

Comonotonic measures of multivariate risks

Ivar Ekeland, Alfred Galichon, Marc Henry

▶ **To cite this version:**

Ivar Ekeland, Alfred Galichon, Marc Henry. Comonotonic measures of multivariate risks. Mathematical Finance, 2012, 22 (1), pp.109-132. 10.1111/j.1467-9965.2010.00453.x . hal-01053550

HAL Id: hal-01053550

https://sciencespo.hal.science/hal-01053550

Submitted on 31 Jul 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

aptara	MAFI	mafi_453	**Dispatch:** September 9, 2010	**CE:** WRV
	Journal	**MSP No.**	**No. of pages:** 24	**PE:** Angélique

Mathematical Finance, Vol. xx, No. x (xxxx 2010), 1–24

COMONOTONIC MEASURES OF MULTIVARIATE RISKS

IVAR EKELAND

University of British Columbia

ALFRED GALICHON

Department of Economics, École polytechnique

MARC HENRY

Université de Montréal

We propose a multivariate extension of a well-known characterization by S. Kusuoka of regular and coherent risk measures as maximal correlation functionals. This involves an extension of the notion of comonotonicity to random vectors through generalized quantile functions. Moreover, we propose to replace the current law invariance, subadditivity, and comonotonicity axioms by an equivalent property we call *strong coherence* and that we argue has more natural economic interpretation. Finally, we reformulate the computation of regular and coherent risk measures as an optimal transportation problem, for which we provide an algorithm and implementation.

KEY WORDS: regular risk measures, coherent risk measures, comonotonicity, maximal correlation, optimal transportation, strongly coherent risk measures.

1. INTRODUCTION

The notion of coherent risk measure was proposed by Artzner et al. (1999) as a set of axioms to be verified by a real-valued measure of the riskiness of an exposure. In addition to monotonicity, positive homogeneity and translation invariance, the proposed coherency axioms include subadditivity, which is loosely associated with hedging. Given this interpretation, it is natural to require the risk measure to be additive on the subsets of risky exposures that are *comonotonic*, as this situation corresponds to the worse-case scenario for the correlation of the risks. In Kusuoka (2001), Kusuoka showed the remarkable result that law invariant coherent risk measures that are also comonotonic additive are defined by the integral of the quantile function with respect to a positive

The authors thank two anonymous referees and an associate Editor who helped improving the paper. The authors are grateful to Guillaume Carlier, Rose-Anne Dana, Nicole El Karoui, Jean-Charles Rochet, Ludger Rüschendorf, and Nizar Touzi, as well as participants to the workshop "Multivariate and Dynamic Risk Measures" (IHP, Paris, October 2008), the SAS seminar at Princeton University, the Statistics Department Risk seminar at Columbia University for helpful discussions and comments. Galichon and Ekeland gratefully acknowledge support from Chaire EDF-Calyon "Finance and Développement Durable" and FiME, Laboratoire de Finance des Marchés de l'Energie (www.fime-lab.org), and Galichon that of Chaire Société Générale "Risques Financiers" and Chaire Axa "Assurance et Risques Majeurs".

Manuscript received October 2007; final revision received November 2009.

Address correspondence to Alfred Galichon, Department of Economics, École polytechnique, 91128 Palaiseau, France; e-mail: alfred.galichon@polytechnique.edu.

DOI: 10.1111/j.1467-9965.2010.00453.x
© 2010 Wiley Periodicals, Inc.

measure, a family that includes Expected Shortfall (also known as Conditional Value at Risk, or Expected Tail Loss).

The main drawback of this formulation is that it does not properly handle the case when the numéraires in which the risky payoffs are labeled are not perfect substitutes. This situation is commonly met in Finance. In a two-country economy with floating exchange rates, the fact that claims on payoffs in different currencies are not perfectly substitutable is known as the *Siegel paradox*; in the study of the term structure of interest rates, the fact that various maturities are (not) perfect substitutes is called the (failure of the) *pure expectation hypothesis*. The technical difficulty impeding a generalization to the case of a multivariate risk measure is that the traditional definition of comonotonicity relies on the order in $\mathbb{R}$. When dealing with portfolios of risk that are not perfectly substituable, as Jouini, Meddeb, and Touzi did in Jouini, Meddeb, and Touzi (2004) for coherent risk measures, and Rüschendorf (2006) for law invariant convex risk measures, the right notion of multivariate comonotonicity is not immediately apparent.

This work circumvents these drawbacks to generalize Kusuoka's result to multivariate risk portfolios, and proposes a simplifying reformulation of the axioms with firm decision theoretic foundations. First, we propose an alternative axiom called *strong coherence*, which is equivalent to the axioms in Kusuoka (2001) and which, unlike the latter, extends to the multivariate setting. We then make use of a variational characterization of Kusuoka's axioms and representation in order to generalize his results to the multivariate case. We show that multivariate risk measures that satisfy strong coherence have the same representation as in Kusuoka (2001), which we discuss further later.

The work is organized as follows. The first section motivates a new notion called strong coherence which is shown to be intimately related to existing risk measures axioms, yet appears to be more natural. The second section shows how the concept of comonotonic regular risk measures can be extended to the case of multivariate risks, by introducing a proper generalization of the notion of comonotonicity and giving a representation theorem. The third section discusses in depth the relation with Optimal Transportation Theory, and shows important examples of actual computations.

1.1. Notations and Conventions

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, which is *standard* in the terminology of (Jouini, Schachermayer, and Touzi 2006), that is $\mathbb{P}$ is nonatomic and $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is separable. Let $X : \Omega \to \mathbb{R}^d$ be a random vector; we denote the distribution law of X by $\mathcal{L}_X$, hence $\mathcal{L}_X = X\#\mathbb{P}$, where $X\#\mathbb{P} := \mathbb{P}X^{-1}$ denotes the *push-forward* of probability measure $\mathbb{P}$ by X. The *equidistribution class* of X is the set of random vectors with distribution with respect to $\mathbb{P}$ equal to $\mathcal{L}_X$ (reference to $\mathbb{P}$ will be implicit unless stated otherwise). As explained in the Appendix, essentially one element in the equidistribution class of X has the property of being the gradient of a convex function; this random element is called the (generalized) *quantile function* associated with the distribution $\mathcal{L}_X$ and denoted by Q_X (in dimension 1, this is the quantile function of distribution $\mathcal{L}_X$ in the usual sense). We denote by $\mathcal{M}(\mathcal{L}, \mathcal{L}')$ the set of probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals $\mathcal{L}$ and $\mathcal{L}'$. We call $L^2_d(\mathbb{P})$ (abbreviated in L^2_d) the equivalence class of $\mathcal{F}$-measurable functions $\Omega \to \mathbb{R}^d$ with a finite second moment modulo $\mathbb{P}$ -negligible events. We call $\mathcal{P}_2(\mathbb{R}^d)$ the set of probability on $\mathbb{R}^d$ with finite second moment. Finally, for two elements X, Y of L^2_d, we write $X \sim Y$ to indicate equality in distribution, that is $\mathcal{L}_X = \mathcal{L}_Y$. We also write $X \sim \mathcal{L}_X$. Define c.l.s.c.$(\mathbb{R}^d)$ as the class of convex lower

semi-continuous functions on $\mathbb{R}^d$, and the *Legendre-Fenchel conjugate* of $V \in \text{c.l.s.c.}(\mathbb{R}^d)$ as $V^*(x) = \sup_{y\in\mathbb{R}^d}[x \cdot y - V(y)]$. In all that follows, the dot "$\cdot$" will denote the standard scalar product in $\mathbb{R}^d$ or L^2_d. $\mathcal{M}_d(\mathbb{R})$ denotes the set of $d \times d$ matrices, and $\mathcal{O}_d(\mathbb{R})$ the orthogonal group in dimension d. For $M \in \mathcal{M}_d(\mathbb{R})$, M^T denotes the matrix transpose of M. For a function $V : \mathbb{R}^d \to \mathbb{R}$ differentiable at x, we denote $\nabla V(x)$ the gradient of V at x; this is the vector $(\frac{\partial V(x)}{\partial x_1}, \ldots, \frac{\partial V(x)}{\partial x_d}) \in \mathbb{R}^d$. When V is twice differentiable at x, we denote D^2V the Hessian matrix of V that is the matrix $(\frac{\partial^2 V(x)}{\partial x_i \partial x_j})_{1\le i,j\le d}$. By Aleksandrov's theorem, a convex function is (Lebesgue-) almost everywhere differentiable on the interior of its domain (see Villani 2003, pp. 58–59), so ∇V and D^2V exist almost everywhere. For a functional Φ defined on a Banach space, we denote $D\Phi$ its Fréchet derivative.

2. STRONG COHERENCE: A NATURAL AXIOMATIC CHARACTERIZATION

In this section, we advocate a very simple axiomatic setting, called *strong coherence* which will be shown to be equivalent to the more classical axiomatic framework described in the next section. We argue that this axiom has more intuitive appeal than the classical (equivalent) axioms.

2.1. Motivation: Structure Neutrality

The regulating instances of the banking industry are confronted with the problem of imposing rules to the banks to determine the amount of regulatory capital they should budget to cover their risky exposure. A notable example of such a rule is the Value-at-Risk, imposed by the Basel II committee, but a number of competing rules have been proposed. We call $X \in L^2_d$ the vector of random losses[1] of a given bank. Note that contrary to a convention often adopted in the literature, we chose to account positively for net losses: X is a vector of effective losses. Also note that we have supposed that the risk is multivariate, which means that there are multiple numéraires, which, depending on the nature of the problem, can be several assets, several term maturities, or several non-monetary risks of different nature. We suppose that these multiple num éraires cannot be easily exchanged into one another: *the problem is intrinsiquely multivariate*. This could be the case if the firm (or the regulator) is unable or unwilling to define a monetary equivalent for the various dimensions of its risks. For instance, an oil company is likely to be unable to estimate a dollar amount to price its environmental risk; similarly, a pharmaceutical company may be unwilling to give a monetary estimate for the health hazard its product carry.

