Partial Symbolic Transfer Entropy

Angeliki Papana*
Department of Economics, University of Macedonia, Greece
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

November 7, 2013

Abstract

In this paper, we introduce the partial symbolic transfer entropy (PSTE), an extension of the symbolic transfer entropy that accounts only for the direct causal effects among the components of a multivariate system. It is an information theoretic measure, and as such does not suffer from model mis-specification bias. The PSTE is defined on the ranks of vectors that are formed from the reconstructed vectors, instead of the original time series values. The statistical significance of

*Corresponding author

PSTE is assessed by randomization test making use of surrogate time series. The PSTE is evaluated on multivariate time series of different types of coupled and uncoupled systems and compared with conditional Granger causality index (CGCI). It is shown that the PSTE is not affected by the existence of outliers, it is directly applicable to time series that are non-stationary in mean and in variance, and it is also not affected by data filtering. As a real application, the causal effects among three economic indexes are investigated. Computations of PSTE and CGCI on both the initial returns and the VAR filtered returns, and only of PSTE on the original indexes, showed consistency of the PSTE in estimating the causal effect.

1 Introduction

The investigation of interactions among the components of a multivariate system addresses two major aspects: the detection of the couplings direction and the quantification of the coupling strength. When evaluating the causal influence between two time series from a multivariate data set, it is necessary to take the effects of the remaining variables into account. Multivariate analysis is required to distinguish between direct and indirect causal effects.

The concept of linear Granger causality is a fundamental tool for the investigation of dynamic interactions from multivariate time series [12]. Linear Granger causality is based on the conception that causes always precede their effects and is implemented by fitting autoregressive models. However, the selected model should be appropriately matched to the underlying dynamics of the examined system, otherwise model mis-specification may lead to spurious causalities.

For the identification of the causal inter-relationships of time series, information theoretic approaches have been developed. Their advantage is that they are sensitive to nonlinear signal properties and they are model-free. On the other hand, information measures usually require more data than linear model-based methods, such as linear Granger causality. Information measures have also more free parameters and may suffer from their mis-specification.

In order to implement the linear Granger causality, weak stationarity of the data is assumed. However, stationarity is difficult to justify for real data. It emerges from many studies that processes with non constant mean or/ and variance are a common feature in practice. Pre-processing of the data (e.g.

detrending, differencing, filtering) can be used to deal with non-stationarity (e.g. see [5, 30]).

Several methods for time series analysis in presence of non constant variance have been proposed in the literature, e.g. model fitting allowing for a non constant variance and tests of heteroscedasticity [31, 15]. Causality measures, such as transfer entropy and linear Granger causality, are theoretically invariant under a rather broad class of transformations [4]. However, in the case of real data, these theoretically invariant transformations, may impact causal inference.

Causality can be subdivided into short-run and long-run causality, e.g. as modeled using error correction models [22, 7]. A cointegration test can be viewed as an indirect test of long-run dependence. Cointegration between two variables implies the existence of long-run causality in at least one direction [8]. In that way, we can separate short and long-run relationships among variables. Testing for cointegration and causality should be considered jointly.

In financial applications, most causality tests are not implemented on the raw data but on the (log) returns. For example, the modified test of nonlinear Granger causality introduced in [13] is usually applied on the residual series from a VAR model. It is however considered that linear filtering of the data before the application of a causality test can lead to serious distortions (e.g. see [20, 14]). On the other hand, [10] claims that the calculation of information-theoretic quantities is typically improved by diminishing long-range second-order temporal structure using VAR filters, provided that the interactions between time series are not purely linear. The influence of filtering on the different causality tests remains an open issue for further investigation.

A causality measure that can be applied to non-stationary data is essential for the detection of the causal effects in real data sets. In this work, we extend the bivariate information causality measure symbolic transfer entropy (STE) [28] to the multivariate case, called partial symbolic transfer entropy (PSTE), in order to introduce a direct causality measure applicable to non-stationary time series, while also not affected by the filtering of the data. A corrected version of the STE and PSTE have been introduced recently [18, 19], but here we consider the initial definition used in different applications [16, 17, 23].

The PSTE is evaluated on multivariate time series of known coupled and uncoupled systems, on stationary and non-stationary time series in mean and in variance, on time series with outliers, and on VAR filtered time series. For comparison, the conditional Granger causality index (CGCI) is also considered. As a real application, the causal effects among three economic time series are investigated.

In Sec. 2, the multivariate causality measures partial symbolic transfer entropy and conditional Granger causality index are presented, and their statistical significance is discussed. In Sec. 3, the two causality measures are evaluated in a simulation study. The performance of the measures in a real application with three financial time series is presented in Sec. 4. Finally, the conclusions of this study are discussed in Sec. 5.

2 Materials and Methods

2.1 Partial symbolic transfer entropy

The symbolic transfer entropy (STE) is an extension of transfer entropy defined on rank-points formed by the reconstructed vectors of the variables [28]. As it is estimated based on the ranks of the reconstructed vectors of the time series instead of the time series values, the PSTE is applicable to time series that are non-stationary in the level (mean), since slow drifts do not have a direct effect on the ranks.

