Granger causality and transfer entropy for financial returns

Ewa M. Syczewska

Zbigniew R. Struzik

Warsaw School of Economics, Department of Economic Analyses, Institute of Econometrics, Madalinskiego 6/8, 02-513 Warsaw, Poland
and
University of Tokyo, RIKEN Brain Science Institute, Wakoshi, Japan

7 Sympozjum FENS 2014, 14-16 maja 2014 r., Lublin

Granger causality and transfer entropy – two approaches to mutual causation

Causality is usually posed using two alternative scenarios:

1. the autoregressive, spectral based, causal *Granger predictive modelling [Granger C.W.J., Econometrica 37(3) (1969)], and*
2. *the information-theoretic oriented, Kullback-Leibler (K-L) divergence based,* Transfer Entropy formulation [Schreiber T., Phys. Rev. Lett. 85(2) (2000)].

Both measures concern directional causation, between coupled variables, or among multiple variables.
The two formulations are equivalent for Gaussian variables [Barnett L. et al., Phys. Rev. Lett. 103(23) (2009)]; subsequent generalisation by Hlavackova-Schindler [Appl. Math. Sci. 73(5) (2011)].

Are GC and TE equivalent in general case?

In many complex phenomena observations and models are non-Gaussian, have non-linear correlations. In that case there are no general criteria of applicability of the two methods.
The Transfer Entropy is widely used in biomedical analyses, hence studies address generalization of Gaussianity to distributions characteristic for such data (e.g. in Hlavackova-Schindler: generalized normal, log-normal, multivariate Weinman exponential distribution).
Typical empirical distribution of financial instruments returns is non-normal, asymmetric; in financial econometrics skewed t-Student distribution, or mixtures of Gaussian distributions, are often used.
To apply the TE method to checking of interdependence between *financial time series, we need to address the question of equivalency of GC and TE for such empirical distributions. Applicability and confidence estimates for causation inference for such data will be reviewed for as far as they exist, and the possible directions for improvement will be discussed.*

Granger causality according to Granger:

The notion of Granger causality ([1], [2]) also known as Wiener-Granger causality, was formalized in an autoregressive econometric model framework. Granger himself has changed his view of this concept over time. [3]:

> *Proponents of different definitions cannot agree on the basic properties required of a causal relationship: does the cause occur temporally before the effect, or can they occur simultaneously; or possibly even the cause occur after the effect? Can you have cause and effect between deterministic processes? An important question is: Does the causal definition apply to a model or to a data generating process? Sometimes a model builder will assume the model is true and so is identical to the DGP, but this is an unrealistic assumption, particularly when comparing alternative models.*

Granger causality according to Granger:

He argues that his intention in (Granger 1980) was to define G-causality

> *based on the two precepts 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 distributions. An implication of these statements was that using the cause produce a superior forecasts of the effect. Unfortunately, this implication became definition for many writing, including Hoover. However, it does provide suitable, post-sample tests as discussed in Granger (1980). G-causality is concerned with the DGP and the tests with the models.*[1]

[1] See [3] pp. 69–70.

a) Granger and Sims tests of G-causality

The two tests of G-causality – Granger test and Sims test – in simplest form are performed as joint significance tests for a set of lags of causal variable, or leads and lags of causal variable, in an ADL linear model:

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

with $H_0 : \beta_j = 0$ for all $j = 1, ..., k$ corresponds to lack of G-causality from x to y, and

$$x_t = \sum_{j=1}^{k} \alpha_j y_{t-j} + \sum_{j=-m}^{k} \beta_j x_{t-j} + \varepsilon_t \tag{2}$$

with $H_0 : \beta_j = 0$ for all $j = 1, ..., k$ corresponds to lack of G-causality from x to y.
Sims [4] 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.

b) Granger and Sims tests of G-causality, Chamberlain extension:

Chamberlain [5] extends the Granger and Sims causality definitions using *conditional independence instead of linear prediction:*

> *Definition 1.* $(G) - x_{t+1}$ *is independent of* $y_t, y_{t-1}, \ldots$ *conditional on* $x_t, x_{t-1}, \ldots$ *for all* t.
>
> *Definition 2.* $(S) - y_t$ *is independent of* $x_{t+1}, x_{t+2}, \ldots$ *conditional on* $x_t x_{t-1}, \ldots$ *for all* t,

and shows that both definitions of causality are equivalent.

