Cogn Neurodyn (2012) 6:33–42
DOI 10.1007/s11571-011-9175-8

RESEARCH ARTICLE

More discussions for granger causality and new causality measures

Sanqing Hu · Yu Cao · Jianhai Zhang · Wanzeng Kong · Kun Yang · Yanbin Zhang · Xun Li

Received: 24 March 2011 / Revised: 6 July 2011 / Accepted: 13 September 2011 / Published online: 27 September 2011
© Springer Science+Business Media B.V. 2011

Abstract Granger causality (GC) has been widely applied in economics and neuroscience to reveal causality influence of time series. In our previous paper (Hu et al., in IEEE Trans on Neural Netw, 22(6), pp. 829–844, 2011), we proposed new causalities in time and frequency domains and particularly focused on new causality in frequency domain by pointing out the shortcomings/limitations of GC or Granger-alike causality metrics and the advantages of new causality. In this paper we continue our previous discussions and focus on new causality and GC or Granger-alike causality metrics in time domain. Although one strong motivation was introduced in our previous paper (Hu et al., in IEEE Trans on Neural Netw, 22(6), pp. 829–844, 2011) we here present additional motivation for the proposed new causality metric and restate the previous motivation for completeness. We point out one property of conditional GC in time domain and the shortcomings/limitations of conditional GC which cannot reveal the real strength of the directional causality among three time series. We also show the shortcomings/limitations of directed causality (DC) or normalize DC for multivariate time series and demonstrate it cannot reveal real causality at all. By calculating GC and new causality values for an example we demonstrate the influence of one of the time series on the other is linearly increased as the coupling strength is linearly increased. This fact further supports reasonability of new causality metric. We point out that larger instantaneous correlation does not necessarily mean larger true causality (e.g., GC and new causality), or vice versa. Finally we conduct analysis of statistical test for significance and asymptotic distribution property of new causality metric by illustrative examples.

Keywords Granger causality · New causality · Linear regression model · Prediction

S. Hu (✉) · J. Zhang · W. Kong · K. Yang · Y. Zhang · X. Li
College of Computer Science, Hangzhou Dianzi University, Hangzhou, Zhejiang, China
e-mail: sqhu2000@yahoo.com

Y. Cao
College of Engineering and Computer Science, The University of Tennessee at Chattanooga, Chattanooga, TN 37403, USA
e-mail: yu-cao@utc.edu

Introduction

Given a set of time series, the topic of how to define causality influence among them has kept philosophers busy for over two thousand years and has yet to be completely resolved. In the literature one of the most popular definitions for causality is Granger causality (GC). Due to its simplicity and easy implementation, GC has been widely used in economics and recently in neuroscience. The basic idea of GC was originally conceived (Wiener 1956) and later formalized by Granger in the form of linear regression model (Granger 1969; Geweke 1982). It can be simply stated as follows: if the variance $(\sigma^2_{\epsilon_1})$ of the prediction error for the first time series at the present time is not less than the variance $(\sigma^2_{\eta_1})$ of the prediction error by including past measurements from the second time series in the linear regression model, then the second time series can be said to have a causal (driving) influence on the first time series. Reversing the roles of the two time series one repeats the process to address the question of driving in the opposite direction. GC value of

$\ln(\sigma_{\epsilon_1}^2/\sigma_{\eta_1}^2)$ is defined to describe the strength of the causality which the second time series has on the first one (Freiwald et al. 1999; Hesse et al. 2003; Ding et al. 2006; Oya et al. 2007; Bressler and Anil 2010). From GC value, it is clear that (i) $\ln(\sigma_{\epsilon_1}^2/\sigma_{\eta_1}^2) = 0$ when there is no causal influence from the second time series to the first one and $\ln(\sigma_{\epsilon_1}^2/\sigma_{\eta_1}^2) > 0$ when there is. This GC definition was extended to frequency domain to show causal influence from one time series on the other one in different frequencies (Geweke 1982; Ding et al. 2006). Several GC or Granger-alike causality metrics in frequency domain have been developed such as spectral GC (Geweke 1982; Ding et al. 2006), PDC (Baccal and Sameshima 2001), RPC (Yamashita et al. 2005), and DTF (Kaminski et al. 2001). To show whether the influence is a direct component or mediated by the third time series, conditional GC is defined (Geweke 1984; Freiwald et al. 1999; Hesse et al. 2003; Ding et al. 2006; Oya et al. 2007; Bressler and Anil 2010). In recent years there has been significant interest to discuss causal interactions between brain areas which are highly complex neural networks in both time and frequency domains (Freiwald et al. 1999; Hesse et al. 2003; Roebroeck et al. 2005; Oya et al. 2007; Wang et al. 2007, 2008; Atmanspacher and Rotter 2008; Gow et al. 2008, 2009; Rajagovindan and Ding 2008; Seth 2008; Cadotte et al. 2009; Zhang et al. 2010).

Although there are wide applications of GC and Granger-alike causality metrics in both time and frequency domains, some shortcomings/limitations of these metrics were pointed out in our recent paper (Hu et al. 2011). To overcome these shortcomings/limitations, we proposed new causality metrics in time and frequency domains. In particular, we emphasized that GC or Granger-alike causality metrics in frequency domain cannot reveal real strength of causality at all and new causality metric in frequency domain has advantages over spectral GC or Granger-alike causality metrics.

In this paper we will focus on time domain and show inherent shortcomings of conditional GC and DC or normalized DC. We will demonstrate that the proposed new causality metric better reveal directed real causality from one time series on the other one than conditional GC and DC or normalized DC. Thus, we must be caution in drawing any conclusion based on GC, conditional GC and DC or normalized DC by noting the shortcomings/limitations of GC (Hu et al. 2011). We will show the influence of one of the time series on the other is linearly increased as the coupling strength is linearly increased by computing GC and new causality values via an example. We will also discuss statistical test for significance and asymptotic distribution property of new causality metric.

Granger causality in time domain

In this section we introduce the well-known Granger causality and conditional Granger causality.

