Journal of Machine Learning Research Volume 55 (2016) 1–11 Submitted submitted; Published published

Exploring and measuring non-linear correlations: Copulas, Lightspeed Transportation and Clustering

Gautier Marti GMARTI@HELLEBORECAPITAL.COM
Hellebore Capital Ltd
Ecole Polytechnique

Sébastien Andler SEBASTIEN.ANDLER@ENS-LYON.FR
ENS de Lyon
Hellebore Capital Ltd

Frank Nielsen NIELSEN@LIX.POLYTECHNIQUE.FR
Ecole Polytechnique

Philippe Donnat PDONNAT@HELLEBORECAPITAL.COM
Hellebore Capital Ltd

Editor: Oren Anava, Marco Cuturi, Azadeh Khaleghi, Vitaly Kuznetsov, Alexander Rakhlin

Abstract

We propose a methodology to explore and measure the pairwise correlations that exist between variables in a dataset. The methodology leverages copulas for encoding dependence between two variables, state-of-the-art optimal transport for providing a relevant geometry to the copulas, and clustering for summarizing the main dependence patterns found between the variables. Some of the clusters centers can be used to parameterize a novel dependence coefficient which can target or forget specific dependence patterns. Finally, we illustrate the methodology with financial time series (credit default swaps, stocks, foreign exchange rates). Code and numerical experiments are available online at `https://www.datagrapple.com/Tech` for reproducible research.

Keywords: Correlations, Copulas, Regularized Optimal Transport, Financial Time Series

1. Introduction

Pearson's correlation coefficient which estimates linear dependence between two variables is still the mainstream tool for measuring variable correlations in science and engineering. However, its shortcomings are well-documented in the statistics literature: not robust to outliers; not invariant to monotone transformations of the variables; can take value 0 whereas variables are strongly dependent; only relevant when variables are jointly normally distributed. A large but under-exploited literature in statistics and machine learning has expanded recently to alleviate these issues (Reshef et al., 2011; Székely et al., 2009). An underlying idea to many of the dependence coefficients is to compute a distance $D(P(X,Y), P(X)P(Y))$ between the joint distribution $P(X,Y)$ of variables X, Y and $P(X)P(Y)$ the product of marginal distributions encoding the independence. For example, choosing $D = \mathrm{KL}$ (Kullback-Leibler divergence), we end up with the Mutual Information (MI) measure, well-known in information theory. Thus, one can detect all the dependences between X and Y since the distance will be greater than 0 as soon as $P(X,Y)$ is different from $P(X)P(Y)$. Then, the dependence literature focus has shifted toward the new concept of "equitability" (Kinney and Atwal, 2014): How can one quantify the strength of a statistical association between two variables without bias for relationships of a specific form? Many researchers now aim at designing and proving that their proposed measures are indeed equitable (Ding and Li, 2013; Chang et al., 2016). This is *not* what we look for in this article. But, on the contrary, we want to target specific dependence patterns and ignore others. We want to target dependence which are relevant to such or such problem, and forget about the dependence which are not in the scope of the problems at hand, or even worse which may be spurious associations (pure chance or artifacts in the data). The latter will be detected with an equitable dependence measure since they are deviation from independence, and will be given as much weight as the interesting ones. Rather than using the biases for specific dependence of several coefficients, we propose a dependence coefficient that can be parameterized by a set of *target-dependences*, and a set of *forget-dependences*. Sets of target and forget dependences can be built using expert hypotheses, or by leveraging the centers of clusters resulting from an exploratory clustering of the pairwise dependences. To achieve this goal, we will leverage three tools: copulas, optimal transportation, and clustering. Whereas clustering, the task of grouping a set of objects in such a way that objects in the same group (also called cluster) are more similar to each other than those in different groups, is common knowledge in the machine learning community, copulas and optimal transportation are not yet mainstream tools. Copulas have recently gained attention in machine learning, and several copula-based dependence measures have been proposed for improving feature selection methods (Ghahramani et al., 2012; Lopez-Paz et al., 2013; Chang et al., 2016). Optimal transport may be more familiar to computer scientists working in computer vision since it is the underlying theory of the Earth Mover's Distance (Rubner et al., 2000). Until very recently, optimal transportation distances between distributions were not deemed relevant for machine learning applications since the best computational cost known was super-cubic to the number of bins used for discretizing the distribution supports which grows itself exponentially with the dimension. A mere distance evaluation could take several seconds! In this article, we leverage recent computational breakthroughs detailed in (Cuturi, 2013) which make their use practical in machine learning. We demonstrate it by studying a comprehensive dataset of financial time series. To capture their associations, most quantitative approaches make use of an estimated covariance or correlation matrix (a mixed information of linear dependence perturbed by the possibly heavy-tailed marginals) or a Gaussian copula (which only captures the linear dependence while factoring out properly the marginals). This may have dramatic effect on subsequent analysis: (Marti et al., 2016b) for an example with financial time series clustering. Using the proposed methodology, we can explore the dependence between these time series.

