Identifying causal relationships in case of non-stationary time series

Angeliki Papana[*]
Department of Economics, University of Macedonia, Greece
apapana@uom.edu.gr, angeliki.papana@gmail.com

Catherine Kyrtsou
Department of Economics, University of Macedonia, Greece;
University of Strasbourg, BETA; University of Paris 10, France
ckyrtsou@uom.gr

Dimitris Kugiumtzis
Faculty of Engineering, Aristotle University of Thessaloniki, Greece
dkugiu@gen.auth.gr

Cees Diks
CeNDEF, University of Amsterdam, The Netherlands
C.G.H.Diks@uva.nl

May 13, 2014

Abstract

The standard linear Granger non-causality test is effective only when time series are stationary. In case of non-stationary data, a vector autoregressive model (VAR) in first differences should be used instead. However, if the examined time series are co-integrated, a VAR in first differences will also fail to capture the long-run relationships. The vector error-correction model (VECM) has been introduced

[*]Corresponding author

to correct a disequilibrium that may shock the whole system. The VECM accounts for both short run and long run relationships, since it is fit to the first differences of the non-stationary variables, and a lagged error-correction term is also included. An alternative approach of estimating causality when time series are non-stationary, is to use a non-parametric information-based measure, such as the transfer entropy on rank vectors (TERV) and its multivariate extension partial TERV (PTERV). The two approaches, namely the VECM and the TERV / PTERV, are evaluated on simulated and real data. The advantage of the TERV / PTERV is that it can be applied directly to the non-stationary data, whereas no integration / co-integration test is required in advance. On the other hand, the VECM can discriminate between short run and long run causality.

1 Introduction

The concept of Granger causality [10] has been widely utilized for the investigation of the directed interactions, mainly in economics, e.g. [9], but also in various other fields, e.g. neuroscience [29, 6], climatology [19]. The basic principle of Granger causality analysis is to test whether past values of a variable X (the driving variable) help to explain current values of another variable Y (the response variable). The linear Granger causality test is implemented by fitting autoregressive models, whereas one tests if the prediction of Y could be improved by incorporating information of X (compared to the prediction of Y using only past values of Y). However, one should first examine the basic properties of the variables. If the variables are non-stationary or / and co-integrated, then the test will be misspecified [11].

Stationarity is a common assumption in many time series techniques. A stationary process is a stochastic process whose joint probability distribution does not change when shifted in time (strict stationarity). Weak stationarity (or covariance stationarity) is usually sufficient for most techniques, requiring only the first and second moment not vary with respect to time. The order of integration, denoted as $I(d)$, of a time series states the minimum number of differences required to obtain a covariance stationary series. Various tests for stationarity or equivalently unit root tests have been developed in order to the test the stationarity of a time series, e.g. the (augmented) Dickey-Fuller test [5] and the Phillips-Perron test [22].

Co-integration is a concept that stems from the economic theory which often suggests that certain pairs of economic or financial variables are linked by a long-run relationship [7]. Two or more time series are co-integrated if they share a common stochastic drift, i.e. a certain linear combination of the variables is $I(0)$. There are three main co-integration tests, namely the EngleGranger two-step method [7], the Johansen test [14] and the PhillipsOuliaris test [21].

The cointegration technique made a significant contribution towards testing Granger causality. Although co-integration does not provide any information about the direction of causality, if two variables are cointegrated, there should be causality in at least one direction [11]. To this respect, a cointegration test can be viewed as an indirect test of long-run dependence [7]. Causality in non-stationary time series (in mean) is typically investigated through vector error correction models (VECM) in econometrics, and it is subdivided into short-run and long-run causality. As reported in [12], the regression results with non-stationary variables will be spurious. Further, if the variables are non-stationary and cointegrated, running a regression with first differenced variables will not capture the long run information as the first differenced regression results are for short run relationship. A comparison of the prediction performance of VAR models and VECMs can be found in [17].

The developments in the area of nonlinear dynamics led to the contribution of other fields, such as statistics and physics on econometrics and vice versa. Linear and nonlinear extensions of the Granger causality concept (e.g. [13, 24]) have been utilized in different scientific fields such as the analysis of brain dynamics, while causality measures originally defined for the analysis of biological signals have been applied in financial data, e.g. partial directed coherence [4] has been used to detect the information flow among financial markets in [2].

Information theory is essential for the analysis of information flow among variables of complex systems. Measures from information theory, such as Shannon entropy, (conditional) mutual information [27], mutual coarse-grained information rate [20] and transfer entropy [25], have been extensively used for the detection of the general statistical dependence and the information flow among variables. Their main advantage is that they are model free and make no assumption for the distribution of the data, while are able to detect the overall dependencies and not only the linear ones. The transfer entropy is the most commonly used information causality measure and has been applied also to financial data, e.g. see [18]. Information theory has also been employed for the investigation of non-linear co-integration, e.g. [3].

The transfer entropy on rank vectors (TERV) is an information bivariate causality measure that has recently been introduced for the detection of directed interrelationships [15]. The partial TERV (PTERV) is the extension of TERV in the multivariate case, in order to account only for the direct couplings [16]. The TERV / PTERV is applicable to any type of data (stationary or not) and makes no assumptions prior to the distribution of the data.

In this work, we demonstrate the performance of the TERV / PTERV on bivariate and multivariate time series of known coupled and uncoupled systems and on stationary and non-stationary time series in mean. Further, we compare the performance of TERV / PTERV with the ECM / VECM using different bivariate and multivariate simulation systems, with $I(1)$ co-integrated variables. From the simulation study, the advantages and disadvantages of each method are indicated. As a real application, the TERV / PTERV and the VECMs are applied on financial variables in order to explore the causal affects.

In Sec. 2, some introductory concepts are discussed, i.e. the stationarity of a process and the augmented Dickey-Fuller test for unit root, the Granger causality test for stationary processes, and the notion of co-integration along with the Johansen co-integration test. The causality test based on the (vector) error correction models (ECM / VECM) and the (partial) transfer entropy on rank vectors (TERV / PTERV) are presented in Sec. 3. In Sec. 4, the two causality methods are evaluated on a simulation study. The performance of the causality test is assessed in a real application from financial time series in Sec. 5. Finally, the conclusions of this study are discussed in Sec. 6.

