

# Towards a sharp estimation of transfer entropy for identifying causality in financial time series

Àlex Serès<sup>1</sup>, Alejandra Cabaña<sup>1</sup>, and Argimiro Arratia<sup>2</sup>

<sup>1</sup> Universitat Autònoma de Barcelona, Mathematics, Barcelona, SPAIN,  
alejandro.seres@e-campus.uab.cat, acabana@mat.uab.cat

<sup>2</sup> Universitat Politècnica de Catalunya, Computer Science, Barcelona, SPAIN  
argimiro@cs.upc.edu

**Abstract.** We present an improvement of an estimator of causality in financial time series via transfer entropy, which includes the side information that may affect the cause-effect relation in the system, i.e. a conditional information-transfer based causality. We show that for weakly stationary time series the conditional transfer entropy measure is non-negative and bounded below by the Geweke’s measure of Granger causality. We use  $k$ -nearest neighbor distances to estimate entropy and approximate the distribution of the estimator with bootstrap techniques. We give examples of the application of the estimator in detecting causal effects in a simulated autoregressive stationary system in three random variables with linear and non-linear couplings; in a system of non stationary variables; and with real financial data.

## 1 Introduction

The determination of cause-effect relation among financial time series poses several research challenges: besides the proper detection of the causality, it is important to quantify its strength and the effects of side information that may be present in the system. Also, considering well known observed stylized facts about time series, the measure of causality should be sensible to detection of possible non linear dependence, and adaptable to non stationary time series. In econometrics the standard tool for testing statistical causality is the one proposed by Granger [6, 5], for the bivariate case, and extended by Geweke [3] to a conditional Granger causality test. These tests assume a linear relationship among the causes and effects since it is implemented by fitting autoregressive models. More recently there have been several approaches to testing causality based on non parametric methods, kernel methods and information theory, among others, in order to cope with non linearity and non stationarity, e.g., [2, 17, 11, 1, 16].

The purpose of this work is to contribute to the enhancement of an estimator of bivariate causality proposed in [16], and implemented as software tool TRENTOOL, which has been shown to be robust in detecting the direction and measuring the strength of causality in complex biological systems. Our work and contributions is structured as follows. In Section 2 we present our modified

version of Wibral et al. transfer entropy based causality test by extending the definition to a conditional causality test, and thus, accounting for side information. In Section 3 we show that this conditional transfer entropy is a measure of statistical causality in the same sense as Granger causality (i.e. that the causes precede the effects). Section 4 describes the general steps to estimate transfer entropy. We use the same technique of  $k$ -nearest neighbor distances of [8] for estimating mutual information, but we make use of a bootstrap technique suited for stationary time series [12, 10], for approximating the distribution of the estimator. As a proof of concept, in Section 5, we report three applications of the modified causality measure. Two with simulated data, one to assess the sensitivity of the causality test to linear and non linear couplings in a system of three variables, the other composed of two non stationary variables. The third experiment is our reality check where we use real financial data: we try to determine the possible influence of a market index to another by testing for causality among the German DAX index and the Spanish IBEX.

Due to space restrictions we omit several technical details in this report. These can be found in the extended document [14].

## 2 Granger Causality and Transfer Entropy

Let  $X = \{X_t\}$ ,  $Y = \{Y_t\}$ ,  $Z = \{Z_t\}$  be three stochastic processes defined on a common probability space, from which we try to infer a causal interaction between  $X$  and  $Y$ , and where  $Z$  represents the side information to complete the system. We denote the realization of these random variables at time  $t$  as  $x_t, y_t, z_t$ . Further we use  $\mathbf{X}_t, \mathbf{Y}_t$  and  $\mathbf{Z}_t$  to denote the state space vectors that characterize the processes at time  $t$ , in this case we choose the whole collection of random variables up to time  $t$ , to follow closely the definition given in [16].

**Conditional Granger causality.** It is said that  $X$  does not (Granger) cause  $Y$ , relative to the side information,  $Z$ , if for all  $t \in \mathbb{Z}$ ,  $P(Y_t | \mathbf{X}_{t-1}, \mathbf{Y}_{t-k}, \mathbf{Z}_{t-1}) = P(Y_t | \mathbf{Y}_{t-k}, \mathbf{Z}_{t-1})$ , where  $k \in \mathbb{N}$  is the lag, and  $P(\cdot | \cdot)$  stands for conditional probability distribution. In the bivariate case (unconditional causality) the side information is omitted. To determine conditional Granger causality the following vector regression models are considered

