Granger Causality and Transfer Entropy for Financial Returns

E.M. Syczewska[a,*] and Z.R. Struzik[b,c,d]

[a]Warsaw School of Economics, Department of Economic Analyses, Institute of Econometrics,
Madalińskiego 6/8, 02-513 Warsaw, Poland

[b]RIKEN Brain Science Institute, 2-1 Hirosawa, Wako-shi 351-0198, Japan

[c]Graduate School of Education, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

[d]Institute of Theoretical Physics and Astrophysics, The University of Gdańsk,
Wita Stwosza 57, 80-952 Gdańsk, Poland

Granger causality in its linear form has been shown by Barnett, Barrett and Seth [*Phys. Rev. Lett.* **103**, 238701 (2009)] to be equivalent to transfer entropy in case of Gaussian distribution. Generalizations by Hlaváčková-Schindler [*Appl. Math. Sci.* **5**, 3637 (2011)] are applied to distributions typical for biomedical applications. The financial returns, which are of great importance in financial econometrics, typically do not have Gaussian distribution. Generalizations leading to the concept of nonlinear Granger causality (e.g. causality in variance, causality in risk), known and applied in econometric literature, seem to be less known outside this field. In the paper an overview of some of the definitions and applications is given. In particular, we indicate some recent econometric results concerning application of the tests in linear multivariate framework. We emphasize importance of other variants of Granger causality, and need of development of methods reflecting features of financial variables.

1. Introduction

Granger causality (GC) and transfer entropy (TE) are two approaches to mutual causation. The transfer entropy, TE, introduced by Schreiber [1], builds on the concept of (Shannon) entropy, but aims at detecting dynamic causation links between a pair of variables. The Granger causality, GC, aside of TE, tends to be used by researchers in climatology, physiology, neurophysiology, multimode laser dynamics, analysis of causality in cardio-respiratory interactions, etc. (see [2], p. 5), partially due to equivalence of the GC and TE concepts, proved by Barnett et al. in [3], and by Hlaváčková-Schindler et al. [2]. Let us emphasize that the equivalence was shown only for a linear GC test, and originally only under assumption of Gaussianity. Later Hlaváčková-Schindler [4] extended the analysis for some probability distributions typical for biomedical phenomena (lognormal distribution, Gaussian mixtures, generalized normal distribution, Weinman exponential distribution). Barnett and Bossomaier [5] show TE and GC equivalence for the vector autoregressive model also under assumption of Gaussianity. This does not cover other concepts of Granger causality, and does not necessarily cover all distributions used in applied financial research and practice.

The concentration on the linear form of the GC (and the Gaussianity assumption) in biomedical applications seems to be somewhat misleading. First of all, the Granger causality concept as described in papers by C.W.J. Granger is richer than that (covers linear and nonlinear causality, causality in spectral domain, causality based on information concepts), and has several important generalizations. Second, the question of stationarity and nonstationarity, and of proper choice of distribution has been learned by econometricians the hard way. This is especially important for financial time series — underestimating or overestimating particular values due to a wrong choice of probability distribution for the model can lead to losses or to inefficiency. The proper choice of the distribution is especially crucial for risk assessment and volatility forecasting by financial institutions.

Hlaváčková-Schindler et al. [2] in addition to entropy and mutual information measures, describe multivariate GC in vector autoregression model framework, nonlinear GC tests, and nonparametric GC measures based on correlation integral, but still do not cover the GC-TE equivalence in the general sense. On the other hand, econometricians use both linear and nonlinear GC tests, and tools based on mutual information, entropy etc. It seems that absorption by econometricians of methods aimed at detecting causality, developed in the field of neurophysiology etc. is stronger than that of methods developed in econometrics (especially of nonlinear Granger causality) in the other direction.

2. Typical features of financial variables

Financial variables have specific features, which led to development of particular modeling tools from the field of financial econometrics. Let P_t denote price of a financial instrument at time t. Most financial variables of interest (stock indices, exchange rates, stock and options etc.) are nonstationary. Their returns, defined as difference of natural logarithms of prices, $r_t = \ln(P_t) - \ln(P_{t-1})$, are stationary in mean, but typically show changes of volatility in time (termed volatility clustering). Changing volatility is modeled e.g. with ARCH or GARCH-models, the first equation of which describes (and forecasts) mean

*corresponding author; e-mail: ewams@sgh.waw.pl

of the variable, and the second equation describes dependence between changing volatilities with use of conditional variance of the first equation. The ARCH ("autoregressive with conditional heteroskedasticity") model was introduced by Robert F. Engle [6] in 1982; GARCH (generalized ARCH) — by Timothy Bollerslev [7] in 1986, and many more detailed GARCH-type specifications have been introduced by various authors to cover specific features of financial instruments and particular markets.

The GARCH family models originally estimated by the maximum likelihood method with assumption of Gaussianity of error terms, prove to give more accurate results if non-Gaussian, often skewed, probability distributions are applied, especially for daily and higher frequency data (see Alexander [8]).