2 Materials and methods

The concepts of stationarity and co-integration are discussed here, along with the corresponding tests in order to investigate these properties for a given time series.

2.1 Stationarity

A process is stationary if its statistical properties are constant over time. Weak stationarity implies that the first and second moment of the process is constant over time, i.e. the mean, variance and autocovariance. The stationarity is an important feature of the time series since it is required in order to apply certain statistical tests. If two time series x and y are non-stationary (or equivalently integrated or have a unit root), then modelling the x and y relationship as a simple ordinary least square relationship of the form $y_t = \alpha + \beta x_t + e_t$ will only generate a spurious regression.

If a series is stationary without any differencing it is designated as $I(0)$ (integrated of order 0). The most common technique to transform a non-stationary time series to a stationary one is by differencing, e.g. the first-differenced values of a time series x_t are given by the $\Delta x_t = x_t - x_{t-1}$, where Δ is the differencing operator. We discuss here one of the most popular unit root test, the augmented Dickey-Fuller test [5].

Augmented Dickey-Fuller test for unit root The Augmented Dickey-Fuller test for a unit root assesses tests the null hypothesis of a unit root, $H_0 : \phi = 1$ (i.e. the data needs to be differenced to make it stationary), under the alternative hypothesis, $H_1 : \phi < 1$. To implement the test, the following model is considered

$$x_t = c + \delta t + \phi x_{t-1} + A_1 \Delta x_{t-1} + \ldots + A_p \Delta x_{t-p} + \epsilon_t \tag{1}$$

where Δ is the differencing operator, c is the drift coefficient, δ is the deterministic trend coefficient, p is the number of lagged difference terms used in the model and ϵ_t is a mean zero innovation process. The number of lags can be determined using the Schwartz Bayesian information criterion [26] or the Akaike information criterion [1].

To infer about H_0, the t-statistic is used on the ϕ coefficient. The test statistic is given by the expression

$$t_{DF} = \frac{\hat{\phi} - 1}{SE(\hat{\phi})}, \tag{2}$$

where $\hat{\phi}$ is the estimated ϕ from the fitting and SE is the standard error.

The Dickey-Fuller statistics follow nonstandard distributions under the null hypothesis (even asymptotically). Critical values for a range of sample sizes and significance levels can be calculated using Monte Carlo simulations of the null model with Gaussian innovations. The null hypothesis H_0 is rejected if the test statistic is smaller than the corresponding critical value.

2.2 Granger causality

The Granger causality has been developed in order to quantify the causal effects among time series [10] and it is based on the concept that the cause occurs prior to its effect. The application of Granger causality assumes that the analyzed time series are covariance stationary. Formally, if a future value of a time series X_1 that one would like to predict is improved by using the values of X_2 (instead of using only past values of X_1), then we say that X_2 Granger causes X_1 or equivalently that X_2 is driving X_1.

To estimate the linear Granger causality (in the bivariate case), a vector autoregressive model (VAR) in two variables and of order P is fitted to the time series $\{x_{1,t}\}$ (unrestricted model)

$$x_{1,t+1} = \sum_{j=0}^{P-1} a_{1,j} x_{1,t-j} + \sum_{j=0}^{P-1} b_{1,j} x_{2,t-j} + \epsilon_{1,t+1}, \tag{3}$$

where $a_{1,j}, b_{1,j}$ are the coefficients of the model and ϵ_1 the residuals from fitting the model with variance s_{1U}^2. We also consider the VAR model based on Eq. 3 obtained by omitting the terms regarding the driving variable (restricted model). The variable X_2 Granger causes X_1 if the residuals s_{1R}^2 of the restricted model are significantly larger than the residuals s_{1U}^2 of the unrestricted model, i.e. if $s_{1R}^2 > s_{1U}^2$.

To infer about the causal effects, a parametric significance test can be conducted for the null hypothesis that variable X_2 is not driving X_1 making use of the F-significance test for all P coefficients $b_{2,j}$ [**?**]. The Granger causality can be easily extended to the multivariate case by fitting all the available data to the VAR models.

2.3 Co-integration

Consider an K-dimensional time series $\mathbf{y}_t$, with components that have a unit root. This process is co-integrated if a linear combination $\beta_1 y_{1t} + \ldots + \beta_K y_{Kt}$ of the components of the process is stationary (co-integrating relation). The vector $\beta = (\beta_1 + \ldots + \beta_K)'$ is called the co-integrating vector. The co-integration should be distinguished from the short-term relations.

The extension of the VAR model to include co-integrated variables balances the short-term dynamics of a process with the long-term tendencies. The vector error correction model (VECM) expresses the long run dynamics of the process including error correction terms i.e. the term $\alpha\beta' y_{t-1}$ measures the deviation from the stationary mean at time $t-1$:

$$\Delta y_t = c + \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \epsilon_t \quad (4)$$

where $\Pi = \alpha\beta'$ and c is the drift coefficient. Considering Eq. 1, it holds that $\Pi = \sum_{i=1}^{p} A_i - I$ and $\Gamma_i = -\sum_{j=i+1}^{p} A_j$. If the variables in $\mathbf{y}_t$ are $I(1)$, the terms involving differences are stationary, while the error-correction term in the VEC model introduces long-term stochastic trends.

Johansen co-integration test The Johansen co-integration test is one of the most methods for the identification of the existence of co-integration. Its main advantage is that provides comprehensive testing in the presence of multiple co-integrating relations.

The Johansen method is based on the relationship between the rank of the matrix Π and the size of its eigenvalues. The rank of the matrix Π determines the long-term dynamics. If Π has full rank, the process $\mathbf{y}_t$ is stationary in mean. If the rank of Π is zero, then the error-correction term disappears, and the system is stationary in differences (the VAR model in differences can be used). If the rank of Π is r (within $(0, K)$), then there are r independent co-integrating relations among the variables in $\mathbf{y}_t$. For a given r, the maximum likelihood estimator of β defines the combination of y_{t-1} that yields the r largest canonical correlations of $\Delta\mathbf{y}_t$ with y_{t-1}.

