Modelling Extreme Dependence for Multivariate Data

Damien Bosc
(joint work with Alfred Galichon)

Département d'Economie Ecole Polytechnique & Axa-IM

Motivation

Investigate the notion of 'multivariate' dependence and extreme multivariate dependence.

- univariate dependence:
  dependence between real r.v $\rightarrow$ 'copula world'
- multivariate dependence:
  dependence between two random *vectors* / multivariate laws of probability.

As with bivariate copulas, marginal laws p and q are fixed; but they are multivariate, i.e. laws on $\mathbf{R}^n$.
A *coupling* between p and q is a law of a couple (X, Y) with $X \sim p$, $Y \sim q$.
The set of all couplings between p and q is denoted $\Pi(p, q)$.

Introduction

In the univariate case, the strongest dependence between two random variables is given by upper Fréchet Copula:

$$C(u_1, u_2) = min(u_1, u_2)$$

A couple (X, Y) exhibiting upper Fréchet dependence maximizes the covariance

$$\mathbf{E}(XY) = \sup_{\substack{\widetilde{X} \sim X \\ \widetilde{Y} \sim Y}} \mathbf{E}(\widetilde{X}\widetilde{Y})$$

In higher dimensions, there is no notion of copula between multivariate vectors: no 'natural' notion of Fréchet multivariate dependence exists.

Introduction 2

One possible extension: the *maximum correlation coupling* is the coupling π s.t.

$$\mathbf{E}_{\pi}(X'Y) = \max_{\substack{\widetilde{X}\sim X \\ \widetilde{Y}\sim Y}} \mathbf{E}(\widetilde{X}'\widetilde{Y})$$

This is not fully satisfactory as it involves only component-wise covariances; the notion of cross dependence is not accounted for. Our goal is to define a more general notion of extreme dependence that yields more extremely dependent couplings.

More on cross-covariance

We will focus on the cross-covariance matrix of (X, Y). As a unique vector in $\mathbf{R}^{2n}$ its covariance matrix is

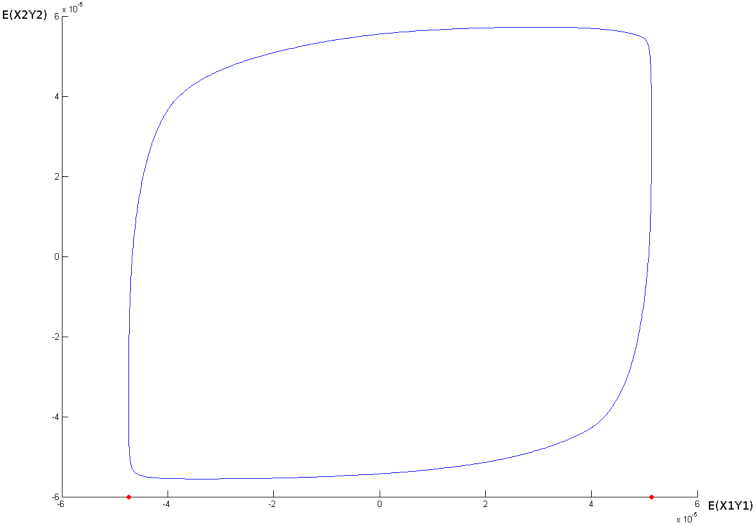
$$Cov((X,Y)) = \left(\begin{array}{c|c} Cov(X) & \mathbf{E}(XY') \\ \hline \mathbf{E}(XY')' & Cov(Y) \end{array}\right)$$

- The diagonal blocks are known (they do not depend on the coupling).
- For convenience we write $\sigma_{X,Y} = \mathbf{E}(XY') = (\mathbf{E}(X_i Y_j))_{i,j}$.
- Example: if $p = q = \mathcal{N}(0, Id_2)$, $X = Y \Rightarrow \sigma_{X,Y} = Id_2$.