Let us consider 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 embedding parameters in order to form the reconstructed vector of the time series X_1 are the embedding dimension m_1, the time delay τ_1 and let h be the number of time steps ahead to address the interaction. The reconstructed vectors of X_1 are defined as $\mathbf{x}_{1,t} = (x_{1,t}, x_{1,t-\tau_1}, \ldots, x_{1,t-(m_1-1)\tau_1})'$, where $t = 1, \ldots, n'$ and $n' = n - \max\{(m_1-1)\tau_1, (m_2-1)\tau_2\}$. The reconstructed vector for X_2 is defined accordingly, with parameters m_2 and τ_2. For each vector $\mathbf{x}_{1,t}$, the ranks of its components assign a rank-point $\hat{\mathbf{x}}_{1,t} = [r_{1,t}, r_{2,t}, \ldots, r_{m_1,t}]$, where $r_{j,t} \in \{1, 2, \ldots, m_1\}$ for $j = 1, \ldots, m_1$, and $\hat{\mathbf{x}}_{2,t}$ is defined accordingly. The STE is defined as

$$\mathrm{STE}_{X_2 \to X_1} = \sum p(\hat{\mathbf{x}}_{1,t+h}, \hat{\mathbf{x}}_{1,t}, \hat{\mathbf{x}}_{2,t}) \log \frac{p(\hat{\mathbf{x}}_{1,t+h}|\hat{\mathbf{x}}_{1,t}, \hat{\mathbf{x}}_{2,t})}{p(\hat{\mathbf{x}}_{1,t+h}|\hat{\mathbf{x}}_{1,t})}, \quad (1)$$

where $p(\hat{\mathbf{x}}_{1,t+h}, \hat{\mathbf{x}}_{1,t}, \hat{\mathbf{x}}_{2,t})$, $p(\hat{\mathbf{x}}_{1,t+h}|\hat{\mathbf{x}}_{1,t}, \hat{\mathbf{x}}_{2,t})$ and $p(\hat{\mathbf{x}}_{1,t+h}|\hat{\mathbf{x}}_{1,t})$ are the joint and conditional distributions defined on the rank vectors. STE can also be expressed in entropy terms as

$$\mathrm{STE}_{X_2\to X_1} = H(\hat{\mathbf{x}}_{2,t},\hat{\mathbf{x}}_{1,t}) - H(\hat{\mathbf{x}}_{1,t+h},\hat{\mathbf{x}}_{2,t},\hat{\mathbf{x}}_{1,t}) + H(\hat{\mathbf{x}}_{1,t+h},\hat{\mathbf{x}}_{1,t}) - H(\hat{\mathbf{x}}_{1,t}), \quad (2)$$

where $H(\cdot)$ is the Shannon entropy defined on the rank-points.

The partial symbolic transfer entropy (PSTE) is the extension of the STE that accounts only for direct causal effects in multivariate systems. It is defined conditioning on the set of the remaining variables $Z = \{X_3, X_4, \ldots, X_K\}$ of a multivariate system of K observed variables

$$\mathrm{PSTE}_{X_2\to X_1|Z} = \sum p(\hat{\mathbf{x}}_{1,t+h},\hat{\mathbf{x}}_{1,t},\hat{\mathbf{x}}_{2,t},\hat{\mathbf{z}}_t)\log\frac{p(\hat{\mathbf{x}}_{1,t+h}|\hat{\mathbf{x}}_{1,t},\hat{\mathbf{x}}_{2,t},\hat{\mathbf{z}}_t)}{p(\hat{\mathbf{x}}_{1,t+h}|\hat{\mathbf{x}}_{1,t},\hat{\mathbf{z}}_t)}, \quad (3)$$

where the rank vector $\hat{\mathbf{z}}_t$ is defined as the concatenation of the rank vectors for each of the embedding vectors of the variables in Z. For the estimation of the PSTE, the joint and conditional distributions are estimated from the sample probabilities.

The PSTE is a measure based on nonparametric estimators from information theoretic arguments. Its definition is built on the probability distributions or equivalently on conditional entropies and quantifies the reduction in conditional uncertainty of $\hat{\mathbf{x}}_{1,t+h}$ when conditioning changes from $\hat{\mathbf{x}}_{1,t}, \hat{\mathbf{z}}_t$ to $\hat{\mathbf{x}}_{2,t}, \hat{\mathbf{x}}_{1,t}, \hat{\mathbf{z}}_t$. Causality is defined in terms of predictive power using an information theoretic statistic rather than linear modeling tools. Given this general criterion of predictive power, this approach is less sensitive to the presence of non-stationarity and accounts for nonlinearity in the data.

2.2 Conditional Granger Causality Index

For comparison reasons, the Conditional Granger Causality Index (CGCI) is also considered in this study [9]. Granger causality by definition states that an observed time series X_2 Granger causes another series X_1 if the knowledge of past values of the X_2 significantly improves the prediction of X_1. Granger causality provides a measure of the strength of the interaction between time series.

In order to estimate the linear Granger causality index, a bivariate autoregressive model of order P is fitted to the time series $\{x_{1,t}\}$ and $\{x_{2,t}\}$

$$x_{i,t+1} = \sum_{j=0}^{P-1} a_{i,j}x_{1,t-j} + \sum_{j=0}^{P-1} b_{i,j}x_{2,t-j} + \epsilon_{i,t+1}, \quad i = 1,2 \quad (4)$$

where $a_{i,j}$ and $b_{i,j}$ are the coefficients of the model and $\epsilon_{i,t}$ the residuals from fitting the model. If the variance s_{1U}^2 of the residuals of the 'unrestricted' model in Eq. 4 for X_1 is statistically significantly less than the variance s_{1R}^2 of the 'restricted' model for X_1 that does not include X_2 (the second sum in Eq. 4), then this is statistical evidence that the variable X_2 Granger causes X_1. The magnitude of the effect of X_2 on X_1 is given by the Granger Causality Index (GCI) as

$$\mathrm{GCI}_{X_2 \to X_1} = \ln(s_{1R}^2 / s_{1U}^2). \tag{5}$$

Considering further the variables $Z = \{X_3, X_4, \ldots, X_K\}$ in the model of Eq. 4 (adding additional terms for each of the variables $X_3, X_4, \ldots, X_K$) and defining the 'restricted' and the 'unrestricted' model in the same way, the conditional Granger causality index (CGCI) is

$$\mathrm{CGCI}_{X_2 \to X_1|Z} = \ln(s_{1R}^2 / s_{1U}^2). \tag{6}$$

The CGCI is a causality measure able to detect the direct causal effects in multivariate systems with linear causal effects.