Assuming that the VECM errors are independent and Gaussian distributed, and given the co-integrating restrictions on the trend or the parameters of the model, the maximum likelihood $L_{max}(r)$ is a function of the co-integration rank r. In the Johansen testing there are two test statistics; the trace statistics and the maximum eigenvalue statistic. The trace statistic tests the null hypothesis that there are at most r co-integrating relationships against the alternative of K co-integrating relations. The test statistics is given as

$$J_{trace} = -n \sum_{i=r+1}^{K} \ln(1 - \hat{\lambda}_i), \tag{5}$$

where $\hat{\lambda}_i$ is the i-largest canonical correlation and n is the sample size. The maximum eigenvalue test considers the null hypothesis that r co-integrating vectors exist, with the alternative that $r+1$ co-integrating vectors exist. The test statistics is given as

$$J_{max} = -n \ln(1 - \hat{\lambda}_{r+1}). \tag{6}$$

Both tests have non-standard asymptotic null distributions, there simulations are used for their critical values.

3 Causality

The linear Granger causality on VAR can be applied to time series that are stationary. If data are not stationary and not co-integrated, then the VAR can fitted to the differenced time series. If data are non-stationary and co-integrated, then the VAR model will give miscellaneous results. If co-integration is detected among the examined time series, then a long-term equilibrium relationships among them exists and one should apply the VECM in order to evaluate the short run properties of the co-integrated time series.

3.1 Causality test based on VECM

The VECMs incorporate information about the short run and long-run relationships of the variables. The assumptions to use the VECM are that the variables are integrated of the same order and a co-integration relationship exists. The regression equation form for the VECM has been given in Eq. 4. An interaction is significant in the long-run, if the coefficient Π of the model is negative and significant. An interaction is Granger significant in the short-run, if the corresponding coefficients Γ_i of the model are significant.

The VECM is efficiently specified when the residuals are normally distributed, exhibit no serial correlation and no heteroscedasticity. The examined null hypothesis is that the residuals are normally distributed. For this, the Jarque-Bera statistic is used

$$JB = \frac{n}{6}(S^2 + \frac{(k-3)^2}{4}) \sim \chi^2(2), \tag{7}$$

where n is the sample size, S is the skewness and k is the kurtosis of the residuals. If H_0 is satisfied, JB follows asymptotically a X^2 distribution with 2 degrees of freedom. H_0 is rejected for large values of JB, i.e. for small p-values of the test.

The Ljung-Box Q-test for residual autocorrelation tests the null hypothesis that the residuals exhibit no autocorrelations for a fixed number of lags L. The test statistic is defined as

$$Q = n(n+2)\sum_{\tau=1}^{L}\frac{\rho(\tau)^2}{n-\tau}) \sim \chi^2(L), \tag{8}$$

where $\rho(\tau)$ is the sample autocorrelation at lag τ.

Finally, we test the hypothesis that there are no auto regressive conditional heteroskedasticity (ARCH) effects in the residuals, using the Engle test. The ARCH(h) model is fitted to the residuals

$$r_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \ldots + \alpha_h r_{t-h}^2 + \epsilon_t. \tag{9}$$

The Lagrange multiplier statistic is employed

$$LM = nR^2 \sim X^2(L), \tag{10}$$

where R^2 is the coefficient of determination, to test the hypothesis that the coefficients of the ARCH model are all equal $\alpha_0 = \alpha_1 = \ldots = \alpha_h$.

3.2 Transfer entropy on rank vectors

Information theory has provided many efficient and effective correlation and causality measures that have been tested in different applications. Their advantage is that they are model free and make no assumptions on the distribution of the data. Further, the causality measures symbolic transfer entropy (STE) [28] and (partial) transfer entropy on rank vectors (TERV) [15, 16], have been recently developed, so that they do not require stationarity and therefre are applicable to any type of data.

The transfer entropy (TE) quantifies the amount of information explained in a variable X_1 at h time steps ahead from the state of a variable X_2 accounting for the concurrent state of X_1. Let us consider first two simultaneously observed time series $\{x_{1,t}\}$, $\{x_{2,t}\}$, $t = 1, \ldots, n$ derived from the dynamical systems X_1 and X_2, respectively. The reconstructed vectors for X_1 and X_2 are defined as $\mathbf{x}_{1,t} = (x_{1,t}, x_{1,t-\tau_1}, \ldots, x_{1,t-(m_1-1)\tau_1})'$,

$\mathbf{x}_{2,t} = (x_{2,t}, x_{2,t-\tau_2}, \dots, x_{2,t-(m_2-1)\tau_2})'$, where $t = 1, \dots, n'$, $n' = n - h - \max\{(m_1 - 1)\tau_1, (m_2 - 1)\tau_2\}$, m_1 and m_2 are the embedding dimensions, τ_1 and τ_2 are the time delays and h is the step ahead to address for the interaction. The TE is expressed as a conditional mutual information

$$\begin{aligned} \mathrm{TE}_{X_2 \to X_1} &= I(x_{1,t+h}; \mathbf{x}_{2,t} | \mathbf{x}_{1,t}) \\ &= \sum p(x_{1,t+h}, \mathbf{x}_{2,t}, \mathbf{x}_{1,t}) \log \frac{p(x_{1,t+h} | \mathbf{x}_{2,t}, \mathbf{x}_{1,t})}{p(x_{1,t+h} | \mathbf{x}_{1,t})}, \end{aligned} \tag{11}$$

where $p(x_{1,t+h}, \mathbf{x}_{2,t}, \mathbf{x}_{1,t})$, $p(x_{1,t+h}|\mathbf{x}_{2,t}, \mathbf{x}_{1,t})$ and $p(x_{1,t+h}|\mathbf{x}_{1,t})$ are the joint and conditional probability distributions.

