Tests for partial correlation between repeatedly observed nonstationary nonlinear timeseries

Kenneth D. Harris1*, Alex E. Yuan2*

1UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK. * Correspondence: kenneth.harris@ucl.ac.uk, alexericyuan@gmail.com.
2Independent Researcher, Seattle WA, USA.

We describe two families of statistical tests to detect partial correlation in vectorial timeseries. The tests measure whether an observed timeseries $Y$ can be predicted from a second series $X$, even after accounting for a third series $Z$ which may correlate with $X$. They do not make any assumptions on the nature of these timeseries, such as stationarity or linearity, but they do require that multiple statistically independent recordings of the 3 series are available. Intuitively, the tests work by asking if the series $Y$ recorded on one experiment can be better predicted from $X$ recorded on the same experiment than on a different experiment, after accounting for the prediction from $Z$ recorded on both experiments.

In many fields of science, observations consist of timeseries. These timeseries are often autocorrelated, meaning that observations from a single timeseries at different times are not statistically independent. Such autocorrelations mean that standard statistical tests, which make an assumption of independent, identically distributed observations, cannot be applied 1–5. Common approaches to this problem include filtering to remove autocorrelations; corrections to standard tests such as the $t$-test for Pearson correlation6–8; and synthesis of surrogate data9. However all these approaches make assumptions such as stationarity or linearity the timeseries in question, which are often untrue or impossible to show in practice.

This problem is greatly ameliorated if one has access to multiple repetitions of the experiment. For example, consider an experiment yielding two timeseries $X$ and $Y$ which we consider as vectors of length $T$. The timeseries are statistically independent if $\mathbb{P}(X, Y) = \mathbb{P}(X)\mathbb{P}(Y)$: in other words, if there is no correlation between the entire history of $X$ and the entire history of $Y$. Given only a single observation of the vectors $X$ and $Y$, we cannot test for violation of independence without making further assumptions: methods such as standard tests for Pearson correlation, which assume independence of timepoints, give erroneously significant “nonsense correlations”10, and corrections to these make further assumptions such as stationarity. However, if the experiment is repeated $N$ times, and we assume that the histories $X_i$ and $Y_i$ ($i = 1 \ldots N$) observed on each repeat are independent vectors, we can employ a “session permutation” method to test for independence10, without any further assumptions. This method essentially asks whether the relationship between $X_i$ and $Y_i$ is larger than that between $X_i$ and $Y_j$ for $i \neq j$, using a permutation test that randomizes across experiments. The test can be based on any measure of similarity between timeseries, including but not limited to Pearson correlation.

Here, we describe two families of tests for partial correlation between repeatedly observed timeseries. We assume that each of $N$ experiments yielded 3 vectorial timeseries $X, Y$ and $Z$. We ask whether a correlation between $X$ and $Y$ exists beyond a common effect of $Z$. We do not assume that time series are stationary or linear, in fact $X, Y$ and $Z$ can be arbitrary vectors rather than timeseries.

The problem

We assume $N$ statistically independent experiments, each of which gives an observation of three vectorial timeseries: $X_i, Y_i$ and $Z_i$, for $i = 1 \ldots N$, considered as matrices of sizes $T \times p_i, T \times q_i$ and $T \times r_i$ respectively. Write $X$ for the collection of all $X_i$ and $Z_i$. We would like to test the null hypothesis that

$$Y_i = Z_iW_i + E_i$$

where $W_i$ is a known matrix and $E_i$ represents the prediction error.
where the \( E_i \) are timeseries independent of each other and of \( X \), and \( W_i \) is an unknown deterministic \( p_i \times r_i \) matrix. This null hypothesis models the idea that \( Y \) may be linearly predicted from \( Z \) but there is no further effect of \( X \) on \( Y \) once the effect of \( Z \) is considered, even though \( X \) and \( Z \) may themselves be correlated.

**Test 1: An approximate test using orthogonal projections**

For a \( T \times n \) matrix \( A \), let \( M \) be the function that maps \( A \) to the \( T \times T \) matrix performing the orthogonal projection onto the columns of \( A \). Specifically, one can take \( M(A) = UU^T \), where \( USV^T \) is the compact singular value decomposition of \( A \) (Ref. 11, p. 82).