Given two stochastic process $X_1(t)$ and $X_2(t)$ which are assumed to be jointly stationary. Their autoregressive representations are described as

$$\begin{cases} X_{1,t} = \sum_{j=1}^{m} \mathbf{a}_{11,j} X_{1,t-j} + \epsilon_{1,t} \\ X_{2,t} = \sum_{j=1}^{m} \mathbf{a}_{22,j} X_{2,t-j} + \epsilon_{2,t} \end{cases} \tag{1}$$

and their joint representations are described as

$$\begin{cases} X_{1,t} = \sum_{j=1}^{m} a_{11,j} X_{1,t-j} + \sum_{j=1}^{m} a_{12,j} X_{2,t-j} + \eta_{1,t} \\ X_{2,t} = \sum_{j=1}^{m} a_{21,j} X_{1,t-j} + \sum_{j=1}^{m} a_{22,j} X_{2,t-j} + \eta_{2,t} \end{cases} \tag{2}$$

where $t = 0, 1, \ldots, N$, the noise terms are uncorrelated over time, ϵ_i and η_i have zero means and variances of $\sigma_{\epsilon_i}^2$ and $\sigma_{\eta_i}^2$, $i = 1, 2$. The covariance between η_1 and η_2 is defined by $\sigma_{\eta_1\eta_2} = \mathrm{cov}(\eta_1, \eta_2)$.

Now consider the first equalities in Eqs. 1 and 2. According to the original formulations (Wiener 1956; Granger 1969), if $\sigma_{\eta_1}^2$ is less than $\sigma_{\epsilon_1}^2$ in some suitable statistical sense, then X_2 is said to have a causal influence on X_1. In this case, the first equality in Eq. 2 is more accurate than that in Eq. 1 to estimate X_1. Otherwise, if $\sigma_{\eta_1}^2 = \sigma_{\epsilon_1}^2$, then X_2 is said to have no causal influence on X_1. In this case, the two equalities are almost same. Such kind of causal influence called Granger causality (GC) (Ding et al. 2006; Geweke 1982), can be defined by

$$F_{X_2 \to X_1} = \ln \frac{\sigma_{\epsilon_1}^2}{\sigma_{\eta_1}^2}. \tag{3}$$

Obviously, $F_{X_2 \to X_1} = 0$ when there is no causal influence from X_2 to X_1 and $F_{X_2 \to X_1} > 0$ when there is. Similarly, the causal influence from X_1 to X_2 is defined by

$$F_{X_1 \to X_2} = \ln \frac{\sigma_{\epsilon_2}^2}{\sigma_{\eta_2}^2}. \tag{4}$$

To show whether the interaction between two time series is direct or is mediated by another recorded time series, conditional Granger causality (Ding et al. 2006; Geweke 1984), was defined by

$$F_{X_2 \to X_1 | X_3} = \ln \frac{\sigma_{\epsilon_3}^2}{\sigma_{\eta_3}^2} \tag{5}$$

where $\sigma_{\epsilon_3}^2$ and $\sigma_{\eta_3}^2$ are variances of two noise terms, ϵ_3 and η_3, of the following two joint autoregressive representations:

$$X_{1,t} = \sum_{j=1}^{m} \mathbf{a}_{11,j} X_{1,t-j} + \sum_{j=1}^{m} \mathbf{a}_{13,j} X_{3,t-j} + \epsilon_{3,t} \tag{6}$$

and $X_{1,t} =$

$$\sum_{j=1}^{m} a_{11,j} X_{1,t-j} + \sum_{j=1}^{m} a_{12,j} X_{2,t-j} + \sum_{j=1}^{m} a_{13,j} X_{3,t-j} + \eta_{3,t}. \tag{7}$$

According to this definition, $F_{X_2 \to X_1 | X_3} = 0$ means that no further improvement in the prediction of X_1 can be expected by including past measurements of X_2. On the other hand, when there is still a direct component from X_2 to X_1, the past measurements of X_1, X_2, X_3 together result in better prediction of X_1, leading to $\sigma^2_{\eta_3} < \sigma^2_{\epsilon_3}$, and $F_{X_2 \to X_1 | X_3} > 0$.

For conditional GC, we point out one important property as follows.

Property 1 Consider the following model

$$\begin{cases} X_{1,t} = a_{11,1} X_{1,t-1} - 0.8 X_{2,t-1} + a_{13,1} X_{3,t-1} + \eta_{1,t} \\ X_{2,t} = a_{21,1} X_{1,t-1} + 0.8 X_{2,t-1} + a_{23,1} X_{3,t-1} + \eta_{2,t} \\ X_{3,t} = a_{31,1} X_{1,t-1} - 0.8 X_{2,t-1} + a_{33,1} X_{3,t-1} + \eta_{3,t} \end{cases} \tag{8}$$

where $0 < a_{i1,1}, a_{i3,1} < 0.7$, $i = 1, 2, 3$ and for simplicity η_1, η_2, η_3 are assumed to be independent white noise processes with zero mean and variances $\sigma^2_{\eta_1} = \sigma^2_{\eta_2} = \sigma^2_{\eta_3} = 1$. Figure 1 shows $F_{X_2 \to X_1 | X_3}$ for Model (8) under different parameters $a_{i1,1}$ and $a_{i3,1}$. When we calculate conditional GC it should be pointed out that for each specific Model (8) we generate a data set of 200 realizations of 10,000 time points. For each realization, we estimate AR models (joint representations Model (6) and Model (7)) with the order of 8 by using the least-squares method and calculate conditional GC where the order 8 fits well for all examples throughout the paper (see Fig. 1a and b from which one can see $F_{X_2 \to X_1 | X_3}$ keeps steady when the order of the estimated models is >8). Then we obtain the average value across all realizations and get $F_{X_2 \to X_1 | X_3}$. From Fig. 1 one can clearly see that $F_{X_2 \to X_1 | X_3}$ has nothing to do with parameters $a_{i1,1}$ and $a_{i3,1}$, $i = 1, 2, 3$ (of course, choices of parameters $a_{i1,1}$ and $a_{i3,1}$ are such that Model (8) does not diverge).

