Persistence landscapes and correlation: an application to financial multivariate time series

Josep Vives *(josep.vives@ub.edu)*

Departament de Matemàtiques i Informàtica

Topological Machine Learning - UB Seminar
May 6, 2022

Reference papers

The talk is based on the papers:

LLOYD L. AROMÍ, YURI A. KATZ AND JOSEP VIVES (2021): *Topological features of multivariate distributions: dependency on the covariance matrix*. Communications in Nonlinear Science and Numerical Simulation 103: 105996.

CARLES CASACUBERTA, DAVID FARRÉ AND JOSEP VIVES (2022): *Bounding the persistence of cycles in a filtered simplicial complex*. Preprint in progress.

Outline

Abstract

- We study the relationship between L^p**-norms of persistence landscapes** of random datasets and the properties of their generating probability distributions.
- We conduct numerical experiments with bivariate time series generated by probability distributions with known properties which reveal that **the increase of L^p-norms due to rising variability can be suppressed by a growing covariance in the system**.
- These results help to understand the puzzling behavior of topological summaries observed on US and European stock markets during the early phase of the global meltdown caused by the COVID-19 pandemic,
- More generally, they could aid in a topological identification of approaching regime changes in complex systems.

Introduction I

- Topological data analysis (TDA) combines deterministic constructs of computational topology with statistical and machine learning methods to study the shape of data.
- The analysis of persistent landscapes is a major TDA tool to describe point clouds (Bubenik (2015)).
- This approach has been successfully used in Finance as well as in many other fields (Gidea and Katz (2018)).

Introduction II

- We prove a functional relationship of the mean of L^p-norms to the variability of a point cloud sampled from a broad class of probability distributions.
- We conduct numerical experiments with bivariate time series spawned by probability distributions with known properties, which reveal that the increase of the average L^p-norms due to rising variability can be suppressed by growing covariance in the system.

Introduction III

- The results help to understand the puzzling behavior of the norms of persistent landscapes observed on US and European stock markets at the inception phase of the global meltdown caused by the COVID-19 pandemic.
- Specifically, contrary to early phases of the technology crash of 2000 and the global financial crisis of 2007-2009, we do not observe any growth of L^p-norms derived from the daily time series of four US and four European equity indices, at the beginning of the most recent market meltdown.

Introduction IV

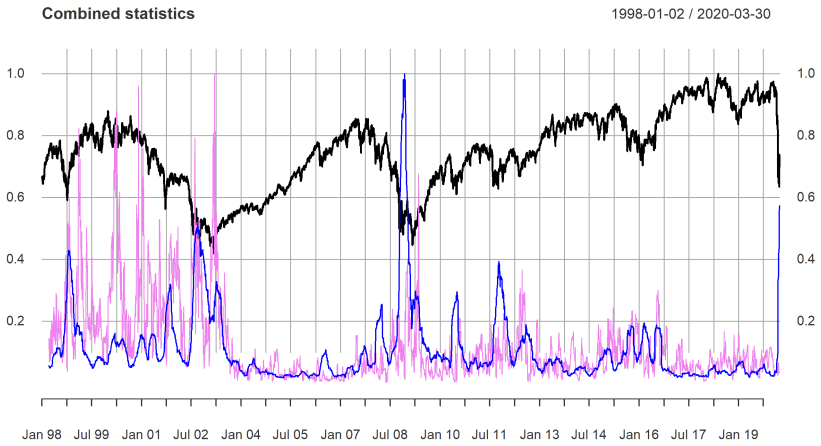

Figure: (Color online) Time series of FTSE 100 (black line), L^1-norm (purple line) and estimated variability (blue line) of the mixture of daily log-returns of four indices.

Introduction V

- Financial markets provide a fascinating example of a complex system. Sudden regime changes – financial crashes – are characterized by increased variance and, often, by growing auto-and cross-correlation between broad market indices.
- Here, we use the sliding window technique to quantify the temporal changes in variability, cross-correlation, and topological features of financial time series under study.

Introduction VI

- The observed cross-correlations between different equity indices were much stronger (> 0.95) at the inception of the global 2020 crash than at the early stages of the two previous crises.
- We conclude that, despite the spike in variability, the strong exogenous shock caused by the COVID-19 pandemic leads to synchronization between all elements of the global financial market system, which essentially eliminates any persistence of homological features of the underlying point clouds.

Introduction VII

Imagine a point cloud in $\mathbb{R}^2$. It will become clear that:

If points are distributed uniformly in a circle, the increasing of variability implies also the increasing of the norm of its persistence landscape.

On the contrary, if points are distributed in an ellipse with a very small second axis, the increase of variability does not generate changes in the norm of its persistence landscape.

Topological Data Analysis (TDA)

- TDA is designed to study and measure certain features of discrete multidimensional data sets, commonly studied as point clouds, embedded in $\mathbb{R}^d$, using a combination of statistical, computational and topological tools (Bubenik (2015), Carlsson (2009), Edelsbrunner and Harer (2009)).
- Calculation of persistence homology (PH) is at the core of TDA. Informally, it is based on the computation of persistence of k-dimensional cycles, e.g., connected components ($k = 0$), loops ($k = 1$), cavities ($k = 2$) and so on, at a wide range of scales.
- Like the shape of a distant landscape is changing at different binocular resolutions, topological features of a dataset are changing with the scale. As the scaling parameter changes, some homologies appear while others disappear.

Persistence Homology I

- Each homology feature is assigned a 'birth' and a 'death' value, and the difference between these two values represents its life or persistence.
- The output of this filtration procedure is captured in a concise form by a persistence Rips diagram. Coordinates of each point on the Rips diagram represent the birth value (x-coordinate) and the death value (y-coordinate) of a k-dimensional cycle.
- As a result, an arbitrary complex multidimensional dataset is projected via a Rips filtration onto the two-dimensional persistence diagram.

