1896.

He, X. and Peng, X. (2018). Surplus-invariant, law-invariant, and conic acceptance sets must be the sets induced by Value-at-Risk. *Operations Research*, **66**(5), 1268–1276.

Herdegen, M. and Khan, N. (2022). Mean-ρ portfolio selection and ρ-arbitrage for coherent risk measures. *Mathematical Finance*, **32**(1), 226–272.

Joe, H. (2014). *Dependence Modeling with Copulas*. London: Chapman & Hall.

Kou, S. and Peng, X. (2016). On the measurement of economic tail risk. *Operations Research*, **64**(5), 1056–1072.

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

Lin, H., Saunders, D. and Weng, C. (2021). Mean-expectile portfolio selection. *Applied Mathematics and Optimization*, **83**, 1585–1612.

Liu, F. and Wang, R. (2021). A theory for measures of tail risk. *Mathematics of Operations Research*, **46**(3), 1109–1128.

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

Marinacci, M. and Montrucchio, L. (2004). Introduction to the mathematics of ambiguity. In *Uncertainty in Economic Theory* (I. Gilboa, Ed.) 46–107. Routledge, New York.

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

Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*. Wiley, England.

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

Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, **2**(3), 21–41.

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 deviation in risk analysis. *Finance and Stochastics*, **10**, 51–74.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica*, **57**(3), 571–587.

Tchen, A. H. (1980). Inequalities for distributions with given marginals. *Annals of Probability*, **8**(4), 814–827.

Tierney, L. (1996). Introduction to general state-space Markov chain theory. In *Markov Chain Monte-Carlo in Practice* (W. Gilks, S. Richardson, and D. Spiegelhalter, Eds), 59–74. Chapman and Hall, London.

Vovk, V. and Wang, R. (2021). E-values: Calibration, combination, and applications. *Annals of Statistics*, **49**(3), 1736–1754.

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

Wang, R. and Wei, Y. (2020). Risk functionals with convex level sets. *Mathematical Finance*, **30**(4), 1337–1367.

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.

A   Proof of Lemma 3

Proof of Lemma 3. Let $C' = C/2 + C^+/2$. Note that $C^+ \geqslant C$, and hence $C' \geqslant C$. Moreover, since $t_{C^+} = s_{C^+}$ is the identity on $(0,1)$, we have $t_{C'} = \min(t_C, t_{C^+}) = t_C$ and $s_{C'} = \max(s_C, s_{C^+}) = s_C$. See Figure 1a-b for an illustration of C and C'. It suffices to show that there exists $\widetilde{C} \geqslant C'$ such that $\widetilde{C}$ has positive density on B_C.

Below, for simplicity, we will use the notation C for C' above. (Alternatively, we can directly assume that the measure C^+ is absolutely continuous with respect to the measure C, and the above argument guarantees that this assumption is without loss of generality.)

Take $(U, V) \sim C$ and $(U', V') \sim C$ such that (U, V) and (U', V') are independent. Construct a random variable $\widetilde{V}$ by

$$\widetilde{V} = V \mathbb{1}_{A^c} + V' \mathbb{1}_A, \text{ where } A = \{U > U', \ V < V'\} \cup \{U < U', \ V > V'\}. \tag{11}$$

Note that since (U, V) and (U', V') are iid, we have, for $v \in (0,1)$,

$$\mathbb{P}(V' \leqslant v, \ A) = \mathbb{P}(V' \leqslant v, \ U > U', \ V < V') + \mathbb{P}(V' \leqslant v, \ U < U', \ V > V')$$
$$= \mathbb{P}(V \leqslant v, \ U' > U, \ V' < V) + \mathbb{P}(V \leqslant v, \ U' < U, \ V' > V) = \mathbb{P}(V \leqslant v, \ A).$$

Hence,

$$\mathbb{P}(\widetilde{V} \leqslant v) = \mathbb{P}(V \leqslant v, A^c) + \mathbb{P}(V' \leqslant v, A) = \mathbb{P}(V \leqslant v, A^c) + \mathbb{P}(V \leqslant v, A) = \mathbb{P}(V \leqslant v) = v$$

implying that $\widetilde{V}$ is uniformly distributed on $[0,1]$. As a consequence, the distribution of $(U, \widetilde{V})$ is a

copula, and we denote it by $\widehat{C}$.