By the conditional GC definition (5), we know that conditional GC $F_{X_2 \to X_1 | X_3}$ explicitly measures the degree to which X_2 predicts X_1 over and above the degree to which X_1 already predicts itself and X_3 predicts X_1, the influence of X_1 on itself (i.e. the coefficients $a_{i1,1}$) and X_3 on X_1 (i.e. the coefficients $a_{i3,1}$) are factored out and quite correctly makes no difference to the result. So, according to the conditional GC definition (5) the above property is true.

In general GC is useful to show whether or not theoretically there is directional interaction between two neurons or among three neurons. When there exists causal influence, a question is arising: does GC value reveals the real strength of causality? to answer this question, let's consider the following simple model

$$\begin{cases} X_{1,t} = a_{12,1} X_{2,t-1} + \eta_{1,t} \\ X_{2,t} = a_{21,1} X_{1,t-1} + \eta_{2,t} \end{cases} \tag{9}$$

where η_1 and η_2 are two independent white noise processes with zero mean and $a_{12,1} a_{21,1} \neq 0$. From Model (9) one can get

$$\begin{aligned} X_{1,t} &= a_{12,1} (\overbrace{a_{21,1} X_{1,t-2} + \eta_{2,t-1}}^{X_{2,t-1}}) + \eta_{1,t} \\ &= a_{12,1} a_{21,1} X_{1,t-2} + a_{12,1} \eta_{2,t-1} + \eta_{1,t}. \end{aligned} \tag{10}$$

So, GC value $F_{X_2 \to X_1} =$

$$\ln \frac{a^2_{12,1} \sigma^2_{\eta_2} + \sigma^2_{\eta_1}}{\sigma^2_{\eta_1}} = -\ln \left(1 - \frac{a^2_{12,1} \sigma^2_{\eta_2}}{a^2_{12,1} \sigma^2_{\eta_2} + \sigma^2_{\eta_1}} \right) \tag{11}$$

$\in [0, +\infty)$ or equivalently,

$$F_{X_2 \to X_1} = \frac{a^2_{12,1} \sigma^2_{\eta_2}}{a^2_{12,1} \sigma^2_{\eta_2} + \sigma^2_{\eta_1}} \in [0, 1] \tag{12}$$

which is only related to the last two noise terms and has nothing to do with the first term. It is noted that all of three terms make contributions to current $X_{1,t}$, and $a_{12,1} a_{21,1} X_{1,t-2}$ must have causal influence on $X_{1,t}$ and must be considered to illustrate real causality. Especially, when $\sigma^2_{\eta_2} = 0$, we have $X_{1,t} = a_{12,1} a_{21,1} X_{1,t-2} + \eta_{1,t}$ and $F_{X_2 \to X_1} = 0$. Since $a_{21,1} X_{1,t-2}$ comes from $X_{2,t-1}$ we can surely know that X_2 has real nonzero causality on X_1 due to $a_{12,1} a_{21,1} \neq 0$. Thus, GC value may not reveal real causality at all. As such, the definition has its inherent shortcomings and/or limitations to illustrate the real strength of causality. Please refer to Remark 1 (Hu et al. 2011) for the detailed discussions.

New causality

Due to the shortcomings and/or limitations of GC, we next give a new definition for causality anaylsis of multivariate time series. Let's consider the following general model:

$$\begin{cases} X_{1,t} = \sum_{j=1}^{m} a_{11,j} X_{1,t-j} + \cdots + \sum_{j=1}^{m} a_{1n,j} X_{n,t-j} + \eta_{1,t} \\ X_{2,t} = \sum_{j=1}^{m} a_{21,j} X_{1,t-j} + \cdots + \sum_{j=1}^{m} a_{2n,j} X_{n,t-j} + \eta_{2,t} \\ \vdots \\ X_{n,t} = \sum_{j=1}^{m} a_{n1,j} X_{1,t-j} + \cdots + \sum_{j=1}^{m} a_{nn,j} X_{n,t-j} + \eta_{n,t} \end{cases} \tag{13}$$

where $X_i (i = 1, \ldots, n)$ are n time series, $t = 0, 1, \ldots, N$, η_i has zero mean and variance of $\sigma^2_{\eta_i}$ and $\sigma_{\eta_i \eta_k} = \mathrm{cov}(\eta_i, \eta_k)$, $i, k = 1, \cdots, n$. Based on Eq. 13, Fig. 2 clearly shows contributions to $X_{k,t}$, which includes $\sum_{j=1}^{m} a_{k1,j} X_{1,t-j}, \ldots, \sum_{j=1}^{m} a_{kn,j} X_{n,t-j}$ and the