Persistence Homology II

- For this procedure to be applied, a finite multidimensional dataset must be encoded in some metric space, forming a 'point cloud'.
- A standard procedure to compute the persistence homology of a point cloud relies on the construction of a filtration of simplicial complexes.
- Two main types of assemblies of complexes can be used: Cech scheme and Vietoris-Rips scheme (Carlsson (2009)). Both are quite equivalent (Jung's theorem).
- Computations in the published paper are done using Vietoris-Rips scheme. But now we know that since the theoretical point of view, and for our purposes, Cech scheme is better. Despite this, two methods are very similar and Vietoris-Rips is more easy to implement.

Cech scheme

- Consider an increasing resolution parameter $\epsilon > 0$. In Cech scheme, two points p and q become connected when ϵ arrives to $\frac{1}{2}d(p, q)$. (In our case, d is Euclidean distance).
- We are interested in loops (cycles in the plane). A cycle starts when a certain number of points p_i of the cloud become connected leaving a hole in the interior, and dies when all the points in the interior are connected with one of the generating points of the cycle.
- Note that triangles can have a short life under Cech scheme (not under the Vietoris-Rips one). But rectangle triangles have no life in any of the schemes.

Persistent landscapes

- A Persistence Homology based summarization instrument is the persistence landscape. It consists of a sequence of piecewise linear functions defined in the re-scaled birth-death coordinates of the underlying Rips diagram.
- The key advantage of persistence landscapes is related to their embedding into a Banach space. Hence, one can apply standard tools of functional analysis and statistics, e.g., compute their means, variances and norms (Bubenik (2015)).

The persistence landscape function

Definition

Let $\mathcal{D} = \{(b_i, d_i)\}_{i \in I}$ be a persistence diagram. For each birth-death point (b_i, d_i) in $\mathcal{D}$, we define a piecewise linear continuous function:

$$f_{(b_i,d_i)}(x) = \begin{cases} x - b_i & \text{if } b_i < x \leq \frac{b_i+d_i}{2} \\ -x + d_i & \text{if } \frac{b_i+d_i}{2} < x < d_i \\ 0 & \text{otherwise.} \end{cases}$$

Then, the function $\Lambda \colon \mathbb{N} \times \mathbb{R} \longrightarrow \mathbb{R}$ given by $\Lambda(k, x) = \mathrm{kmax}\{f_{(b_i,d_i)}(x)\}_{i \in I}$, is called the *persistence landscape* function associated to the persistence diagram $\mathcal{D}$, where kmax denotes the k-th largest value of a set.

Basic Properties

- By definition, if $k > |I|$, then the value of kmax is zero, where I denotes the index set of $\mathcal{D}$.
- Alternatively, a persistence landscape may also be viewed as a sequence of functions $\lambda_1, \lambda_2, \ldots : \mathbb{R} \to \mathbb{R}$, where $\lambda_k(x) = \Lambda(k, x)$ is called the *k-th persistence landscape function* of $\mathcal{D}$.
- Each function $\lambda_k(x)$ is piecewise linear with slope either 0, 1, or -1.

Figure 1

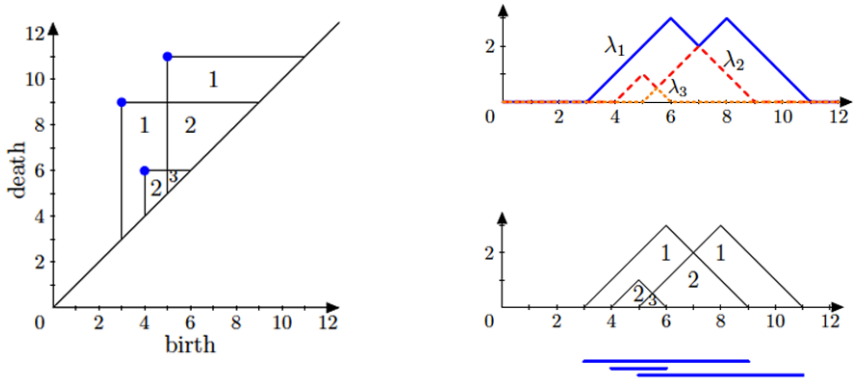

Figure: On the right, two visualizations of the persistence landscape derived from the persistence diagram on the left. The top-right figure, shows the different k-persistence landscape functions for $k = 1, 2, 3$. The bottom-right figure shows the persistence landscape above the barcodes. Figure adapted from Bubenik (2015).

The k-th persistence landscape

For a persistence landscape Λ derived from a persistence diagram $\mathcal{D}$, the k-th persistence landscape function λ_k has the following properties:

a) $\lambda_k(x) \geq 0$ for all x;

b) $\lambda_k(x) \geq \lambda_{k+1}(x)$ for all x.

First statement follows directly from the definition. Indeed, for every birth-death point p in $\mathcal{D}$, we have the associated piecewise linear function f_p which has image $\mathrm{Im}(f_p) \subseteq [0, \infty)$. Thus, for every k we have that $\mathrm{kmax}\{f_{(b_i,d_i)}(x)\}_{i\in I} \geq 0$.

The second property follows directly from the fact that

$$\lambda_k = \mathrm{kmax}\{f_{(b_i,d_i)}\}_{i\in I} \geq (k+1)\mathrm{max}\{f_{(b_i,d_i)}\}_{i\in I} = \lambda_{k+1}.$$

Norms of persistent landscapes I

Persistence landscapes can be understood in terms of elements of a Banach space. Recall that, for a given measure space $(\mathcal{S}, \mathcal{A}, \mu)$ and a function $f\colon \mathcal{S} \to \mathbb{R}$ defined μ-almost everywhere, one defines, for $1 \leq p < \infty$, the *L^p-norms*

$$\|f\|_p = \Big(\int |f|^p \, d\mu \Big)^{1/p}$$

$$\|f\|_\infty = \sup_{x \in \mathcal{S}} |f(x)| = \inf \{a \mid \mu\{s \in \mathcal{S} : f(s) > a\} = 0\}$$

Moreover, for $1 \leq p \leq \infty$, we have the Banach space

$$\mathcal{L}^p(\mathcal{S}) = \{f\colon \mathcal{S} \to \mathbb{R} \mid \|f\|_p < \infty\},$$

and define $L^p(\mathcal{S}) = \mathcal{L}^p(\mathcal{S})/\sim$, where $f \sim g$ if $\|f - g\|_p = 0$.

