Identification of causal relationships in non-stationary time series with an information measure: Evidence for simulated and financial data

Angeliki Papana[1] · **Catherine Kyrtsou**[1] · **Dimitris Kugiumtzis**[3] · **Cees Diks**[4]

Abstract The standard linear Granger causality test, based on the vector autoregressive model (VAR), requires stationarity of the time series. A VAR model is fitted to the first-differences of the time series, when they exhibit trends and are not co-integrated. In the case of co-integration, the vector error-correction model (VECM) is used instead. Alternatively, a nonlinear information causality measure is suggested, called partial transfer entropy on rank vectors (PTERV), which uses locally ranked observations. It is model-free and of a more general purpose, as it can be directly applied to the original time series without pre-testing for stationarity or co-integration. The significance test of the PTERV detects effectively the connectivity structure of complex multivariate systems. In particular, the size and power of this test are comparable to that of the standard linear Granger causality approach (VAR or VECM) when applied to systems with only linear causal effects, while the PTERV test outperforms the linear causality test when nonlinear causal effects exist, as long as the sample size is large enough. The application of PTERV to stock market data and interest rates illustrates that it can be a useful tool in the causality analysis of financial time series.

Keywords Granger causality · non-stationarity · co-integration · VAR · VECM · PTERV · financial time series

Corresponding author: A. Papana
angeliki.papana@gmail.com, apapana@uom.edu.gr

[1]Department of Economics, University of Macedonia, Greece
· [1]Department of Economics, University of Macedonia, Greece
E-mail: ckyrtsou@uom.gr
· [3]Department of Electrical and Computer Engineering,
Aristotle University of Thessaloniki, Greece
E-mail: dkugiu@auth.gr
· [4]Department of Economics, University of Amsterdam, The Netherlands
E-mail: C.G.H.Diks@uva.nl

1 Introduction

Causality is the relationship between cause and effect. Granger causality is the statistical concept of causality that is based on prediction (Granger, 1969). The notion of Granger causality was first introduced by Clive Granger in the area of economics, and nowadays it is developed and extended to various fields, such as neuroscience (Ding et al., 2006; Porta and Faes, 2016), climatology (Stern and Kaufmann, 2014) and physics (Paluš, 2007).

In financial time series, such as interest rates and stock market data, there are strong indications for nonlinear dynamics in financial time series, such as interest rates sand stock market data, both in the mean and in the variance (Brock et al., 1991; Maki, 2003; Jones, 2006). For example, the US short-term interest rate depends in a nonlinear way on the spread between long and short-term interest rates (Granger, 1993; Anderson, 1997).

As a result, nonlinear and non-parametric methods are required to capture the complex dynamics of financial data for questions such as modelling, forecasting, making inference and detecting the interrelationships among variables (Fang and Wolski, 2019; Tiwari and Mutascu, 2014). A review of relevant parametric and nonparametric methods in financial time series can be found in Franke (2008); Zhao (2008).

The concept of Granger causality has been widely utilized for the investigation of directed interactions, mainly in economics (Geweke et al., 1983; Dufour and Taamouti, 2010). Its basic principle is to evaluate whether past values of a variable X (driving variable) help to explain current values of variable Y (response variable). The linear Granger causality test is employed by fitting vector autoregressive models. However, one should first implement a preliminary econometric analysis. If the variables are non-stationary and/or co-integrated, mis-specifications may occur (Granger and Newbold, 1974; Granger, 1988). Causality in non-stationary time series (in mean) is typically investigated through vector error correction models (VECM), addressing the presence of both short- and long-run relationships. A comparison of the prediction performance of VAR models and VECMs can be found in LeSage (1990); Clarke and Mirza (2006).

The developments in the area of econometrics were appreciated in other fields, included statistics and physics. Linear and nonlinear extensions of the Granger causality concept have been also utilized in different areas, such as brain dynamics (Hiemstra and Jones, 1994; Schiff et al., 1996). Linear causality measures originally defined for the analysis of biological signals, such as the partial directed coherence (Baccala and Sameshima, 2001), have been applied to financial data (Allali et al., 2011). Regarding nonlinear Granger causality, the test of Hiemstra and Jones (1994), which is based on the residuals of a fitted VAR model and conditional probabilities, was used broadly in finance. The test was later corrected by Diks and Panchenko (2006) and extended from the bivariate to the multivariate setting by Bai et al. (2010). The Granger causality measures from information theory actually replace the conditional probabilities in the Hiemstra and Jones test, with entropies and mutual information. The main advantage is that better estimates for entropies using k nearest neighbors estimation (Kraskov et al., 2004) can be obtained, allowing the estimation of Granger causality in higher dimensional settings. In particular, the transfer entropy (Schreiber, 2000) is applied to financial data as well (Marschinski and Kantz, 2002; Kwon and Yang, 2008). The main advantage of information-based Granger causality measures is that they are model-free and make no assumption about the distribution of the data, while they are able to detect the overall dependencies and not only the linear ones. In addition, information theory is used in empirical studies dealing with nonlinear co-integration (Aparicio and Escribano, 1998).

While transfer entropy and other nonlinear Granger causality measures require stationarity, the transfer entropy defined on ranks rather than samples, initially introduced as (partial) symbolic transfer entropy (Staniek and Lehnertz, 2008; Papana et al., 2016), and then corrected and termed (partial) transfer entropy on rank vectors (Kugiumtzis, 2012, 2013), can be applied to non-stationary time series (Kugiumtzis, 2013; Papana et al., 2016). Measures defined on ranks, transform sample values to symbols and have been recently employed for investigating Granger causality to economic and financial time series (Matilla-Garcia et al., 2014; Shi et al., 2015; Papana et al., 2016).

The equivalence of transfer entropy and the linear Granger causality has been established for Gaussian processes (Barnett et al., 2009). The extensions of transfer entropy to rank vectors, symbolic transfer entropy (Staniek and Lehnertz, 2008) and transfer entropy on rank vectors (Kugiumtzis, 2012), can also be considered as Granger causality measures. Thus, tests for Granger causality are developed using as test statistic the Granger causality index based on linear VARs, and different variants of the transfer entropy not restricted to linear cause-effect only.

In this work, we propose the use of the partial transfer entropy on rank vectors (PTERV) that is able to overcome some of the existing methodological shortcomings regarding the following aspects: a) nonlinear interdependencies are taken into consideration, b) multivariate analysis is performed, c) estimations are not restricted in cases of low memory, and d) stationarity of each examined time series in either mean or variance is not required. The suggested measure requires the sequence of rank vectors formed from the examined time series be strictly stationary, which is expected to hold for all practical purposes as the rank components define a discrete limited state space. It is not sensitive to outliers and can handle non-stationary time series in mean and variance since it is rank based. These issues have been reported with regard to a measure similar to the PTERV, the Partial Symbolic Transfer Entropy, in Papana et al. (2016). We should also note that the PTERV is not as time consuming as the Partial Symbolic Transfer Entropy. To demonstrate its effectiveness, we perform a simulation study on the basis of bivariate and multivariate time series of different nonlinear coupled systems, with stationary and non-stationary time series in mean, in the presence or absence of co-integration. In the aim to emphasize the superiority of the PTERV compared to the standard linear procedure, we provide results also for the linear parametric methodology, built on a VAR or a VEC model depending on the existence of stationarity and co-integration. An application to stock market and interest rate time series reveals the strength of the PTERV toward detecting causal effects in the case of real datasets.

The structure of the paper is as follows. In Sec. 2, the methodology is discussed, i.e. we present the PTERV and the standard linear causality procedure based on VAR/VEC models. In Sec. 3, a simulation study is performed and results are reported and discussed. In Sec. 4, the corresponding methodology is applied to financial time series, i.e. stock indexes and interest rates. Finally, Sec. 5 concludes the paper.

2 Methodology

In this section, the Partial Transfer Entropy on Rank Vectors, is presented, along with the standard linear procedure for the determination of the directed interdependencies between the variables of a complex system.

