Independence Testing for Temporal Data

Anonymous authors
Paper under double-blind review

Abstract

Temporal data are increasingly prevalent in modern data science. A fundamental question is whether two time series are related or not. Existing approaches often have limitations, such as relying on parametric assumptions, detecting only linear associations, and requiring multiple tests and corrections. While many non-parametric and universally consistent dependence measures have recently been proposed, directly applying them to temporal data can inflate the p-value and result in an invalid test. To address these challenges, this paper introduces the temporal dependence statistic with block permutation to test independence between temporal data. Under proper assumptions, the proposed procedure is asymptotically valid and universally consistent for testing independence between stationary time series, and capable of estimating the optimal dependence lag that maximizes the dependence. Moreover, it is compatible with a rich family of distance and kernel based dependence measures, eliminates the need for multiple testing, and exhibits excellent testing power in various simulation settings.

1 Introduction

Temporal data, often referred to as time series, finds wide applications across diverse domains, such as functional magnetic resonance imaging (fMRI) in neuroscience, dynamic social networks in sociology, financial indices, etc. In a broader context, temporal data can be seen as a type of structural data characterized by inherent underlying patterns. When dealing with temporal data, a fundamental problem is to determine the presence of a relationship between two jointly observed time series.

In the context of standard independent and identically distributed (i.i.d.) data, where observations $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ are drawn independently and identically from the joint distribution F_{XY}, the question simplifies to whether the underlying random variables X and Y are independent, i.e., $F_{XY} = F_X F_Y$. Many recent dependence measures have been proposed to tackle this problem, aiming to achieve valid and universally consistent independence testing. These methods include distance correlation (Szekely et al., 2007; Szekely & Rizzo, 2009; 2014), Hilbert-Schmidt independence criterion (Gretton et al., 2005; Gretton & Gyorfi, 2010; Gretton et al., 2012), multiscale graph correlation (Vogelstein et al., 2019; Shen et al., 2020; Lee et al., 2019), and many others (Heller et al., 2013; Zhu et al., 2017; Pan et al., 2020).

However, the standard testing framework is not applicable to structured data such as time series, because the i.i.d. assumption often does not hold. As a result, standard testing procedures like the permutation test are known to produce inflated p-values and are thus unsuitable for testing structured data (Guillot & Rousset, 2013; DiCiccio & Romano, 2017). Existing research on testing independence for temporal data is limited, often relying on linear measures such as autocorrelation and cross-correlation, which may overlook potential nonlinear relationships (Guochang et al., 2021). A commonly made assumption is to consider the sample data as stationary, meaning that the joint distribution of (X_t, Y_{t-l}) depends only on the lag l and not on any specific time index t. Approaches for addressing the instantaneous time problem, where the goal is to detect whether X_t and Y_t are independent, have been explored in Chwialkowski & Gretton (2014). Moreover, Chwialkowski et al. (2014) investigates the problem of testing between X_t and Y_{t-l} for each lag l separately, employing multiple testing techniques.

In this paper, we propose an aggregated temporal statistic and utilize a block permutation procedure to extend the scope of independence testing beyond the i.i.d. assumption. Given a standard dependence measure

such as distance correlation, our method first calculates a set of cross dependence statistics. These statistics not only facilitate the estimation of the optimal dependence lag, but also enable the computation of the temporal dependence statistic as a weighted aggregation of all cross dependence statistics. Subsequently, we employ a block permutation procedure to derive a p-value for hypothesis testing. Under proper assumptions regarding the choice of the dependence measure, the joint distribution of the temporal data, and the parameters of the block permutation, we establish the asymptotic properties of the temporal dependence, and prove the asymptotic validity and universal consistency of our method. Notably, the proposed temporal dependence method is non-parametric and does not require multiple testing.

Numerically, we show that the proposed approach yields satisfactory testing power when applied to simulated time series with small sample sizes. It is compatible with various dependence measure choices, and numerically superior and more versatile than previously proposed time series testing procedures. Additionally, we present the results of a real-data experiment that utilizes the proposed method to analyze neural connectivity based on fMRI data.

2 Method

2.1 Hypothesis for Testing Temporal Dependence

Given the joint sample data $\{(X_1, Y_1), ..., (X_n, Y_n)\}$, let $\vec{X} = \{X_1, \ldots, X_n\} \in \mathbb{R}^{p\times n}$ and $\vec{Y} = \{Y_1, \ldots, Y_n\} \in \mathbb{R}^{q\times n}$ represent each individual sample data. Here, p and q denote the dimensions and are positive integers, and n is the sample size.

Suppose $(\vec{X}, \vec{Y})$ is strictly stationary, meaning the distribution at any set of indices remains the same. We can represent the distributions of X_t and Y_t at any point t as F_X and F_Y, and represent the distribution of (X_t, Y_{t-l}) as $F_{XY_{-l}}$ for each lag $l \geq 0$.

We aim to test the following independence hypothesis between $\vec{X}$ and $\vec{Y}$:

$$H_0 : F_{XY_{-l}} = F_X F_Y \text{ for each } l \in \{0, 1, ..., L\}$$
$$H_A : F_{XY_{-l}} \neq F_X F_Y \text{ for some } l \in \{0, 1, ..., L\},$$

Here, L is a non-negative integer denoting the maximum lag under consideration. Essentially, the null hypothesis states that X_t is independent of present and past values of Y_{t-l} for all of $l = 0, \ldots, L$. In contrast, the alternative hypothesis suggests $(\vec{X}, \vec{Y}_{-l})$ are dependent for at least one l in the range of $[0, L]$.

This setting is, in fact, a generalization of the standard i.i.d. setting, where it was assumed that $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n) \overset{i.i.d.}{\sim} F_{XY}$, and the null hypothesis simplifies to $F_{XY} = F_X F_Y$ because there is no possible dependence other than $l = 0$. Hence, our subsequent method and theory for testing two time series are also applicable when only one of them is time series or when both are standard i.i.d. data. Moreover, they are applicable to any general structured data that can be assumed stationary.

2.2 Main Algorithm

The proposed method consists of four steps: computation of the cross-lag dependence statistics, estimation of the optimal dependence lag, computation of temporal dependence statistic, and block permutation to obtain the p-value for testing purposes. Details regarding the choice of the dependence measure, block permutation, and computational complexity are discussed in the following subsections.

Input: Two jointly-sampled datasets represented as $\vec{X} \in \mathbb{R}^{p\times n}$ and $\vec{Y} \in \mathbb{R}^{q\times n}$, a given choice of sample dependence measure $\tau_n(\cdot, \cdot) : \mathbb{R}^{p\times n} \times \mathbb{R}^{q\times n} \to \mathbb{R}$, and three positive integers: the lag limit L, the number of blocks B, and the number of random permutations R.

Step 1: Compute the set of cross dependence sample statistics $\{\tau_n(\vec{X}, \vec{Y}_{-l}), l = 0, \ldots, L\}$. Here, $(\vec{X}, \vec{Y}_{-l})$ denotes the sample data with l lags apart, which consists of $(n - l)$ pairs of observations:

$$(\vec{X}, \vec{Y}_{-l}) = \{(X_{1+l}, Y_1), (X_{2+l}, Y_2), \ldots, (X_n, Y_{n-l})\}.$$

Step 2: Estimate the optimal dependence lag:

$$\hat{L}^* = \arg\max_{l \in [0,L]} \left(\frac{n-l}{n}\right) \cdot \tau_n(\vec{X}, \vec{Y}_{-l}).$$

Here, the weight $\left(\frac{n-l}{n}\right)$ simply weights each cross dependence statistic based on the number of observations it uses.

