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

Argimiro Arratia, **Alejandra Cabaña** and
Àlex Serés

Universitat Politècnica de Catalunya, Spain
Universitat Autònoma de Barcelona, Spain

MIDAS Workshop
Riva del Garda, September 2016

Introduction

The determination of cause-effect relations among (financial) time series poses several challenges:

- the proper detection of the causality
- the quantifyfication of its strength, and
- the effects of *side information* that might be present in the system.

In econometrics the standard tool for testing statistical causality are Granger's and Geweke's tests for (conditional) Granger causalitythat assume a linear relation among the causes and effects.

There are several recent approaches to testing causality based on non parametric methods, kernel methods and information theory, among others, that cope with non linearity and non stationarity[1], but disregarding the presence of side information (conditional causality).

We present a modification of Wibral et al. transfer entropy based causality test onto a conditional causality test, and thus, accounting for side information. 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).

[1] Diks& Wolski (2015) *Nonlinear Granger causality...* Jr. Appl. Econometrics; Zaremba et.al. (2014) *Measures of causality in complex datases with applications to financial data*, Entropy 16; Marinazzo et.al (2008), *Kernel method for nonlinear Granger causality*, Phys. Rev. Lett 100; Barnett et.al. (2009) *Granger Causality and TE are equivalent in the Gaussian case*, Phys. Rev. Lett. 103; Wibral et.al.(2013), *Measuring information-transfer delays*, PLoS ONE 8(2)

Let's fix notation

Let X_t, Y_t, Z_t be stochastic processes defined on a common probability space.

We try to infer a causal interaction between X and Y, and where Z represents the side information to complete the system.

X_t will be understood as the random variable associated with time t, and x_t its realization.

Consider $\boldsymbol{X_t}, \boldsymbol{Y_t}$ and $\boldsymbol{Z_t}$ 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, though a finite and well chosen collection would suffice.

On State-Space representations

A state space model for a (possibly multivariate) time series $\{\boldsymbol{\eta}_t, t = 1, 2, \ldots\}$ consists of:
The *state equation* that expresses the observation $\boldsymbol{\eta}_t$ as a linear function of a state variable $\boldsymbol{\xi}_t$, plus noise:

$$\boldsymbol{\eta}_t = G_t \boldsymbol{\xi}_t + W_t, \quad t = 1, 2, \ldots$$

and the *observation equation*, that determines the state vector over time:

$$\boldsymbol{\xi}_{t+1} = F_t \boldsymbol{\xi}_t + V_t, \quad t = 1, 2, \ldots$$

$\{W_t\}$ and $\{V_t\}$ are uncorrelated and uncorrelated to $\boldsymbol{\xi}_1$.

Causality

Causality is an asymmetric, binary relation that connects one process (the cause) to another (the effect).

Wiener principle: "For two simultaneously measured signals, if we can predict the first signal better by using the past information from the second one than by using the information without it, then we call the second signal causal to the first one". This statement highlights two central issues that other definitions of causality abide:

- A *cause* occurs strictly before its *effect* (there is a **lag**)
- The *cause* has unique information about the future values of its *effect*.

Currently the most accepted measure of causality is the one made by Clive Granger, conceived in the context of linear autoregressive models, and was built from these two properties.

We shall say that X does not (Granger) cause Y, relative to the side information, Z, if for all $t \in \mathbb{Z}$ and lag $k \in \mathbb{N}$:

$$P(Y_t | \boldsymbol{X_{t-1}}, \boldsymbol{Y_{t-k}}, \boldsymbol{Z_{t-1}}) = P(Y_t | \boldsymbol{Y_{t-k}}, \boldsymbol{Z_{t-1}})$$

The standard measure of Granger causality is based on comparing

$$Y_t = L_Y(\boldsymbol{Y}_{t-1}) + L_{XY}(\boldsymbol{X}_{t-1}) + L_{ZY}(\boldsymbol{Z}_{t-1}) + \epsilon_{Y,t}$$

where L_Y, L_{XY}, L_{ZY} are linear functions and $\epsilon_{Y,t}$ is the residual, with:

$$Y_t = \widetilde{L}_Y(\boldsymbol{Y_{t-1}}) + \widetilde{L}_{ZY}(\boldsymbol{Z_{t-1}}) + \widetilde{\epsilon}_{Y,t}$$

Then we can quantify the usefulness of including X in explaining Y by comparing the variances of the residuals[2]:

$$\mathcal{F}_{X\to Y|Z} = \log \frac{\mathbf{Var}(\widetilde{\epsilon}_{Y,t})}{\mathbf{Var}(\epsilon_{Y,t})} (\geq 0)$$

The corresponding (ML based) estimator will have (asymptotically) a χ^2 distribution under the null hypothesis

$$\mathcal{F}_{X\to Y|Z} = 0$$

and a non-central χ^2 distribution under the alternative hypothesis $\mathcal{F}_{X\to Y|Z} > 0$.