Further, the statistical significance of each of the two approaches is discussed. The examined null hypothesis for the causality test is H_0: Y_2 does not Granger cause Y_1, and it is realized either with the tests developed for VAR and VECM or the PTERV using the corresponding test statistic. The statistical significance is then assessed based on a parametric or non-parametric test in order to decide whether H_0 is accepted, i.e. whether a positive value of the test statistic is due to bias or causality.

Finally, we clarify the difference between direct and indirect causality and the notion of spurious causality providing some indicative examples.

2.1 Partial Transfer Entropy on Rank Vectors

The Partial Transfer Entropy on Rank Vectors (PTERV) is a non-parametric causality measure that stems from information theory and constitutes a robust, efficient alternative to the standard linear multi-step test procedure described in the following subsection (Kugiumtzis, 2012, 2013). The PTERV is chosen because it is model-free, requires no assumptions about the distribution of the data, remains sensitive to nonlinear effects and is not affected by non-stationarity (Kugiumtzis, 2013). Although interesting, its effectiveness on financial applications has not been investigated so far.

The PTERV utilizes rank points in contrast to the original transfer entropy (TE) that uses the original time delayed vectors. Given the multivariate time series of K variables $\mathbf{y}_t = \{y_{1,t}, y_{2,t}, ..., y_{K,t}\}$, $t = 1, \dots, n$, the time delayed vectors from each of the K variables, e.g. for Y_1, are $\mathbf{y}_{1,t} = (y_{1,t}, y_{1,t-\tau}, \dots, y_{1,t-(m-1)\tau})'$ for $t = (m-1)\tau+1, \dots, n-1$, where m is the embedding dimension and τ is the time delay. From $\mathbf{y}_{1,t}$, we form the corresponding rank-points $\hat{\mathbf{y}}_{1,t} = (r_1, \dots, r_m)'$, where $r_j \in \{1, \dots, m\}$ for $j = 1, \dots, m$ are the ranks of the corresponding amplitude values of $\mathbf{y}_{1,t}$ when arranged in ascending order. The future response of one step ahead at time t is given by the rank $\hat{y}^1_{1,t}$ of the future point $y^1_{1,t} = (y_{1,t+1})$, when sorting the observations of the joint vector $[y^1_{1,t}, \mathbf{y}_{1,t}]$.

As an example, let us consider the embedding dimension $m = 2$ and the time delay $\tau = 1$. In this case, the embedding vector $\mathbf{y}_{1,t} = (y_{1,t}, y_{1,t-1})'$ contains only two values, therefore the corresponding rank vector $\hat{\mathbf{y}}_{1,t}$ can be either $(1,2)$ or $(2,1)$. Similarly, the rank $\hat{y}^1_{1,t}$ is extracted from the joint vector $[y^1_{1,t}, \mathbf{y}_{1,t}] = (y_{1,t+1}, y_{1,t}, y_{1,t-1})'$, which can be any of the 6 triplets $(1,2,3)$, $(1,3,2)$, $(2,1,3)$, $(2,3,1)$, $(3,1,2)$, and $(3,2,1)$, and thus $y^1_{1,t}$ can take any of the values 1, 2 and 3.

Suppose we are interested in the causality from Y_2 to Y_1. Collecting the rank vectors of each of the confounding variables $Y_3, \ldots, Y_K$ (denoted all together as Z) in the vector $\hat{\mathbf{z}}_t$, allows for the definition of the PTERV as

$$\begin{aligned}\text{PTERV}_{Y_2 \to Y_1 | Z} &= I(\hat{y}^1_{1,t}; \hat{\mathbf{y}}_{2,t} | \hat{\mathbf{y}}_{1,t}, \hat{\mathbf{z}}_t) \\ &= \sum_{\hat{y}^1_{1,t}, \hat{\mathbf{y}}_{2,t}, \hat{\mathbf{y}}_{1,t}, \hat{\mathbf{z}}_t} p(\hat{y}^1_{1,t}, \hat{\mathbf{y}}_{2,t}, \hat{\mathbf{y}}_{1,t}, \hat{\mathbf{z}}_t) \log \frac{p(\hat{y}^1_{1,t} | \hat{\mathbf{y}}_{2,t}, \hat{\mathbf{y}}_{1,t}, \hat{\mathbf{z}}_t)}{p(\hat{y}^1_{1,t} | \hat{\mathbf{y}}_{1,t}, \hat{\mathbf{z}}_t)} \quad (1)\end{aligned}$$

where $I(X;Y|Z)$ is the mutual information of X and Y conditioned on Z, while $p(X,Y)$ and $p(X|Y)$ denote the joint probability mass function of X and Y, and the conditional probability mass function of X on Y, respectively, which is evaluated over all possible values of X and Y in (1). The subscript "t" in the equation of mutual information enters the name of the rank vectors, and does not denote dependence of the PTERV on the exact time step t. Note that the possible values of the vector variables, being arguments in the probabilities in (1), are the possible combinations of ranks of each rank vector. The probability masses in Eq. 1 are estimated through the corresponding relative frequencies of the observed rank vectors.

In our procedure, we need a significance test for the PTERV, i.e. to test the null hypothesis of no direct Granger causal effect from Y_2 to Y_1, which is expressed as H_0: $\text{PTERV}_{Y_2 \to Y_1|Z} = 0$.

Analytic approximations of the asymptotic null distribution for the PTERV have been investigated by Kugiumtzis (2013) and Papapetrou and Kugiumtzis (2014), who provided evidence that the Gamma distribution with parameters as given by Goebel et al. (2005) attains best convergence to the null distribution. Kugiumtzis (2013) further reports that often the approximation with resampling is superior to the parametric approximation. The resampling regards a randomization test using time-shifted surrogates (Quiroga et al., 2002). The surrogate time series are formed by time-shifting the driving variable, while the remaining variables stay intact. Considering the driving time series $\{y_{2,1}, \ldots, y_{2,n}\}$ and a random integer d ($1 + a < d < n - a$, where a is a small integer to account for autocorrelation effects), the first d values of the time series are moved to the end, so that the time-shifted time series is $\{y_{2,d+1}, \ldots, y_{2,n}, y_{2,1}, \ldots, y_{2,d}\}$. If the original PTERV value, q_0, lies at the tail of the distribution of the PTERV values, $q_1, \ldots, q_M$, using the M time-shifted time series, then H_0 is rejected. If r_0 is the rank of q_0, when ranking in ascending order the list $q_0, q_1, \ldots, q_M$, then the p-value of the one sided test is given by $1 - \frac{(r_0-0.326)}{M+1+0.348}$ (Yu and Huang, 2001). We employ and compare the performance of the PTERV assessing its statistical significance based on both the parametric (Gamma) and the non-parametric (randomization) significance test.

2.2 Standard (linear) procedure for Granger causality

The standard procedure for investigating linear Granger causality, encompasses three different scenarios. We consider the multivariate time series of K variables $\mathbf{y}_t$, $t = 1, \ldots, n$. First, a unit-root test, such as the augmented Dickey-Fuller test (Dickey and Fuller, 1979), is performed in order to determine whether each time series is covariance stationary. Then, the following alternative cases are implemented.

Case 1. Stationary time series and short-run causality

We use the conditional Granger causality index (CGCI) (Geweke, 1982). More specifically, a vector autoregressive model (VAR) is fitted to the time series

$$\mathbf{y}_t = \mathbf{c} + \sum_{r=1}^{P} A_r \mathbf{y}_{t-r} + \mathbf{e}_t, \tag{2}$$

where P is the order of the model, $\mathbf{c}$ is the $K \times 1$ constant vector, A_r are the $K \times K$ coefficient matrices of the model and $\mathbf{e}_t$ is a white Gaussian random vector with identity covariance matrix. Let us denote s_{1U}^2 and s_{1R}^2 the variances of the residuals of the unrestricted model (2) with respect only to the response variable Y_1 and the restricted model, derived from the unrestricted model by omitting the terms regarding the driving variable Y_2, respectively. Then, the CGCI is given as:

$$\text{CGCI}_{Y_2 \to Y_1|Z} = \ln(s_{1R}^2 / s_{1U}^2), \tag{3}$$

where $Z = \{Y_3, \ldots, Y_K\}$ contains the $K-2$ confounding variables.