Norms of persistent landscapes II

Hence, we can define the norm of persistence landscape, as follows.

Definition

Let $\Lambda\colon \mathbb{N}\times\mathbb{R}\to\mathbb{R}$ be a persistence landscape function. Suppose that on $\mathbb{N}\times\mathbb{R}$ we use the product of the counting measure on $\mathbb{N}$ and the Lebesgue measure on $\mathbb{R}$. Then, for $1\leq p<\infty$, we define

$$\|\Lambda\|_p=\sum_{k=1}^{\infty}\|\lambda_k\|_p,$$

where $\lambda_k(t)=\Lambda(k,t)$, and $\|\lambda_k\|_p$ denotes the standard L^p-norm of λ_k.

Norms of persistence landscapes III

- Thus, we can endow the space of persistent landscapes with the previous norm and the set of persistence landscapes becomes a subset of the Banach space $L^p(\mathbb{N} \times \mathbb{R})$.
- Ana alternative would be

$$\|\Lambda\|_p = (\sum_{k=1}^{\infty} \|\lambda_k\|_p^p)^{\frac{1}{p}},$$

- From my experience, for practical purposes, to compute the L^1-norm of λ_1 is enough. In general, other alternatives don't add much more to the analysis.
- The L^1-norm of λ_1 is the area under the highest chain of the landscape.

L^p-norms of randomized point clouds I

- Our goal is to characterize the behavior of the mean of L^p-norms of persistence landscapes derived from random point clouds sampled from a broad class of multivariate distributions with known statistical properties.
- For a given probability space $(\Omega, \mathcal{F}, P)$, let X denote a multivariate random variable with distribution F and corresponding persistence landscape Λ. By this we mean that, for $\omega \in \Omega$, $\mathbb{X}(\omega)$ is a data set and $\Lambda(\omega)$ is the corresponding persistence landscape. In this sense, we interpret the persistence landscape as a Banach space valued random variable $\Lambda(\omega) \colon \Omega \to L^p(\mathbb{N} \times \mathbb{R})$.

Some statistical properties of persistence landscapes

- Under some conditions on the distribution F, the behavior of the expected L^p-norm of a persistence landscape can be described by a *non linear dependency* on a scaling factor of F.
- For a given persistence landscape Λ we have introduced before a p-norm defined as the infinite sum of $\|\lambda_k\|_p$, where λ_k is the k-th persistence landscape function of Λ.
- For a finite data set, there exists t such that $\lambda_k = 0$ for all $k \geq t$ and $\|\lambda_k\|_p = 0$ for all p, hence $\sum_{k=1}^{\infty} \|\lambda_k\|_p < \infty$.

Proposition 1

Proposition

Let $X\colon \Omega \to \mathbb{R}^d$ denote a multivariate random variable with distribution $D(\mu, \Sigma)$, and let $X^1, \ldots, X^N$ be identical copies of X such that $\mathbb{X} := (X^1, \ldots, X^N)$ describes an N-point data set in $\mathbb{R}^d$. Assume μ and Σ are finite. Then,

$$\mathbb{E}(\|\Lambda\|_p) \leq C(N) \cdot N \cdot tr(\Sigma)^{\frac{p+1}{2p}},$$

where $\|\Lambda\|_p$ denotes the p-norm of the persistence landscape Λ, and $C(N)$ describes the number of non zero k-persistence landscapes.

Proof I

Let $\{(b_i, d_i)\}_{i \in I}$ the different cycles appeared when resolution ϵ increases. Define

$$\epsilon_b = \min_s \{b_s\}, \quad \epsilon_d = \max_l \{d_l\},$$

to be the minimum and the maximum values at which a first cycle is born and the last cycle dies, respectively. Now, for each $x \in (\epsilon_b, \epsilon_d)$, we consider the piecewise linear function

$$\tilde{f}(x) := f_{(\epsilon_b, \epsilon_d)}(x) = \begin{cases} x - \epsilon_b & \text{if } x \in (\epsilon_b, \frac{\epsilon_b + \epsilon_d}{2}] \\ -x + \epsilon_d & \text{if } x \in (\frac{\epsilon_b + \epsilon_d}{2}, \epsilon_d). \end{cases}$$

Proof II

It is enough to bound the 1-persistence landscape function λ_1, and since $\lambda_k \geq \lambda_{k+1}$ we can bound $\sum_{k\geq 1} \|\lambda_k\|_p$ because the sum if finite.