Returns of financial instruments typically have skewed and leptokurtic distributions. According to Mandelbrot, who first noticed volatility clustering in financial returns time series [9], returns should be modelled with use of Pareto-Lévy processes[†]. In finance and in econometrics returns are most often described with use of the following distributions: Gaussian mixtures, t-Student, or more complicated: general error distribution (GED in short) or generalizations of the last two (see Osińska [12], pp. 170–172). The GED (General Error Distribution) has been introduced in [13] by Subbotin, and later generalized by Theodossiou [14] to a Skewed GED. The reason is that some stable distributions are not suitable for statistical testing of an econometric model, and that more complex distributions are intended to better reflect actual features of financial variables. The skewed generalized t distribution can be expressed as (see Osińska [12]):

$$f_{\mathrm{SGDT}}(x) = C \left\{ 1 + \left(\frac{k}{\nu - 2} \right) \right.$$
$$\left. \times \left[\theta (1 + \mathrm{sgn}(x - \mu)) \lambda \right]^{-k} \left| \frac{x + \mu}{\sigma} \right|^k \right\}^{\frac{\nu + 1}{k}}, \quad (1)$$

where $B(\cdot, \cdot)$ — the beta function, $k, \nu, \lambda, \mu, \sigma$ — parameters of the distribution ($k > 0; \nu > 2; -1 < \lambda < 1; \mu$ — the mean; σ — the standard deviation); $\mathrm{sgn}(z) = \frac{z}{|z|}$; the coefficients are defined as

$$C = 0.5 B \left(\frac{1}{k}, \frac{\nu}{k} \right)^{-1.5} B \left(\frac{3}{k}, \frac{\nu - 2}{k} \right)^{0.5} S(\lambda) \sigma^{-1},$$

$$\theta = \left(\frac{k}{\nu - 2} \right)^{1/k} B(\frac{1}{k}, \frac{\nu}{k})^{0.5} B(\frac{3}{k}, \frac{\nu - 2}{k})^{0.5} S(\lambda)^{-1},$$

$$S(\lambda) = \left[1 + 3\lambda^2 - 4\lambda^2 B \left(\frac{2}{k}, \frac{\nu - 1}{k} \right)^2 \right.$$
$$\left. \times B \left(\frac{3}{k}, \frac{\nu - 2}{k} \right)^{-1} B \left(\frac{1}{k}, \frac{\nu}{k} \right)^{-1} \right]^{0.5}.$$

The GED probability density function is (see Tsay [15], p. 122):

$$f_{\mathrm{GED}}(x) = \frac{\nu \exp(-0.5|x/\lambda|^\nu)}{\lambda 2^{(1+1/\nu)} \Gamma(1/\nu)}, \quad (2)$$

where $\Gamma(\cdot)$ is the gamma function, and $\lambda = [2^{(-2/\nu)} \Gamma(1/\nu)/\Gamma(3/\nu)]^{1/2}$. This distribution has heavy tails when $\nu < 2$ and reduces to a Gaussian distribution when $\nu = 2$.

To illustrate discrepancies, especially in the tails, between Gaussian and the empirical distribution of the daily returns of the USDPLN exchange rate and the WIG20 returns, see Figure for daily quotations from stooq.pl (Jan. 4, 2000–Oct. 31, 2014)[‡]. The results of the Granger causality test (based on the VAR model) for the two variables [16] show that during crisis, Oct. 01, 2007–Feb. 27, 2009, the null hypothesis of no causality from USDPLN to WIG20 is weakly rejected, the same for causality from WIG20 towards USDPLN. The results change somewhat after the crisis. Neither the returns nor the error terms of the VAR models are Gaussian.
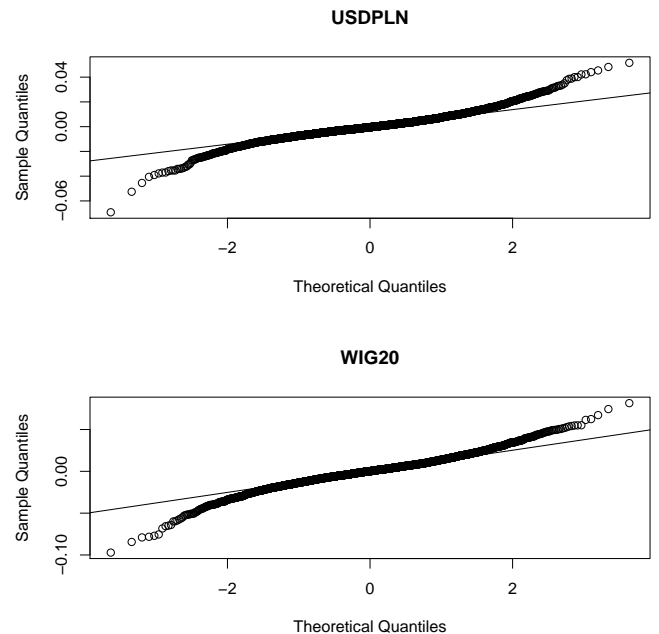

Fig. 1. Comparison of the quantiles of daily returns with theoretical Gaussian distribution.

