

RESEARCH ARTICLE

Assessment of resampling methods for causality testing: A note on the US inflation behavior

Angeliki Papana^{1*}, Catherine Kyrtsov^{1,2}, Dimitris Kugiumtzis³, Cees Diks⁴

1 Department of Economics, University of Macedonia, Thessaloniki, Greece, **2** CAC IXXI-ENS Lyon, Lyon, France; University of Paris 10, Paris, France; University of Strasbourg, BETA, Strasbourg, France, **3** Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece, **4** Center for Nonlinear Dynamics in Economics and Finance (CeNDEF), Amsterdam School of Economics, University of Amsterdam, Amsterdam, The Netherlands

* angeliki.papana@gmail.com



OPEN ACCESS

Citation: Papana A, Kyrtsov C, Kugiumtzis D, Diks C (2017) Assessment of resampling methods for causality testing: A note on the US inflation behavior. PLoS ONE 12(7): e0180852. <https://doi.org/10.1371/journal.pone.0180852>

Editor: Zhong-Ke Gao, Tianjin University, CHINA

Received: December 5, 2016

Accepted: June 6, 2017

Published: July 14, 2017

Copyright: © 2017 Papana et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The matlab codes for generating the corresponding simulation time series of the manuscript are provided as a Supplementary File. The financial time series from the real applications can be downloaded from the Federal Reserve Bank of Saint Louis at the following link: <https://fred.stlouisfed.org/categories>.

Funding: The research project is implemented within the framework of the Action “Supporting Postdoctoral Researchers” of the Operational Program “Education and Lifelong Learning” (Action’s Beneficiary: General Secretariat for Research and Technology), and is co-financed by

Abstract

Different resampling methods for the null hypothesis of no Granger causality are assessed in the setting of multivariate time series, taking into account that the driving-response coupling is conditioned on the other observed variables. As appropriate test statistic for this setting, the partial transfer entropy (PTE), an information and model-free measure, is used. Two resampling techniques, time-shifted surrogates and the stationary bootstrap, are combined with three independence settings (giving a total of six resampling methods), all approximating the null hypothesis of no Granger causality. In these three settings, the level of dependence is changed, while the conditioning variables remain intact. The empirical null distribution of the PTE, as the surrogate and bootstrapped time series become more independent, is examined along with the size and power of the respective tests. Additionally, we consider a seventh resampling method by contemporaneously resampling the driving and the response time series using the stationary bootstrap. Although this case does not comply with the no causality hypothesis, one can obtain an accurate sampling distribution for the mean of the test statistic since its value is zero under H_0 . Results indicate that as the resampling setting gets more independent, the test becomes more conservative. Finally, we conclude with a real application. More specifically, we investigate the causal links among the growth rates for the US CPI, money supply and crude oil. Based on the PTE and the seven resampling methods, we consistently find that changes in crude oil cause inflation conditioning on money supply in the post-1986 period. However this relationship cannot be explained on the basis of traditional cost-push mechanisms.

Introduction

Connectivity analysis of multivariate time series is a rapidly growing branch of interest with applications in different fields, such as economy, climatology and brain dynamics. A variety of methods have been developed that uncover complex dynamical structures, i.e. analysis of

the European Social Fund (ESF) and the Greek State. The publication fees of this manuscript will be covered by the University of Amsterdam.

Competing interests: The authors have declared that no competing interests exist.

complex networks from multivariate time series [1–3] and characterization of the complexity of multivariate time series using entropy measures [4, 5].

The investigation of the causal relationships between the variables of a multivariate dynamical system or stochastic process allows us to better understand its structure. In the estimation of direct causality, effects from the remaining variables should be taken into account. We note that by causality we mean Granger causality, either linear and/or nonlinear.

Various causality measures have been recently developed based on information theory. Their advantage is that they are model-free and detect both linear and nonlinear causal effects. The transfer entropy (TE) is the most popular information causality measure, a non-parametric measure that quantifies the amount of information transferred between two random processes [6]. The TE has been proven to be equivalent to the standard linear Granger causality for Gaussian variables [7]. The TE is extended to the partial transfer entropy (PTE) that estimates direct causal effects in multivariate time series [8, 9], and they both have been modified to work on symbol or rank vectors derived from the time series measurements [10–14]. The TE and its variants have been used to investigate the coupling in complex systems [15], also in combination with other methods [16]. In order to avoid the curse of dimensionality, progressive selection of lagged variables [17, 18] and graphical models [19] have been combined with information causality measures. Information causality measures are mainly applied to neuroscience and physiology (e.g. see [20, 21]), as well as finance (e.g. see [22]).

Theoretically, a causality measure should be zero if there are no causal interactions and positive otherwise. However, its value may deviate from the true value (bias) due to the estimation method, the selection of parameters, the finite sample size, the level of noise, as well as the system complexity. In particular for the TE and PTE, the bias can be large stemming from the estimation of conditional densities [23]. In the presence of varying bias, a significance test is more appropriate, than arbitrary thresholding, for deciding the presence of weak coupling. This issue is particular relevant when constructing causality networks with binary connections from direct Granger causality on real multivariate time series [24–28]. Different randomization and bootstrap methods have been employed to correct the bias of the causality measures (see e.g. [29]) and more specifically of transfer entropy (see e.g. [14, 30]).

When using linear Granger causality measures, their asymptotic distribution under the null hypothesis H_0 of no causal effect is known [31–33]. For information causality measures, parametric tests are only developed when the time series are discrete-valued [13, 34]. When the asymptotic distribution of a test statistic cannot be established, resampling techniques are employed for the construction of its empirical null distribution. The resampled time series should satisfy the H_0 and also capture the statistical properties of the original time series. Resampled time series can be generated by bootstrapping and randomization.

Bootstrapping is a statistical technique introduced in [35] that aims to estimate the properties of a test statistic when sampling from an approximate distribution. The empirical distribution of the statistic is formed by its values computed on samples drawn with replacement from the original sample. For time series, bootstraps must be carried out in a way that they suitably capture the dependence structure of the data generation process consistent to the H_0 , and be otherwise random (see e.g. [36–38]). For our hypothesis testing, the bootstraps are used to form the null distribution of the causality measure. In a bivariate setting, this is done by bootstrapping the two variables independently or contemporaneously or by only bootstrapping one of the two variables [39].

Statistical tests based on randomization utilize randomized data, which are shuffled samples of the original data, to empirically estimate the expected probability distribution of the estimator. The randomization methods are designed to preserve the dependence structure consistent with H_0 , when randomly shuffling the time series. The time-shifted surrogates [40], as well as

other types of surrogates such as the twin surrogates [41], have been extensively suggested in applications, e.g. see [42–45].

In this work, we make an explorative study on resampling time series for the H_0 of no causal effect and compare seven resampling techniques with regard to the size and power of the significance test, using the PTE as test statistic. Specifically, we combine two resampling techniques: 1) the time-shifted surrogates [40] and 2) the stationary bootstrap [38], with three independence settings of the time series adapted for the non-causality test (giving six resampling methods): A) resampling only the time series of the driving variable, B) resampling independently the driving and the response time series, and C) resampling separately the driving and the response time series, while destroying the dependence of the future and past of the response variable. To the best of our knowledge, schemes B) and C) in conjunction with randomization or bootstrap have not been considered in any methodological study or application. We also introduce a new (seventh) method by bootstrapping contemporaneously the driving and the response time series. In this case, the bootstrap PTE values are centered to zero since the H_0 of no causal effects is not satisfied.

The empirical distribution of PTE, as well as the size and power of the significance test, for the seven resampling methods are assessed in a simulation study. Some first results on the aforementioned resampling methods have been already presented in [46] and [47]. Here, we extend the study of the examined resampling methods in order to establish their performance.

Finally, to demonstrate the performance of PTE in conjunction with the seven resampling methods using real data, we investigate the possible sources of the US inflation in the post-Volcker era utilizing two 3-variate systems built on the Consumer Price Index for All Urban Consumers, the core CPI, the money supply and the price of crude oil. Empirical results support evidence in favor of a statistically significant direct causal relationship between oil prices and US inflation obeying dynamics which are not comparable with the oil episodes occurred in the 1970s.

Materials and methods

Partial transfer entropy

The TE quantifies the amount of information explained in a response variable Y at one time step ahead from the state of a driving variable X accounting for the concurrent state of Y . Let $\{x_t, y_t\}, t = 1, \dots, n$ be the observed time series of two variables. We define the reconstructed state space vectors of the variables as $\mathbf{x}_t = (x_t, x_{t-\tau}, \dots, x_{t-(m-1)\tau})'$ and $\mathbf{y}_t = (y_t, y_{t-\tau}, \dots, y_{t-(m-1)\tau})'$, where m is the embedding dimension and τ the time delay. The TE from X to Y constitutes the conditional mutual information $I(y_{t+1}; \mathbf{x}_t | \mathbf{y}_t)$ given as [6]

$$\begin{aligned} TE_{X \rightarrow Y} &= I(y_{t+1}; \mathbf{x}_t | \mathbf{y}_t) = \sum p(y_{t+1}, \mathbf{x}_t, \mathbf{y}_t) \log \frac{p(y_{t+1} | \mathbf{x}_t, \mathbf{y}_t)}{p(y_{t+1} | \mathbf{y}_t)} \\ &= H(\mathbf{x}_t, \mathbf{y}_t) - H(y_{t+1}, \mathbf{x}_t, \mathbf{y}_t) + H(y_{t+1}, \mathbf{y}_t) - H(\mathbf{y}_t), \end{aligned} \tag{1}$$

where TE is expressed either based on the probability distributions, $p(\cdot)$ (here being defined for the discretized variables), or the entropy terms, $H(\cdot)$, where $H(\mathbf{x}) = -\int f(\mathbf{x}) \log f(\mathbf{x}) d\mathbf{x}$ is the differential entropy of the vector variable \mathbf{x} with probability density function $f(\mathbf{x})$. We note that m and τ are set to be similar for both variables as suggested in [29].