The symbolic transfer entropy (STE) is an extension of TE defined on rank points. For each reconstructed vector, e.g. $\mathbf{x}_{1t}$, we form the corresponding rank-points $\hat{\mathbf{x}}_{1t} = (r_1, \dots, r_m)'$, where $r_j \in \{1, \dots, m\}$ and $j = 1, \dots, m$, by arranging the amplitude values of $\mathbf{x}_{1t}$ in an ascending order. The STE is defined similarly to TE, however the rank points are used in its definition instead of using the reconstructed vectors.

The transfer entropy on rank vectors (TERV) is a correction of the STE, so that the future of the response is defined for more than one time steps and therefore capture the information flow from the driving system over a longer time horizon. For its definition, assume the vector $\hat{\mathbf{x}}_{1,t}^h = (x_{1,t+1}, \dots, x_{1,t+h})'$ instead of using $\hat{x}_{1,t+h}$. In [15], the results show that TERV improves the performance of STE and at some cases performs better than TE, e.g. in the presence of noise. The partial TERV (PTERV) is the extension of TERV in the multivariate case, in order to account for the presence of confounding variables $Z = \{Z_1, \dots, Z_K\}$

$$\mathrm{PTERV}_{X_2 \to X_1 | Z} = \sum p(\hat{\mathbf{x}}_{1,t}^h, \hat{\mathbf{x}}_{2,t}, \hat{\mathbf{x}}_{1,t}, \hat{\mathbf{z}}_t) \log \frac{p(\hat{\mathbf{x}}_{1,t}^h | \hat{\mathbf{x}}_{2,t}, \hat{\mathbf{x}}_{1,t}, \hat{\mathbf{z}}_t)}{p(p(\hat{\mathbf{x}}_{1,t}^h | \hat{\mathbf{x}}_{1,t}, \hat{\mathbf{z}}_t)}. \tag{12}$$

The PTERV has also been tested in case of non-stationary data and proved to be robust in the presence of drifts in the time series [16].

Statistical significance of TERV / PTERV In [16], the parametric approximation of the null distribution H_0 of no coupling for PTERV is discussed and it is concluded that resampling methods are required to sufficiently approximate it. The statistical significance of the PTERV is assessed by a randomization test making use of time-shifted surrogates [23]. The surrogate time series are formed by time-shifting the time series of the driving variable, while the other time series are intact.

Considering the driving time series $\{x_{1,1}, \ldots, x_{1,n}\}$ and a random integer d ($d < n$), the first d values of the time series are moved to the end, so that the time-shifted time series is $\{x_{1,d+1}, \ldots, x_{1,n}, x_{1,1}, \ldots, x_{1,d}\}$. If the original TERV / PTERV value (q_0) lies at the tail of the distribution of the TERV / PTERV values ($q_1, \ldots, q_M$) using the time-shifted time series, then the null hypothesis of no causal effects 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 $p-$value $= \frac{1-(r_0-0.326)}{M+1+0.348}$ [30].

4 Simulations and Results

The effectiveness of the two causality methods, i.e. ECM / VECM and TERV / PTERV, in detecting direct causal effects is assessed based on a simulation study. First, the TERV / PTERV is applied on different types of time series, bivariate and multivariate, stationary or non-stationary, with with linear and/or nonlinear couplings. Results from TERV / PTERV are compared with those from the linear Granger causality test. Then, non-stationary and co-integrated time series are generated, so that ECM / VECMs can be also applied. Results from both methods are compared.

We consider 100 realizations of different simulation systems for time series lengths $n = 256, 512, 1024, 2048$. For the TERV / PTERV, the time lag τ for all variables is set to one and at cases to $\tau = m$. The embedding dimension m is set to be the same for all variables and for each simulation system is set equal the true model order from the equations of each system. The number of time steps ahead h is set to 1. The order P of the VAR model or the ECM / VECM is equal to m.

The performance of the causality methods is quantified by the percentage of statistically significant couplings in the 100 realizations. For all methods, the couplings are always regarded to be conditioned on the remaining variables, if the system is multivariate.

System 1. A stationary, linear, uncoupled, bivariate model

$$x_{1,t} = 0.1x_{1,t-1} - 0.3x_{1,t-2} + \epsilon_{1,t}$$

$$x_{2,t} = 0.7x_{2,t-1} + \epsilon_{2,t}$$

where $\epsilon_{i,t}$ is Gaussian white noise (the same stands for all the systems).

The TERV successfully indicates that no couplings exists for $h = 1$, $m = 2$, $\tau = 1$, giving low percentages of rejection of H_0 (Table 1). Since the assumptions for the ECM are not satisfied, the standard linear Granger causality (GC) is used for system 1. Both methods give similar results (Table 1).

Table 1: Percentages of rejecting the non-causality hypothesis based on TERV ($m = 2$) and on GC (P=2), for the simulation system 1.

	TERV ($m = 2$)				GC ($P = 2$)			
n	256	512	1024	2048	256	512	1024	2048
$X_1 \to X_2$	6	8	6	6	7	4	2	5
$X_2 \to X_1$	3	3	8	3	5	3	5	4

In order to further check the performance of the TERV in case of non-stationarity, we generate the corresponding non-stationary time series from system 1 by integrating the original time series. The TERV gives similar results as for the stationary case, indicating low percentages of rejections of the non-causality hypothesis (varying from 2% to 7% for the different n).

System 2. A non-stationary, linear, bivariate model with bidirectional couplings $X_1 \leftrightarrow X_2$, generated by integrating the variables from the following system

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

We are considering system 2 for two reasons; first, in order to evaluate the performance of TERV when bidirectional coupling is present, and secondly to investigate the effect of mis-specifying the embedding dimension m. From the definition of TERV, we can only set $m \geq 2$, however the true order model of system 2 is one. The TERV (for $m = 2$) seems to be effective only for large time series lengths (see Table 2).

The GC is applied to the original stationary time series from system 2. It effectively denotes the true couplings for all time series lengths, giving similar results for $P = 1$ and $P = 2$ (Table 2).