2.3 Statistical significance of PSTE and CGCI

The statistical significance of the PSTE is assessed by a randomization test for the null hypothesis of no causal effects making use of time-shifted surrogates [25]. Time-shifted surrogates are used to form the null distribution when it is not known analytically, as is the case with PSTE. The surrogate time series are formed by time-shifting the time series of the driving variable by a random time step, while the other time series are intact. By this, the driving and the response time series become independent and the causal effects are destroyed. Explaining further time-shifting, we draw a random integer d (with d less than the time series length), and the first d values of the driving time series are moved to the end.

To examine the null hypothesis H_0 of no causal effects, the PSTE is estimated from the original data (let us denote it q_0) and from M surrogate data series (let us denote them $q_1, \ldots, q_M$). H_0 is rejected if q_0 lies at the tail of the distribution of $q_1, \ldots, q_M$. The p-values for the two-sided test are derived by rank ordering. Letting the original value have rank i in the ordered list of $M + 1$ values, the p-value is $2i/(M + 1)$ if $i < M + 1$ and

$2(M+1-i)/(M+1)$ if $i \geq M+1$ (actually the correction of the rank approximation of the cumulative density function in [32] is applied).

The statistical significance of the CGCI can be given by means of a parametric test, i.e. the F-test for the null hypothesis that the coefficients for the driving variable in the unrestricted model are zero [6]. For example, applying the F-significance test for each of the P coefficients $b_{i,j}$ in Eq.4, constitutes the parametric significance test for CGCI to test the null hypothesis that variable X_2 is not driving X_1.

3 Simulations and Results

The PSTE and the CGCI are estimated from 100 realizations of the simulation systems, for different coupling strengths and for all directions. In this section, the simulation systems and the results of the simulation study are presented.

3.1 Simulation study

The PSTE is evaluated on multivariate time series of different types of coupled and uncoupled systems, which are stationary or non-stationary in mean (cointegrated or not) or in variance, coupled systems with outliers, with linear or/ and nonlinear causal effects. In order to evaluate the sensitivity of the PSTE on filtering, it is also estimated on VAR filtered time series. Specifically, the following simulation systems are considered:

1. A stationary system in three variables with linear ($X_2 \to X_3$) and nonlinear causal effects ($X_1 \to X_2$, $X_1 \to X_3$) ([11], Model 7)

$$\begin{aligned} x_{1,t} &= 3.4x_{1,t-1}(1-x_{1,t-1})^2\exp(-x_{1,t-1}^2) + 0.4\epsilon_{1,t} \\ x_{2,t} &= 3.4x_{2,t-1}(1-x_{2,t-1})^2\exp(-x_{2,t-1}^2) + 0.5x_{1,t-1}x_{2,t-1} + 0.4\epsilon_{2,t} \\ x_{3,t} &= 3.4x_{3,t-1}(1-x_{3,t-1})^2\exp(-x_{3,t-1}^2) + 0.3x_{2,t-1} + 0.5x_{1,t-1}^2 + 0.4\epsilon_{3,t} \end{aligned}$$

where $\epsilon_{i,t}$, $i = 1,2,3$, are Gaussian white noise terms with unit variance.

2. A stationary system in three variables, with only nonlinear causal effects ($X_1 \to X_2$, $X_1 \to X_3$)

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

$$
\begin{aligned}
x_{2,t} &= 0.3x_{2,t-1} + 0.5x_{2,t-2}x_{1,t-1} + \epsilon_{2,t} \\
x_{3,t} &= 0.3x_{3,t-1} + 0.5x_{3,t-2}x_{1,t-1} + \epsilon_{3,t}
\end{aligned}
$$

The model of only the two first variables was introduced in [3]. The term of variable product in the second and third equation causes the variables X_2 and X_3 to have marginal distributions with long tails.

3. A stationary system of three coupled Hénon maps with nonlinear causal effects ($X_1 \to X_2$, $X_2 \to X_3$)

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

with equal coupling strengths c for $X_1 \to X_2$ and $X_2 \to X_3$, for $c = 0,\ 0.05,\ 0.3,\ 0.5$. The time series of this system become completely synchronized for coupling strengths $c \geq 0.7$.

4. A stationary system with outliers, from the three coupled Hénon maps (system 3), where outliers have been randomly added to each variable under the standard uniform distribution. The number of outliers constitute 1% of the total number of data points.

5. A non-stationary system in level (mean), from the three coupled Hénon maps (system 3), where a stochastic trend $\eta_t = \eta_{t-1} + \epsilon_t$ is added to each variable, where ϵ_t, is Gaussian white noise with unit variance. The stochastic trends that are added to each variable are not co-integrated.

6. A non-stationary system in level (mean), from the three coupled Hénon maps (system 3) where a deterministic trend $\eta_t = a \cdot t$ is added to each variable, and a is a constant. The value of a is randomly set for each realization of the system, whereas a values are from a normal distribution with mean 0.01 and standard deviation 0.02. The deterministic trends are co-integrated.

7. A non-stationary system in variance, from system 2, where a 'standardized' time series from the GARCH(1,1) (to have zero mean and st. deviation 1) is added to each variable:

$$
\begin{aligned}
x_t &= \sigma_t * \varepsilon_t \\
\sigma_t^2 &= \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2,
\end{aligned}
$$

where ε_t is Gaussian white noise with unit variance, $\alpha_0 = 0.2$, $\alpha_1 = 0.9$ and $\beta_1 = 0.1$.