To infer about the existence of causality, a parametric significance test can be conducted for the null hypothesis that Y_2 is not driving Y_1 making use of the F-significance test for all P coefficients $A_r(1,2), r = 1, \ldots, P$ (Brandt and Williams, 2007). Equivalently, we can employ the likelihood ratio chi-squared test, which basically uses the CGCI as test statistic (Brandt and Williams, 2007).

In the alternative to Case 1, at least one time series is non-stationary. In this scenario, we apply the Johansen co-integration test (Johansen, 1991) (considering both the trace and the eigenvalue tests) leading to the two following cases.

Case 2. Non-stationary time series and absence of co-integration

The time series are non-stationary but not co-integrated. To obtain stationary series, their first (logarithmic) differences are computed and then the CGCI is implemented as in Case 1. The existence of short-run causality is identified based on the p-values from the parametric significance test on the coefficients of the VAR model.

Case 3. Non-stationary and co-integrated time series

The properties of non-stationarity and co-integration imply the use of a VECM. The linear combination of the variables $\beta_1 y_{1,t} + \ldots + \beta_K y_{K,t}$ is stationary, where $\beta = (\beta_1, \ldots, \beta_K)'$ is the co-integrating coefficient vector.

The VECM expresses the long-run dynamics of the process and has the following form

$$\Delta \mathbf{y}_t = \mathbf{C}\mathbf{y}_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta \mathbf{y}_{t-i} + \mathbf{e}_t. \tag{4}$$

where $\mathbf{C}\mathbf{y}_{t-1}$ measures the deviation from the stationary mean at time $t-1$. The coefficients in Γ_i account for the short-run dynamics, while the matrix $\mathbf{C}$ determines the long run effect. The long- and short- run causal effects are derived for each response variable (Y_1) from another driving variable (Y_2) by testing the significance of the corresponding coefficients in the matrix $\mathbf{C}$ and the matrices Γ_i, $i = 1, \ldots, p$, respectively.

For the significance of the VECM coefficients, parametric tests are performed based on the F-statistic for Γ_i (short-run effects) and the Johansen test for the error correction terms (long-run effects).

Although the assumption of Gaussian residuals may not be required for the VECM, the possible sources of residual non-Gaussianity are of great interest, specifically when nonlinearity seems to be the driver of the data generating process of the examined time series.

2.3 Direct vs indirect Granger causality and spurious causality

Direct causality measures, such as the PTERV, exploit all the available information of the data and detect the causal influence of one variable on another one, conditioning on the confounding variables of a complex system. The advantage of partial causality tests in multivariate systems can be stressed in the following example. Considering a trivariate system composed of the variables X, Y and Z, where X drives Y ($X \to Y$) and Y drives Z ($Y \to Z$). A bivariate causality test will identify the two true direct causal effects, i.e. $X \to Y$ and $Y \to Z$, as well as the indirect causality from X to Z. In contrast, a direct causality test will indicate only the true causal couplings $X \to Y$ and $Y \to Z$.

While the indirect causality cannot be considered as completely wrong, as it is implied by other direct causality effects in the system, any other causality between two variables is regarded as spurious causality. For example, in the trivariate system, all other causal-effects than the direct ones $X \rightarrow Y$, $Y \rightarrow Z$, and the indirect one $X \rightarrow Z$, if detected by the causality test are spurious. Spurious causality can occur due to the inadequacy of the causality test but there may be other sources of spurious causality, such as causal effects of an unobserved variables and instantaneous interactions.

3 Simulations and Results

The effectiveness of both methodological procedures in detecting causal effects is assessed based on a simulation study. Different types of systems are considered; bivariate and multivariate, stationary and non-stationary, co-integrated and nonlinearly coupled. Since co-integration does not affect the performance of PTERV, only co-integrated systems with one co-integration relationship are simulated. The simulation systems are defined so that by construction one variable may have (or not) a causal influence on another one, while causality is succeeded using temporal precedence of the variables. The coupling strength can be taken into account in the coefficient of the corresponding lagged term of the driving variable in the equation of the response variable.

We perform 100 realizations of each simulation system for various time series lengths $n = 256, 512, 1024, 2048$ in order to balance the computational cost of the simulation study and the quality of the approximation of the size and power of the PTERV. Results for 1000 realizations for System 1 have also been obtained and do not significantly differ from the displayed outcomes.

For the PTERV, the time lag τ for all variables is set to 1 (as initially used in the TE definition). The embedding dimension m is common for all variables and equals the true model order from the equations of each simulation system. The order P of the VAR / VECM is equal to m. The statistical significance of the linear method is given by the parametric test, while for the PTERV both parametric and non-parametric tests of significance are employed.

The performance of the causality methods is quantified by the percentage of statistically significant couplings within the 100 realizations for each causality effect. The couplings are always regarded to be conditioned on the remaining variables, if the system is multivariate. For consistency, we keep the notations CGCI, VECM and PTERV also for the bivariate simulation systems.

We note that for the linear procedure, we perform for each realization of the simulation systems stationarity and co-integration tests in order to determine the respective case, assuming that the properties of each system are unknown. Nonetheless, the suitable linear procedure is considered for each simulation system, based on the generating data process and therefore independently of the outcome of the co-integration test. However, the effectiveness of the co-integration tests is reported.

System 1.

A non-stationary, nonlinear, multivariate system, with the unidirectional couplings $X_1 \to X_2$, $X_2 \to X_3$, is generated by integrating the time series from three coupled Hénon maps with coupling strength $c = 0.2$

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

We set $m = P = 2$. The initial values of the three variables, required for the simulation, are randomly chosen.

The results in Table 1 indicate that the PTERV detects the true causalities, even for small time series lengths, while the percentage of significant values for the uncoupled directions are low. Similar conclusions are drawn from the two different significance tests.

Table 1 Percentages of rejections of the non-causality hypothesis H_0 based on the PTERV ($m = 2$) (parametric / non-parametric statistical significance) and the CGCI ($P = 2$), for the simulation System 1. When the same percentage has been found for the different significance tests for the PTERV, a single number is displayed in the cell. The true couplings are given in bold. All the indicated couplings are conditioned on the third variable of the system.

	PTERV ($m=2$)				CGCI ($P=2$)			
n	256	512	1024	2048	256	512	1024	2048
$\mathbf{X_1} \to \mathbf{X_2}$	51 / 54	81	99	100	46	91	100	100
$X_2 \to X_1$	8 / 6	7 / 5	2 / 4	7 / 9	7	5	8	32
$\mathbf{X_2} \to \mathbf{X_3}$	56 / 54	91	100 / 99	100	22	60	93	100
$X_3 \to X_2$	7 / 10	5 / 7	13 / 9	16	20	10	16	18
$X_1 \to X_3$	5 / 9	10 / 11	7 / 10	8 / 9	6	8	9	7
$X_3 \to X_1$	5	18 / 10	6 / 3	8 / 9	10	12	4	6

The CGCI is applied to the corresponding stationary time series, before integration or equivalently after taking the first-differences of the non-stationary time series (Case 2). It correctly indicates the couplings, however also spurious causal relationships are observed, e.g. $X_2 \to X_1$ for $n = 2048$ (32%) and $X_3 \to X_2$ for $n = 256$ (20%) (see Table 1).