Table 2: As Table 1 but for system 2.

	TERV ($m = 2$)					GC ($P = 2$)			
n	256	512	1024	2048	4096	256	512	1024	2048
$X_1 \to X_2$	13	16	31	58	86	97	99	100	100
$X_2 \to X_1$	0	3	9	14	41	99	100	100	100

System 3. A non-stationary, linear, bivariate model with unidirectional coupling $X_2 \leftrightarrow X_1$, generated by integrating the variables of the system

$$\begin{aligned} x_{1,t} &= 0.6x_{1,t-1} + 0.3x_{2,t-2} + \epsilon_{1,t} \\ x_{2,t} &= 0.1x_{2,t-2} + \epsilon_{2,t} \end{aligned}$$

For System 3, the embedding dimension is set equal to the true order of the model, i.e. $m = 2$. Although the TERV fails to detect the coupling for $h = 1$ (percentages of rejection of H_0 vary from 3 to 8% for all n), it correctly detects the coupling $X_2 \to X_1$ for $h = 2$ (see Table 3). The sensitivity of TERV on the free parameter h is displayed in this example.

Table 3: As Table 1 but for system 3.

	TERV ($h = 2$, $m = 2$)				GC ($P = 2$)			
n	256	512	1024	2048	256	512	1024	2048
$X_1 \to X_2$	5	3	10	3	5	2	6	5
$X_2 \to X_1$	29	54	92	100	100	100	100	100

On the other hand, the GC correctly indicates the true coupling even for small time series lengths, when the original stationary system is considered (Table 3). However, if P is mis-specified and is set to 1, then the GC fails to come up with the correct coupling and specifically no couplings are denoted.

System 4. A non-stationary, nonlinear, multivariate system, with unidirectional couplings $X_1 \to X_2$ and $X_2 \to X_3$, generated by integrating the variables from three coupled Hénon maps, with coupling strength $c = 0.2$

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

The PTERV ($m = 2$) correctly indicates the true direct couplings for system 4, even for small time series lengths, while the percent of significant values are low for the uncoupled directions (see Table 4).

Table 4: As Table 1 but for system 4. All the indicated couplings are conditioned on the third variable of the system.

	TERV ($m = 2$)				GC ($P = 2$)			
n	256	512	1024	2048	256	512	1024	2048
$X_1 \to X_2$	54	81	99	100	46	91	100	100
$X_2 \to X_1$	6	5	4	9	7	5	8	32
$X_2 \to X_3$	54	91	99	100	22	60	93	100
$X_3 \to X_2$	10	7	9	16	20	10	16	18
$X_1 \to X_3$	9	11	10	9	6	8	9	7
$X_3 \to X_1$	5	10	3	9	10	12	4	6

The GC is again applied to the corresponding stationary time series of system 4. The GC ($P = 2$) correctly indicates the couplings, however also spurious couplings 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 4).

System 5. A non-stationary, linear, multivariate system, with unidirectional couplings $X_1 \to X_3$, $X_2 \to X_1$, $X_2 \to X_3$, $X_4 \to X_2$, generated by integrating the variables of a VAR(5) model

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

The PTERV ($m = 5$) correctly indicates only the true coupling $X_1 \to X_3$, however fails to detect the other ones. No spurious couplings are detected (see Table 5). The PTERV seems to be ineffective for large embedding dimensions probably due to the high dimensionality of the vectors, i.e. m^*K + 1, where K is the number of variables.

In this case, the GC ($P = 5$) outperforms the PTERV, indicating the true couplings for all n, while no spurious couplings are observed (Table 5). Again, the stationary time series are considered for the GC.

Table 5: As Table 4 but for system 5.

	TERV ($m = 5$)				GC ($P = 5$)			
n	256	512	1024	2048	256	512	1024	2048
$X_1 \to X_2$	0	0	0	0	5	4	8	8
$X_2 \to X_1$	5	10	12	7	100	100	100	100
$X_1 \to X_3$	21	53	97	100	100	100	100	100
$X_3 \to X_1$	2	1	0	0	4	4	6	4
$X_1 \to X_4$	0	0	0	0	7	3	3	2
$X_4 \to X_1$	8	6	4	1	5	3	4	7
$X_2 \to X_3$	2	3	1	0	100	100	100	100
$X_3 \to X_2$	0	0	0	0	4	3	4	6
$X_2 \to X_4$	1	2	0	0	7	8	7	8
$X_4 \to X_2$	1	10	5	12	100	100	100	100
$X_3 \to X_4$	0	0	0	0	3	8	4	4
$X_4 \to X_3$	6	3	0	1	4	5	6	7

The following systems are generated in order to be non-stationary ($I(1)$) and co-integrated, in order to apply the ECM / VECM methodology.

System 6. A non-stationary, linear, bivariate system, with unidirectional coupling $X_2 \to X_1$, with co-integrated variables

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

System 6 is co-integrated since variable X_2 is non-stationary ($X_2 \sim I(1)$) and therefore $X_1 \sim I(1)$ as well, and there is a linear combination of them which is stationary: $x_{1,t} - 0.7x_{2,t} = x_{1,t} - 0.7(x_{2,t-1} + e_{2,t}) = e_{1,t} - 0.7e_{2,t} \sim I(0)$.

The TERV ($m = 2$) is effectively applied to system 6 (see Table 6). The coupling $X_2 \to X_1$ is denoted even for small time series lengths, even though the embedding dimension is mis-specified (we set $m = 2$, however based on the equations of the system, one should set m equal to 1).

The assumptions for using the ECM are satisfied for system 6, i.e. the variables are $I(1)$ and co-integrated. Short run causality is searched based on the coefficients Γ_i of Eq. 4, while long-run causality is addressed regarding the coefficient Π. Based on the ECM ($P = 1$), only long-run causality is detected

Table 6: Percentages of rejection of the non-causality hypothesis based on TERV ($m = 2$) and of short-run causality based on ECM ($P = 1$), for the simulation system 6.