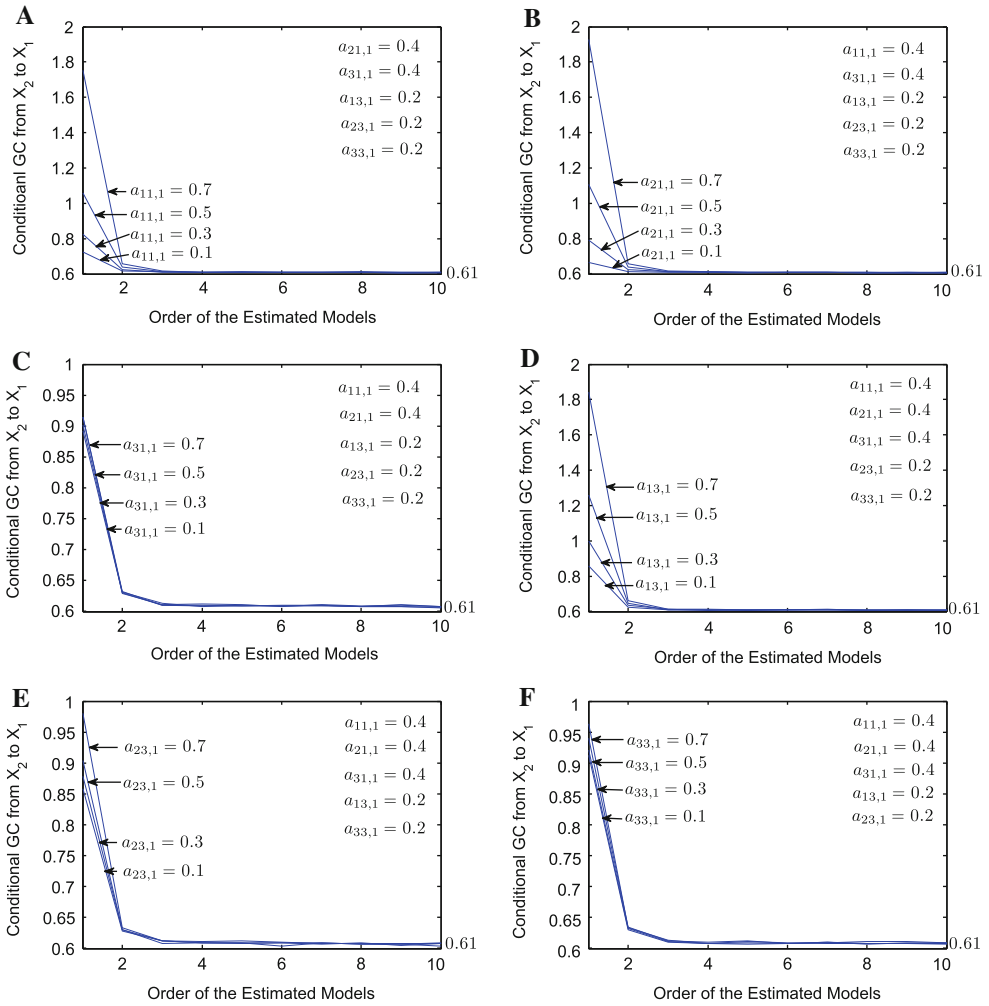

Fig. 1 **a** $F_{X_2\to X_1|X_3}$ as a function of the order of the estimated models for Model (8) when $a_{21,1} = a_{31,1} = 0.4$, $a_{13,1} = a_{23,1} = a_{33,1} = 0.2$ and $a_{11,1}$ changes from 0.1 to 0.7. **b** $F_{X_2\to X_1|X_3}$ as a function of the order of the estimated models for Model (8) when $a_{11,1} = a_{31,1} = 0.4$, $a_{13,1} = a_{23,1} = a_{33,1} = 0.2$ and $a_{21,1}$ changes from 0.1 to 0.7. **c** $F_{X_2\to X_1|X_3}$ as a function of the order of the estimated models for Model (8) when $a_{11,1} = a_{21,1} = 0.4$, $a_{13,1} = a_{23,1} = a_{33,1} = 0.2$ and $a_{31,1}$ changes from 0.1 to 0.7. **d** $F_{X_2\to X_1|X_3}$ as a function of the order of the estimated models for Model (8) when $a_{11,1} = a_{21,1} = a_{31,1} = 0.4$, $a_{23,1} = a_{33,1} = 0.2$ and $a_{13,1}$ changes from 0.1 to 0.7. **e** $F_{X_2\to X_1|X_3}$ as a function of the order of the estimated models for Model (8) when $a_{11,1} = a_{21,1} = a_{31,1} = 0.4$, $a_{13,1} = a_{33,1} = 0.2$ and $a_{23,1}$ changes from 0.1 to 0.7. **f** $F_{X_2\to X_1|X_3}$ as a function of the order of the estimated models for Model (8) when $a_{11,1} = a_{21,1} = a_{31,1} = 0.4$, $a_{13,1} = a_{23,1} = 0.2$ and $a_{33,1}$ changes from 0.1 to 0.7. From **a–f** one can see that *a*) $F_{X_2\to X_1|X_3}$ keeps steady and converges to 0.61 when the order $p > 4$. *b* $F_{X_2\to X_1|X_3}$ has nothing to do with parameters $a_{i1,1}$ and $a_{i3,1}$

noise term $\eta_{k,t}$ where the influence from $\sum_{j=1}^{m} a_{kk,j}X_{k,t-j}$ is causality from X_k's own past values. Each contribution plays an important role in determining $X_{k,t}$. If $\sum_{j=1}^{m} a_{ki,j}X_{i,t-j}$ occupies larger portion among all those contributions, then X_i has stronger causality on X_k, or vice versa. Thus, a good definition for causality from X_i to X_k in time domain should be able to describe what proportion X_i occupies among all these contributions. This is a general guideline for proposing any causality method (i.e., all contributions must be considered). For Eq. 10, let's define

$$\bar{\eta}_t \stackrel{\Delta}{=} a_{12,1}\eta_{2,t-1} + \eta_{1,t} \tag{14}$$

which is summation of two noise terms and each noise term makes contributions to $\bar{\eta}$. To describe what proportion η_2 occupies in $\bar{\eta}$, we define

$$\frac{a_{12,1}^2 \sum_{t=1}^{N} \eta_{2,t-1}^2}{a_{12,1}^2 \sum_{t=1}^{N} \eta_{2,t-1}^2 + \sum_{t=1}^{N} \eta_{1,t}^2} = \frac{a_{12,1}^2 \sigma_{\eta_2}^2}{a_{12,1}^2 \sigma_{\eta_2}^2 + \sigma_{\eta_1}^2} \tag{15}$$

which is the same as GC defined in Eq. 12. Therefore, here in nature GC is actually defined based on the noise model

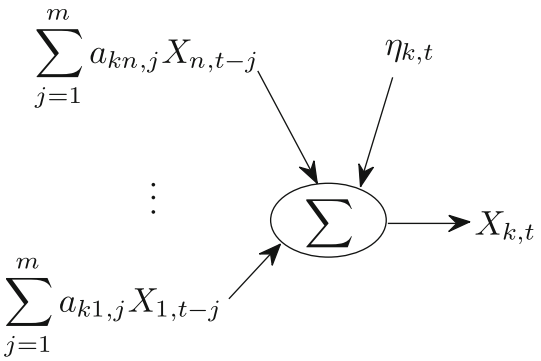

Fig. 2 Contributions to $X_{k,t}$

(14) and follows the above guideline. Motivated by this idea, we can naturally extend the noise model (14) to the kth equation of Model (13) and define a new direct causality from X_i to X_k as follows:

$$n_{X_i \overset{D}{\to} X_k} = \frac{\sum_{t=m}^{N} \left(\sum_{j=1}^{m} a_{ki,j} X_{i,t-j} \right)^2}{\sum_{h=1}^{n} \sum_{t=m}^{N} \left(\sum_{j=1}^{m} a_{kh,j} X_{h,t-j} \right)^2 + \sum_{t=m}^{N} \eta_{k,t}^2}. \tag{16}$$