$$\begin{aligned} Y_t &= L_Y(\mathbf{Y}_{t-1}) + L_{XY}(\mathbf{X}_{t-1}) + L_{ZY}(\mathbf{Z}_{t-1}) + \epsilon_{Y,t} \\ Y_t &= \tilde{L}_Y(\mathbf{Y}_{t-1}) + \tilde{L}_{ZY}(\mathbf{Z}_{t-1}) + \tilde{\epsilon}_{Y,t} \end{aligned}$$

where  $L_Y, L_{XY}, L_{ZY}, \tilde{L}_Y, \tilde{L}_{ZY}$  are linear functions and  $\epsilon_{Y,t}, \tilde{\epsilon}_{Y,t}$  are the residuals of the regression of  $Y_t$  on  $\mathbf{X}_t$  and information  $\mathbf{Z}_t$ , and without  $\mathbf{X}_t$ . Then one can quantify the usefulness of including  $\mathbf{X}$  in explaining  $Y$  using Geweke's test based on the variances of the residuals [3]:  $\mathcal{F}_{X \rightarrow Y | Z} = \log \frac{\text{var}(\tilde{\epsilon}_{Y,t})}{\text{var}(\epsilon_{Y,t})}$ .

Note that the residual variance of the second regression will always be larger or equal to the one from the first, so  $\mathcal{F}_{X \rightarrow Y | Z} \geq 0$ . For its statistical treatment,

it is known that the corresponding maximum likelihood estimator will have a  $\chi^2$  distribution under the null hypothesis  $\mathcal{F}_{X \rightarrow Y|Z} = 0$ , and a non-central  $\chi^2$  distribution under the alternative hypothesis  $\mathcal{F}_{X \rightarrow Y|Z} > 0$  [5, 3].

**Transfer entropy.** The transfer entropy of process  $X$  to  $Y$ , conditioned to the side information  $Z$ , can be obtained from mutual information as

$$TE_{X \rightarrow Y|Z} = I(Y^+; X^- | Y^- \oplus Z^-)$$

where  $A^-$  denotes all information available of the past of  $A$ ,  $A^+$  its immediate future, and  $A \oplus B$  represents the concatenation of random vectors  $A$  and  $B$ . Wibral et al. in [16] argue extensively on the inadequacy of this formula in the bivariate case for truly capturing cause-effect relations, and go on to propose a *transfer entropy with self prediction optimality at lag  $u$* , which extended to account for the side information  $Z$ , has the form:

$$TE_{SPO:X \rightarrow Y,u|Z} = I(Y_t; \mathbf{X}_{t-u} | \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1})$$

Then they show,

**Theorem 1.** *For two discrete-time random processes  $X, Y$  with a state-space representation  $\mathbf{X}, \mathbf{Y}$  coupled from  $X$  to  $Y$  via a non-zero delay  $\delta$ ,  $TE_{SPO:X \rightarrow Y,u|Z}$  is maximal for  $u = \delta$ . This also holds in presence of additional coupling from  $Y$  to  $X$ .  $\square$*

This approach suits every requisite that we need to correctly estimate information transfer, but in practice the associated estimator resulted in an unreliable predictor due to the amount of information about  $X_{t-\delta}$  present in  $\mathbf{X}_{t-u}$ , for  $1 < u < \delta$ , which causes it to assign a larger value to the information “rebounds” produced by the approximations assumed. (By rebound it is meant the contribution of information due to indirect causality, e.g.  $X \rightarrow Z$  and  $Z \rightarrow Y$ .) This can be corrected by using the state space representation for the target  $Y_t$ :

$$TE_{SPO:X \rightarrow Y,u|Z} = I(\mathbf{Y}_t; \mathbf{X}_{t-u} | \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}) \quad (1)$$

By the properties of the state-space representation we are not losing any condition that we imposed to the former estimator, and we obtain an advantage: we are compensating the rebound effect of  $X_{t-\delta}$  in  $\mathbf{X}_{t-u}$ , because we swap the numerical value of  $Y_t$  for a state vector that contains information of the past which, supposing that the interaction lag is constant for all  $t$  considered, accounts for a more precise value for the unwanted rebounds, with a clear apex at the true delay. On the other hand we have a disadvantage, related to the method used to estimate conditioned densities. As a result of using an element of a larger dimension we get a bigger error in the estimation.

### 3 Transfer Entropy is a measure of Causality

Barnett et al. in [1] showed that for random variables with a Gaussian distribution, measuring transfer entropy is equivalent to measuring Granger causality.

This result easily extends to the conditional causality concept. Formally, it can be shown that  $\mathcal{F}_{X \rightarrow Y|Z} = 2TE_{X \rightarrow Y|Z}$ .

