Multivariate comonotonicity, stochastic orders and risk measures

Alfred Galichon
(Ecole polytechnique)

Brussels, May 25, 2012

Based on collaborations with:
– A. Charpentier (Rennes) – G. Carlier (Dauphine)
– R.-A. Dana (Dauphine) – I. Ekeland (Dauphine)
– M. Henry (Montréal)

This talk will draw on four papers:

[CDG]. "Pareto efficiency for the concave order and multivariate comonotonicity". Guillaume Carlier, Alfred Galichon and Rose-Anne Dana. *Journal of Economic Theory, 2012*.

[CGH] ""Local Utility and Multivariate Risk Aversion". Arthur Charpentier, Alfred Galichon and Marc Henry. Mimeo.

[GH] "Dual Theory of Choice under Multivariate Risks". Alfred Galichon and Marc Henry. *Journal of Economic Theory, forthcoming*.

[EGH] "Comonotonic measures of multivariate risks". Ivar Ekeland, Alfred Galichon and Marc Henry. *Mathematical Finance, 2011*.

Introduction

Comonotonicity is a central tool in decision theory, insurance and finance.

Two random variables are « comonotone » when they are maximally correlated, i.e. when there is a nondecreasing map from one to another. Applications include risk measures, efficient risk-sharing, optimal insurance contracts, etc.

Unfortunately, no straightforward extension to the multivariate case (i.e. when there are several numeraires).

The goal of this presentation is to investigate what happens in the multivariate case, when there are several dimension of risk. Applications will be given to:
– Risk measures, and their aggregation
– Efficient risk-sharing
– Stochastic ordering.

1 Comonotonicity and its generalization

1.1 One-dimensional case

Two random variables X and Y are comonotone if there exists a r.v. Z and nondecreasing maps T_X and T_Y such that

$$X = T_X(Z) \text{ and } Y = T_Y(Z).$$

For example, if X and Y are sampled from empirical distributions, $X(\omega_i) = x_i$ and $Y(\omega_i) = y_i$, $i = 1, ..., n$ where

$$x_1 \leq ... \leq x_n \text{ and } y_1 \leq ... \leq y_n$$

then X and Y are comonotonic.

By the rearrangement inequality (Hardy-Littlewood),

$$\max_{\sigma \text{ permutation}} \sum_{i=1}^{n} x_i y_{\sigma(i)} = \sum_{i=1}^{n} x_i y_i.$$

More generally, X and Y are comonotonic if and only if

$$\max_{\tilde{Y} =_d Y} \mathbb{E}\left[X\tilde{Y}\right] = \mathbb{E}\left[XY\right].$$

Example. Consider

ω	ω_1	ω_2
$\mathbb{P}(\omega)$	$1/2$	$1/2$
$X(\omega)$	$+1$	-1
$Y(\omega)$	$+2$	-2
$\tilde{Y}(\omega)$	-2	$+2$

X and Y are comonotone.

$\tilde{Y}$ has the same distribution as Y but is not comonotone with X.

One has

$$\mathbb{E}[XY] = 2 > -2 = \mathbb{E}\left[X\tilde{Y}\right].$$

Hardy-Littlewood inequality. The probability space is now $[0,1]$. Assume $U(t) = \phi(t)$, where ϕ is nondecreasing.

Let P a probability distribution, and let

$$X(t) = F_P^{-1}(t).$$

For $\tilde{X} : [0,1] \to \mathbb{R}$ a r.v. such that $\tilde{X} \sim P$, one has

$$\mathbb{E}[XU] = \int_0^1 \phi(t) F_P^{-1}(t) dt \geq \mathbb{E}\left[\tilde{X}U\right].$$

Thus, letting

$$\begin{aligned} \varrho(X) &= \int_0^1 \phi(t) F_X^{-1}(t) dt = \max\left\{\mathbb{E}[\tilde{X}U],\ \tilde{X} =_d X\right\} \\ &= \max\left\{\mathbb{E}[X\tilde{U}],\ \tilde{U} =_d U\right\}. \end{aligned}$$

A geometric characterization. Let μ be an absolutely continuous distribution; two random variables X and Y are comonotone if for some random variable $U \sim \mu$, we have

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

Geometrically, this means that X and Y have the same projection of the *equidistribution class* of μ=set of r.v. with distribution μ.