	TERV ($m = 2$)				short-run causality			
n	256	512	1024	2048	256	512	1024	2048
$X_1 \to X_2$	2	4	6	6	7	4	5	7
$X_2 \to X_1$	100	100	100	100	1	8	7	3

(see Table 7), for the correct coupling $X_2 \to X_1$, while no short-run dynamics are noted (Table 6). The ECM model is effectively specified, since the residuals of the ECM are normally distributed, do not have autocorrelations or ARCH effects.

Table 7: Long-run causality based on ECM ($P = 1$), for the simulation system 6. For each equation, we consider the corresponding depending variable, e.g. in Eq.1 the dependent variable is X_1.

	long-run causality			
n	256	512	1024	2048
Eq.1	6	2	4	2
Eq.2	100	100	100	100

System 7. A non-stationary, linear, multivariate system, with unidirectional couplings $X_2 \to X_1$ and $X_3 \to X_1$, with co-integrated variables, defined as

$$\begin{aligned}
x_{1,t} &= 0.4x_{1,t-1} + 0.4x_{2,t-1} + 0.5x_{3,t-1} + 0.2x_{1,t-2} \\
&\quad -0.2x_{2,t-2} - 0.2x_{1,t-3} + 0.15x_{2,t-3} + 0.1x_{3,t-3} + \epsilon_{1,t} \\
x_{2,t} &= 0.6x_{2,t-1} + 0.2x_{2,t-2} + 0.2x_{2,t-3} + \epsilon_{2,t} \\
x_{3,t} &= 0.4x_{3,t-1} + 0.3x_{3,t-2} + 0.3x_{3,t-3} + \epsilon_{3,t}.
\end{aligned}$$

The variables of system 7 are $I(1)$ and co-integrated with one co-integration relationship (see [8], Model 8, p.78).

The PTERV ($m = 3$) detects the direct couplings for the time series lengths $n = 1024$ and 2048, while the corresponding percentages of significant PTERV values increase with n (see Table 8).

Table 8: As Table 6 but for system 7.

	TERV ($m = 2$)				short-run causality			
n	256	512	1024	2048	256	512	1024	2048
$X_1 \to X_2$	1	1	1	3	0	0	0	0
$X_2 \to X_1$	3	5	34	85	95	100	100	100
$X_2 \to X_3$	3	2	2	2	0	0	0	0
$X_3 \to X_2$	2	2	2	2	0	0	0	0
$X_1 \to X_3$	1	0	1	0	0	0	0	0
$X_3 \to X_1$	3	6	44	99	13	8	11	10

Based on the VECM ($P = 3$), only the coupling $X_2 \to X_1$ is detected in the short-run (see Table 8), while long-run relationships are present only when variable X_1 is the dependent variable, however small percentages are obtained for all n (Table 9). The ECM model is effectively specified.

Table 9: As Table 7 but for system 7.

	long-run causality			
n	256	512	1024	2048
Eq.1	15	14	25	21
Eq.2	4	3	2	4
Eq.3	1	3	2	4

System 8. Finally, we consider three co-integrated random walks

$$\begin{aligned} r_{1,t} &= w_t + \epsilon_{1,t} \\ r_{2,t} &= 0.3w_t + \epsilon_{2,t} \\ r_{3,t} &= 0.6w_t + \epsilon_{3,t} \end{aligned}$$

with a common stochastic drift $w_t = w_{t-1} + \epsilon_t$, where a coupled system with linear ($X_2 \to X_3$) and non-linear couplings ($X_1 \to X_2$ and $X_1 \to X_3$) is superimposed

$$\begin{aligned} x_{1,t} &= 3.4x_{1,t-1}(1 - x_{1,t-1}^2)\exp -x_{1,t-1}^2 + 0.4\delta_{1,t} \\ x_{2,t} &= 3.4x_{2,t-1}(1 - x_{2,t-1}^2)\exp -x_{2,t-1}^2 + 0.5x_{1,t-1}x_{2,t-1} + 0.4\delta_{2,t} \\ x_{3,t} &= 3.4x_{3,t-1}(1 - x_{3,t-1}^2)\exp -x_{3,t-1}^2 + 0.5x_{2,t-1} + 0.5x_{1,t-1}^2 + 0.4\delta_{3,t} \end{aligned}$$

where ϵ_t, $\epsilon_{i,t}$ and $\delta_{i,t}$ are Gaussian innovations with zero mean and variance one. System 8 is non-stationary and co-integrated by construction.

The PTERV ($m = 2$) is effective only for large time series lengths, while the corresponding percentages of significant PTERV values increase with n (see Table 10). For large $n = 2048$, the spurious coupling $X_3 \to X_1$ is also indicated.

Table 10: As Table 6 but for system 8.

	TERV ($m = 2$)				short-run causality			
n	256	512	1024	2048	256	512	1024	2048
$X_1 \to X_2$	10	6	14	35	0	0	0	0
$X_2 \to X_1$	5	9	14	6	31	40	66	93
$X_2 \to X_3$	6	18	25	64	0	0	0	0
$X_3 \to X_2$	1	9	6	8	0	0	0	0
$X_1 \to X_3$	7	11	22	39	0	0	0	0
$X_3 \to X_1$	8	19	11	37	21	38	43	63

The VECM ($P = 2$) fails to detect the true couplings, while the spurious couplings $X_2 \to X_1$ and $X_3 \to X_1$ are obtained in the short run (Table 10). For this system, the residuals from the VECM seem to have sifnificant autocorrelations, especially for large n. Regarding the long run causality, high percentages of rejection of the non-causality hypothesis are obtained when variable X_3 is the dependent one, while lower percentages are denoted when X_1 is the dependent variable (see Table 11).

Table 11: As Table 7 but for system 8.

	long-run causality			
n	256	512	1024	2048
Eq.1	34	33	20	20
Eq.2	10	11	11	13
Eq.3	61	68	80	80

5 Application

The causality test are considered in two applications, where the directed relationships among financial time series are investigated. The TERV / PTERV is estimated on the original financial time series (prices) in both applications, whereas the ECM / VECM is applied on the logarithmic time series in order to avoid spurious couplings due to the variability of the data.