Nevertheless, we want to evaluate our capacity to infer causality against financial data, which is widely known to behave neither as stationary nor to be Gaussian. For this motive it is very important to confirm that when measuring information transfer between processes we are, indeed, testing for statistical causality. We have not found any explicit proof of this proposition, but there are some signs in the literature. One is an argument asserting that non-zero Granger causality means non-zero transfer entropy [11], but it is a bit vague. Next is an indication for a lower bound for mutual information [8], which resulted really helpful in proving that for weakly stationary variables, not necessarily Gaussian, the linear influences detected with Granger causality are no more than the ones detected for Gaussian variables, and therefore we will detect them at the very least. We state without proof the following proposition,

**Proposition 1.** *Given two stochastic processes  $\{X_t\}, \{Y_t\}$ , and supposing that  $X$  causes  $Y$  in the Granger sense, then  $TE_{X \rightarrow Y|Z} \geq 0$   $\square$*

Now, conditional transfer entropy relates to Granger causality as follows:

**Theorem 2.** *Let  $\{X_t\}, \{Y_t\}, \{Z_t\}$  be three jointly distributed, weakly stationary stochastic processes, defined in a common probability space. Then the standard measure of Granger causality and transfer entropy are related by:*

$$2TE_{X \rightarrow Y|Z} \geq \mathcal{F}_{X \rightarrow Y|Z}$$

*Proof.* It suffices to prove that conditional mutual information is minimized by Gaussian distributions. Since  $TE_{X \rightarrow Y|Z}$  is only a specific form of a conditional mutual information the result will follow.

As proposed in [8] we will set a minimization problem on the mutual information for continuous variables using Lagrange multipliers. If we use  $\mu(x, y) = \mu'(x, y|z)$ , then the formula that we need to minimize does not change:

$$\begin{aligned} I(X; Y|Z) &= \iint \mu'(x, y|z) \log \frac{\mu'(x, y|z)}{\mu'_x(x|z)\mu'_y(y|z)} dx dy \\ &= \iint \mu(x, y) \log \frac{\mu(x, y)}{\mu_x(x)\mu_y(y)} dx dy \end{aligned}$$

Recall that  $\iint \mu(x, y) dx dy = 1$ ,  $\mu_x(x) = \int \mu(x, y) dy$  and  $\mu_y(y) = \int \mu(x, y) dx$  and that  $C$  is the covariance matrix of the variables  $X, Y$  conditioned on  $Z$ , so that  $\int x^2 dx \mu_X(x) = C_{xx} = \sigma_x$ ,  $\int y^2 dy \mu_Y(y) = C_{yy} = \sigma_y$ , and  $\iint xy dx dy \mu(x, y) = C_{xy}$ . Then the Lagrangian can be written as:

$$\begin{aligned} \mathcal{L} &= - \iint \mu(x, y) \log \frac{\mu(x, y)}{\mu_x(x)\mu_y(y)} dx dy - \lambda_1 \left( \iint \mu(x, y) dx dy - 1 \right) \\ &\quad - \lambda_2 \left( \iint x^2 \mu(x, y) dx dy - \sigma_x \right) - \lambda_3 \left( \iint y^2 \mu(x, y) dx dy - \sigma_y \right) \\ &\quad - \lambda_4 \left( \iint xy \mu(x, y) dx dy - \sigma_{xy} \right) \end{aligned}$$

Now impose  $\frac{\partial \mathcal{L}}{\partial \mu(x,y)} = 0$ , since the first and second order moments are finite, we can swap the partial derivative with the integral sign, thus obtaining:

$$0 = - \iint \frac{\partial}{\partial \mu(x,y)} \left[ \mu(x,y) \log \frac{\mu(x,y)}{\mu_x(x)\mu_y(y)} + \lambda_1 \mu(x,y) + \lambda_2 x^2 \mu(x,y) + \lambda_3 y^2 \mu(x,y) + \lambda_4 xy \mu(x,y) \right] dx dy + \frac{\partial}{\partial \mu(x,y)} \left[ \lambda_1 + \lambda_2 \sigma_x + \lambda_3 \sigma_y + \lambda_4 \sigma_{xy} \right],$$

which is zero if the derivative inside the integral is zero, as it is positive. Then:

$$0 = - \log \frac{\mu(x,y)}{\mu_x(x)\mu_y(y)} - \mu(x,y) \frac{1}{\frac{\mu(x,y)}{\mu_x(x)\mu_y(y)}} \frac{1}{\mu_x(x)\mu_y(y)} - \lambda_1 - \lambda_2 x^2 - \lambda_3 y^2 - \lambda_4 xy$$