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

Next way of G-causality testing is in multivariate VAR models, in which all equations have a similar form: – in bivariate case: $\mathbf{Y} = [Y_1, Y_2]^T$

$$\mathbf{Y}_t = \mathbf{A}_0 + \mathbf{A}_1\mathbf{Y}_{t-1} + ... + \mathbf{A}_1\mathbf{Y}_{t-1} + \varepsilon_t \quad (3)$$

the test of non-causality for a stationary case can be performed as the Wald test of joint insignificance of all lags of Y_2 in the equation for Y_1.

In VAR/VECM models with nonstationary/cointegrated variables:

Economic, especially financial variables often are nonstationary. The VAR model is transformed according to Johansen's [7] method, the Johansen's trace and maximum eigenvalue tests for cointegration are performed. Cointegrating relationship means that there exists a *stationary linear combination of non-stationary variables. This corresponds to a long-run stable equilibrium relationship (described as attractor by Granger).* [2]*. See Jin-Lung Lin [6] for an overview of causality-testing pitfalls and strategy both in case of stationary multivariate model and in case of nonstationary variables – strategy of cointegration and causality testing: recommends using the Johansen test in case of cointegration, and applying first differencing before causality testing when the variables are nonstationary and not cointegrated.*

[2] Notion of cointegration is generalized to include nonlinear specifications.

Note that the linear version of causality is not the only one. See more general definition ([16], who analyse the test of Granger causality in variance in presence of causality in mean):

> *Let* $\mathbf{z}_t = [z_{1t}, z_{2t}]_0^T$, $t = 1, 2, \ldots$ *denote a bivariate stationary and ergodic stochastic process. The notion of Granger causality can be classified into cathegories related to the* r *th conditional moment. Let* $z_{t-1}^0 = z_{t-1}, z_{t-2}, \ldots$ *denote past of the process.*
>
> *If* $E(\mathbf{z}_{1t}^r | z_{t-1}^0) \neq E(\mathbf{z}_{1t}^r | z_{1,t-1}^0)$, *then we may say that* z_{2t} *Granger - causes* z_{1t} *in the* r*th moment. [...] causality in mean and causality in variance are the two types of Granger causality that appear to be most relevant in economic and financial applications.*

Tests for causality in variance, proposed e.g. by Cheung and Ng [17] and by Hong [18] were aimed at revealing volatility spillovers across financial markets.

The Cheung and Ng methodology

1. For each of 2 series, estimate a model with conditional mean and conditional variance specified as ARMA(p,q) and GARCH(1,1) models, respectively. (In presence of causality in mean, modify the conditional mean to account for this dynamics.)
2. Retrieve the squared standardized residuals, $\hat{e}_{it}^2 = (z_{it} - \hat{\mu}_{it})^2 / \hat{h}_{it}^2$. Estimate the sample cross-covariance function between $\hat{e}_{1t}^2$ and $\hat{e}_{2t}^2$, $\hat{C}_{12}(j)$, and compute the sample cross-correlation function:

$$\hat{\rho}_{12}(j) = [\hat{C}_{11}(0)\hat{C}_{22}(0)]^{-0.5}\hat{C}_{12}(j)$$

   where $\hat{C}_{11}(0)$ and $\hat{C}_{22}(0)$ are the sample variances of the two variables;
3. Compute Cheung and Ng test statistic, S, based on the first M squared sample cross-correlations, $S = T\sum_{j=1}^{M} \hat{\rho}_{12}^2(j) \sim^{asy} \chi^2(M)$, under H_0 of no causality between the series. A small-sample version of S statistic is also available (weighted sum of squared cross-correlations with Bartlett weights).

As the S-test may not be fully efficient due to equal weighting to each of the sample cross-correlations, Hong (2001) proposes a weghting scheme $k(\cdot)$ that gives a larger weight to a lower lag order j; uses various kernels – with compact support (Bartlett, Parzen and Tukey-Hanning), and with unbounded support (Daniell, and Quadratic Spectral). Under appropriate regularity conditions, $Q_1 \to N(0, 1)$ under the null of non causality. Under a general class of causality in variance alternatives, $Q_1 \to \infty$ hence the Hong test is one-sided.

Pantelidis on effect of causality in mean on causality-in-variance test

Pantelidis and Pittis [16] assume that $\mathbf{z}_t = [z_{1t}, z_{2t}]_0^T$, $t = 1, 2, \ldots$ follow a bivariate VAR(1) process, with error terms given by a bivariate GARCH(1,1) process:

1. When moderate causality-in-mean are not accounted for, the causality-in-variance tests are oversized for all kernels, regardless of the choice of bandwidth parameter M; the unweighted S test statistics less sensitive than the Hong test statistics;
2. When causality-in-mean effects are filtered out by estimating the appropriate model for the conditional mean, then the causality-in-variance tests perform well, with empirical sizes close to nominal values for all bandwiths.
3. Pantelidis and Pittis results are similar in case of strong causality-in-mean;
4. For the case of strong causality-in-mean and cointegrated nonstationary variables—e.g., a pair of exchange rates or stock market prices—similar result: causality-in-variance test has good size properties if causality-in-mean effects are taken into account.