In the first application, we study the direct and indirect relationships in international stock markets. We consider the Morgan Capital International's market capitalization weighted index data of three Stanley developed markets. The data consist of daily measurements from 5/3/2004-5/3/2009 from Germany (denoted as X_1), Greece (X_2) and USA (X_3) (see Fig. 2). The logarithmic time series are denoted by Y_1, Y_2 and Y_3, respectively.

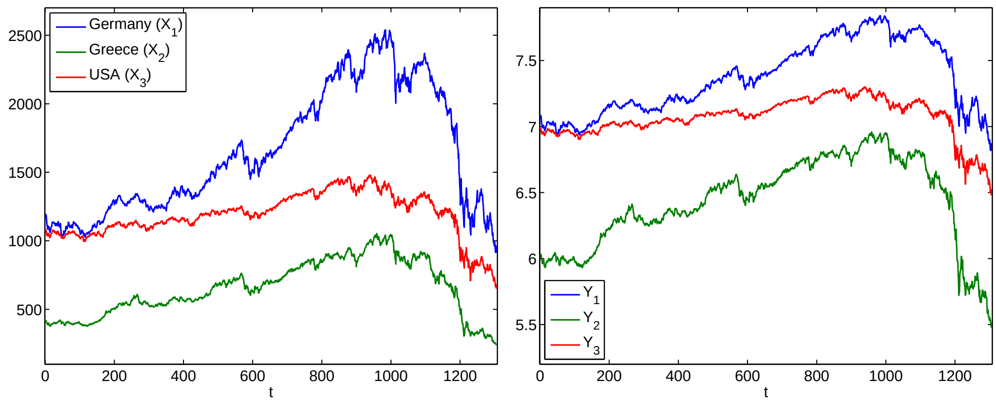

Figure 1: Time series of (a) original prices of developed markets and (b) after taking the natural logarithm of them.

The Augmented Dickey-Fuller test indicated that both the original and the logarithmic time series are non-stationary ($I(1)$). The results of the test for both data sets are displayed in Table 12.

Table 12: Results from Augmented Dickey-Fuller test for application 1.

var.	p-val.	stat.	crit. value	var.	p-val.	stat.	crit. value
X_1	0.990	1.055	-3.414	Y_1	0.990	1.749	-3.414
X_2	0.990	1.052	-3.414	Y_2	0.990	2.052	-3.414
X_3	0.990	1.362	-3.414	Y_3	0.990	2.066	-3.414

The Johansen co-integration test is also applied to the data for lag equal to 2 (based on BIC). One co-integrating relationship is indicated for the original

Table 13: Results from Johansen co-integration test for application 1.

original prices						
null	trace stat.	crit 90%	crit 95%	eigen. stat	crit 90%	crit 95%
$r \leq 0$	38.892	32.065	35.012	24.696	21.873	24.252
$r \leq 1$	14.195	16.162	18.398	13.173	15.001	17.148
$r \leq 2$	1.023	2.705	3.841	1.023	2.705	3.841
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

prices, however no co-integration is present when applied to the logarithmic prices. The results from the co-integration tests are presented in Table 13.

The PTERV is estimated for $h = 1$, $m = 2$ and $\tau = 1$. Results indicate the driving of USA on Greece (p-value from surrogate test is 0.007) and of USA on Germany (p-value= 0.007), while a bidirectional coupling is found between Germany and Greece (p-value = 0.007). Since no co-integration is detected for the logarithmic prices, the CG is applied on the logarithmic prices after differencing in order to have stationary time series. Results suggest the driving of USA on Greece (p-value from F-test = 0), the driving of USA on Germany (p-value= 0) and the driving of Germany on Greece (p-value= 0.025).

We note that the PTERV gives similar results for embedding dimension $m = 3$, while for larger m, no couplings are indicated. Equivalent results are displayed when applied to the logarithmic time series.

++ comment on the results

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: 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). The logarithmic time series are denoted by Z_1, Z_2 and Z_3, respectively.

The Augmented Dickey-Fuller test indicated that both the original and the logarithmic time series are non-stationary ($I(1)$). The results of the test for both data sets are displayed in Table 14.

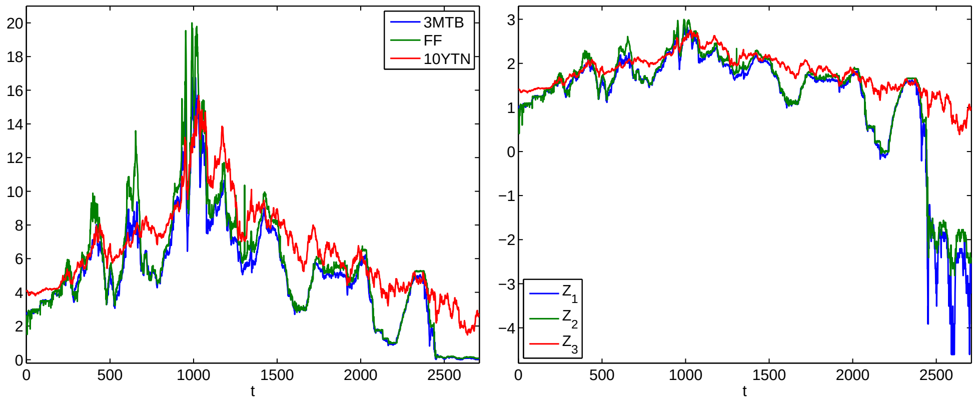

Figure 2: Time series of (a) interest rates and (b) after taking the natural logarithm of them.

Table 14: Results from Augmented Dickey-Fuller test for application 2.

var.	p-val.	stat.	crit. value	var.	p-val.	stat.	crit. value
X_1	0.256	-2.679	-3.414	Z_1	0.528	-2.131	-3.414
X_2	0.253	-2.686	-3.414	Z_2	0.945	-0.976	-3.414
X_3	0.605	-1.976	-3.414	Z_3	0.523	-2.142	-3.414

The Johansen co-integration test is applied to the original prices for lag equal to 2 (based on BIC), and no co-integration is found. On the other hand, when applied to the logarithmic prices (BIC indicates here to use lag equal to 6), one co-integrating relationship is indicated. The results from the co-integration tests are presented in Table 15.