2. Background on Copulas and Optimal Transport

2.1 Copulas

Copulas are functions that couple multivariate distribution functions to their univariate marginal distribution functions. In this article, we will only consider bivariate copulas, but most of the results and the methodology presented hold in the multivariate setting, at the cost of a much higher computational burden which is for now a bit unrealistic.

Theorem 1 (Sklar's Theorem) *Let* $X = (X_i, X_j)$ *be a random vector with a joint cumulative distribution function* F*, and having continuous marginal cumulative distribution functions* F_i, F_j *respectively. Then, there exists a unique distribution* C *such that* $F(X_i, X_j) = C(F_i(X_i), F_j(X_j))$*.* C*, the copula of* X*, is the bivariate distribution of uniform marginals* $U_i, U_j := F_i(X_i), F_j(X_j)$*.*

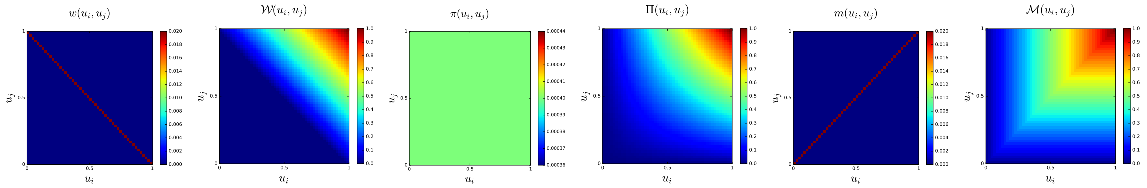


Figure 1: Copulas measure (left column) and cumulative distribution function (right column) heatmaps for negative dependence (first row), independence (second row), i.e. the uniform distribution over $[0, 1]^2$, and positive dependence (third row)

Copulas are central for studying the dependence between random variables: their uniform marginals jointly encode all the dependence. They allow to study scale-free measures of dependence and are *invariant to monotonous transformations of the variables*. Some copulas play a major role in the measure of dependence, namely $\mathcal{W}$ and $\mathcal{M}$ the Fréchet-Hoeffding copula bounds, and the independence copula $\Pi(u_i, u_j) = u_i u_j$ (depicted in Figure 1).

Proposition 1 (Fréchet-Hoeffding copula bounds) *For any copula* $C : [0,1]^2 \to [0,1]$ *and any* $(u_i, u_j) \in [0,1]^2$ *the following bounds hold:*

$$\mathcal{W}(u_i, u_j) \leq C(u_i, u_j) \leq \mathcal{M}(u_i, u_j), \tag{1}$$

where $\mathcal{W}(u_i, u_j) = \max\{u_i + u_j - 1, 0\}$ *is the copula for countermonotonic random variables and* $\mathcal{M}(u_i, u_j) = \min\{u_i, u_j\}$ *is the copula for comonotonic random variables.*

Many correlation coefficients can actually be expressed as a distance between the data copula and one of these reference copulas. For example, the Spearman (rank) correlation ρ_S which is usually understood as $\rho_S(X_i, X_j) = \rho(F_i(X_i), F_j(X_j))$, i.e. the linear dependence of the probability integral transformed variables (rank-transformed data), can also be viewed as an average distance between the copula C of (X_i, X_j) and the independence copula Π: $\rho_S(X_i, X_j) = 12 \int\int_{[0,1]^2} (C(u_i, u_j) - u_i u_j)\, du_i du_j$. Moreover, since $|u_i - u_j|/\sqrt{2}$ is the distance between point (u_i, u_j) to the diagonal (the measure of the positive dependence copula), one can rewrite $\rho_S(X_i, X_j) = 12 \int\int_{[0,1]^2} (C(u_i, u_j) - u_i u_j)\, du_i du_j = 12 \int\int_{[0,1]^2} u_i u_j dC(u_i, u_j) - 3 = 1 - 6 \int\int_{[0,1]^2} (u_i - u_j)^2 dC(u_i, u_j)$. Thus, Spearman correlation can also be viewed as measuring a deviation from the monotonically increasing dependence to the data copula using a quadratic distance. *We will leverage this idea to propose our dependence-parameterized dependence coefficient.*

Notice that when working with empirical data, we do not know a priori the margins F_i for applying the probability integral transform $U_i := F_i(X_i)$. Deheuvels in (Deheuvels, 1979) has introduced a practical estimator for the uniform margins and the underlying copula, the empirical copula transform.