In order to be aplicable in financial econometrics, question of TE and GC equivalence should include generalization to the types of distributions typically applied in financial econometric research.

In this paper, examples of causation inference in the case of realistic data models will be reviewed, and the possible directions for improvement will be discussed.

[†]Description of the Pareto-Lévy, or stable distributions can be found in Feller [10], Vol. II, Chapter VI. S.T. Rachev, among others, is a strong proponent of Lévy distribution (see e.g., [11]).

[‡]QQ-plot is a graph of percentiles of the empirical versus theoretical distribution.

3. Granger causality and transfer entropy — two approaches to mutual causation

Causality is usually posed using two alternative scenarios: the autoregressive, spectral based, causal Granger predictive modelling (Granger [17]), and the information theoretic oriented, Kullback–Leibler (K–L) divergence based, Transfer Entropy formulation (Schreiber [1]). Both measures aspire to infer and quantify mutual directional causation, between coupled variables, or among multiple variables.

There is a need to precisely analyse possible equivalence of the two methods especially for complex phenomena where the observations and models are usually distributed following non-Gaussian statistics and possess intra-and inter-observable non-linear correlations. As empirical distributions of financial time series differ significally from the Gaussian distribution, in order to apply TE methods to such data we need to address the question of equivalency of the GC and TE approaches for such skewed and leptokurtic distributions. First we try to clarify concepts of the Granger causality, as defined and tested in econometric literature, next we show that concepts based on information theory, such as mutual information, entropy and transfer entropy are known and used by econometricians to address causality in risk and in volatility, among others. Next we shall analyse reasoning shown in the papers by Hlaváčková-Schindler et al. [2, 4, 18] and Barnett, Barrett, Seth [3, 19], to check intuitions behind their proofs of the linear GC and TE equivalence.

4. Diversity of Granger causality concepts

First the notion of Granger causality needs to be clarified. It can be understood as improvement, by use of past values of one variable, of forecasts of another, but has also other interpretations. The Granger causality ([17, 20]) was intended by Granger as a more general notion based on two assumptions: "that for a time series, the cause preceded the effect and a causal series had information about the effect that was not contained in any other series according to the conditional distribution"[§]. Granger himself devoted several articles and book chapters to issues of causality and its role in economic analysis and forecasts. Aside from the linear time-series approach to causality, Granger [17] discussed spectral approach as early as in 1969[¶]. In financial analysis, nonlinear causality concepts and tests (e.g., risk causality, causality in variance) are becoming more and more important. The concept of the Granger causality is more complicated than the simplest linear version.

The general definition of the Granger causality in mean is the following (Osińska [12], pp. 40–41). Let $\mathcal{F}(Y|X)$

denote a distribution of Y conditional on X, and let $\mathcal{J}_t$ define a set of all information avaiable at time t; $\mathcal{J}_t \setminus X_t$ — all information at time t aside from the information contained in X_t. If for all forecast horizons k

$$\mathcal{F}(Y_{t+k}|\mathcal{J}_t) = \mathcal{F}_t(Y_{k+1}|\mathcal{J}_t \setminus X_t) \tag{3}$$

then X_t is not a G-cause for Y.

To make this definition operational, the set of available information has to be described more precisely. Usually, for weakly stationary time series X_t, Y_t, past observations of the variables are taken into account. Let $X_t^- = \{x_{t-1}, x_{t-2}, ..., x_{t-k}, ...\}$ (past values of the X variable) and similarly for Y. Let $\mathcal{J}_t$ be a set of information (σ-algebra) based on the past values of X and Y. If a variance of prediction error for a predictor based on all the information differs from that for the predictor based on all information but the past of X,

$$\sigma^2(Y_t|\mathcal{J}_t) < \sigma^2(Y_t|\mathcal{J}_t \setminus X_t^-) \tag{4}$$

then the X is said to Granger-cause the Y.

The instantaneous G-causality means that Y can be predicted better with use of the past and current value of X than without them:

$$\sigma^2(Y_t|\mathcal{J}_t \cup X_t) < \sigma^2(Y_t|\mathcal{J}_t \setminus (X_t^- \cup X_t)). \tag{5}$$

This definition can be implemented in framework of a linear regression model or multivariate vector autoregresion model. In the papers of Barnett et al. [3] and Hlaváčková-Schindler [4] (who follows their notation) namely this linear version of GC is compared to the transfer entropy approach.