When N is large enough,

$$\sum_{t=m}^{N} \eta_{k,t}^2 = \sum_{t=1}^{N} \eta_{k,t}^2 - \sum_{t=1}^{m} \eta_{k,t}^2 = N\sigma_{\eta_k}^2 - \sum_{t=1}^{m} \eta_{k,t}^2 \approx N\sigma_{\eta_k}^2.$$

Then, Eq. 16 can be approximated as

$$\frac{\sum_{t=m}^{N} \left(\sum_{j=1}^{m} a_{ki,j} X_{i,t-j} \right)^2}{\sum_{h=1}^{n} \sum_{t=m}^{N} \left(\sum_{j=1}^{m} a_{kh,j} X_{h,t-j} \right)^2 + N\sigma_{\eta_k}^2}. \tag{17}$$

Throughout the paper, we always assume that N is large enough so that $n_{X_i \overset{D}{\to} X_k}$ is always defined as Eq. 17.

New causality based on Eq. 10 can be written as

$$n_{X_2 \overset{D}{\to} X_1} = \frac{\sum_{t=1}^{N} \left(a_{12,1} X_{2,t-1} \right)^2}{\sum_{t=1}^{N} \left(a_{12,1} X_{2,t-1} \right)^2 + \sum_{t=1}^{N} \eta_{1,t}^2} = \frac{\sum_{t=1}^{N} \left(a_{12,1} X_{2,t-1} \right)^2}{\sum_{t=1}^{N} \left(a_{12,1} X_{2,t-1} \right)^2 + N\sigma_{\eta_1}^2}, \tag{18}$$

which describes what proportion X_2 occupies among two contributions in X_1 (see Eq. 10). Note that for (10) GC $F_{X_2 \to X_1}$ is proposed based on Model (14) and describes what proportion η_2 occupies among two contributions in $\bar{\eta}$ (see Eq. 15). Thus GC actually reveals causal influence from η_2 to $\bar{\eta}$, it does not reveal causal influence from X_2 to X_1 at all by noting that $\bar{\eta}$ is only partial information of X_1, i.e., the noise terms $a_{12,1}\eta_{2,t-1} + \eta_{1,t}$. One can see that new causality definition based on Eq. 10 is a natural extension of GC definition based on Eq. 14 as far as the concept of proportion is concerned (in this way, our new causality definition has a rather sound conceptual or theoretical basis). But, as said above, the equivalent GC describes what the proportion η_2 occupies in $\bar{\eta}$ of Eq. 14 and new causality describes what the proportion X_2 occupies in X_1 of Eq. 10. Thus, they are two totally different concepts, that is, the equivalent GC reveals causal influence from η_2 to $\bar{\eta}$, but new causality indeed reveals causal influence from X_2 to X_1. Obviously, one cannot use causal influence value from η_2 to $\bar{\eta}$ (involving partial information of X_1, i.e., the noise terms $a_{12,1}\eta_{2,t-1} + \eta_{1,t}$) to express causality influence value from X_2 to X_1 (involving complete information, i.e.,the noise terms $a_{12,1}\eta_{2,t-1} + \eta_{1,t}$ plus $a_{12,1}a_{21,1}X_{1,t-2}$). Any causality definition (like traditional GC) only using partial information of X_1 inevitably leads to misinterpretation result. In addition to some important points shown in Remark 2 (Hu et al. 2011), next we give some more comments for new causality in the following remark.

Remark 1 (1) Let the transfer function of Model (13) be $H(f) = [H_{ij}(f)]_{n \times n}$. RPC (Yamashita et al. 2005) is used to reveal causality influence from X_j to X_i at frequency f and is defined as

$$R_{i \leftarrow j}(f) = \frac{|H_{ij}(f)|^2 \sigma_{\eta_j}^2}{S_{X_i X_i}(f)} \tag{19}$$

where the power spectrum

$$S_{X_i X_i}(f) = \sum_{j=1}^{n} |H_{ij}(f)|^2 \sigma_{\eta_j}^2, \ (i = 1, \ldots, n). \tag{20}$$

Equation 20 indicates that the power spectrum of $X_{i,t}$ at frequency f can be decomposed as to n terms $|H_{ij}(f)|^2 \sigma_{\eta_j}^2, (j = 1, \ldots, n)$, each of which can be interpreted as the power contribution of the jth innovation $\eta_{j,t}$ transferring to $X_{i,t}$ via the transfer function $H_{ij}(f)$. RPC $R_{i \leftarrow j}(f)$ can be regarded as a ratio of the power contribution of the innovation $\eta_{j,t}$ on the power spectrum of $X_{i,t}$ to the power spectrum $S_{X_i X_i}(f)$. From this point of view, this ratio also provides a strong motivation (or theoretical basis) to define the proposed new causality metric Eq. 16 for multivariate time series. However, as pointed out in Remark 6 (Hu et al. 2011) RPC has inherent shortcomings/limitations and cannot reveal real causality influence at all.