Definition 1 (Empirical Copula Transform) *Let* (X_i^t, X_j^t), $t = 1, \ldots, T$, *be* T *observations from a random vector* (X_i, X_j) *with continuous margins. Since one cannot directly obtain the corresponding copula observations* $(U_i^t, U_j^t) := (F_i(X_i^t), F_j(X_j^t))$, *where* $t = 1, \ldots, T$, *without knowing a priori* F_i, *one can instead estimate the empirical margins* $F_i^T(x) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}(X_i^t \leq x)$, *to* *obtain the* T *empirical observations* $(\tilde{U_i^t}, \tilde{U_j^t}) := (F_i^T(X_i^t), F_j^T(X_j^t))$. *Equivalently, since* $\tilde{U_i^t} = R_i^t/T$, R_i^t *being the rank of observation* X_i^t, *the empirical copula transform can be considered as the normalized rank transform.*

Notice that the empirical copula transform is fast to compute, sorting arrays of length T can be done in $O(T \log T)$, consistent and converges fast to the underlying copula (Ghahramani et al., 2012).

As motivated in the introduction, we want to compare and summarize the pairwise empirical dependence structure (empirical bivariate copulas) of many variables. This brings the following questions: How can we compare two such copulas? What is a relevant representative of a set of empirical copulas? Which geometries are relevant for clustering these empirical distributions, and which are not?

2.2 Optimal Transport

In (Marti et al., 2016a), authors illustrate in a parametric setting using Gaussian copulas that common divergences (such as Kullback-Leibler, Jeffreys, Hellinger, Bhattacharyya) are not relevant for clustering these distributions, especially when dependence is high. These information divergences are only defined for absolutely continuous measures whereas some copulas have no density (e.g. the one for positive dependence). In practice, when working with frequency histograms, it gets worse: One has to pre-process the empirical measures with a kernel density estimator before computing these divergences. On the contrary, optimal transport distances are well-defined for both discrete (e.g. empirical) and continuous measures.

The idea of optimal transport is intuitive. It was first formulated by Gaspard Monge in 1781 as a problem to efficiently level the ground: Given that work is measured by the distance multiplied by the amount of dirt displaced, what is the minimum amount of work required to level the ground? Optimal transport plans and distances give the answer to this problem.

In practice, empirical distributions can be represented by histograms. We follow notations from (Cuturi, 2013). Let r, c be two histograms in the probability simplex $\Sigma_m = \{x \in \mathbb{R}_+^m : x^\top 1_m = 1\}$. Let $U(r,c) = \{P \in \mathbb{R}_+^{m \times m} \mid P1_m = r, P^\top 1_m = c\}$ be the transportation polytope of r and c, that is the set containing all possible transport plans between r and c.

Definition 2 (Optimal Transport) *Given a* $m \times m$ *cost matrix* M, *the cost of mapping* r *to* c *using a transportation matrix* P *can be quantified as* $\langle P, M \rangle_F$, *where* $\langle \cdot, \cdot \rangle_F$ *is the Frobenius dot-product. The optimal transport between* r *and* c *given transportation cost* M *is thus:*

$$d_M(r,c) := \min_{P \in U(r,c)} \langle P, M \rangle_F. \tag{2}$$

Whenever M belongs to the cone of distance matrices, the optimum of the transportation problem $d_M(r,c)$ is itself a distance.