To a random vector of losses X, one associates a number $\varrho(X)$ which measures the intensity of the risk incurred. The unit in which ϱ is to be thought of as some extra currency unity, or alternatively a non-monetary score; it is not assumed to be one of the monetary units associated with the various dimensions of the vector of the risks. This score is used by investors to compare the risks of two companies, or by regulators to set limits to risky exposures for regulated firms. An important desirable feature of the rule proposed by the regulator is to avoid regulatory arbitrage. Here, a regulatory arbitrage

[1] In this paper, we have chosen to restrict ourselves to the case where risks are in $L^2(\mathbb{R}^d)$ for notational convenience, but all results in the paper carry without difficulty over to the case where the risks are in $L^p(\mathbb{R}^d)$ for $p \in (1, +\infty)$.

would be possible if the firms could split their risk into several different subsidiaries S_i, $i = 1, \ldots, N$ with independent legal existence, so that the shareholder's economic risk remained the same $X = X_1 + \cdots + X_N$, but such that the amount of the shareholder's capital which is required to be budgeted to cover their risk were strictly inferior after the split, namely such that $\varrho(X) > \varrho(X_1) + \cdots + \varrho(X_N)$. To avoid this, we shall impose the requirement of *subadditivity*, that is

$$\varrho(X_1 + \cdots + X_N) \leq \varrho(X_1) + \cdots + \varrho(X_N)$$

for all possible dependent risk exposures $(X_1, \ldots, X_N) \in (L^2_d)^N$. We now argue that the regulator is only interested in the amount and the intensity of the risk, not in its operational nature: the capital budgeted should be the same for a contingent loss of 1% of the total capital at risk no matter how the loss occurred (whether on the foreign exchange market, the stock market, the credit market, etc.) This translates mathematically into the requirement that the regulatory capital to budget should only depend on the distribution of the risk X, that is, the rule should satisfy the *law invariance* property:

DEFINITION 2.1. A functional $\varrho : L^2 \to \mathbb{R}$ is called law-invariant if $\varrho(X) = \varrho(Y)$ when $X \sim Y$, where $\sim$denotes equality in distribution.

By combining together subadditivity and law invariance, we get the natural requirement for the capital budgeting rule, that $\varrho(\tilde{X}_1 + \cdots + \tilde{X}_N) \leq \varrho(X_1) + \cdots + \varrho(X_N)$ for all $X, \tilde{X}$ in $(L^2_d)^N$ such that $X_i \sim \tilde{X}_i$ for all $i = 1, \ldots, N$. However, to prevent giving a premium to conglomerates, and to avoid imposing an overconservative rule to the regulated firms, one is led to impose the inequality to be sharp and pose the *structure neutrality* axiom

$$\varrho(X_1) + \cdots + \varrho(X_N) = \sup_{\tilde{X}_i \sim X_i} \varrho(\tilde{X}_1 + \cdots + \tilde{X}_N).$$

This requirement is notably failed by the Value-at-Risk, which leads to the fact that *the Value-at-Risk as a capital budgeting rule is not neutral to the structure of the firm*. This result should be read in the perspective of the corporate finance literature on the optimal structure of the firm, originating in the celebrated Modigliani-Miller theorem, according to which the value of the firm does not depend on the structure of its capital. This point is explained in detail in (Galichon forthcoming), where an explicit construction is provided. **Q1** We introduce the axiom of strong coherence to be satisfied by a measure of the riskiness of a portfolio of risk exposures (potential losses) $X \in L^2_d$.

DEFINITION 2.2 (Strong coherence). For $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, a functional $\varrho : L^2_d \to \mathbb{R}$ is called a strongly coherent risk measure if (i) it is convex continuous, and (ii) it is structure neutral: for all $X, Y \in \mathcal{L}^2_d$,

$$\varrho(X) + \varrho(Y) = \sup \left\{ \varrho(\tilde{X} + \tilde{Y}) \, : \, \tilde{X} \sim X; \; \tilde{Y} \sim Y \right\}.$$

The convexity axiom can be justified by a risk aversion principle: in general, one should prefer to diversify risk. The structure neutrality axiom, being defined as a supremum over all correlation structures, can be interpreted as a provision against worst-case scenarios, and may be seen as unduly conservative. However, this axiom is no more conservative than the set of axioms defining a regular coherent risk measure as we shall see.

As we shall see also, strongly coherent risk measures satisfy all the classical axioms of coherent risk measures (recalled in Definition 3.1) let alone monotonicity, for which the multivariate extension is not obvious. In particular, these measures satisfy positive homogeneity and translation invariance. They also satisfy law invariance, which can be seen by taking $Y = 0$ in the definition earlier.

We now show that strongly coherent risk measures are represented by maximal correlation functionals with respect to a given random vector or scenario.

2.2. Characterization of Strongly Coherent Risk Measures

We are now going to show that the strong coherence property essentially characterizes a class of risk measures known as *maximal correlation risk measures*, which we shall first recall the definition of.

2.2.1. Maximal Correlation Measures. We first define *maximal correlation risk measures* (in the terminology of Rüschendorf who introduced them in the multivariate case, see, e.g. Rüschendorf 2006). These measures will generalize the variational formulation for coherent regular risk measures given in (3.1).

DEFINITION 2.3 (Maximal correlation measures). A functional $\varrho_\mu : L_d^2 \to \mathbb{R}$ is called a maximal correlation risk measure with respect to a baseline distribution μ if for all $X \in L_d^2$,

$$\varrho_\mu(X) := \sup\left\{\mathbb{E}[X \cdot \tilde{U}] : \ \tilde{U} \sim \mu\right\}.$$

Our notion of maximal correlation risk measure is essentially the same as Rüschendorf's, with a few minor variants: Rüschendorf defines his measures on L_d^∞ instead of L_d^2, and imposes the extra requirements $U_i \geq 0$ and $E_\mu[U_i] = 1$ for $i = 1, \ldots, d$, which we do not impose for now.

REMARK 2.4 (Geometric interpretation). The maximum correlation measure with respect to measure μ is the support function of the equidistribution class of μ.

EXAMPLE 2.5 (Multivariate expected shortfall). An interesting example of univariate risk measure within the class of maximal correlation risk measures is the expected shortfall, also known as conditional value at risk. This risk measure can be generalized to the multivariate setting by defining the α-expected shortfall of a risk exposure X as the maximal correlation measure when the baseline risk U is a Bernoulli random vector (i.e., its distribution $\mathcal{L}_U$ is determined by $U = (1/\alpha, \ldots, 1/\alpha)^T$ with probability α and 0 with probability $1 - \alpha$). In such case, one can easily check that if $\mathcal{L}_X$ is absolutely continuous, then defining $W(x) = \max(\sum_{i=1}^d x_i - c, 0)$, with c given by requirement $Pr(\sum_{i=1}^d x_i \geq c) = \alpha$, it follows that W is convex and ∇W exists $\mathcal{L}_X$ almost everywhere and pushes $\mathcal{L}_X$ to $\mathcal{L}_U$ as in Proposition B.1; therefore, the maximal correlation measure is given in this case by $E[(\sum_{i=1}^d X_i)1\{\sum_{i=1}^d X_i \geq c\}]$. In other words, the maximum correlation measure in this example is the (univariate) α-expected shortfall for $Y = \sum_{i=1}^d X_i$.

EXAMPLE 2.6. With a more complex baseline risk, other important examples where explicit or numerical computation is possible include the cases when (1) the baseline risk

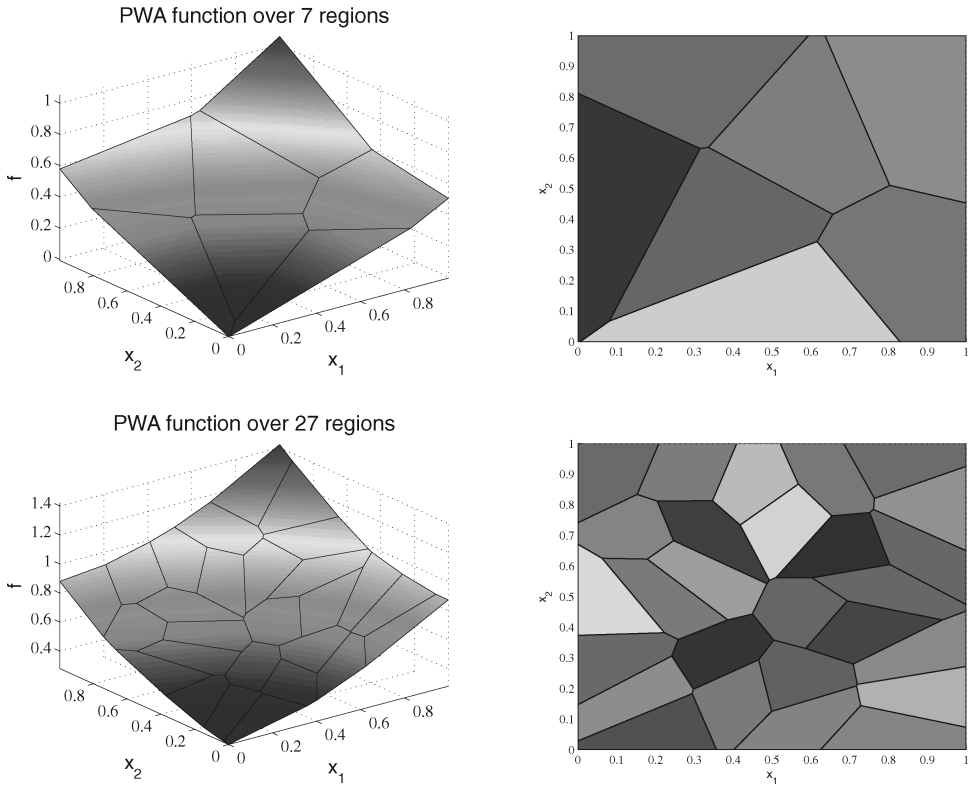


FIGURE 2.1. Mapping the uniform to a discrete discrete distribution in dimension $d = 2$. Upper row: seven atom points, lower row: 27 atom points. Left column: the potential $V(u) = w^*(u)$. Right column: the corresponding partition of the space $\mathcal{U}$.

and the risk to be measured are both Gaussian, or (2) the baseline risk is uniform on $[0, 1]^d$ and the risk to be measured has a discrete distribution. Both these cases are treated in detail in Section 4. **Q2**

Let us first recall the following lemma, which emphasizes the symmetry between the roles played by the equivalence class of X and U in the definition earlier.

