amse
école d'économie d'aix-marseille
aix-marseille school of economics

Working Papers / Documents de travail

Local Linear Dependence Measure for Functionally Correlated Variables

Loann D. Desboulets
Costin Protopopescu

WP 2018 - Nr 53

Noname manuscript No.
(will be inserted by the editor)

Local Linear Dependence Measure for Functionally Correlated Variables

L. D. Desboulets · C. Protopopescu

This version 07/11/2018

Abstract We propose a new correlation measure for functionally correlated variables based on local linear dependence. It is able to detect non-linear, non-monotonic and even implicit relationships. Applying the classical linear correlation in a local framework combined with tools from Principal Components Analysis the statistic is capable of detecting very complex dependences among the data. In a first part we prove that it meets the properties of independence, similarity invariance and dependence and the axiom of continuity. In a second part we run a numerical simulation over a variety of dependences and compare it to other dependence measures in the literature. The results indicate that we outperform existing coefficients. We also show better stability and robustness to noise.

Keywords Local Correlation · Pearson Coefficient · PCA · Non-Parametric Statistic · Implicit Dependence · Non-Monotonic · Non-Linear

Mathematics Subject Classification (2010) MSC 2010

1 Introduction

Assessing statistical dependence among variables has been a major subject for a long time. Strongly related to regression problems, correlation can be used to understand the relationship between numerical variables. Even though correlation coefficients are basic statistics in regression analysis, they are still useful

L. D. Desboulets
Aix-Marseille Univ., CNRS, EHESS, Centrale Marseille, AMSE
5-9 Boulevard Maurice Bourdet, 13001 Marseille
E-mail: loann.DESBOULETS@univ-amu.fr

C. Protopopescu
Aix-Marseille Univ., CNRS, EHESS, Centrale Marseille, AMSE
5-9 Boulevard Maurice Bourdet, 13001 Marseille
E-mail: costin.protopopescu@univ-amu.fr

in summarizing lots of informations contained in the data. First introduced by Pearson (Pearson, 1895) the product-moment correlation coefficient is still a very popular tool in statistics however it is limited to linear correlations. Through the years it has been improved in many ways. Former improvements were to detect monotonic correlations. The two most common monotonic coefficients are Spearman's ρ_s (Spearman, 1904) and Kendall's τ (Kendall, 1938). They use ranks instead of the raw variables giving them the ability to measure a monotonic link between two variables. Both of them can be considered as special cases of a more general correlation coefficient latter proposed by Daniels (1944). Hoeffding's d (Hoeffding, 1948) was a successful attempt of non-monotonic correlation coefficients. Major drawbacks of this one arise from its sensitivity to noisy variables and its lack of interpretation as it belongs to $[-\frac{1}{2}; 1]$, 1 standing for complete dependence. It is not the only non-linear correlation coefficient, in his Analysis Of Variance (ANOVA), Pearson introduced the correlation ratio, described in Saporta (2006)[1]. It can be seen as the maximal correlation between X and the denoised variable Y. More recent developments deal with non-linear dependence. Even if it is not in the scope of this paper one should notice that most of them apply in multivariate settings. Bach and Jordan (2002) and Akaho (2006) developed the Kernel Canonical Correlation Analysis (KCCA) based on Reproducing Kernel Hilbert Spaces (RHKS) using kernel methods for generalizing the simple linear covariance. The idea of using RHKS is not new, it was already proposed in Grindea and Postelnicu (1977) for the General Coefficient (GC) but they did not explain how to estimate the transformations in the RHKS. Solutions for computing this coefficient with splines instead of kernels can be found in Hall and Miller (2009). Székely et al. (2007) proposed the Distance Correlation (DC) which uses correlation on Euclidean distances instead of ranks or raw variables. Naturally DC can be extended to more general distances (Lyons, 2013). Another coefficient is the Maximal Information Coefficient (MIC) of Reshef et al. (2011) that uses mutual information (MI), closely related to entropy in information theory.

There exists also tests for nonlinear independence from which the statistic assessing the correlation may not be easily compared between each other. We can cite the Hilbert-Schmidt Independence Criterion (HSIC) of Gretton et al. (2005, 2007) based on RHKS in the same spirit as KCCA, the Continuous ANOVA (CANOVA) of Wang et al. (2015) based on local permutations and the Maximal Local Correlation (MLC) of Chen et al. (2010). This is closely linked to our method since it uses local information in the form of local correlation integrals as defined in Grassberger and Procaccia (1983). His drawback is that their coefficient does not lie in a specific range. These methods are related to non-linear correlation, a comparison showing equivalences between DC and HSIC can be found in Sejdinovic et al. (2013). We do not include these methods in our simulation as they do not produce genuine correlation coefficients. In their paper, Chen et al. (2010) compare the MLC to this kind of measures on

[1] See pages 82-83

the ground of false positive rates making it more like a test for independence than a genuine coefficient. For that same reason we do not include MLC as well. In the end we will compare to classical coefficients: Pearson's ρ, Spearman's ρ_s and Kendall's τ, and to a panel of non-linear coefficients most widely used in practice: Hoeffding's d, Correlation Ratio η, DC, KCCA and MIC. For the sake of simplicity we do not include marginal variations of these coefficients existing in the literature.

In this paper, we extend the concept of correlation into a new framework to estimate non-monotonic and possibly more general functional relationships such as a circle or any implicit dependence. Our new method is aimed at detecting any statistical dependence between two variables. It is based on two separate statistics: the local correlation from a local sample and a sparsity measure using the concepts in Principal Component Analysis first introduced by Pearson (1901).

2 New Method

If there exists a non-trivial dependence between two variables X and Y it is hard to detect it globally (in the whole sample) as it can be any implicit dependence. Since any continuous function can be approximated using an affine function in the neighbourhood of a given point, measuring an unknown relationship can be achieved as a local linear approximation. We also benefit from local neighbours[2] in an unsupervised manner to allow for a broader class of dependencies rather than just functional relationships.