The partial transfer entropy (PTE) is the multivariate extension of transfer entropy (TE) in [8, 9]. The PTE accounts for the direct coupling of X to Y conditioning on the remaining

variables of a multivariate system, collectively denoted by Z . It is defined as

$$\begin{aligned} \text{PTE}_{X \rightarrow Y|Z} &= I(y_{t+1}; \mathbf{x}_t | \mathbf{y}_t, \mathbf{z}_t) \\ &= H(\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t) - H(y_{t+1}, \mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t) + H(y_{t+1}, \mathbf{y}_t, \mathbf{z}_t) - H(\mathbf{y}_t, \mathbf{z}_t). \end{aligned} \tag{2}$$

The estimation of PTE relies on the estimation of the joint probability density functions in the expression of the entropies. Different types of estimators for the TE and PTE exist, such as histogram-based (e.g. by discretizing the variables to equidistant intervals [48]), kernel-based [49] and using correlation sums [50]. In this paper, we choose the nearest neighbor estimator [51], which is specifically effective for high-dimensional data [18]. This estimator uses the distances between the reconstructed state space vectors to estimate the joint and marginal densities. For each reference point, viewed in the largest state space, the distance length ϵ is defined as the distance to the k -th nearest neighbor. Then densities, at projected subspaces, are locally formed by the number of points within ϵ from each reference point. Thus, the free parameter in the estimation of entropies is the number of neighbors k .

Theoretically, the causality measures including the PTE should be zero in the case of no causal effects. However, various issues such as the estimation method for the entropies and subsequently densities, the selection of the embedding parameters, the finite sample size and the inherent dynamics of each subsystem [29] may introduce bias. In order to determine whether a PTE value indicates a weak coupling or it is not statistically significant, resampling methods are employed.

Resampling methods

Our examined null hypothesis H_0 is that there is no direct causal effect from X to Y or more specifically that $\text{PTE}_{X \rightarrow Y|Z} = 0$, i.e. $I(y_{t+1}; \mathbf{x}_t | \mathbf{y}_t, \mathbf{z}_t) = 0$. In order to generate resampled time series representing the H_0 , we consider two resampling techniques, i.e. 1) the time-shifted surrogates and 2) the stationary bootstrap, and combine them with three independence settings. Thus six resampling methods (cases 1A to 2C) are formulated to test the H_0 . In addition, we introduce a seventh resampling method that is based on the stationary bootstrap and does not directly comply with the H_0 .

Resampling techniques. 1) **Time-shifted surrogates.** Let us consider two variables X and Y and their corresponding time series $\{x_1, \dots, x_n\}$ and $\{y_1, \dots, y_n\}$. The time-shifted surrogates are generated, so that they preserve the dynamics of the original time series, i.e. $\{x_1, \dots, x_n\}$, while the couplings between X and Y are destroyed [40]. They are formed by cyclically time-shifting the components of a time series. In more details, for the time series $\{x_1, \dots, x_n\}$, an integer d is randomly chosen and the d first values of the time series are moved to the end, giving the time-shifted surrogate time series $\{x_i^*\} = \{x_{d+1}, \dots, x_n, x_1, \dots, x_d\}$. The random number, d , is randomly drawn from the discrete uniform distribution in the range $[0.05n; 0.95n]$ in order to maintain disruption of the time order of the original time series even in the presence of strong autocorrelation.

2) **The stationary bootstrap.** The stationary bootstrap was introduced in [38] to adapt bootstrap on correlated data. By construction, the stationary bootstrap does not destroy the time dependence of the data. This method tries to replicate the correlations by resampling blocks of data. The lengths of the resampled blocks have a geometric distribution. For a fixed probability p , block lengths L_i are generated with probability $p(L_i = k) = (1 - p)^{(k-1)} p$ for $k = 1, 2, \dots$. The starting time points of the blocks L_i are drawn from the discrete uniform distribution on $\{1, \dots, n - k\}$. A bootstrap time series $\{x_i^*\}$ is formed by first starting with a random block as defined above $B_{L_i, L_i} = \{x_{L_i}, x_{L_i+1}, \dots, x_{L_i+L_i-1}\}$, and blocks are added until length n is reached.

Independence settings. The three independence settings presented below regard both time-shifted surrogate and stationary bootstrapped time series.

A. The first setting is to resample only the time series of the driving variable X . This constitutes the standard approach for the surrogate test for the significance of causality measures [18, 40, 52, 53]. The intrinsic dynamics of the variable X is preserved in the resampled time series $\{x_t^*\}$ but the coupling between X^* and Y is destroyed. So, H_0 is approximated and $\text{PTE}_{X^* \rightarrow Y|Z} = 0$. The variables X and Y as well as X and Z are independent, however the pair of variables (Y, Z) preserves its interdependence.

B. This second scheme resamples both the driving variable X and the response variable Y , i.e. the resampled time series $\{x_t^*\}$ and $\{y_t^*\}$ are generated. Again, the intrinsic dynamics of both X and Y are preserved but the coupling between them is destroyed, so that $\text{PTE}_{X^* \rightarrow Y^*|Z} = 0$. Here, independence holds for all variable pairs (X, Y) , (Y, Z) and (X, Z) . Nevertheless, there is still no complete independence between all arguments in the definition of PTE, as y_{t+1} preserves by construction of $\{y_t^*\}$ its dependence on \mathbf{y}_t .

C. The third scheme establishes complete independence of all the terms involved in the definition of PTE, i.e. in addition to the resampling of X and Y , also y_{t+1} is resampled separately. Technically, we first form the reconstructed vectors of X and Y and then we randomly shuffle them independently for each time series. In this way, the time dependence is destroyed between y_{t+1} , \mathbf{x}_t and \mathbf{y}_t and therefore they become independent. Further, \mathbf{z}_t becomes independent of $\mathbf{x}_t, \mathbf{y}_t$ but not of y_{t+1} .

The seventh resampling method uses stationary bootstrap to resample contemporaneously the driving and the response time series (X, Y) . The resampled time series are not consistent to H_0 because the coupling of X and Y is not destroyed. In order to obtain an accurate sampling distribution of the mean of the test statistic one can take into consideration that the mean value of the test statistic is zero under H_0 . The idea is that $\sqrt{n}(\text{PTE} - E(\text{PTE}))$, where $E(\text{PTE})$ is the mean of PTE, can be distributed similarly for series that comply to H_0 ($E(\text{PTE}) = 0$) and series that do not ($E(\text{PTE}) > 0$); it is assumed that $\sqrt{n}(\text{PTE} - E(\text{PTE}))$ tends to the normal distribution with zero mean and known variance [38]. Since our goal is to compare the different resampling methods, no results for this approximation of the true distribution are discussed. By centering the distribution of the bootstrap PTE values around zero, we get an approximation of the null distribution of PTE. Thus, this resampling method can be employed to test H_0 , provided that the null distribution of the bootstrap values of the test statistic is shifted to have mean zero. It is labelled as 2D to stress that it the fourth setting for the stationary bootstrap.

Simulation study

We apply the significance test for the PTE with the seven resampling methods to multiple realizations of various simulation systems. Specifically, we estimate the PTE from 1000 realizations per simulation system. For each realization and each resampling method, $M = 100$ resampled time series are generated. Let us denote q_0 the PTE value from one realization of a system and q_1, q_2, \dots, q_M the PTE values from the resampled time series for this particular realization and for a specific resampling method. The rejection of H_0 of no causal effects is decided by the rank ordering of the PTE values computed on the original time series, q_0 , and the resampled time series, q_1, \dots, q_M . For the one-sided test, if r_0 is the rank of q_0 when ranking the list q_0, q_1, \dots, q_M in ascending order, the p -value of the test is $1 - [(r_0 - 0.326)/(M + 1 + 0.348)]$, by applying the correction in [54].

The simulation systems we considered in this study are:

1. Three coupled Hénon maps, with nonlinear couplings ($X_1 \rightarrow X_2 \rightarrow X_3$)

$$\begin{aligned} x_{1,t} &= 1.4 - x_{1,t-1}^2 + 0.3x_{1,t-2} \\ x_{2,t} &= 1.4 - cx_{1,t-1}x_{2,t-1} - (1-c)x_{2,t-1}^2 + 0.3x_{2,t-2} \\ x_{3,t} &= 1.4 - cx_{2,t-1}x_{3,t-1} - (1-c)x_{3,t-1}^2 + 0.3x_{3,t-2}, \end{aligned}$$

with equal coupling strengths c for $X_1 \rightarrow X_2$ and $X_2 \rightarrow X_3$. We set $c = 0$ (uncoupled case), $c = 0.3$ (moderate coupling) and $c = 0.5$ (strong coupling). We note that the time series of this system become completely synchronized for coupling strengths $c \geq 0.7$.

2. A vector autoregressive process of 4 variables and order 5, VAR(5), with linear couplings ($X_4 \rightarrow X_2 \rightarrow X_1 \rightarrow X_3$ and $X_2 \rightarrow X_3$)

$$\begin{aligned} x_{1,t} &= 0.8x_{1,t-1} + 0.65x_{2,t-4} + \epsilon_{1,t} \\ x_{2,t} &= 0.6x_{2,t-1} + 0.6x_{4,t-5} + \epsilon_{2,t} \\ x_{3,t} &= 0.5x_{3,t-3} - 0.6x_{1,t-1} + 0.4x_{2,t-4} + \epsilon_{3,t} \\ x_{4,t} &= 1.2x_{4,t-1} - 0.7x_{4,t-2} + \epsilon_{4,t}, \end{aligned}$$

where $\epsilon_i, i = 1, \dots, 4$, are independent to each other Gaussian white noise processes with unit standard deviation (Eq (12) in [55]).

3. Five coupled Hénon maps, with nonlinear couplings ($X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_4 \rightarrow X_5$) defined similarly to system 1. We consider again equal coupling strengths c , and set $c = 0$ (uncoupled case), $c = 0.2$ (moderate coupling) and $c = 0.4$ (strong coupling).

We consider two time series lengths: $n = 512$ and 2048 . The calculation of the PTE relies on the phase space reconstruction [56, 57]; specifically for PTE see [8]. Since all the simulation systems are discrete in time we set the time delay τ equal to one, while the embedding dimension m is identical for all variables, which is reported to be the best strategy [29], and for each system it is set according to its complexity, i.e. taking into account the maximum delay in the equations of each system. The number of nearest neighbors for the estimation of the probability distributions equals 10 (the choice of k does not substantially affect the estimation of PTE [53, 58]).

To investigate the performance of the significance tests for the PTE with the different resampling methods, we use the sensitivity of the PTE, i.e the percentage of rejection of H_0 when there is true direct causality, as well as the specificity of the PTE, i.e. the percentage of no rejection of H_0 when there is no direct causality, at the significance level $\alpha = 0.05$. The notation $X_2 \rightarrow X_1|Z$ denotes the Granger causality from X_2 to X_1 , accounting for the presence of confounding variables $Z = X_3, \dots, X_K$, where K is the number of observed variables. For brevity, we use the notation $X_2 \rightarrow X_1$ instead of $X_2 \rightarrow X_1|Z$, implying the conditioning on the confounding variables. The same holds for the remaining pairs of variables.