Entropy and information-based measures of dependence

There are several measures and definitions of entropy. Bruzda [8] describes Granger, Maasoumi and Racine [9] conditions 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, equals 0 for independent variables;
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 = correlation coefficient ρ or a simple function of ρ;
5. Fulfills conditions of a distance;
6. Is invariant to continuous and strictly monotonous transformations of variables.

She next describes measures used to detect nonlinear dependencies between variables:

- ρ^* – maximum correlation coefficient;
- R – mutual information coefficient;
- I_2 – generalized mutual information; and
- S_p – entropy.

The last measure was proposed by Granger, Maasoumi and Racine [9] and Maasoumi and Racine [10], based on joint and marginal distributions of the variables:

$$S_\rho = \frac{1}{2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left[(p(x,y))^{0.5} - (p_1(x)p_2(y))^{0.5}\right]^2 dxdy \qquad (4)$$

In case of *standard Gaussian distributions, this reduces to*

$$S_\rho = 1 - \left(1 - \rho^2\right)^{1.25} / \left(1 - 0.5\rho^2\right)^{1.5} \tag{5}$$

*Bruzda comments on her simulation results [**?**] 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.*

- *The mutual information coefficient was able to detect all cases of nonlinearity;*
- *The maximum correlation coefficient did almost as well, and was able to detect nonlinearity in variance;*
- *The entropy measure could distinguish between GARCH-type nonlinearity and bilinear nonlinearity.*
- *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.*

In a recent study Orzeszko [19] applies measure of dependence based on mutual information to rates of return of stock exchange indices. **Mutual information** (or relative entropy) for two stochastic variables X, Y is defined as:

$$I(X,Y) = \int\int p(x,y) log \left(\frac{p(x,y)}{p_1(x)p_2(y)} \right) dxdy \tag{6}$$

where $p(x,y)$ – density of joint distribution, $p_1(x), p_2(y)$ – marginal density distributions (see [20], [21], [22]), and is related to the Shannon entropy, H, defined as:

$$H(X,Y) = -\int log[p(x,y)]p(x,y)dxdy \tag{7}$$

$$H(X) = -\int log[p_1(x)]p_1(x)dx \tag{8}$$

$$H(Y) = -\int log[p_2(y)]p_2(y)dy \tag{9}$$

by the following relationship:

$$I(X,Y) = H(X) + H(Y) - H(X,Y) \tag{10}$$

Mutual information coefficient

Mutual information, $I(X, Y)$, may be used to identify both linear and nonlinear dependencies, to choice of lags number in a model, and to nonlinear processes forecasts. It is nonnegative, equal to zero only for independent variables.
MI can be normalized to give a mutual information coefficient:

$$R(X, Y) = 1 - exp[-2I(X, Y)]^{0.5} \tag{11}$$

for which (see Granger and Terasvirta, Granger and Lin 1994):

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 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 $Y = X_{t-j}, j = 1, 2, ...$, to detect autodependencies between a variable and its lags, not necessarily linear; Orzeszko [19] proves 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. First filters log returns with an ARMA-GARCH t-Student distribution model, and applies $R(lag)$ (and checkes for significance) for both a series and model errors. In some cases this indicated dependencies not explained by the ARMA-GARCH model.

(Linear) G-causality and transfer entropy equivalence only under Gaussianity?

Note that in bivariate **Gaussian case**, the information coefficient reduces to correlation coefficient, other information measures are supposed to reduce to a simple function of ρ. Barnett et al. [**?**] give an elegant proof of equivalence of transfer entropy and G-causality, but only for Gaussian stationary processes. For Y, X, Z,

> *transfer entropy of Y to X given Z is defined as the difference between the entropy of X conditioned on its own past and the past of Z, and its entropy conditioned, in addition, on the past of Y*

namely

$$\tau_{Y\to X|Z} = H(X|X_{t-1}, X_{t-2}, Z_{t-1}, Z_{t-2}, ...) - H(X|X_{t-1}, Z_{t-1}, X_{t-2}, Z_{t-2}, ..., Y_{t-1}, Y_{t-2}, ...) \tag{12}$$

where $H(\cdot)$ denotes entropy and $H(\cdot|\cdot)$ conditional entropy.