Projecting couplings in the plane

- A simple example is to consider two bivariate laws $X = \binom{X_1}{X_2}$ and $Y = \binom{Y_1}{Y_2}$.
- One can project any coupling between X and Y in the plane by considering the coordinates $(\mathbf{E}(X_1 Y_1), \mathbf{E}(X_2 Y_2))$.
- We obtain the image of $\Pi(p, q)$: this is the set of attainable covariances, called the *covariogram*.

Attainable covariances



Conic ordering on cross-covariance matrices

→ Loewner (strict) order on symmetric matrices:

$$A \succ B, \text{ iff } A - B \in S_n^{++}$$

→ If $\pi = \mathcal{L}_{(X,Y)}$, symmetric part of the cross-covariance matrix:

$$\Sigma_\pi = (\sigma_\pi + \sigma'_\pi)/2$$

→ Σ_π is *maximal* if for all $\Sigma_{\pi'}$, $\Sigma_{\pi'} - \Sigma_\pi \notin S_n^{++}$

Problem

Which couplings π yield a Σ_π *maximal* for $\succ$?

We denote these *maximal couplings*: for instance the maximal correlation coupling is maximal.

Main result on maximizing cross-covariance matrices

Variational characterization of maximality

$$\Sigma_\pi \text{ extremal iff } \exists A \in S_n^+ \backslash \{0\} \text{ such that}$$

$$\Sigma_\pi \cdot A = \sup_{\pi' \in \Pi(p,q)} \Sigma_{\pi'} \cdot A \tag{1}$$

In other words (1) $\Leftrightarrow$ $\mathbf{E}_\pi(X'AY)$ is maximum $\Leftrightarrow$ under π :

$$(X, AY) \text{ has the maximum correlation}$$

Ex : two bivariate vectors sharing the *same* first component are not necessarily maximally correlated. But this is a maximal coupling as it satisfies (1) with $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$.
$\rightarrow$ As a consequence, we thus have many extremely dependent couplings.

Covariogram and maximal couplings

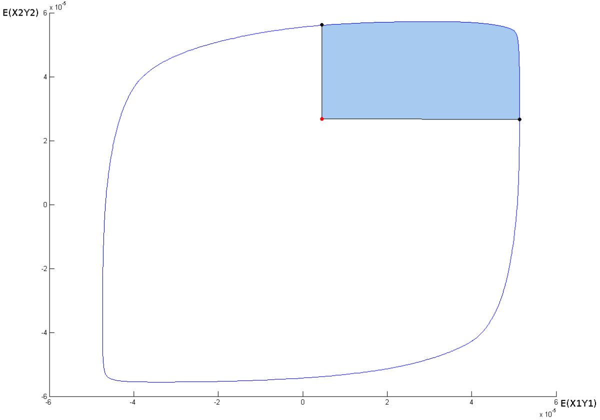


Figure: Positive orthant order

Covariogram and maximal couplings (2)

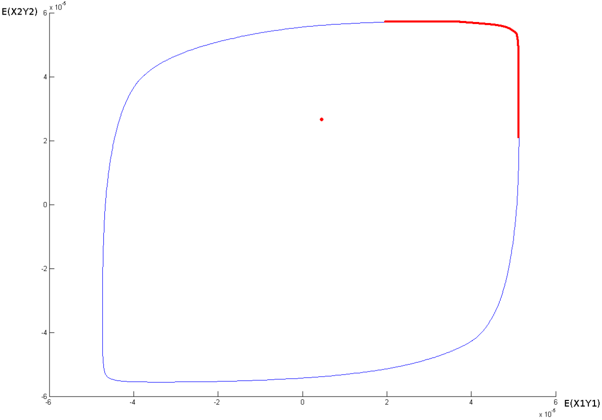


Figure: Location of maximal couplings

Covariogram and its boundary

Boundary of special interest: the boundary of the complete covariogram is composed of couplings π for which there exists some $A \neq 0$ such that

$$\mathbf{E}(X'AY) = \sup_{\pi' \in \Pi(p,q)} \mathbf{E}_{\pi'}(X'AY) \qquad (2)$$

or equivalently

$$\sigma_\pi \cdot A = \sup_{\pi'} \sigma_{\pi'} \cdot A$$

(1) is a particular case of (2) with constraints on the form of A.