For the aforementioned system, the PTERV outperforms the CGCI. We note that BIC suggests to set $P = 3$ for which a loss of power and size occurs compared to the presented results for $P = 2$. The Augmented Dickey-Fuller unit root test does not always indicate that all three time series are non-stationary. Finally, in 87% of the simulated data the Johansen co-integration test confirms the absence of co-integration. A topic for further investigation is whether a non-parametric stationarity and co-integration test would be more appropriate for such data that contain strong nonlinearity.

System 2.

This is a non-stationary, stochastic system in three variables, with unidirectional nonlinear coupling $X_1 \to X_2$ and linear coupling $X_1 \to X_3$, specified as

$$\begin{aligned} x_{1,t} &= x_{1,t-1} + \epsilon_{1,t} \\ x_{2,t} &= 0.1x_{1,t-1}^2 - 0.2x_{2,t-2} + \epsilon_{2,t} \\ x_{3,t} &= 0.3x_{1,t-1} - 0.1x_{3,t-1} + \epsilon_{3,t}, \end{aligned}$$

where $\epsilon_{i,t}$, $i = 1, 2, 3$ are standard normal white noise innovations with identity covariance matrix and independent to each other (the same applies for all the following systems). The time series derived from the equations for random initial values are integrated. Thus, the time series are non-stationary but non co-integrated as there is no linear combination of the variables that is stationary. We also set $P = m = 2$.

The PTERV effectively shows the true couplings, although spurious causalities arise for large n, e.g. $X_2 \to X_3$. Therefore, short-run causality can be again obtained based on the suggested non-parametric measure directly from the original non-stationary data.

Table 2 As Table 1 but for System 2.

	PTERV ($m = 2$)				CGCI ($P = 2$)			
n	256	512	1024	2048	256	512	1024	2048
$\mathbf{X_1 \to X_2}$	74 / 75	89 / 88	92	98	88	93	93	99
$X_2 \to X_1$	5	0 / 1	7 / 8	5 / 6	7	8	11	8
$X_2 \to X_3$	6 / 3	14	17 / 16	32 / 33	100	100	100	100
$X_3 \to X_2$	4	6 / 4	15 / 16	29 / 27	99	100	100	100
$\mathbf{X_1 \to X_3}$	36 / 32	67	98 / 96	100	43	40	52	58
$X_3 \to X_1$	7	2 / 3	2	2	100	100	100	100

Since the variables are not co-integrated by construction, the linear CGCI is applied in the first log-differences. However, the Johansen co-integration test, reveals that almost 50% of the realizations of this system present one co-integrating relationship based on the trace statistic. The linear measure fails to identify the causal links for this simulation system, since almost all resulting couplings are bidirectional (see Table 2). When nonlinearities are present, the linear methods such as the co-integration test and the CGCI do not seem to be effective.

System 3.

This is a non-stationary, nonlinear (chaotic), bivariate system, with co-integrated variables and unidirectional coupling $X_1 \to X_2$. It is generated by superimposing (adding) to each of the variables X_1 and X_2 at time t of System 1 (i.e. two coupled Hénon maps), two co-integrated random walks $r_{1,t}$ and $r_{2,t}$, respectively, defined as

$$r_{1,t} = w_t + \epsilon_{1,t}$$
$$r_{2,t} = 0.3w_t + \epsilon_{2,t},$$

with a common stochastic drift $w_t = w_{t-1} + \epsilon_t$, where ϵ_t, $\epsilon_{1,t}$, $\epsilon_{2,t}$ are standard normal white noise innovations. Thus, each of the two time series $y_{1,t}, y_{2,t}$ of System 3 is generated by the sum $y_{i,t} = x_{i,t} + r_{i,t}$, $i = 1, 2$ and $t = 1, \ldots, n$, where the first term is nonlinear (chaotic) and the second term is a random walk. We consider the coupling strength $c = 0.2$, while we set $m = P = 2$.

As the Table 3 reports, the nonlinear method identifies the correct causal effects with power that increases with n.

Table 3 As Table 1 but for System 3, where for the linear procedure short- and long-run causality is assessed by the fit of VECM ($P = 2$). For long-run causality, the percentages of not rejecting the null hypothesis that the number of co-integrating relations r is ≤ 0 or ≤ 1 are reported.

PTERV	causality			
n	256	512	1024	2048
$\mathbf{Y_1 \to Y_2}$	8 / 10	8 / 11	22 / 26	39 / 41
$Y_2 \to Y_1$	8 / 7	5	14 / 16	7 / 9

VECM	short-run causality			
n	256	512	1024	2048
$\mathbf{Y_1 \to Y_2}$	92	100	100	100
$Y_2 \to Y_1$	28	67	85	100
	long-run causality			
$r <= 0$	0	0	0	0
$r <= 1$	49	60	54	46

On the other hand, the VECM suggests that there is short-run causality in both directions while long-run causality is obtained for almost half of the examined samples, with one co-integration relation (see Table 3). As n increases, the percentage of significant short-run causalities for the spurious link $Y_2 \to Y_1$ increases. We note that for $n = 2048$, the residuals of the VECM are not normally distributed in 66% of the realizations (for both VECMs with Y_1 and Y_2 being the dependent variable, respectively). Further, the residuals have statistically significant autocorrelation at least for lag one. Finally, heteroskedasticity, based on the Engle test, is detected in 11% (for Y_1) and 12% (for Y_2) of the cases. The comparative advantage of a model-free method is apparent. The usually observed drawbacks of modelmis-specifications seem to be effectively overcome with the application of the PTERV.

System 4.

We consider three co-integrated random walks, i.e.

$$r_{1,t} = w_t + \epsilon_{1,t}$$

$$r_{2,t} = 0.3w_t + \epsilon_{2,t}$$
$$r_{3,t} = 0.6w_t + \epsilon_{3,t}$$

with a common stochastic drift $w_t = w_{t-1} + \epsilon_t$, where a coupled system with linear ($Y_2 \rightarrow Y_3$) and nonlinear couplings ($Y_1 \rightarrow Y_2$ and $Y_1 \rightarrow Y_3$) is superimposed via

$$y_{1,t} = 3.4y_{1,t-1}(1 - y_{1,t-1}^2)e^{-y_{1,t-1}^2} + 0.4\delta_{1,t}$$
$$y_{2,t} = 3.4y_{2,t-1}(1 - y_{2,t-1}^2)e^{-y_{2,t-1}^2} + 0.5y_{1,t-1}y_{2,t-1} + 0.4\delta_{2,t}$$
$$y_{3,t} = 3.4y_{3,t-1}(1 - y_{3,t-1}^2)e^{-y_{3,t-1}^2} + 0.5y_{2,t-1} + 0.5y_{1,t-1}^2 + 0.4\delta_{3,t}$$

where ϵ_t, $\epsilon_{i,t}$ and $\delta_{i,t}$ are Gaussian innovations with zero mean and variance one. The considered system is given as $x_{i,t} = r_{i,t} + y_{i,t}$ for $i = 1, 2, 3$. The causality tests are performed for $m = P = 2$.

As previously mentioned, the selection of long samples maximizes the performance of the PTERV. The corresponding percentages of significant PTERV values increase with n (see Table 4). For $n = 2048$, the spurious coupling $X_3 \rightarrow X_1$ and the indirect coupling $X_1 \rightarrow X_3$ are also detected.

Table 4 As Table 3 but for System 4.