2.1 Lambda measure

As it was said we will be measuring correlation locally in the sample. The simpler way to summarize information about the local correlations is to average them. Our measure does not rely on a particular coefficient, any of the mentioned in the previous section can be used. The most simple is to use the linear correlation coefficient. This one lies in the range $[-1; 1]$ so we take its absolute value to avoid symmetry effects. We define the average local correlation as:

$$\bar{\rho}_h(X, Y) = \frac{1}{n} \sum_{i=1}^{n} |\rho_{i,h}(N_h(P_i(X_i, Y_i)))| \tag{1}$$

with $\rho_{i,h}$ the correlation coefficient in the sub-sample of (X, Y), defined by a circle of radius h around the considered point $P_i(X_i, Y_i)$:

$$N_h(P_i(X_i, Y_i)) = \{P(X, Y) \mid d(P_i(X_i, Y_i), P(X, Y)) \leq h\}$$

[2] Closely related to k-nearest neighbours.

where $d(\cdot)$ is a distance metric.[3].

One can note that the statistic λ_h is a measure of the goodness of fit of the local approximation of the conditional mean of the data generating process. It will be close to 1 when there is a noise-free dependence, providing h was optimally set. The sequence of $\rho_{i,h}$ contains the information of local linearities and is directly related with a piecewise linear approximation of the true dependence (in the general case, intimately related with the projection on the class of piecewise linear functions). The average local correlation has the ability to look for non-monotonic dependence on its own, but in order to be more robust to noisy data it has to be improved.

2.2 Sparsity measure

The noise-to-signal ratio is higher in local samples than in the whole sample so we overcome this issue by introducing a second measure. Principal Component Analysis Pearson (1901), hereafter PCA, provides a solution to measure the signal-to-noise ratio with different clues and interpretations. We use the singular value decomposition (SVD) of the covariance matrix $\mathbf{C}(X, Y)$ to extract all lower dimensional representation of the data as well as the variances of each of these representations. The lower dimensional planes are computed by projecting the data on the eigenvectors $\mathbf{V}$ while the variances are contained in the eigenvalues $\mathbf{L}$. From these matrices we can compute the matrix of "Squared Loadings" $\mathbf{\Lambda}$ which contains the information on the direction and the variance of the projection.

$$\mathbf{C} = \mathbf{V}\mathbf{L}\mathbf{V}'$$
$$\mathbf{\Lambda} = (\mathbf{V}\mathbf{L}^{1/2})^2$$

If the data are independent, which is equivalent to an equal distribution in the 2 dimensions, then each component has the same relative importance. If the data lie on a lower dimensional space, such as a line, the last component will have a small eigenvalue. Square loadings describe the information of the linear compression of the dataset. In the 2-dimensional case, it is very straightforward to see that the first component will measure the signal and the second will measure the noise, under the assumption that h is optimally chosen. Let $\mathbf{\Lambda}^{(i)}$ be local "squared loadings" matrix computed in the local sub-sample around the i^{th} observation. It is natural to consider as a measure of the relative importance of the second component with respect to the first one the statistic:

$$s_{i,h} = \frac{\sum_{j=1}^{2} \mathbf{\Lambda}_{j,2}^{(i)}}{\sum_{j=1}^{2} \mathbf{\Lambda}_{j,1}^{(i)}}$$

[3] The distances are computed on standardized variables to avoid scale effects. In our simulations we will use the Euclidean distance.

The sequence of $s_{i,h}$ is a numerical indicator containing information about the noise in the data that we can average just as we did with the previous one:

$$s_h(X,Y) = \frac{1}{n}\sum_{i=1}^{n} s_{i,h} \tag{2}$$

Even for a non optimal choice of h, which might result in measuring the opposite direction, $s_h \in [0,1]$ because the first component always attests to the most variance along its direction whichever it is. This ensures our final statistic to be bounded in $[0,1]$.
What is useful is that the behaviour of s_h is anti-correlated with the one of λ_h. More exactly, for a strong dependency between X and Y we have $\lambda_h \to 1$ and $s_h \to 0$, and respectively for independence (or a weak dependency) $\lambda_h \to 0$ and $s_h \to 1$. So to improve on efficiency and stability we combine $\bar{\rho}$ and s into the local linear dependence measure:

$$\lambda_h := (\bar{\rho}_h)^{s_h}$$

This value will approach 1 faster and so reveal stronger dependency sooner than the average local correlation $\bar{\rho}$ itself and still remains 0 when the data are independent. Other combinations of $\bar{\rho}$ and s could have been employed, but we stick to this one because of its performance.

2.3 Dependence Measures Axioms

Correlation coefficients can come in very different forms as it has been mentioned in the introduction. For instance some like the HSIC, CANOVA, MLC are not upper bounded. Some may take negative values like the Pearson's ρ, Spearman's ρ_s and Kendall's τ. This kind of properties are very crucial as it can change or limit their interpretation. The question of what is the system of exhaustive desirable properties of dependence measures has been studied by Rényi (1959). We recall his set of axioms in the same way Móri and Székely (2018) did:

(A) Existence: $\mathcal{M}(X,Y)$ is defined for all non-constant random variables X and Y.
(B) Symmetry: $\mathcal{M}(X,Y) = \mathcal{M}(X,Y)$.
(C) Boundedness: $0 \leq |\mathcal{M}(X,Y)| \leq 1$.
(D) Independence: $\mathcal{M}(X,Y) = 0$ if and only if X and Y are independent.
(E) Strict Dependence: $\mathcal{M}(X,Y) = 1$ if X and Y ($X = g(Y)$ or $Y = f(X)$, where $g(x)$ and $f(x)$ are Borel measurable functions)
(F) Invariance: If the Borel measurable functions $f(x)$ and $g(x)$ map the real axis in a one-to-one way onto itself, $\mathcal{M}(f(X), g(Y)) = \mathcal{M}(X,Y)$.
(G) Gaussian Case: If the joint distribution of X and Y is normal, then $\mathcal{M}(X,Y) = |\rho(X,Y)|$ where $\rho(X,Y)$ is the correlation coefficient of X and Y.

Because of its structure our statistic inherits many properties from the Pearson's ρ coefficient. And on the other hand, it is not affected by the PCA component in a way that contradicts these axioms. For the first set of axioms: using PCA does not affect symmetry (B). The combination of $\bar{\rho}$ and s ensures boundedness between 0 and 1 (C). One can note that axiom (D) is violated by coefficients that do not detect non-linear dependence. For example Pearson's ρ does not respect the "only if" in case of a quadratic dependence on a symmetric support. Axiom (G) is satisfied only by $\bar{\rho}$ alone. So in the end we satisfy axioms (A)-(B)-(C)-(D)-(E).