8. The VAR filtered residuals for system 1. The order of the VAR filter is selected based on Schwarz's Bayesian Information Criterion (BIC) [27], for each realization.

The time series lengths $n = 512$ and 2048 are considered in the simulation study, in order to test the effectiveness of the measure on small and large time series lengths. The same embedding parameters are considered for each variable of a coupled system. The embedding dimension m for the estimation of PSTE is set based on the complexity of each coupled system, the time delay τ is set to 1 and the time step ahead h is set to 1. For the estimation of CGCI, the Akaike Information Criterion (AIC) [1] and the Bayesian Information Criterion (BIC) [27] are considered to determine the optimal model order P of the VAR model.

3.2 Results from simulation study

The performance of each causality measure is quantified by the percentages of significant causality values in the 100 realizations for all the ordered couples of variables in the system, i.e. the percentage of rejections of the null hypothesis of no causal effects. For both measures, the causal effects are always regarded conditioned on the remaining variable (i.e. on the third variable).

System 1 For the first simulation system, the optimal choice for the embedding dimension m would be 1, since the maximum delay in the equations of the system is 1. By definition, however, we can only set $m \geq 2$ to estimate the PSTE. For $m = 2$, the PSTE correctly detects the direct linear causal effect $X_2 \rightarrow X_3$ for $n = 2048$ and the nonlinear causal effect $X_1 \rightarrow X_2$ for both time series lengths. For these directions, the power of the test increases with n. However, the PSTE fails to detect the nonlinear causal effect $X_1 \rightarrow X_3$ (see Table 1). The percentages of significant PSTE values at the direction of no causal effects are low (between 1 and 8%). The failure of PSTE to detect the relationship $X_1 \rightarrow X_3$, is probably due to the fact that the effect of X_2 on X_3 is much greater than the effect of X_1 on X_3. The weak causal effect of X_1 on X_3 might be arising from the small values of the variable X_1, that get even smaller by squaring (x_1^2 is included in the equation of the system).

Table 1: Percentages of statistically significant PSTE ($m = 2$) and CGCI ($P = 2$) values for the simulation system 1.

PSTE	$X_1 \to X_2$	$X_2 \to X_1$	$X_2 \to X_3$	$X_3 \to X_2$	$X_1 \to X_3$	$X_3 \to X_1$
$n = 512$	13	5	66	5	2	5
$n = 2048$	68	5	100	6	6	8
CGCI	$X_1 \to X_2$	$X_2 \to X_1$	$X_2 \to X_3$	$X_3 \to X_2$	$X_1 \to X_3$	$X_3 \to X_1$
$n = 512$	12	2	100	7	7	4
$n = 2048$	7	7	100	4	7	5

The CGCI also fails to detect all the causal effects of the first coupled system, for order models $P = 1$, 2 and 3 (BIC suggested to set $P = 1$ and AIC $P = 1$, 2 and 3 for the different realizations). It captures only the linear causal effect $X_2 \to X_3$, while fails to take into consideration the nonlinear connectivity $X_1 \to X_2$ and $X_1 \to X_3$ (see Table 1 for $P = 2$). The percentages of significant CGCI values at the direction of no causal effects are again low (e.g. between 4 and 7% for $P = 2$). The CGCI is a linear measure and therefore it is expected to present low power in the case of nonlinear couplings. One requirement for linear Granger causality is the separability, i.e. the information about a causative factor should be unique to that variable, which may not hold for nonlinear systems [29].

System 2 The second simulation system is a stationary system with only nonlinear causal effects ($X_1 \to X_2$ and $X_1 \to X_3$), whereas the variables X_2 and X_3 come from distributions with long tails. The maximum delay in the equations of this system is 2, and therefore we set $m = 2$. One realization of system 2, for $n = 512$ is displayed in Fig. 1a.

The PSTE correctly detects the nonlinear direct causality for system 2 for $m = 2$, giving though low percentages of significant PSTE values for $n = 512$ (see Table 2). Again, the power of the test increases with the time series length n. The percentages of significant PSTE values at the direction of no causal effects are between 1 and 6%.

The CGCI is not able to detect the two nonlinear interactions (see Table 2). On the other hand, it indicates the spurious causal effects $X_2 \to X_1$ (almost 40%), $X_2 \to X_3$ (almost 70%), $X_3 \to X_2$ (almost 68%) and $X_3 \to X_1$ (almost 43%). The estimated BIC and AIC values vary from 1 to 10 for the different realization of this coupled systems. The CGCI is estimated for P

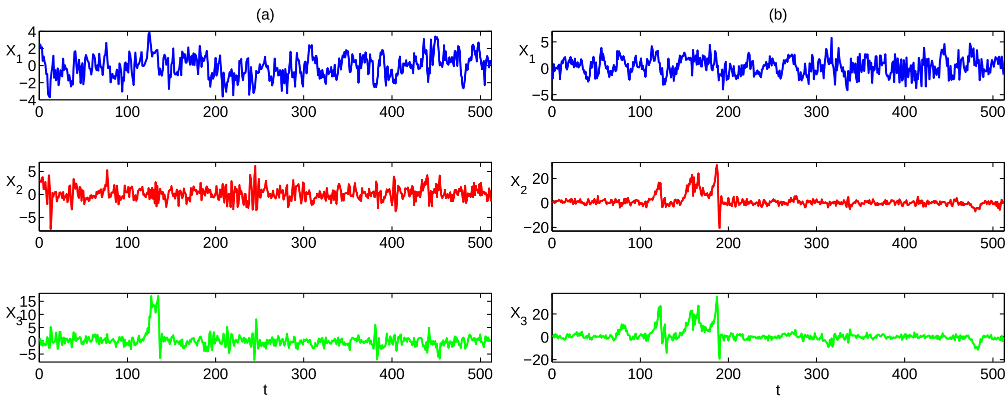

Figure 1: (a) One realization of system 2, (b) one realization of system 7.