LEMMA 2.7. *For any choice of* $U \sim \mu$, *with* $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, *one has*

$$\varrho_\mu(X) = \sup\left\{\mathbb{E}[\tilde{X} \cdot U] : \ \tilde{X} \sim X\right\},$$

and U *is called the baseline risk associated with* ϱ_μ. *It follows that* ϱ_μ *is law invariant.*

Proof. See (2.12) in Rüschendorf (2006). □

2.2.2. Characterization. We now turn to our first main result, which is a characterization of strongly coherent risk measures. We first prove a useful intermediate characterization in Proposition 2.9. We shall use Lemma A.4 from Jouini et al. (2006), which we

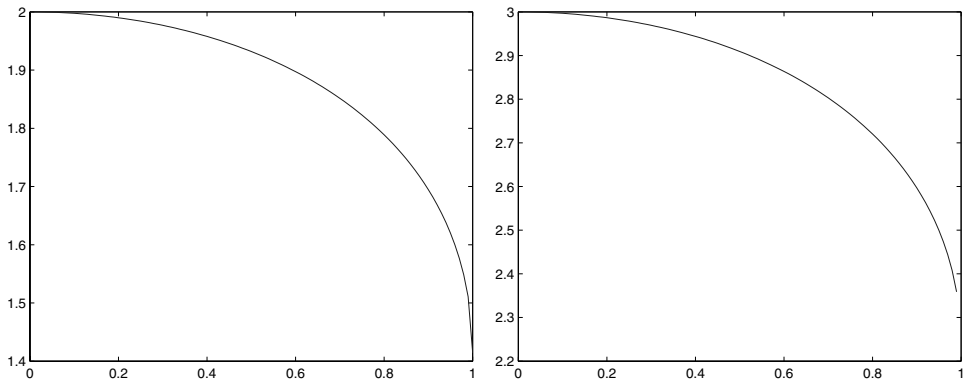


FIGURE 2.2. The value of the risk measure in the Gaussian case, plot against ρ. Left panel: $\sigma_1 = \sigma_2 = 1$. Right panel: $\sigma_1 = 1, \sigma_2 = 2$.

quote here for the reader's convenience. Denote by $\mathcal{A}$ the set of bimeasurable bijections σ from $(\Omega, \mathcal{A}, \mathbb{P})$ into itself which preserve the probability, so that $\sigma \# \mathbb{P} = \mathbb{P}$. Recall that $(\Omega, \mathcal{F}, \mathbb{P})$ was assumed to be a probability space which does not have atoms, and such that $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is separable.

LEMMA 2.8. *Let C be a norm closed subset of $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Then the following are equivalent:*

(1) C is law invariant, that is $X \in C$ and $X \sim Y$ implies that $Y \in C$
(2) C is transformation invariant, that is for any $X \in C$ and any $\sigma \in \mathcal{A}$, we have $X \circ \sigma \in C$

As an immediate consequence, we have the following result.

PROPOSITION 2.9. *A convex continuous functional $\varrho : L^2_d \to \mathbb{R}$ is a strongly coherent risk measure if and only if we have*

$$\varrho(X) + \varrho(Y) = \sup \{\varrho(X \circ \sigma + Y \circ \tau) \,:\, \sigma, \tau \in \mathcal{A}\}. \tag{2.1}$$

Proof. Clearly $X \circ \sigma \sim X$ and $Y \circ \tau \sim Y$. Hence,

$$\sup \{\varrho(X \circ \sigma + Y \circ \tau) \,:\, \sigma, \tau \in \mathcal{A}\} \leq \sup \left\{\varrho(\tilde{X} + \tilde{Y}) \,:\, \tilde{X} \sim X,\ \tilde{Y} \sim Y\right\} \tag{2.2}$$

To prove the converse, take any $\varepsilon > 0$ and some $X' \sim X$ and $Y' \sim Y$ such that

$$\varrho(X' + Y') \geq \sup \left\{\varrho(\tilde{X} + \tilde{Y}) \,:\, \tilde{X} \sim X,\ \tilde{Y} \sim Y\right\} - \varepsilon$$

Consider the set $\{X \circ \sigma \,:\, \sigma \in \mathcal{A}\}$ and denote by C its closure in L^2. It is obviously transformation invariant. By the preceding Lemma, it is also law invariant. Because $X \in C$ and $X' \sim X$, we must have $X' \in C$, meaning that there exists a sequence $\sigma_n \in \mathcal{A}$ with $\| X \circ \sigma_n - X' \| \longrightarrow 0$. Similarly, there must exist a sequence $\tau_n \in \mathcal{A}$ with

$\|Y \circ \tau_n - Y'\| \longrightarrow 0$. Since ϱ is continuous, it follows that, for n large enough, we have

$$\begin{aligned}\sup\{\varrho(X \circ \sigma + Y \circ \tau) : \sigma, \tau \in \mathcal{A}\} &\geq \varrho(X \circ \sigma_n + Y \circ \tau_n) \geq \varrho(X' + Y') - \varepsilon \\ &\geq \sup\left\{\varrho(\tilde{X} + \tilde{Y}) : \tilde{X} \sim X,\ \tilde{Y} \sim Y\right\} - 2\varepsilon\end{aligned}$$

and because this holds for any $\varepsilon > 0$, the converse of (2.2) holds □

We can now state our main result.

THEOREM 2.10. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space which does not have atoms, and such that $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is separable. Let ϱ be a functional defined on L^2_d. Then the following propositions are equivalent:*

(i) ϱ is a strongly coherent risk measure;
(ii) ϱ is a maximal correlation risk measure.

Before we turn to the proof, note that this representation implies immediately that strongly coherent risk measures are in particular positive homogenous and translation invariant, as announced earlier.

Proof. We first show (i)⇒(ii). As the proof is quite long, we will punctuate it with several lemmas. □

By the preceding proposition and law invariance, it is enough to prove that

$$\varrho(X) + \varrho(Y) = \sup_{\sigma \in \mathcal{A}} \varrho(X + Y \circ \sigma). \tag{2.3}$$

Call ϱ^* the Legendre transform of ϱ in L^2_d.

LEMMA 2.11. *ϱ^* is law-invariant.*

Proof. For $\sigma \in \mathcal{A}$, one has $\varrho^*(X^* \circ \sigma) = \sup_{X \in L^2_d}\{\langle X^* \circ \sigma, X\rangle - \varrho(X)\}$, so $\varrho^*(X^* \circ \sigma) = \sup_{X \in L^2_d}\{\langle X^*, X \circ \sigma^{-1}\rangle - \varrho(X)\} = \sup_{X \in L^2_d}\{\langle X^*, X \circ \sigma^{-1}\rangle - \varrho(X \circ \sigma^{-1})\} = \varrho^*(X^*)$. □

LEMMA 2.12. *If the functions f_i, $i \in I$, are l.s.c. convex functions, then*

$$\left(\sup_i f_i\right)^* = \left(\inf_i f_i^*\right)^{**}.$$

Proof. For a given l.s.c. convex function f, $f \leq (\sup_i f_i)^*$ is equivalent to $f^* \geq \sup_i f_i$, hence to $f \geq f_i$ for all i, hence to $f^* \leq f_i^*$ for all i, hence $f \leq \inf_i f_i^*$, hence, as f is l.s.c. convex, to $f \leq (\inf_i f_i^*)^{**}$. □

Applying Lemma 2.12 to the structure neutrality equation, one has

$$\begin{aligned}\varrho^*(X^*) + \varrho^*(Y^*) &= \left(\inf_{\sigma\in\mathcal{A}} \sup_{X,Y} \left\{\langle X,\, X^*\rangle + \langle Y,\, Y^*\rangle - \varrho(X + Y\circ\sigma)\right\}\right)^{**} \\ &= \left(\inf_{\sigma\in\mathcal{A}} \sup_{Y} \left\{\langle Y,\, Y^*\rangle + \varrho^*(X^*) - \langle Y\circ\sigma,\, X^*\rangle\right\}\right)^{**} \\ &= \left(\varrho^*(X^*) + \inf_{\sigma\in\mathcal{A}} \sup_{Y} \langle Y,\, Y^* - X^*\circ\sigma^{-1}\rangle\right)^{**}.\end{aligned}$$

The term in $\sup_Y(\ldots)$ on the right-hand side is 0 if $Y^* = X^*\circ\sigma^{-1}$ and $+\infty$ otherwise. Hence the previous formula becomes

$$\varrho^*(X^*) + \varrho^*(Y^*) = \varphi^{**}(X^*,\, Y^*), \tag{2.4}$$

where we have defined

$$\varphi(X^*,\, Y^*) = \begin{cases} \varrho^*(X^*) & \text{if } X^* \sim Y^* \\ +\infty & \text{otherwise.}\end{cases} \tag{2.5}$$

Now suppose $\varphi(X^*,\, Y^*) < \infty$, hence that $\varrho^*(X^*) = \varrho^*(Y^*) < \infty$ and $X^* \sim Y^*$. As $\varphi \geq \varphi^{**}$, it follows that $\varrho^*(X^*) \geq \varrho^*(X^*) + \varrho^*(Y^*)$ hence $\varrho^*(Y^*) = \varrho^*(X^*) \leq 0$, and $\varphi(X^*,\, Y^*) \leq 0$.

Suppose $\varphi(X^*,\, Y^*) < \infty$ and $\varphi(X^*,\, Y^*) - \varphi^{**}(X^*,\, Y^*) < \varepsilon$. Replacing in (2.4), one finds that

$$0 \leq -\varrho^*(X^*) = -\varrho^*(Y^*) \leq \varepsilon.$$

LEMMA 2.13. *φ^{**} is valued into $\{0, +\infty\}$.*

Proof. As $\varphi^* = \varphi^{***}$, one has

$$\begin{aligned}\varphi^*(X,\, Y) &= \sup_{(X^*,\, Y^*)} \{\langle X,\, X^*\rangle + \langle Y,\, Y^*\rangle - \varphi^{**}(X^*,\, Y^*)\} \\ &= \sup_{(X^*,\, Y^*)} \{\langle X,\, X^*\rangle + \langle Y,\, Y^*\rangle - \varphi(X^*,\, Y^*)\}.\end{aligned}$$

Taking a maximizing sequence $(X_n^*,\, Y_n^*)$ in the latter expression, one has necessarily $\varphi(X_n^*,\, Y_n^*) - \varphi^{**}(X_n^*,\, Y_n^*) \longrightarrow 0$. From the previous remark, $\varrho^*(X_n^*) = \varrho^*(Y_n^*) \longrightarrow 0$, hence $\varphi(X_n^*,\, Y_n^*) \longrightarrow 0$. Therefore,

$$\varphi^*(X,\, Y) = \sup_{(X^*,\, Y^*):\varphi(X^*,\, Y^*)=0} \left\{\langle X,\, X^*\rangle + \langle Y,\, Y^*\rangle\right\}$$

which is clearly positively homogeneous of degree 1. Its Legendre transform φ^{**} can therefore only take values 0 and $+\infty$. □

Therefore, there is a closed convex set K such that φ^{**} is the indicator function of K, that is

$$\varphi^{**}(X^*,\, Y^*) = \begin{cases} 0 & \text{if } (X^*,\, Y^*) \in K \\ +\infty & \text{otherwise}\end{cases} \tag{2.6}$$

and condition (2.4) implies that

$$\varrho^*(X^*) + \varrho^*(Y^*) = \begin{matrix} 0 & \text{if } (X^*,\, Y^*) \in K. \\ +\infty & \text{otherwise} \end{matrix} \tag{2.7}$$

Note that if $\varrho^*(X^*) < \infty$, then $\varphi(X^*,\, Y^*) = \varrho^*(X^*)$ for all $Y^* \sim X^*$, and then $\varphi^{**}(X^*,\, Y^*) \leq \varphi(X^*,\, Y^*) < \infty$. This implies that $\varphi^{**}(X^*,\, Y^*) = 0$, hence that $\varrho^*(X^*) = 0$. Therefore, ϱ^* is also an indicator function: there exists a closed convex set C such that

$$\varrho^*(X^*) = \begin{matrix} 0 \text{ if } X^* \in C \\ +\infty \text{ otherwise.} \end{matrix} \tag{2.8}$$