and hence  $\mu(x,y) = \mu_x(x)\mu_y(y)e^{-(1+\lambda_1)}e^{-\lambda_2 x^2 - \lambda_3 y^2 - \lambda_4 xy}$ , where  $\lambda_1, \lambda_2 (\geq 0), \lambda_3 (\geq 0), \lambda_4$  are constants fixed by the constraints. Therefore a probability density that follows this equation is consistent with a Gaussian. To prove that it is the only one, integrate the expression of  $\mu(x,y)$  over  $y$  and make the change of variables  $x = -i\zeta/c$ :

$$\begin{aligned} \int \mu(x,y) dy &= \int \mu_x(x)\mu_y(y)e^{-(1+\lambda_1)}e^{-\lambda_2 x^2 - \lambda_3 y^2 - \lambda_4 xy} dy \Rightarrow \\ \mu_X(x) &= \int \mu_x(x)\mu_y(y)e^{-(1+\lambda_1)}e^{-\lambda_2 x^2 - \lambda_3 y^2 - \lambda_4 xy} dy \Rightarrow \\ e^{(1+\lambda_1)}e^{\lambda_2 x^2} &= \int \mu_y(y)e^{-\lambda_3 y^2 - \lambda_4 xy} dy \Rightarrow \\ e^{(1+\lambda_1)}e^{-\lambda_2 \zeta^2/c^2} &= \int e^{i\lambda_4 y\zeta/c} [\mu_Y(y)e^{-\lambda_3 y^2}] dy \end{aligned}$$

The left hand side of this last equality is a Gaussian density function, and the right hand side, its Fourier transform,  $\mu_Y(y)e^{-\lambda_3 y^2}$ . Since the Fourier transform of a Gaussian is another Gaussian, and  $\mu_Y(y)$  cannot be a constant if it is a density function, then  $\mu_Y(y)$  is also Gaussian. The same argument (integrating over  $x$ ) shows that  $\mu_X(x)$  is Gaussian, thus the joint distribution of the random variables  $X, Y$  that minimizes its mutual information conditioned to  $Z$ ,  $\mu(x,y) = \mu'(x,y|z)$ , is a Gaussian distribution.  $\square$

## 4 Practical estimation of transfer entropy

In this section we outline our methodology used to estimate transfer entropy. Our methodology follows the procedures outlined in [16], but we introduce some key modifications: aside from our particular estimator which deals with conditional mutual information, we test for null transfer entropy by obtaining a bootstrap approximation of the distribution of the estimator, a non parametric procedure.

**The state-space representation.** First we need to reconstruct the corresponding state-space of the interacting systems from a scalar time series. For this purpose we use Takens delay embedding [15] with the parameters suggested by Ragwitz [13]. Delay embedding states of the systems under consideration can be written as delay vectors of the form:  $\mathbf{X}_t^{(m)} = (X_t, X_{t-\tau}, X_{t-2\tau}, \dots, X_{t-(m-1)\tau})$ , where  $m$  and  $\tau$  denote the embedding dimension and Takens' embedding delay respectively.

**Embedding lag.** To enhance the predictability between the state variables it is key to choose the correct variables. Ragwitz' criterion yields delay embedding states that provide optimal self prediction for a large class of systems. It is based in a close study of the properties of an assumed multivariate Langevin process from which the data proceeds, assuming that the univariate series behaves locally like a Markov chain.

As for the delay parameter  $\tau$ , in the limit of infinite precision of the data and infinitely long time series, all values are equivalent. In a practical situation, however, a good choice of the delay is crucial. If  $\tau$  is too large, successive elements of the embedding vector are almost independent and the vectors fill an amorphous cloud in the state-space. If  $\tau$  is too small, successive elements of the embedding vector are strongly correlated and all vectors are clustered around the "diagonal" of the state space. We will need to search for meaningful neighbors of these vectors in the next two steps, and in both extreme cases they are difficult to obtain. Then, the simplest reasonable estimate of an optimal delay is the first zero of the autocorrelation function  $\rho(k) = \gamma(k)/\gamma(0)$ , where  $\gamma(k) = Cov(X_t, X_{t+k})$ ,  $k \in \mathbb{Z}$ . To discern the first zero, our program performs the test

$$H_0 : \rho(k) = 0 \quad \text{versus} \quad H_1 : \rho(k) \neq 0$$

**Embedding dimension.** To select the dimension of the vectors of the reconstructed state-space we use the false nearest neighbors method [9, 7]. This method compares distances of  $m$ -valued state-space vectors to its  $k$ -nearest neighbor ( $knn$ ) with distances for  $(m + 1)$ -valued state-space vectors from the same data. These distances are really easy to calculate considering that all vectors have a similar form once we have the estimated delay. Only the  $knn$  search may prove to be a problem if the dimension becomes too big. If the distances grow too much with the growth in dimension, the neighbors are deemed as false, and we try again increasing the dimension by one, until no false neighbor is found. In this way the data itself dictate the embedding dimension, in a way that it approximates correctly the dimension of the manifold in which the states of the system reside.