4.1. Test of significance of (lagged) predictor

The test procedure is the following: run a regression of Y on its lags and lags of X, and check whether lags of X are significant (i.e. their parameters are different from zero).

$$y_t = \sum_{j=1}^{k} \alpha_j y_{t-j} + \sum_{j=1}^{k} \beta_j x_{t-j} + \varepsilon_t, \tag{6}$$

where: y_t — observations of a variable of interest; x_{t-j} — observations of causal variable; α_j, β_j — parameters of the regression; ε_t — error term. The restricted version of the (6) model with the null restrictions imposed is

$$y_t = \sum_{j=1}^{k} \alpha_j y_{t-j} + \eta_t. \tag{7}$$

The null hypothesis $H_0 : \beta_j = 0$ for all $j = 1, ..., k$ corresponds to lack of G-causality from X to Y, the alternative $H_1 : \beta_j \neq 0$ for some $j \in \{1, ..., k\}$ means that the X Granger-causes Y. The joint significance test is applied, either in the form of χ^2 or Fisher F distribution[**]. This is called the Granger test of the Granger causality.

Let $\hat{\sigma}^2(\varepsilon)$ and $\hat{\sigma}^2(\eta)$ denote estimators of error term variance for (6) and (7). The Wald test is computed as

---

[§]See [21] pp. 69–70.
[¶]See e.g., Osińska and Stawicki [22] for an example of empirical application of causality across spectral frequency bands.

[**]See e.g. Feller [10], Vol. II, chapter II, sect. 3, for description of the both statistics.

$$T^W = T[\hat{\sigma}^2(\varepsilon) - \hat{\sigma}^2(\eta)]/\hat{\sigma}^2(\eta), \qquad (8)$$

the likelihood-ratio test as

$$T^{LR} = T\ln[\hat{\sigma}^2(\varepsilon)/\hat{\sigma}^2(\eta)] \qquad (9)$$

and the Lagrange Multiplier test as

$$T^{LM} = T[\hat{\sigma}^2(\varepsilon) - \hat{\sigma}^2(\eta)]/\hat{\sigma}^2(\varepsilon), \qquad (10)$$

where T — sample size; the T^W is preferable; for small sample, the F-form of the tests (right-hand side expressions are multiplied by $(T-k)/q$, instead of T, where k — number of restriction, q — number of parameters of the model) has better properties (see [12], pp. 77-78, 84).

In this framework, instantaneous Granger causality can be tested: the current X is included on the right-hand side with a parameter β_0, the null hypothesis corresponds to $\beta_j = 0$ for $j = 0, 1, 2, ..., k$. For example, Malliaris and Urrutia [23] analyzed daily closing data for index returns of the main stock exchanges (Sydney, Tokyo, Hong Kong, Singapore, London and New York), to study potential dependencies between the markets before, during and after the crash of October 1987. They test G-causality and instantaneous version. The null hypothesis of no causality cannot be rejected for the pre-crash period; for the month of the market crash, the χ^2 test statistics increased, indicating dependence. They detect bidirectional causality between New York and London, New York and Hong Kong, London and Singapore, London and Tokyo, London and Sydney, Hong Kong and Sydney, and Tokyo and Sydney.

Barnett et al. [3] start with a linear regression of the "predictee" variable on its lags and lags of "predictor". They use the ratio of the residual variance of errors in the restricted equation to the residual variance of unrestricted equation. In this they follow Geweke [24], the same formula as in (9).

4.2. The Sims test of the Granger causality

Let us briefly mention another kind of linear test: Sims [25] defines X to be strict exogenuous relative to Y if the linear predictor of Y_t based on past and future values of $X : ..., x_{t-1}, x_t, x_{t+1}, ...$ is identical to the linear predictor based only on current and past values of x, and has shown those two definitions to be equivalent. The Sims test of G-causality from X towards Y is performed as a joint significance test of leads $x_{t+1}, x_{t+2}, ..., x_{t+k}$ in a regression of X on current value and lags of the Y variables and leads and lags of the X variable. Chamberlain [26] extends the Granger and Sims causality definitions using conditional independence instead of linear prediction and shows that Granger and Sims definitions of causality are equivalent.

4.3. Testing G-causality in VAR and VECM framework — stationary vs. nonstationary series

Next approach is to check causality for multivariate variables. Let $\mathbf{Y}$ denote column vector of variables of interest: $\mathbf{Y} = [Y_1, Y_2, \ldots, Y_m]^T$. The vector autoregressive model (VAR in short), a generalization of the ARMA models of Box and Jenkins [27], is build as an autoregression of a (column) vector $\mathbf{Y}$ on its lags:

$\mathbf{Y}_t = \mathbf{A}_0 + \mathbf{A}_1\mathbf{Y}_{t-1} + ... + \mathbf{A}_k\mathbf{Y}_{t-k} + \boldsymbol{\mathcal{E}}_t$, where $\mathbf{Y}$ — vector of variables of the model, $\mathbf{A}_i$ — matrices of parameters, k — number of lags, $\boldsymbol{\mathcal{E}}$ — (column) vector of error terms, $\boldsymbol{\mathcal{E}}_t = [\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{mt}]^T$. The model can be estimated with maximum likelihood method.