Entropic relaxation

Perform an *entropic relaxation* of the problem (2) :

$$W(A,T) = \max_{\pi} \mathbf{E}(X'AY) + T\,Ent(\pi) \tag{3}$$

$Ent(\pi)$ is the *entropy* of π, defined as $-\mathbf{E}_{\pi}\big(\log(\pi(X,Y))\big)$.
Homogeneity in (A,T) : we set the temperature at 1. We then aim at finding $\widehat{A}$ s.t

$$\sigma_{\hat{\pi}} = \sigma_{\pi(\widehat{A})} \qquad , \qquad \pi(A) \text{ solution of } (3)$$

Remark that :

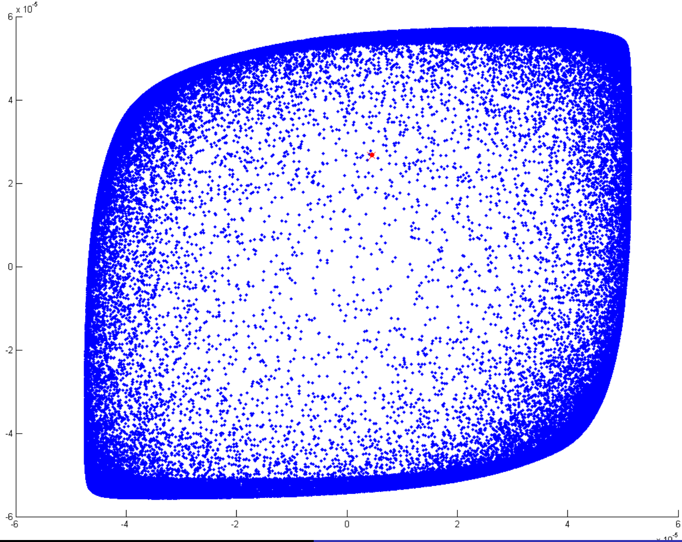
$$\nabla_A W(A,1) = \sigma_{\pi(A)}$$

It thus amounts to solve :

$$\min_{A} W(A,1) - \sigma_{\hat{\pi}} \cdot A \tag{4}$$

which is a *convex minimization problem*.

Filled covariogram



Algorithm : Iterative Proportional Fitting Procedure

The solution π of (3) can be shown to take the form:

$$\log \pi(x,y) \quad = x'Ay + u(x) + v(y), \quad u \in L^1(dp), v \in L^1(dq)$$

u and v must be adjusted so that $\pi \in \Pi(p,q)$.
This is the purpose of IPFP (Deming & Stephan 1940, Von Neumann 1950). Recursion scheme :

$$\begin{cases} e^{u_{n+1}(x)} & = & \dfrac{p(x)}{\int e^{x'Ay+v_n(y)}dy} \\ e^{v_{n+1}(x)} & = & \dfrac{q(y)}{\int e^{x'Ay+u_{n+1}(x)}dx} \end{cases}$$

- $\pi_{2n} \propto e^{x'Ay+u_n(x)+v_n(y)}$ has first marginal p
- $\pi_{2n+1} \propto e^{x'Ay+u_n(x)+v_{n+1}(y)}$ has second marginal q
- $\pi_n \to \pi \in \Pi(p,q)$ in total variation

Examples with sector indices

We consider the time series of daily returns on S&P 500 and DJ EUROSTOXX subsectors: construction, health care and financials.

We introduce $X = \binom{X_1}{X_2}$, $Y = \binom{Y_1}{Y_2}$ with,

$$\begin{aligned} X_1 = & \quad \text{return on the S\&P 500 health care sector} \\ X_2 = & \quad \text{return on the S\&P 500 financial sector} \end{aligned}$$

Y is defined in the same manner for the DJ Eurostoxx.