**Mutual Information estimation** Once we have the parameters involved in the state space representation of both uni-valued series of data, we apply the formula (1) to find the transfer entropy from  $\mathbf{X}$  to  $\mathbf{Y}$ :

$$TE_{SPO:X \rightarrow Y,u} = I(\mathbf{Y}_t; \mathbf{X}_{t-u} | \mathbf{Y}_{t-1}) = I(\mathbf{Y}_t; \mathbf{X}_{t-u}, \mathbf{Y}_{t-1}) - I(\mathbf{Y}_t; \mathbf{Y}_{t-1})$$

(We are not considering a third variable  $\mathbf{Z}_t$  symbolizing side information to simplify notation, the inclusion of which would mean four mutual information terms in the previous formula instead of two.)

To estimate the mutual information  $I(A; B)$  between two multivariate time series,  $A$  and  $B$ , we use the  $k$ -nearest neighbor estimator suggested in [8]. When we use this method the only assumption behind is a certain smoothness of the underlying probability functions, so it fits in as a non-parametric technique. This estimator has the form

$$\hat{I}(A, B) = \Psi(k) - \langle \Psi(n_A + 1) + \Psi(n_B + 1) \rangle + \Psi(N)$$

where  $N$  is the length of the series used,  $n_A$  is the number of points in  $A$  whose pairwise distance is below a certain given threshold (similarly define  $n_B$ ), and  $\Psi(x)$  is the digamma function:  $\Psi(x) = \Gamma(x)^{-1} \frac{d\Gamma(x)}{dx}$ . The digamma function satisfies the recursion  $\Psi(x + 1) = \Psi(x) + 1/x$  and  $\Psi(1) = -C$ , with  $C$  being the Euler-Mascheroni constant.

**Bootstrap sampling** Once we have the estimators of the Transfer Entropy between our time series as a function of the lag, we approximate their distribution via a bootstrap experiment. Take bootstrap replications of the series and recompute the estimators as usual. Bootstrapping stationary time series has been widely studied in [12] and [10] and we follow their procedures. This allows us to assess (for each lag) whether there is no causation (transfer entropy equal to zero) by looking at the bootstrap confidence intervals. We think that this is a more natural way of approximating the distribution of the statistic of the transfer entropy estimator, than the test conducted in [16] which consist in comparing the observed TE with that of surrogate data obtained by permutations of the source.

## 5 Simulations and applications in finance

We present several experiments that we have performed to assess the efficiency of our method to estimate transfer entropy, as well an application to real data.

**Example 1.** Consider an autoregressive stationary system in three variables with one linear coupling ( $Y \rightarrow Z$ ), and two non-linear ones ( $X \rightarrow Y$  and  $X \rightarrow Z$ ), proposed in [4]. We will use different lags in each interaction to be able to clearly differentiate them, as well as the “rebounds” produced in an indirect cause, specifically  $\delta_{X \rightarrow Y} = 10$ ,  $\delta_{X \rightarrow Z} = 5$ ,  $\delta_{Y \rightarrow Z} = 15$ . The system satisfies:

$$\begin{aligned} X_t &= 3.4X_{t-1}(1 - X_{t-1})^2 e^{X_{t-1}^2} + 0.4\epsilon_{1,t} \\ Y_t &= 3.4Y_{t-1}(1 - Y_{t-1})^2 e^{Y_{t-1}^2} + 0.5X_{t-10}^2 + 0.4\epsilon_{2,t} \\ Z_t &= 3.4Z_{t-1}(1 - Z_{t-1})^2 e^{Z_{t-1}^2} + 0.3Y_{t-15} + 0.5X_{t-5}Z_{t-1} + 0.4\epsilon_{3,t} \end{aligned}$$