For stationary $\mathbf{Y}$, the test of non-causality can be performed in the context of bivariate VAR as the Wald test of joint insignificance of all lags of one variable in the equation explaining another. It can be generalized to VAR with three variables, in which we test GC from X to Y and *vice versa*, with Z as exogenous regressor.

For nonstationary variables, the accepted practice is first to check whether there exists a linear stationary combination of the $Y_1, Y_2, \ldots, Y_m$ variables, so-called cointegrating relationship which corresponds to a stable dynamic long-run equilibrium relationship (see the famous paper by Engle and Granger [28], and Johansen [29] for the appropriate tests)[**]. The VAR regression should then include (stationary) terms corresponding to such a relationship. For a bivariate case, suppose that $\hat{u}_t$ corresponds to a cointegration relationship between the two variables, then the model appropriate for GC-test has the form:

$$y_t = \sum_{j=1}^{k} \alpha_{1j}y_{t-j} + \sum_{j=1}^{k} \beta_{1j}x_{t-j} + \gamma_1\hat{u}_t + \varepsilon_1 t, \qquad (11)$$

$$x_t = \sum_{j=1}^{k} \alpha_{2j}y_{t-j} + \sum_{j=1}^{k} \beta_{2j}x_{t-j} + \gamma_2\hat{u}_t + \varepsilon_2 t, \qquad (12)$$

where $\alpha_{1j}, \beta_{1j}, \gamma_1$ — parameters of the first equation, $\alpha_{2j}, \beta_{2j}, \gamma_2$ — parameters of the second equation. The null hypothesis $\beta_{1j} = 0$ for $j = 1, 2, \ldots, k$ corresponds to lack of causality from X to Y, and the null hypothesis $\alpha_{1j} = 0$ for $j = 1, 2, \ldots, k$ to lack of G-causality from Y to X.

Recently it was shown by Toda and Yamamoto [31] and Bauer and Maynard [32] that in case of possibly nonstationary variables it is advisable to apply overparameterised VAR model to the pair of variables in question, namely: to choose number of lags, k, in the VAR model according to the Akaike (or Bayesian Schwarz) Information Criteria, and estimate the VAR model using $k + r$ lags for both variables, where r denotes maximum integration order for the variables[††]. The Wald test statistics is then applied *only to the first k* parameters of the lagged causal variable.

5. Entropy and information-based measures of dependence

In discussing potential transfer entropy and linear Granger causality equivalence, we should bear in mind

———

[**]Bossomaier, Barnett and Harré [30] describe briefly Granger causality test in the context of the VAR model (they suggest to apply local linear approximation to a nonstationary series).

[††]A variable is integrated of order r, if it is nonstationary, but its r-th differences are stationary. Order of integration is the minimum number of differences required to achieve stationarity, see [28].

that there are several measures and definitions of entropy, starting with the Shannon definition, Rényi entropy and others. All can be used to construct measures of mutual information, slightly differing in properties; however the Shannon entropy, $\mathcal{H}$, seems to be most widely used. Hlaváčková et al. [2] write: "Besides Shannon and Rényi entropy, other entropy definitions (Tsallis, etc.) are studied in the mathematical literature, but Shannon entropy is the only one possessing all the desired properties of an information measure." (p. 6). Jizba et al. [33] note that the Rényi transfer entropy is not generally positive, and the Rényi entropy emphasizes only parts of the probability density function.

The measures based on information theory are known to and applied by econometricians. Bruzda [34] describes conditions formulated by Granger, Maasoumi and Racine [35] for the ideal measure of functional dependence of two stochastic variables:
1) Is well defined for discrete and for continuous variables;
2) Is normalized to [0, 1] or [−1, 1] interval, and for independent variables is equal to 0;
3) If there is a measurable function $f : Y = f(X)$, then an absolute value of this measure $= 1$;
4) For Gaussian bivariate distribution of (X, Y), the measure is either equal to correlation coefficient ρ or is a simple function of ρ;
5) Fulfills conditions of a distance;
6) Is invariant to continuous and strictly monotonous transformations of variables.

Let S be a discrete random variable, taking values $s_1, s_2, \ldots, s_m$ with corresponding probability $p_i, i = 1, 2, \ldots, m$. The (Shannon) entropy is an "average amount of information gained from a measurement of one particular value" (see [2]):

$$\mathcal{H}(S) = -\sum_{i=1}^{m} p_i \log p_i, \tag{13}$$

next the joint entropy $\mathcal{H}(X, Y)$ of two discrete random variables is defined as

$$\mathcal{H}(X, Y) = -\sum_{i=1}^{m_X} \sum_{j=1}^{m_Y} p(x_i, y_j) \log p(x_i, y_j), \tag{14}$$

where $x_i, i = 1, \ldots, m_X$ — values of the (discrete) variable X, $y_j, j = 1, \ldots, m_Y$ — values of the (discrete) variable Y, $p(x_i, y_j)$ — the joint probability that X is in state x_i and Y is in state y_j. The joint entropy is expressed in terms of conditional entropy $\mathcal{H}(Y|X)$:

$$\mathcal{H}(Y|X) = \mathcal{H}(X, Y) - \mathcal{H}(Y). \tag{15}$$

In the discrete case the conditional entropy is equal to

$$\mathcal{H}(Y|X) = \sum_{i=1}^{m_X} \sum_{j=1}^{m_Y} p(x_i, y_j) p(y_j|x_i), \tag{16}$$

where $p(y_j|x_i)$ denotes conditional probability.