System 1. The PTE is negatively biased; the mean PTE values from the 1000 realizations at all directions are negative when $c = 0$ (Table 1). For $c = 0.3$ and $c = 0.5$, it is larger when direct couplings exist ($X_1 \rightarrow X_2, X_2 \rightarrow X_3$) and raises with n . Regarding the indirect coupling $X_1 \rightarrow X_3$, the PTE slightly increases with n as c increases, reaching the highest mean value for $c = 0.5$ (mean $\text{PTE}_{X_1 \rightarrow X_3} = 0.0004$ for $n = 512$ and $\text{PTE}_{X_1 \rightarrow X_3} = 0.0071$ for $n = 2048$). For the rest of the couplings, the PTE is negative at the same level regardless of c or n . The occurrence of many negative values of the PTE indicates the need for a significance test.

We evaluate how the null distribution of the PTE from the seven resampling methods differs with respect to the original PTE values. For $c = 0$, all of them correctly indicate the absence

Table 1. Mean PTE values from 1000 realizations of system 1 for $n = 512$ and 2048 , highlighted at the directions of the true couplings.

$n = 512$	$X_1 \rightarrow X_2$	$X_2 \rightarrow X_1$	$X_2 \rightarrow X_3$	$X_3 \rightarrow X_2$	$X_1 \rightarrow X_3$	$X_3 \rightarrow X_1$
$c = 0$	-0.0059	-0.0062	-0.0062	-0.0061	-0.0058	-0.0056
$c = 0.3$	0.0802	-0.0042	0.0885	-0.0064	-0.0045	-0.0074
$c = 0.5$	0.2324	-0.0071	0.1557	-0.0044	0.0004	-0.0079
$n = 2048$	$X_1 \rightarrow X_2$	$X_2 \rightarrow X_1$	$X_2 \rightarrow X_3$	$X_3 \rightarrow X_2$	$X_1 \rightarrow X_3$	$X_3 \rightarrow X_1$
$c = 0$	-0.0086	-0.0088	-0.0087	-0.0085	-0.0087	-0.0088
$c = 0.3$	0.1736	-0.0024	0.1725	-0.0059	-0.0039	-0.0094
$c = 0.5$	0.3649	-0.0026	0.2601	-0.0049	0.0071	-0.0078

<https://doi.org/10.1371/journal.pone.0180852.t001>

of couplings as the percentage of rejection at $\alpha = 0.05$ is not larger than 5% (Table 2). Considering $c = 0.3$, the true couplings are identified again. However, spurious and indirect couplings are indicated as well for the setting A and less for B. Additionally, similar performance is observed when the coupling strength is strong ($c = 0.5$) and large percentages are obtained for the indirect coupling $X_1 \rightarrow X_3$ in all schemes.

The sensitivity of PTE is assessed from the two true causal links, i.e. $X_1 \rightarrow X_2$ and $X_2 \rightarrow X_3$ since we calculate the proportion of ‘positives’ (true causal links) that are correctly identified. A high sensitivity is established by a high percentage of significant PTE values over the 1000 realizations for these two couplings, which means that the PTE correctly detects the true causal effects. Similarly, the specificity of PTE is decided by the percentage of the significant PTE for

Table 2. Percentage of significant PTE values for system 1 for $n = 512 / 2048$, for all resampling methods. A single number is displayed when the same percentage corresponds to both n . The true couplings are highlighted.

$c = 0$	$X_1 \rightarrow X_2$	$X_2 \rightarrow X_1$	$X_2 \rightarrow X_3$	$X_3 \rightarrow X_2$	$X_1 \rightarrow X_3$	$X_3 \rightarrow X_1$
1A	5.7 / 4.2	5.6 / 5.2	4.7 / 4.9	5.3 / 5.6	5.8 / 5.5	5.5 / 5.2
1B	5.2 / 4.8	4.6 / 5.6	4 / 5.2	4.3 / 6.6	4.6 / 5	5.8 / 5.5
1C	0.7 / 0	0.8 / 0	0.4 / 0	0.7 / 0	0.3 / 0	0.5 / 0
2A	4.4 / 3.8	3.1 / 3.9	3.4 / 4.1	4.5 / 4.5	4.5 / 4.3	4.1 / 5.1
2B	1.9 / 0.4	1.9 / 0.7	1.8 / 0.6	2.1 / 0.3	1.9 / 0.5	2.4 / 1
2C	0.6 / 0	0.6 / 0	0.3 / 0	0.5 / 0	0.4 / 0	0.1 / 0
2D	0.6 / 0	0.7 / 0	0.3 / 0	0.7 / 0	0.2 / 0	0.2 / 0
$c = 0.3$	$X_1 \rightarrow X_2$	$X_2 \rightarrow X_1$	$X_2 \rightarrow X_3$	$X_3 \rightarrow X_2$	$X_1 \rightarrow X_3$	$X_3 \rightarrow X_1$
1A	100	11.8 / 40.1	100	9.5 / 17.2	12.8 / 34	6.1 / 5.5
1B	100	9 / 37.2	100	2.7 / 1.8	5.4 / 6.7	5 / 4.3
1C	100	0.9 / 0.5	86.3 / 100	0	0.2 / 0	0.4 / 0.1
2A	100	8.7 / 32.8	100	6.9 / 13.5	8.9 / 28	4.7 / 4.1
2B	100	2.9 / 13.7	100	0.9 / 0.3	1.2 / 1.7	1.2 / 0.5
2C	100	0.8 / 0.6	99.9 / 100	0	0 / 0.1	0.3 / 0.1
2D	100	1.2 / 0.9	100	0.1 / 0	0.2 / 0.3	0.4 / 0.1
$c = 0.5$	$X_1 \rightarrow X_2$	$X_2 \rightarrow X_1$	$X_2 \rightarrow X_3$	$X_3 \rightarrow X_2$	$X_1 \rightarrow X_3$	$X_3 \rightarrow X_1$
1A	100	8.1 / 33.8	100	10.2 / 21.5	31 / 96.3	6.2 / 8.3
1B	100	4.3 / 30.4	100	1.7 / 1.4	9.1 / 67.3	4.5 / 4.8
1C	100	0.7 / 0.4	100	0	1.9 / 25.4	0.1
2A	100	5.1 / 29.2	100	7.7 / 17	24.1 / 94.7	4 / 7.1
2B	100	2 / 11	100	0.8 / 0.2	5.2 / 53.3	1.3 / 0.8
2C	100	0 / 0.2	100	0	1.2 / 24.3	0 / 0.1
2D	100	0.2 / 0.6	100	0 / 0.1	1.4 / 11.6	0.1

<https://doi.org/10.1371/journal.pone.0180852.t002>

the remaining couples, for which there is no true causal effect. A low percentage of significant PTE values signifies a large proportion of ‘negatives’ (no causal links) correctly identified.

Concerning the first six resampling methods, the percentage of erroneously rejected H_0 for non-existing or indirect couplings tends to increase with c and the time series length n , the most robust being 1C and 2C. It turns out that when the resampled time series become more independent (from A to C), the percentage of spurious couplings decreases. This is so because the null distribution for the test is somewhat more spread and displaced to the right as the resampling changes from the least independent scheme (setting A) to the most independent one (setting C) (Fig 1).

The resampling method 2D seems to be the most effective one as it attains the same highest percentage of rejection for true direct couplings and the lowest percentage of rejection for no direct coupling. We note that the green dots are not displayed in Fig 1a because they exceed the axis and we kept the same range of PTE values (y-axis) in all subfigures in order to be able to straightforwardly compare the different cases.

We are interested in the spread of the resulting surrogate null distribution. Thus, we display some indicative results for the mean value of the means and standard deviations of the

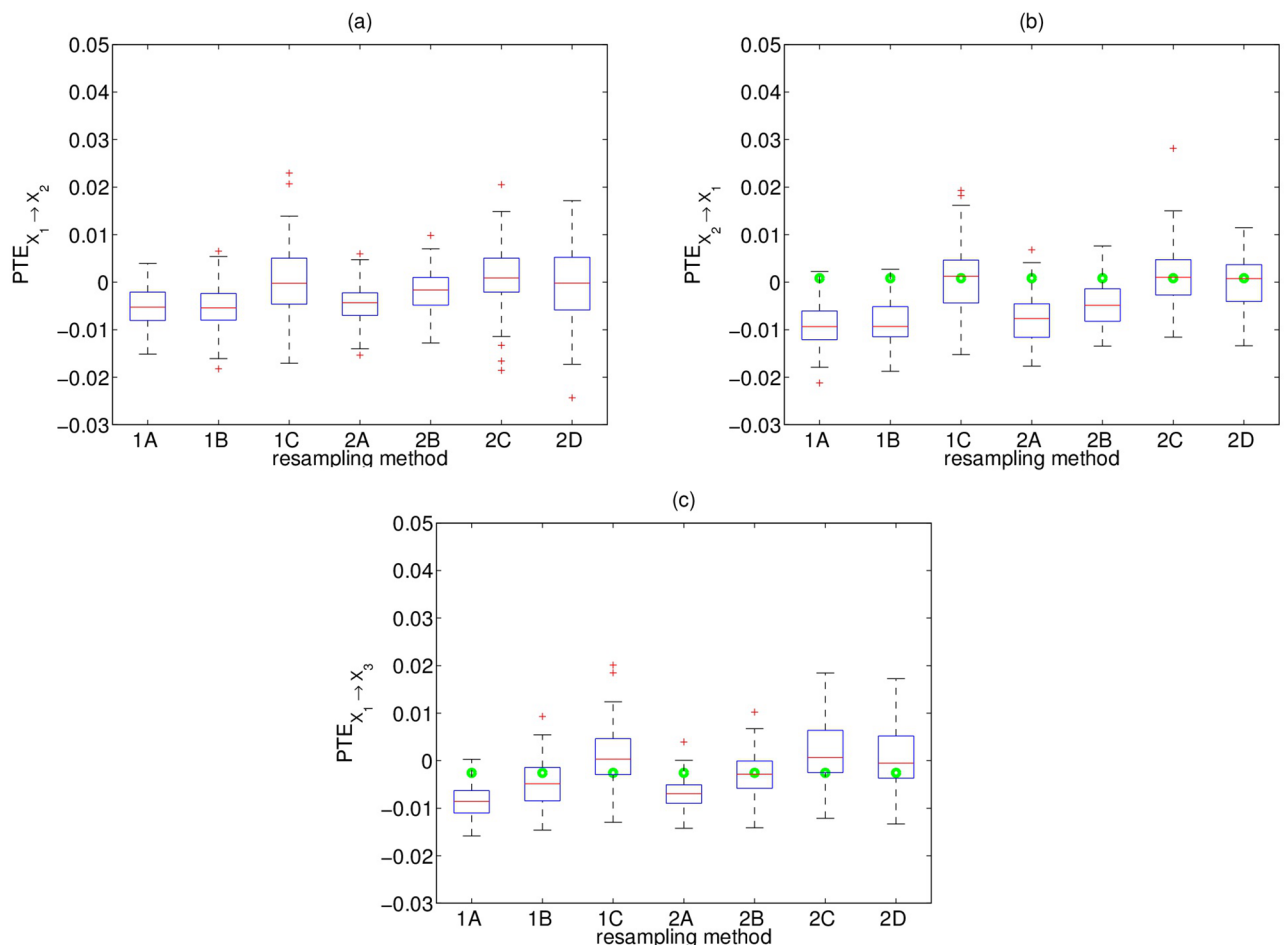


Fig 1. Boxplots of surrogate/bootstrapped PTE values and original PTE value from one realization of system 1 for $c = 0.3$ and $n = 2048$, for the directions (a) $X_1 \rightarrow X_2$ (direct coupling), (b) $X_2 \rightarrow X_1$ (no coupling) and (c) $X_1 \rightarrow X_3$ (indirect coupling). The dots at the same level denote the PTE value on the original data, and in (a) the value is 0.19 and not displayed. The central mark, the bottom and top edges on each box indicate the median, the 25th and 75th percentile, respectively. Outliers are denoted with the ‘+’.