Lightspeed transportation. Optimal transport distances suffer from a computational burden scaling in $O(m^3 \log m)$ which has prevented their widespread use in machine learning: A mere distance computation between two high-dimensional histograms can take several seconds. In (Cuturi, 2013), Cuturi provides a solution to this problem: He restrains the polytope $U(r,c)$ of all possible transport plans between r and c to a Kullback-Leibler ball $U_\alpha(r,c) \subset U(r,c)$, where $U_\alpha(r,c) = \{P \in U(r,c) \mid \mathrm{KL}(P \| rc^\top) \leq \alpha\}$. He then shows that it amounts to perform an

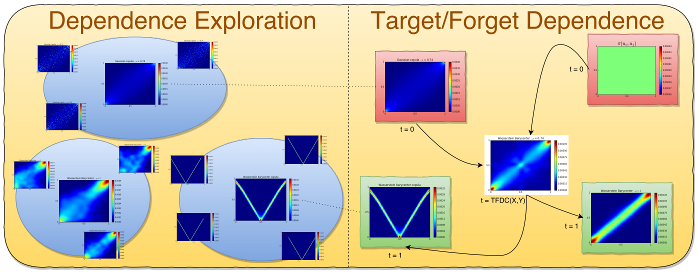


Figure 2: Exploration (left panel) and measure (right panel) of non-linear correlations. Exploration consists in finding clusters of similar copulas, visualizing their centroids, and eventually using them to assess the dependence of given variables represented by their copula

entropic regularization (recently generalized to many more regularizers in (Muzellec et al., 2016; Dessein et al., 2016)) of the optimal transportation problem whose solution is smoother and less deterministic. The regularized optimal transportation problem is now strictly convex, and can be solved efficiently using the Sinkhorn-Knopp iterative algorithm which exhibits linear convergence. Its solution is the Sinkhorn distance (Cuturi, 2013):

$$d_{M,\alpha}(r,c) := \min_{P \in U_\alpha(r,c)} \langle P, M \rangle_F, \tag{3}$$

and its dual $d_M^\lambda(r,c)$: $\forall \alpha > 0, \exists \lambda > 0$,

$$d_{M,\alpha}(r,c) = d_M^\lambda(r,c) := \langle P^\lambda, M \rangle_F, \tag{4}$$

where $P^\lambda = \mathrm{argmin}_{P \in U(r,c)} \langle P, M \rangle_F - \frac{1}{\lambda} h(P)$, and h is the entropy function.

In the following, we will leverage the dual-Sinkhorn distances for comparing, clustering and computing the clusters centers (Cuturi and Doucet, 2014) of a set of copulas at full speed.

3. A methodology to explore and measure non-linear correlations

We propose an approach to explore and measure non-linear correlations between N variables X_1, …, X_N in a dataset. These N variables can be, for instance, time series or features. The methodology presented (which is summarized in Figure 2) is twofold, and consists of: (i) an exploratory part of the pairwise dependence between variables, (ii) the parameterization and use of a novel dependence coefficient.

3.1 Using transportation of copulas as a measure of correlations

In this section, we leverage and extend the idea presented in our short introduction to copulas: correlation coefficients can be viewed as a distance between the data-copula and the Fréchet-Hoeffding bounds or the independence copula. The distance involved is usually an ℓ_p Minkowski metric distance. In the following, we will:

- replace the ℓ_p distance by an optimal transport distance between measures,

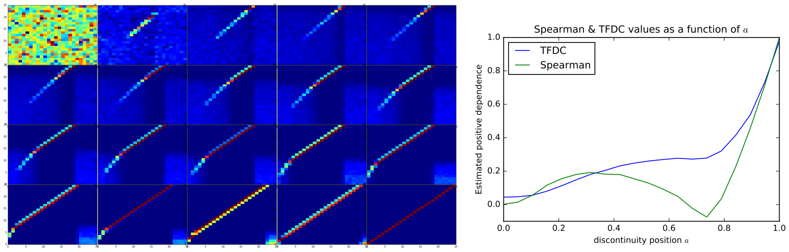


Figure 3: Empirical copulas for (X, Y) where $X = Z\mathbf{1}_{Z<a} + \epsilon_X \mathbf{1}_{Z>a}$, $Y = Z\mathbf{1}_{Z<a+0.25} + \epsilon_Y \mathbf{1}_{Z>a+0.25}$, $a = 0, 0.05, \ldots, 0.95, 1$, and where Z is uniform on $[0, 1]$ and ϵ_X, ϵ_Y are independent noises (left figure). Top left is an empirical copula for independence ($a = 0$), bottom right is the copula for perfect positive dependence ($a = 1$). Parameter a is increasing from top to bottom, and from left to right; TFDC and Spearman coefficients estimated between X and Y as a function of a (right figure). For $a = 0.75$, Spearman coefficient yields a negative value, yet $X = Y$ over $[0, a]$

- parameterize a dependence coefficient with other copulas than the Fréchet-Hoeffding bounds or the independence one.