Examples with sector indices

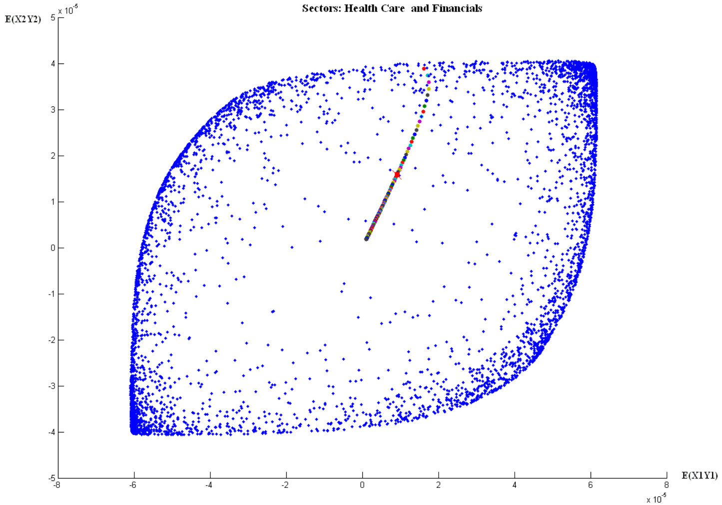
Multivariate discrete laws are defined from historical data X_t and Y_t:

$$p = \frac{1}{N} \sum_{t=1}^{N} \delta_{X_t} \quad , \quad q = \frac{1}{N} \sum_{t=1}^{N} \delta_{Y_t}$$

For historical data spanning 5 years between september 2004 and september 2009 one gets the following results :

# of components	2	3
$\widehat{A}$	$\begin{pmatrix} 0.23 & -0.14 \\ -0.10 & 0.40 \end{pmatrix}$	$\begin{pmatrix} 0.25 & -0.139 & -0.37 \\ -0.39 & 0.44 & -0.80 \\ -0.57 & -0.15 & 0.86 \end{pmatrix}$
error $= \frac{\lvert\lvert\sigma_A - \sigma_{\hat{\pi}}\rvert\rvert}{\lvert\lvert\sigma_{\hat{\pi}}\rvert\rvert}$	$\approx 0.1\%$	< 0.2 %

Example of trajectory



Analysis of the optimal coupling

The empirical coupling $\hat{\pi}$ is thus associated with a $\pi(\widehat{A}, T = 0)$ which maximizes

$$\mathbf{E}_\pi(X'\widehat{A}Y),\ \pi \in \Pi(p, q)$$

Singular value decomposition on $\widehat{A}$ yields $\widehat{A} = USV'$ where U, V are unitary and S diagonal nonnegative.
$(\tilde{X}, \tilde{Y}) = (\sqrt{S}U'X, \sqrt{S}VY)$ has maximum correlation.
Ex : with 3 components, we obtain

$$\begin{aligned} \tilde{X} &= \begin{pmatrix} -0.42X_1+0.95X_2-0.019X_3 \\ -0.64X_1-0.27X_2+0.26X_3 \\ 0.11X_1+0.06X_2+0.35X_3 \end{pmatrix} \\ \tilde{Y} &= \begin{pmatrix} -0.30Y_1+0.99Y_2-0.13Y_3 \\ -0.67Y_1-0.16Y_2+0..28Y_3 \\ 0.12Y_1+0.08Y_2+0.34Y_3 \end{pmatrix} \end{aligned}$$

$\rightarrow$ two *indices* most correlated to one another.

Conclusion

- We proposed a new notion of extreme dependence in the multivariate case
- Maximality of cross-covariance matrices can be generalized to more general conic orders
- To every coupling between multivariate laws (historical, simulated ...) we can associate an extremal coupling
- It yields a natural construction of indices of maximum correlation: 'foreign risk' indices
- This can apply to stress testing: build scenarios on multivariate laws with increasing dependence.