By comparison of (2.7) and (2.8), one finds that

$$K = C \times C.$$

By duality, (2.8) becomes

$$\begin{aligned} \varrho\,(X) &= \sup_{X^* \in C} \langle X^*,\, X \rangle \\ C &= \left\{ X^* \mid \varrho^*(X^*) = 0 \right\}. \end{aligned} \tag{2.9}$$

Condition (2.5) then implies that φ is an indicator function: there exists a set K_0 (in general, neither a closed nor a convex set) such that

$$\varphi(X^*,\, Y^*) = \begin{matrix} 0 \text{ if } (X^*,\, Y^*) \in K_0 \\ +\infty \text{ otherwise.} \end{matrix}$$

By comparison with formulas (2.5) and (2.6), one finds that

$$\begin{aligned} (X^*,\, Y^*) \in K_0 &\Longleftrightarrow X^* \in C,\ Y^* \in C \text{ and } X^* \sim Y^* \\ K &= \overline{co}\ K_0 \end{aligned}$$

LEMMA 2.14. *Denote by $\mathcal{E}(C)$ the set of strongly exposed points of C, and $\overline{K_0}$ the closure of K_0 for the norm topology in $L^2 \times L^2$. Then*

$$\mathcal{E}(C) \times \mathcal{E}(C) \subset \overline{K_0}.$$

Proof. Recall (cf. Fabian et al. 2001) that X^* is *strongly exposed* in C if there is a continuous linear form X such that any maximizing sequence for X in C converges strongly to X^*:

$$\left.\begin{matrix} X_n^* \in C \\ \langle X,\, X_n^* \rangle \longrightarrow \sup_C \langle X,\, X^* \rangle \end{matrix}\right\} \Longrightarrow \left\| X_n^* - X^* \right\| \longrightarrow 0.$$

For $\varepsilon > 0$, denote by $T_C(X, \varepsilon)$ the set of $Y^* \in C$ such that $\sup_{Z^* \in C} \langle X,\, Z^* \rangle - \langle X,\, Y^* \rangle \leq \varepsilon$. Then $X^* \in C$ is strongly exposed by X if and only if $\sup_{Z^* \in C} \langle X,\, Z^* \rangle = \langle X,\, X^* \rangle$ and

$\delta[T_C(X, \varepsilon)]$ tends to 0 when $\varepsilon \longrightarrow 0$, where δ denotes the diameter,

$$\delta\left[T_C(X, \varepsilon)\right] := \sup\left\{\left\| X_1^* - X_2^* \right\| \mid X_1^* \in T_C(X, \varepsilon),\ X_2^* \in T_C(X, \varepsilon)\right\}.$$

Going back to the problem, it is clear that if X^* and Y^* are strongly exposed in C, then (X^*, Y^*) is strongly exposed in $C \times C$:

$$\mathcal{E}(C) \times \mathcal{E}(C) \subset \mathcal{E}(C \times C) = \mathcal{E}(K).$$

We claim that every strongly exposed point of K necessarily belongs to $\overline{K_0}$ (the closure is still the norm closure). Indeed, suppose there exists $(X_1^*, X_2^*) \in \mathcal{E}(K)$ such that $(X_1^*, X_2^*) \notin \overline{K_0}$. Then there exists $\varepsilon > 0$ such that $K_0 \cap B((X_1^*, X_2^*), \varepsilon) = \varnothing$, where $B((X_1^*, X_2^*), \varepsilon)$ is the ball of center (X_1^*, X_2^*) and radius $\varepsilon > 0$. As (X_1^*, X_2^*) is strongly exposed, there exists a linear form (X_1, X_2) strongly exposing it, and one can choose $\eta > 0$ small enough to ensure $\delta[T_K((X_1, X_2), \eta)] < \varepsilon$. Since $T_K((X_1, X_2), \eta)$ contains (X_1^*, X_2^*), one concludes that $K_0 \cap T_K((X_1, X_2), \eta) = \varnothing$, thus

$$K_0 \subset \left\{(Y_1^*, Y_2^*) \in K \mid \langle X_1, Y_1^* \rangle + \langle X_2, Y_2^* \rangle \geq \langle X_1, X_1^* \rangle + \langle X_2, X_2^* \rangle + \eta\right\}.$$

But the right-hand side is a closed convex set, so by taking the closed convex hull of the left-hand side, one gets

$$\overline{co}(K_0) \subset \left\{(Y_1^*, Y_2^*) \in K \mid \langle X_1, Y_1^* \rangle + \langle X_2, Y_2^* \rangle \geq \langle X_1, X_1^* \rangle + \langle X_2, X_2^* \rangle + \eta\right\}$$

and taking $(Y_1^*, Y_2^*) = \left(X_1^*, X_2^*\right) \in K$ leads to a contradiction.

Therefore, $\mathcal{E}(K) \subset \overline{K_0}$, and one has $\mathcal{E}(C) \times \mathcal{E}(C) \subset \mathcal{E}(K) \subset \overline{K_0}$. □

By a celebrated theorem of Bishop and Phelps (see again Fabian et al. 2001), there is a dense subset H of L^2 (in fact, a dense G_δ) such that, for every $X \in H$, the maximum of $\langle X^*, X \rangle$ for $X^* \in C$ is attained at a strongly exposed point. Going back to (2.9), take some $X \in H$, and let $X^* \in C$ be such that

$$\varrho(X) = \langle X^*, X \rangle$$

with $X^* \in \mathcal{E}(C)$. Now take another $Y \in H$, and another point $Y^* \in \mathcal{E}(C)$ such that $\varrho(Y) = \langle Y^*, Y \rangle$. One has $(X^*, Y^*) \in \mathcal{E}(C) \times \mathcal{E}(C)$, and it results from the previous lemma that $(X^*, Y^*) \in \overline{K_0}$. This implies the existence of a sequence $(X_n^*, Y_n^*) \in K_0$ such that (X_n^*, Y_n^*) converges to (X^*, Y^*) in norm. By the definition of K_0, one should have $X_n^* \sim Y_n^*$, that is $Y_n^* = X_n^* \circ \sigma_n$ for $\sigma_n \in \mathcal{A}$. Hence, $\varrho(Y) = \langle Y^*, Y \rangle = \lim_n \langle Y_n^*, Y \rangle = \lim_n \langle X_n^* \circ \sigma_n, Y \rangle = \lim_n \langle X_n^*, Y \circ \sigma_n^{-1} \rangle$. But by the Cauchy-Schwartz inequality,

$$\left|\left\langle X_n^*, Y \circ \sigma_n^{-1}\right\rangle - \left\langle X^*, Y \circ \sigma_n^{-1}\right\rangle\right| \leq \left\| Y \circ \sigma_n^{-1} \right\|_2 \left\| X_n^* - X^* \right\|_2,$$

which tends to 0 as $\| Y \circ \sigma_n^{-1} \|_2 = \| Y \|_2$. Therefore, $\varrho(Y) = \lim_n \langle X_n^*, Y \circ \sigma_n^{-1} \rangle = \lim_n \langle X^*, Y \circ \sigma_n^{-1} \rangle \leq \sup_{\tilde{Y} \sim Y} \langle X^*, \tilde{Y} \rangle$. But one has also $\varrho(Y) = \sup_{Y^* \in C} \langle Y^*, Y \rangle \geq \sup_{\sigma \in \mathcal{A}} \langle X^* \circ \sigma, Y \rangle = \sup_{\sigma \in \mathcal{A}} \langle X^*, Y \circ \sigma \rangle \geq \sup_{\tilde{Y} \sim Y} \langle X^*, \tilde{Y} \rangle$, therefore

$$\varrho(Y) = \sup_{\tilde{Y} \sim Y} \langle X^*, \tilde{Y} \rangle \quad \forall Y \in \Omega$$

The functions $\rho(Y)$ and

$$\sup_{\tilde{Y} \sim Y} \langle X^*, \tilde{Y} \rangle = \sup_{\tilde{X} \sim Y} \langle \tilde{X}^*, Y \rangle$$

are both convex, finite and lsc on L^2, and hence continuous. Since they coincide on a dense subset, they coincide everywhere. This proves the direct implication (i)⇒(ii) of the theorem.

We now turn to the converse. Let ϱ_μ be a maximal correlation risk measure with respect to baseline measure μ. Then ϱ_μ is clearly convex. Take X and Y in L^2_d. By Proposition B.1 in the Appendix, there exist two convex functions ϕ_1 and ϕ_2 such that for $U \sim \mu$, one has $\nabla\phi_1(U) \sim X$ and $\nabla\phi_2(U) \sim Y$, and $\varrho_\mu(X) = E[U \cdot \nabla\phi_1(U)]$, $\varrho_\mu(Y) = E[U \cdot \nabla\phi_2(U)]$. Thus $\varrho_\mu(X) + \varrho_\mu(Y) = E[U \cdot (\nabla\phi_1(U) + \nabla\phi_2(U))]$, but for all $\tilde{U} \sim U$, $E[U \cdot (\nabla\phi_1(U) + \nabla\phi_2(U))] \geq E[\tilde{U} \cdot (\nabla\phi_1(U) + \nabla\phi_2(U))]$, hence $\varrho_\mu(X) + \varrho_\mu(Y) = \sup\{\varrho(\tilde{X} + \tilde{Y}) : \tilde{X} \sim X, \tilde{Y} \sim Y\}$. Thus, ϱ_μ is strongly coherent, which completes the proof of Theorem 2.10. □

3. A MULTIVARIATE GENERALIZATION OF KUSUOKA'S THEOREM

In this section, we recall the existing axiomatization leading to the representation result of Kusuoka (2001), where risk measures for univariate risks that are subadditive, law invariant and comonotonic additive are represented by maximal correlation functionals. We then propose a way to generalize these axioms to the case where risk measures deal with multivariate risks, by showing how to generalize the only problematic axiom, namely comonotonic additivity. We then give a representation result which extends Kusuoka's to the multivariate case.

3.1. Coherent and Regular Risk Measures

To describe the existing axiomatic framework, we first recall the following definitions valid in the univariate case, from Artzner et al. (1999), and existing results.

DEFINITION 3.1 (Coherent; Convex risk measures). A functional $\varrho : L^2 \to \mathbb{R}$ is called a coherent risk measure if it satisfies the following four properties (MON), (TI), (CO), and (PH) as follows:

(i) Monotonicity (MON): $X \leq Y \Rightarrow \varrho(X) \leq \varrho(Y)$.
(ii) Translation invariance (TI): $\varrho(X + m) = \varrho(X) + m\varrho(1)$.
(iii) Convexity (CO): $\varrho(\lambda X + (1 - \lambda)Y) \leq \lambda\varrho(X) + (1 - \lambda)\varrho(Y)$ for all $\lambda \in [0, 1]$.
(iv) Positive homogeneity (PH): $\varrho(\lambda X) = \lambda\varrho(X)$ for all $\lambda \geq 0$.