It is easily deduced that $\mathrm{dom}(\lambda_1) = \mathrm{dom}(\tilde{f}) =: S \subset \mathbb{R}$, since $\lambda_1 = \max\{f_{p_i}\}_{i\in I}$. Moreover, $\tilde{f}(x) \geq \lambda_1(x) \geq 0$ for all $x \in S$. Thus,

$$\|\lambda_1\|_p^p \leq \|\tilde{f}\|_p^p = \frac{(\epsilon_d - \epsilon_b)^{p+1}}{2^p(p+1)}.$$

Proof III

Now we can write the following chain of inequalities:

$$\|\lambda_1\|_p^p \leq \frac{(\epsilon_d - \epsilon_b)^{p+1}}{2^p(p+1)} \overset{(2)}{\leq} \max_{1\leq i\leq N}(d^{p+1}(X^i,\mu)) = \max_{1\leq i\leq N}\Big(\sum_{j=1}^{d}(X_j^i-\mu)^2\Big)^{\frac{p+1}{2}}.$$

Inequality (2) comes directly from the fact that

$$(\epsilon_d - \epsilon_b) \leq 2\cdot \max_{1\leq i\leq N}(d(X^i,\mu)) = 2r.$$

In other words, the distance between two points of $\mathbb{X}$ inside the disc $D(\mu,r) = \{x\in\mathbb{R}^d \mid d(x,\mu)\leq r\}$ will never surpass $2r$.
Taking expected values,

$$\mathbb{E}[\|\lambda_1\|_p] \leq \mathbb{E}\Big[\max_{1\leq i\leq N}\Big(\sum_{j=1}^{d}(X_j^i-\mu)^2\Big)^{\frac{p+1}{2p}}\Big] \leq \mathbb{E}\Big[\sum_{i=1}^{N}\Big(\sum_{j=1}^{d}(X_j^i-\mu)^2\Big)^{\frac{p+1}{2p}}\Big].$$

Proof IV

From here, setting $\alpha := 2p/(p+1)$ one can prove directly the inequality

$$\mathbb{E}\Big[\sum_{i=1}^{N}\Big(\sum_{j=1}^{d}(X_j^i-\mu)^2\Big)^{\frac{1}{\alpha}}\Big] \leq N\cdot \mathrm{tr}(\Sigma)^{\frac{1}{\alpha}}.$$

Indeed, if $p=1$ then $\alpha=1$ and it follows directly. If $p>1$, then $\alpha>1$ and we can set $\beta>1$, such that $\frac{1}{\alpha}+\frac{1}{\beta}=1$. By the Hölder inequality we have that,

$$\mathbb{E}\Big[\sum_{i=1}^{N}\Big(\sum_{j=1}^{d}(X_j^i-\mu)^2\Big)^{\frac{1}{\alpha}}\Big] \leq N\cdot \Big(\mathbb{E}\Big[\sum_{j=1}^{d}(X_j^i-\mu)^2\Big]\Big)^{\frac{1}{\alpha}}\Big(\mathbb{E}[1^\beta]\Big)^{\frac{1}{\beta}}.$$

Proof V

The number of loops described by an N point data set $L_1(N)$, can be roughly bounded by the sum

$$L_1(N) \leq \sum_{i=4}^{i=N} \binom{N}{i},$$

where we consider the total number of loops to be bounded by grouping each different cluster of $i = 4, 5, \ldots, N$ points that can determine a loop; in the Vietoris-Rips scheme the minimum size of these clusters is 4, in Cech scheme is 3. This allows us to guarantee that $\sum_{k\geq 1} \|\lambda_k\|$ has a finite number of non zero terms $C(N)$. So we can write the following:

$$\mathbb{E}(\|\Lambda\|_p) = \mathbb{E}\Big(\sum_{i=1}^{C(N)} \|\lambda_i\|_p\Big) = \sum_{i=1}^{C(N)} \mathbb{E}(\|\lambda_i\|_p) \overset{(1)}{\leq} C(N) \cdot \mathbb{E}(\|\lambda_1\|_p) \leq C(N) \cdot N \cdot \mathrm{tr}(\Sigma)^{\frac{p+1}{2p}}.$$

Inequality (1) is due to the property $\lambda_k \geq \lambda_{k+1}$ mentioned before.

Proposition II

This describes the change of L^p-norms in response to scaling a point cloud by a certain factor h.

Proposition

Let X denote a finite point cloud in $\mathbb{R}^d$ with associated persistence landscape Λ_X, and $H\colon \mathbb{R}^d \to \mathbb{R}^d$ be a homotecy such that $H(X) = h \cdot X$. Then,

$$\|\Lambda_{H(X)}\|_p = h^{\frac{p+1}{p}} \cdot \|\Lambda_X\|_p,$$

where $\|\Lambda_{H(X)}\|_p$ denotes the p-norm of the persistence landscape associated to the scaled point cloud $H(X)$.

Proof I

Consider $\mathcal{D}$, $\mathcal{D}'$ to be the corresponding persistence diagrams for $X, H(X)$, respectively. Then, every $P_i \in \mathcal{D}$ becomes $hP_i \in \mathcal{D}'$.

Indeed, for any k-simplex $\sigma = [p_0, \ldots, p_k]$ in the Vietoris-Rips complex attached to X at parameter α, it has to be so that $d(p_i, p_j) \leq \alpha$ for all i, j. So, when we apply H we have that $\sigma = [p_0, \ldots, p_k]$ becomes $\sigma' = [hp_0, \ldots, hp_k]$ and if σ is formed at α, then σ' is formed at $h \cdot \alpha$.

Thus, $|\mathcal{D}| = |\mathcal{D}'|$ and every point in $\mathcal{D}$ is scaled by h in $\mathcal{D}'$.

Proof II

Furthermore, for every f_{P_i} (mentioned in Section 2.2) associated to the persistence landscape of X, we have that $f_{P_i} \mapsto f_{hP_i}$ and so, Λ_X becomes $h\Lambda_X$.

Recall that we denote $\Lambda_X(k, x) = \lambda_k(x)$ to be the k-th persistence landscape function of Λ_X, so we can write $\Lambda_{H(X)}(k, x) = h\Lambda_X(k, x) = \lambda_k^{'}(x)$.

Moreover, we have the associated domains in $\mathbb{R}^2$, $\Omega_k = \{(x, y) \mid x \in \mathrm{dom}(\lambda_k),\ \ y \leq \lambda_k(x)\}$ and for $H(X)$, the new scaled domain is $\Omega_k^{'} = \{(x, y) \mid x \in \mathrm{dom}(\lambda_k^{'}),\ \ y \leq \lambda_k^{'}(x)\}$.