We first verify $\widehat{C} \geqslant C$. Using the fact that (U, V) and (U', V') are iid, we get, for $(u, v) \in [0, 1]^2$,

$$\mathbb{P}((U, V') \leqslant (u, v),\ U < U',\ V > V') - \mathbb{P}((U, V) \leqslant (u, v),\ U < U',\ V > V')$$

$$= \mathbb{P}(U \leqslant u,\ V' \leqslant v < V,\ U < U')$$

$$= \mathbb{P}(U' \leqslant u,\ V \leqslant v < V',\ U' < U)$$

$$\geqslant \mathbb{P}(U \leqslant u,\ V \leqslant v < V',\ U > U')$$

$$= \mathbb{P}((U, V) \leqslant (u, v),\ U > U',\ V < V') - \mathbb{P}((U, V') \leqslant (u, v),\ U > U',\ V < V').$$

As a consequence,

$$\mathbb{P}((U, V') \leqslant (u, v),\ A) \geqslant \mathbb{P}((U, V) \leqslant (u, v),\ A),$$

and hence

$$\mathbb{P}((U, \widetilde{V}) \leqslant (u, v)) = \mathbb{P}((U, V) \leqslant (u, v),\ A^c) + \mathbb{P}((U, V') \leqslant (u, v),\ A) \geqslant \mathbb{P}((U, V) \leqslant (u, v)),$$

which gives the order $\widehat{C}(u, v) \geqslant C(u, v)$. Intuitively, this is because $\widehat{C}$ is obtained from C via a continuum of CI transfers (see Figure 1b).

Finally, we verify the statement on the positive density on B_C. For $(s, t) \in [0, 1]^2$,

$$\widehat{C}(s, t) = \mathbb{E}\left[\mathbb{P}\left((U, \widetilde{V}) \leqslant (s, t) \mid U, V\right)\right]$$

$$= \int_{[0,1]^2} \mathbb{P}\left((U, \widetilde{V}) \leqslant (s, t) \mid (U, V) = (u, v)\right) dC(u, v)$$

$$= \int_{[0,s] \times [0,1]} \mathbb{P}\left(\widetilde{V} \leqslant t \mid (U, V) = (u, v)\right) dC(u, v)$$

$$= \int_{[0,s] \times [0,1]} \left(\mathbb{P}\left(V' \leqslant t,\ U' > u,\ V' < v\right) + \mathbb{P}\left(V' \leqslant t,\ U' < u,\ V' > v\right)\right) dC(u, v)$$

$$+ \int_{[0,s] \times [0,t]} \left(\mathbb{P}\left(U' \leqslant u,\ V' \leqslant v\right) + \mathbb{P}\left(U' \geqslant u,\ V' \geqslant v\right)\right) dC(u, v).$$

Write $t \wedge v = \min(t, v)$. We have $\widehat{C} = F + G$ where

$$F(s, t) = \int_{[0,s] \times [0,1]} \left(\mathbb{P}\left(V' \leqslant t \wedge v,\ U' > u\right) + \mathbb{P}\left(v < V' \leqslant t,\ U' < u\right)\right) dC(u, v),$$

and

$$G(s, t) = \int_{[0,s] \times [0,t]} \left(\mathbb{P}\left(U' \leqslant u,\ V' \leqslant v\right) + \mathbb{P}\left(U' \geqslant u,\ V' \geqslant v\right)\right) dC(u, v).$$

Note that F and G are the distribution functions of two Borel measures on $[0,1]^2$. Below we will show that F has a positive density on a subset of B_C (this is shown in Figure 1c), and then we use another construction to obtain positive density on B_C.