PTERV	causality			
n	256	512	1024	2048
$\mathbf{X_1 \rightarrow X_2}$	11 / 10	12 / 6	14	22 / 35
$X_2 \rightarrow X_1$	3 / 5	5 / 9	8 / 14	12 / 6
$\mathbf{X_2 \rightarrow X_3}$	13 / 6	17 / 18	40 / 25	61 / 64
$X_3 \rightarrow X_2$	5 / 1	5 / 9	3 / 6	3 / 8
$X_1 \rightarrow X_3$	8 / 7	14 / 11	21 / 22	43 / 39
$X_3 \rightarrow X_1$	6 / 8	9 / 19	18 / 11	33 / 37

VECM	short-run causality			
n	256	512	1024	2048
$\mathbf{X_1 \rightarrow X_2}$	9	30	59	84
$X_2 \rightarrow X_1$	6	5	7	14
$\mathbf{X_2 \rightarrow X_3}$	2	4	4	7
$X_3 \rightarrow X_2$	7	13	14	30
$X_1 \rightarrow X_3$	47	85	98	100
$X_3 \rightarrow X_1$	25	52	73	97
	long-run causality			
$r <= 0$	0	0	0	0
$r <= 1$	0	0	0	0
$r <= 2$	66	72	66	71

The VECM (Case 3) correctly indicates the link $X_1 \rightarrow X_2$ but fails to reveal $X_2 \rightarrow X_3$. The spurious relationships $X_1 \rightarrow X_3$ and $X_3 \rightarrow X_1$ are found in the short-run (Table 4). For this system, the VECM residuals present significant autocorrelations, especially for large n (e.g. for $n = 2048$, we obtain residual autocorrelations in 100%, 98% and 69% of the realizations for X_1, X_2 and X_3 as dependent variable, respectively). Further, we get a high percentage of heteroskedastic effects for X_2 as dependent variable (e.g. 36% for $n = 2048$).

Regarding the long-run causality, two co-integration relations are suggested with a percentage between 66% and 72% for all sample sizes. Since the VECM is not adequately fitted to the data, the nonlinear methodology seems to be a more appropriate alternative for indicating the existence of causality among the variables.

Although the PTERV is proposed for the detection of the causality when nonlinearities exist, we also provide evidence for its performance when only linear couplings are present.

System 5

A non-stationary, linear, bivariate model with bidirectional couplings $X_1 \leftrightarrow X_2$, is generated by integrating the variables from the system given by

$$\begin{aligned} x_{1,t} &= -0.7 + 0.7x_{1,t-1} + 0.2x_{2,t-1} + \epsilon_{1,t} \\ x_{2,t} &= \;\; 1.3 + 0.2x_{1,t-1} + 0.2x_{2,t-1} + \epsilon_{2,t}. \end{aligned}$$

The variables are not co-integrated. System 5 is taken into consideration for three reasons; first, in order to evaluate the performance of the PTERV only in the presence of linear couplings, secondly when a bidirectional coupling exists, and finally to investigate the effect of mis-specifying the embedding dimension m. From the definition of the PTERV, we can only set $m \geq 2$. However, the true order model of this system is one.

The PTERV (for $m = 2$) seems to be robust only for large time series lengths based on the non-parametric significance test, while for the parametric one, it gives high percentages of significant values also for smaller n (see Table 5). The CGCI is applied to the first-differenced time series, since data are non-stationary and non co-integrated (Case 2). It effectively denotes the true couplings for all time series lengths, giving similar results when $P = 1$ and $P = 2$ (see Table 5 for $P = 2$).

Table 5 As Table 1 but for System 5.

	PTERV ($m = 2$)					CGCI ($P = 2$)			
n	256	512	1024	2048	4096	256	512	1024	2048
$\mathbf{X_1 \rightarrow X_2}$	53 / 13	48 / 16	35 / 31	51 / 58	84 / 86	97	99	100	100
$\mathbf{X_2 \rightarrow X_1}$	47 / 0	38 / 3	22 / 9	18 / 14	37 / 41	99	100	100	100

System 6

A non-stationary, linear, multivariate system, with unidirectional couplings $X_2 \rightarrow X_1$ and $X_3 \rightarrow X_1$, and co-integrated variables, defined as

$$x_{1,t} = 0.4x_{1,t-1} + 0.4x_{2,t-1} + 0.5x_{3,t-1} + 0.2x_{1,t-2}$$

$$-0.2x_{2,t-2} - 0.2x_{1,t-3} + 0.15x_{2,t-3} + 0.1x_{3,t-3} + \epsilon_{1,t}$$
$$x_{2,t} = 0.6x_{2,t-1} + 0.2x_{2,t-2} + 0.2x_{2,t-3} + \epsilon_{2,t}$$
$$x_{3,t} = 0.4x_{3,t-1} + 0.3x_{3,t-2} + 0.3x_{3,t-3} + \epsilon_{3,t}.$$

The variables of this system are $I(1)$ and co-integrated with one co-integration relationship (see (Sharp, 2010), Model 8, p.78). We set $m = P = 3$.

The parametric significance test for the PTERV spuriously indicates all couplings to be significant for $n = 1024$ and 2048, while the non-parametric one correctly detects the direct couplings only for $n = 1024$ and 2048. The corresponding percentages of significant PTERV values increase with n (Table 6). Based on the standard linear procedure of VECM (Case 3), the short-run causal relations $X_2 \to X_1$ and $X_3 \to X_1$ are correctly captured, with a confidence that increases with the sample size (see Table 6). The results from the Johansen trace test suggest the existence of long-run relationships, while two co-integration relations are indicated in most cases.

Table 6 As Table 3 but for System 6.

PTERV	causality			
n	256	512	1024	2048
$X_1 \to X_2$	0 / 1	0 / 1	87 / 0	90 / 6
$\mathbf{X_2 \to X_1}$	0 / 4	0 / 9	96 / 25	100 / 93
$X_2 \to X_3$	0 / 0	0 / 3	93 / 1	88 / 7
$X_3 \to X_2$	0 / 0	0 / 2	86 / 0	82 / 5
$X_1 \to X_3$	0 / 5	0	95 / 1	92 / 5
$\mathbf{X_3 \to X_1}$	0 / 5	0 / 7	100 / 33	100 / 97

VECM	short-run causality			
n	256	512	1024	2048
$X_1 \to X_2$	6	6	10	7
$\mathbf{X_2 \to X_1}$	63	97	100	100
$X_2 \to X_3$	4	4	5	5
$X_3 \to X_2$	8	6	5	6
$X_1 \to X_3$	4	9	5	2
$\mathbf{X_3 \to X_1}$	22	48	76	96
	long-run causality			
$r <= 0$	0	0	0	0
$r <= 1$	80	92	85	90
$r <= 2$	87	91	90	91

System 7

Finally, we consider a system that is non-stationary in variance, generated by superimposing three integrated generalized autoregressive conditional heteroskedastic processes of order (1,1), IGARCH (1,1), to the system of three coupled Hénon maps (System 1). The IGARCH (1,1) is given as:

$$z_{i,t} = \sigma_{i,t}\epsilon_{i,t}$$
$$\sigma^2_{i,t} = 0.2 + 0.9\epsilon^2_{i,t-1} + 0.1\sigma^2_{i,t-1}, \quad i = 1, 2, 3$$

where $\epsilon_{i,t}$ is a Gaussian white noise with unit variance. Specifically, each $z_{i,t}$ of IGARCH(1,1) ($i = 1, 2, 3$) is first standardized to have zero mean and standard deviation equal to one. Then, the standarized series, denoted as $zs_{i,t}$, are multiplied by the factor $g = 0.2$ and are added to each $x_{i,t}, i = 1, 2, 3$ of the coupled Hénon map, so that the derived time series of System 7 are given as $y_{i,t} = x_{i,t} + gzs_{i,t}, i = 1, 2, 3$. In this example, variables are not co-integrated and the nonlinear causality is known, i.e., $Y_1 \to Y_2$ and $Y_2 \to Y_3$.

The CGCI ($P = 2$) correctly suggests the causal links, however an increased percentage of significant causality appears for $Y_2 \to Y_1$ and $Y_3 \to Y_2$ for large samples (Table 7). On the other hand, the PTERV ($m = 2$), adequately identifies the connectivity of the system, both with parametric and non-parametric significance tests.

Table 7 As Table 1 but for System 7.