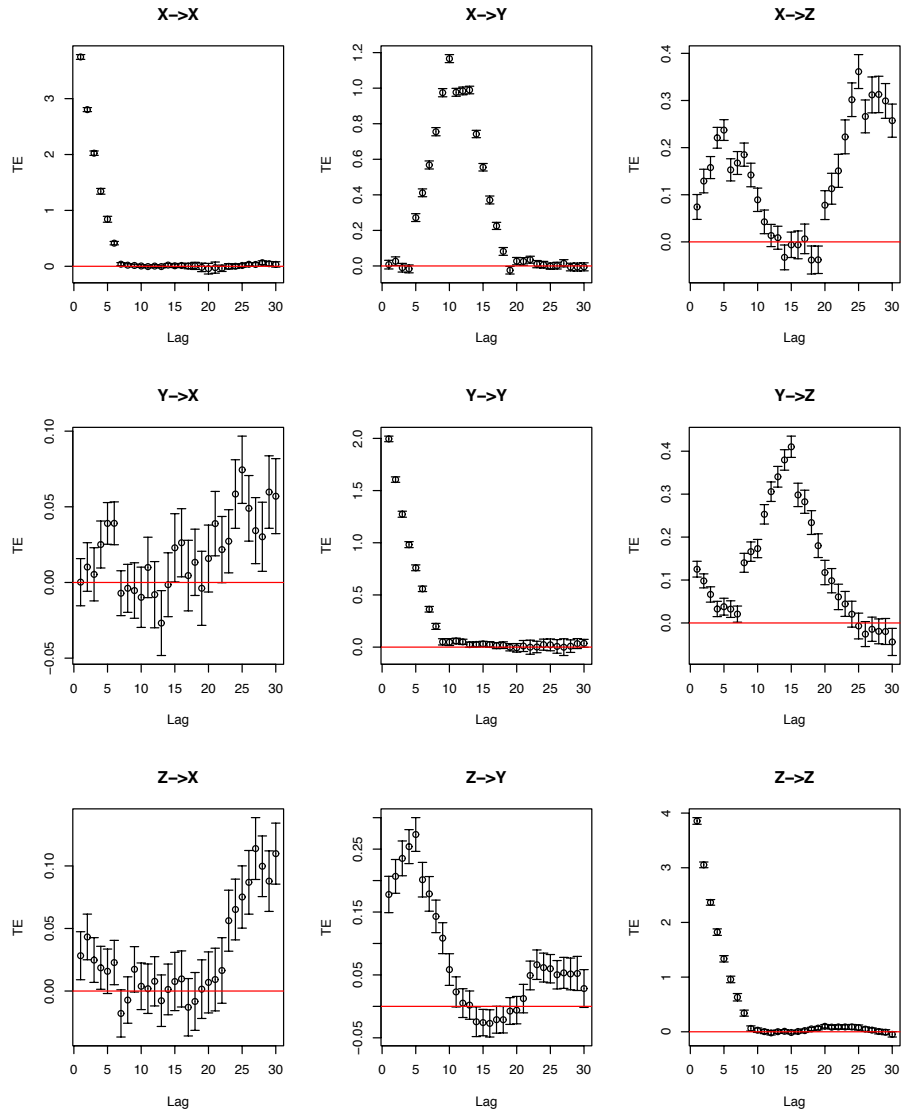
where  $\epsilon_{1,t}, \epsilon_{2,t}, \epsilon_{3,t}$  are Gaussian white noise terms with identity covariance matrix. To simulate it, we have taken the 30 initial terms of each series as 0, and begin applying the formula from there for a total of  $10^4$  steps, and the last  $10^3$  were used to test the transfer entropy estimator, as the arbitrary initial conditions should have little effect then. We have performed the geometric time series bootstrap 60 times for each lag, with average sizes of length 100. We then have drawn 90% confidence intervals. We have estimated up to lag 30 to save computation time, as it should be enough to recover the causal interactions. The results are presented in Figure 1.

Observe that we have recovered all the true delays as clearly distinct maximums of the transfer entropy versus the lag plots. Its important to highlight that the scale in the  $y$ -axis, that is the transfer entropy magnitude, is different in each graph, as our estimator depends on the embedding dimension, so each test is unique and we must look at the error bars generated in the bootstrap process to look at it with perspective. That being said, variation in TE depending on extremely different embedding dimension reaches a 30% of the larger value, and in this graph we can make conclusions about the strength of the interactions as we see differences of an order of magnitude. That is why, as expected, we can consider null the  $Z \rightarrow X$  and  $Y \rightarrow X$  interactions, as the measured TE for the lag values statistically different to zero is really small. There is a hint of a pattern, but it should be no more than a residual influence in the estimator, motivated by the real influence in the opposite direction. The  $Z \rightarrow Y$  interaction is also fairly small, but really interesting:  $Z$  does not influence directly  $Y$ , but  $X$  influences both, first  $Z$  at  $u = 5$  and  $Y$  at  $u = 10$ , therefore  $Z_{t-5}$  has important information on  $Y$  via  $X$ . Here we have a good example of when we would need to apply the whole TE formula, conditioning to a known overarching series.