Step 3: Compute the temporal dependence sample statistic:

$$\mathrm{T}_n(\vec{X}, \vec{Y}) = \sum_{l=0}^{L} \left(\frac{n-l}{n}\right) \cdot \tau_n(\vec{X}, \vec{Y}_{-l}).$$

Step 4: Compute the p-value using block permutation:

$$\text{p-val} = \sum_{r=1}^{R} \mathrm{I}(\mathrm{T}_n(\vec{X}, \vec{Y}) > \mathrm{T}_n(\vec{X}, \vec{Y}_{\pi_B}))/R,$$

where $\mathrm{I}(\cdot)$ is the 0-1 indicator function, and π_B is a randomly generated block permutation for each r.

Output: The temporal dependence statistic T, the corresponding p-value, and the estimated optimal dependence lag $\hat{L}^*$.

The null hypothesis is rejected if the p-value is less than a pre-specified Type 1 error level, such as 0.05.

2.3 Choice of Dependence Measure

While the algorithm can accommodate any dependence measure as the choice of $\tau_n(\cdot, \cdot)$, it is essential for the chosen measure to be well-behaved and satisfy the required assumptions outlined in Section 3.1. This ensures consistency in detecting dependence between temporal data, both in terms of performance and subsequent theory. In our experiments, we employed distance correlation, Hilbert-Schmidt independence criterion, and multiscale graph correlation. All of these measures meet the necessary assumptions, and the resulting tests appear valid and consistent in our numerical experiments.

As the proposed temporal statistic is essentially an aggregation of the underlying dependence measure, its effectiveness in capturing dependence is contingent upon the choice of dependence measure. It is well known that each dependence measure has its own unique strengths. Therefore, our usage of distance and kernel based statistics in this paper should be viewed as an illustration of the validity and consistency properties of the proposed temporal test.

Some examples of other dependence measures include correlation coefficients (Fukumizu et al., 2007; Bießmann et al., 2010), Chatterjee's rank correlation (Chatterjee, 2021; Shi et al., 2021; 2022), the HHG method (Heller et al., 2013; 2016), projection correlation (Zhu et al., 2017), ball covariance (Pan et al., 2020), as well as recent high-dimensional dependence statistics (Zhu et al., 2020; Huang & Huo, 2022; Shen & Dong, 2024; Xu et al., 2024; Zhou et al., 2024). All of these dependence measures can be directly incorporated into our temporal testing framework by simply modifying the cross-dependence statistics in Step 1. Such adaptations may offer better testing power for certain dependence structures.

For instance, using the correlation coefficient with block permutation will only detect linear associations in temporal data, while a universally consistent dependence measure can detect all possible dependencies with a sufficiently large sample size; dependence measures that are better at detecting nonlinear or high-dimensional dependencies in standard i.i.d. data will also perform better under such dependencies in the case of temporal data, requiring a smaller sample size to achieve perfect testing power; rank-based dependence measures can be more robust against data noise.

2.4 The Block Permutation Test

The standard permutation test is widely used for independence testing (Good, 2005). In a standard permutation, $\pi(\cdot)$ randomly permutes the indices $1, 2, \ldots, n$, resulting in $\vec{Y}_\pi$ and $\vec{X}$ that are mostly independent (except for a few indices that do not change position, which are asymptotically negligible as n increases). Given sufficiently many random permutations, this process allows the permuted test statistics to estimate the true null distribution.

However, the above is only true under the standard i.i.d. setting, and it no longer holds when there exists structural dependence within the sample sequence, such as when (X_t, Y_t) are dependent with (X_{t-1}, Y_{t-1}). Specifically, the permuted statistics would under-estimate the true null distribution, leading to an inflation of the testing power. This issue has been noted in Guillot & Rousset (2013); DiCiccio & Romano (2017), which can affect any dependence measure that relies on the standard permutation test.

To ensure validity of the test, we employ a block permutation procedure (Politis, 2003) denoted as $\pi_B(\cdot)$, where B denotes the number of blocks. The construction of $\pi_B(\cdot)$ proceeds as follows:

We partition the index list into B consecutive blocks. For $j = 1, \ldots, B$, block j consists of indices

$$B_j = (\lceil \frac{n}{B} \rceil * (j-1) + 1, \lceil \frac{n}{B} \rceil * (j-1) + 2, \ldots, \lceil \frac{n}{B} \rceil * j - 1).$$

Note that for the last block, the last few indices may exceed n, in which case the indices wrap around and restart from 1.

As an example, consider a sample size of $n = 100$ and $B = 20$ blocks, with each block containing 5 indices. Then the first block would be $(Y_1, Y_2, ..., Y_5)$, the second block would be $(Y_6, Y_7, ..., Y_{10})$, etc. During the block permutation process, each block is shifted to another position. For instance, the first block might be permuted to the fourth block, resulting in $\pi_B(1) = 16, \pi_B(2) = 17, \pi_B(3) = 18, \pi_B(1) = 19, \pi_B(1) = 20$. This shuffling of blocks ensures a randomized distribution of data while maintaining the block structure.

2.5 Parameter Choice and Computational Complexity

The choice of the maximum lag, denoted as L, is typically determined based on subject matter considerations. For example, if the signal from one region of the brain can only influence another region within a range of 20 time steps, then setting $L = 20$ would be appropriate. Similarly, when collecting daily stock trading data for two stocks, choosing $L = 30$ indicates that we are examining the dependence structure within the past month.

As for the number of blocks, we used $B = 20$ in our experiments, which is sufficient for our purposes. For the number of permutation, we used $R = 1000$ replicates. Assuming that the dependence measure can be computed in $O(n^2)$ time complexity (which is the case for distance correlation), the temporal independence test has a time complexity of $O(n^2RL)$.

3 Supporting Theory

In this section, we establish the asymptotic properties of the test statistics and the resulting tests, which include asymptotic convergence, validity, and consistency. All the proofs can be found in the appendix. We begin by outlining the necessary assumptions for the theoretical results, followed by detailed elaborations on each assumption.

3.1 Assumptions

- The observed data $\{(X_t, Y_t)\}_{t=1}^n$ is strictly stationary, non-constant, and the underlying distribution $F_{XY_{-l}}$ has finite moments for any lag $l \geq 0$.

- There exists a maximum dependence lag M such that for all $l \geq M$, the two time series are almost independent for large n, so are each time series within itself:

$$\sup |F_{XY_{-l}} - F_X F_Y| = O(\frac{1}{n}),$$
$$\sup |F_{XX_{-l}} - F_X F_X| = O(\frac{1}{n}),$$
$$\sup |F_{YY_{-l}} - F_Y F_Y| = O(\frac{1}{n}).$$

- The maximum dependence lag M and the maximum lag under consideration L are non-negative integers that satisfies $L \geq M$ and $L = o(n)$, i.e., they may increase together with n but at a slower pace.