<https://doi.org/10.1371/journal.pone.0180852.g001>

Table 3. Mean value of all means and standard deviations (std) over all realizations of system 1 of the surrogate PTE values for the direction $X_1 \rightarrow X_2$ and time series length $n = 512$ for each resampling case.

$c = 0$	mean	std	$c = 0.3$	mean	std	$c = 0.5$	mean	std
1A	-0.0059	0.0085	1A	-0.0056	0.0075	1A	-0.0037	0.0087
1B	-0.0059	0.0086	1B	-0.0042	0.0100	1B	-0.0009	0.0190
1C	0.0009	0.0107	1C	0.0000	0.0115	1C	0.0008	0.0115
2A	-0.0049	0.0086	2A	-0.0043	0.0076	2A	-0.0025	0.0086
2B	-0.0019	0.0087	2B	-0.0013	0.0079	2B	0.0003	0.0087
2C	0.0022	0.0104	2C	0.0023	0.0113	2C	0.0023	0.0112
2D	0.0000	0.0103	2D	0.0000	0.0135	2D	0.0000	0.0182

<https://doi.org/10.1371/journal.pone.0180852.t003>

surrogate PTE values over all the realizations for the direction $X_1 \rightarrow X_2$ and for time series length $n = 512$ in Table 3. The more independent setting we consider (from A to B to C), the greater the median and the mean (as shown in Fig 1 and Table 3, respectively) and the larger the spread of the distribution of the surrogate PTE values, while case 2D features one of the greatest spreads.

System 2. The mean PTE values from 1000 realizations of the second system are all positive and the PTE for the directions of the true couplings is larger, with the exception of $X_2 \rightarrow X_3$ being at the level of no direct coupling and not significantly increasing with n (Table 4). The level of the PTE for the uncoupled directions varies from 0.0014 to 0.0097 and decreases with n .

The true couplings $X_2 \rightarrow X_1$, $X_1 \rightarrow X_3$, $X_4 \rightarrow X_2$ are well established by the significance test (Table 5). The weak coupling $X_2 \rightarrow X_3$ is detected only by the setting A (1A and 2A), with the power of the test increasing with n . No spurious causalities are identified by the first six resampling methods (percentage of significant PTE varies from 0% to 6% at the uncoupled directions), however method 2D identifies wrongly the couplings $X_2 \rightarrow X_4$ and $X_3 \rightarrow X_4$, giving much higher percentage than the nominal size 5%. The surrogate/bootstrap PTE values seem to increase as the resampled time series become more independent. This can be clearly observed when comparing settings A and B, as shown in Fig 2 for the strong coupling $X_2 \rightarrow X_1$ and Fig 3 for the weak coupling $X_2 \rightarrow X_3$. The bootstrap PTE values for method 2D are centered around zero by construction, while the surrogate/bootstrap PTE values for the other six resampling methods are positively biased. Their distribution becomes wider as the resampling method gets more independent (A to C), with method 2D having the wider one. The latter performs poorly because the distribution of the bootstrap PTE values is much wider compared to the other ones and the original PTE value falls within the tail of this distribution (Fig 3, case 2D).

System 3. The mean PTE values from 1000 realizations of the third system are presented in Table 6. Slightly negative PTE values are obtained at the uncoupled directions, while some

Table 4. As Table 1 but for system 2.

	$X_1 \rightarrow X_2$	$X_2 \rightarrow X_1$	$X_1 \rightarrow X_3$	$X_3 \rightarrow X_1$	$X_1 \rightarrow X_4$	$X_4 \rightarrow X_1$
$n = 512$	0.0044	0.0914	0.0757	0.0032	0.0057	0.0038
$n = 2048$	0.0026	0.1232	0.0960	0.0014	0.0038	0.0021
	$X_2 \rightarrow X_3$	$X_3 \rightarrow X_2$	$X_2 \rightarrow X_4$	$X_4 \rightarrow X_2$	$X_3 \rightarrow X_4$	$X_4 \rightarrow X_3$
$n = 512$	0.0056	0.0052	0.0097	0.1002	0.0069	0.0033
$n = 2048$	0.0058	0.0029	0.0064	0.1348	0.0045	0.0014

<https://doi.org/10.1371/journal.pone.0180852.t004>

Table 5. As Table 2 but for system 2.

	$X_1 \rightarrow X_2$	$X_2 \rightarrow X_1$	$X_1 \rightarrow X_3$	$X_3 \rightarrow X_1$	$X_1 \rightarrow X_4$	$X_4 \rightarrow X_1$
1A	0.4 / 0	100	100	0.6 / 0.3	0.1 / 0	4.6 / 3.2
1B	0	100	99.4 / 100	0	0	0
1C	0	100	100	0	0	0
2A	0.4 / 0	100	100	0.5	0.1 / 0	2.8 / 3.7
2B	0	100	100	0	0.2 / 0	0 / 0
2C	0	100	99.7 / 100	0	0	0
2D	2.3 / 3.8	100	100	0.8 / 0.5	8.2 / 15.7	1.9 / 1.8
	$X_2 \rightarrow X_3$	$X_3 \rightarrow X_2$	$X_2 \rightarrow X_4$	$X_4 \rightarrow X_2$	$X_3 \rightarrow X_4$	$X_4 \rightarrow X_3$
1A	18.8 / 62.4	1.1 / 0.2	3.5 / 2	100	0.8 / 0	6 / 4.2
1B	0 / 0.1	0	2.1 / 1.3	99.9 / 100	0.4 / 0	0
1C	3.7 / 10.1	0	0	100	0	0.9 / 0
2A	11.7 / 60.1	0.6 / 0.1	2.6 / 3.2	100	0.4 / 0	3.1
2B	0 / 0.2	0	3.1	100	0.8 / 0.1	0
2C	3.4 / 18.5	0	0	100	0	0.2 / 0
2D	2.7 / 24	4.7 / 6.7	21.6 / 37.8	100	15.7 / 24.2	0.8 / 0.6

<https://doi.org/10.1371/journal.pone.0180852.t005>

positive ones come up for the directions of the true couplings. Positive values are estimated for large coupling strength and indirect causal effects (e.g. $X_2 \rightarrow X_4$), but they are much smaller compared to those for direct causal effects.

No couplings are found in the uncoupled case ($c = 0$) for system 3 (Table 7). A Table including the percentage of significant PTE values for system 3 for all the directions is available as a Supporting file (S1 Table). The percentage of significant PTE values range from 0% to 5.6% for all the resampling methods and both time series lengths. The PTE is also effective when couplings are present. When $c = 0.2$, its sensitivity increases with n , and when $c = 0.4$ the highest sensitivity tends to be obtained even for small n .

The results for method 2D are similar to methods 1C and 2C. All the true couplings are well identified, while spurious couplings are found at a percentage higher to 5% only in three instances for $c = 0.4$ and $n = 2048$: $X_1 \rightarrow X_3$ (5.8%), $X_2 \rightarrow X_4$ (9.4%) and $X_3 \rightarrow X_5$ (15.4%).

As resampled time series become less dependent, we observe a loss in the power of the test for $n = 512$, especially when couplings are not very strong. Regarding the size of the test, for $c = 0.2$ the percentage of rejections for indirect (e.g. $X_2 \rightarrow X_4$) or no coupling (e.g. $X_5 \rightarrow X_4$) is modestly above the 5% level only for 1A and 2A, while for $c = 0.4$ is substantially higher for 1A and 2A and lower for 1B and 2B. For example, we obtain for scheme 1A and $n = 2048$: 50.5% for $X_1 \rightarrow X_3$ (indirect coupling), 22.2% for $X_2 \rightarrow X_1$ (no coupling), 56.8% for $X_2 \rightarrow X_4$ (indirect coupling), 19.7% for $X_3 \rightarrow X_2$ (no coupling), 62.2% for $X_3 \rightarrow X_5$ (indirect coupling), 22.9% for $X_4 \rightarrow X_3$ (no coupling) and 14.1% for $X_5 \rightarrow X_4$ (no coupling). Respective results are indicated by the scheme 2A. When considering more independent resampled time series, the corresponding percentages of indirect and no couplings decrease, e.g. for method 1B and $n = 2048$: 27.5% for $X_1 \rightarrow X_3$, 20% for $X_2 \rightarrow X_1$, 21.4% for $X_2 \rightarrow X_4$, 3.7% for $X_3 \rightarrow X_2$, 28% for $X_3 \rightarrow X_5$, 4.1% for $X_4 \rightarrow X_3$ and 4.7% for $X_5 \rightarrow X_4$. Similar results are observed for 2A. The correct test size, i.e. the probability of falsely rejecting the null hypothesis being close to $\alpha = 0.05$, is attained only with the resampling methods of type C; the percentage of the significant PTE values for the uncoupled cases varies from 0% to 4.7% for both 1C and 2C and both n , while spurious causality is detected for cases A and B. As n and c increase, the percentage of those spurious indications increases.