Let \( P_i = I - M(Z_i) \), so \( P_i \) is the \( T \times T \) projection matrix orthogonal to all columns of \( Z_i \). Note that if the null hypothesis is true, \( P_i Y_i = P_i E_i \). Let \( \rho(X; Y) \) be a measure of the predictability of a vector timeseries \( Y \) from a second timeseries \( X \), for example the Pearson coefficient of multiple linear regression. For a statistical test, you might first think to ask if \( \rho(X_i; P_i Y_i) \) is bigger than \( \rho(X_j; P_j Y_j) \) for \( i \neq j \), i.e. if the relationship of \( E_i \) to \( X \) is stronger when they come from the same experiment. But \( \rho(X_i; P_i Y_i) = \rho(X_i; P_i E_i) \) is not identically distributed to \( \rho(X_j; P_j Y_j) = \rho(X_j; P_j E_j) \) for \( j \neq i \), and even if \( E \) is independent of both \( X_i \) and \( X_j \) and \( P_i \) can be dependent under the null hypothesis but \( X_i \) and \( P_i \) cannot.

Instead, let \( P_{i,j} = I - M([Z_i, Z_j]) \) where \([Z_i, Z_j]\) is a horizontal concatenation; thus \( P_{i,j} \) is the \( T \times T \) projection matrix orthogonal to the columns of both \( Z_i \) and \( Z_j \). Note that \( P_{j,i} = P_{i,j} \) because the orthogonal projection onto a given subspace is a unique matrix (Ref. 11, p. 82). We will compare \( \rho(X_i; P_{i,j} Y_i) \) to \( \rho(X_j; P_{i,j} Y_j) \). Specifically, define

\[
G_i = \frac{1}{N} \sum_{j=1}^{N} \rho(X_i; P_{i,j} Y_i) - \rho(X_j; P_{i,j} Y_j)
\]

Since \( P_{i,j} Y_i = P_{i,j} E_i \), it follows that \( G_i = \frac{1}{N} \sum_{j=1}^{N} \rho(X_i; P_{i,j} E_i) - \rho(X_j; P_{i,j} E_i) \). Under the null hypothesis, the \( E_i \) are independent, and thus the \( G_i \) are conditionally independent given \( X \), although they need not be identically distributed.

Under the null, the expectation of \( \sum_{i=1}^{N} G_i \) conditional on \( X \) is zero. To see this, first use the fact that \( P_{j,i} = P_{i,j} \) to rewrite it as:

\[
\sum_{i=1}^{N} G_i = \frac{1}{N} \sum_{i,j=1}^{N} \rho(X_i; P_{i,j} E_i) - \rho(X_j; P_{i,j} E_j)
\]

Next, because \( E_i \) and \( E_j \) are identically distributed and independent of everything else,

\[
\mathbb{E}[\sum_{i=1}^{N} G_i | X] = \frac{1}{N} \sum_{i,j=1}^{N} \mathbb{E}[\rho(X_i; P_{i,j} E_i) - \rho(X_j; P_{i,j} E_j) | X] = 0
\]

Therefore, under the null, the statistics \( G_i \) are statistically independent conditional on \( X \), with expectations that sum to zero. They are marginally

---

**Figure 1: detecting a genuine correlation.** In this simulation \( X_i \) and \( Y_i \) are both equal to a sum of two step functions, plus independent additive noise. One step function is constant across experiments. \( Z_i \) equals the constant step function, and after removing it from \( Y_i \), the residual \( P_i Y_i \) reliably more correlated to \( X_i \) than to \( X_j \), and statistical significance is detected.
Harris

Partial correlations in repeatedly observed timeseries

3

exchangeable by independence of experiments, and are thus marginally identically distributed. Nevertheless, they are not necessarily either conditionally or marginally Gaussian. To test the null hypothesis in such circumstances, a t-test that the mean of $G_i$ is zero seems to be a reasonable approximate approach, provided the distributions of the $G_i$ are close to symmetric $^{12,13}$; this can be checked by usual methods such as histograms or QQ plots.

An illustration of Test 1

We illustrate this test with a simple example (Figs. 1, 2) in which $X$ and $Y$ are nearly identical, and are given by a sum of two step functions, plus independent additive noise. The first step function $S_0$ is the same for each observation $i$, being 0 for the first half of the timeseries and 1 for the second half. The second step function $S_1$ is a short pulse which occurs at a different time for different experiments $i$, but always the same time for $X$ and $Y$. For each $i$, we simulate $X_i = S_0 + S_1 + e_i^{(X)}$ and $Y_i = S_0 + S_1 + e_i^{(Y)}$, where $e_i^{(X)}$ and $e_i^{(Y)}$ are both iid sequences of normal random variables with mean of zero and standard deviation of 0.05. We consider two cases: that $Z_i = S_0$ for all $i$, or that $Z_i = S_1$.

If $Z_i = S_0$, there is a partial correlation of $X_i$ and $Y_i$ given $Z_i$, because the time of the pulse varies between experiments, but $Z_i$ is always the same. Thus, $\rho(X_i; P_{ij}Y_i)$ is reliably larger than $\rho(X_i; P_{ij}Y_i)$, and the test finds statistical significance (Figure 1).