Table 2: Percentages of statistically significant PSTE ($m = 2$) and CGCI ($P = 2$) values for the simulation system 2.

PSTE	$X_1 \to X_2$	$X_2 \to X_1$	$X_2 \to X_3$	$X_3 \to X_2$	$X_1 \to X_3$	$X_3 \to X_1$
$n = 512$	20	2	6	3	19	1
$n = 2048$	86	4	2	6	86	5
CGCI	$X_1 \to X_2$	$X_2 \to X_1$	$X_2 \to X_3$	$X_3 \to X_2$	$X_1 \to X_3$	$X_3 \to X_1$
$n = 512$	3	41	61	55	3	40
$n = 2048$	1	41	78	81	5	45

orders from 1 to 10, however the results are similar for all P values.

System 3 For the third coupled system (Hénon maps), the PSTE is estimated for $m = 2$. For the uncoupled case ($c = 0$), the PSTE indicates no causal effects, while for weakly coupled system ($c = 0.05$), gives very low percentages. For coupling strength $c = 0.3$ and for strongly coupled systems ($c = 0.5$), it correctly detects the causal effects. The power of the test increases with n. For $c = 0.5$ and $n = 2048$, along with 100% significant PSTE for the true couplings, there is high percentage of significant PSTE also for false couplings, approximately 30% for $X_2 \to X_1$ and $X_3 \to X_2$ (see Table 3), while for $X_1 \to X_3$ and $X_3 \to X_1$ the percentage varies from 1% to 9%. For $m = 3$, the PSTE indicates the indirect causal effects $X_1 \to X_3$ and the spurious causal effects $X_2 \to X_1$, $X_3 \to X_2$, but only for $c = 0.5$ and

$n = 2048$.

Table 3: Percentages of statistically significant PSTE ($m = 2$) values for the simulation system 3.

$n = 512$	$X_1 \to X_2$	$X_2 \to X_1$	$X_2 \to X_3$	$X_3 \to X_2$	$X_1 \to X_3$	$X_3 \to X_1$
$c = 0$	6	9	6	4	3	8
$c = 0.05$	9	2	7	1	5	9
$c = 0.3$	19	7	18	8	4	5
$c = 0.5$	67	16	79	7	3	7
$n = 2048$	$X_1 \to X_2$	$X_2 \to X_1$	$X_2 \to X_3$	$X_3 \to X_2$	$X_1 \to X_3$	$X_3 \to X_1$
$c = 0$	3	2	3	3	1	1
$c = 0.05$	6	5	3	4	2	3
$c = 0.3$	88	6	98	8	7	4
$c = 0.5$	100	31	100	31	7	0

The CGCI has a poor performance for the coupled Hénon maps for $P = 2$. It correctly detects the direct causal effects for $c = 0.3$ and 0.5, however also falsely detects with high confidence the indirect causality $X_1 \to X_3$ and the spurious causal effects $X_2 \to X_1$ and $X_3 \to X_2$ for $n = 2048$ and $c = 0.5$ (only $X_3 \to X_2$ for $c = 0.3$) (see Table 4). Results for $P = 3$ seem to improve the performance of the CGCI, since it correctly indicates the causal relationship for $c = 0.3$ and $c = 0.5$, and indicates only the indirect causal effect $X_1 \to X_3$ for $c = 0.5$ and $n = 2048$ (52%).

Table 4: Percentages of statistically significant CGCI ($P = 2$) values for the simulation system 3.

$n = 512$	$X_1 \to X_2$	$X_2 \to X_1$	$X_2 \to X_3$	$X_3 \to X_2$	$X_1 \to X_3$	$X_3 \to X_1$
$c = 0$	19	13	13	7	10	10
$c = 0.05$	13	12	8	8	14	10
$c = 0.3$	99	9	96	31	7	10
$c = 0.5$	100	9	100	21	5	6
$n = 2048$	$X_1 \to X_2$	$X_2 \to X_1$	$X_2 \to X_3$	$X_3 \to X_2$	$X_1 \to X_3$	$X_3 \to X_1$
$c = 0$	11	12	10	11	10	14
$c = 0.05$	29	20	20	10	11	10
$c = 0.3$	100	14	100	43	9	8
$c = 0.5$	100	65	100	52	8	7

System 4 For the coupled Hénon map system with the addition of outliers (1% of n), the PSTE performs similarly as without outliers. Indicative results are displayed in Table 5, for $c = 0.3$ and $c = 0.5$. We notice that the percentages of significant PSTE values at the directions $X_1 \to X_3$ and $X_3 \to X_1$ are between 3% and 10%.

Table 5: Percentages of statistically significant PSTE ($m = 2$) values for the simulation system 4.

$n = 512$	$X_1 \to X_2$	$X_2 \to X_1$	$X_2 \to X_3$	$X_3 \to X_2$	$X_1 \to X_3$	$X_3 \to X_1$
$c = 0.3$	16	5	17	7	6	7
$c = 0.5$	69	15	67	6	3	8
$n = 2048$	$X_1 \to X_2$	$X_2 \to X_1$	$X_2 \to X_3$	$X_3 \to X_2$	$X_1 \to X_3$	$X_3 \to X_1$
$c = 0.3$	88	9	98	9	4	3
$c = 0.5$	100	37	100	35	8	10

On the other hand, the CGCI is significantly affected by the existence of outliers, performing poorly for $P = 2$ and 3. The direct causal effects $X_1 \to X_2$ and $X_2 \to X_3$ are detected only for $P = 2$, $c = 0.5$ and $n = 2048$. The significance test with CGCI identifies the spurious causalities $X_2 \to X_1$ and $X_3 \to X_2$ for the coupling strengths $c = 0.3$ and 0.5.