Proposition 1 *λ satisfies axioms (A)-(B)-(C)-(D)-(E).*

Proof Let $\bar{\rho}_h(X,Y)$ defined as in equation 1 and $s_h(X,Y)$ as in equation 2.

(A) For any pair (X,Y) non-constant for which the covariance matrix $\mathbf{C}(X,Y)$ exists $\rho_h(X,Y)$ and $s_h(X,Y)$ are well defined and so is λ_h.
(B) Given the symmetry of the covariance and of the distance both $\rho_h(X,Y)$ and $s_h(X,Y)$ are symmetric and so is λ_h.
(C) By Schwarz inequality we have $|\rho_{i,h}| \leq 1$ and so its average $\rho_h(X,Y)$. This ensures $0 \leq \lambda_h(X,Y) \leq 1$.
(D) Independence of (X,Y) is equivalent to local independence of (X_i,Y_i). $\lambda_h(X,Y) = 0 \Leftrightarrow \rho_h(X,Y) = 0 \Leftrightarrow |\rho_{i,h}(X_i,Y_i)| = 0 \ \forall i = 1,...,n$. This implies either local independence inside each $N_h(P_i)$ which concludes the proof. Or local symmetry in each $N_h(P_i)$ which is impossible because it would be broken through translation.
(E) $\lambda_h(X,Y) = 1 \Leftrightarrow \rho_h(X,Y) = 1$ or $s_h(X,Y) = 0$. And $\rho_h(X,Y) = 1 \Leftrightarrow |\rho_{i,h}(X_i,Y_i)| = 1 \ \forall i = 1,...,n$. This implies that there is a perfect local linear correlation in each $N_h(P_i)$ which is equivalent with a local linear approximation of the function defining the dependence.

Even if this system of axioms may be considered as exhaustive for characterizing a dependence measure, and it exists at least one measure verifying all these axioms, i.e. the maximal correlation measure (Renyi, 1959), one may ask some complementary question about defining a "minimalist" system of axioms that we can expect to be satisfied by all acceptable dependency measures. More recently Móri and Székely (2018) who also developed the Distance Correlation (DC) formulated a set of four axioms for which dependence measures should be in line with. In their paper, they investigated to which extent Pearson's ρ, Spearman's ρ_s, Kendall's τ, Correlation Ratio η, DC and MIC fit these axioms. For the ease of discussion, we recall them here as they did:

(i) Independence: $\mathcal{M}(X,Y) = 0$ if and only if X and Y are independent.
(ii) Similarity Invariance: $\mathcal{M}(X,Y)$ is invariant with respect to all similarity transformations of the Hilbert space denoted H; that is $\mathcal{M}(LX,MY) = \mathcal{M}(X,Y)$ where L, M are similarity transformations of H.
(iii) Similarity Dependence: $\mathcal{M}(X,Y) = 1$ if and only if $Y = LX$ with probability 1, where L is a similarity transformation of H.

(iv) Continuity: $\mathcal{M}(X,Y)$ is continuous; that is, if $(X_n,Y_n) \in S, n = 1,2,...$ such that for some positive constant K we have $E(|X_n|^2+|Y_n|^2) \leq K, n = 1,2,...$ and (X_n,Y_n) converges weakly (in distribution) to (X,Y) then $\mathcal{M}(X_n,Y_n) \to \mathcal{M}(X,Y)$.

In their paper, they insist on (iv), concerning the weak continuity of the dependence measure. Not fulfilling this axiom means that as the sample size increases the empirical values of a dependence measure do not necessarily converge to the population value. And in order to justify the introduction of these new axioms, they show that maximal correlation measure does not satisfy (iv).

For the second set of axioms, one may note that the absolute value ρ and so $\bar{\rho}_h$ to verify all the four previous axioms. More precisely, (i) is equivalent to (D). Just as DC we did not fulfil axiom (F) but a restricted version that is axiom (ii). The main difference is that the dependence measure has to stay the same under one-to-one transformations as long as they preserve distances. (iii) is a restricted version of (E). So we satisfy axioms (i)-(ii)-(iii)-(iv).

Proposition 2 *λ satisfies axioms (i)-(ii)-(iii)-(iv).*

Proof Let $\bar{\rho}_h(X,Y)$ defined as in equation 1 and $s_h(X,Y)$ as in equation 2.

(i) Equivalent to axiom (D).
(ii) Similarity invariance preserves distances so $\bar{\rho}_h(X,Y)$ and $s_h(X,Y)$ remain unchanged.
(iii) Restricted version of (E).
(iv) Since Pearson's ρ is a bilinear application we have $\bar{\rho}_h(X,Y)$ absolutely continuous so we only need to prove the continuity for $s_h(X,Y)$. Dauxois et al. (1982) proved that if (X_n,Y_n) converges weakly (in distribution) to (X,Y) then eigenvectors $\mathbf{V_n}$ and eigenvalues $\mathbf{L_n}$ converge to $\mathbf{V}$ and $\mathbf{L}$ respectively. Since s_h is a continuous function of these quantities the continuity is guaranteed for $s_h(X_n,Y_n)$ and so for $\lambda_h(X_n,Y_n)$.

Móri and Székely (2018) shows that DC measure satisfies all these axioms too. The other ones fail for at least one of the axioms. MIC violates (iii) if the cumulative distribution function of X is not continuous and (iv) is not satisfied and Correlation Ratio η violates (i), (ii) and (iii) (see Móri and Székely (2018)). As it will be highlighted in section 3 of this paper, DC detects less the dependence between independent random variables, as opposed for example to MIC and KCCA (DC's extension).

These theoretical axioms alone are not sufficient to assess a dependence measure behaviour. Numerical simulations will complete this analysis in some empirical cases that were not considered here, e.g. circular dependence.

3 Simulation Study

3.1 Design

A numerical simulation is conducted to assess the behaviour of the new statistic in finite sample. We compute the mentionned coefficients under different dependences among the data following the methodology in Chen et al. (2010) and Wang et al. (2015).
The simulation is made with $n = 300$ observations[4], $X \sim$ i.i.d. $U([0, 1])$ and a Gaussian noise $\epsilon \sim$ i.i.d. $\mathcal{N}(0, \sigma)$. We compare the correlation measures for various scenarios.[5]. Results of the simulations are reported in Table 1.