The mutual information (MI in short) between two variables is defined as (see [2, 36–38]):

$$I(X, Y) = \mathcal{H}(X) + \mathcal{H}(Y) - \mathcal{H}(X, Y). \tag{17}$$

Mutual information is the mutual reduction of uncertainty of one variable by knowing the other ([2]). It is nonnegative, equal to zero only for independent variables.

It can be normalized to give a mutual information coefficient (MIC in short), defined as:

$$R(X, Y) = 1 - \exp\left(-2I(X, Y)\right)^{0.5} \tag{18}$$

for which (see Granger and Teräsvirta [36], Granger and Lin [37]):
1) $0 \leq R(X, Y) \leq 1$;
2) $R(X, Y) = 0 \iff X$ and Y are independent;
3) $R(X, Y) = 1 \iff Y = f(X)$, where f is an invertible function;
4) $R(X, Y)$ is invariant to data transformation, i.e., $R(X, Y) = R(h_1(X), h_2(Y))$, where functions h_1, h_2 are strictly monotonous;
5) for bivariate Gaussian process (X, Y) with correlation coefficient $\rho(X, Y)$, $R(X, Y)$ simplifies to $|\rho(X, Y)|$.
Both $I(X, Y)$ and $R(X, Y)$ can be applied to $X = X_t$ and $Y = X_{t-j}, j = 1, 2, \ldots$, to detect autodependencies between a variable and its lags, not necessarily linear.

5.1. Example: $R(X, Y)$ applied to detecting nonlinearity

Bruzda [38] presents simulation experiment for the nonlinear MA model, ARCH(2), GARCH(1,1), GARCH-M, three variants of bilinear processes and two variants of linear processes: Gaussian white noise and Gaussian stationary AR(1) process. Her results are the following: (1) The mutual information coefficient was able to detect all cases of nonlinearity; (2) The maximum correlation coefficient did almost as well, and was able to detect nonlinearity in variance; (3) The entropy measure could distinguish between GARCH-type nonlinearity and bilinear nonlinearity. (4) For a nonlinear moving-average process MA(1), both the mutual information coefficient and entropy measure cut off after 1 lag, hence perhaps can be used as a tool for detecting number of lags in such a model.

5.2. Example: mutual information measures applied to rates of return

Orzeszko [39] with use of the mutual information and the mutual information coefficient for rates of return of stock exchange indices, proves existence of some autodependencies in rates of return for the BUX, DAX, CAC20, DJIA, FTSE, HangSeng, Nasdaq, Nikkei, SP500, and WIG20, MWIG40 and SWIG80 for a period 2001/01/02–2011/06/30. The log returns were first filtered with an ARMA-GARCH model with t-Student distribution, and mutual corelation coefficients were applied (and checked for significance) for both a series and model errors. In some cases this indicated dependencies not explained by the ARMA-GARCH model.

6. G-causality and transfer entropy — equivalence only under Gaussianity?

Let us address the question of GC and TE equivalence, starting with description of line of reasoning of Barnett, Barrett and Seth [3]. First, they formulate the measure of GC as the variance ratio of the restricted and unrestricted linear model error terms, in our notation — equation (7)

and (6), and the fraction similar to that in (9). Next they extend this proportion to multivariate variables. More precisely, for jointly distributed random variables X, Y let $\Sigma(X)$ denote a covariance matrix of X, $\Sigma(Y)$ for Y, and $\Sigma(X, Y) = [\text{cov}(X_i, Y_i)]$ — cross-covariance matrix for the two variables. They introduce partial covariance matrix, $\Sigma(Y|X)$ and show that under certain conditions this equals the covariance matrix of the error term in the linear regression of Y on X.

As a measure of Granger causality they use the ratio of determinants of the two conditional covariance matrices: in the numerator, the partial covariance matrix of Y conditioned only on its own past (this corresponds to a restricted multivariate model), in the denominator, the determinant of the partial covariance matrix of Y conditioned on its past and the past of X (this corresponds to the unrestricted model).

$$\mathcal{F}_{X \to Y} = \ln \left[\frac{\det(\Sigma(Y|Y^-))}{\det(\Sigma(Y|Y^- \cup X^-))} \right].$$

If an additional variable, Z, is included on the right-hand side of the multivariate model, the measure of GC from X to Y conditional on Z becomes in this notation:

$$\mathcal{F}_{X \to Y|Z} = \ln \left[\frac{\det(\Sigma(Y|Y^- \cup Z^-))}{\det(\Sigma(Y|Y^- \cup X^- \cup Z^-))} \right],$$

where $\det(\cdot)$ denotes determinant of a matrix, $Y^- \cup Z^-$ — past values of the Y and Z variable, etc.