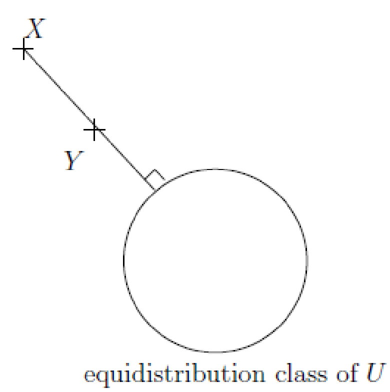

FIGURE 1. The equidistribution class of U is the circle and two μ-comonotone random vectors X and Y have the same L^2 projection $\tilde{U}$ on the equidistribution class of U with distribution μ.

1.2 Multivariate generalization

Problem: what can be done for risks which are multidimensional, and which are not perfect substitutes?

Why? risk usually has several dimension (price/liquidity; multicurrency portfolio; environmental/financial risk, etc).

Concepts used in the univariate case do not directly extend to the multivariate case.

The variational characterization given above will be the basis for the generalized notion of comonotonicity given in [EGH].

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

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

equivalentely:

X and Y are μ-comonotonic if there exists two convex functions V_1 and V_2 and a random variable $U \sim \mu$ such that

$$\begin{aligned} X &= \nabla V_1 (U) \\ Y &= \nabla V_2 (U) . \end{aligned}$$

Note that in dimension 1, this definition is consistent with the previous one.

Monge-Kantorovich problem and Brenier theorem

Let μ and P be two probability measures on $\mathbb{R}^d$ with second moments, such that μ is absolutely continuous. Then

$$\sup_{U \sim \mu, X \sim P} E\left[\langle U, X\rangle\right]$$

where the supremum is over all the couplings of μ and P if attained for a coupling such that one has $X = \nabla V\left(U\right)$ almost surely, where V is a convex function $\mathbb{R}^d \to \mathbb{R}$ which happens to be the solution of the dual Kantorovich problem

$$\inf_{V(u)+W(x) \geq \langle x,u\rangle} \int V\left(u\right) d\mu\left(u\right) + \int W\left(x\right) dP\left(x\right).$$

Call $Q_P\left(u\right) = \nabla V\left(u\right)$ the μ-quantile of distribution P.

Comonotonicity and transitivity.

Puccetti and Scarsini (2010) propose the following definition of comonotonicity, called c-comonotonicity: X and Y are c-comonotone if and only if

$$Y \in argmax_{\tilde{Y}} \left\{ \mathbb{E}[X \cdot \tilde{Y}],\ \tilde{Y} \sim Y \right\}$$

or, equivalently, iff there exists a convex function u such that

$$Y \in \partial u\,(X)$$

that is, whenever u is differentiabe at X,

$$Y = \nabla u\,(X)\,.$$

However, this definition is not transitive: if X and Y are c-comonotone and Y and Z are c-comonotone, and if the distributions of X, Y and Z are absolutely continuous, then X and Z are not necessarily c-comonotome.

This transivity (true in dimension one) may however be seen as desirable.

In the case of μ-comonotonicity, transitivity holds: if X and Y are μ-comonotone and Y and Z are μ-comonotone, and if the distributions of X, Y and Z are absolutely continuous, then X and Z are μ-comonotome.

Indeed, express μ-comonotonicity of X and Y: for some $U \sim \mu$,

$$\begin{aligned} X &= \nabla V_1 (U) \\ Y &= \nabla V_2 (U) \end{aligned}$$

and by μ-comonotonicity of Y and Z, for some $\tilde{U} \sim \mu$,

$$\begin{aligned} Y &= \nabla V_2 \left(\tilde{U}\right) \\ Z &= \nabla V_3 \left(\tilde{U}\right) \end{aligned}$$

this implies $\tilde{U} = U$, and therefore X and Z are μ-comonotone.