System 5 The simulations systems 5, 6 and 7 are non-stationary, therefore only the PSTE is estimated. The first non-stationary system considered is the coupled Hénon map system with the addition of non-cointegrated stochastic trend. One realization of system 5, for $n = 512$ and $c = 0$ is displayed in the Fig. 2a.

The PSTE performs worse in the case of system 5, compared to the results from system 3. The percentages of statistically significant PSTE values for the coupled Hénon maps with addition of not co-integrated stochastic trends for $m = 2$ are lower at the directions of direct causal effects compared with those from system 3. Representative results are displayed in Table 6, for $c = 0.3$ and 0.5. However, these percentages at the directions of direct causal effects seem to increase with n, indicating that PSTE requires larger time series lengths to effectively recognize causalities. Regarding $X_1 \to X_3$ and $X_3 \to X_1$, the obtained results vary between 2% and 10%.

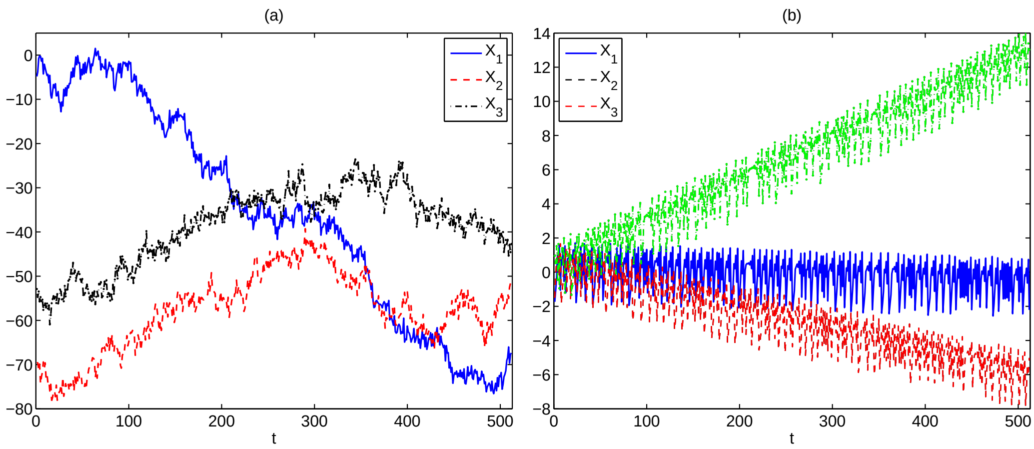

Figure 2: (a) One realization of system 5 (three coupled Hénon maps with addition of non-cointegrated stochastic trends), (b) one realization of system 6 (three coupled Hénon maps with addition of cointegrated deterministic trends), for $n = 512$.

System 6 The sixth simulation system is the coupled Hénon map with the addition of co-integrated deterministic trend, which is a non-stationary system in mean. One realization of the system for $n = 512$ for the uncoupled case ($c = 0$) is displayed in the Fig. 2b. The addition of the co-integrated deterministic trend does not affect the performance of PSTE, and the results are very similar to these for system 5.

System 7 The simulation system 7 is a non-stationary system in variance, with only nonlinear causal effects ($X_1 \to X_2$ and $X_1 \to X_3$). One realization of system 7, for $n = 512$ is displayed in Fig. 1b. The PSTE seems to be effective only for large time series lengths for $m = 2$. The percentages of significant PSTE values at the directions $X_1 \to X_2$ and $X_1 \to X_3$ increase with n (see Table 7). At the directions of no causal effects low percentages are obtained (between 2% - 5%). The performance of the PSTE improved when we increased the data size to $n = 4096$, e.g. the percentage of significant PSTE increased to 38% for $X_1 \to X_2$ and 54% for $X_1 \to X_3$.

Having the variance of input noise in the GARCH term at the same amplitude as the original system, the effect of the non-stationarity in variance turns out to be very strong. Therefore, we also examined the following cases:

Table 6: Percentages of statistically significant PSTE ($m = 2$) values for the simulation system 5.

$n = 512$	$X_1 \to X_2$	$X_2 \to X_1$	$X_2 \to X_3$	$X_3 \to X_2$	$X_1 \to X_3$	$X_3 \to X_1$
$c = 0.3$	4	4	9	7	4	4
$c = 0.5$	22	10	30	10	10	2
$n = 2048$	$X_1 \to X_2$	$X_2 \to X_1$	$X_2 \to X_3$	$X_3 \to X_2$	$X_1 \to X_3$	$X_3 \to X_1$
$c = 0.3$	8	5	16	4	6	2
$c = 0.5$	77	28	93	22	3	5

adding to system 2, standardized realizations of the GARCH(1,1) system multiplied by $c = 0.2$ and by 0.5. In these cases, the PSTE indicates much higher percentages at the directions of direct causality. The percentages remain low at the directions of no causal effects, at all three cases.

Table 7: Percentages of statistically significant PSTE ($m = 2$) values for the simulation system 7 (standardized realizations of GARCH(1,1) multiplied by c are added to each variable of system 2).

$n = 512$	$X_1 \to X_2$	$X_1 \to X_3$	$n = 2048$	$X_1 \to X_2$	$X_1 \to X_3$
$c = 1$	5	9	$c = 1$	24	17
$c = 0.5$	5	11	$c = 0.5$	46	61
$c = 0.2$	14	16	$c = 0.2$	83	73

System 8 It is a common practise in financial applications, to estimate causality measures or apply causality tests to the VAR residuals of the data in order to specify the underlying nature of the causal effects. However, the influence of the filtering on the different causality measures/ tests has not been fully investigated so far. For this reason, we consider the VAR filtered residuals of the simulation system 1 (system 8), and estimate the PSTE and the CGCI. The PSTE has similar performance for systems 8 and 1, revealing the nonlinear causal effect $X_1 \to X_2$. However, it fails to pick up the nonlinear causal effect $X_1 \to X_3$ (as for system 1). Additionally, the PSTE seems to require large time series lengths in order to be effective (see Table 8). The percentages remain low at the directions of no causal effects at all cases. As expected, the CGCI finds no causal phenomena when estimated on the VAR filtered data.