A functional which only satisfies (MON), (TI), and (CO) is called a *convex risk measure*.

Even though these definitions are mostly standard, note that since we have considered risk measures associated with random vectors of potential *losses*, the definition of monotonicity takes an nondecreasing form, unlike the definition in most of the literature on coherent risk measures. Compared to the traditional presentation in the literature, the expression of translation invariance is adapted to take into account the fact that we did not impose the scaling convention $\varrho(1) = 1$. Also note (as we have a multivariate generalization in mind) that, let alone monotonicity (which we shall discuss separately below), all these axioms admit a straightforward generalization to the case of risks $X \in L^2_d$. The expression for (CO) and (PH) will remain unchanged; for (TI) the natural extension to dimension d will be given in (3.2).

A representation of coherent risk measures was given in the original work of Artzner et al. (1999), whereas representation of convex risk measures was proposed in Föllmer and Schied (2004). These were extended to the multivariate setting by Jouini et al. (2004) who characterize coherent acceptance sets, that is, sets in $\mathbb{R}^n$ that cancel the risk associated with an $\mathbb{R}^d$ valued random vector, and consider aggregation issues, and Burgert and Rüschendorf (2006) who characterize convex real valued measures for multivariate risks, and Rüschendorf (2006), who characterizes those of the latter that are law invariant, and proposes maximal correlation risk measures as an example. The idea of introducing a variational characterization of comonotonic additivity as well as the generalization of Kusuoka's axiomatic approach it allows constitute the essential novelties of this section.

3.1.1. Regularity. In the case of univariate risks, *comonotonic additivity* is used in addition to law invariance to define *regular risk measures* (see Föllmer and Schied 2004, section 4.7):

DEFINITION 3.2 (Comonotonicity; Regularity). Two random variables X and Y are comonotonic (or synonymously, comonotone) if there exits a random variable U and two increasing functions ϕ and ψ such that $X = \phi(U)$ and $Y = \psi(U)$ hold almost surely.

A functional $\varrho : L^2 \to \mathbb{R}$ is called a regular risk measure if it satisfies:

- Law invariance (LI), and
- Comonotonic additivity (CA): $\varrho(X + Y) = \varrho(X) + \varrho(Y)$ when X, Y are comonotonic.

Note that comonotonic additivity implies translation invariance, as any random variable is comonotonic with the constant. Informally speaking, law invariance suggests that the risk measure is a functional of the quantile function $F_X^{-1}(t) = \inf\{x : F_X(x) \geq t\}$ associated with the distribution. Positive homogeneity and comonotonic additivity together suggest that this representation is linear $\varrho(X) := \int_0^1 \phi(t) F_X^{-1}(t)\, dt$. Finally, subadditivity suggests that the weights $\phi(t)$ are increasing with respect to t. Precisely, Kusuoka has shown the following in Kusuoka (2001), Theorem 7.

PROPOSITION 3.3 (Kusuoka). *A coherent risk measure ϱ is regular if and only if for some increasing and nonnegative function ϕ on* [0, 1], *we have*

$$\varrho(X) := \int_0^1 \phi(t) F_X^{-1}(t)\, dt,$$

where F_X denotes the cumulative distribution functions of the random variable X, and its generalized inverse $F_X^{-1}(t) = \inf\{x : F_X(x) \geq t\}$ is the associated quantile function.

3.1.2. Variational Characterization. By the Hardy–Littlewood–Pólya inequality shown in lemma 11 of Kusuoka (2001), we can write a variational expression for coherent regular risk measures

$$\int_0^1 \phi(t) F_X^{-1}(t)\, dt = \max\left\{\mathbb{E}[X\tilde{U}] : \ \tilde{U} \sim \mu\right\}, \tag{3.1}$$

where μ if the probability distribution of ϕ, and the maximum is taken over the equidistribution class of μ. The reader is referred to Dana (2005) and Cheridito and Li (2009)

for a nice treatment of this variational problem and the dual representation of Schur convex functions in the univariate case. As we shall see, variational characterization 3.1 will be key when generalizing to the multivariate setting.

3.2. A Multivariate Notion of Comonotonicity

We now turn to an extension of the concept of comonotonicity. Note first that a valid definition of comonotonicity in dimension one is the following: two random variables X and Y are comonotonic if and only if one can construct almost surely $Y = T_Y(U)$ and $X = T_X(U)$ for some third random variable U, and T_X, T_Y nondecreasing functions. In other words, X and Y are comonotonic whenever there is a random variable U such that $\mathbb{E}[UX] = \max\{\mathbb{E}[X\tilde{U}] : \tilde{U} \sim U\}$ and $\mathbb{E}[UY] = \max\{\mathbb{E}[Y\tilde{U}] : \tilde{U} \sim U\}$. This variational characterization will be the basis for our generalized notion of comonotonicity.

To simplify our exposition in the remainder of the paper, we shall make the following assumption.

ASSUMPTION. *In the remainder of the paper, we shall assume that the baseline distribution of risk μ is absolutely continuous with respect to Lebesgue measure.*

DEFINITION 3.4 (μ-comonotonicity). Let μ be a probability measure on $\mathbb{R}^d$ that is absolutely continuous. Two random vectors X and Y in L^2_d are called μ-comonotonic if for some random vector $U \sim \mu$, we have

$$U \in argmax_{\tilde{U}} \left\{\mathbb{E}[X \cdot \tilde{U}],\ \tilde{U} \sim \mu\right\} \text{ and}$$
$$U \in argmax_{\tilde{U}} \left\{\mathbb{E}[Y \cdot \tilde{U}],\ \tilde{U} \sim \mu\right\}.$$

In particular, every random vector X is μ-comonotonic with constant vectors $Y = y$. Note that the geometric interpretation of this definition is that X and Y are μ-comonotonic if and only if they have the same L^2 projection on the equidistribution class of μ. We next give a few useful lemmas. We start with a result securing the existence of a μ -comonotonic pair with given marginals.

LEMMA 3.5. *Let μ be a probability measure on $\mathbb{R}^d$ that is absolutely continuous. Then given two probability distributions P and Q in $\mathcal{P}_2(\mathbb{R}^d)$, there exists a pair (X, Y) in $(L^2_d)^2$ such that $X \sim P$, $Y \sim Q$, and X and Y are μ-comonotonic.*

Proof. By Brenier's theorem (Proposition B.1 in the Appendix), there exists $U \sim \mu$ and two convex functions ϕ_1 and ϕ_2 such that $X = \nabla\phi_1(U) \sim P$ and $Y = \nabla\phi_2(U) \sim Q$. Then X and Y are μ-comonotonic. □

We then provide a useful characterization of μ-comonotonicity.

LEMMA 3.6. *Let μ be probability measure on $\mathbb{R}^d$ that is absolutely continuous. Then two random vectors X and Y in L^2_d are μ-comonotonic if*

$$\varrho_\mu(X + Y) = \varrho_\mu(X) + \varrho_\mu(Y),$$

where $\varrho_\mu(X) := \sup\{\mathbb{E}[X \cdot \tilde{U}] : \tilde{U} \sim \mu\}$ is the maximal correlation risk measure, defined in Definition 2.3.

Proof. There exists $U \sim \mu$ such that $\varrho_\mu(X+Y) = \mathbb{E}[(X+Y)\cdot U]$. We have $\mathbb{E}[(X+Y)\cdot U] = \mathbb{E}[X\cdot U] + \mathbb{E}[Y\cdot U]$, and both inequalities $\mathbb{E}[X\cdot U] \le \varrho_\mu(X)$ and $\mathbb{E}[Y\cdot U] \le \varrho_\mu(Y)$ hold, thus $\mathbb{E}[X\cdot U] + \mathbb{E}[Y\cdot U] \le \varrho_\mu(X) + \varrho_\mu(Y)$ with equality if and only both inequalities above are actually equalities, which is the equivalence needed. □

This lemma implies in particular that maximal correlation functionals with baseline measure μ are μ-comonotone additive. Thus, combining with Theorem 2.10, this establishes that strongly coherent risk measures are μ-comonotone additive for some μ.

We next show that *in dimension 1*, the notion of μ-comonotonicity is equivalent to the classical notion of comonotonicity, regardless of the choice of μ (provided it is absolutely continuous).

LEMMA 3.7. *In dimension $d = 1$, let μ be probability measure on $\mathbb{R}^d$ that is absolutely continuous. Then X and Y are μ-comonotonic if and only if they are comonotonic in the classical sense, that is, if and only if there exists a random variable Z and two nondecreasing functions f and g such that $X = f(Z)$ and $Y = g(Z)$ holds almost surely.*

Proof. Suppose that X and Y are μ-comonotonic. Then there is a $U \sim \mu$ such that $U \in \mathit{argmax}_{\tilde{U}}\{\mathbb{E}[X\tilde{U}],\ \tilde{U} \sim \mu\}$ and $U \in \mathit{argmax}_{\tilde{U}}\{\mathbb{E}[Y\tilde{U}],\ \tilde{U} \sim \mu\}$. This implies in particular the existence of two increasing functions f and g such that $X = f(U)$ and $Y = g(U)$ holds almost surely. Hence X and Y are comonotonic in the classical sense. Conversely, suppose that X and Y are comonotonic in the classical sense. There exists a random variable Z and two increasing functions f and g such that $X = f(Z)$ and $Y = g(Z)$ holds almost surely. Let F_Z be the cumulative distribution function of Z, and F_μ the one associated with μ. Defining $U = F_\mu^{-1}(F_Z(Z))$, one has $U \sim \mu$, and denoting $\varphi = f \circ F_\mu^{-1} \circ F_Z$ and $\phi = g \circ F_\mu^{-1} \circ F_Z$, one has $X = \varphi(U)$ and $Y = \phi(U)$. Thus, X and Y are μ-comonotonic. □

In dimension one, one recovers the classical notion of comonotonicity regardless of the choice of μ as shown in the previous lemma. However, in dimension greater than one, the comonotonicity relation crucially depends on the baseline distribution μ, unlike in dimension one. The following lemma makes this precise.