On the other hand, the autoregressive interaction of all the variables, represented in the diagonal, has been recovered at a maximum in the  $u = 1$  lag, with decreasing intensity as the influence of the past of the series declines and stabilizing at 0 after. The  $X \rightarrow Y$  interaction is prominently recovered in a similar fashion, at  $u = 10$ , also showing the decreasing remnants of information still present at values around the true delay. The linear interaction  $X \rightarrow Y$  is also displayed with a maximum at  $u = 15$ . And the interaction  $X \rightarrow Z$  shows two interesting properties of our estimator, aside of successfully detecting the influence at  $u = 5$  with a distinct local maximum. We see how we detect a maximum at  $u = 25$ , which corresponds to the interaction  $X \rightarrow Y \rightarrow Z$ , which means that we successfully detect indirect interactions, as it displays a strength similar to  $Y \rightarrow Z$ . The other important detail is to check how the variances of the indirect interaction are bigger than those of the direct one.

We repeat the same analysis using a linear Granger causality estimator, provided in the R library `MSBVAR`. We used the regressions up to lag  $u = 15$ , as we did not want the indirect influences to have enough steps to affect the test over the direct interactions that are completed at this lag. The results obtained are shown in Table 1.





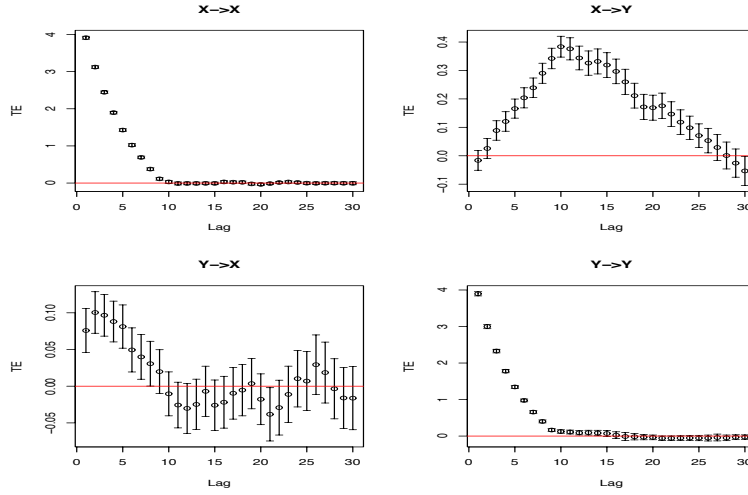
**Fig. 1.** Graphs representing the estimated transfer entropy depending on the lag of the supposed cause, for each of the nine possible interactions of the system.

The test recovers swiftly the  $Y \rightarrow Z$  linear interaction, and also the  $X \rightarrow Y$  quadratic interaction, but fails at detecting any interaction in  $X \rightarrow Z$ , which is highly non-linear at true lag  $u = 5$ . To better check the responsiveness of the test we have repeated it considering various smaller lag values for  $X$ , and we

**Table 1.** Granger test statistics and  $p$ -values for different interactions in the system

	F-statistic	p-value
$Y \rightarrow X$	1.0111734	0.44062932
$Z \rightarrow X$	1.5610944	0.07815942
$X \rightarrow Y$	278.3554360	0.00000000
$Z \rightarrow Y$	1.2341100	0.23930984
$X \rightarrow Z$	0.7986152	0.68012265
$Y \rightarrow Z$	18.8199782	0.00000000

never obtained a favorable  $p$ -value. Our TE test shows the various interactions in the system (direct or indirect), justifying its appropriateness when dealing with non-linear systems.  $\square$



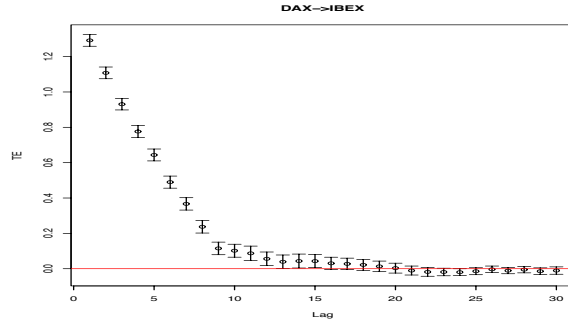
**Fig. 2.** Graphs representing the estimated transfer entropy depending on the lag. This time both variables are non stationary.