The simulation study is conducted over 7 different models:

- Linear $f_1(x) = 2 + 0.8x$
- Non-Continuous $f_2(x) = (0.5x)\mathbb{1}_{x<0.5} + (-0.5 + 0.5x)\mathbb{1}_{x\geq 0.5}$
- Monotonic $f_3(x) = e^{-10x^2}$
- Quadratic $f_4(x) = 4(x - 0.5)^2$
- Non-Monotonic $f_5(x) = 6x - 8log(1 + x) + 2e^{-50(x-0.3)^2} + 0.6e^{-100(x-0.8)^2}$
- Non-Functional $f_6(x) = -x^2 + 1$
- Independence $f_7(x) = 0.6$

3.2 Results

*** INSERT TABLE 1 HERE ***

First of all it is important to note that among the 10 coefficients λ_h is always in the top 3. Moreover it is the only one to always be on top. This means that our new statistic is the only measure able to detect dependence correctly in all considered cases.

First, in the noiseless cases ($\sigma = 0.0$) we expect to measure the real abilities of the correlation coefficients and this is almost what one observes. In the first row we see that under a linear model all coefficients were able to measure perfect correlation, except η for which it is known the kernel regression can misfits the linear functions. For the non-continuous case it is important to note that the monotonic coefficients ρ, ρ_s and τ detected negative correlation while the simulated model is a clearly positive dependence. It denotes the risk of using these coefficients as well as interpreting their signs without a deeper investigation of the data. This is a not a problem with non-linear coefficient ranging from 0 to 1. Under models f_3 to f_5 both λ and MIC detect the perfect dependence. It is more surprising that KCCA (resp. 0.87 and 0.54) and DC (resp. 0.74 and 0.46) perform less in cases of non-continuity (f_2) and high

[4] Results for $n = 500$ can be found in Appendix A Table 2.
[5] These are illustrated in Appendix B Figure 2

non-linearities (f_5). Things become more complicated in the case of implicit dependence (f_6). MIC (0.67) and DC (0.20) do not manage to characterize a perfect dependence while KCCA (0.99) and λ (1.00) still perform well. We omit model f_7 because most of the coefficients are not defined for constant variables.

For the noisy cases ($\sigma = 0.1$), more related to realistic data, we find similar behaviours. As expected linear and monotonic coefficients namely ρ, ρ_s and τ are not able to detect dependences beyond their respective scopes. We notice that Spearman's ρ_s is more efficient than Kendall's τ in models f_1 to f_3. Non-linear coefficients have heterogeneous performances depending on the type of dependence. Hoeffding's d is very sensitive to the noise. Previously it only managed to find perfect dependence (1.00) in cases f_1 and f_3, but now it decreased to 0.51 and 0.46 respectively. It has still very low values for all other cases. DC is powerful in monotonic cases, and surprisingly good in the non-continuous model f_2, having almost the same performance as λ. However it underperforms at detecting non-monotonic dependences. KCCA and MIC are strong competitors in this setting. The only important thing to notice is that KCCA fails at the non-continuous case (f_2) as well as MIC fails at the non-functional one (f_6) whereas λ's performance remained constant throughout all the cases. One can notice that the performance of $\bar{\rho}$ has sharply decreased as it was expected. The need for the PCA based correction s is clearly highlighted here. It makes the coefficient robust to the locally amplified noise and allows to detect dependences correctly.

When the signal-to-noise ratio is much lower ($\sigma = 0.3$), as it is often the case in empirical work, we observe more clearly the robustness to noise of KCCA and λ as opposed to DC and MIC. It is also important to focus on the case of independence (f_7) DC, KCCA and MIC all found quite high values (resp. 0.09, 0.15, 0.20) and this was also true before (for $\sigma = 0.1$, resp. 0.14, 0.18, 0.23). This is a major drawback because one would not distinguish between independence and low signal. Nevertheless even if η and λ coefficients have reasonably low values (resp. 0.03 and 0.08) best ones are Pearson's ρ and Hoeffding's d (resp. -0.01 and 0.00).

Second, we also compare our new statistic to Pearson's ρ, DC, KCCA and MIC in Figure 1 (for $n = 300$, $\sigma = 0.1$) and Figure 3 (for $n = 300$, $\sigma = 0.3$). We repeated the simulation for a given model 20 times to obtain a cloud of points. If the cloud lies on the 45° line then λ is equal to the other measure. If the cloud lies above this line then our statistic is greater.

*** INSERT FIGURE 1 HERE ***

Figure 1 (a) shows the behaviour of λ against the Pearson's ρ. We see that for the linear case (f_1) and the independence case (f_7) we are close to the 45° line. It is a desirable property that in appropriate circumstances λ reduces to ρ. The slight bias arises because of the PCA component of the statistic that implies $\lambda \geq \rho$. The most interesting cases are the ones for which Pearson's

ρ is almost zero whereas λ did measure the dependence (vertical line). The circular model (f_6) produces a perfect symmetry so Pearson's ρ is zero in average while λ is almost 1. This is also true for the quadratic model (f_4). The very non-linear model (f_5) has a slight negative trend so Pearson's ρ is around -0.1 while in fact the dependence is much stronger, this is reflected in λ. In the case of the non-continuous piecewise model (f_2) the Pearson's ρ is completely misleading, it measures a negative and relatively weak correlation. On the contrary λ has high values indicating a relatively strong dependence.

It was expected that λ would perform better than Pearson's ρ in non-linear cases, but it is also interesting to know how we perform in contrast to non-linear dependence measures. As we have seen in Table 1 the overall performance of η and Hoeffding's d is quite poor so it is not useful to compare to them. On the contrary MIC, KCCA and even DC were quite good competitors in certain circumstances.

Figure 1 (b) shows values for λ against DC. In monotonic cases (f_1 and f_3) both are quite high and close to each other. For independence (f_7) DC is higher in average which means λ detects independence better. In Appendix B Figure 3 (b) one can note that DC do not distinguish between independence and the non-continuous model (f_2). DC detects non-linear dependence in models f_4 and f_5 but λ has greater values. DC fails completely at detecting dependence for the circular model (f_6).