Importance of μ. In dimension one, one recovers the classical notion of comotonicity regardless of the choice of μ. However, in dimension greater than one, the comonotonicity relation crucially depends on the baseline distribution μ, unlike in dimension one. The following lemma from [EGH] makes this precise:

Lemma. Let μ and ν be atomless probability measures on $\mathbb{R}^d$. Then:
- In dimension $d = 1$, μ-comonotonicity always implies ν-comonotonicity.
- In dimension $d \geq 2$, 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 classes.

2 Applications to risk measures

2.1 Coherent, regular risk measures (univariate case)

Following Artzner, Delbaen, Eber, and Heath, recall the classical risk measures axioms:

Recall axioms:
Definition. A functional $\varrho : L^2_d \to \mathbb{R}$ is called a *coherent risk measure* if it satisfies the following properties:
- Monotonicity (MON): $X \leq Y \Rightarrow \varrho(X) \leq \varrho(Y)$
- Translation invariance (TI): $\varrho(X+m) = \varrho(X)+m\varrho(1)$
- Convexity (CO): $\varrho(\lambda X + (1-\lambda)Y) \leq \lambda\varrho(X) + (1-\lambda)\varrho(Y)$ for all $\lambda \in (0,1)$.
- Positive homogeneity (PH): $\varrho(\lambda X) = \lambda\varrho(X)$ for all $\lambda \geq 0$.

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

- Law invariance (LI): $\varrho(X) = \varrho(\tilde{X})$ when $X \sim \tilde{X}$.
- Comonotonic additivity (CA): $\varrho(X+Y) = \varrho(X) + \varrho(Y)$ when X, Y are comonotonic, i.e. weakly increasing transformation of a third randon variable: $X = \zeta_1(U)$ and $Y = \zeta_2(U)$ a.s. for ζ_1 and ζ_2 nondecreasing.

Result (Kusuoka, 2001). 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 (thus $Q_X(t) = F_X^{-1}(t)$ is the associated quantile).

ϱ is called a *Spectral risk measure*. For reasons explained later, also called *Maximal correlation risk measure*.

Leading example: **Expected shortfall** (also called **Conditional VaR** or **TailVaR**): $\phi(t) = \frac{1}{1-\alpha} \mathbf{1}_{\{t \geq \alpha\}}$. Then

$$\varrho(X) := \frac{1}{1-\alpha} \int_{\alpha}^{1} F_X^{-1}(t) dt.$$

Kusuoka's result, intuition.

- Law invariance $\Rightarrow \varrho(X) = \Phi\left(F_X^{-1}\right)$

- Comonotone additivity+positive homogeneity $\Rightarrow \Phi$ is linear w.r.t. F_X^{-1}:
  $\Phi\left(F_X^{-1}\right) = \int_0^1 \phi(t) F_X^{-1}(t) dt.$

- Monotonicity $\Rightarrow \phi$ is nonnegative

- Subadditivity $\Rightarrow \phi$ is increasing

Unfortunately, this setting does not extend readily to multivariate risks. We shall need to reformulate our axioms in a way that will lend itself to easier multivariate extension.

2.2 Alternative set of axioms

Manager supervising several N business units with risk $X_1, ..., X_N$.
Eg. investments portfolio of a **fund of funds**. True economic risk of the fund $X_1 + ... + X_N$.

Business units: portfolio of (contingent) losses X_i report a summary of the risk $\varrho(X_i)$ to management.

Manager has limited information:
1) does not know what is the **correlation** of risks - and more broadly, the dependence structure, or **copula** between $X_1, ..., X_N$. Maybe all the hedge funds in the portfolio have the same risky exposure; maybe they have independent risks; or maybe something inbetween.

2) aggregates risk by summation: reports $\varrho(X_1) + ... + \varrho(X_N)$ to shareholders.

Reported risk: $\varrho(X_1)+...+\varrho(X_N)$; **true risk:** $\varrho(X_1+ ... + X_N)$.