LEMMA 3.8. *Let μ and ν be probability measures on $\mathbb{R}^d$ that is absolutely continuous. Then*

1. *In dimension $d = 1$, μ-comonotonicity always implies ν-comonotonicity.*
2. *In dimension $d \ge 2$, μ-comonotonicity implies ν-comonotonicity if and only if $\nu = T\#\mu$ for some location-scale transform $T(u) = \lambda u + u_0$ where $\lambda > 0$ and $u_0 \in \mathbb{R}^d$. In other words, comonotonicity is an invariant of the location-scale family transformation classes.*

Proof. In dimension one, all the notions of μ-comonotonicity coincide with the classical notion of comonotonicity, as remarked above. Let $d \ge 2$, and suppose that μ-comonotonicity implies ν-comonotonicity. Consider $U \sim \mu$, and let ϕ be the convex function (defined up to an additive constant) such that $\nabla\phi\#\nu = \mu$. Then there exists a random vector $V \sim \nu$ such that $U = \nabla\phi(V)$ almost surely. Consider some arbitrary symmetric positive endomorphism Σ acting on $\mathbb{R}^d$. Then the map $u \to \Sigma(u)$ is the gradient of a convex function (namely the associated quadratic form $u \to \frac{1}{2} < u, \Sigma(u) >$),

therefore the random vectors U and $\Sigma(U)$ are μ -comonotonic. By hypothesis, it follows that U and $\Sigma(U)$ are also ν-comonotonic, hence there exists a convex function ζ such that $\Sigma(U) = \nabla\zeta(V)$ holds almost surely. Therefore, the equality $\Sigma \circ \nabla\phi(v) = \nabla\zeta(v)$ holds for almost every v. By differentiating twice (which can be done almost everywhere, by Aleksandrov's theorem), we get that $\Sigma \circ D^2\phi(v) = D^2\zeta(v)$ hence $\Sigma \circ D^2\phi$ is almost everywhere a symmetric endomorphism. This being true regardless of the choice of Σ, it follows that the matrix of $D^2\phi$ in any orthonormal basis of $\mathbb{R}^d$ is almost everywhere a diagonal matrix, hence there exists a real valued map $\lambda(u)$ such that $D^2\phi(u) = \lambda(u)u$, with $\lambda(u) > 0$. But this implies $\partial_{u_i}\partial_{u_j}\phi(u) = 0$ for $i \neq j$ and $\partial^2_{u_i}\phi(u) = \lambda(u)$ for all i. Therefore, $\partial_{u_j}\lambda(u) = \partial_{u_j}\partial^2_{u_i}\phi(u) = 0$. Hence $\lambda(u) = \lambda$ a strictly positive constant. It follows that $\nabla\phi(u) = \lambda u + u_0$. The converse holds trivially. □

REMARK 3.9. A close inspection of the proof of this lemma reveals that the essential reason of the discrepancy between dimension one and higher is the simple fact that the general linear matrix group $\mathcal{G}l_d(\mathbb{R})$ is Abelian if and only if $d = 1$.

We can now define a concept which generalizes comonotonic additivity to the multidimensional setting.

DEFINITION 3.10 (μ-comonotonic additivity; μ-regularity). A functional $\varrho : L^2_d \to \mathbb{R}$ is called a μ-regular risk measure if it satisfies

(i) Law invariance (LI), and
(ii) μ-comonotonic additivity (μ-CA): $\varrho(X + Y) = \varrho(X) + \varrho(Y)$ when X, Y are μ-comonotonic.

As every random vector is comonotonic with constants, it implies that a μ-comonotonic additive functional ϱ is in particular translation invariant in the following multivariate sense

$$\varrho(X + my) = \varrho(X) + m\varrho(y) \text{ for all } m \in \mathbb{R} \text{ and } y \in \mathbb{R}^d. \tag{3.2}$$

3.3. A Multivariate Extension of Kusuoka's Theorem

We now show that maximal correlation is equivalent to the combination of subadditivity, law invariance, μ-comonotonic additivity and positive homogeneity. Further, the probability measure μ involved in the definition of comonotonic additivity shall be precisely related to the one which is taken as a baseline scenario of the maximal correlation measure.

We have seen earlier (Lemma 3.6) that maximal correlation risk measures defined with respect to a distribution μ are μ-comonotonic additive. When the measure is also law invariant and coherent, we shall see that the converse holds true, and this constitutes our second main result, which is a multivariate extension of Kusuoka's theorem. Note that while Kusuoka's theorem was stated using the axioms of subadditivity and positive homogeneity in addition to others, we only need the weaker axiom of convexity in addition to the same others.

THEOREM 3.11. *Let ϱ be a l.s.c. risk measure on L^2_d with the properties of convexity (CO), and μ-regularity, that is law invariance (LI) and μ-comonotonic additivity (μ-CA). Then ϱ is strongly coherent. Equivalently, ϱ is a maximal correlation risk measure, namely there exists $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ such that $\varrho = \varrho_\nu$, where ϱ_ν is a maximal correlation measure with respect to baseline scenario ν, and μ and ν are related by location-scale transformation, that is $\nu = T\#\mu$ where $T(u) = \lambda u + u_0$ with $\lambda > 0$ and $u_0 \in \mathbb{R}^d$.*

Proof. Combining the convexity and law invariance axioms imply $\varrho(\tilde{X} + \tilde{Y}) \leq \varrho(X) + \varrho(Y)$ for all X, Y, $\tilde{X}$, $\tilde{Y}$ in L^2_d, thus $\varrho(X) + \varrho(Y) \geq \sup\{\varrho(\tilde{X} + \tilde{Y}) : \tilde{X} \sim X;\ \tilde{Y} \sim Y\}$. But by Lemma 3.5, there exists a μ-comonotonic pair (X, Y). By μ-comonotonic additivity, one has $\varrho(X) + \varrho(Y) = \varrho(X + Y)$, therefore the previous inequality is actually an equality, and

$$\varrho(X) + \varrho(Y) = \sup\left\{\varrho(\tilde{X} + \tilde{Y}) : \tilde{X} \sim X;\ \tilde{Y} \sim Y\right\}$$

therefore ϱ is strongly coherent. By Theorem 2.10, it results that there exists $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ such that $\varrho = \varrho_\nu$. But by the comonotonic additivity of ϱ and lemma 3.6, any two vectors X and Y which are μ-comonotonic are also ν-comonotonic. By lemma 3.8, this implies that there is a location-scale map T such that $\nu = T\#\mu$, so that the result follows. □

Because it allows a natural generalization of well-known univariate results, this theorem makes a strong point in arguing that our notion of comonotonic additivity is the right one when considering multivariate risks.

3.4. Extending Monotonicity

We extend the concept of monotonicity with reference to a partial order $\preceq$ defined on $\mathbb{R}^d$ in the following way.

DEFINITION 3.12 ($\preceq$-monotonicity). A functional $\varrho : L^2 \to \mathbb{R}$ is said to be $\preceq$-monotone if it satisfies

$$(\preceq\text{-}MON)\text{: } X \preceq Y \text{ almost surely} \Rightarrow \varrho(X) \leq \varrho(Y).$$

We have the following result.

PROPOSITION *3.13. Let ϱ_μ be the maximal correlation risk measure with respect to baseline distribution μ. Let $(Supp\,\mu)^0$ be the polar cone of the support of μ. For a cone $C \subset \mathbb{R}^d$, denote $\preceq_C$ the partial order in $\mathbb{R}^d$ induced by C, namely $x \preceq_C y$ if and only if $y - x \in C$. Then ϱ_μ is monotone with respect to $\preceq_C$ if and only if $C \subset -(Supp\,\mu)^0$.*

Proof. If X and U are μ-comonotonic, then $D\varrho_X(Z) = E[U \cdot Z]$, but the property that $E[U \cdot Z] \geq 0$ for all Z almost surely included in C is equivalent to $C \subset -(Supp\,\mu)^0$. □

Note that in dimension $d = 1$, with $C = \mathbb{R}_+$, one recovers the usual notion of monotonicity. In higher dimension, we get in particular that if μ is supported in $\mathbb{R}^d_+$, then ϱ_μ is monotone with respect to the strong order of $\mathbb{R}^d$. Finally, note also that the concept of monotonicity proposed here is a somewhat weak one, as it deals only with almost

sure domination between X and Y. A stronger concept of monotonicity would involve stochastic ordering of X and Y; we do not pursue this approach here.

4. NUMERICAL COMPUTATION

In this section, we show explicit examples of computation of the maximal correlation risk measure. We start by the Gaussian case, where closed-form formulas are available. To handle more general cases we shall show that the problem may be thought of as an auction mechanism, an intuition we shall develop and use to derive an efficient computational algorithm.

4.1. Gaussian Risks

We now consider the case where the baseline risk U is Gaussian with distribution $\mu = N(0, \Sigma_U)$, with Σ_U a positive definite matrix of size d, and we study the restriction of ϱ_μ to the class of Gaussian risks.

Note (cf. Rachev and Rüschendorf 1998 I, Ex. 3.2.12) that the linear map $u \to A_X u$ where

$$A_X = \Sigma_U^{-1/2}\left(\Sigma_U^{1/2}\Sigma_X\Sigma_U^{1/2}\right)^{1/2}\Sigma_U^{-1/2},$$

sends the probability measure $N(0, \Sigma_U)$ to the probability measure $N(0, \Sigma_X)$; further, A_X is positive semidefinite, so this map is the gradient of convex function $u \to \frac{1}{2}u' A_X u$. Hence, we have the following straightforward matrix formulation of comonotonicity.

LEMMA 4.1. *Consider two Gaussian vectors $X \sim N(0, \Sigma_X)$ and $Y \sim N(0, \Sigma_Y)$ with Σ_X and Σ_Y invertible. Then X and Y are μ-comonotonic if and only if*

$$E[XY^T] = \Sigma_U^{-1/2}\left(\Sigma_U^{1/2}\Sigma_X\Sigma_U^{1/2}\right)^{1/2}\left(\Sigma_U^{1/2}\Sigma_Y\Sigma_U^{1/2}\right)^{1/2}\Sigma_U^{-1/2}. \tag{4.1}$$

In particular, in the case $\mu = N(0, I_d)$, X and Y are μ-comonotonic if and only if $E[XY^T] = \Sigma_X^{1/2}\Sigma_Y^{1/2}$.

Proof. If X and Y are μ-comonotonic, then there exists $U \sim N(0, \Sigma_U)$ such that $X = A_X U$ and $Y = A_Y U$, and the result follows. Conversely, if equality (4.1) holds, then denoting $U = A_X^{-1} X$ and $V = A_Y^{-1} Y$, we get that (1) $U \sim N(0, \Sigma_U)$ and $V \sim N(0, \Sigma_U)$, and (2) $E[UV^T] = A_X^{-1} E[XY^T] A_Y^{-1} = \Sigma_U$, therefore by the Cauchy–Schwartz inequality, $U = V$ almost surely. Thus, X and Y are μ-comonotonic. □

We now derive the value of correlation risk measures at Gaussian risks. Still by Rachev and Rüschendorf (1998) I, Ex. 3.2.12, we have immediately:

PROPOSITION *4.2. When the baseline risk U is Gaussian with distribution $\mu = N(0, \Sigma_U)$, we have for a Gaussian vector $X \sim N(0, \Sigma_X)$:*

$$\varrho_\mu(X) = tr\left[\left(\Sigma_U^{1/2}\Sigma_X\Sigma_U^{1/2}\right)^{1/2}\right].$$

In particular, in the case $\mu = N(0, I_d)$, ϱ_μ is the trace norm: $\varrho_\mu(X) = tr[\Sigma_X^{1/2}]$.