Proof III

Then,

$$\|\Lambda_{H(X)}\|_p^p = \sum_{k\geq 1}\Big(\int_{\Omega_k'} 1\cdot d(x,y)\Big) = h^{p+1}\cdot\sum_{k\geq 1}\Big(\int_{\Omega_k} 1\cdot d(s,t)\Big) = h^{p+1}\cdot\|\Lambda_X\|_p^p.$$

We note that this relation results from a simple change of variables,

$$x = h\cdot s \to dx = h\cdot ds$$
$$y = h\cdot t \to dy = h^p\cdot dt.$$

Theorem I

Theorem

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X\colon \Omega \to \mathbb{R}^d$ denote a multivariate random variable with distribution $D(\mu, \Sigma)$, and let $X^1, \ldots, X^N$ be identical copies of X such that $\mathbb{X} := (X^1, \ldots, X^N)$ describes a random N-point data set in $\mathbb{R}^d$. Assume μ, Σ are finite and for a certain scaling factor h, $hD(\mu, \sigma^2) \sim D(h\mu, h^2\Sigma)$. Then,

$$\mathbb{E}(\|\Lambda_{h^2\Sigma}(\omega)\|_p) = h^{\frac{p+1}{p}} \cdot \mathbb{E}(\|\Lambda_{\Sigma}(\omega)\|_p),$$

where $\Lambda_{h^2\Sigma}$, Λ_{Σ} denote the persistence landscapes for random point clouds with corresponding variance-covariance matrix $h^2\Sigma$ and Σ, respectively.

Proof I

Let H denote a homotecy such that $H\colon \mathbb{R}^d \to \mathbb{R}^d$, so $\mathbb{X} \mapsto h \cdot \mathbb{X}$. From the assumptions made, we can safely deduce that for every point $X^i \in \mathbb{X}$ we have $H(X^i) = h \cdot X^i \sim D(\mu, h^2\Sigma)$.

We use indistinguishably $\Lambda_{\mathbb{X}}(\omega) = \Lambda_{\Sigma}(\omega)$ and $\Lambda_{h\cdot\mathbb{X}}(\omega) = \Lambda_{h^2\Sigma}(\omega)$.

Let $\mathbb{X}_1, \ldots, \mathbb{X}_n$ denote independent identically distributed copies of $\mathbb{X}$, and let $\Lambda^1_{\mathbb{X}}, \ldots, \Lambda^n_{\mathbb{X}}$ be the corresponding persistence landscapes. Proposition 2 implies that

$$\frac{1}{n}\sum_{i=1}^{n} \|\Lambda^i_{h^2\Sigma}(\omega)\|_p = h^{\frac{p+1}{p}} \frac{1}{n}\sum_{i=1}^{n} \|\Lambda^i_{\Sigma}(\omega)\|_p, \tag{1}$$

since for every landscape Λ^i_{Σ} we have the corresponding *scaled* persistence landscape $\Lambda^i_{h^2\Sigma}$.

Proof II

From Proposition 1 we have that the expected values $\mathbb{E}(\|\Lambda^i_\Sigma\|_p)$, $\mathbb{E}(\|\Lambda^i_{h^2\Sigma}\|_p)$ are finite. Hence, we can apply the Strong Law of Large Numbers, and we obtain

$$\frac{1}{n}\sum_{i=1}^{n}\|\Lambda^i_{h^2\Sigma}(\omega)\|_p \longrightarrow \mathbb{E}(\|\Lambda_{h^2\Sigma}\|_p) \text{ a.s.}$$

$$h^{\frac{p+1}{p}}\frac{1}{n}\sum_{i=1}^{n}\|\Lambda^i_{\Sigma}(\omega)\|_p \longrightarrow h^{\frac{p+1}{p}}\cdot\mathbb{E}(\|\Lambda_{\Sigma}\|_p) \text{ a.s.}$$

So, taking limits in (1), we obtain

$$\mathbb{E}(\|\Lambda_{h^2\Sigma}\|_p) = h^{\frac{p+1}{p}}\cdot\mathbb{E}(\|\Lambda_{\Sigma}\|_p).$$

Theorem II

Theorem

Let (X, Y) be a 2D-random vector sampled from a generic bi-variate distribution $\mathcal{D}(\mu, \Sigma)$ with finite mean-vector $\mu = (\mu_x, \mu_y)$ and finite variance-covariance matrix Σ:

$$\Sigma = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_y\sigma_x & \sigma_y^2 \end{pmatrix}.$$

Here, $\rho \in (-1, 1)$ is the correlation coefficient. Assume $\sigma_x \geq \sigma_y$. Let θ_1 and θ_2 be the eigenvalues of the matrix Σ. Since Σ is symmetric and positive semi-definite, we have $\theta_1 \geq \theta_2 \geq 0$. For a fixed number of points N, consider the associated point cloud $\mathbb{X}$ and the corresponding persistence landscape Λ. Then, we have

$$\mathrm{E}[\|\Lambda\|_1] \leq L_1(N)^2 \cdot N \cdot \theta_2.$$

Proof I

Without loss of generality, using Principal Component Analysis, we can write $\mathbb{X}$ as a centered point cloud such that the points (X_i, Y_i), for $i = 1, \ldots, N$, are sampled from a bivariate generic distribution $D(0, \Theta)$ where

$$\Theta = \begin{pmatrix} \theta_1 & 0 \\ 0 & \theta_2 \end{pmatrix}.$$

The new covariance structure allows to interpret the X and Y axis as the axis of maximum and minimum variability, respectively. Now, we have that both variables are uncorrelated with the variance-covariance structure Θ, while the underlying topology of $\mathbb{X}$ remains unchanged.

Furthermore, we can encapsulate $\mathbb{X}$ into a rectangle of basis $2 \max_i |X_i|$ and height $2 \max_i |Y_i|$ with $\max_i |Y_i| \leq \max_i |X_i|$.