Using the optimal transport distance between copulas, we now propose a dependence coefficient which is parameterized by two sets of copulas: *target* copulas and *forget* copulas.

Definition 3 (Target/Forget Dependence Coefficient) *Let $\{C_l^-\}_l$ be the set of forget-dependence copulas. Let $\{C_k^+\}_k$ be the set of target-dependence copulas. Let C be the copula of (X_i, X_j). Let d_M be an optimal transport distance parameterized by a ground metric M. We define the Target/Forget Dependence Coefficient as:*
$\mathrm{TFDC}\left(X_i, X_j; \{C_k^+\}_k, \{C_l^-\}_l\right) :=$

$$\frac{\min_l d_M(C_l^-, C)}{\min_l d_M(C_l^-, C) + \min_k d_M(C, C_k^+)} \in [0, 1]. \tag{5}$$

Using this definition, we obtain: $\mathrm{TFDC}\left(X_i, X_j; \{C_k^+\}_k, \{C_l^-\}_l\right) = 0 \Leftrightarrow C \in \{C_l^-\}_l$, $\mathrm{TFDC}\left(X_i, X_j; \{C_k^+\}_k, \{C_l^-\}_l\right) = 1 \Leftrightarrow C \in \{C_k^+\}_k$.

Example. A standard correlation coefficient can be obtained by setting the forget-dependence set to the independence copula, and the target-dependence set to the Fréchet-Hoeffding bounds. How does it compare to the Spearman correlation? In Figure 3, we display how the two coefficients behave on a simple numerical experiment: $X = Z\mathbf{1}_{Z<a} + \epsilon_X \mathbf{1}_{Z>a}$, $Y = Z\mathbf{1}_{Z<a+0.25} + \epsilon_Y \mathbf{1}_{Z>a+0.25}$, where Z is uniform on $[0, 1]$ and ϵ_X, ϵ_Y are independent noises. That is $X = Y$ over $[0, a]$. Notice that for $a = 0.75$, Spearman coefficient takes a negative value. We may thus prefer the monotonically increasing behaviour of the TFDC to the Spearman one.

3.2 How to choose, design and build targets?

We now propose two alternatives for choosing, designing and building the *target* and *forget* copulas: an exploratory data-driven approach and an hypotheses testing approach.

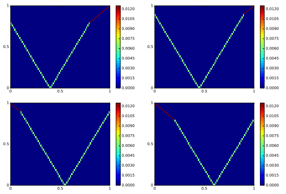
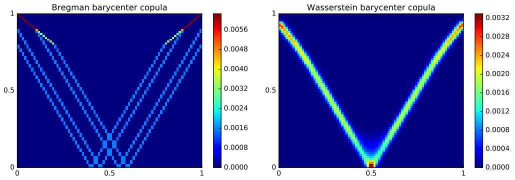


Figure 4:
4 copulas describing the dependence between $X \sim \mathcal{U}([0,1])$ and $Y \sim (X \pm \epsilon_i)^2$, where ϵ_i is a constant noise specific for each distribution. X and Y are countermonotonic (more or less) half of the time, and comonotonic (more or less) half of the time



Bregman barycenter copula

Wasserstein barycenter copula

Figure 5:
Barycenter of the 4 copulas from Figure 4 for: (left) Bregman geometry (Banerjee et al., 2005) (which includes, for example, squared Euclidean and Kullback-Leibler distances); (right) Wasserstein geometry.

3.2.1 DATA-DRIVEN: CLUSTERING OF COPULAS

Assume we have N variables $X_1, \ldots, X_N$, and T observations for each of them. First, we compute $\binom{N}{2} = O(N^2)$ empirical copulas which represent the dependence structure between all the couples (X_i, X_j). Then, we summarize all these distributions using a center-based clustering algorithm, and extract the clusters centers using a fast computation of Wasserstein barycenters (Cuturi and Doucet, 2014). A given center represents the mean dependence between the couples (X_i, X_j) inside the corresponding cluster. Figure 4 and 5 illustrate why a Wasserstein W_2 barycenter, i.e. the minimizer $\mu^\star$ of $\frac{1}{N}\sum_{i=1}^{N} W_2^2(\mu, \nu_i)$ (Agueh and Carlier, 2011) where $\{\nu_1, \ldots, \nu_N\}$ is a set of N measures (here, bivariate empirical copulas), is more relevant to our needs: we benefit from robustness against small deformations of the dependence patterns.