	PTERV ($m = 2$)				CGCI ($P = 2$)			
n	256	512	1024	2048	256	512	1024	2048
$\mathbf{Y_1 \to Y_2}$	25 / 33	72	96 / 98	100	46	83	100	100
$Y_2 \to Y_1$	4 / 6	3 / 5	1	0 / 1	4	5	14	23
$\mathbf{Y_2 \to Y_3}$	44 / 37	77 / 71	98 / 99	100	27	52	86	100
$Y_3 \to Y_2$	10	7	11 / 6	12 / 11	9	15	16	28
$Y_1 \to Y_3$	7 / 8	3 / 2	3 / 8	9 / 12	10	12	10	6
$Y_3 \to Y_1$	2 / 5	4 / 7	3 / 6	0 / 2	7	10	11	9

To sum-up, the PTERV indicates appropriately Granger causality when nonlinearities are present and outperforms the standard linear procedure. Although it has been developed to address features of nonlinear data, it performs well also in the existence of linear couplings, such as for the simulation Systems 5 and 6. When nonlinear couplings are assumed, as in the first four systems, the PTERV overpowers the linear method. The PTERV has low power for small time series lengths but it improves with increasing n. Further, it seems to be able to handle data that are non-stationary in variance. Finally, no major differences in the performance of the PTERV are reported, when the estimation of its statistical significance is parametric or non-parametric.

As a closing remark, we have to mention that the setting in general was favorable for the linear procedure, since the PTERV on the integrated (non-stationary) time series is compared to the CGCI on the original (stationary) time series. A more realistic scenario, not applied here, would be to form the non-stationary time series by adding a stochastic trend to the stationary time series, e.g. as in Kugiumtzis (2013). Then, the first-differencing would not produce the original stationary time series and the results of CGCI would deviate, depending on the success of detrending.

4 Application

In the effort to evaluate the performance of the methodological procedures, presented in Section 2, on real data, we proceed with an application to two sets of financial time series: stock indexes and interest rates. Since the respective variables obey unknown nonlinear features, non-parametric methods are required for their analysis.

The PTERV is estimated on the original time series (prices) in both applications, the VECM is applied to the logarithmic time series in order to avoid spurious couplings due to the variability of the data (Tsay, 2005; Lütkepohl and Xu, 2011) and finally for the CGCI, the return series are calculated. To gain insight from the PTERV measure and avoid eventual biases induced by high dimension, we chose to have three time series at each of the two real data sets. We recall that, by definition, the PTERV is not affected by the logarithmic transformation or any other monotonic transformation.

In the first application, we study the causal relationships among international stock markets. We consider the Morgan Capital International's market capitalization weighted index data of three markets consisting of daily measurements from March 5, 2004 until March 5, 2009 for Germany (X_1), Greece (X_2) and US (X_3) (https://www.msci.com/market-cap-weighted-indexes) (see Fig. 1). The logarithmic time series are denoted by Y_1, Y_2 and Y_3, respectively.

The PTERV ($m = 2$) is estimated on the prices and shows that the US drives Greece and Germany (p-values = 0.007, see Table 8). Further, between Germany and Greece a bidirectional coupling is statistically significant (p-value = 0.007).

Table 8 The estimated p-values from PTERV and CGCI for the first application. The rows indicate the driving variables and the columns the respective response ones.

PTERV	X_1	X_2	X_3	CGCI	YY_1	YY_2	YY_3
X_1	-	0.007	0.086	YY_1	-	0.057	0.224
X_2	0.007	-	0.510	YY_2	0.025	-	0.115
X_3	0.007	0.007	-	YY_3	0	0	-

Following the linear test procedure, the Augmented Dickey-Fuller test is implemented. The results in Table 9 show that the logarithmic time series are non-stationary, $I(1)$. Then, the Johansen co-integration test is applied to the log-transformed data for two lags (based on BIC). The test confirms that there is not co-intergration (Table 10). Since the time series are $I(1)$ and non co-integrated, the CGCI is applied to the logarithmic returns, denoted as YY_1, YY_2, YY_3, respectively (Case 2). It appears that the US drives Greece and Germany (p-values < 0.0001), while Germany causes Greece (p-value$= 0.025$) (Table 8).

In the second application, we consider weekly measurements of interest rates (in percent, not seasonally adjusted) for the period 5/1/1962 - 22/11/2013

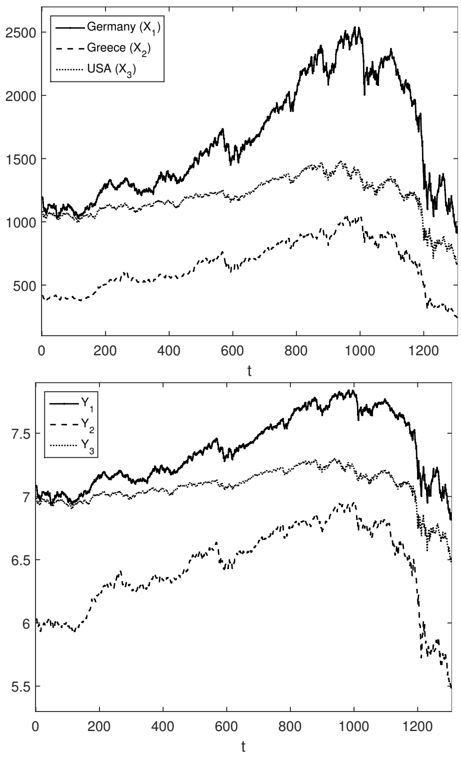

Fig. 1 (a) Original prices of developed markets and (b) logarithmic prices.

Table 9 Augmented Dickey-Fuller test results for the first application.

var.	p-val.	stat.	crit. value
Y_1	0.990	1.749	-3.414
Y_2	0.990	2.052	-3.414
Y_3	0.990	2.066	-3.414

Table 10 Results from Johansen co-integration test for application 1.

logarithmic prices						
null	trace stat.	crit 90%	crit 95%	eigen. stat	crit 90%	crit 95%
$r \leq 0$	51.391	32.065	35.012	29.325	21.873	24.252
$r \leq 1$	22.066	16.162	18.398	15.933	15.001	17.148
$r \leq 2$	6.133	2.705	3.841	6.133	2.705	3.841

such as the 3-Month Treasury Bill of the Secondary Market Rate (3MTB), the effective Federal Funds Rate (FF) and the 10-Year Treasury Constant Maturity Rate (10YTN) (https://fred.stlouisfed.org/categories/22). The logarithmic time series are denoted by Z_1, Z_2 and Z_3, respectively (see Fig. 2).

The PTERV is estimated for $m = 2$. The empirical findings indicate two drivings: the 3MTB on the FF (p-value from surrogate test is 0.007) and of

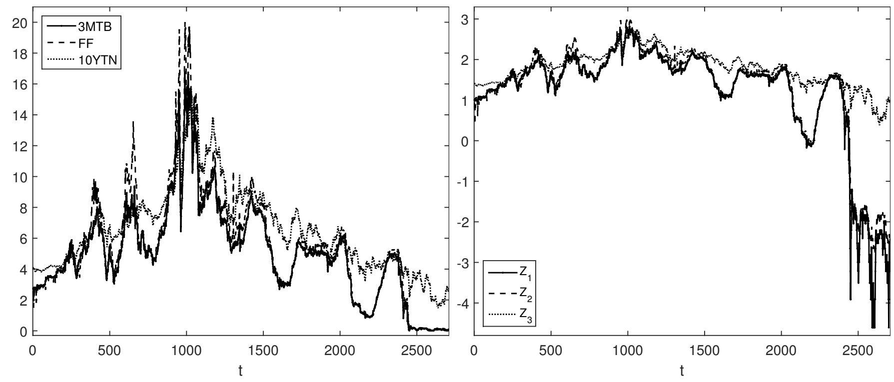

Fig. 2 (a) Interest rates in levels and (b) logarithmic series.

the 10YTN on the FF (p-value= 0.007). Furthermore, a bidirectional coupling exists between 3MTB and 10YTN (p-value= 0.007). Results are displayed in Table 11. We note that the PTERV leads to identical findings for embedding dimension $m = 3$, while for larger m, no couplings occur.