(2) Consider the following two models

$$\begin{cases} X_{1,t} = -0.99X_{2,t-1} + \eta_{1,t} \\ X_{2,t} = 0.1X_{2,t-1} - 0.99X_{3,t-1} + \eta_{2,t} \\ X_{3,t} = 0.1X_{2,t-1} - 0.99X_{3,t-1} + \eta_{3,t} \end{cases} \tag{21}$$

and

$$\begin{cases} \bar{X}_{1,t} = -0.99\bar{X}_{2,t-1} + \bar{\eta}_{1,t} \\ \bar{X}_{2,t} = 0.1\bar{X}_{2,t-1} + \bar{\eta}_{2,t} \\ \bar{X}_{3,t} = 0.1\bar{X}_{2,t-1} + \bar{\eta}_{3,t} \end{cases} \quad (22)$$

where η_1, η_2 and η_3 are three independent white noise processes with zero mean and variances $\sigma^2_{\eta_1} = 1 = \sigma^2_{\eta_3}$, $\sigma^2_{\eta_2} = 0.1, \eta_{1,t} = \bar{\eta}_{1,t}, \eta_{2,t} = \bar{\eta}_{2,t}, \eta_{3,t} = \bar{\eta}_{3,t}$, and the initial conditions $X_{1,0} = \bar{X}_{1,0}, X_{2,0} = \bar{X}_{2,0}$ and $X_{3,0} = \bar{X}_{3,0}$. From both of Models (21) and (22) it can be seen that there are no direct causality from X_3 to X_1, so, $F_{X_2 \to X_1|X_3} = F_{X_2 \to X_1}$. Moreover, based on Property 1, $F_{X_2 \to X_1|X_3}$ are same for both of Models (21) and (22). We can obtain $F_{X_2 \to X_1|X_3} = 0.092$ for both of Models (21) and (22), $n_{X_2 \overset{\mathrm{D}}{\to} X_1} = 0.824$ for Model (21), and $n_{X_2 \overset{\mathrm{D}}{\to} X_1} = 0.090$ for Model (22). Figure 3 shows trajectories $-0.99X_2, -0.99\bar{X}_2$, and η_1 for one realization of Model (21) and Model (22). From Fig. 3a and c one can clearly see that amplitudes of $-0.99X_2$ are much larger than that of η_1 and the contribution from $-0.99X_{2,t-1}$ occupies much larger portion compared to that from $\eta_{1,t}$, as a result, the causal influence from X_2 to X_1 occupies a major portion compared to the influence from η_1 and the real strength of causality from X_2 to X_1 should have higher value. This fact is real. Our causality value $n_{X_2 \overset{\mathrm{D}}{\to} X_1} = 0.824$ for Model (21) is consistent with this fact. Similarly, from Fig. 3b and c one can clearly see that amplitude of $-0.99\bar{X}_2$ is much smaller than that of η_1 and the contribution from $-0.99\bar{X}_{2,t-1}$ occupies much smaller portion compared to that from $\eta_{1,t}$, as a result, the causal influence from $\bar{X}_2$ to $\bar{X}_1$ occupies a rather small portion compared to the influence from η_1 and the real strength of causality from $\bar{X}_2$ to $\bar{X}_1$ should have smaller value. This fact is also real. Our causality value $n_{X_2 \overset{\mathrm{D}}{\to} X_1} = 0.090$ for Model (22) is consistent with this fact. However, conditional GC always equals to 0.092 for both of Models (21) and Eq. 22 and does not reflect such kind of changes at all, and violates above two real facts. These results show that the conditional GC definition (5) does not reveal real strength of direct causality from X_2 to X_1 at all for Models (21) and (22), our new causality definition very reasonably reflects the real strength of the direct causality.

(3) An alternate time-domain metric namely direct causality (DC) has been proposed earlier (Kaminski et al. 2001) which quantifies causality based on the AR coefficients

$$DC_{X_k \to X_i} = \sum_{j=1}^{m} a^2_{ik,j}$$

or the following normalized measure

$$DC_{X_k \to X_i} = \frac{\sum_{j=1}^{m} a^2_{ik,j}}{\sum_{k=1}^{n}\sum_{j=1}^{m} a^2_{ik,j}}. \quad (23)$$

Since the AR coefficients themselves represent the coupling strength, now a question is arising: would this normalized measure suffice to reveal the real causality of two time series? Unfortunately, the answer is no. Let's take a look at the following model:

$$\begin{cases} X_{1,t} = 0.1X_{1,t-1} + 0.9(-X_{2,t-1} + X_{2,t-2} - X_{2,t-3} \\ \quad +X_{2,t-4}) + 0.9X_{3,t-1} + \eta_{1,t} \\ X_{2,t} = \eta_{2,t} \\ X_{3,t} = -0.1X_{1,t-1} + 0.1X_{2,t-1} - 0.9X_{3,t-1} + \eta_{3,t} \end{cases} \quad (24)$$

where η_1, η_2 and η_3 are three independent white noise processes with zero mean and variances $\sigma^2_{\eta_1} = 0.1 = \sigma^2_{\eta_2}, \sigma^2_{\eta_3} = 1$. For Models (24) we can obtain $DC_{X_2 \to X_1} = 0.78$ and $n_{X_2 \overset{\mathrm{D}}{\to} X_1} = 0.11$. Figure 4 shows trajectories of $X_{1,t}$, $0.9(-X_{2,t-1} + X_{2,t-2} - X_{2,t-3} + X_{2,t-4})$, $0.9X_{3,t}$, and $\eta_{1,t}$ for one realization of Model (24). From Fig. 4a–d one can clearly see that amplitudes of $0.9X_3$ are much larger than that of $\eta_{1,t}$ and $0.9(-X_{2,t-1} + X_{2,t-2} - X_{2,t-3} + X_{2,t-4})$, and the contribution from $0.9X_3$ occupies much larger portion compared to that from $\eta_{1,t}$ and $0.9(-X_{2,t-1} + X_{2,t-2} - X_{2,t-3} + X_{2,t-4})$. As a result, the causal influence from X_2 to X_1 occupies a small portion compared to the influence from $0.9X_3$ and the real strength of causality from X_2 to X_1 should have smaller value. This fact is real. Our causality value $n_{X_2 \overset{\mathrm{D}}{\to} X_1} = 0.11$ for Model (24) is consistent with this fact. However, $DC_{X_2 \to X_1} = 0.78$ for Model (24) violates the real fact. These results show that the direct causality definition (23) does not reveal real strength of direct causality from X_2 to X_1 at all for Model (24), our new causality definition reflects the real strength of the direct causality very reasonably.