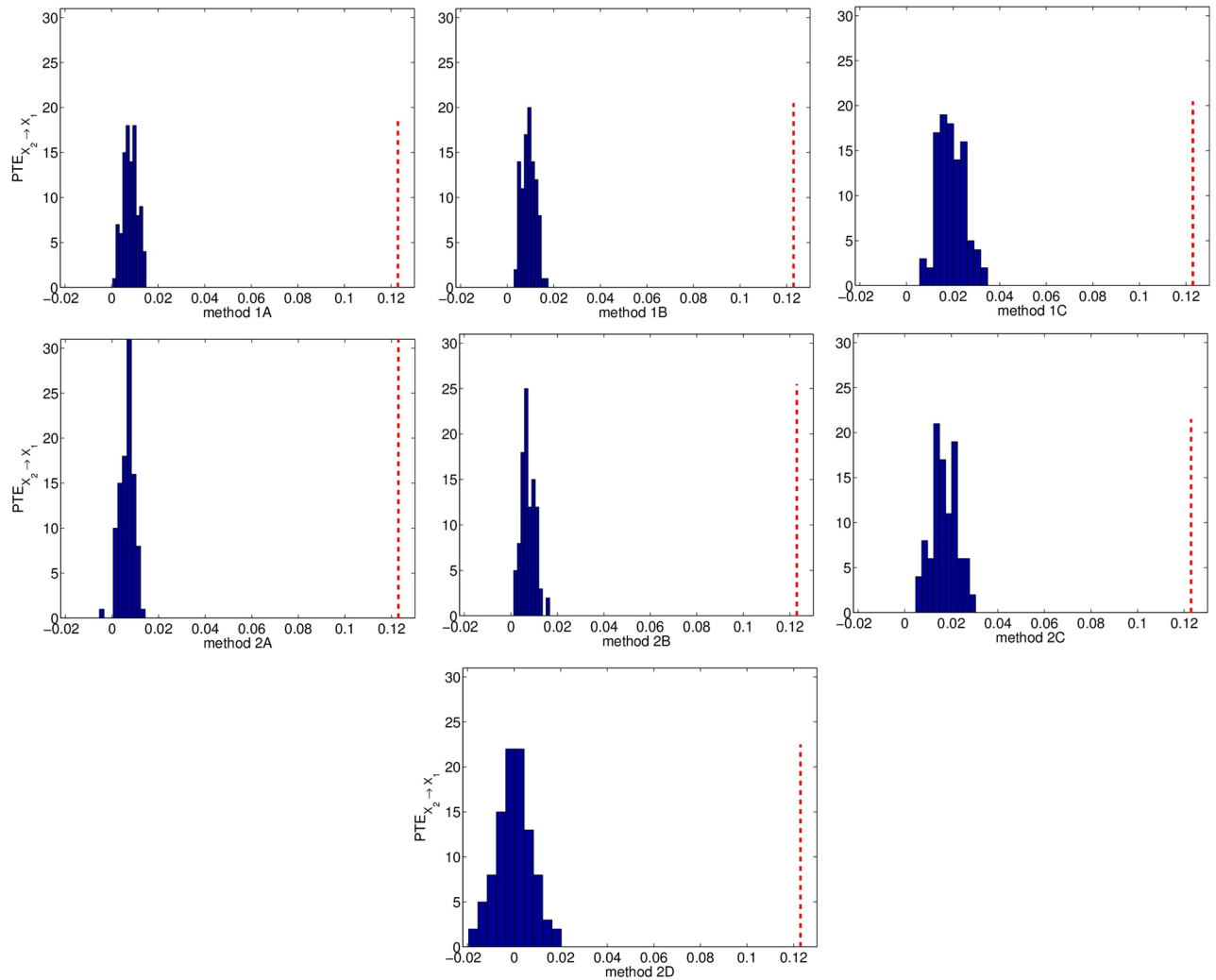


Fig 2. Distribution of surrogate/bootstrap PTE values and original PTE value (vertical dotted line) from one realization of system 2 with $n = 2048$, for the direction $X_2 \rightarrow X_1$.

<https://doi.org/10.1371/journal.pone.0180852.g002>

Application

In the effort to provide further evidence on the possible sources of US inflation in the post-Volcker era, we will try to gain insights from the application of the PTE by employing the aforementioned resampling methods. For this reason, we create two 3-variate systems of real economic variables, the first one consisting of monthly observations for the US Consumer Price Index for All Urban Consumers (CPI), the money supply (M2, Billions of Dollars) and the crude oil prices (West Texas Intermediate—Cushing, Oklahoma, Dollars per Barrel) while the second one is obtained by replacing CPI with the core CPI (Fig 4). The data are not seasonally adjusted and the sample spans from 01-01-1986 to 01-02-2014. We used the longer available sample at the time the application is implemented in order to ensure PTE accuracy. Since in the post-2009 period, US inflation reached very low values in association with the QE strategy of the Federal Reserve, we strongly believe that our findings over the period of interest (i.e. before the crisis of 2007–2009) are not qualitatively affected. To assess the impact of restricting the sample until 2009, we re-estimated the PTE for both systems. In the first case, we observe a

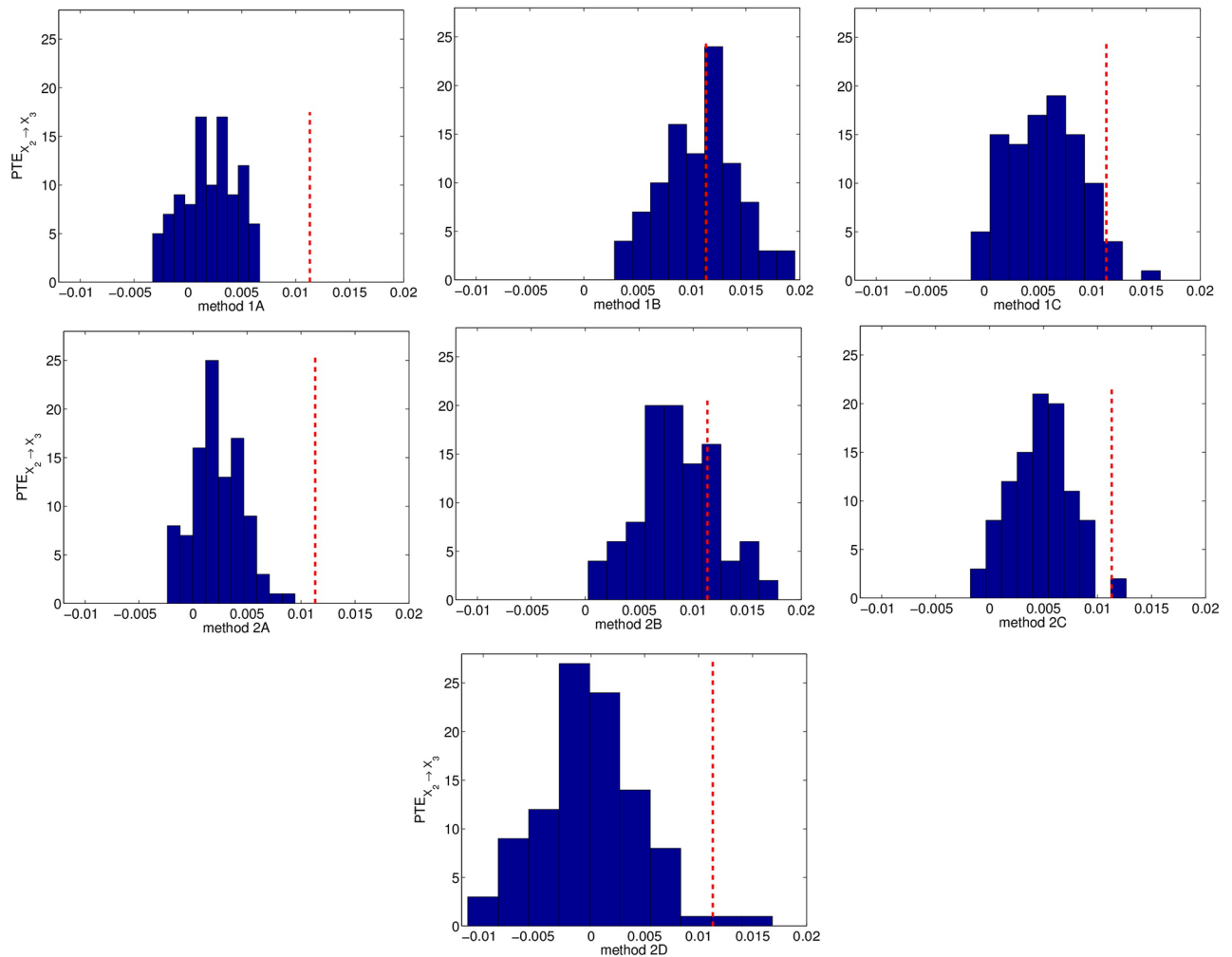


Fig 3. As Fig 2 but for the direction $X_2 \rightarrow X_3$.

<https://doi.org/10.1371/journal.pone.0180852.g003>

feedback between CPI inflation and crude oil changes, while for the 2nd system identical causal relationships appear. Prices are transformed into growth rates by using their first logarithmic differences to give inflation (Y_1) in the case of CPI, core inflation (Y_{11}), M2 returns (Y_2) and oil price changes (Y_3).

For the assessment of the statistical significance of the PTE we look back at the seven resampling methods mentioned above. The embedding dimension m for the estimation of the PTE is set equal to one ($m = 1$), often used in log differenced data expecting to have very short memory [59] and the number of nearest neighbors is ten ($k = 10$).