3.2.2 TARGETS AS HYPOTHESES FROM AN EXPERT

One can specify dependence hypotheses, generate the corresponding copulas, then measure and rank correlations with respect to them. For example, one can answer to questions such as: Which are the pairs of assets that are usually positively correlated for small variations but uncorrelated otherwise? In (Durante et al., 2009), authors present a method for constructing bivariate copulas by changing the values that a given copula assumes on some subrectangles of the unit square. They discuss some applications of their methodology including the construction of copulas with different tail dependencies. Building *target* and *forget* copulas is another one. In the Experiments section, we illustrate its use to answer the previous question and other dependence queries.

4. Experiments

4.1 Exploration of financial correlations

We illustrate the first part of the methodology with three different datasets of financial time series. These time series consist in the daily returns of stocks (40 stocks from the CAC 40 index comprising the French highest market capitalizations), credit default swaps (75 CDS from the iTraxx Crossover index comprising the most liquid sub-investment grade European entities) and foreign exchange rates (80 FX rates of major world currencies) between January 2006 and August 2016. We display

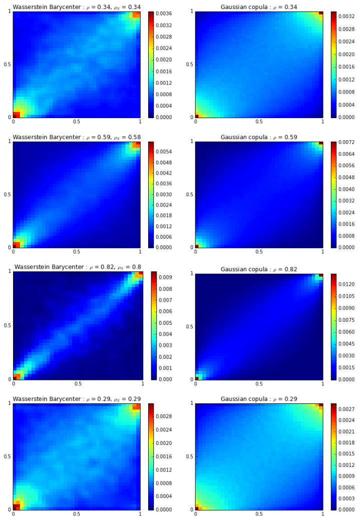


Figure 6:
Stocks: More mass in the bottom-left corner, i.e. lower tail dependence. Stock prices tend to plummet together. Otherwise, empirical copulas are similar to the Gaussian ones.

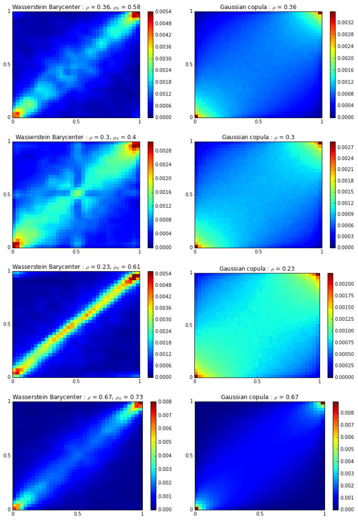


Figure 7:
Credit default swaps: More mass in the top-right corner, i.e. upper tail dependence. Insurance cost against the default of companies tends to soar in distressed market.

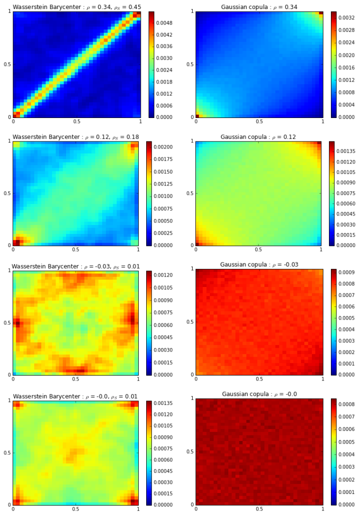


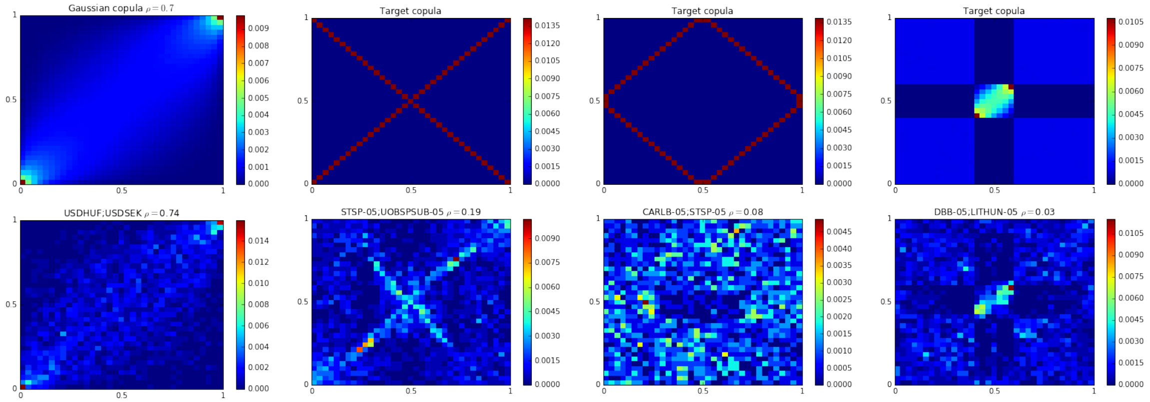
Figure 8:
FX rates: Empirical copulas show that dependence between FX rates are various, and strongly non-linear: Elliptical copulas and comonotonic measures are thus ill-suited.