Proof II

The main idea is the following. Consider a rectangle of four points in the vertices. Let α the basis and β the height, where $\alpha > \beta$. The loop of this four points starts at α and dies at $\sqrt{\alpha^2 + \beta^2}$. Then, the life of the $1D$ homology of this set of points is given by

$$\epsilon_d - \epsilon_b \leq \sqrt{\alpha^2 + \beta^2} - \alpha \leq \beta.$$

Finally it is clear that for any set of points of the point cloud $\beta \leq 2 \max_i |Y_i|$.

The idea can be generalized to any subset of points of the point cloud.

Proof III

Recall now that for each point (b_i, d_i) in the persistence diagram, we have an associated *triangle* in the first persistence landscape function with the area given by $(d_i - b_i)^2/4$. Hence, we can bound $\|\lambda_1\|_1$ of $\mathbb{X}$ as follows:

$$\|\lambda_1\|_1 \leq L_1(N) \cdot \max_i \frac{(d_i - b_i)^2}{4} \leq L_1(N) \cdot \max_i |Y_i|^2 \leq L_1(N) \cdot \Sigma_{i=1}^N |Y_i|^2.$$

From our proof of Proposition 1 we have that

$$\|\Lambda\|_1 \leq L_1(N) \cdot \|\lambda_1\|_1,$$

so taking expectations in the inequality above proves the result.

Remark I

Note that from Theorem 2 it is easy to see that as $\rho \to \pm 1$, the expected value $\mathrm{E}[\|\Lambda\|_1] \to 0$. Indeed, the eigenvalues of Σ are:

$$\theta_1 = \frac{\sigma_x^2 + \sigma_y^2}{2} + \sqrt{\left(\frac{\sigma_x^2 + \sigma_y^2}{2}\right)^2 - (1-\rho^2)\sigma_x^2\sigma_y^2},$$

$$\theta_2 = \frac{\sigma_x^2 + \sigma_y^2}{2} - \sqrt{\left(\frac{\sigma_x^2 + \sigma_y^2}{2}\right)^2 - (1-\rho^2)\sigma_x^2\sigma_y^2}.$$

Thus, if $\rho = 0$, the eigenvalues are σ_x^2, σ_y^2, and if $\rho = \pm 1$, we obtain $\sigma_x^2 + \sigma_y^2$ and 0.

Remark II

- The boundary on $E[\|\Lambda\|_1]$ that follows from Theorem 2 is rough. Nonetheless, its dependency on θ_2 points to suppression of PH in systems with strong covariance.
- If we have a point cloud $\mathbb{X}$ from an embedding of an $d-$dimensional random vector in $\mathbb{R}^d$, we have an analogous covariance matrix which is a diagonal matrix composed by $\theta_1 \geq \cdots \geq \theta_d \geq 0$ eigenvalues. Similarly to the 2-dimensional case, the persistence of loops is controlled by θ_2 and the results of Theorem 2 is still valid.

Remark III

Theorem 2 can be extended to h-dimensional cycles for any $h \geq 0$. For any $h \geq 0$, the persistence of the h-cycles is controlled by the $(h+1)$-th eigenvalue, provided $0 \leq h \leq m-1$. That is, connected components are bounded by $L_0(N)^2 N\theta_1$, loops are bounded by $L_1(N)^2 N\theta_2$, cavities are bounded by $L_2(N)^2 N\theta_3$ and so on. Here,

$$L_r(N) = \sum_{i=2^{h+1}}^{i=N} \binom{N}{i} \leq 2^N.$$

This formula is for Vietoris-Rips scheme. In the Cech case, the sum starts a little before. In any case, the bound 2^N is universal.

Uncorrelated bivariate Normal and Gamma distributions I

To illustrate the dependency described in Theorem 1, we use Monte Carlo simulations. We construct a random point cloud in the $2D$ metric space by sampling 50 points from the bi-variate Normal and bi-variate Gamma distributions. Each distribution has a correlation matrix

$$\Sigma_i = \sigma_i^2 Id \quad (i = 1, 2),$$

so that points have i.i.d coordinates.

These distributions have bounded first and second moments and different symmetry, which allows us to illustrate the growth of the mean of L^p-norms with rising variability of underlying distributions, described by Theorem 1.

Uncorrelated bivariate Normal and Gamma distributions II

- For each realization of the generated data set we associate the corresponding Rips filtration, compute the persistence diagram, and the corresponding L^p-norm. We collect the values of L^p-norms obtained at each realization and calculate their mean-values at the end of the first simulation.
- We run these simulations 10 times, sequentially increasing the variance of each distribution. For the uncorrelated bivariate Normal distribution, we increase $\sigma_1 = \sigma_2$ from 1 to 10. For the uncorrelated bi-variate Gamma distribution we chose the shape parameter $k1 = k2 = 2$, and increase $\sigma_1 = \sigma_2$ from $\sqrt{2}$ to $10\sqrt{2}$. Next Figure shows the predicted behavior of the mean of L^p-norms ($p = 1, 2, 3$) for a relatively low number of iterations (1000).