- As the sample size n increases, both the number of blocks B and the number of observations per block $\frac{n}{B}$ increase to infinity. Moreover, $\frac{n}{B} \geq M$ for sufficiently large n.

- The sample dependence measure has the following form:

$$\tau_n(\vec{X}, \vec{Y}) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_n(i,j)}{n^2},$$

  where each $\gamma_n(i,j)$ is a function of (X_i, X_j, Y_i, Y_j), and remaining sample pairs may also be used but with a weight of $O(1/n)$.

- In the standard i.i.d. setting where $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n) \overset{i.i.d.}{\sim} F_{XY}$, there exists a population statistic $\tau(X, Y)$ defined solely based on the joint distribution F_{XY}. When $i \neq j$, each term in the sample statistic satisfies:

$$\mathbb{E}(\gamma_n(i,j)) = \tau(X, Y) + o(1).$$

  Moreover, the population statistic $\tau(X, Y)$ is non-negative and equals 0 if and only if X and Y are independent, i.e., $F_{XY} = F_X F_Y$.

The first assumption is a common one in time series research. The key distinction from the standard i.i.d. setting is that the samples are no longer independent, but remain identically distributed. For non-stationary data, there exist many common techniques to remove trends and process them into approximately stationary processes (Cleveland et al., 1990; Hastie et al., 2009; Enders, 2010; Shumway & Stoffer, 2010; Box et al., 2015). Some examples include differencing, where one computes the difference between consecutive observations; detrending via linear regression or polynomial fitting and subtracting the trend component from the original series; seasonal adjustment by decomposition; log / square root / Box-Cox transformation to stabilize variance; smoothing via moving averages to reduce noise and short-term fluctuations; filtering to remove specific frequencies from the data.

The second and third assumptions require that the time series exhibit independence for sufficiently large lags beyond M, and that the maximum lag to be examined, L, must be no less than M. Such an assumption shares similarity with the mixing property, where a stochastic process is mixing if its values at widely-separated times are asymptotically independent (Pham & Tran, 1985; McDonald et al., 2011; Ziemann & Tu, 2022). Hence, our results can also be considered approximately true for mixing time series.

The fourth assumption imposes a regularity condition on block permutation. In theory, choices for B can be $log(n)$ or $\sqrt{n}$, while a practical choice like $B = 20$ is sufficient for our simulations. This resembles the Bayes optimal condition for K-nearest-neighbor, where K is required to increase to infinity but slower than n.

The remaining assumptions regarding the dependence measure are satisfied by a variety of distance and kernel measures that have been recently proposed. For example, distance covariance satisfies the two assumptions, with

$$\gamma_n(i,j) = \{d(X_i, X_j) - \mu_{X_i} - \mu_{X_j} + \mu_X\}\{d(Y_i, Y_j) - \mu_{Y_i} - \mu_{Y_j} + \mu_Y\}.$$

Here, $d(\cdot,\cdot)$ is the Euclidean distance, μ_{X_i} denotes the mean of all distance pairs relative to X_i within $\vec{X}$, and μ_X is the mean of the whole pairwise distance matrices of $\vec{X}$. Furthermore, the population distance covariance is defined in terms of characteristic functions and equals 0 if and only if $F_{XY} = F_X F_Y$ in the standard i.i.d. settings. Indeed, many dependence measures that are universal consistent in the standard i.i.d. setting satisfy this assumption. For example, the Hilbert-Schmidt independence criterion utilizes the same formulation (Shen & Vogelstein, 2021; Sejdinovic et al., 2013) on the Gaussian kernel. Additionally, the unbiased distance covariance and distance correlation, as well as the multiscale graph correlation – a truncated version of distance correlation where large distance pairs may be unused – also satisfy this assumption.

3.2 Convergence of the Sample Statistics

We begin by proving the convergence of the sample cross dependence to the population cross dependence:

Theorem 1. *The cross dependence sample statistic satisfies:*

$$\mathbb{E}(\tau_n(\vec{X},\vec{Y}_{-l})) - \tau(X,Y_{-l}) = o(1),$$
$$Var(\tau_n(\vec{X},\vec{Y}_{-l}))) = O(\frac{1}{n-l}).$$

Therefore, for each $l \in \{0,...,L\}$, we have

$$\tau_n(\vec{X},\vec{Y}_{-l}) \overset{n\to\infty}{\to} \tau(X,Y_{-l})$$

in probability.

Theorem 1 shows that both the bias and variance of the cross dependence statistic diminish to 0 as the sample size n increases. Consequently, this guarantees that the aggregated temporal dependence statistic and the estimated optimal lag also converge to their corresponding population forms in probability.

Theorem 2. *The temporal dependence sample statistic satisfies:*

$$\mathrm{T}_n(\vec{X},\vec{Y}) \overset{n\to\infty}{\to} \sum_{l=0}^{L} \tau(X,Y_{-l}).$$

The estimated optimal dependence lag satisfies:

$$\hat{L}^* \overset{n\to\infty}{\to} \arg\max_{l\in[0,L]} \tau(X,Y_{-l}).$$

3.3 Validity and Consistency for Testing Temporal Independence

In this subsection we establish the validity and consistency of the method. Specifically, if $\vec{X}$ and $\vec{Y}$ are independent, the power of the test equals the Type 1 error level α. Conversely, if $\vec{X}$ and $\vec{Y}$ are dependent, the power of the test converges to 1, and the method can consistently detect any dependence.

Given $\mathrm{T}_n(\vec{X},\vec{Y})$ as the observed test statistic, let $F_{T_n^B}(z)$ be the empirical distribution of the block-permuted statistics $\{\mathrm{T}_n(\vec{X},\vec{Y}_{\pi_B})\}$, and denote $z_{n,\alpha}$ as the critical value where:

$$F_{T_n^B}(z)(z_{n,\alpha}) = 1-\alpha.$$

The following theorem establishes the asymptotic validity of our block permutation test:

Theorem 3 (Asymptotic Validity)**.** *Under the null hypothesis that $\vec{X}$ and $\vec{Y}$ are independent for all lags $l \in [0,L]$, the test statistic satisfies:*

$$\mathrm{T}_n(\vec{X},\vec{Y}) \overset{n\to\infty}{\to} 0.$$

Moreover, the block-permutation test is asymptotically valid, i.e.,

$$Prob(\mathrm{T}_n(\vec{X},\vec{Y}) \geq z_{n,\alpha}) \overset{n\to\infty}{\to} \alpha.$$

The next theorem proves that the method is universally consistent against any alternative.