Table 8: Percentages of statistically significant PSTE ($m = 2$) values for the simulation system 8.

	$X_1 \to X_2$	$X_1 \to X_2$	$X_2 \to X_3$	$X_3 \to X_2$	$X_1 \to X_3$	$X_3 \to X_1$
$n = 512$	11	3	3	8	6	1
$n = 2048$	33	2	9	3	7	3
$n = 4096$	73	6	11	6	5	4

4 Application

In this section, we aim to study the direct causal relationships among three main financial variables. Specifically, the data consist of the following time series: the 3-Month Treasury Bill of Secondary Market Rate (denoted as X_1), the 10-Year Treasury Constant Maturity Rate (X_2) and the Chicago Board Options Exchange (CBOE) Volatility Index or VIX (X_3). Data are daily measurements from 05/01/2004 up to 18/5/2012 (see Fig. 3). The monetary policy is well represented by the 3-month Treasury Bill, the financial uncertainty by the well-known fear index, the VIX (option-implied expected volatility on the $S\&P500$ index with horizon of 30 calendar days) and the long-term rate by the 10 year Treasury Note.

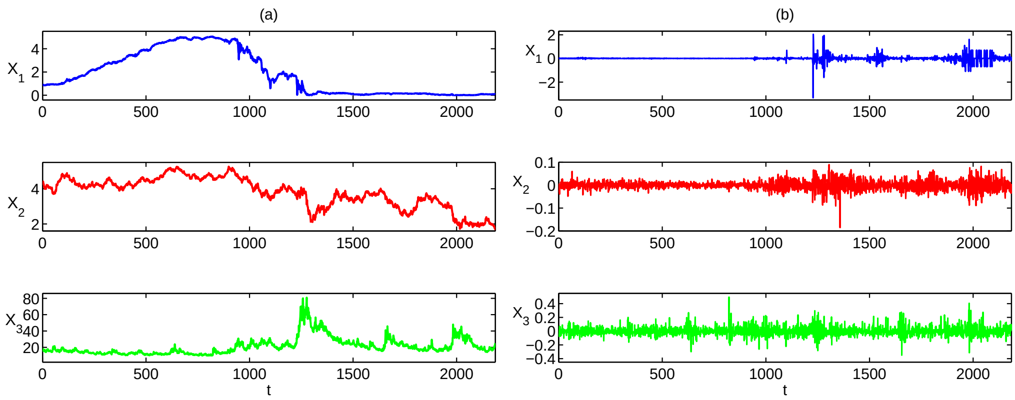

Figure 3: Time series of (a) original prices and (b) the returns of the economic variables.

We highlight here, that the linear Granger causality test is a preliminary step in order to ensure that VAR filtering will not destroy structure in case that no linear causality is present. Therefore, the CGCI is estimated on the returns and the VAR residuals, since it cannot be directly applied to the original data (prices) because of the stationarity condition. The time series of returns and the filtered returns are stationary in level. For the estimation of the CGCI, the BIC suggests the order models $P = 1$ and 2, while the AIC gives much larger order models varying from $P = 7$ to 14 for the different data sets. Since the AIC usually overestimates the order models and the CGCI is not robust for large order models, we set $P = 1$ up to 5.

As expected, the CGCI indicates no causal effects after the VAR filtering. Based on the returns, it recognises the couplings $X_1 \to X_2$, $X_1 \to X_3$, $X_2 \to X_1$ for different P values (see Table 9). For larger P values, fewer couplings are emerged, e.g. for $P = 6$ to 10, only the causal effect $X_1 \to X_3$ is found for the return series.

Table 9: Direct causal effects based on the CGCI values for the financial application.

CGCI	returns
$P = 1$	$X_1 \to X_3$, $X_2 \to X_1$
$P = 2$	$X_1 \to X_2$, $X_1 \to X_3$, $X_2 \to X_1$
$P = 3$	$X_1 \to X_2$, $X_1 \to X_3$, $X_2 \to X_1$
$P = 4$	$X_1 \to X_2$, $X_1 \to X_3$
$P = 5$	$X_1 \to X_3$

The PSTE is estimated on the original prices of the economic variables, the returns, as well as the VAR filtered returns for $m = 2$ and 3. It consistently indicates the direct causal effect $X_2 \to X_1$ for all data sets for $m = 2$. Only for the VAR residuals, the additional coupling $X_3 \to X_1$ is obtained. For $m = 3$, except $X_2 \to X_1$ and $X_3 \to X_1$, for the VAR filtered returns, no causal effects are detected for the rest of data sets (see Table 10).

Table 10: Direct causal effects based on the PSTE values for the financial application.

PSTE	$m = 2$	$m = 3$
prices	$X_2 \to X_1$	-
returns	$X_2 \to X_1$	-
VAR filtered returns	$X_2 \to X_1$, $X_3 \to X_1$	$X_2 \to X_1$, $X_3 \to X_1$

5 Conclusions

The PSTE is a nonlinear measure designed to detect only direct causal effects. It has been shown to be effective for identifying causality when nonlinear interrelations are present, for both stationary and non-stationary systems in mean and variance. It is also not affected by the presence of outliers, since it uses ranks from the reconstructed vectors of the data and not the time series values. On the other hand, the PSTE requires large time series lengths in order to attain high power. Therefore, the stability of the results based on the PSTE is expected to be lost by increasing m, unless large data sets are considered (see [24]).