However, if $Z_i = S_1$, there is not a partial correlation of $X_i$ and $Y_i$ given $Z_i$, since the correlation between $X_i$ and $Y_i$ can be explained by a common dependence on $Z_i$. Thus, $\rho(X_i; P_{ij}Y_i)$ and $\rho(X_i; P_{ij}Y_i)$ are comparable, and the test finds no statistical significance (Figure 2).

To demonstrate that it is essential to use $P_{ij}Y_i$ rather than simply $P_{ij}Y_i$, we repeat the simulation with $Z_i = S_1$, but comparing $\rho(X_i; P_{ij}Y_i)$ with $\rho(X_i; P_{ij}Y_i)$. This invalid test erroneously finds a significant partial correlation where none exists (Figure 3).

**Test 2: An exact test using projection and permutation**

A random permutation approach may be used for a test that provably controls the type 1 error rate. Let $h$ be a function that permutes the order of a sequence. For example, if $a = (7,8,9)$, we might have $h(a) = (8,9,7)$. We will use the notation $h(a)_{ij}$ to refer to the $j$th element of $h(a)$, so that $h(a)_{ij} = 8$ in the example.

Let $Y$ be the sequence of all $Y_i$. Similarly, let $E = (E_{ij}, \ldots, E_{in})$ be the sequence of all $E_{ij}$, and recall that $X$ is the sequence of all pairs $(X_i, Z_i)$. Let $P = I - M([Z_{ij}Z_{ij} \ldots Z_{in}])$ where $[Z_{ij}Z_{ij} \ldots Z_{in}]$ is a block matrix, meaning that $P$ is the $T \times T$ projection matrix orthogonal to the columns of all the $Z_i$. We define the statistic

$$T(X,Y) = \frac{1}{n} \sum_{i=1}^{n} \rho(PX_i, PY_i)$$

Under the null hypothesis, $Y_i = Z_iW_i + E_i$, so $PY_i = PE_i$ and
Using partial correlations in repeatedly observed timeseries

Let

\[ T(X, h(Y)) = \frac{1}{n} \sum_{i=1}^{n} \rho(PX_i, PE_{h(i)}) \]

To test the null we may compare the above statistic with an ensemble consisting of values of the form

\[ T(X, h(Y)) = \frac{1}{n} \sum_{i=1}^{n} \rho(PX_i, PE_{h(i)}) \]

in which each \( X_i \) is compared to a \( Y \) from a session chosen randomly without replacement. Because \( P \) is orthogonal to all the \( Z_i \) under the null

\[ T(X, h(Y)) = T(X, h(E)) = \frac{1}{n} \sum_{i=1}^{n} \rho(PX_i, PE_{h(i)}) \]

Also under the null, all \( E_i \) are independent of each other and of \( X \). Thus, the distribution of the pair \( (X, h(E)) \) is the same for all permutations \( h \), including the identity permutation.

We can therefore obtain a valid permutation test by comparing \( T(X, Y) \) to a null ensemble \( T(X, h(Y)) \) with randomly chosen permutations \( h \). The proof of this follows the standard proof for the validity of permutation tests\(^{14,15} \). Let \( O(E) \) be the orbit of \( E \) under the set of all permutations of the \( n \) experiments. Then the probability \( \mathbb{P}[E | X, O(E)] \) is equal for each member of \( O(E) \). Thus if we let \( h_1, \ldots, h_m \) be independent random permutations, then the \( m+1 \) random variables \( T(X, Y) = T(X, h(Y)) \) and \( T(X, h_l(Y)) = T(X, h_l(E)) \), \( l = 1 \ldots m \), are independent and identically distributed on \( (X, O(E)) \). If we let \( R \) be the rank of \( T(X, Y) \) within these numbers, breaking any ties randomly, then \( R \) is a random integer uniformly distributed between \( 1 \) and \( m+1 \), and

\[ p = \frac{R}{m+1} \]

is an exact p-value, obeying \( \mathbb{P}(p \leq \alpha) = \alpha \) for any \( \alpha \) in the set of attainable p-values, i.e. \( 1/(m+1), 2/(m+1), \ldots, 1 \).