The PTERV is estimated for $h = 1$, $m = 2$ and $\tau = 1$. Results indicate the driving of 3MTB on FF (p-value from surrogate test is 0.007) and of 10YTN on FF (p-value= 0.007), while a bidirectional coupling is found between 3MTB and 10YTN (p-value = 0.007). The VECM is applied on the logarithmic prices for order model $P = 6$. All the couplings are found to be bidirectional, while long run causality is also detected when FF is the dependent variable.

We note here that the PTERV gives the same results for embedding dimension $m = 3$, while for larger m, no couplings are indicated. Equivalent results are displayed when applied to the logarithmic time series. On the other hand, based on VECM, the number of couplings increases with the P.

Table 15: Results from Johansen co-integration test for application 2.

original prices						
null	trace stat.	crit 90%	crit 95%	eigen. stat	crit 90%	crit 95%
$r \leq 0$	188.729	32.065	35.012	157.908	21.873	24.252
$r \leq 1$	30.821	16.162	18.398	26.325	15.001	17.148
$r \leq 2$	4.496	2.705	3.841	4.496	2.705	3.841
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

++ comment on the results

6 Conclusions

The TERV / PTERV is applicable to any type of data and no assumptions should be made prior to its estimation. It can effectively detect the direction of both linear and non-linear couplings. It is sensitive to its free parameters (step ahead h, the embedding dimension m and the time delay τ). Its performance improves with the time series length and does not discriminate between long and short run causality. Based on the simulation study, its performance improves with the time series length, while it is not efficient for large embedding dimensions.

The Granger causality test has proved to be effective in different applications, however it cannot be applied when co-integration exists. The ECM / VECMs is effective for the detection of directional interactions in bivariate / multivariate systems when data are non-stationary (with the same degree of integration) and co-integrated. It can discriminate between long- and short-run causality. It is efficient even for small time series lengths. However, it seems to be ineffective when only non-linear couplings are present.

++ comment on results from application

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.

References

[1] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716723, 1974.

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

[3] F.M. Aparicio and A. Escribano. Information-theoretic analysis of serial dependence and cointegration. *Studies in Nonlinear Dynamics and Econometrics*, 3(3):119–140, 1998.

[4] L.A. Baccalá and K. Sameshima. Partial directed coherence: a new concept in neural structure determination. *Biological Cybernetics*, 84(6):463–474, 2001.

[5] D.A. Dickey and W.A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366):427–431, 1979.

[6] M. Ding, Y. Chen, and S.L. Bressler. *Granger causality: Basic theory and application to neuroscience. In Schelter, B., Winterhalder, M., Timmer, J. (Eds.), Handbook of Time Series Analysis.* Wiley-VCH, Berlin, 2006.

[7] R.F. Engle and C.W.J. Granger. Co-integration and error correction: Representation, estimation and testing. *Econometrica*, 55(2):251–276, 1987.

[8] Sharp G.D. *Lag length selection for vector error correction models.* PhD thesis, Rhodes university, 2010.

[9] J. Geweke, R. Meese, and W.T. Dent. Comparing alternative tests of causality in temporal systems: analytic results and experimental evidence. *Journal of Econometrics*, 21:161–194, 1983.

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

[11] C.W.J. Granger. Some recent developments in a concept of causality. *Journal of Econometrics*, 39:199–211, 1988.

[12] C.W.J. Granger and P. Newbold. Spurious regressions in econometrics. *Journal of Econometrics*, 2:110–120, 1974.

[13] C. Hiemstra and J.D. Jones. Testing for linear and nonlinear granger causality in the stock pricevolume relation. *Journal of Finance*, 49:1639–1664, 1994.

[14] S. Johansen. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59(6):15511580, 1991.

[15] D. Kugiumtzis. Transfer entropy on rank vectors. *Journal of Nonlinear Systems and Applications*, 3(2):73–81, 2012.

[16] D. Kugiumtzis. Partial transfer entropy on rank vectors. *The European Physical Journal Special Topics*, 222(2):401–420, 2013.

[17] J.P. LeSage. A comparison of the forecasting ability of ECM and VAR models. *The Review of Economics and Statistics*, 72(4):664–671, 1990.

[18] R. Marschinski and H. Kantz. Analysing the information flow between financial time series - an improved estimator for transfer entropy. *European Physical Journal B*, 30:275–281, 2002.

[19] T.J. Mosedale, D.B. Stephenson, M. Collins, and T.C. Mills. Granger causality of coupled climate processes: ocean feedback on the north atlantic oscillation. *J. Climate*, 19:1182–1194, 2006.

[20] M. Paluš, V. Komárek, Z. Hrnčíř, and K. Štěrbová. Synchronization as adjustment of information rates: Detection from bivariate time series. *Physical Review E*, 63:046211, 2001.

[21] P.C.B. Phillips and S. Ouliaris. Asymptotic properties of residual based tests for cointegration. *Econometrica*, 58:165–193, 1990.

[22] P.C.B Phillips and P. Perron. Testing for a unit root in time series regression. *Biometrika*, 75:335–346, 1988.

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

[24] S.J. Schiff, P. So, T. Chang, R.E. Burke, and T. Sauer. Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble. *Physical Review E*, 54:6708–6724, 1996.

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

[26] G.E. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464, 1978.

[27] C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.

[28] M. Staniek and K. Lehnertz. Symbolic transfer entropy. *Physical Review Letters*, 100(15):158101, 2008.

[29] F. Wendling, F. Bartolomei, J. Bellanger, and P. Chauvel. Interpretation of interdependencies in epileptic signals using a macroscopic physiological model of the EEG. *Clin. Neurophysiol.*, 112:1201–1218, 2001.

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