some of the clustering centroids obtained for each asset class in the left column, and on their right we display their corresponding Gaussian copulas parameterized by the estimated linear correlations. Notice in Figures 6, 7, 8 the strong difference between the empirical copulas and the Gaussian ones which are still widely used in financial engineering due to their convenience. Notice also the difference between asset classes: Though estimated correlations are $\rho = 0.34$ for the topmost copulas, they have much dissimilar peculiarities.

4.2 Answering dependence queries

Inspired by the previous exploration results, we may want to answer such questions: (A) Which pair of assets having $\rho = 0.7$ correlation has the nearest copula to the Gaussian one? Though such questions can be answered by computing a likelihood for each pairs, our methodology stands out for dealing with non-parametric dependence patterns, and thus for questions such as: (B) Which pairs of assets are both positively and negatively correlated? (C) Which assets occur extreme variations while those of others are relatively small, and conversely? (D) Which pairs of assets are positively correlated for small variations but uncorrelated otherwise?

Considering a cross-asset dataset which comprises the SBF 120 components (index including the CAC 40 and 80 other highly capitalized French entities), the 500 most liquid CDS worldwide, and 80 FX rates, we display in Figure 9 the empirical copulas (below their respective targets) which best answer questions A,B,C,D.



(A) (B) (C) (D)

Figure 9: Target copulas (simulated or handcrafted) and their respective nearest copulas which answer questions A,B,C,D

4.3 Power of TFDC

In this experiment, we compare the empirical power of TFDC to well-known dependence coefficients such as Pearson linear correlation (cor), distance correlation (dCor) (Székely et al., 2009), maximal information coefficient (MIC) (Reshef et al., 2011), alternating conditional expectations (ACE) (Breiman and Friedman, 1985), maximum mean discrepancy (MMD) (Gretton et al., 2012), copula maximum mean discrepancy (CMMD) (Ghahramani et al., 2012), randomized dependence coefficient (RDC) (Lopez-Paz et al., 2013). Statistical power of a binary hypothesis test is the probability that the test correctly rejects the null hypothesis (H0) when the alternative hypothesis (H1) is true. In the case of dependence coefficients, we consider (H0): X and Y are independent; (H1): X and Y are dependent. Following the numerical experiment described in (Simon and Tibshirani, 2014; Lopez-Paz et al., 2013), we estimate the power of the aforementioned dependence measures with simulated pairs of variables with different relationships (considered in (Reshef et al., 2011; Lopez-Paz et al., 2013)), but with varying levels of noise added. By design, TFDC aims at detecting the simulated dependence relationships. Thus, this dependence measure is expected to have a much higher power than coefficients such as MIC. Results are displayed in Figure 10.

5. Discussion

It is known by risk managers how dangerous it can be to rely solely on a correlation coefficient to measure dependence. That is why we have proposed a novel approach to explore, summarize and measure the pairwise correlations which exist between variables in a dataset. The experiments show the benefits of the proposed method: It allows to highlight the various dependence patterns that can be found between financial time series, which strongly depart from the Gaussian copula widely used in financial engineering. Though *answering dependence queries* as briefly outlined is still an art, we plan to develop a rich language so that a user can formulate complex questions about dependence, which will be automatically translated into copulas in order to let the methodology provide these questions accurate answers.