Following Schreiber [1], Barnett et al. [3] define transfer entropy of Y to X as "the difference between the entropy of Y conditioned on its own past and the past of Z, and its entropy conditioned, in addition, on the past of X", hence the notation:

$$\mathcal{T}_{X \to Y|Z} = \mathcal{H}(Y|Y^- \cup Z^{-1}) - \mathcal{H}(Y|Y^- \cup X^- \cup Z^-),$$

where $\mathcal{H}(\cdot)$ denotes entropy and $\mathcal{H}(\cdot|\cdot)$ — conditional entropy. They show that for multivariate Gaussian random variable, its entropy is a function of natural logarithm of the covariance matrix determinant, and that conditional entropy $\mathcal{H}(Y|X)$ is proportional to the determinant of the corresponding partial covariance matrix. Hence

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

where the first expression corresponds to the measure of Granger causality, the second — to the transfer entropy. Note that Barnett, Barret and Seth in this 2009 paper do not cover other variants of entropy nor the questions of entropy estimation.

Hlaváčková-Schindler [4] turns to the question of analytical expressions for entropy of several multivariate distributions — aside from the Gaussian distribution, she investigates generalize-normal distribution, the Weinman exponential distribution and the multivariate log-normal distribution, showing that for all three the equivalence of $\mathcal{F}_{X \to Y|Z}$ and $\mathcal{T}_{X \to Y|Z}$ holds. The starting point is still the linear version of the Granger causality. The proof goes along comparison of the corresponding variances and TE measure for a Gaussian distribution. More detailed analysis — covering also nonlinear versions of the Granger causality, and problems of estimation of entropy measures — is performed in [18]. The starting point is also the GC definition based on forecasting interpretation, and its linear test. Hlaváčková-Schindler next addresses question of several methods of entropy estimation. In this chapter, she does not investigate question of other multivariate distributions and their influence on TE estimation, and on possible equivalence.

Hlaváčková-Schindler et al. [2] describe several other measures of dependence, and variants of Granger causality measures (both linear and nonlinear). They note among others, the notion of Granger causality based on generalized correlation integrals, used to construct the Hiemstra–Jones test [40] and its generalizations.

Barnett and Seth [19] build the GC measure in the time domain as the likelihood ratio, i.e. again as ratio of appropriate covariance matrices, but provide at the same time tools for GC measure in the frequency domain. Their MATLAB toolbox seems to be promising, however does not include tools for Granger causality testing in the nonstationary case of cointegrated variables. In Barnett and Bossomaier [5] paper, equivalence of GC and TE is shown. In addition, they mention that the VAR linear model does not necessarily provide the best model for a variable in question, sometimes diagnostic tools indicate e.g., heteroskedasticity, thus suggesting the GARCH models; they say: "Now it may be far from clear how one should define a Granger-like predictive statistic for such models". In our opinion, the answer is to apply procedures well known and often applied in financial econometrics — to estimate an ARMA-GARCH or similar model, and then apply the GC measure to the residuals (as was mentioned in subsection 5.2).

7. Concluding remarks

In the context of biomedical data analysis, the assumption of Gaussianity and even stationarity may in some cases be fulfilled, but it is not quite realistic in the case of financial variables. If we take into account such features and note that in general $\mathcal{H}(f(X)) \leq \mathcal{H}(X)$, with equality only for Gaussian distribution; that estimating of covariance matrix can be more numerically demanding for data showing heteroskedasticity and long-term dependence; that distribution of (error terms in models) of financial returns are better approximated by asymmetric distributions such as (1) or (2) — then we note that the TE and GC in the context of linear models, does not exhaust the topic of the Granger causality.

In our opinion, other variants of the Granger causality methods (nonlinear, nonparametric) developed in the field of econometrics, deserve more attention outside of this area. It seems that they are starting to appear recently in neuroscience and similar journals.

For financial applications, TE and GC should be compared also for (skewed and leptokurtic) distributions used in financial returns modeling, such as skewed t distribution (1), or the GED (2) of Subbotin [13] and its skewed generalization of Theodossiou [14]. It is known that with increase of observations frequency, the financial

variables distribution similarity to the Gaussian distribution decreases with increase of observations frequency (monthly data can be described with a Gaussian distribution, for daily data it would not work well especially in modelling and forecasting such features as measures of risk and volatility, needed to make investment decisions). The more accurate choice of a distribution is crucial for several financial decisions and forecasts.

Studies concerning the Shannon entropy representation for skewed distributions, as the one by Arellano-Valle et al. [41], seem to be promising base for further research. Also availability of software such as MATLAB software by Seth [42] or [19] (and several packages in R) will be helpful.

In addition, question of precision and robustness of measures (depending e.g. on version of entropy estimates) should be addressed in future research.