**Example 2.** Next we propose a similar (but non stationary) system. Consider two processes ( $X \rightarrow Y$ ) with a true delay  $\delta_{X \rightarrow Y} = 10$  and random variance that follows an IGARCH(1,1) model:

$$\begin{aligned}
 X_t &= 0.7X_{t-1} + \sigma_{1,t}\epsilon_{1,t} \\
 Y_t &= 0.3Y_{t-1} + 0.5X_{t-10}Y_{t-2} + \sigma_{2,t}\epsilon_{2,t} \\
 \sigma_{i,t}^2 &= 0.2 + 0.9\epsilon_{i,t-1}^2 + 0.1\sigma_{i,t-1}^2
 \end{aligned}$$

The results are displayed in Figure 2 and the analysis is similar to the previous one. We recover perfectly the autoregressive tendencies, and we obtain a fairly clear view of the  $X \rightarrow Y$  interaction, similar to the  $X \rightarrow Z$  interaction in Example 1, but now it takes a larger time to fade to zero, so we can sense the distortion generated by the variance, although we can still identify the true delay with precision.  $\square$

**Example 3.** As a real world application of our estimator we take a look at the relationship between two major European stock indices, the German DAX-30 and the Spanish IBEX-35, between January 1st, 2011 and June 24th, 2016, that constitute approximately 1200 data points, as we have a daily frequency data feed. We have tested for transfer entropy the log-returns ( $\log(P_t/P_{t-1})$ ), where  $P_t$  is the adjusted close price of index at day  $t$ ) of the DAX to the log-returns of IBEX, as we suppose that the territorially more influential stock market affects the smaller, especially in the current context where Germany holds a major portion of Spain’s debt. The results are represented in Figure 3. We can see a strong influence detected in lag  $u = 1$  that fades quickly in 8 days. It is logical to think that real stock markets react to their sphere of influence quickly enough that daily data does not allow for precise detection of the interaction lag, as we expect it to be much smaller, but we can definitely conclude that there is a causal influence present.  $\square$



**Fig. 3.** *Transfer entropy of the DAX-30 index to the IBEX-35 depending on the lag of the interaction*

## Acknowledgments

A. Cabaña acknowledges support of MINECO projec MTM2015-69493-R.  
A. Arratia acknowledges support of MINECO project APCOM (TIN2014-57226-P), and Gen. Cat. SGR2014-890 (MACDA).

## References

1. Barnett, L., Barnett, A.B., Seth, A.K.: Granger causality and transfer entropy are equivalent for gaussian variables. *Phys. Rev. Lett* 103, 238701 (2009)
2. Diks, C., Wolski, M.: Nonlinear granger causality: Guidelines for multivariate analysis. *Journal of Applied Econometrics* (2015)
3. Geweke, J.: Measurement of linear dependence and feedback between multiple time series. *J. J. Amer. Stat. Assoc.* 77, 304–313 (1982)
4. Gourévitch, B., Bouquin-Jeannès, R.L., Faucon, G.: Linear and nonlinear causality between signals: methods, examples and neurophysiological applications. *Biological Cybernetics* 95, 349–369 (2006)
5. Granger, C.W.J.: Testing for causality: A personal viewpoint. *J. Econ. Dyn. Control* 2, 329–352 (1980)
6. Granger, C.: Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438 (1969)
7. Hegger, R., Kantz, H.: Improved false nearest neighbor method to detect determinism in time series data. *Physical Review E* 60(4), 4970 (1999)
8. Karasov, A., Stögbauer, H., Grassberger, P.: Estimating mutual information. *Phy. Rev. E* 69, 066138 (2004)
9. Kennel, M., Brown, R., Abarbanel, H.: Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A* 45(6), 3403–3411 (1992)
10. Kunsch, H.R.: The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* 17, 1217–1241 (1989)
11. Marinazzo, D., Pellicoro, M., Stramaglia, S.: Kernel method for nonlinear granger causality. *Phys. Rev. Lett.* 100, 144103 (2008)
12. Politis, D.N., Romano, J.P.: The stationary bootstrap. *Journal of the American Statistical Association* 89, 1303–1313 (1994)
13. Ragwitz, M., Kantz, H.: Markov models from data by simple nonlinear time series predictors in delay embedding spaces. *Phys. Rev. E* 65, 056201 (2002)
14. Serès-Cabasés, A.: Causality via transfer entropy. Bachelor Sc. Thesis, Universitat Autònoma de Barcelona (2016)
15. Takens, F.: Dynamical systems and turbulence. In: *Lecture Notes in Mathematics*, vol. 898, Warwick 1980 Symp., pp. 366–381. Springer (1981)
16. Wibral, M., Pampu, N., Priesemann, V., Siebenhühner, F., Seiwert, H., Linder, M., Lizier, J., Vicente, R.: Measuring information-transfer delays. *PLoS ONE* 8(2), e55809 (2013)
17. Zaremba, A., Aste, T.: Measures of causality in complex datasets with application to financial data. *Entropy* 16, 2309–2349 (2014)