The empirical findings from the application of the PTE on the 1st 3-variate system consistently reveal the coupling oil (Y_3) \rightarrow inflation (Y_1). The fact that this linkage becomes statistically insignificant when the core CPI inflation is taken into account, is an indication that the observed inflation in the post-1986 period cannot be interpreted with means of traditional cost-push mechanisms. Table 8 displays the connectivity results based on each of the seven resampling methods, where statistically significant probabilities are given in bold (when p -value < 0.05). Aside from the link $Y_3 \rightarrow Y_1$, few additional links also appear. Since we do not obtain any consistency, these sparse links may be due to estimation biases, the method of

Table 6. As Table 1 but for system 3.

	n = 512			n = 2048		
	c = 0	c = 0.2	c = 0.4	c = 0	c = 0.2	c = 0.4
$X_1 \rightarrow X_2$	-0.0012	0.0104	0.0510	-0.0027	0.0274	0.1096
$X_2 \rightarrow X_1$	-0.0014	-0.0009	-0.0042	-0.0028	-0.0009	-0.0003
$X_1 \rightarrow X_3$	-0.0010	-0.0019	-0.0011	-0.0028	-0.0033	0.0014
$X_3 \rightarrow X_1$	-0.0016	-0.0016	-0.0031	-0.0028	-0.0035	-0.0034
$X_1 \rightarrow X_4$	-0.0013	-0.0023	-0.0023	-0.0027	-0.0037	-0.0043
$X_4 \rightarrow X_1$	-0.0012	-0.0015	-0.0023	-0.0027	-0.0036	-0.0043
$X_1 \rightarrow X_5$	-0.0009	-0.0022	-0.0030	-0.0030	-0.0039	-0.0046
$X_5 \rightarrow X_1$	-0.0015	-0.0012	-0.0025	-0.0027	-0.0036	-0.0039
$X_2 \rightarrow X_3$	-0.0012	0.0123	0.0576	-0.0027	0.0286	0.1079
$X_3 \rightarrow X_2$	-0.0012	-0.0002	0.0003	-0.0028	-0.0025	-0.0014
$X_2 \rightarrow X_4$	-0.0011	-0.0001	0.0028	-0.0027	-0.0027	0.0028
$X_4 \rightarrow X_2$	-0.0015	-0.0011	-0.0020	-0.0029	-0.0034	-0.0036
$X_2 \rightarrow X_5$	-0.0008	-0.0009	-0.0003	-0.0029	-0.0035	-0.0026
$X_5 \rightarrow X_2$	-0.0012	-0.0010	-0.0017	-0.0029	-0.0035	-0.0039
$X_3 \rightarrow X_4$	-0.0009	0.0135	0.0510	-0.0026	0.0300	0.1015
$X_4 \rightarrow X_3$	-0.0011	0.0004	0.0015	-0.0025	-0.0020	-0.0006
$X_3 \rightarrow X_5$	-0.0012	-0.0000	0.0028	-0.0029	-0.0028	0.0034
$X_5 \rightarrow X_3$	-0.0010	-0.0005	-0.0002	-0.0027	-0.0032	-0.0026
$X_4 \rightarrow X_5$	-0.0009	0.0122	0.0446	-0.0029	0.0284	0.0928
$X_5 \rightarrow X_4$	-0.0014	0.0002	0.0025	-0.0028	-0.0020	-0.0007

<https://doi.org/10.1371/journal.pone.0180852.t006>

Table 7. As Table 2 but for system 3 for the true couplings, an indirect coupling ($X_2 \rightarrow X_4$) and an uncoupled case ($X_5 \rightarrow X_4$).

c = 0	$X_1 \rightarrow X_2$	$X_2 \rightarrow X_3$	$X_3 \rightarrow X_4$	$X_4 \rightarrow X_5$	$X_2 \rightarrow X_4$	$X_5 \rightarrow X_4$
1A	4.5 / 5.1	5.8 / 5.6	5.6 / 5.4	5.4 / 4.1	4.9 / 4.6	3.8 / 4.8
1B	4.5 / 4.3	5.8 / 5.6	5.9 / 5.5	5.2 / 4.8	4.9 / 4.5	3.9 / 4.4
1C	1.9 / 0.6	2 / 0.5	2.2 / 0.5	2.1 / 0.5	2.2 / 0.4	1.5 / 0.6
2A	4.4 / 4.3	4.8 / 5.5	5.1 / 4.8	4.9 / 4.2	4.7 / 4.4	4.3
2B	3.3 / 2.6	3.6 / 3.3	3.7 / 2.9	3 / 2.9	2.9 / 2.3	3.2 / 3
2C	1 / 0.7	1.4 / 0.3	1.5 / 0.5	1.3 / 0.2	2.1 / 0.6	0.9 / 0.4
2D	1.9 / 0.8	2.4 / 0.9	2.3 / 0.9	1.6 / 0.8	1.7 / 0.6	1.4 / 0.7
c = 0.2	$X_1 \rightarrow X_2$	$X_2 \rightarrow X_3$	$X_3 \rightarrow X_4$	$X_4 \rightarrow X_5$	$X_2 \rightarrow X_4$	$X_5 \rightarrow X_4$
1A	58 / 100	51.8 / 100	57 / 100	52.7 / 100	6.5 / 6.6	8.1 / 10.8
1B	57.5 / 100	50.6 / 100	54.5 / 100	49.2 / 100	4.9	5.6 / 7
1C	34.3 / 100	17.5 / 100	18.9 / 100	16.6 / 100	0.5 / 0	0.5 / 0
2A	57.1 / 100	56.9 / 100	62.1 / 100	57 / 100	7.7 / 7.1	8.7 / 11.3
2B	49.8 / 100	52.1 / 100	58.1 / 100	52.2 / 100	4.9 / 2.4	6.2 / 4.2
2C	30.6 / 100	24.2 / 99.8	26 / 99.9	24.3 / 99.8	0.5 / 0	1 / 0.1
2D	31.3 / 100	34.4 / 100	38.9 / 100	33.5 / 100	3.2 / 0.8	3.4 / 0.8
c = 0.4	$X_1 \rightarrow X_2$	$X_2 \rightarrow X_3$	$X_3 \rightarrow X_4$	$X_4 \rightarrow X_5$	$X_2 \rightarrow X_4$	$X_5 \rightarrow X_4$
1A	100	99.7 / 100	99.8 / 100	99.4 / 100	14 / 56.8	14.1 / 23
1B	100	99.8 / 100	99.6 / 100	99.1 / 100	5.9 / 21.4	5 / 4.7
1C	100	85.2 / 100	87.7 / 100	84 / 100	0.4 / 0.6	0.8 / 0.2
2A	100	99.9 / 100	100	99.8 / 100	18.2 / 59.5	16.9 / 25
2B	100	99.9 / 100	99.9 / 100	99.8 / 100	11 / 26.3	9.6 / 6.4
2C	99.8 / 100	97.1 / 100	97.6 / 100	95.1 / 100	1.5 / 2.7	2.4 / 0.3
2D	99.8 / 100	99.1 / 100	99.1 / 100	98.5 / 100	5.1 / 9.4	5.1 / 2.2

<https://doi.org/10.1371/journal.pone.0180852.t007>

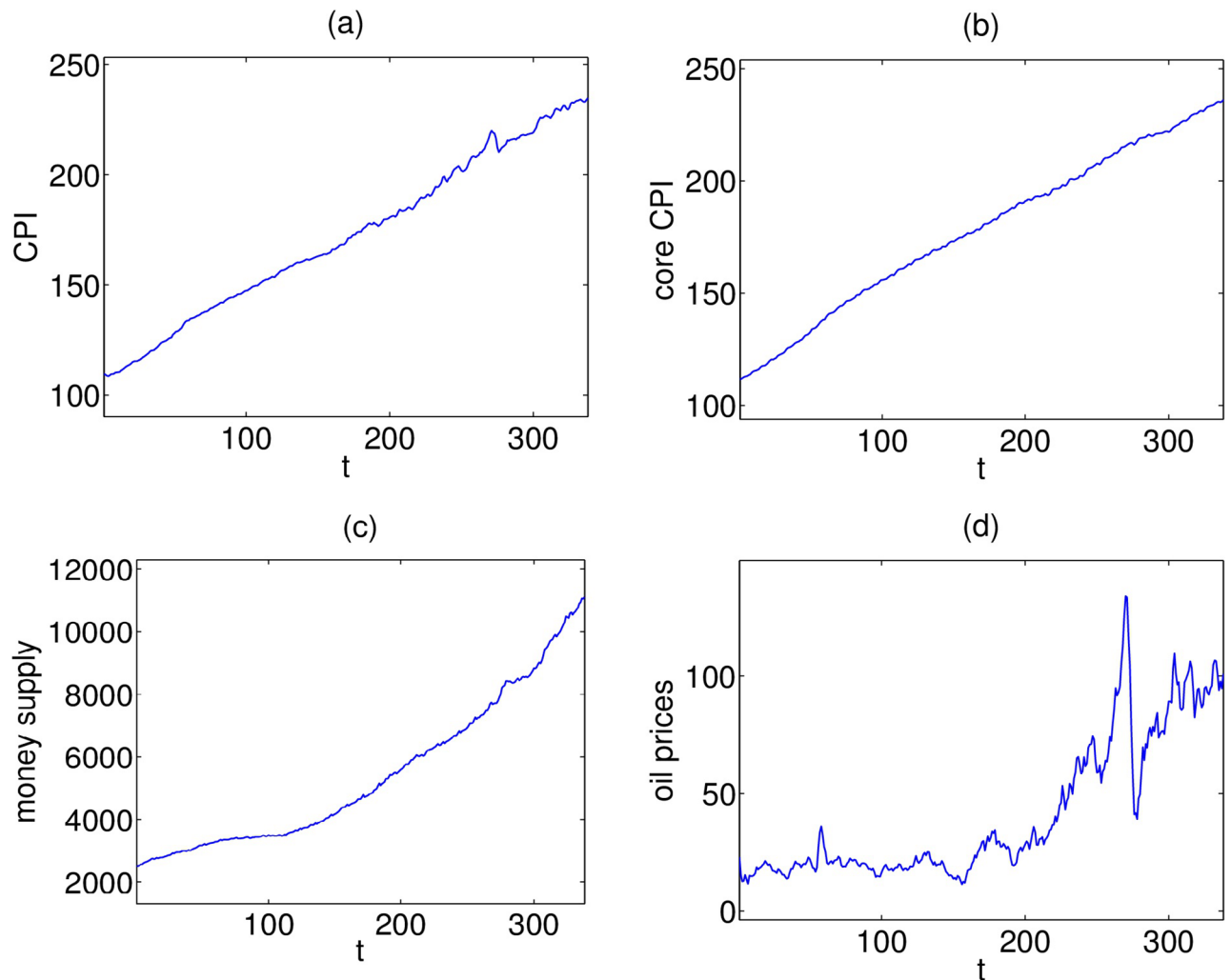


Fig 4. Monthly observations of the (a) CPI, (b) core CPI, (c) money supply and (d) oil prices, respectively.

<https://doi.org/10.1371/journal.pone.0180852.g004>

assessing its statistical significance or existence of high noise in the data. Table 9 presents the results for the 2nd 3-variate system, where the CPI inflation has been replaced by the core CPI inflation. As it can be seen, the influence of crude oil to core inflation is not statistically significant and new relationships emerge. We detect a persistent causal feedback between core

Table 8. The p-values from PTE based on the seven resampling methods for the 1st 3-variate system including the CPI inflation. Conditioning on the third variable is implied. Significant causal effects are denoted in bold.

p-value	$Y_1 \rightarrow Y_2$	$Y_2 \rightarrow Y_1$	$Y_2 \rightarrow X_3$	$Y_3 \rightarrow Y_2$	$Y_1 \rightarrow X_3$	$Y_3 \rightarrow Y_1$
1A	0.4211	0.3323	0.2533	0.0560	0.0363	0.0067
1B	0.0757	0.0165	0.3816	0.0165	0.9933	0.0165
1C	0.5296	0.2928	0.2435	0.2139	0.0363	0.0067
2A	0.1744	0.1152	0.2040	0.2435	0.0856	0.0067
2B	0.2928	0.2237	0.2336	0.3323	0.1053	0.0165
2C	0.1645	0.1053	0.3224	0.3224	0.0659	0.0067
2D	0.2237	0.1349	0.1843	0.2435	0.0757	0.0461

<https://doi.org/10.1371/journal.pone.0180852.t008>

Table 9. As Table 8 but for the 2nd 3-variate system including the core CPI inflation.