Requirement: management does not understate risk to shareholders. Summarized by

$$\varrho(X_1) + ... + \varrho(X_N) \geq \varrho(\tilde{X}_1 + ... + \tilde{X}_N) \qquad (*)$$

whatever the joint dependence $(X_1, ..., X_N) \in (L_d^\infty)^2$.

But no need to be overconservative:

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

where $\sim$ denotes equality in distribution.

Definition. A functional $\varrho : L_d^2 \to \mathbb{R}$ is called a *strongly coherent risk measure* if it is convex continuous and for all $(X_i)_{i \leq N} \in \left(\mathcal{L}_d^2\right)^N$,

$$\varrho(X_1)+...+\varrho(X_N) = \sup\left\{\varrho(\tilde{X}_1 + ... + \tilde{X}_N) \,:\, \tilde{X}_i \sim X_i\right\}.$$

A representation result.

The following result is given in [EGH].

Theorem. The following propositions about the functional ϱ on L_d^2 are equivalent:
(i) ϱ is a strongly coherent risk measure;
(ii) ϱ is a *max correlation risk measure*, namely there exists $U \in L_d^2$, such that for all $X \in L_d^2$,

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

(iii) There exists a convex function $V : \mathbb{R}^d \to \mathbb{R}$ such that
$\varrho(X) = E[U \cdot \nabla V(U)]$

Idea of the proof . One has $\varrho(X)+\epsilon\varrho(Y) = \sup\left\{\varrho(X+\epsilon\tilde{Y}) :\right.$
But $\varrho(X+\epsilon\tilde{Y}) = \varrho(X) + \epsilon D\varrho_X(\tilde{Y}) + o(\epsilon)$

By the Riesz theorem (vector case) $D\varrho_X(\tilde{Y}) = E\left[m_X.\tilde{Y}\right]$, thus

$$\varrho(X)+\epsilon\varrho(Y) = \sup\left\{\varrho(X) + \epsilon E\left[m_X.\tilde{Y}\right] + o(\epsilon) : \ \tilde{Y} \sim Y\right\}$$

thus

$$\varrho(Y) = \sup\left\{E\left[m_X.\tilde{Y}\right] : \ \tilde{Y} \sim Y\right\}$$

therefore ϱ is a maximum correlation measure.

3 Application to efficient risk-sharing

Consider a risky payoff X (for now, univariate) to be shared between 2 agents 1 and 2, so that in each contingent state:

$$X = X_1 + X_2$$

X_1 and X_2 are said to form an allocation of X.

Agents are risk averse in the sense of stochastic dominance: Y is preferred to X if every risk-averse expected utility decision maker prefers Y to X:

$X \leq_{cv} Y$ iff $\mathbb{E}[u(X)] \leq \mathbb{E}[u(Y)]$ for all concave u

Agents are said to have concave order preferences. These are incomplete preferences: it can be impossible to rank X and Y.

One wonders what is the set of efficient allocations, i.e. allocations that are not dominated w.r.t. the concave order for every agent.

Dominated allocations. Consider a random variable X (aggregate risk). An allocation of X among p agents is a set of random variables $(Y_1, ..., Y_p)$ such that

$$\sum_i Y_i = X.$$

Given two allocations of X, Allocation (Y_i) dominates allocation (X_i) whenever

$$\mathbb{E}\left[\sum_i u_i(Y_i)\right] \geq \mathbb{E}\left[\sum_i u_i(X_i)\right]$$

for every continuous concave functions $u_1, ..., u_p$. The domination is strict if the previous inequality is strict whenever the u_i's are strictly concave.

Comonotone allocations. In the single-good case, it is intuitive that efficient sharing rules should be such that in “better” states of the world, every agent should be better of than in “worse” state of the world – otherwise there would be some mutually agreeable transfer.

This leads to the concept of comonotone allocations. The precise connection with stochastic dominance is due to Landsberger and Meilijson (1994). Comonotonicity has received a lot of attention in recent years in decision theory, insurance, risk management, contract theory, etc. (Landsberger and Meilijson, Ruschendorf, Dana, Jouini and Napp...).