Since V' is uniform on $[0,1]$, we know $\mathbb{P}(V' \leqslant t \wedge v) = t \wedge v$, and hence

$$F(s,t) = \int_{[0,s] \times [0,1]} \left(t \wedge v - \mathbb{P}\left(V' \leqslant t \wedge v,\ U' < u\right) + \mathbb{P}\left(v < V' \leqslant t,\ U' < u\right) \right) \mathrm{d}C(u,v)$$
$$= \int_{[0,s] \times [0,1]} \left(t \wedge v + C(u,t) - 2C(u, t \wedge v) \right) \mathrm{d}C(u,v).$$

Using (see Section 2.12 of Joe (2014))

$$C_{2|1}(v|u) = \frac{\partial C}{\partial u}(u,v) \quad \text{and} \quad C_{1|2}(u|v) = \frac{\partial C}{\partial v}(u,v) \qquad \text{almost everywhere,} \tag{12}$$

we get

$$\frac{\partial F}{\partial s}(s,t) = \int_{[0,1]} \left(t \wedge v + C(s,t) - 2C(s, t \wedge v) \right) C_{2|1}(\mathrm{d}v|s).$$

Exchanging the order of the derivative and the integral (guaranteed by the dominated convergence theorem), we get

$$\frac{\partial^2 F}{\partial s \partial t}(s,t) = \int_{[0,1]} \left(\left(1 - C_{1|2}(s|t)\right) \mathbb{1}_{\{t<v\}} + C_{1|2}(s|t)\mathbb{1}_{\{t>v\}} \right) C_{2|1}(\mathrm{d}v|s)$$
$$= \left(1 - C_{1|2}(s|t)\right)\left(1 - C_{2|1}(t|s)\right) + C_{1|2}(s|t)C_{2|1}(t|s).$$

Therefore, we have $\frac{\partial^2 F}{\partial s \partial t}(s,t) > 0$ as soon as $(C_{1|2}(s|t), C_{2|1}(t|s))$ is not $(0,1)$ or $(1,0)$. Equivalently, $\frac{\partial^2 F}{\partial s \partial t}(s,t) > 0$ if the support of C includes either (s',t) to the left of (s,t) and (s,t') below (s,t), or (s',t) to the right of (s,t) and (s,t') above (s,t). Since the support of C includes the diagonal line, F has positive density on the set

$$B_C^* = \{(s,t) : t_C^*(s) \leqslant t \leqslant s_C^*(s)\},$$

where $t_C^*(s)$ is the essential infimum of $C_{2|1}(\cdot|s)$ and $s_C^*(s)$ is the essential supremum of $C_{2|1}(\cdot|s)$. Since $\widehat{C} = F + G$, we obtain that $\widehat{C}$ has positive density on B_C^*, possibly plus a non-absolutely continuous component coming from G.

Generally, the set B_C^* may be different from the set B_C. To obtain a positive density on B_C, we apply the above procedure again with $\widehat{C}$ in place of C, starting from (11). This time, we arrive

at a new copula $\widetilde{C} \geqslant \widehat{C}$ such that $\widetilde{C}$ has positive density on $B^*_{\widehat{C}}$ (this is shown in Figure 1d). We claim $B_C \subseteq B^*_{\widehat{C}}$. To show this, take $(s, t) \in B_C$. Assume $t \geqslant s$, and the case $s < t$ is symmetric. By definition of B_C, there exists (s', t') in the support of C such that $s' \leqslant s$ and $t' \leqslant t$. This implies that $(s', t) \in B^*_C$. Since $\widehat{C}$ has positive density on B^*_C, we know that (s', t) and (s, s) are both in the support of $\widehat{C}$. This further implies $(s, t) \in B^*_{\widehat{C}}$, and hence $B_C \subseteq B^*_{\widehat{C}}$. Therefore, we conclude that $\widetilde{C}$ has positive density on B_C. $\qquad\square$