Table 11 The estimated p-values for PTERV ($m = 2$) for the second application.

PTERV	3MTB	FF	10YTN
3MTB	-	0.007	0.007
FF	0.076	-	0.322
10YTN	0.007	0.007	-

Regarding the standard linear procedure, the Augmented Dickey-Fuller test shows that the data are non-stationary ($I(1)$) (see Table 12). The Johansen co-integration test is then employed for $P = 6$ (based on BIC), suggesting one co-integrating relationship (see Table 13). Thus, the VECM is built on the logarithmic prices for $P = 6$ (Case 3). Analytical results are reported in Table 14, including the coefficients of the model for each lag and the respective probability. In the short-run, we capture the bidirectional couplings 3MTB $\leftrightarrow$ FF and 3MTB $\leftrightarrow$ 10YTN together with the unidirectional FF $\rightarrow$ 10YTN (p-values < 0.05).

We recall that the outcomes of the VECM are sensitive to the selection of P. As P increases, the number of causal links increases as well. It also turns out that the VECM residuals are not normally distributed based on the Jarque-Bera test of composite normality (p-value < 0.001). In the same line, the Ljung-Box Q-test indicates residuals autocorrelation (p-value < 0.0001). Finally, the Engle test detects ARCH effects (p-value < 0.0001).

Table 12 Augmented Dickey-Fuller test results for the second application.

var.	p-val.	stat.	crit. value
Z_1	0.528	-2.131	-3.414
Z_2	0.945	-0.976	-3.414
Z_3	0.523	-2.142	-3.414

Table 13 Johansen co-integration test results for the second application.

logarithmic prices						
null	trace stat.	crit 90%	crit 95%	eigen. stat	crit 90%	crit 95%
$r \leq 0$	114.692	32.065	35.012	103.470	21.873	24.252
$r \leq 1$	11.221	16.162	18.398	9.601	15.001	17.148
$r \leq 2$	1.620	2.705	3.841	1.620	2.705	3.841

Table 14 Results from VECM ($P = 6$) for the second application.

		Eq. 1		Eq. 2		Eq. 3	
Variable	lag	coef.	prob.	coef.	prob.	coef.	prob.
Y_1	1	0.0338	0.1095	0.0886	0.0000	0.0057	0.2044
Y_1	2	0.0021	0.9195	0.0236	0.0672	0.0100	0.0253
Y_1	3	0.0182	0.3798	0.1201	0.0000	0.0167	0.0002
Y_1	4	-0.0052	0.8045	0.0950	0.0000	0.0090	0.0446
Y_1	5	-0.0048	0.8175	0.0692	0.0000	0.0128	0.0041
Y_1	6	0.0523	0.0123	0.0106	0.4107	0.0018	0.6882
Y_2	1	-0.0359	0.2713	-0.1767	0.0000	-0.0057	0.4096
Y_2	2	0.1336	0.0000	-0.0959	0.0000	-0.0180	0.0095
Y_2	3	-0.0631	0.0487	-0.1132	0.0000	-0.0048	0.4804
Y_2	4	0.0205	0.5043	-0.0111	0.5558	-0.0042	0.5244
Y_2	5	0.2408	0.0000	-0.0152	0.4139	0.0019	0.7728
Y_2	6	0.0041	0.8902	-0.0041	0.8224	0.0187	0.0033
Y_3	1	0.6259	0.0000	0.0752	0.1844	0.2041	0.0000
Y_3	2	-0.0922	0.3279	0.1507	0.0092	0.0316	0.1157
Y_3	3	0.0027	0.9773	-0.0320	0.5812	-0.0067	0.7407
Y_3	4	0.0438	0.6418	-0.0004	0.9944	0.0341	0.0900
Y_3	5	0.0331	0.7252	0.0365	0.5282	-0.0306	0.1277
Y_3	6	-0.0432	0.6412	-0.0586	0.3031	0.0027	0.8925
ec term	1	0.014664	0.0000	-0.0062	0.0000	0.0008	0.0571

5 Conclusions

In this paper we propose the use of a nonlinear Granger causality measure, the partial transfer entropy on rank vectors (PTERV), as an alternative to the standard linear procedure (VAR / VECM) for the investigation of directed inter-relationships among financial variables. Therefore, we present and compare the two procedures in the aim to detect causality in non-stationary time series. Both are evaluated for different simulation systems and their performance is assessed based on the percentages of rejection of the non-causality hypothesis from different realizations.

The PTERV can be applied to any type of data and no assumptions need to be made regarding the nature of the involved time series (e.g. their volatility) prior to its estimation. This means that it can be directly employed to non-stationary series. In addition to that, it is not affected by any monotonic transformation of the data as well as by the existence of co-integration, because it is model-free. We should note that the PTERV, as the majority of the Granger causality measures, requires the separability assumption, i.e. the assumption that cause and effect are separable. Further, the PTERV has not been designed for systems with time dependent causal relations, however, such cases can potentially be faced by estimating the PTERV on relatively short rolling windows, tracing in this way possible changes of the causality relations across time.

The PTERV outperforms the linear approach (VAR or VECM) in the case of nonlinear causal effects and tends to avoid possible spurious causalities stemming from first-differencing. Large samples are required for the PTERV; its performance improves with the time series length and with relatively low embedding dimensions. Additionally, it is a convenient alternative procedure to the VECM methodology in cases where the VECM is poorly fitted to the data. However, the PTERV indicates both long- and short-run causality. We leave for future work the combination of PTERV with some non-parametric/nonlinear co-integration test in order to be able to extract information for the existence of long-run causality (Breitung, 2001; Choi and Saikkonen, 2010).

On the other hand, the standard linear procedure consists of three cases arising from the outcome of stationarity and co-integration tests. When the data are not co-integrated, Granger causality tests, such as CGCI, can be applied (Case 1 and Case 2). The CGCI has been vastly used in the literature and found very effective in different applications. However, it gives poor results in the case of nonlinear structures. When data are non-stationary and co-integrated, the VECM is utilized (Case 3). Its main advantage is that we can discriminate between long- and short-run causality, while it is efficient even for small time series lengths. Nevertheless, it fails to capture the true couplings in the presence of nonlinearity. The validity of the results depends on whether the fitted model is well-specified.

To see whether the conclusions from the simulation experiment hold, both procedures have been applied to real data. In the first application to national capitalization indices, both methods highlight that the US drives Germany and Greece. In addition, the VECM indicates the causal link Germany $\rightarrow$ Greece, while the PTERV unveils a bidirectional interaction between Greece and Germany supporting empirical evidence about the strong connectivity among the European stock markets.

In the second application where interest rates are examined, the detection of bidirectional couplings, via the linear procedure, leads to an over-simplistic view of money and capital markets, since it is not able to take into account the impact of the financial crisis of 2007-2009 on the traditional interest rate channel. In contrast, the PTERV accurately describes the effect of the 3MTB and the 10YTN on the FF as well as the feedback mechanism between the 3MTB and 10YTN. According to Kyrtsou et al. (2014) and Papana et al. (2016), this kind of connectivity can be explained by the stance of monetary policy and the behavioral aspects of investors' expectations. In the same line, Guo et al. (2011) find that in the aftermath of the financial crisis, a reversal of the causal relationship between bond yields is observed, conditional to the stock market signals, so that long-term yields cause shorter-term yields and then the US policy rate.

Acknowledgements 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.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Allali, A., Oueslati, A., and Trabelsi, A. (2011). Detection of information flow in major international financial markets by interactivity network analysis. *Asia-Pacific Financial Markets*, 18(3):319–344.

Anderson, H. M. (1997). Transaction costs and non-linear adjustment towards equilibrium in the US treasury bill market. *Oxf. Bull. Econ. Stat.*, 59(4):465–484.