Figure 1 (c) shows values for λ against KCCA. Both are very close except for the piecewise model (f_2) and the very non-linear model (f_5) where λ outperforms KCCA. For the independence case (f_7) KCCA is greater than λ in average meaning KCCA detects more dependence than λ when there is none. When the proportion of noise is increased (see Appendix B Figure 3 (c)) KCCA values dramatically decrease and λ outperforms in all cases. This means that our statistic is more robust to the noise than KCCA is.

Figure 1 (d) shows values for λ against MIC. We can observe horizontal clouds of points for models f_1, f_2, f_3, f_4 and f_6 showing that there is a great variance in MIC, on the contrary λ shows a better stability. As it was true for other measures λ is always greater or equal to MIC. The latter has also a too large value for independence, even more than preceding measures. Moreover MIC is less robust to the noise, when the noise variance is increased its values decrease even more than for KCCA (see Appendix B Figure 3 (d)).

4 Summary and Concluding Remarks

The local linear dependence measure (denoted λ) presented here has proven to be able to measure (implicit) functional dependences and to be robust to noisy samples. The new statistic λ inherits linear properties of Pearson's ρ and extends them into a local framework allowing for approximation of any classical or implicit dependence. On the theoretical side our statistic satisfies the desirable properties of dependence measures as defined in Móri and Székely (2018). On the empirical side we compare to historical correlation measures

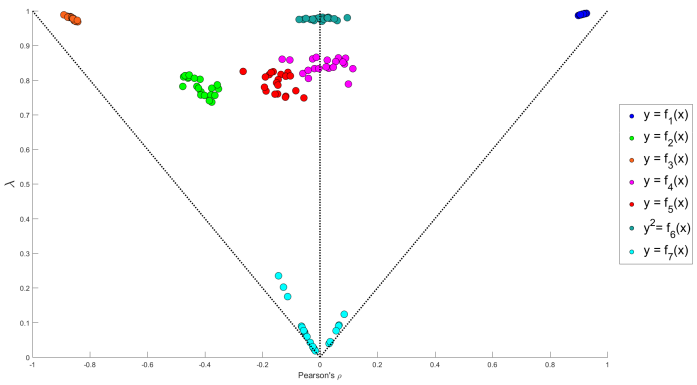
such as Pearson's ρ, Spearman's ρ_s and Kendall's τ. We found that it can detect dependence where these ones cannot and performs as best under appropriate cases. We also compare to non-linear coefficients such as Hoeffding's d, Correlation Ratio η, Distance Correlation (DC), Kernel Canonical Correlation Analysis (KCCA) and Maximum Information Coefficient (MIC). We found that it outperforms them under a variety of models and is much more robust in presence of noise in the data than them.

Table 1: Simulation for different dependences and different levels of noise

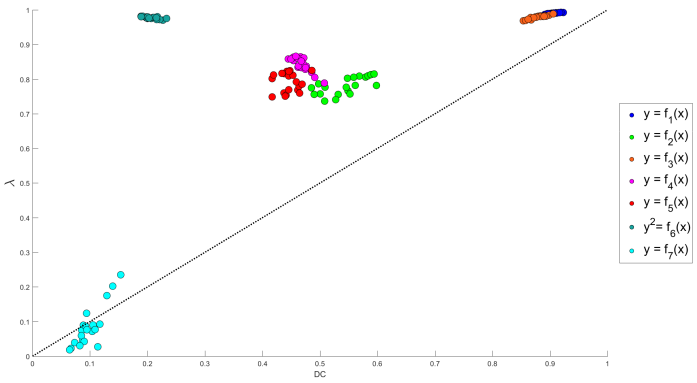
Functions	ρ	ρ_s	τ	d	η	DC	KCCA	MIC	$\bar{\rho}_h$	λ_h
$\sigma = 0.0$										
$y = f_1(x) + \epsilon$	**1.00**	**1.00**	**1.00**	**1.00**	0.87	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
$y = f_2(x) + \epsilon$	−0.51	−0.50	0.00	0.06	0.50	0.74	0.87	**1.00**	**0.98**	**1.00**
$y = f_3(x) + \epsilon$	−0.91	**-1.00**	**-1.00**	**1.00**	0.84	0.94	0.99	**1.00**	0.98	**1.00**
$y = f_4(x) + \epsilon$	−0.07	−0.08	−0.06	0.25	0.56	0.49	**1.00**	**1.00**	0.98	**1.00**
$y = f_5(x) + \epsilon$	−0.11	−0.04	−0.01	0.09	0.36	0.46	0.54	**1.00**	**0.80**	**0.91**
$y^2 = f_6(x) + \epsilon$	0.01	0.00	0.00	0.06	0.00	0.20	**0.99**	0.67	**0.96**	**1.00**
$\sigma = 0.1$										
$y = f_1(x) + \epsilon$	**0.93**	0.92	0.77	0.51	0.76	0.92	**0.93**	0.82	0.76	**0.94**
$y = f_2(x) + \epsilon$	−0.45	−0.46	−0.22	0.03	0.41	**0.60**	0.53	**0.71**	0.45	**0.73**
$y = f_3(x) + \epsilon$	−0.88	−0.87	−0.70	0.46	0.77	**0.90**	**0.96**	0.83	0.51	**0.88**
$y = f_4(x) + \epsilon$	0.02	0.05	0.03	0.13	0.52	0.46	**0.95**	**0.77**	0.57	**0.83**
$y = f_5(x) + \epsilon$	−0.11	−0.02	−0.01	0.07	0.35	0.48	**0.52**	**0.91**	0.45	**0.85**
$y^2 = f_6(x) + \epsilon$	0.02	0.02	0.01	0.04	0.00	0.21	**0.93**	0.55	**0.67**	**0.96**
$y = f_7(x) + \epsilon$	**-0.03**	−0.04	**-0.02**	**0.00**	0.05	0.14	0.18	0.23	0.04	0.07
$\sigma = 0.3$										
$y = f_1(x) + \epsilon$	0.64	**0.65**	0.46	0.15	0.37	0.60	**0.65**	0.53	0.41	**0.69**
$y = f_2(x) + \epsilon$	−0.23	−0.21	−0.13	0.02	0.12	**0.31**	0.29	**0.30**	0.15	**0.30**
$y = f_3(x) + \epsilon$	−0.71	**-0.73**	−0.53	0.22	0.53	0.71	**0.80**	0.57	0.48	**0.73**
$y = f_4(x) + \epsilon$	−0.05	−0.03	−0.02	0.04	0.34	0.38	**0.74**	**0.47**	0.37	**0.72**
$y = f_5(x) + \epsilon$	−0.08	−0.03	−0.01	0.04	0.28	0.35	**0.38**	**0.60**	0.25	**0.83**
$y^2 = f_6(x) + \epsilon$	0.00	0.00	0.00	0.01	0.01	0.17	**0.68**	0.30	**0.43**	**0.80**
$y = f_7(x) + \epsilon$	**-0.01**	−0.04	−0.04	**0.00**	**0.03**	0.09	0.15	0.20	0.05	0.08