Uncorrelated bivariate Normal and Gamma distributions III

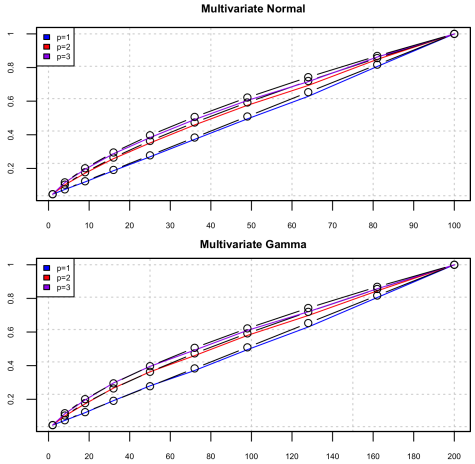

Figure: (Color online) Plots of Monte Carlo simulations, 1000 realizations, of the dependency of L^p-norms on the variance of uncorrelated bi-variate Normal and Gamma distributed data sets. Dots depict the mean-values of p-norms computed for each value of the variance. Colored lines

Uncorrelated bivariate Normal and Gamma distributions IV

Thus, one could expect to find an increase in values of L^1-norms whenever the variability in the underlying system is growing. However, this expectation is not always supported by observations. To qualitatively explain this anomaly, we conduct numeric experiments with point clouds sampled from the bivariate Normal distribution, which reveal that increase of L^1-norms due to rising variability can be suppressed by growing cross-correlation in the system.

Correlated bivariate Normal distribution I

- Theorem 1 assumes there is no correlation in the underlying point cloud. In this section we perform Monte Carlo simulations of the behavior of L^1-norms of persistence topological landscapes derived from $2D$ point clouds, which are sampled from the correlated bivariate Normal distribution.
- The choice of this model is motivated by simplicity of its correlation structure. Specifically, we wish to test how a growing correlation coefficient ρ of two white noises influences values of L^1-norms.
- We use the mvrnorm function from the R-package MASS to sample 50 random data points from the bivariate Normal distribution $\mathcal{N}(\mu, \Sigma)$, where μ and Σ as given in Theorem 2.

Correlated bivariate Normal distribution II

- Thereby, we construct a random point cloud in the $2D$ metric space. We repeat this procedure multiple times. For each realization of the generated data set we associate the corresponding Rips filtration, compute the persistence diagram, and the L^1-norm of persistence landscape. We collect the values of L^1-norms obtained at each realization and calculate their mean-values at the end of the first simulation.
- We run these simulations 10 times, first with $\sigma_x = \sigma_y = 1$ and correlation factor ρ sequentially increased from 9.9% to 99.0%. In this case we find that the average values of the L^1-norm converge with the number of repetitions towards a smoothly decreasing function of the correlation coefficient.

Correlated bivariate Normal distribution III

- In the second experiment, we keep $\sigma_x = 1$ and increase σ_y linearly from 0.3 to 3 in 10 steps, while simultaneously increasing ρ from 9.9% to 99.0%.
- Finally, we run the experiment with $\sigma_x = \sigma_y$, both increasing linearly from 0.3 to 3 in 10 steps, simultaneously with ρ increasing from 9.9% to 99.0%.
- For simplicity, in all experiments we set $\mu = 0$. Next Figure demonstrates the outcome of these experiments.

Correlated bivariate Normal distribution IV

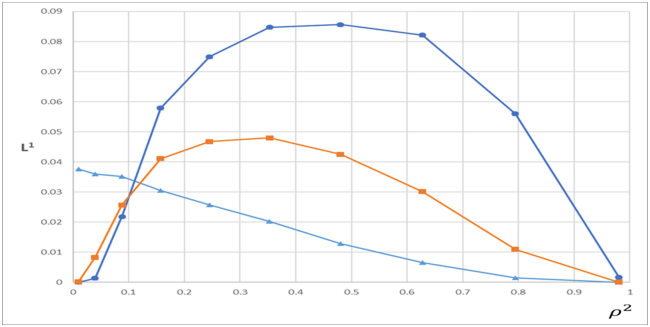

Figure: Dependency of the mean of L^1-norm, obtained with $10,000$ repetitions, on the square of the correlation strength. The line marked with triangles represents the case with standard deviations $\sigma_x = \sigma_y = 1$. The line marked with filled squares represents the case with $\sigma_x = 1$ and σ_y growing linearly and simultaneously with ρ^2 in 10 steps from 0.3 to 3. The line marked with filled circles is derived in the case with both σ_x and σ_y linearly growing in 10 steps from 0.3 to 3.

Empirical results

- To be applied, TDA requires an embedding of the data set into some Euclidean, metric space. Any portfolio-like mixture of financial time series naturally defines the dimensionality of such a space, whereas the size of the sliding window determines the size of a point cloud.
- Here, we use a relatively short window (50 trading days) to explore the evolution of daily log-returns of four US and four European equity indices.
- Consequently, we derive the transient topological summaries of two ordered in time sets of $4D$ point clouds, with 50 data points each, one for US and one for Europe.
- Specifically, we compute the time evolution of L^p-norms of persistence landscapes derived via Rips filtration from these point clouds. For details of the time-resolved TDA methodology see Gidea and Katz (2018).

- We use Yahoo Finance API to obtain time series of four broad US equity indices – DJI, S&P 500, NASDAQ, and Russell 2000 – as well as four European indices – FTSE 100 (UK), DAX 30 (Germany), CAC 40 (France), IBEX 35 (Spain) - between January 1, 1998 and April 24, 2020. We drop 3.9% of the corresponding data points to synchronize trading dates on different European markets.
- We study the stationary time series formed by the daily log-returns of these indices, $ln(p_{i,t+1}/p_{i,t})$, where i identifies the index and t determines the trading day, and group them into two portfolio-like mixtures - one with four US indices, another with four European indices.

- We use a sliding window of 50 trading days with the sliding step set to one day to derive the temporal changes in covariance and cross-correlation matrices, respectively.
- We use the same sliding window technique to obtain two ordered in time sets of $4D$ point clouds, each consisting of 50 points describing the temporal changes of daily log-returns of US and European indices; we denote them by X_n^i, where $i = 1, 2$ determines the geographical portfolio and n is the size of the sliding window.
- In our analysis we consider the persistence of $1D$ loops exclusively. This is due to lengthy computations of Vietoris-Rips complexes for higher dimensions.