In contrast, the CGCI has a poor performance in case of nonlinear causal couplings as expected. The present study showed the inadequacy of CGCI also in the presence of long tails and outliers.

When real time series are used, the PSTE indicates that the 10-year Treasury bond drives the short-term interest rate ($X_2 \to X_1$). This dominant relationship is not affected by the non-stationarity of data (in the case of using original prices). Besides, the PSTE does not indicate any loss of information by taking their first logarithmic differences. The suggested relationship $X_2 \to X_1$, that the PSTE highlights, is nonlinear since it is achieved on VAR residual series. The consistency of the empirical finding over the short and long-term via the use of both returns and prices emphasizes the direct impact of expectations in the design of monetary policy.

It is well documented that financial time series are prone to present stylized facts such as non-stationarity in mean or in variance, heteroscedasticity, nonlinearity and outliers [2, 21]. When stationarity condition does not hold, data are treated before estimating Granger causality measures. However, in line with [26, 4] it has been shown that filtering can lead to spurious results when applying Granger causality measures. It turns out that the PSTE is a measure performing well with either non-stationary or stationary data in mean and variance. As such, it constitutes a powerful tool when real data with complex underlying properties are studied.

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*, 19:716–723, 1974.

[2] C. Alexander. *Practical Financial Econometrics*. John Wiley and Sons, Ltd, 2008.

[3] M. Baghli. A model-free characterization of causality. *Economics Letters*, 91:380–388, 2006.

[4] L. Barnett and A.K. Seth. Behaviour of Granger causality under filtering: Theoretical invariance and practical application. *Journal of Neuroscience Methods*, 201:404–419, 2011.

[5] T. Bossomaier, L. Barnett, and M. Harre. Information and phase transitions in socio-economic systems. *Complex Adaptive Systems Modeling*, 2013. in press.

[6] P. T. Brandt and J. T. Williams. *Multiple Time Series Models*, chapter 2, pages 32–34. Sage Publications, 2007.

[7] J. Cheng, L. Taylor, and W. Weng. The links between international parity conditions and Granger causality: A study of exchange rates and prices. *Applied Economics*, 42:3491–3501, 2010.

[8] R.F. Engle and C.W.J. Granger. Cointegration and error correction: Representation, estimation and testing. *Econometrica*, 5:251–276, 1987.

[9] J. Geweke. Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association*, 77(378):304–313, 1982.

[10] G. Gomez-Herrero. *Brain Connectivity Analysis with EEG*. PhD thesis, 2010.

[11] B. Gourévitch, R. Le Bouquin-Jeannés, and G. Faucon. Linear and nonlinear causality between signals: Methods, examples and neurophysiological applications. *Biological Cybernetics*, 95:349–369, 2006.

[12] J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424–438, 1969.

[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. Karagianni and C. Kyrtsou. Analysing the dynamics between US inflation and Dow Jones index using nonlinear methods. *Studies in Nonlinear Dynamics and Econometrics*, 15(2):4, 2011.

[15] C.S. Kim and J.Y. Park. Cointegrating regressions with time heterogeneity. *Econometric Reviews*, 29:397–438, 2010.

[16] A. M. Kowalski, M. T. Martin, A. Plastino, and L. Zunino. Information flow during the quantum-classical transition. *Physics Letters A*, 374(17–18):1819–1826, 2010.

[17] S.-W. Ku, U. Lee, G.-J. Noh, I.-G. Jun, and G. A. Mashour. Preferential inhibition of frontal-to-parietal feedback connectivity is a neurophysiologic correlate of general anesthesia in surgical patients. *PLoS ONE*, 6(10):e25155, 10 2011.

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

[19] D. Kugiumtzis. Partial transfer entropy on rank vectors. *European Physical Journal Special Topics*, 2013. to appear.

[20] C. Kyrtsou. Don't bleach highly complex data: A multivariate study, 2005. Mimeo, University of Macedonia, Thessaloniki, Greece.

[21] C. Kyrtsou and A.G. Malliaris. The impact of information signals on market prices when agents have non-linear trading rules. *Economic Modelling*, 26(1):167–176, 2009.

[22] H.Y. Lee, K.S. Lin, and J. L. Wu. Pitfalls in using Granger causality tests to find an engine of growth. *Applied Economics Letters*, 9:411–414, 2002.

[23] M. Martini, T. A. Kranz, T. Wagner, and K. Lehnertz. Inferring directional interactions from transient signals with symbolic transfer entropy. *Physical Review E*, 83:011919, 2011.

[24] A. Papana, C. Kyrtsou, D. Kugiumtzis, and C. Diks. Direct causality in multivariate time series, 2013. manuscript.

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

[26] T. Sargent. *Macroeconomic Theory, 2nd Edition*. Academic Press, San Diego, 1987.

[27] G. Schwartz. Estimating the dimension of a model. *The Annals of Statistics*, 5(2):461–464, 1978.

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

[29] G. Sugihara, R. May, H. Ye, C.-H. Hsieh, E. Deyle, M. Fogarty, and S. Munch. Detecting causality in complex ecosystems. *Science*, 2012.

[30] W. W. S. Wei. *Time Series Analysis Univariate & Multivariate Methods (Second Edition)*. Addison-Wesley, 2006.

[31] K.L. Xu and P.C.B. Phillips. Adaptive estimation of autoregressive models with time-varying variances. *Journal of Econometrics*, 142:265–280, 2008.

[32] G.-H. Yu and C.-C. Huang. A distribution free plotting position. *Stochastic Environmental Research And Risk Assessment*, 15(6):462–476, 2001.