Bold values indicates the three best coefficients, or best ones if there are tied values.
Values for h are optimally set via grid search.

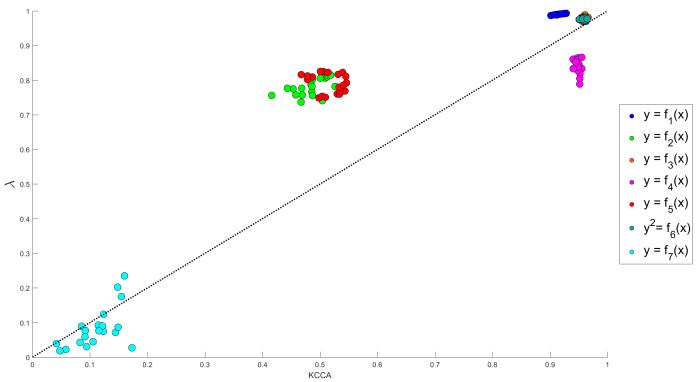
Fig. 1: λ against other measures ($n = 300$ and $\sigma = 0.1$)



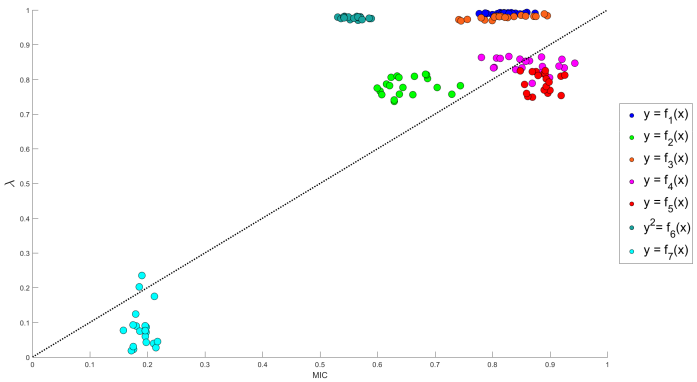
(a) λ versus Pearson's ρ



(b) λ versus DC



(c) λ versus KCCA



(d) λ versus MIC

Appendix A Supplementary Tables

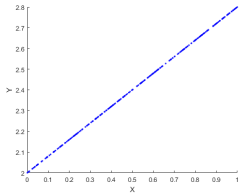
Table 2: Simulation for different dependences ($n = 500$)

Functions	ρ	ρ_s	τ	d	η	DC	KCCA	MIC	$\bar{\rho}_h$	λ_h
$\sigma = 0.0$										
$y = f_1(x) + \epsilon$	**1.00**	**1.00**	**1.00**	**1.00**	0.89	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
$y = f_2(x) + \epsilon$	−0.51	−0.50	0.00	0.06	0.54	0.73	0.86	**1.00**	**1.00**	**1.00**
$y = f_3(x) + \epsilon$	−0.91	**-1.00**	**-1.00**	**1.00**	0.88	0.94	0.99	**1.00**	0.99	**1.00**
$y = f_4(x) + \epsilon$	−0.10	−0.03	0.01	0.25	0.66	0.50	**1.00**	**1.00**	0.97	**1.00**
$y = f_5(x) + \epsilon$	−0.16	−0.03	0.01	0.10	0.47	0.47	0.53	**1.00**	**0.79**	**0.92**
$y^2 = f_6(x) + \epsilon$	0.03	0.02	0.00	0.06	0.00	0.21	**0.99**	0.64	**0.96**	**1.00**
$\sigma = 0.1$										
$y = f_1(x) + \epsilon$	**0.91**	**0.92**	0.75	0.48	0.74	0.90	**0.91**	0.82	0.88	**0.99**
$y = f_2(x) + \epsilon$	−0.43	−0.46	−0.23	0.04	0.34	**0.55**	0.47	**0.66**	0.40	**0.77**
$y = f_3(x) + \epsilon$	−0.86	−0.82	−0.65	0.40	0.82	**0.88**	**0.96**	0.84	0.82	**0.98**
$y = f_4(x) + \epsilon$	−0.03	−0.01	−0.01	0.13	0.62	0.49	**0.95**	**0.89**	0.63	**0.83**
$y = f_5(x) + \epsilon$	−0.14	−0.05	−0.01	0.06	0.36	0.42	**0.47**	**0.88**	0.44	**0.83**
$y^2 = f_6(x) + \epsilon$	−0.01	−0.01	0.00	0.03	0.00	0.19	**0.91**	0.55	**0.68**	**0.97**
$y = f_7(x) + \epsilon$	−0.08	−0.09	−0.06	**0.00**	**0.01**	0.12	0.13	0.21	**0.04**	0.08
$\sigma = 0.3$										
$y = f_1(x) + \epsilon$	**0.64**	0.63	0.45	0.14	0.35	0.60	**0.64**	0.39	0.53	**0.83**
$y = f_2(x) + \epsilon$	−0.21	−0.22	−0.14	0.02	0.11	**0.26**	0.23	**0.26**	0.23	**0.47**
$y = f_3(x) + \epsilon$	**-0.68**	−0.66	−0.47	0.17	0.50	**0.68**	**0.79**	0.49	0.58	**0.87**
$y = f_4(x) + \epsilon$	0.00	−0.01	0.00	0.05	0.35	0.36	**0.73**	**0.44**	0.37	**0.70**
$y = f_5(x) + \epsilon$	−0.14	−0.08	−0.04	0.04	0.38	**0.39**	0.38	**0.69**	0.26	**0.84**
$y^2 = f_6(x) + \epsilon$	0.00	0.00	0.00	0.01	0.00	0.15	**0.64**	0.29	**0.46**	**0.83**
$y = f_7(x) + \epsilon$	0.02	0.02	**0.01**	**0.00**	**0.00**	0.05	0.06	0.18	0.02	0.04

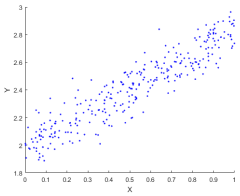
Bold values indicates the three best coefficients, or best ones if there are tied values.
Values for h are optimally set via grid search.