Aparicio, F. M. and Escribano, A. (1998). Information-theoretic analysis of serial dependence and cointegration. *Stud. Nonlinear Dyn. Econ.*, 3(3):119–140.

Baccala, L. A. and Sameshima, K. (2001). Partial directed coherence: A new concept in neural structure determination. *Biol. Cybern.*, 84(6):463–474.

Bai, Z., Wong, W.-K., and Zhang, B. (2010). Multivariate linear and nonlinear causality tests. *Math Comput Simul*, 81(1):5–17.

Barnett, L., Barrett, A., and Seth, A. (2009). Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys. Rev. Lett.*, 103(23):238701.

Brandt, P. T. and Williams, J. T. (2007). *Multiple Time Series Models, Volume 2*, pages 32–34. Sage Publications.

Breitung, J. (2001). Rank tests for nonlinear cointegration. *J. Bus. Econ. Stat.*, 19(3):331–340.

Brock, W. A., Hsieh, D. A., and LeBaron, B. (1991). *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence.* MIT Press, Cambridge, Mass.

Choi, I. and Saikkonen, P. (2010). Tests for nonlinear cointegration. *Econometric Theory*, 26(3):682–709.

Clarke, J. A. and Mirza, S. (2006). A comparison of some common methods for detecting Granger noncausality. *J. Stat. Comput. Simul.*, 76(3):207–231.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Assoc.*, 74(366):427–431.

Diks, C. and Panchenko, V. (2006). A new statistic and practical guidelines for nonparametric Granger causality testing. *J. Econ. Dyn. Control*, 30(9-10):1647–1669.

Ding, S. L., Chen, M., and Bressler, Y. (2006). Granger causality: Basic theory and application to neuroscience. Schelter B., Winterhalder M., Timmer J. (edit.), Handbook of time series analysis.

Dufour, J.-M. and Taamouti, A. (2010). Short and long run causality measures: Theory and inference. *J. Econom.*, 154(1):42–58.

Fang, Z. and Wolski, M. (2019). Human capital, energy and economic growth in China: Evidence from multivariate nonlinear Granger causality tests. *Empir. Econ.*, pages 1–26.

Franke, J. (2008). *Non-parametric Concepts for Financial Time Series*, pages 279–302. Springer Berlin Heidelberg, Berlin, Heidelberg.

Geweke, J. (1982). Measurement of linear dependence and feedback between multiple time series. *J. Am. Stat. Assoc.*, 77(378):304–313.

Geweke, J., Meese, R., and Dent, W. T. (1983). Comparing alternative tests of causality in temporal systems: analytic results and experimental evidence. *J. Econom.*, 21:161–194.

Goebel, B., Dawy, Z., Hagenauer, J., and Mueller, J. C. (2005). An approximation to the distribution of finite sample size mutual information estimates. *IEEE International Conference on Communications*, 2:1102–1106.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:24–36.

Granger, C. W. J. (1988). Some recent developments in a concept of causality. *J. Econom.*, 39:199–211.

Granger, C. W. J. (1993). Modelling non-linear relationships between long-memory variables. Technical report, University of California, San Diego.

Granger, C. W. J. and Newbold, P. (1974). Spurious regressions in econometrics. *J. Econom.*, 2:110–120.

Guo, K., Zhou, W.-X., Cheng, S.-W., and Sornette, D. (2011). The US stock market leads the federal funds rate and treasury bond yields. *PloS One*, 6(8):e22794.

Hiemstra, C. and Jones, J. D. (1994). Testing for linear and nonlinear Granger causality in the stock price - volume relation. *J. Finance*, 49:1639–1664.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vector in Gaussian vector autoregressive models. *Econometrica*, 59(6):1551–1580.

Jones, C. S. (2006). A nonlinear factor analysis of S&P 500 index option returns. *J. Finance*, 61(5):2325 – 2363.

Kraskov, A., Stogbauer, H., and Grassberger, P. (2004). Estimating mutual information. *Phys. Rev. E*, 69(6):066138.

Kugiumtzis, D. (2012). Transfer entropy on rank vectors. *J. Nonlinear Sci. Appl.*, 3(2):73–81.

Kugiumtzis, D. (2013). Partial transfer entropy on rank vectors. *Eur. Phys. J. Spec. Top.*, 222(2):401–420.

Kwon, O. and Yang, J. S. (2008). Information flow between stock indices. *EPL*, 82(6):68003.

Kyrtsou, C., Malliaris, A., and Mikropoulou, C. (2014). Dynamic analysis of the US short- and long-term interest rates relationships: on the role of monetary policy. In *1st Conference of the Society for Economic Measurement, University of Chicago's Booth School of Business*.

LeSage, J. P. (1990). A comparison of the forecasting ability of ECM and VAR models. *Rev. Econ. Stat.*, 72(4):664–671.

Lütkepohl, H. and Xu, F. (2011). Forecasting annual inflation with seasonal monthly data: Using levels versus logs of the underlying price index. *J. Time Ser. Econ.*, 3(1):article 7.

Maki, D. (2003). Nonparametric cointegration analysis of the nominal interest rate and expected ination rate. *Econ. Lett.*, 81:349–354.

Marschinski, R. and Kantz, H. (2002). Analysing the information flow between financial time series - an improved estimator for transfer entropy. *Eur. Phys. J. B*, 30:275–281.

Matilla-Garcia, M., Marian, M. R., and Dore, M. I. (2014). A permutation entropy based test for causality: The volume-stock price relation. *Physica A*, 398:280–288.

Paluš, M. (2007). From nonlinearity to causality: Statistical testing and inference of physical mechanisms underlying complex dynamics. *Contemp. Phys.*, 48(6):307–348.

Papana, A., Kyrtsou, C., Kugiumtzis, D., and Diks, C. (2016). Detecting causality in non-stationary time series using partial symbolic transfer entropy: Evidence in financial data. *Comput. Econ.*, pages 1–25.

Papapetrou, M. and Kugiumtzis, D. (2014). Investigating long range correlation in DNA sequences using significance tests of conditional mutual information. *Comp. Biol. Chem.*, 53:32–42.

Porta, A. and Faes, L. (2016). Wiener-Granger causality in network physiology with applications to cardiovascular control and neuroscience. *Proceedings of the IEEE*, 104(2):282–309.

Quiroga, Q. R., Kraskov, A., Kreuz, T., and Grassberger, P. (2002). Performance of different synchronization measures in real data: A case study on electroencephalographic signals. *Phys. Rev. E*, 65:041903.

Schiff, S. J., So, P., Chang, T., Burke, R. E., and Sauer, T. (1996). Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble. *Phys. Rev. E*, 54:6708–6724.

Schreiber, T. (2000). Measuring information transfer. *Phys. Rev. Lett.*, 85(2):461–464.

Sharp, G. D. (2010). *Lag Length Selection for Vector Error Correction Models*. PhD thesis, Rhodes University.

Shi, W., Shang, P., and Lin, A. (2015). The coupling analysis of stock market indices based on cross-permutation entropy. *Nonlinear Dyn.*, 79(4):2439–2447.

Staniek, M. and Lehnertz, K. (2008). Symbolic transfer entropy. *Phys. Rev. Lett.*

Stern, D. I. and Kaufmann, R. K. (2014). Anthropogenic and natural causes of climate change. *Clim. Change*, 122(1-2):257–269.

Tiwari, A. K. and Mutascu, M. (2014). A revisit on the tax burden distribution and GDP growth: fresh evidence using a consistent nonparametric test for causality for the USA. *Empir. Econ.*, 46(3):961–972.

Tsay, R. S. (2005). *Analysis of Financial Time Series, 2nd Edition.* John Wiley and Sons, INC.

Yu, G. H. and Huang, C. C. (2001). A distribution free plotting position. *Stoch. Environ. Res. Risk Assess.*, 15(6):462–476.

Zhao, Z. (2008). Parametric and nonparametric models and methods in financial econometrics. *Statistics Surveys*, 2:1–42.