Proof. One has $\varrho_\mu(X) = \max\{\mathbb{E}[\tilde{X} \cdot U];\ \tilde{X} \sim X\} = \mathbb{E}[A_X U U^T]$, thus because of the previous results, $\varrho_\mu(X) = \mathbb{E}[U^T \Sigma_U^{-1/2} (\Sigma_U^{1/2} \Sigma_X \Sigma_U^{1/2})^{1/2} \Sigma_U^{-1/2} U] = \mathrm{tr}((\Sigma_U^{1/2} \Sigma_X \Sigma_U^{1/2})^{1/2})$. □

In dimension 2, we have the formula $tr(\sqrt{S}) = \sqrt{tr(S) + 2\sqrt{\det S}}$, so we get a closed form expression:

EXAMPLE 4.3. When $d = 2$, and $\mu = N(0, I_2)$, we have for $\Sigma_X = \begin{pmatrix} \sigma_1^2 & \varrho\sigma_1\sigma_2 \\ \varrho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$ the following expression $\varrho_\mu(X) = \sqrt{\sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2\sqrt{1 - \varrho^2}}$.

4.2. Kantorovich Duality and Walras Auction

We now see how optimal transportation duality permits the computation of maximal correlation risk measures. More precisely, we shall see that the algorithm we shall propose to compute numerically the maximal correlation risk measures is to be thought of intuitively as a *Walrasian auction*, as we shall explain. We refer to Rachev and Rüschendorf (1998) and Villani (2003) for overviews of the theory and applications of optimal transportation, including recent results. Consider a baseline distribution μ, and recall the expression for the maximal correlation risk measure $\varrho_\mu(X)$ of a random vector $X \in \mathbb{R}^d$: $\varrho(X) = \sup\{\mathbb{E}[X \cdot \tilde{U}] :\ \tilde{U} \sim \mu\}$. This problem is the problem of computing the maximal transportation cost of mass distribution μ to mass distribution $\mathcal{L}_X$ with cost of transportation $c(u, x) = u \cdot x$.

The problem has a dual expression according to Monge–Kantorovich duality (or duality of optimal transportation). We have (theorem 2.9 page 60 of Villani 2003):

$$\varrho_\mu(X) = \min_{V \in \mathrm{c.l.s.c.}(\mathbb{R}^d)} \left(\int V d\mu + \int V^* d\mathcal{L}_X \right). \tag{4.2}$$

The function V that achieves the minimum in (4.2) exists by theorem 2.10(iii) and when $\mathcal{L}_X$ is absolutely continuous, one has $\nabla V^*(X) \sim \mu$ and $\varrho_\mu(X) = \mathbb{E}[X \cdot \nabla V^*(X)]$. In the sequel we shall make the law invariance of ϱ_μ and the symmetry between the roles played by the distributions of X and U explicit in the notation by writing

$$\varrho_\mu(\mathcal{L}_X) := \varrho(\mu, \mathcal{L}_X) := \varrho_\mu(X).$$

4.2.1. Law-Invariant, Convex Risk Measures. Following Rüschendorf (2006), theorem 2.3, the maximum correlation risk measures are the building blocks of more general convex risk measures. One has the following result, which was proven by Rüschendorf in the cited paper.

PROPOSITION 4.4. *Let ϱ be a convex measure. Then ϱ is law invariant if and only if there exists a penalty function α such that*

$$\varrho(X) = \sup_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \varrho_\mu(X) - \alpha(\mu).$$

Furthermore, $\alpha(\mu)$ can be chosen as $\alpha(\mu) = \sup\{\varrho_\mu(X) :\ X \in L_d^2, \varrho(X) \geq 0\}$.

4.2.2. Dual Representations of the Risk Measure. The following lemma provides an expression of the conjugate of the maximal correlation risk measure.

LEMMA 4.5. *For $W : \mathbb{R}^d \to \mathbb{R}$ convex and lower semicontinuous, one has*

$$\sup_{P \in \mathcal{P}_2(\mathbb{R}^d)} \left\{ \varrho_\mu(P) + \int W dP \right\} = \int (-W)^* d\mu.$$

Proof. One has $\int (-W)^* d\mu = \int \sup_y \{u \cdot y + W(y)\} d\mu(u)$, thus $\int (-W)^* d\mu = \sup_{\tau(\cdot)} \int u \cdot \tau(u) + W(\tau(u)) d\mu(u)$ where the supremum is over all measurable maps $\tau : \mathbb{R} \to \mathbb{R}$. Grouping by equidistribution class, one has

$$\begin{aligned} \int (-W)^* d\mu &= \sup_P \left[\sup_{\tau \# \mu = P} \int u \cdot \tau(u) \, d\mu(u) + \int W dP \right] \\ &= \sup_P \left\{ \varrho_\mu(P) + \int W dP \right\}. \end{aligned}$$

□

4.2.3. General Equilibrium Interpretation. We now consider then $\varrho(\mu, \mathcal{L}_X)$ for two probability distributions on $\mathbb{R}^d$, and we interpret μ as a distribution of consumers (e.g., insurees) and $\mathcal{L}_X$ as a distribution of goods (e.g., insurance contracts) in an economy. Consumer with characteristics u derives utility from the consumption of good with attributes x equal to the interaction $u \cdot x$ of consumer characteristics and good attributes. Consumer x maximizes utility $u \cdot x$ of consuming good x minus the *price* $V^*(x)$ of the good. Hence his *indirect utility* is $\sup_{x \in \mathbb{R}^d} [u \cdot x - V^*(x)] = V^{**}(u) = V(u)$. According to equation (4.2), the *total surplus* in the economy $\mathbb{E}[X \cdot U]$ is maximized for the pair (V, V^*) of convex lower semi-continuous functions on $\mathbb{R}^d$ that minimizes

$$\Phi(V) := \int V d\mu + \int V^* d\mathcal{L}_X.$$

The functional Φ is convex and its Fréchet derivative, when it exists, is interpreted as the *excess supply* in the economy, with value at h equal to $D\Phi(h) = \int h \, d(\mu - \nu_V)$, where $\nu_V := \nabla V^* \# \mathcal{L}_X$. Indeed, the convexity of the map $V \to \Phi(V)$ follows from the identity established above in Lemma 4.5

$$\Phi(V) = \sup_{\nu \in \mathcal{P}_2(\mathbb{R})^d} \left\{ \varrho(\mathcal{L}_X, \nu) + \int V d(\mu - \nu) \right\},$$

thus this map is the supremum of functionals that are linear in V. The supremum is attained for $\nu = \nu_V$, hence it follows that $D\Phi_V(h) = \int h \, d(\mu - \nu_V)$.

Hence, excess supply is zero when the indirect utility V and the prices V^* are such that $\nu_V = \mu$. With our economic interpretation above, this can be seen as a *Walrasian welfare theorem*, where the total surplus is maximized by the set of prices that equates excess supply to zero.

This general, equilibrium interpretation of maximal correlation risk measures extends to the method of computation of the latter through a gradient algorithm to minimize the

convex functional Φ. This algorithm can be interpreted as a *Walrasian tâtonnement* that adjusts prices to reduce excess supply $D\Phi_V$. This algorithm is described in more detail and implemented fully in the case of discretely distributed risks later.

4.3. Discrete Risks

We now consider the restriction ϱ_μ to the class of risks whose distribution is discrete. We have in mind in particular the empirical distribution of a sample of recorded data of the realization of the risk. The procedure we shall now describe consists in the computation of the generalized quantile of the discrete distribution, which opens the way for econometric analysis of maximal correlation risk measures.

4.3.1. Representation. Let $X \sim P_n$, where $P_n = \sum_{k=1}^n \pi_k \delta_{Y_k}$ is a discrete distribution supported by $\{Y_1, \ldots, Y_n\}$, n distinct points in $\mathbb{R}^d$. For instance if P_n is the empirical measure of the sample $\{Y_1, \ldots, Y_n\}$, then $\pi_k = 1/n$. We are looking for $\varphi : [0, 1]^d \to \mathbb{R}^d$ such that

(i) for (almost) all $u \in \mathbb{R}^d$, $\varphi(u) \in \{Y_1, \ldots, Y_n\}$;
(ii) for all $k \in \{1, \ldots, n\}$, $\mu(\varphi^{-1}\{Y_k\}) = \pi_k$, that is, φ pushes forward μ to P_n;
(iii) $\varphi = \nabla V$, where $V : \mathbb{R}^d \to \mathbb{R}$ is a convex function.

It follows from the Monge–Kantorovich duality that there exist weights $(w_1, \ldots, w_n) \in \mathbb{R}^n$, such that $V(u) = w^*(u) : \max_k\{\langle u, Y_k\rangle - w_k\}$ is the solution. Introduce the functional $\Phi_\mu : \mathbb{R}^n \to \mathbb{R}$, $\Phi_\mu(w) = \int w^*(u) d\mu(u)$. The numerical implementation of the method is based on the following result.

PROPOSITION *4.6*. *There exist unique (up to an additive constant) weights $w_1, \ldots, w_n$ such that for $w^*(u) = \max_k\{\langle u, Y_k\rangle - w_k\}$, the gradient map $\varphi = \nabla w^*$ satisfies (i), (ii) and (iii) above. The function $w \to \Phi_\mu(w) + \sum_{k=1}^n \pi_k w_k$ is convex, and reaches its minimum at $w = (w_1, \ldots, w_n)$ defined earlier.*

Proof. By the Knott–Smith optimality criterion (theorem 2.12(i) page 66 of Villani 2003), there exists a convex function w on the set $\{Y_1, \ldots, Y_n\}$ such that the optimal pair in (4.2) is (w, V), where V is the Legendre–Fenchel conjugate of w, that is, the function $V(u) = \sup_{x \in \{Y_1, \ldots, Y_n\}}(u \cdot x - w(x)) = \max_k(u \cdot Y_k - w_k)$, where $w_k = w(Y_k)$ for each $k = 1, \ldots, n$. Note that the subdifferential ∂V is a singleton except at the boundaries of the sets $\mathcal{U}_k = \{u : \arg\max_i\{\langle u, Y_i\rangle - w_i\} = k\}$, so ∇V is defined $\mathcal{L}_U$ almost everywhere. Since for all k, and all $u \in \mathcal{U}_k$, $Y_k \in \partial V(u)$, ∇V satisfies (i). Finally, by Brenier's Theorem (theorem 2.12(ii) page 66 of Villani 2003), ∇V pushes $\mathcal{L}_U$ forward to P_n, hence it also satisfies (iii). The function Φ_μ: $w \to \int w^*(u) d\mu(u)$ is convex, which follows from the equality

$$\int w^*(u) d\mu(u) = \max_{\sigma(.)} \int \left\langle u, Y_{\sigma(u)} \right\rangle - w_{\sigma(u)}\, d\mu(u)$$

where the maximum is taken over all measurable functions $\sigma : \mathbb{R}^d \to \{1, \ldots, n\}$. □

4.3.2. The Tâtonnement Algorithm. The problem is therefore to minimize the convex function $w \to \Phi_{\mu,\pi}(w) = \Phi_\mu(w) + \sum_{k=1}^n \pi_k w_k$, which can be done using a gradient approach. To the best of our knowledge, the idea of using the Monge-Kantorovich duality to compute the weights using a gradient algorithm should be credited to Aurenhammer et al. (see Aurenhammer, Hoffmann, and Aronov (1998) and also Rüschendorf and Uckelmann (2000)). However, by the economic interpretation seen above, the algorithm's dynamics is the time-discretization of a "tâtonnement process," as first imagined by Léon Walras (1874) and formalized by Paul Samuelson (1947) (see Samuelson 1947). Hence to emphasize the economic interpretation, we shall refer to the algorithm as "Tâtonnement Algorithm."