Appendix B Supplementary Figures

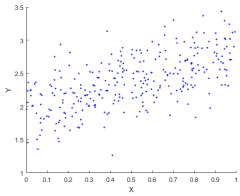
Fig. 2: Simulated Cases with increasing noises level $\epsilon(\sigma)$



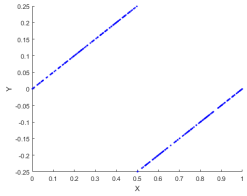
(a) $y = f_1(x) + \epsilon(0.0)$



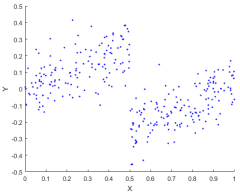
(b) $y = f_1(x) + \epsilon(0.1)$



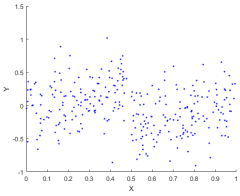
(c) $y = f_1(x) + \epsilon(0.3)$



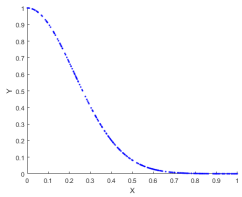
(d) $y = f_2(x) + \epsilon(0.0)$



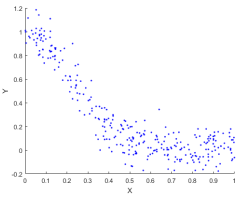
(e) $y = f_2(x) + \epsilon(0.1)$



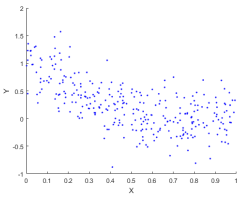
(f) $y = f_2(x) + \epsilon(0.3)$



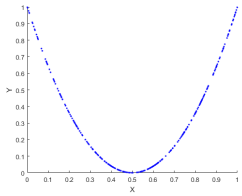
(g) $y = f_3(x) + \epsilon(0.0)$



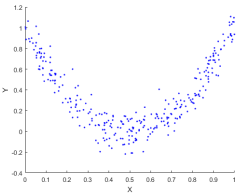
(h) $y = f_3(x) + \epsilon(0.1)$



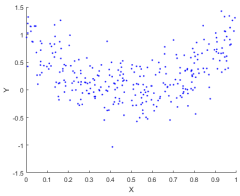
(i) $y = f_3(x) + \epsilon(0.3)$



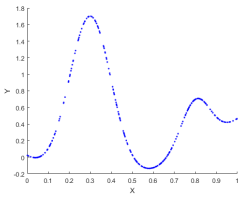
(j) $y = f_4(x) + \epsilon(0.0)$



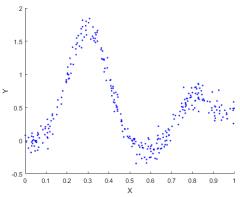
(k) $y = f_4(x) + \epsilon(0.1)$



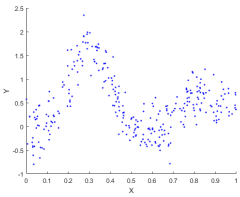
(l) $y = f_4(x) + \epsilon(0.3)$



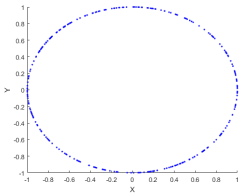
(m) $y = f_5(x) + \epsilon(0.0)$



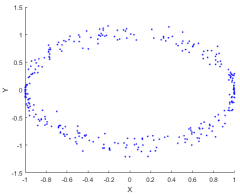
(n) $y = f_5(x) + \epsilon(0.1)$



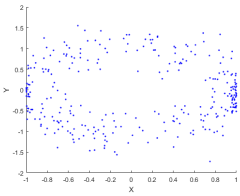
(o) $y = f_5(x) + \epsilon(0.3)$



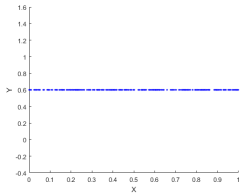
(p) $y^2 = f_6(x) + \epsilon(0.0)$



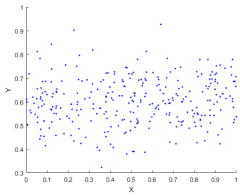
(q) $y^2 = f_6(x) + \epsilon(0.1)$



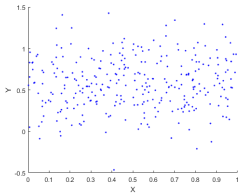
(r) $y^2 = f_6(x) + \epsilon(0.3)$



(s) $y = f_7(x) + \epsilon(0.0)$

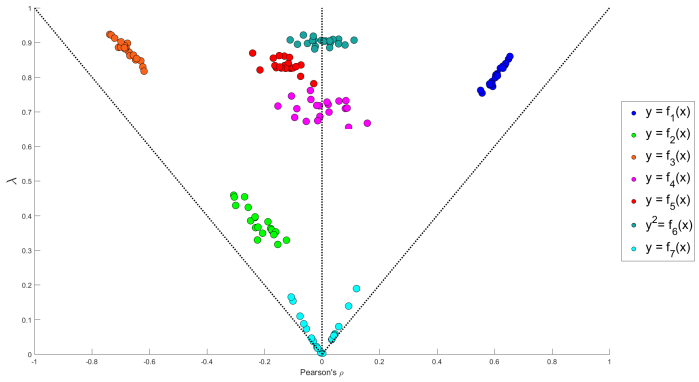


(t) $y = f_7(x) + \epsilon(0.1)$

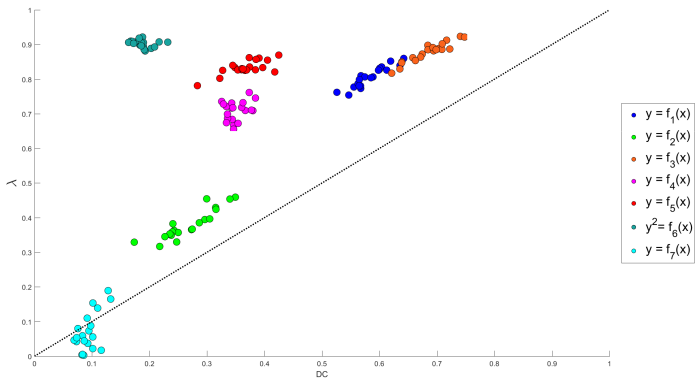


(u) $y = f_7(x) + \epsilon(0.3)$

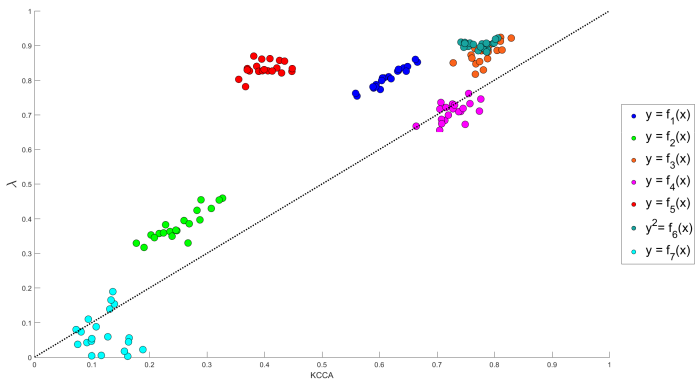
Fig. 3: λ against other measures ($n = 300$ and $\sigma = 0.3$)



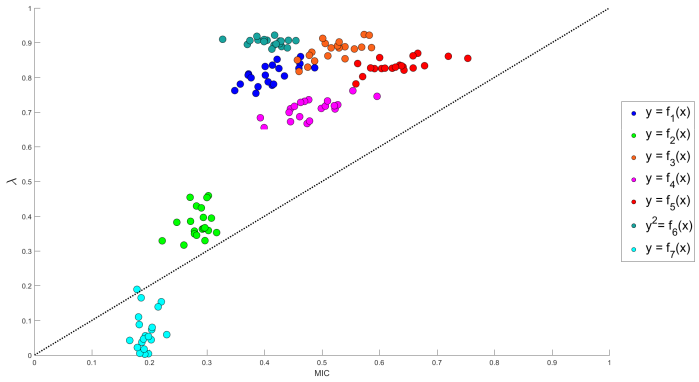
(a) λ versus Pearson's ρ



(b) λ versus DC



(c) λ versus KCCA



(d) λ versus MIC

References

Akaho S (2006) A kernel method for canonical correlation analysis. arXiv preprint cs/0609071

Bach FR, Jordan MI (2002) Kernel independent component analysis. Journal of machine learning research 3(Jul):1–48

Chen YA, Almeida JS, Richards AJ, Müller P, Carroll RJ, Rohrer B (2010) A nonparametric approach to detect nonlinear correlation in gene expression. Journal of Computational and Graphical Statistics 19(3):552–568

Daniels HE (1944) The relation between measures of correlation in the universe of sample permutations. Biometrika 33(2):129–135, URL http://www.jstor.org/stable/2334112