**An illustration of Test 2**

In Fig. 4 we illustrate Test 2 (the permutation test just now described) by applying it to the same example data as shown in Figs. 1 and 2. The results are similar to those obtained with Test 1, with Test 2 detecting the correlation that cannot be explained by \( Z \), and appropriately failing to detect the correlation that is explained by \( Z \).
Discussion
We have described two families of statistical tests for partial correlation between vectorial timeseries. The timeseries are not assumed stationary or linear, and in fact the tests apply to any data consisting of vectors depending on an index $t$. The tests require multiple observations of these timeseries. Intuitively, the tests measure how much better one can predict $Y_i$ from $X_i$ measured in the same experiment, compared to predicting $Y_i$ from $X_j$ measured in a different experiment. To remove a possible common effect of the confounding variable $Z$, the first test projects out both $Z_i$ and $Z_j$ from the T-dimensional vector $Y_i$. The second test projects out the all the matrices $Z_1 \ldots Z_n$, this potentially reduces the power of the test but allows a proof that the p-value is valid even for small numbers of experiments.

The tests depend on a user-supplied measure $\rho(X,Y)$ of the degree of association between two vectorial timeseries. They make no assumptions about this measure, so anything can be used. When dealing with 1-dimensional timeseries (as in the examples presented here), Pearson correlation is a natural choice. In higher dimensions one could for example use the fraction of variance explained by multiple linear regression or ridge regression. If $Y$ is high dimensional, it may be advisable to use regularization or methods such as reduced rank regression or canonical correlation analysis. Using cross-validation to assess performance might improve statistical power, but this is not required for a valid test.

The tests may find use in multiple scientific fields where time-dependent experiments are performed repeatedly. An example in neuroscience would be if one makes multivariate recordings from the brains of subjects performing a behavioral task, and would like to test if some behavioral variables correlate with brain activity even after the effect of other behavioral variables are controlled for. Unlike the “pseudosession method” in which the current tests require improve statistical power, but this is not required for the method of surrogate data. However to apply the test, all timeseries must have the same number of time points, which can be achieved by truncating them to have the same length if necessary. Furthermore, unlike the “pseudosession method” this method can be used when the predictor variables $X$ and $Z$ are not randomly generated, but produced by the subject themselves according to an unknown distribution.

REFERENCES
1. Box, G.E.P. (2008). Time series analysis forecasting and control / George E.P. Box, Gwilym M. Jenkins, Gregory C. Reinsel. 4th ed. (John Wiley).
2. Granger, C.W.J., and Newbold, P. (1974). Spurious regressions in econometrics. J. Econom. 2, 111–120.
3. Haugh, L.D. (1976). Checking the independence of two covariance-stationary time series: a univariate residual cross-correlation approach. J. Am. Stat. Assoc. 71, 378–385.
4. Phillips, P.C.B. (1986). Understanding spurious regressions in econometrics. J. Econom. 33, 311–340.
5. Yule, G.U. (1926). Why do we Sometimes get Nonsense-Correlations between Time-Series?--A Study in Sampling and the Nature of Time-Series. J. R. Stat. Soc. 89, 1–63. 10.2307/2341482.
6. Cliff, O.M., Novelli, L., Fulcher, B.D., Shine, J.M., and Lizier, J.T. (2021). Assessing the Significance of Directed and Multivariate Measures of Linear Dependence Between Time Series. Phys. Rev. Res. 3, 013145. 10.1103/PhysRevResearch.3.013145.
7. Clifford, P., Richardson, S., and Hemon, D. (1989). Assessing the Significance of the Correlation between Two Spatial Processes. Biometrics 45, 123–134. 10.2307/2532039.
8. Bartlett, M.S. (1946). On the Theoretical Specification and Sampling Properties of Autocorrelated Time-Series. Suppl. J. R. Stat. Soc. 8, 27–41. 10.2307/2983611.
9. Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., and Farmer, J.D. (1992). Testing for nonlinearity in time series: the method of surrogate data. Phys. D 58, 77.
10. Harris, K.D. (2020). Nonsense correlations in neuroscience. bioRxiv, 2020.11.29.402719. 10.1101/2020.11.29.402719.
11. Golub, G.H., and Loan, C.F.V. (2013). Matrix Computations (Johns Hopkins University Press) 10.56021/9781421407944.
12. Cressie, N. (1980). Relaxing assumptions in the one sample t-test. Aust. J. Stat. 22, 143–153. 10.1111/j.1467-842X.1980.tb01161.x.
13. Efron, B. (1969). Student’s t-Test Under Symmetry Conditions. J. Am. Stat. Assoc. 64, 1278–1302. 10.2307/2286068.
14. Lehmann, E.L., and Romano, J.P. (2022). Testing Statistical Hypotheses (Springer International Publishing) 10.1007/978-3-030-70578-7.
15. Hemertik, J., and Goeman, J. (2018). Exact testing with random permutations. TEST 27, 811–825. 10.1007/s11749-017-0571-1.