Theorem 4 (Testing Consistency)**.** *Under the alternative hypothesis that $\vec{X}$ and $\vec{Y}$ are dependent for some lag $l \in [0, L]$, the test statistic satisfies*

$$\mathrm{T}_n(\vec{X}, \vec{Y}) \overset{n\to\infty}{\to} c > 0.$$

Moreover, the block-permutation test is asymptotically consistent, i.e.,

$$Prob(\mathrm{T}_n(\vec{X}, \vec{Y}) \geq z_{n,\alpha}) \overset{n\to\infty}{\to} 1.$$

4 Simulations

We estimated the testing power of the proposed approach through simulations on various temporal dependence structures. Specifically, we considered three different implementations of the proposed temporal dependence statistic, which utilized distance correlation (`DCorr`), Hilbert-Schmidt independence criterion (`HSIC`), and multiscale graph correlation (`MGC`). For comparison, we included `ShiftHSIC` (Chwialkowski & Gretton, 2014), `WildHSIC` (Chwialkowski et al., 2014), and the widely recognized `Ljung-Box` test (Ljung & Box, 1978) using traditional cross-correlations. Each simulation was repeated 300 times, with 1000 permutations and a Type 1 error level of $\alpha = 0.05$ used to compute the p-values. The testing power is measured by how often the p-value is lower than 0.05 out of the 300 Monte-Carlo simulations. Analysis of `ShiftHSIC` and `WildHSIC` was performed using MATLAB code[1] and `wildBootstrap`[2].

4.1 Testing Power Evaluation

Independence First, we check the validity of the tests by generating two independent, stationary autoregressive time series with a lag of one:

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} \phi & 0 \\ 0 & \phi \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ \eta_t \end{bmatrix}.$$

Here, (ϵ_t, η_t) are standard normal noise terms. As shown in Figure 1, the proposed methods maintain a testing power close to $\alpha = 0.05$ across varying n and ϕ, regardless of the statistic used.

Linear Dependence Next, we assess our methods' ability to capture linear relationships in the following simulation:

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} 0 & \phi \\ \phi & 0 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ \eta_t \end{bmatrix}.$$

As this represents a straightforward linear relationship, the `Ljung-Box` test, based on auto-correlation, is expected to perform best. This is indeed the case in the left panel of Figure 2. Our proposed methods using `DCorr`, `MGC`, and `HSIC` follow closely, quickly converging to perfect power around $n = 100$. In contrast, the other competitors do not perform well in this scenario. This is not surprising, as the `ShiftHSIC` method is designed to detect whether X_t and Y_t are dependent at lag 0, whereas the linear dependence here is of lag 1. The `WildHSIC` method used a wild bootstrap method to estimate the null distribution, which can be inaccurate at small sample size.

Nonlinear Dependence The next simulation considers a nonlinear dependent model:

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} \epsilon_t Y_{t-1} \\ \eta_t \end{bmatrix}.$$

In the right panel of Figure 2, our proposed methods utilizing `DCorr`, `MGC`, and `HSIC` demonstrate superior performance compared to other competing methods. Notably, the `HSIC` and `MGC` implementations exhibit better finite-sample power, as these two dependence measures are better at identifying nonlinear relationships than `DCorr`. In contrast, all other tests fail to detect dependence in this scenario.

[1] `https://github.com/kacperChwialkowski/HSIC/`
[2] `https://github.com/kacperChwialkowski/wildBootstrap`

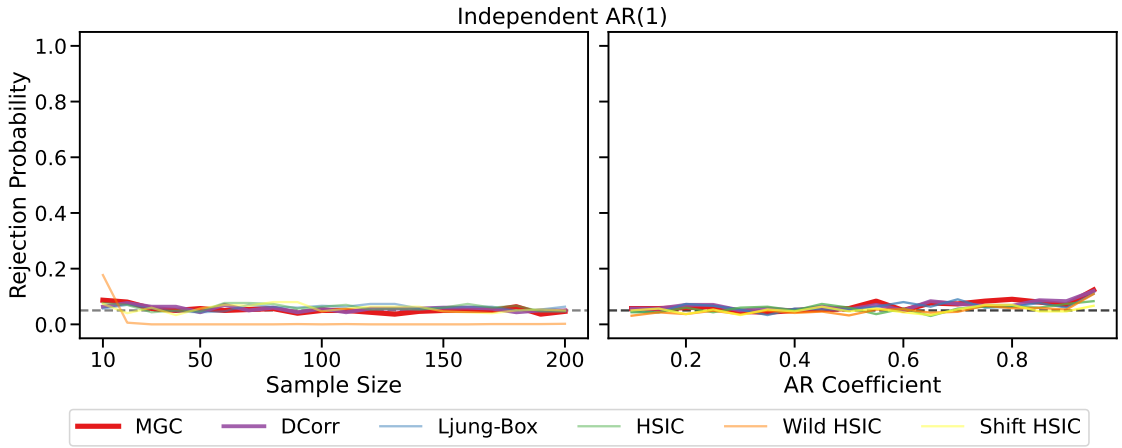

Figure 1: This figure illustrates the validity of the tests using two independent time series. In the left panel, the testing power is computed as the sample size increases, with an AR coefficient of $\phi = 0.5$. The right panel keeps the sample size at $n = 1200$ while varying the AR coefficient ϕ, with the noise variance appropriately adjusted by $(1 - \phi^2)$, based on the same simulation as in Chwialkowski & Gretton (2014). The dashed black line represents the significance level $\alpha = 0.05$.

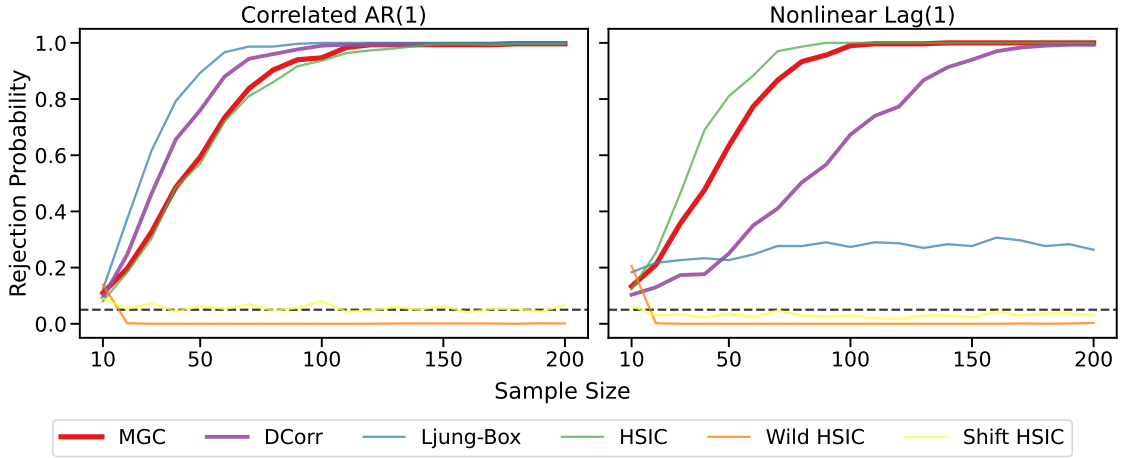

Figure 2: The testing power for linear (left panel) and nonlinear (right panel) simulations based on 300 replicates.

Extinct Gaussian This simulation uses the same extinct Gaussian process from Chwialkowski & Gretton (2014), where

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} \phi & 0 \\ 0 & \phi \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ \eta_t \end{bmatrix},$$

and we set $n = 1200$. Here, the (ϵ_t, η_t) pair are dependent and drawn from an Extinct Gaussian distribution with two additional parameters: e (extinction rate) and r (radius). Both variables are initially drawn from independent standard normal, and U is sampled from standard uniform. If either $\epsilon_t^2 + \eta_t^2 > r$ or $U > e$ holds, then (ϵ_t, η_t) are returned; otherwise, they are discarded and the process is repeated. In this process, the dependence between ϵ_t and η_t increases with extinction rate e. Therefore, we expect power to increase with the extinction rate, which is indeed the case as shown in Figure 3. While all methods, except `Ljung-Box`, are consistent and eventually achieve perfect power, our proposed method using `MGC` stands out as the best performer.