[2] Geweke, *Mesurements of linear dependence and feedback between multiple time series*, JASA 77 (1982)

Entropy

Entropy is a measure of the disorder of a system.
In information theory, the Shannon entropy $H(X)$ of a random variable X with density (or mass function) $p(x)$ is defined as the expected value of the logarithm of the random variable $1/p(X)$:

$$H(X) = -\mathbf{E}\log(p(X))$$

Likewise, joint and conditional entropy are defined respectively as

$$H(X,Y) = -\mathbf{E}\log(p(X,Y)), \qquad H(X|Y) = -\mathbf{E}\log(p(X|Y))$$

and chain rules are easily deduced, such as

$$H(X,Y) = H(X)+H(Y|X) \quad H(X,Y|Z) = H(X|Z)+H(Y|X,Z)$$

Mutual Information

Given two random variables, their *mutual information* $I(X;Y)$ measures the deviation from the system where both variables are independent, as the Kullback-Lieber distance:

$$I(X;Y) = \mathbf{E} \log \frac{p(X,Y)}{p(X)p(Y)}$$

hence it can be written in terms of entropy in several ways,

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) = H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X,Y) \end{aligned}$$

The conditional mutual information of X and Y given Z is simply defined as

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z)$$

Transfer Entropy

The transfer entropy with self predicting optimality of process X to Y, conditioned to the side information Z, is defined as

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

This is an adaption of the definition in Wibral *et. al.*[3] that allows the incorporation of more information on the target and the presence of side information.

- If X and Y are coupled via a delay δ, then $TE_{SPO:X\longrightarrow Y,u|Z}$ is maximal for $u = \delta$.

[3] *Measuring information-transfer delays*, PLoS ONE 8 (2013)

TE is a measure of causality

- Adapting Barnet & Seth[4] to the conditional case, we see that if $\{X_t\}$,$\{Y_t\}$ and $\{Z_t\}$ are Gaussian Processes,

$$\mathcal{F}_{X\to Y|Z} = 2\,TE_{X\longrightarrow Y|Z}$$

- For $\{X_t\}$,$\{Y_t\}$, if X Granger-causes Y, then

$$TE_{X\longrightarrow Y|Z} \geq 0$$

- Given $\{X_t\}$,$\{Y_t\}$, stationary stochastic processes defined in a common probability space, then the standard measure of Granger Causality and transfer entropy are related by

$$2\,TE_{X\longrightarrow Y|Z} \geq \mathcal{F}_{X\to Y|Z}$$

[4] *Granger Causality and Transfer Entropy are equivalent in the Gaussian case*, PRL (2009

Embedding Parameters

In order to reconstruct the state-space of the system from scalar time series, each of them can be written as a delay vector of the form[5]

$$\boldsymbol{X}_t^{(m)} = (X_t, X_{t-\tau}, X_{t-2\tau}, \ldots, X_{t-m\tau})$$

m is the embedding dimension $\longrightarrow$ *false nearest neighbours* [6]

τ is the embedding delay $\longrightarrow$ first zero of the ACF [7]

---

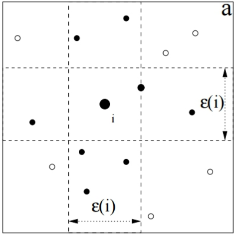
[5] Takens (1981), *Detecting strange attractors in Turbulence*, in LNMath 898, Springer

[6] Hegger & Krantz (1999), *Improved false nearest neighbours method to detect determinism in time series data*, Phys. Rev. E 60(4)

[7] Ragwiz et.a. (2002) *Markov models from data by simple nonlinear time series predictors in delay embedding spaces*, Phys. Rev. E 65

Estimating MI

Following Karasov et.al[8] the MI of random variables A and B is estimated approximating the joint density of the pairs $c_t = (a_t, b_t)$ and that of the marginal densities.



If N is the length of the series, n_A is the number of points in A whose pairwise distance is below certain given threshold (similarly define n_B), and $\Psi(x)$ is the digamma function $\Psi(x) = \Gamma(x)^{-1}\frac{d\Gamma(x)}{dx}$, then ...