<i>p</i> -value	$Y_{11} \rightarrow Y_2$	$Y_2 \rightarrow Y_{11}$	$Y_2 \rightarrow X_3$	$Y_3 \rightarrow Y_2$	$Y_{11} \rightarrow X_3$	$Y_3 \rightarrow Y_{11}$
1A	0.0363	0.0067	0.3816	0.0461	0.4507	0.3816
1B	0.9933	0.0067	0.0067	0.0067	0.0067	0.0067
1C	0.2040	0.0659	0.4803	0.0264	0.5888	0.3915
2A	0.0165	0.0067	0.4013	0.0067	0.6776	0.3816
2B	0.0067	0.0067	0.6480	0.0461	0.7171	0.6184
2C	0.0067	0.0067	0.5592	0.0165	0.6381	0.5000
2D	0.0363	0.0067	0.3816	0.3816	0.4507	0.3816

<https://doi.org/10.1371/journal.pone.0180852.t009>

inflation (Y_{11}) and money supply (Y_2), while oil (Y_3) causes money supply (Y_2). The former bidirectional causality underscores the role of US monetary policy in controlling inflation during Great Moderation. This has been achieved through changes in the policy rate and monetary authority open market operations. Thus, via buying and selling bonds in the market, the Federal Reserve adjusted monetary base that in turns affected accordingly the monetary aggregates, such as M2. The latter unidirectional link from oil to money supply is in line with [60] conclusion that the demand-fuelled oil price rises in the 2000s have been accommodated by economic policy.

The relationship between crude oil and consumer price index has been determined dynamically over the past 50 years. The strength of the linkage seems to vary conditionally to several factors including the nature of oil shocks, the response of monetary policy and the rigidities in the labor market. In the 1970s the oil price shocks of 1973 and 1979 were associated with significant reductions in OPEC supply. In the early of middle 1980s starts a phase of stability for the US economy, known as The Great Moderation, characterized by low volatility in inflation and output. Oil prices however become more volatile again from the second half of the 1990s until mid-2008. While the oil shock episodes in 1973 and 1979 coincide with an increase in the US inflation and the beginning of rising unemployment, the variation of these two variables becomes smaller in size during the episodes of 1999–2000 and 2002–2007.

Whereas the stable core CPI in the post-1984 period, [61] show that the relative contribution of oil shocks to CPI inflation has increased since oil price changes have passed through the energy component of CPI. This lack of significant second-round effects on core inflation via cost-push mechanisms puts forward the difference in the effects of oil prices in the 1970s and the 2000s. Oil prices are not only affected by disturbances in supply. Oil shocks can be the consequence of technological changes or financial innovation able to affect consumers' demand for oil. According to [62] the oil price increase between 2009 and mid-2008 was driven by global demand shocks and as such it was not associated with recessionary dynamics of the US economy. Going further, [63] defines oil price fluctuations as symptoms of the underlying oil demand and oil supply shocks and conclude that disentangling between these two sources can prevent from unnecessary monetary policy interventions.

Conclusion

This study stems from the necessity to derive an effective causality test for the investigation of the connectivity structure of a multivariate complex system. Specifically, we investigate how the performance of a (direct) causality test is affected by the scheme generating the resampled data [29, 39, 47]. Our contribution is two-fold, with respect to the methodology and the application. Regarding the methodology, we introduce new resampling methods for the non-causality test. Regarding the application, we obtain coherent results based on the partial transfer

entropy (PTE) and all the aforementioned resampling methods, highlighting the complex nature of oil shocks through their impact on inflation.

The importance of assessing the statistical significance for the partial transfer entropy (PTE) has been explored via a simulation study. In the absence of direct coupling $X \rightarrow Y|Z$, by definition, the mutual information of X and Y conditioned on Z should be theoretically zero, i.e. $I(Y; X|Z) = 0$. The formulation of more independent resampled data (settings B and C) compared to the standard technique (setting A), all consistent to the null hypothesis $I(Y; X|Z) = 0$, seems to account better for the bias of the test statistic and helps prevent false detection of coupling in the case of the nonlinear coupled systems. The size and the power of the test are improved with settings B and C, especially if the direct couplings are strong. However, for large n and c , settings B and C may also give spurious couplings, such as for $X_2 \rightarrow X_4$ for System 3. We should also underline that the performance of PTE is affected by the number of observed variables [53]. On the other hand, when the coupled system is linear, independence setting A seems to be more efficient in identifying weak couplings. The method 2D is also effective for the nonlinear simulation systems and less effective for the linear coupled system, detecting spurious couplings.

It turns out that the PTE estimated on resampled time series increases with increasing level of randomness; i.e. the surrogate PTE values increase going from setting A to C. In addition, the spread of the surrogate PTE distribution gets larger, implying that smaller PTE values on the original time series are likely to be found statistically not significant and consequently less spurious couplings are detected. Figs 1–3 display the distribution of the surrogate PTE values for systems 1 and 2 with respect to each resampling scheme in order to visualize these findings. When we detect the true causality with high probability, we may also get spurious couplings. In order to avoid the detection of false connectivity, we may have a loss in sensitivity. This higher specificity comes at the cost of lower sensitivity, and vice versa. Thus, optimality is not achieved for any of the first six resampling methods, but it becomes clear that the significance test for the PTE gets more conservative as resampling is more random. Regarding method 2D, the bootstrap PTE values are centered by construction around zero and therefore it focuses on the spread of the distribution of the PTE on the bootstrapped data rather than the bias. For linear systems, the bias is larger and method 2D performs worse.

We note that the seven resampling methods have comparable computational cost as randomization procedures are involved at all cases in the same way. Further, they can be utilized for any test statistic in order to examine the null hypothesis of no causal effects. Ongoing research aims at further investigating the performance of various causality measures, gaining insight from the significant impact that the selection of alternative resampling techniques may have.

In the context of the application, using the PTE with all examined statistical significance tests, we confirm the stability of core inflation over the post-Volcker era including the period of Great Moderation. The strong causal influence of crude oil on the total CPI inflation and the absence of link with the core CPI inflation clearly highlight the contribution of oil demand shocks as opposite to the oil supply shocks in the 2000s that the US economy experienced in the 1970s.

Supporting information

S1 Table. Percentage of significant PTE values for system 3 for $n = 512/2048$, for all resampling methods. A single number is displayed when the same percentage corresponds to both n . The true couplings are highlighted.

(DOCX)

S1 Dataset. The matlab codes for generating the corresponding simulation time series of the manuscript are provided as a Supplementary File. The financial time series from the real applications can be downloaded from the Federal Reserve Bank of Saint Louis at the following link: <https://fred.stlouisfed.org/categories>. (ZIP)

Acknowledgments

The research project is implemented within the framework of the Action “Supporting Postdoctoral Researchers” of the Operational Program “Education and Lifelong Learning” (Action’s Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

Author Contributions

Conceptualization: AP CK DK CD.

Data curation: AP.

Formal analysis: AP DK CD.

Funding acquisition: AP CK CD.

Investigation: AP DK.

Methodology: AP DK.

Project administration: AP.

Resources: AP.

Software: AP DK.

Supervision: AP CK.

Validation: AP CK DK CD.

Visualization: AP CK DK CD.

Writing – original draft: AP.

Writing – review & editing: AP CK DK CD.

References

1. Bullmore ET, Fornito A, Zalesky A. Fundamentals of brain network analysis. Academic Press, Elsevier, First edition. 2016.
2. Gao Z-K, Small M, Kurths J. Complex network analysis of time series. Europhysics Letters. 2016; 116: 50001. <https://doi.org/10.1209/0295-5075/116/50001>
3. Fieguth P. “Complex Systems”. An Introduction to Complex Systems. Springer International Publishing. 2017; 245–269.
4. Ahmed MU, Mandic DP. Multivariate multiscale entropy: a tool for complexity analysis of multichannel data. Physical Review E. 2011; 84: 061918. <https://doi.org/10.1103/PhysRevE.84.061918>
5. Schütze H, Martinetz T, Anders S, Mamlouk AM. A multivariate approach to estimate complexity of fMRI time series. Lecture Notes in Computer Science. 2012; 7553: 540–547. https://doi.org/10.1007/978-3-642-33266-1_67
6. Schreiber T. Measuring information transfer. Physical Review Letter. 2000; 85(2):461–464. <https://doi.org/10.1103/PhysRevLett.85.461>