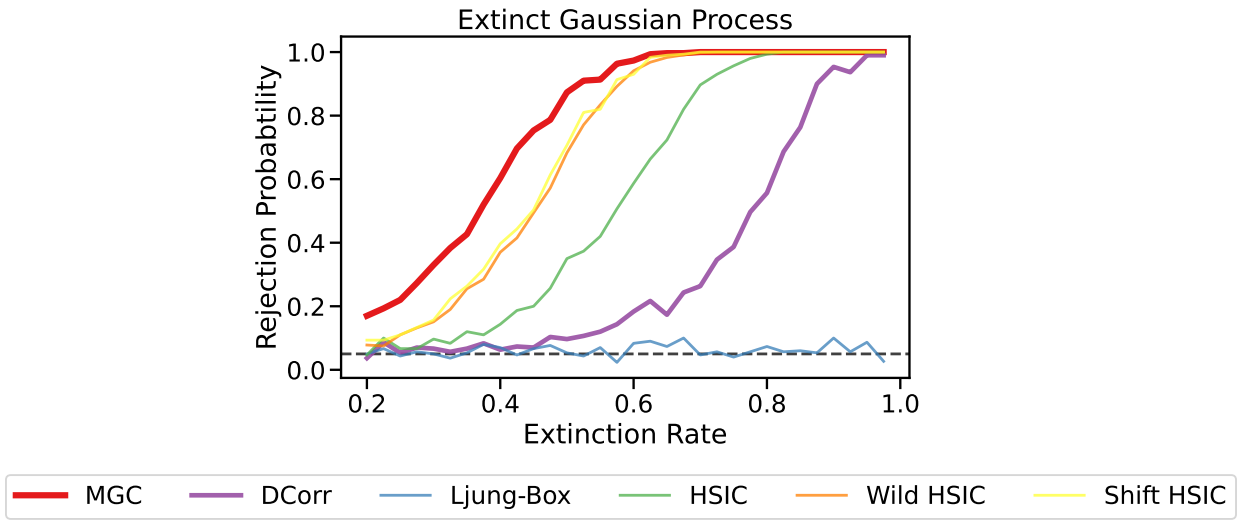

Figure 3: The testing power for the extinct gaussian simulation based on 300 replicates.

4.2 Optimal Dependence Lag Estimation

In this subsection, we evaluate the method's performance in estimating the optimal dependence lag in both linear and nonlinear settings. The linear setting is

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} 0 & \phi_1 \\ \phi_1 & 0 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} 0 & \phi_3 \\ \phi_3 & 0 \end{bmatrix} \begin{bmatrix} X_{t-3} \\ Y_{t-3} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ \eta_t \end{bmatrix},$$

where we set $\phi_3 = 0.8 > \phi_1 = 0.1$ such that the true optimal dependence lag equals 3. The nonlinear simulation is

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} \epsilon_t Y_{t-3} \\ \eta_t \end{bmatrix}.$$

In both simulations, the true optimal dependence lag equals 3. Figure 4 shows that the proposed method using either `DCorr` or `MGC` consistently estimates the optimal dependence lag as the sample size increases, and `MGC` outperforms `DCorr` in the nonlinear setting.

4.3 Multivariate Simulations

In this subsection, we revisit the testing power and dependence lag estimation in both linear and nonlinear settings, maintaining a fixed sample size of $n = 100$ and increasing the dimensionality p, to evaluate performance for multivariate data.

For testing power evaluation, we use the following multivariate linear setting:

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} 0 & \phi D \\ \phi D & 0 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ \eta_t \end{bmatrix},$$

where $\phi = 0.65$, $D \in \mathbb{R}^{p \times p}$ is a diagonal matrix where the elements are $D_{ii} = 1/i$, and ϵ_t, η_t are standard normal of dimension p. In a similar manner, we use the following multivariate nonlinear setting:

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} D(\epsilon_t \odot Y_{t-1}) \\ \eta_t \end{bmatrix},$$

where $\odot$ denotes element-wise multiplication. We intentionally design the matrix D as a decaying weight, reflecting a meaningful multivariate simulation where additional dimensions contain weaker dependence signals.