Theorem (Landsberger and Meilijson). Any allocation of X is dominated by a comonotone allocation. Moreover, this dominance can be made strict unless X is already comonotone. Hence the set of efficient allocations of X coincides with the set of comonotone allocations.

This result generalizes well to the multivariate case. Up to technicalities (see [CDG] for precise statement), efficient allocations of a random vector X is the set of

μ-comonotone allocations of X, hence (X_i) solves

$$\begin{aligned} X_i &= \nabla u_i (U) \\ \sum_i X_i &= X \end{aligned}$$

for convex functions $u_i : \mathbb{R}^d \to \mathbb{R}$, with $U \sim \mu$. Hence

$$X = \nabla u (U)$$

with $u = \sum_i u_i$. That is

$$U = \nabla u^* (X) ,$$

hence efficient allocations are such that

$$X_i = \nabla u_i \circ \nabla u^* (X) .$$

This result opens the way to the investigation of testable implication of efficiency in risk-sharing in an risky endowment economy.

4 Application to stochastic orders

Quiggin (1992) shows that the notion of monotone mean preserving increases in risk (hereafter MMPIR) is the weakest stochastic ordering that achieves a coherent ranking of risk aversion in the rank dependent utility framework. MMPIR is the mean preserving version of Bickel-Lehmann dispersion, which we now define.

Definition. Let Q_X and Q_Y be the quantile functions of the random variables X and Y. X is said to be *Bickel-Lehmann less dispersed*, denoted $X \succsim_{BL} Y$, if $Q_Y(u) - Q_X(u)$ is a nondecreasing function of u on $(0, 1)$. The mean preserving version is called monotone mean preserving increase in risk (MMPIR) and denoted $\precsim_{MMPIR}$.

MMPIR is a stronger ordering than concave ordering in the sense that $X \succsim_{MMPIR} Y$ implies $X \succsim_{cv} Y$.

The following result is from Landsberger and Meilijson (1994):

Proposition (Landsberger and Meilijson). A random variable X has Bickel-Lehmann less dispersed distribution than a random variable Y if and only iff there exists Z comonotonic with X such that $Y =_d X + Z$.

The concept of μ-comonotonicity allows to generalize this notion to the multivariate case as done in [CGH].

Definition. A random vector X is called μ-Bickel-Lehmann less dispersed than a random vector Y, denoted $X \succsim_{\mu BL} Y$, if there exists a convex function $V : \mathbb{R}^d \to \mathbb{R}$ such that the μ-quantiles Q_X and Q_Y of X and Y satisfy $Q_Y(u) - Q_X(u) = \nabla V(u)$ for μ-almost all $u \in [0, 1]^d$.

As defined above, μ-Bickel-Lehmann dispersion defines a transitive binary relation, and therefore an order. Indeed, if $X \succsim_{\mu BL} Y$ and $Y \succsim_{\mu BL} Z$, then $Q_Y(u) - Q_X(u) =$

$\nabla V(u)$ and $Q_Z(u) - Q_Y(u) = \nabla W(u)$. Therefore, $Q_Z(u) - Q_X(u) = \nabla(V(u) + W(u))$ so that $X \succsim_{\mu BL} Z$. When $d = 1$, this definition simplifies to the classical definition.

[CGH] propose the following generalization of the Landsberger-Meilijson characterization .

Theorem. A random vector X is μ-Bickel-Lehmann less dispersed than a random vector Y if and only if there exists a random vector Z such that:

(i) X and Z are μ-comonotonic, and

(ii) $Y =_d X + Z$.

Conclusion

We have introduced a new concept to generalize comonotonicity to higher dimension: "μ-comonotonicity". This concept is based on Optimal Transport theory and boils down to classical comonotonicity in the univariate case.

We have used this concept to generalize the classical axioms of risk measures to the multivariate case.

We have extended existing results on equivalence between efficiency of risk-sharing and μ-comonotonicity.

We have extended existing reults on functions increasing with respect to the Bickel-Lehman order.

Interesting questions for future research: behavioural interpretation of mu? computational issues? empirical testability? case of heterogenous beliefs?