7. Barnett L, Barrett AB, Seth AK. Granger causality and transfer entropy are equivalent for Gaussian variables. *Physical Review Letters*. 2009; 103:238701. <https://doi.org/10.1103/PhysRevLett.103.238701> PMID: 20366183
8. Papan A, Kugiumtzis D, Larsson PG. Detection of direct causal effects and application to electroencephalogram analysis. *International Journal of Bifurcation and Chaos*. 2012; 22(9):1250222. <https://doi.org/10.1142/S0218127412502227>
9. Vakorin VA, Krakovska OA, McIntosh AR. Confounding effects of indirect connections on causality estimation. *Journal of Neuroscience Methods*. 2009; 184:152–160. <https://doi.org/10.1016/j.jneumeth.2009.07.014> PMID: 19628006
10. Amigó JM, Monetti R, Aschenbrenner T, Bunk W. Transcripts: An algebraic approach to coupled time series. *Chaos*. 2012; 22(1). PMID: 22462981
11. Amigó JM, Keller K, Unakafova VA. Ordinal symbolic analysis and its application to biomedical recordings. *Philosophical Transactions of the Royal Society A—Mathematical Physical and Engineering Sciences*. 2015; 373 (2034).
12. Kugiumtzis D. Transfer entropy on rank vectors. *Journal of Nonlinear Systems and Applications*. 2012; 3(2):73–81.
13. Kugiumtzis D. Partial transfer entropy on rank vectors. *The European Physical Journal Special Topics*. 2013; 222(2):401–420. <https://doi.org/10.1140/epjst/e2013-01849-4>
14. Staniek M, Lehnertz K. Symbolic transfer entropy. *Physical Review Letters*. 2008; 100:158101. <https://doi.org/10.1103/PhysRevLett.100.158101> PMID: 18518155
15. Abdul Razak F, Jensen HJ. Quantifying ‘causality’ in complex systems: Understanding transfer entropy. *PLoS ONE*. 2014; 9(6). <https://doi.org/10.1371/journal.pone.0099462> PMID: 24955766
16. Hirata Y, Amigo JM, Matsuzaka Y, Yokota R, Mushiake H, Aihara K. Detecting causality by combined use of multiple methods: Climate and brain examples. *PLoS ONE*. 2016; 11(7):e0158572. <https://doi.org/10.1371/journal.pone.0158572> PMID: 27380515
17. Kugiumtzis D. Direct-coupling information measure from nonuniform embedding. *Physical Review E*. 2013; 87:062918. <https://doi.org/10.1103/PhysRevE.87.062918>
18. Vlachos I, Kugiumtzis D. Non-uniform state space reconstruction and coupling detection. *Physical Review E*. 2010; 82:016207. <https://doi.org/10.1103/PhysRevE.82.016207>
19. Runge J, Heitzig J, Petoukhov V, Kurths J. Escaping the curse of dimensionality in estimating multivariate transfer entropy. *Physical Review Letters*. 2012; 108:258701. <https://doi.org/10.1103/PhysRevLett.108.258701> PMID: 23004667
20. Porta A, Faes L. Wiener-Granger causality in network physiology with applications to cardiovascular control and neuroscience. *Proceedings of the IEEE*. 2016; 104(2):282–309. <https://doi.org/10.1109/JPROC.2015.2476824>
21. Wibral M, Vicente R, Lizier JT. *Directed information measures in neuroscience*. Springer-Verlag Berlin Heidelberg; 2014.
22. Yang F, Duan P, Shah SL, Chen T. *Capturing connectivity and causality in complex industrial processes*. Springer International Publishing; 2014.
23. Kaiser A, Schreiber T. Information transfer in continuous processes. *Physica D: Nonlinear Phenomena*. 2002; 166:43–62. [https://doi.org/10.1016/S0167-2789\(02\)00432-3](https://doi.org/10.1016/S0167-2789(02)00432-3)
24. Bullmore E, Sporns O. Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*. 2009; 10(3):186–198. <https://doi.org/10.1038/nrn2575> PMID: 19190637
25. Hlinka J, Hartman D, Vejmelka M, Runge J, Marwan N, Kurths J, et al. Reliability of inference of directed climate networks using conditional mutual information. *Entropy*. 2013; 15(6):2023–2045. <https://doi.org/10.3390/e15062023>
26. Koutlis C, Kugiumtzis D. Discrimination of coupling structures using causality networks from multivariate time series. *Chaos*. 2016; 26(9). <https://doi.org/10.1063/1.4963175> PMID: 27781444
27. Schmidt C, Pester B, Schmid-Hertel N, Witte H, Wismuller A, Leistriz L. A Multivariate Granger Causality Concept towards Full Brain Functional Connectivity. *PLoS ONE*. 2016; 11(4). <https://doi.org/10.1371/journal.pone.0153105>
28. Sun XQ, Shen HW, Cheng XQ. Trading network predicts stock price. *Scientific Reports*. 2014; 4.
29. Papan A, Kugiumtzis D, Larsson PG. Reducing the bias of causality measures. *Physical Review E*. 2011; 83:036207. <https://doi.org/10.1103/PhysRevE.83.036207>
30. Marschinski R, Kantz H. Analysing the information flow between financial time series. *The European Physical Journal B*. 2002; 30:275–281. <https://doi.org/10.1140/epjb/e2002-00379-2>
31. Brandt PT, Williams JT. 2. In: *Multiple Time Series Models*. Sage Publications; 2007. p. 32–34.

32. Schelter B, Winterhalder M, Eichler M, Peifer M, Hellwig B, Guschlbauer B, et al. Testing for Directed Influences in Neuroscience using Partial Directed Coherence. *Journal of Neuroscience Methods*. 2006; 152:210–219. <https://doi.org/10.1016/j.jneumeth.2005.09.001> PMID: 16269188
33. Takahashi DY, Baccala LA, Sameshima K. Connectivity Inference between Neural Structures via Partial Directed Coherence. *Journal of Applied Statistics*. 2007; 34(10):1259–1273. <https://doi.org/10.1080/02664760701593065>
34. Goebel B, Dawy Z, Hagenauer J, Mueller JC. An approximation to the distribution of finite sample size mutual information estimates. *IEEE International Conference on Communications*. 2005;2:1102–1106.
35. Efron B. Bootstrap methods: Another look at the jackknife. *Annals of Statistics*. 1979; 7(1):1–26. <https://doi.org/10.1214/aos/1176344552>
36. Künsch HR. The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*. 1989; 17:1217–1241. <https://doi.org/10.1214/aos/1176347265>
37. Politis D, Romano JP, Lai T. Bootstrap confidence bands for spectra and cross-spectra. *IEEE Transactions on Signal Processing*. 1992; 40:1206–1215. <https://doi.org/10.1109/78.134482>
38. Politis D, Romano JP. The stationary bootstrap. *Journal of American Statistical Association*. 1994; 89(428):1303–1313. <https://doi.org/10.1080/01621459.1994.10476870>
39. Diks C, DeGoede J. A general nonparametric bootstrap test for Granger causality. *Global Analysis of Dynamical Systems*. 2001; p. 391–403.
40. Quiñero Quiroga R, Kraskov A, Kreuz T, Grassberger P. Performance of different synchronization measures in real data: A case study on electroencephalographic signals. *Physical Review E*. 2002; 65:041903. <https://doi.org/10.1103/PhysRevE.65.041903>
41. Thiel M, Romano MC, Kurths J, Rolf M, Kliegl R. Twin surrogates to test for complex synchronisation. *Europhysics Letters*. 2006; 75(4):535–541. <https://doi.org/10.1209/epl/i2006-10147-0>
42. Faes L, Porta A, Nollo G. Testing frequency-domain causality in multivariate time series. *IEEE Transactions on Biomedical Engineering*. 2010; 57(8):1897–1906. <https://doi.org/10.1109/TBME.2010.2042715>
43. Vicente R, Wibral M, Lindner M, Pipa G. Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *Journal of Computational Neuroscience*. 2011; 30:45–67. <https://doi.org/10.1007/s10827-010-0262-3>
44. Adhikari BM, Sathian K, Epstein CM, Lamichhane B, Dhamala M. Oscillatory activity in neocortical networks during tactile discrimination near the limit of spatial acuity. *NeuroImage*. 2014; 91:300–310. <https://doi.org/10.1016/j.neuroimage.2014.01.007>
45. Porta A, Faes L, Nollo G, Bari V, Marchi A, De Maria B, et al. Conditional self-entropy and conditional joint transfer entropy in heart period variability during graded postural challenge. *PLoS ONE*. 2015; 10(7):e0132851. <https://doi.org/10.1371/journal.pone.0132851>
46. Papana A, Kyrtsov C, Kugiumtzis D, Diks C. Comparison of resampling techniques for the non-causality hypothesis. In: *Topics in statistical simulation—Research papers from the 7th International Workshop on Statistical Simulation*. vol. 114. Springer Proceedings in Mathematics and Statistics; 2014. p. 419–430.
47. Papana A, Kyrtsov C, Kugiumtzis D, Diks C. Assessment of resampling methods for causality testing. CeNDEF Working Paper 14-08, University of Amsterdam; 2014.
48. Moddemeijer R. On estimation of entropy and mutual information of continuous distributions. *Signal Processing*. 1989; 16:233–248. [https://doi.org/10.1016/0165-1684\(89\)90132-1](https://doi.org/10.1016/0165-1684(89)90132-1)
49. Silverman BW. *Density estimation for statistics and data analysis*. Chapman and Hall; 1986.
50. Diks C, Manzan S. Tests for serial independence and linearity based on correlation integrals. *Studies in Nonlinear Dynamics and Econometrics*. 2002; 6(2):1–22.
51. Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. *Physical Review E*. 2004; 69(6):066138. <https://doi.org/10.1103/PhysRevE.69.066138>
52. Faes L, Porta A, Nollo G. Mutual nonlinear prediction as a tool to evaluate coupling strength and directionality in bivariate time series: Comparison among different strategies based on k nearest neighbors. *Physical Review E*. 2008; 78:026201. <https://doi.org/10.1103/PhysRevE.78.026201>
53. Papana A, Kyrtsov C, Kugiumtzis D, Diks C. Simulation study of direct causality measures in multivariate time series. *Entropy*. 2013; 15(7):2635–2661. <https://doi.org/10.3390/e15072635>
54. Yu GH, Huang CC. A distribution free plotting position. *Stochastic Environmental Research and Risk Assessment*. 2001; 15(6):462–476. <https://doi.org/10.1007/s004770100083>
55. Winterhalder M, Schelter B, Hesse W, Schwab K, Leistriz L, Klan D, et al. Comparison of linear signal processing techniques to infer directed interactions in multivariate neural systems. *Signal Processing*. 2005; 85(11):2137–2160. <https://doi.org/10.1016/j.sigpro.2005.07.011>

56. Diks C. Nonlinear time series analysis: Methods and applications, Vol. 4. World Scientific; 1999.
57. Kantz H, Schreiber T. Nonlinear time series analysis, Vol. 7. Cambridge University press; 2004.
58. Papan A, Kugiumtzis D. Evaluation of mutual information estimators for time series. *International Journal of Bifurcation and Chaos*. 2009; 19(12):4197–4215. <https://doi.org/10.1142/S0218127409025298>
59. Kwon O, Yang J-S. Information flow between stock indices. *Europhysics Letters*. 2008; 82(6):68003. <https://doi.org/10.1209/0295-5075/82/68003>
60. Summers PM. What Caused the Great Moderation? Some Cross-Country Evidence. Federal Reserve Bank of Kansas City, *Economic Review*, 3rd Quarter. 2005:5–32.
61. Blanchard JO, Gali J. The Macroeconomic Effects of Oil Shocks: Why are the 2000s So Different from the 1970s?. NBER Working Papers 13368, National Bureau of Economic Research, Inc. 2007.
62. Hamilton JD. Causes and Consequences of the Oil Shock of 2007–08. *Brookings Papers on Economic Activity*. 2009; 1:215–261. <https://doi.org/10.1353/eca.0.0047>
63. Kilian L. Oil price volatility: Origins and effects. WTO Staff Working Papers ERSD-2010-02, World Trade Organization (WTO), Economic Research and Statistics Division. 2010.