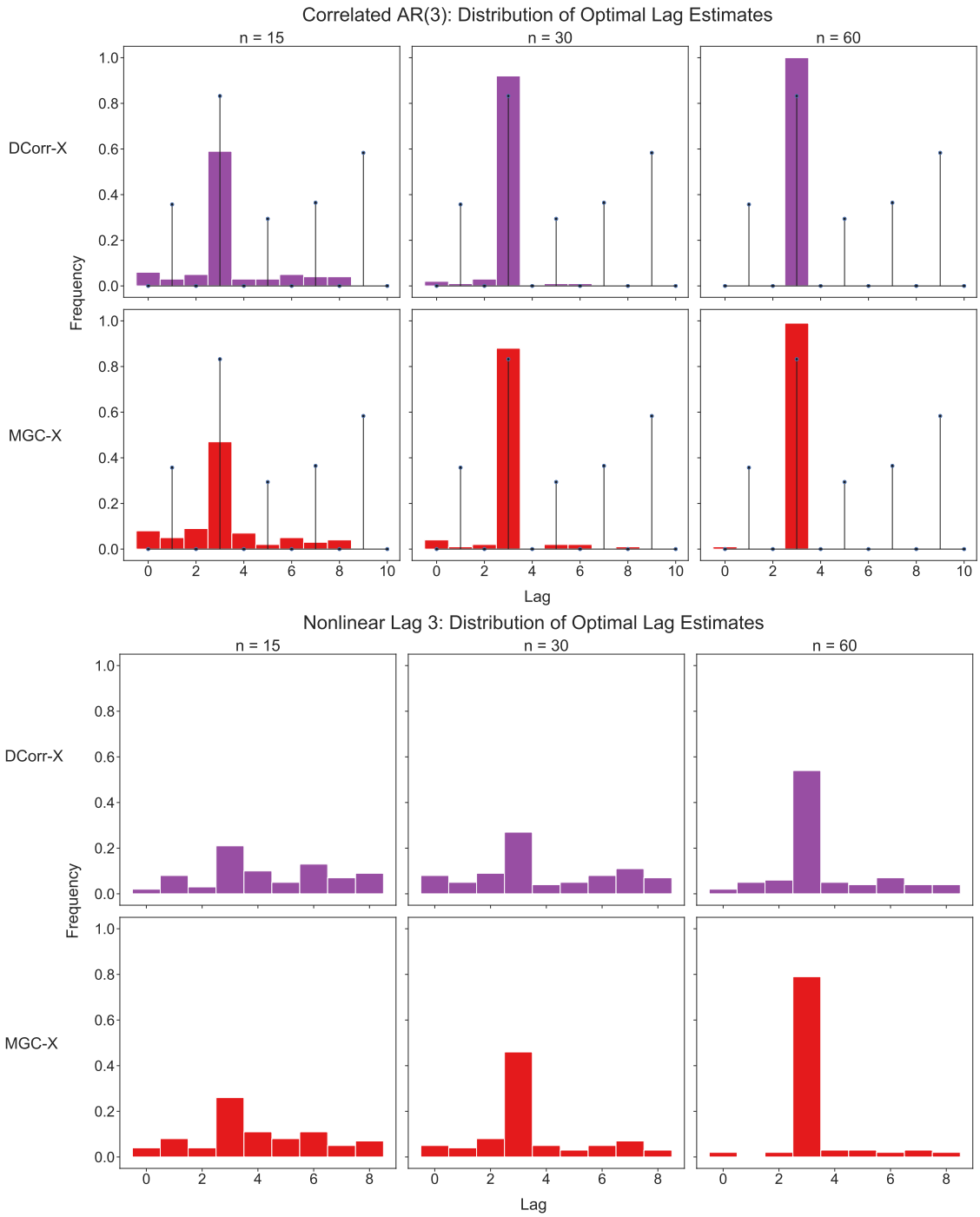

Figure 4: This figure displays the performance of our proposed method using both `MGC` and `DCorr` for estimating the optimal dependence lag $\hat{L}^*$ in linear and nonlinear relationships. The colored bar above lag l shows the empirical frequency of $\hat{L}^* = j$, with red representing `MGC` and purple representing `DCorr`. The probability is estimated based on 100 trials. The first row shows `DCorr` estimation performance at sample sizes $n = 15, 30, 60$ for linear relationships, while the second row shows the `MGC` performance on the same data. The third row displays `DCorr` estimation for nonlinear relationships, and the last row presents the same for `MGC`.

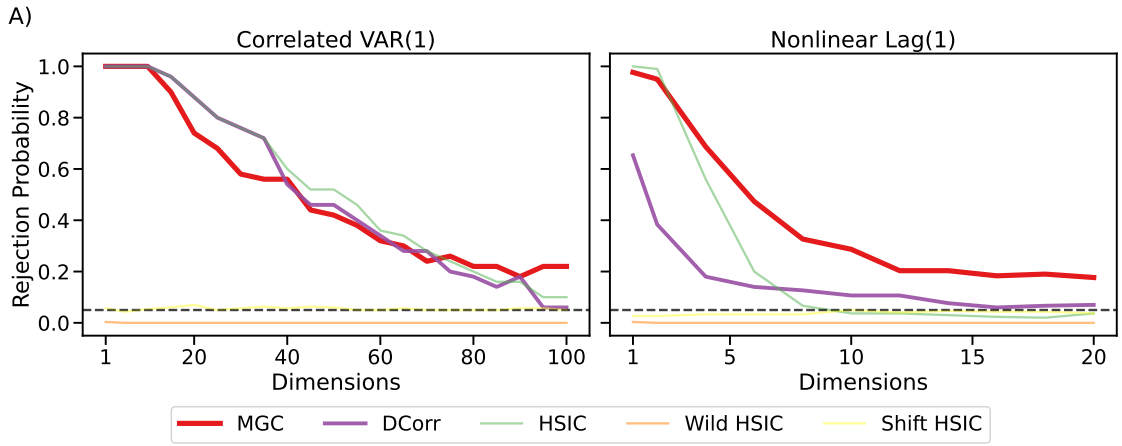

Figure 5: This figure shows the testing power for multivariate simulations, with a constant sample size of $n = 100$ while increasing the dimensionality.

Figure 5 illustrates the testing power as dimensionality increases. At a fixed sample size, all testing powers gradually decrease as p increases. The proposed method using any of `MGC`, `DCorr`, or `HSIC` maintains relatively stable power with slow degradation in the case of linear dependence. The same trend is observed for nonlinear dependence, although the degradation is faster, with the `MGC` statistic performing the best. It is worth emphasizing that due to the consistent property, if we fix p and let n increase, the testing power for our method shall increase to 1.

Similarly, we extend the optimal lag estimation into the following two multivariate settings:

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} 0 & \phi_1 D \\ \phi_1 D & 0 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} 0 & \phi_3 D \\ \phi_3 D & 0 \end{bmatrix} \begin{bmatrix} X_{t-3} \\ Y_{t-3} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ \eta_t \end{bmatrix}.$$

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} D\epsilon_t Y_{t-3} \\ \eta_t \end{bmatrix},$$

where $\phi_1 = 0.1$ and $\phi_3 = 0.65$. These settings are similar to those in Section 4.2, with the addition of increasing dimension and D. The true optimal lag remains 3. The estimation accuracy, as shown in Figure 6, demonstrates successful detection at small p, with accuracy gradually degrading as p increases.

5 Analyzing Connectivity in the Human Brain

This study is based on data from an individual (Subject ID: 100307) of the Human Connectome Project (HCP), which can be downloaded online[3]. The human cortex is parcellated into 180 parcels per hemisphere using the HCP multi-modal parcellation atlas (Glasser et al., 2016). For this study, 22 parcels were selected as regions of interest (ROIs), representing various locations across the cortex. These parcels are denoted as $X^{(1)}, \ldots, X^{(22)}$. Each parcel consists of a contiguous set of vertices whose fMRI signal is projected on the cortical surface. Averaging the vertices within a parcel yields a univariate time series $X^{(u)} = (X_1^{(u)}, \ldots, X_n^{(u)})$, where $n = 1200$ in this particular case. The selected ROIs, their parcel number in the HCP multi-modal parcellation (Glasser et al., 2016), and assigned network are listed in Table 1.

As the temporal dependence method using `MGC` performed well in our simulations, we simply use the `MGC` implementation in this analysis. In the left panel of Figure 7, we present the optimal dependence lag for each interdependency, ranging up to $L = 10$. Meanwhile, the right panel of Figure 7 displays the log-scale p-values of temporal dependence for each pair of parcels. Generally, we observe strong relationships with small lags within the same region, such as an optimal lag of usually 0 within the "DMN" region with significant p-values. In contrast, inter-region dependencies are less significant and typically exist at longer lags.

[3] `https://www.humanconnectome.org/study/hcp-young-adult/data-releases`

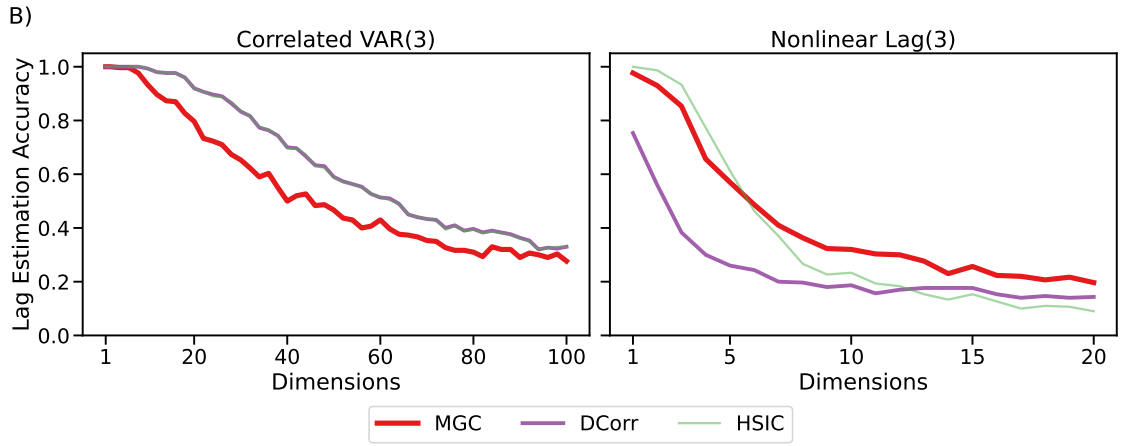

Figure 6: This figure shows the accuracy of estimating the true optimal lag in multivariate simulations, maintaining a constant sample size of $n = 100$ while increasing the dimensionality. Note that in the left panel, the lines representing `DCorr` and `HSIC` overlap due to their almost identical performance.