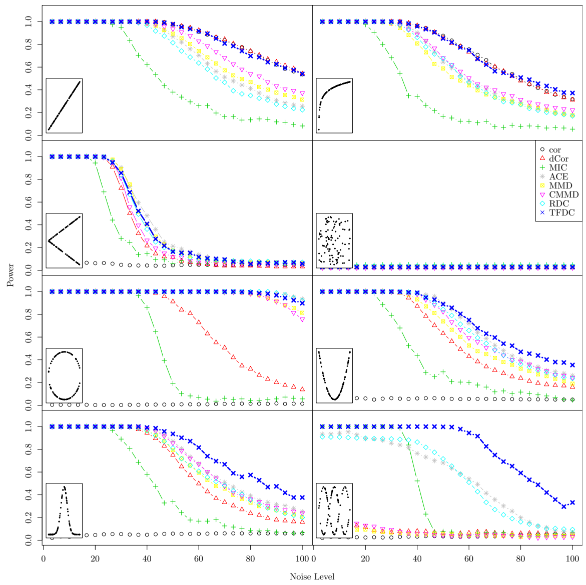


Figure 10: Power of several dependence coefficients as a function of the noise level in eight different scenarios. Insets show the noise-free form of each association pattern. The coefficient power was estimated via 500 simulations with sample size 500 each

References

Martial Agueh and Guillaume Carlier. Barycenters in the Wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.

Arindam Banerjee, Srujana Merugu, Inderjit S Dhillon, and Joydeep Ghosh. Clustering with Bregman divergences. *The Journal of Machine Learning Research*, 6:1705–1749, 2005.

Leo Breiman and Jerome H Friedman. Estimating optimal transformations for multiple regression and correlation. *Journal of the American statistical Association*, 80(391):580–598, 1985.

Yale Chang, Yi Li, Adam Ding, and Jennifer Dy. A robust-equitable copula dependence measure for feature selection. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, pages 84–92, 2016.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, pages 2292–2300, 2013.

Marco Cuturi and Arnaud Doucet. Fast computation of Wasserstein barycenters. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*, pages 685–693, 2014.

Paul Deheuvels. La fonction de dépendance empirique et ses propriétés. un test non paramétrique dindépendance. *Acad. Roy. Belg. Bull. Cl. Sci.(5)*, 65(6):274–292, 1979.

Arnaud Dessein, Nicolas Papadakis, and Jean-Luc Rouas. Regularized optimal transport and the rot mover's distance. *arXiv preprint arXiv:1610.06447*, 2016.

Adam Ding and Yi Li. Copula correlation: An equitable dependence measure and extension of pearson's correlation. *arXiv preprint arXiv:1312.7214*, 2013.

Fabrizio Durante, Susanne Saminger-Platz, and Peter Sarkoci. Rectangular patchwork for bivariate copulas and tail dependence. *Communications in StatisticsTheory and Methods*, 38 (15):2515–2527, 2009.

Zoubin Ghahramani, Barnabás Póczos, and Jeff G Schneider. Copula-based kernel dependency measures. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, pages 775–782, 2012.

Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(Mar):723–773, 2012.

Justin B Kinney and Gurinder S Atwal. Equitability, mutual information, and the maximal information coefficient. *Proceedings of the National Academy of Sciences*, 111(9):3354–3359, 2014.

David Lopez-Paz, Philipp Hennig, and Bernhard Schölkopf. The randomized dependence coefficient. In *Advances in Neural Information Processing Systems*, pages 1–9, 2013.

Gautier Marti, Sébastien Andler, Frank Nielsen, and Philippe Donnat. Optimal transport vs. Fisher-Rao distance between copulas for clustering multivariate time series. In *2016 IEEE Statistical Signal Processing Workshop*, Palma de Mallorca, Spain, 2016a.

Gautier Marti, Frank Nielsen, Philippe Donnat, and Sébastien Andler. On clustering financial time series: a need for distances between dependent random variables. In Frank Nielsen, Frank Critchley, and Christopher T. J. Dodson, editors, *Computational Information Geometry*, pages 149–174, 2016b.

Boris Muzellec, Richard Nock, Giorgio Patrini, and Frank Nielsen. Tsallis regularized optimal transport and ecological inference. *arXiv preprint arXiv:1609.04495*, 2016.

David N Reshef, Yakir A Reshef, Hilary K Finucane, Sharon R Grossman, Gilean McVean, Peter J Turnbaugh, Eric S Lander, Michael Mitzenmacher, and Pardis C Sabeti. Detecting novel associations in large data sets. *science*, 334(6062):1518–1524, 2011.

Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, 40(2):99–121, 2000.

Noah Simon and Robert Tibshirani. Comment on "Detecting Novel Associations In Large Data Sets" by Reshef Et Al, Science Dec 16, 2011. *arXiv preprint arXiv:1401.7645*, 2014.

Gábor J Székely, Maria L Rizzo, et al. Brownian distance covariance. *The annals of applied statistics*, 3(4):1236–1265, 2009.