Dauxois J, Pousse A, Romain Y (1982) Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. Journal of multivariate analysis 12(1):136–154

Grassberger P, Procaccia I (1983) Characterization of strange attractors. Physical review letters 50(5):346

Gretton A, Bousquet O, Smola A, Schölkopf B (2005) Measuring statistical dependence with hilbert-schmidt norms. In: International conference on algorithmic learning theory, Springer, pp 63–77

Gretton A, Borgwardt KM, Rasch M, Schölkopf B, Smola AJ (2007) A kernel method for the two-sample-problem. In: Advances in neural information processing systems, pp 513–520

Grindea S, Postelnicu V (1977) Some measures of association. In: Proceedings of the Fifth Conference on Probability Theory (Brasov, 1974). Editura Acad. RSR, Bucharest, pp 197–203

Hall P, Miller H (2009) Using generalized correlation to effect variable selection in very high dimensional problems. Journal of Computational and Graphical Statistics 18(3):533–550

Hoeffding W (1948) A non-parametric test of independence. The annals of mathematical statistics pp 546–557

Kendall MG (1938) A new measure of rank correlation. Biometrika 30(1/2):81–93

Lyons R (2013) Distance covariance in metric spaces. The Annals of Probability 41(5):3284–3305

Móri TF, Székely GJ (2018) Four simple axioms of dependence measures. Metrika pp 1–16

Pearson K (1895) Note on regression and inheritance in the case of two parents. Proceedings of the Royal Society of London 58:240–242

Pearson K (1901) On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2(11):559–572

Rényi A (1959) On measures of dependence. Acta mathematica hungarica 10(3-4):441–451

Reshef DN, Reshef YA, Finucane HK, Grossman SR, McVean G, Turnbaugh PJ, Lander ES, Mitzenmacher M, Sabeti PC (2011) Detecting novel associ-

ations in large data sets. science 334(6062):1518–1524
Saporta G (2006) Probabilites Statistique et analyse des donnees. Editions Technip, Paris, 2006
Sejdinovic D, Sriperumbudur B, Gretton A, Fukumizu K (2013) Equivalence of distance-based and rkhs-based statistics in hypothesis testing. The Annals of Statistics pp 2263–2291
Spearman C (1904) " general intelligence," objectively determined and measured. The American Journal of Psychology 15(2):201–292
Székely GJ, Rizzo ML, Bakirov NK (2007) Measuring and testing dependence by correlation of distances. The annals of statistics 35(6):2769–2794
Wang Y, Li Y, Cao H, Xiong M, Shugart YY, Jin L (2015) Efficient test for nonlinear dependence of two continuous variables. BMC bioinformatics 16(1):260