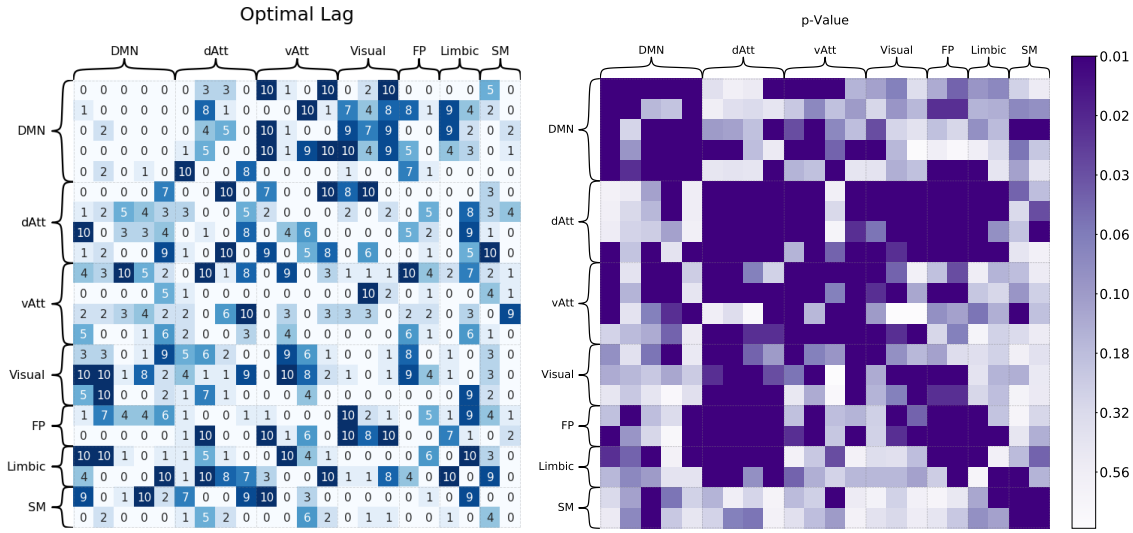

Figure 7: This displays the results of applying the temporal dependence method using `MGC` to resting-state fMRI data.

ROI ID	Network	Shorthand	Parcel Key	Parcel Name
1	Visual Network	Visual	1	V1
2	Visual Network	Visual	23	MT
3	Visual Network	Visual	18	FFC
4	Somatomotor Network	SM	53	3a
5	Somatomotor Network	SM	24	A1
6	Dorsal Attention Network	dAtt	96	6a
7	Dorsal Attention Network	dAtt	117	API
8	Dorsal Attention Network	dAtt	50	MIP
9	Dorsal Attention Network	dAtt	143	PGp
10	Ventral Attention Network	vAtt	109	MI
11	Ventral Attention Network	vAtt	148	PF
12	Ventral Attention Network	vAtt	60	p32pr
13	Ventral Attention Network	vAtt	38	23c
14	Limbic Network	Limbic	135	TF
15	Limbic Network	Limbic	93	OFC
16	FrontoParietal Network	FP	83	p9-46v
17	FrontoParietal Network	FP	149	PFm
18	Default Mode Network	DMN	150	PGi
19	Default Mode Network	DMN	65	p32pr
20	Default Mode Network	DMN	161	32pd
21	Default Mode Network	DMN	132	TE1a
22	Default Mode Network	DMN	71	9p

Table 1: This table displays the parcellation information for the parcels used in our analysis. The first column shows the numeric order of the parcels as they appear in the following figures. The second and third columns show the shorthand and full names, respectively, of the network to which each parcel belongs. The last two columns display the parcel number and name according to the Glasser et al. (2016) parcellation.

6 Conclusion

This paper introduces a new independence testing procedure for temporal data. The method combined the strengths of nonparametric dependence measures, the specialized cross-lag statistic for time series, and the block permutation procedure. As a result, it provides an asymptotically valid and universally consistent approach with outstanding numerical performance. While the exposition of this manuscript is focused on time series data, this work marks an important step in extending independence testing to structural data beyond the realm of standard i.i.d. data, making them more attractive and broadly applicable.

There are several avenues for future research that warrant exploration. Firstly, although we have demonstrated the asymptotic validity of the block permutation test, its computational efficiency remains a challenge when dealing with large sample sizes. Recent studies (Zhang et al., 2018; Shen et al., 2022) have investigated faster testing procedures by approximating the null distribution of distance and kernel correlations under the standard i.i.d. setting. Extending such approaches to structural data could significantly enhance computational scalability.

Secondly, dependence measures are commonly employed in dimension reduction techniques, such as screening (Fan & Lv, 2008; Li et al., 2012), especially in high-dimensional data settings. However, little attention has been given to the temporal domain. While it is straightforward to utilize dependence measures for dimension reduction in multivariate time series, delving into their theoretical properties and their relationships with other standard tools, such as independence component analysis, could provide valuable insights.

Thirdly, causal inference in time series data is an important task (Haufe et al., 2010; Winkler et al., 2016). While it is widely recognized that correlation does not imply causality, recent research has demonstrated the utility of dependence and conditional dependence tests in causal inference (Cai et al., 2022; Laumann et al.,

2023). Therefore, extending this framework to encompass conditional independence and causal inference may significantly advance the understanding of causal inference in time series data.

References

F. Bießmann, F. C. Meinecke, A. Gretton, A. Rauch, G. Rainer, N. K. Logothetis, and K. Müller. Temporal kernel CCA and its application in multimodal neuronal data analysis. *Machine Learning*, 79:5–27, 2010.

G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, 2015.

Zhanrui Cai, Runze Li, and Yaowu Zhang. A distribution free conditional independence test with applications to causal discovery. *Journal of Machine Learning Research*, 23:1–41, 2022.

S. Chatterjee. A new coefficient of correlation. *Journal of the American Statistical Association*, 116(536): 2009–2022, 2021.

Kacper Chwialkowski and Arthur Gretton. A kernel independence test for random processes. In *31st International Conference on Machine Learning*, pp. 1422–1430, 2014.

Kacper P Chwialkowski, Dino Sejdinovic, and Arthur Gretton. A wild bootstrap for degenerate kernel tests, 2014.

R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning. Stl: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1):3–73, 1990.

C. J. DiCiccio and J. P. Romano. Robust permutation tests for correlation and regression coefficients. *Journal of the American Statistical Association*, 112(519):1211–1220, 2017.