4.3.3. The Algorithm. Initialize the prices $w^0 = 0$. At each step m, compute $\Phi_{\mu,\pi}(w^m)$ and the excess demand $\nabla\Phi_{\mu,\pi}(w^m)$. For a well chosen elasticity parameter ϵ^m, update the prices proportionally to excess demand

$$w^{m+1} = w^m + \epsilon^m \nabla\Phi_{\mu,\pi}(w^m).$$

Go to next step, or terminate the algorithm when the excess demand becomes smaller than a prescribed level.

This algorithm requires the evaluation of the function and its gradient. For this we shall need to compute in turns, for each k (1) $\mathcal{U}_k = \{u : \arg\max_i \{\langle u, Y_i\rangle - w_i\} = k\}$; (2) $p_k = \mu(\mathcal{U}_k)$; and (3) u_k the barycenter of $(\mathcal{U}_k, \mu)$ (that is $u_k = \mu(\mathcal{U}_k)^{-1} \int_{\mathcal{U}_k} z d\mu(z)$.) Then we get the value of $\Phi_{\mu,\pi}(w)$: $\Phi_{\mu,\pi}(w) = \sum(\langle u_k, Y_k\rangle - w_k)p_k + w_k\pi_k$ and the value of its gradient $\nabla\Phi_{\mu,\pi}(w) = \pi - p$, that is, $\frac{\partial\Phi_{\mu,\pi}(w)}{\partial w_k} = \pi_k - p_k$. We have implemented these calculations in Matlab using a modified versions of the publicly available Multi-Parametric Toolbox (MPT).[2] All the programs are available upon request.

5. CONCLUSION

In comparison with existing literature on the topic on multidimensional risk exposures, this work proposes a multivariate extension of the notion of comonotonicity, which involves simultaneous optimal rearrangements of two vectors of risk. With this extension, we are able to generalize Kusuoka's result and characterize subadditive, comonotonic additive and law invariant risk measures by maximal correlation functionals, which we show can be conveniently computed using optimal transportation methods. We also show that the properties of law invariance, subadditivity and comonotonic additivity can be summarized by an equivalent property, that we call *strong coherence*, and that we argue has a more natural economic interpretation. Further, we believe that this paper illustrates the enormous potential of the theory of optimal transportation in multivariate analysis and higher dimensional probabilities. We do not doubt that this theory will be included in the standard probabilistic toolbox in a near future.

APPENDIX A: ILLUSTRATIONS

The Tâtonnement algorithm was implemented with the use of the Multi-Parametric Toolbox, and we derived the general quantile ∇V that achieves the optimal transportation

[2] MPT is available online at http://control.ee.ethz.ch/ mpt/.

of the uniform distribution on the unit cube in $\mathbb{R}^d$ and the empirical distribution of a sample of uniformly distributed random vectors in the unit cube in $\mathbb{R}^d$. The following illustrations show the Monge–Kantorovitch potential V, also interpreted as the buyer's indirect utility in our general equilibrium interpretation in the case of samples of sizes 7 and 27, respectively. The potential V is piecewise affine, and the algorithm also requires to determine the regions over which it is affine, and their volume and center of mass. The corresponding partition is given opposite each potential plot. For illustration purposes, the dimension of the space d is taken equal to 2, but the generalized quantiles and corresponding partitions can be derived in higher dimensions.

APPENDIX B: RESULTS ON OPTIMAL TRANSPORTATION

In this appendix, we recall basic results in Optimal Transportation theory. Roughly put, this theory characterizes the properties of the couplings of two random variables which achieve maximal correlation. We state the following basic result, due to Brenier (cf. Villani 2003, Theorem 2.12, in which a proof is given).

PROPOSITION B.1. *Let ϱ be a maximal correlation risk measure with respect to baseline risk U. Then if both $\mathcal{L}_U$ and $\mathcal{L}_X$ are absolutely continuous, there exist a convex functions $V : \mathbb{R}^d \to \mathbb{R}$ and $W : \mathbb{R}^d \to \mathbb{R}$ which are Legendre-Fenchel conjugate of each other, i.e. $W = V^*$, and*

$$\varrho(X) = E[U \cdot \nabla V(U)],$$
$$\varrho(X) = E[X \cdot \nabla W(X)],$$

where the map ∇V pushes forward $\mathcal{L}_U$ to $\mathcal{L}_X$, and conversely ∇W pushes forward $\mathcal{L}_X$ to $\mathcal{L}_U$, and $\nabla W = (\nabla V)^{-1}$. When only $\mathcal{L}_U$ is absolutely continuous, then only those among the statements above involving V alone hold, and similarly, when only $\mathcal{L}_X$ is absolutely continuous then only those among the statements above involving W alone hold.

As $Q_X = \nabla V$ pushes forward measure μ on the distribution of X, it can be seen in some sense as a natural extension of a univariate quantile function (where—in which case $Q_X = F_X^{-1}$) to the multivariate setting.

REFERENCES

ARTZNER, P., F. DELBAEN, J.-M. EBER, and D. HEATH (1999): Coherent Measures of Risk, *Math. Finance* 9, 203–228.

AURENHAMMER, F., F. HOFFMANN, and B. ARONOV (1998): Minkowski-type Theorems and Least-squares Clustering, *Algorithmica* 20, 61–76.

BORWEIN, J., and A. LEWIS (2006): *Convex Analysis and Nonlinear Optimization*, 2nd ed. , New York: Springer. **Q3**

BURGERT, C., and L. RÜSCHENDORF (2006): Consistent Risk Measures for Portfolio Vectors, *Insur. Math. Econ.* 38, 289–297.

BARRIEU, P., and N. EL KAROUI (2005): Inf-convolution of Risk Measures and Optimal Risk Transfer, *Finance Stoch.* 9(2), 269–298. **Q4**

CHERIDITO, P., and T. LI (2009): Risk Measures on Orlicz Hearts, *Math. Finance* 19(2), 189–214.

DANA, R.-A. (2005): A Representation Result for Concave Schur Concave Functions, *Math. Finance* 15(4), 613–634.

DELBAEN, F. (2002): Coherent Risk Measures on General Probability Spaces, *Advances in Finance and Stochastics: Essays in Honour of Dieter Sondermann*, Berlin: Springer, pp. 1–37.

FABIAN, M., P. HABALA, P. HAJEK, V. MONTESINOS SANTALUCIA, J. PELANT, and V. ZIZLER (2001): *Functional Analysis and Infinite-Dimensional Geometry*, Springer: CMS Books in Mathematic.

FÖLLMER, H., and A. SCHIED (2004): *Stochastic Finance*, de Gruyter.

FRITTELLI, M., and E. ROSAZZA GIANIN (2005): Law Invariant Convex Risk Measures, *Adv. Math. Econ.* 7, 33–46.

GALICHON, A. (forthcoming): The VaR at Risk, *Int. J. Theor. Appl. Finance*.

JOUINI, E., M. MEDDEB, and N. TOUZI (2004): Vector Valued Coherent Risk Measures, *Finance Stoch.* 4, 531–552.

JOUINI, E., W. SCHACHERMAYER, and N. TOUZI (2006): Law Invariant Risk Measures have the Fatou Property, *Adv. Math. Econ.* 9, 49–71.

KUSUOKA, S. (2001): On Law Invariant Coherent Risk Measures, *Adv. Math. Econ.* 3, 83–95.

RACHEV, S., and L. RÜSCHENDORF (1998): *Mass Transportation Problems. Volume I: Theory*, and *Volume II: Applications*, New York: Springer.

RÜSCHENDORF, L. (2006): Law Invariant Convex Risk Measures for Portfolio Vectors, *Stat. Decis.* 24, 97–108.

RÜSCHENDORF, L. (2007): Monge–Kantorovich Transportation Problem and Optimal Couplings, *Jahresbericht der DMV* 3, 113–137.

RÜSCHENDORF, L., and L. UCKELMANN (2000): Numerical and Analytical Results for the Transportation Problem of Monge-Kantorovich, *Metrika. Int. J. Theor. Appl. Stat.* 51, 245–258.

SAMUELSON, P. (1947): *Foundations of Economic Analysis*, Cambridge, MA: Harvard University Press.

SCHMEIDLER, D. (1989): Subjective Probability and Expected Utility without Additivity, *Econometrica* 57, 571–587.

VILLANI, C. (2003): *Topics in Optimal Transportation*, Providence: American Mathematical Society.

YAARI, M. (1987): The Dual Theory of Choice under Risk, *Econometrica* 55, 95–115.

Queries

Q1 Author: Please update reference "Galichon (forthcoming)".

Q2 Author: Please provide citations for Figures 1 and 2.

Q3 Author: Reference "Borwein and Lewis (2006)" is not cited in the text. Please provide citation or else delete from the list.

Q4 Author: Reference "Barrieu and El Karoui (2005)" is not cited in the text. Please provide citation or else delete from the list.

Q5 Author: Reference "Delbaen (2002)" is not cited in the text. Please provide citation or else delete from the list.

Q6 Author: Reference "Frittelli and Rosazza Gianin (2005)" is not cited in the text. Please provide citation or else delete from the list.

Q7 Author: Reference "Ruschendorf (2007)" is not cited in the text. Please provide citation or else delete from the list.

Q8 Author: Reference "Schmeidler (1989)" is not cited in the text. Please provide citation or else delete from the list.

Q9 Author: Reference "Yaari et al. (1987)" is not cited in the text. Please provide citation or else delete from the list.