- The algorithm computes the corresponding Rips filtration $R(X_n^i, \epsilon)$, $\epsilon > 0$, the persistence diagram $\mathcal{D}_1(X_n^i)$, the persistent landscape $\Lambda(X_n^i)$, and the relevant functional L^1-norm $\|\Lambda(X_n^i)\|_1$ per point cloud.
- In all computations, we use the R package TDA and C++ library 'GUDHI' (Fasy et al. (2015); GUDHI software: `http://gudhi.gforge.inria.fr/`).

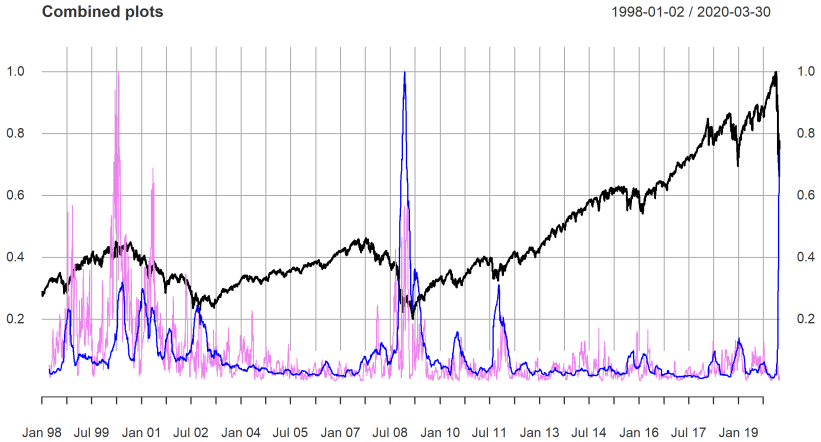

Figure: (Color online) Combined time series of S&P 500 (black line), L^1-norm (purple line) and estimated variability (blue line) of the mixture of daily log-returns of four indices; see text for details.

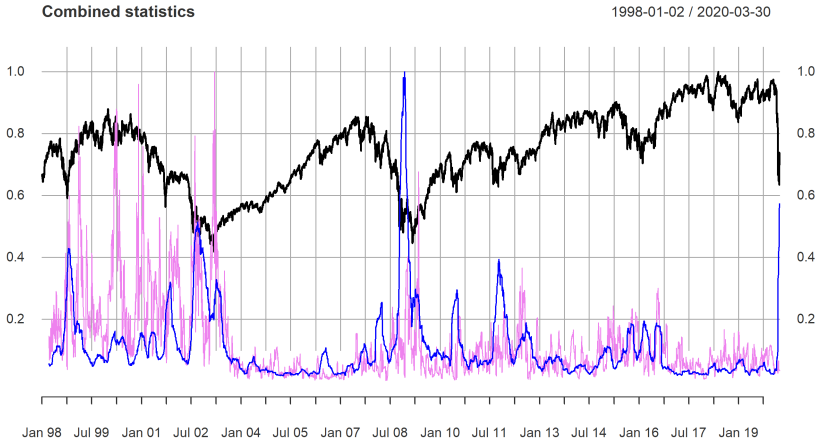

Figure: (Color online) Time series of FTSE 100 (black line), L^1-norm (purple line) and estimated variability (blue line) of the mixture of daily log-returns of four indices; see text for details.

- The striking differences in time series of L^1-norms at the *early stages* of the Technology Crash of 2000, the Global Financial Crash of 2008-2009, and the latest one, are clearly visible on these figures.
- First, notice that spikes in variability during two prior systemic market meltdowns were always accompanied by jumps in L^1-norms. Remarkably, this behavior is not observed at the beginning of COVID-19 crisis.
- Another noticeable observation is related to exceptionally strong cross-correlations between indices, which are above 95% for all of them, throughout the early stage of the pandemic crisis.
- This finding reflects the fundamentally different character of the current global financial crisis, which is triggered by a strong exogenous shock - COVID-19 pandemic.

Conclusions I

- It is widely recognized that market instability and growing variability at early and later stages of the Technology Crash of 2000 and the Global Financial Crisis of 2008 - 2009 were due to *endogenous* economic forces.
- These regime-changes are reflected by spikes in values of functional norms of persistence landscapes derived from the corresponding point clouds, formed by the 4D mixtures of daily log-returns of major US and European indices.
- Informally, stationarity of daily log-returns allows to use the ergodic argument and qualitatively explain this behavior as the real-world manifestation of the theorem which established proportionality of the mean of L^1-norms of persistence landscapes to the variance of the underlying probability distribution.

Conclusions II

- Noticeably, we do not find the similar pattern at the very beginning of the latest market meltdown. Despite the growing variability in the global financial system, there are no spikes in the time series of L^1-norms.
- As we have shown here, qualitatively, this puzzling behavior can be explained by the unusually strong covariance observed in the global financial system. The latter is related to the *exogenous* shock, caused by the COVID-19 pandemic, which synchronized the time series forming the point clouds. A simultaneous drop of global equity markets translates into an isometry, nullifying persistence homology of a noisy financial system.

Conclusions III

- To summarize, the study of the statistical properties of norms of persistence landscapes of points clouds generated by a broad class of random distributions improve the empirical analysis of financial time series (for example, daily log-returns), and helps to distinguish between endogenous and exogenous economic shocks.
- In particular we have seen that these topological features are very sensitive to an interplay between variability and covariance in a system.

References

- GUNAR CARLSSON (2009): *Topology and data*. Bulletin of the American Mathematical Society 46 (2):255–308.
- HERBERT EDELSBRUNNER AND JOHN HARER (2009): *Computational Topology: An Introduction*. American Mathematical Society.
- PETER BUBENIK (2015): *Statistical topological data analysis using persistence landscapes*. Journal of Machine Learning Research 16 (1):77–102.
- MARIAN GIDEA AND YURI KATZ (2018): *Topological data analysis of financial time series: landscapes of crashes*. Physica A, 491 (1):820–834.

The End

Thank you for the attention

Gràcies