W. Enders. *Applied Econometric Time Series*. John Wiley & Sons, 2010.

J. Fan and J. Lv. Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B*, 70(5):849–911, 2008.

K. Fukumizu, F. R. Bach, and A. Gretton. Statistical consistency of kernel canonical correlation analysis. *Journal of Machine Learning Research*, 8:361–383, 2007.

Matthew F. Glasser, Timothy S. Coalson, Emma Claire Robinson, Carl D. Hacker, John W. Harwell, Essa Yacoub, Kâmil Uğurbil, Jesper L. R. Andersson, Christian F. Beckmann, Mark Jenkinson, Stephen M. Smith, and David C. Van Essen. A multi-modal parcellation of human cerebral cortex. *Nature*, 536: 171–178, 2016.

P. Good. *Permutation, Parametric, and Bootstrap Tests of Hypotheses*. Springer, 2005.

A. Gretton and L. Gyorfi. Consistent nonparametric tests of independence. *Journal of Machine Learning Research*, 11:1391–1423, 2010.

A. Gretton, R. Herbrich, A. Smola, O. Bousquet, and B. Scholkopf. Kernel methods for measuring independence. *Journal of Machine Learning Research*, 6:2075–2129, 2005.

A. Gretton, K. Borgwardt, M. Rasch, B. Scholkopf, and A. Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13:723–773, 2012.

Gilles Guillot and François Rousset. Dismantling the mantel tests. *Methods in Ecology and Evolution*, 4(4): 336–344, 2013.

Wang Guochang, Wai Keung Li, and Ke Zhu. New hsic-based tests for independence between two stationary multivariate time series. *Statistica Sinica*, 31(1):269–300, 2021.

T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media, 2009.

Stefan Haufe, Klaus-Robert Müller, Guido Nolte, and Nicole Krämer. Sparse causal discovery in multivariate time series. In *Proceedings of Workshop on Causality: Objectives and Assessment at NIPS 2008*, volume 6, pp. 97–106, 2010.

R. Heller, Y. Heller, and M. Gorfine. A consistent multivariate test of association based on ranks of distances. *Biometrika*, 100(2):503–510, 2013.

Ruth Heller, Yair Heller, Shachar Kaufman, Barak Brill, and Malka Gorfine. Consistent distribution-free k-sample and independence tests for univariate random variables. *Journal of Machine Learning Research*, 17(29):1–54, 2016.

C. Huang and X. Huo. A statistically and numerically efficient independence test based on random projections and distance covariance. *Frontiers in Applied Mathematics and Statistics*, 7:779841, 2022.

Felix Laumann, Julius von Kügelgen, Junhyung Park, Bernhard Schölkopf, and Mauricio Barahona. Kernel-based independence tests for causal structure learning on functional data. *Entropy*, 25(12):1597, 2023.

Y. Lee, C. Shen, C. E. Priebe, and J. T. Vogelstein. Network dependence testing via diffusion maps and distance-based correlations. *Biometrika*, 106(4):857–873, 2019.

R. Li, W. Zhong, and L. Zhu. Feature screening via distance correlation learning. *Journal of American Statistical Association*, 107:1129–1139, 2012.

G. M. Ljung and G. E. P. Box. On a measure of a lack of fit in time series models. *Biometrika*, 65(2): 297–303, 1978.

D. McDonald, C. Shalizi, and M. Schervish. Estimating beta-mixing coefficients. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pp. 516–524, 2011.

W. Pan, X. Wang, H. Zhang, H. Zhu, and J. Zhu. Ball covariance: A generic measure of dependence in banach space. *Journal of the American Statistical Association*, 115(529):307–317, 2020.

T. D. Pham and L. T. Tran. Some mixing properties of time series models. *Stochastic Processes and their Applications*, 19(2):297–303, 1985.

Dimitris N Politis. The impact of bootstrap methods on time series analysis. *Statistical Science*, 18(2): 219–230, 2003.

D. Sejdinovic, B. Sriperumbudur, A. Gretton, and K. Fukumizu. Equivalence of distance-based and rkhs-based statistics in hypothesis testing. *Annals of Statistics*, 41(5):2263–2291, 2013.

C. Shen and Y. Dong. High-dimensional independence testing via maximum and average distance correlations. *https://arxiv.org/abs/2001.01095*, 2024.

C. Shen and J. T. Vogelstein. The exact equivalence of distance and kernel methods in hypothesis testing. *AStA Advances in Statistical Analysis*, 105(3):385–403, 2021.

C. Shen, C. E. Priebe, and J. T. Vogelstein. From distance correlation to multiscale graph correlation. *Journal of the American Statistical Association*, 115(529):280–291, 2020.

C. Shen, S. Panda, and J. T. Vogelstein. The chi-square test of distance correlation. *Journal of Computational and Graphical Statistics*, 31(1):254–262, 2022.

H. Shi, M. Drton, and F. Han. On azadkia-chatterjee's conditional dependence coefficient, 2021.

H. Shi, M. Drton, and F. Han. On the power of chatterjee's rank correlation. *Biometrika*, 109(2):317–333, 2022.

R. H. Shumway and D. S. Stoffer. *Time Series Analysis and Its Applications: With R Examples*. Springer Science & Business Media, 2010.

G. Szekely and M. Rizzo. Brownian distance covariance. *Annals of Applied Statistics*, 3(4):1233–1303, 2009.

G. Szekely and M. Rizzo. Partial distance correlation with methods for dissimilarities. *Annals of Statistics*, 42(6):2382–2412, 2014.

G. Szekely, M. Rizzo, and N. Bakirov. Measuring and testing independence by correlation of distances. *Annals of Statistics*, 35(6):2769–2794, 2007.

J. T. Vogelstein, Q. Wang, E. Bridgeford, C. E. Priebe, M. Maggioni, and C. Shen. Discovering and deciphering relationships across disparate data modalities. *eLife*, 8:e41690, 2019.

Irene Winkler, Danny Panknin, Daniel Bartz, Klaus-Robert Müller, and Stefan Haufe. Validity of time reversal for testing granger causality. *IEEE Transactions on Signal Processing*, 64(11):2746–2760, 2016.

K. Xu, Y. Zhou, L. Zhu, and R. Li. Reducing multivariate independence testing to two bivariate means comparisons. *https://arxiv.org/abs/2402.16053*, 2024.

Qinyi Zhang, Sarah Filippi, Arthur Gretton, and Dino Sejdinovic. Large-scale kernel methods for independence testing. *Statistics and Computing*, 28(1):113–130, 2018.

Y. Zhou, K. Xu, L. Zhu, and R. Li. Rank-based indices for testing independence between two high-dimensional vectors. *The Annals of Statistics*, 52(1):184–206, 2024.

Changbo Zhu, Shun Yao, Xianyang Zhang, and Xiaofeng Shao. Distance-based and rkhs-based dependence metrics in high dimension. *The Annals of Statistics*, 48(6):3366–3394, 2020.

L. Zhu, K. Xu, R. Li, and W. Zhong. Projection correlation between two random vectors. *Biometrika*, 104 (4):829–843, 2017.

Ingvar Ziemann and Stephen Tu. Learning with little mixing. In *Advances in Neural Information Processing Systems*, volume 35, pp. 4626–4637, 2022.