(4) It should be noted that the influence of one of the time series on the other is linearly increased as the coupling strength is linearly increased. Let's consider the following bivariate AR process with 5 levels of coupling strength:

$$\begin{cases} X_{1,t} = -0.1X_{1,t-1} + c \times 0.1(-X_{2,t-1} \\ \quad +X_{2,t-2} - X_{2,t-3} + X_{2,t-4}) + \eta_{1,t} \\ X_{2,t} = -0.5X_{1,t-1} + 0.5X_{2,t-1} + \eta_{2,t} \end{cases} \quad (25)$$

where η_1 and η_2 are two independent white noise processes with zero mean and variances $\sigma^2_{\eta_1} = 1 = \sigma^2_{\eta_2}$. From Eq. 25 one can see that the coupling strength of X_2 on X_1 is linearly increased as the parameter c increases. The true causality of X_2 on X_1 should increase as the parameter c increases. The estimated causality measures (new

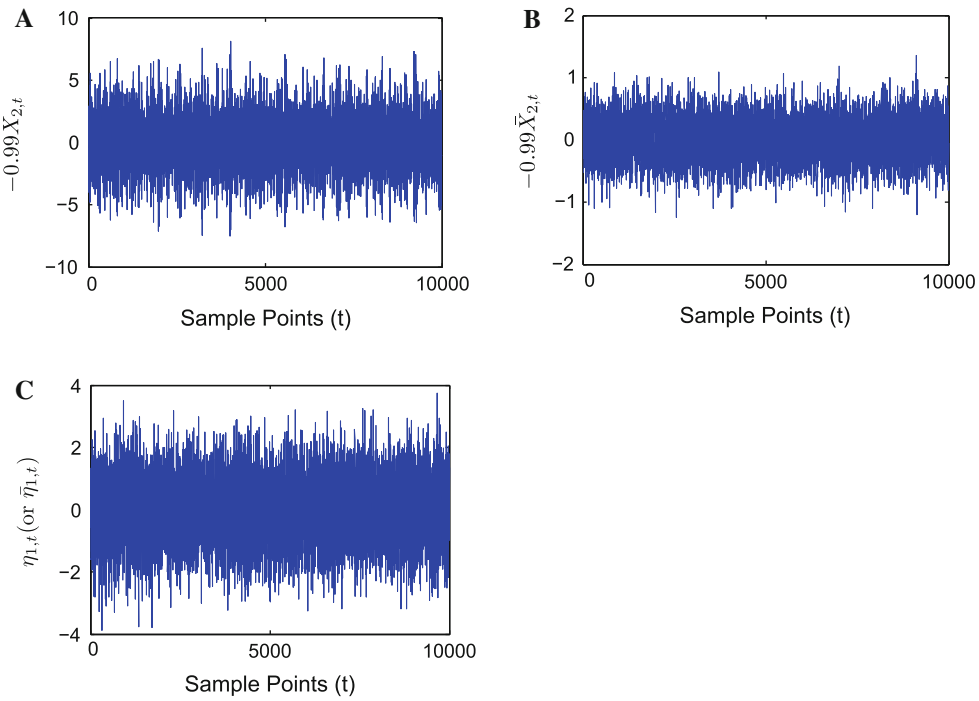

Fig. 3 For one realization of Model (21) and Model (22) where $X_{i,0} = \bar{X}_{i,0}$ and $\eta_{i,t} = \bar{\eta}_{i,t}, i = 1, 2, 3$, trajectories for $-0.99X_{2,t}$, $-0.99\bar{X}_{2,t}$, and $\eta_{1,t}$ are plotted: **a** $-0.99X_{2,t}$'s trajectory for Model (21). **b** $-0.99\bar{X}_{2,t}$'s trajectory for Model (22). **c** $\eta_{1,t}$'s or $\bar{\eta}_{1,t}$'s trajectory

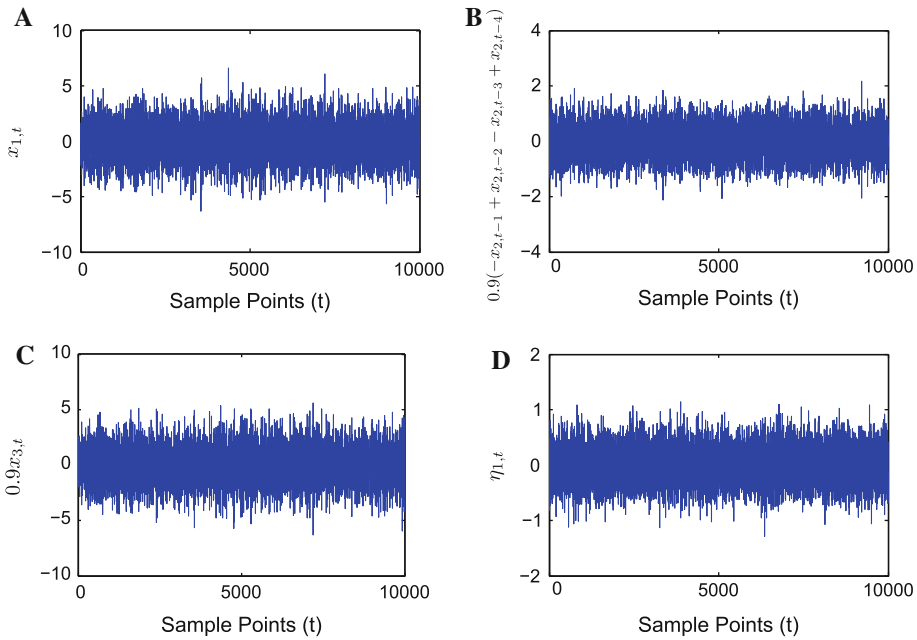

Fig. 4 For one realization of Model (24), trajectories for $X_{1,t}$, $0.9(-X_{2,t-1} + X_{2,t-2} - X_{2,t-3} + X_{2,t-4})$, $0.9X_{3,t}$, and $\eta_{1,t}$ are plotted: **a** Trajectory for $X_{1,t}$. **b** Trajectory for $0.9(-X_{2,t-1} + X_{2,t-2} - X_{2,t-3} + X_{2,t-4})$. **c** Trajectory for $0.9X_{3,t}$. **d** Trajectory for $\eta_{1,t}$

causality and GC) against the parameter c are shown in Fig. 5a and b, respectively, from which one can clearly see that both of two metrics indicate the increasing true causality as the true coupling strength (i.e., the parameter c) increases.