[8] *Estimating mutual information*, Phys. Rev E 69 (2004)

mutual information is estimated by

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

Then

$$\widehat{TE}_{SPO\ \ X \longrightarrow Y,u} = \hat{I}(\boldsymbol{Y}_t; \boldsymbol{X}_{t-u}, \boldsymbol{Y}_{t-1}) - \hat{I}(\boldsymbol{Y}_t; \boldsymbol{Y}_{t-1})$$

The distribution of $\widehat{TE}_{SPO\ \ X \longrightarrow Y,u}$ is approximated by (stationary) Boostrap, and confidence intervals are thus obtained. This allows us to assess (for each lag) whether there is no causation (transfer entropy equal to zero).

Example 1

AR stationary system with one linear coupling ($Y \to Z$), and two non-linear $X \to Y$ and $X \to Z$

$$X_t = 3.4X_{t-1}(1 - X_{t-1})^2 e^{X_{t-1}^2} + 0.4\epsilon_{1,t}$$

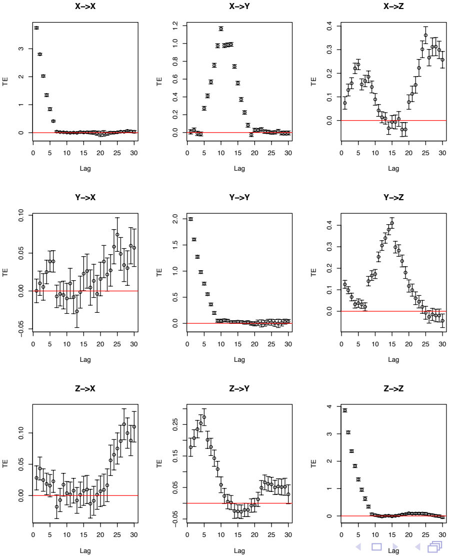
$$Y_t = 3.4Y_{t-1}(1 - Y_{t-1})^2 e^{Y_{t-1}^2} + 0.5X_{t-10}^2 + 0.4\epsilon_{2,t}$$

$$Z_t = 3.4Z_{t-1}(1 - Z_{t-1})^2 e^{Z_{t-1}^2} + 0.3Y_{t-15} + 0.5X_{t-5}Z_{t-1} + 0.4\epsilon_{3,t}$$

Geweke's test for (linear) Granger Causality

	p-value	
$Y \to X$	0.44	
$Z \to X$	0.08	
$X \to Y$	0.00	interaction detected
$Z \to Y$	0.23	
$X \to Z$	0.68	interaction not detected, lag 5
$Y \to Z$	0.00	linear interaction detected

Example 1

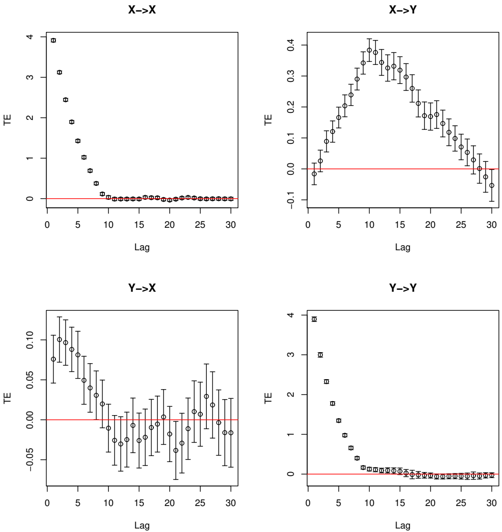


Example 2

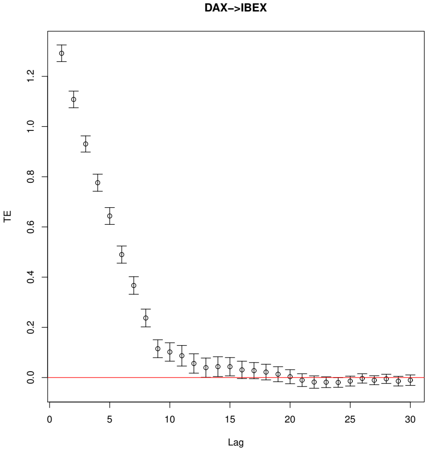
Non stationary system, $X \to Y$ with a true delay $\delta_{X \to Y} = 10$ and random variance that follows an IGARCH(1,1):

$$X_t = 0.7X_{t-1} + \sigma_{1,t}\epsilon_{1,t}$$
$$Y_t = 0.3Y_{t-1} + 0.5X_{t-10}Y_{t-2} + \sigma_{2,t}\epsilon_{2,t}$$
$$\sigma^2_{i,t} = 0.2 + 0.9\epsilon^2_{i,t-1} + 0.1\sigma^2_{i,t-1}$$

Example 2



Example 3 : German DAX-30 & Spanish IBEX-35



Log-returns were considered from =1/01/2011 to 06/24/2016