References

[1] Th. Schreiber, *Phys. Rev. Lett.* **85**, 461 (2000).

[2] K. Hlávačková-Schindler, M. Paluš, M. Vejmelka, J. Bhattacharya, *Phys. Rep.* **441**, 1 (2007).

[3] L. Barnett, A. Barrett, A. Seth, *Phys. Rev. Lett.* **103** 238701 (2009).

[4] K. Hlaváčková-Schindler, *Appl. Math. Sci.* **5**, 3637 (2011).

[5] L. Barnett, T. Bossomaier, *Phys. Rev. Lett.* **109**, 138105 (2012).

[6] R.F. Engle, *Econometrics* **50**, 987 (1982).

[7] T. Bollerslev, *J. Econometrics* **31**, 307 (1986).

[8] C. Alexander, *Market Risk Analysis. Volume II. Practical Financial Econometrics*, John Wiley, Chichester 2008.

[9] B. Mandelbrot, *J. Bus.* **36**, 394 (1963).

[10] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. II, John Wiley, New York 1957.

[11] S.T. Rachev, Y.S. Kim, M.L. Bianchi, *Frank J. Fabozzi Series: Financial Models with Lévy Processes and Volatility Clustering*, John Wiley, Hoboken, NJ 2011.

[12] M. Osińska, *Ekonometryczna analiza zależności przyczynowych*, Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, Toruń, 2008.

[13] M.T. Subbotin, *Matematicheskii Sbornik* **31**, 296 (1923).

[14] P. Theodossiou, *SSRN* **11**, 1 (2000).

[15] R.S. Tsay, *Analysis of Financial Time Series*, 3rd ed., John Wiley, Hoboken 2007.

[16] E.M. Syczewska, Metody Ilościowe w Badaniach Ekonomicznych **XV**/4, 169 (2014).

[17] C.W.J. Granger, *Econometrica* **37**, 424 (1969).

[18] K. Hlaváčková-Schindler, *Causality in Time Series: Its Detection and Quantification by Means of Information Theory*, in: *Information Theory and Statistical Learning*, Ed. F. Emmert-Streib, M. Dehmer, Springer-Verlag, New York 2009, p. 183.

[19] L. Barnett, A.K. Seth, *J. Neurosci. Meth.* **223**, 50 (2014).

[20] C.W.J. Granger, *J. Econ. Dyn. Control* **2** 329 (1980).

[21] C.W.J. Granger, *J. Econometrics* **112**, 69 (2003).

[22] M. Osińska, J. Stawicki, *Testing for causality across spectral frequency bands*, in: *Some aspects of the dynamic econometric modelling*, Ed. Z. Zieliński, Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń 1993, p. 135.

[23] A.G. Malliaris, J.L. Urrutia, *J. Financ. Quant. Anal.* **27**, 353 (1992).

[24] J.F. Geweke, *J. Am. Stat. Assoc.* **77**, 304 (1982).

[25] C.A. Sims, *Am. Econ. Rev.* **62**, 540 (1972).

[26] G. Chamberlain, *Econometrica* **50**, 569 (1982).

[27] G.E.P. Box, G.M. Jenkins, *Time Series Analysis. Forecasting and control*, Holden-Day, San Francisco 1976.

[28] C.W.J. Granger, R.F. Engle, *Econometrica* **55**, 251 (1987).

[29] S. Johansen, *Likelihood-based inference in cointegrated vector autoregressive models*, Oxford University Press, Oxford 1995.

[30] T. Bossomaier, L. Barnett, M. Harré, *Complex Adaptive Systems Modeling* **1**, 9 (2013).

[31] H.Y. Toda, T. Yamamoto, *J. Econometrics* **66**, 225 (1995).

[32] D. Bauer, A. Maynard, *J. Econometrics* **169**, 293 (2012).

[33] P. Jizba, H. Kleinert, M. Shefaat, *Physica A* **391**, 2971 (2012).

[34] J. Bruzda, *Procesy nieliniowe i zależności długookresowe w ekonomii. Analiza kointegracji nieliniowej*, Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, Toruń 2007.

[35] C.W.J. Granger, E. Maasoumi, J. Racine, *J. Time Series Analysis* **25**, 649 (2004).

[36] C.W.J. Granger, T. Teräsvirta, *Modelling Nonlinear Economic Relationships*, Oxford University Press, Oxford 1993.

[37] C.W.J. Granger, J.-L. Lin, *J. Time Series Analysis* **15**, 371 (1994).

[38] J. Bruzda, *AUNC* **34**, 183 (2004).

[39] W. Orzeszko, *Przegląd Statystyczny* **59**, 369 (2012).

[40] C. Hiemstra, J.D. Jones, *J. Financ.* **49**, 1639 (1994).

[41] R. Arellano-Valle, J.E. Contreras-Reyes, M.G. Genton, *Scand. J. Stat.* **40**, 42 (2012).

[42] A.K. Seth, *J. Neurosci. Meth.* **186**, 262 (2010).