For multivariate Gaussian random variable

$$H(X) = 0.5\log(|\Sigma(X)|) + 0.5n\log(2\pi e) \tag{13}$$

and Barnett et al. [11] show that the conditional entropy for two jointly multivariate Gaussian variables

$$H(X|Y) = 0.5\log(|\Sigma(X|Y)|) + 0.5n\log(2\pi e) \tag{14}$$

hence is proportional to Granger causality measure. But they themselves note that the *result may depend on estimate of conditional entropy (in this case based on covariance matrix).*

Comments

- The transfer entropy measure and G-causality are often applied in context of biology, neurology and medical data analysis, where the assumption of Gaussianity and even (mean) stationarity may be fulfilled.
- Let us emphasize the assumption of Gaussianity is **not realistic in case of financial variables**: Logarithmic returns of a financial instrument $r_t = lnP_t - lnP_{t-1}$, are characterized by skewness, leptokurtic distribution, the effect of variance clustering.
- Note that in general $H(f(X)) \leq H(X)$, with equality only for Gaussian distribution,
- –that estimating of covariance matrix can be more numerically demanding in case of heteroskedasticity, long-term dependence, etc.,
- –that distribution of (error terms in models) of financial returns are better approximated by some asymmetric distribution, e.g. t-Student or a mixture of normal distributions.

Perhaps a good research strategy would be to relax assumption of Gaussianity, apply the Monte Carlo or bootstrap method (giving also critical values estimated as quantiles of a distribution), and first describe relation between entropy transfer and G-causality in stationary case, but for nonlinear ARCH behaviour (for returns of financial instruments).

Sir C.W.J. Granger (Sept. 4, 1934-May 27, 2009)

- Clive W.J. Granger was born in Swansea, Wales;
- a graduate of Nottingham University (BSc in mathematics; PhD in statistics in 1959), spend one yeat at Princeton University, since 1974 at the University of San Diego, California, as Economics professor; there he started collaboration with Dr. Robert F. Engle, 'together building one of the world's most prominent econometrics programmes"[3]"; after retirement in 2003 he was spending time working at University of Canterbury, New Zealand, where he once received a call from Stokholm...;
- ...in 2003, was awarded the Sveriges Riksbank Prize in Economic Sciencies in Memory of Alfred Nobel ("for methods of analyzing economic time series with common trends (cointegration)", together with Robert F. Engle ("for methods of analysing economic time series with time-varying volatility (ARCH)";
- In 2004, was inducted into the Order of Knight Bachelor by Queen Elizabeth II;
- In 2005, the University of Nottingham bestowed upon him yet another in a series of rare honors, renaming the schools economics and geography department the Sir Clive Granger Building.

References

Granger, C.W.J., 1969, Investigating causal relations by econometric models and cross-spectral methods, Econometrica 37, 424-438.

Granger, C.W.J., 1980, Testing for causality: a personal viewpoint, J. Econ. Dyn. Control, 2, 329–352.

Granger, C.W.J., 2003, Commentary. Some aspects of causal relationships, Journal of Econometrics, 112, 69–71.

Sims, C.A., 1972, Money, Income and Causality, The American Economic Review, 62, pp. 540–552.

Chamberlain, G., 1982, The General Equivalence of Granger and Sims causality, Econometrica, Vol. 50, No. 3, pp. 569-581.

Lin, J.-L., 2008, Notes on testing causality, Institute of Economics, Academia Sinica.

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

Bruzda, J., 2007, Procesy nieliniowe i zalenoci dugookresowe w ekonomii. Analiza kointegracji nieliniowej, Wydawnictwo Naukowe Uniwersytetu Mikoaja Kopernika, Toru.

Granger, C.W.J., Maasoumi, E., Racine, J., 2004, A Dependence Metric for Possibly Nonlinear Processes, Journal of Time Series Analysis, 25, 649–669.

Maasoumi, E., Racine, J., 2002, Entropy and Preictability of Stock Market Returns, Journal of Econometrics, 107, 291–312.

Barnett, L., Barrett, A.B., A.K. Seth, 2009, Granger causality and transfer entropy are equivalent for Gaussian variables, preprint, Centre for Computational Neuroscience and Robotics, School of Informatics, University of Sussex.

Geweke J.F., 1982, Measurement of linear dependence and feedback between multiple time series, J. Am. Stat. Assoc., 77, 304–324.

Geweke J.F., 1984, Measures of conditional linear dependence and feedback between time series, J. Am. Stat. Assoc., 79, 907–915.

Hoover, K.D., 2001, Causality in Macroeconomics, Cambridge University Press, Cambridge.

Toda, H.Y., P.C.B. Phillips, 1991, Vector Autoregression and Causality: A Theoretical Overview and Simulation Study, Cowles Foundation for

Thank you!