(5) Figure 5c shows the increasing instantaneous correlation (i.e., zero-lag influence) of two time series as the parameter c increases. Combining Fig. 5a and b with c we can conclude that the increasing instantaneous correlation leads to the reasonable increasing new causality and GC. However, in general this conclusion may not be true. It is well known that in general there may be no relationship between true causality and correlation. In other words, larger correlation does not necessarily mean larger true causality, or vice versa. Therefore, it is meaningless to discuss whether new causality or GC will be affected by correlation.

Statistical test for significance

Since new causality metric has a highly nonlinear relation to the time series data, it makes tests of significance difficult to perform. In this section, we use the same

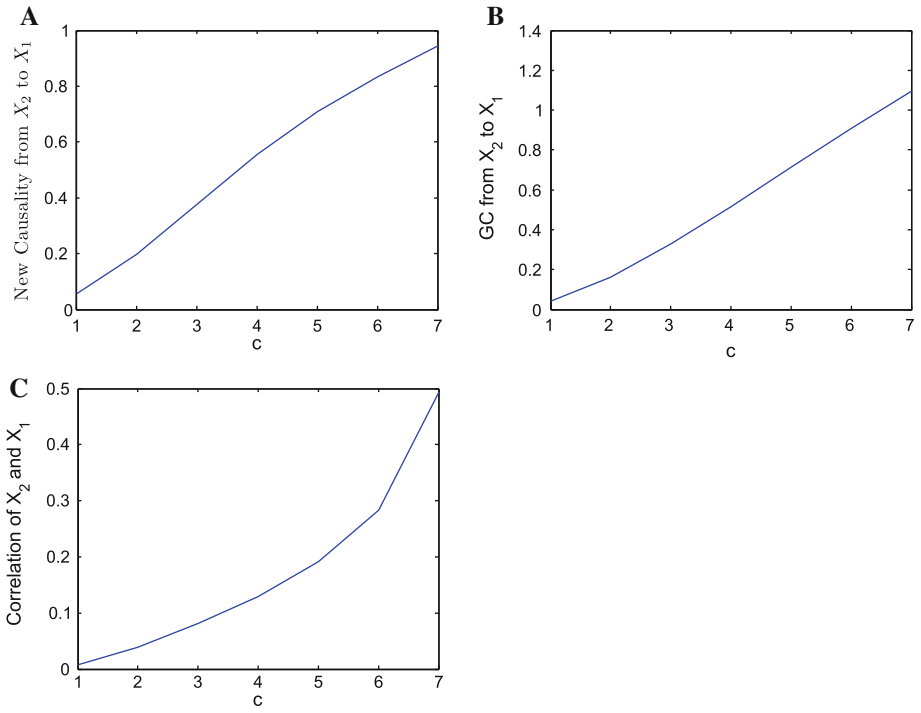

Fig. 5 Causality (or correlation) vs the parameter c for Model (25). **a** New causality of X_2 on X_1 vs the coupling strength (i.e., the parameter c). **b** GC of X_2 on X_1 vs the coupling strength (i.e., the parameter c). **c** Correlation of X_2 and X_1 vs. the coupling strength (i.e., the parameter c)

technology (Kaminski et al. 2001) to deal with this problem; that is, we first create a surrogate data set (Theiler et al. 1992) for each channel, we then fit a model to this surrogate data set and calculate new causality metric. Repeating this process many time, we can obtain an empirical distribution for the metric. Based on this distribution we can then evaluate the significance of the causal metric derived from the actual data. We use the following model to illustrate the process.

$$\begin{cases} X_{1,t} = 0.5X_{1,t-1} + 0.5X_{2,t-1} + \eta_{1,t} \\ X_{2,t} = -0.5X_{1,t-1} + 0.5X_{2,t-1} + \eta_{2,t} \end{cases} \tag{26}$$

where η_1 and η_2 are two independent white noise processes with standard normal distribution. For a simulated data of a realization of Model (26), we produce surrogate data (i.e., perform random and independent shuffling of X_1 and X_2). We then estimate a joint regression model (2) with $m = 6$ to fit the shuffled data and calculate new causality value. Carrying out the procedure for 1,000 such independently shuffled data sets we can construct an empirical distribution for the new causality values. Since the shuffling procedure destroys all the temporal structure in the data, this empirical distribution gives the variability for the new causality values when the null hypothesis of no causal influence is true.

Figure 6 shows Histogram of new causality values based on the original simulated data and 1000 surrogate data. One can see that the new causality value (=0.275) based on the original simulated data is statistically significant ($P < 0.001$).

Asymptotic distribution property

It is well known that the asymptotic distribution of GC metric under the null hypothesis of zero causality is known (it is a χ^2-distribution) (Geweke 1982). Now a question is arising: what is asymptotic distribution for new causality metric under the null hypothesis of zero causality? For new causality metric can we derive the similar asymptotic distribution property as GC metric? In general it is difficult to discuss this issue because of complexity of new causality definition. But in the following example we find that new causality metric has $\chi^2(1)$-distribution as GC metric.

$$\begin{cases} X_{1,t} = -0.5X_{1,t-1} + \eta_{1,t} \\ X_{2,t} = 0.5X_{1,t-1} - 0.5X_{2,t-1} + \eta_{2,t} \end{cases} \tag{27}$$

where η_1 and η_2 are two independent white noise processes with standard normal distribution, $t = 1, 2, \ldots, 10,000$ (i.e., $N = 10{,}000$). For each realization of Model (27) we set $m = 1$ to estimate the autoregression Model (1) and the joint regression Model (2), then based on the two estimated models we calculate new causalty $n_{X_2 \to X_1}$ and GC $F_{X_2 \to X_1}$. This process is repeated by $N = 10{,}000$ times. Finally we plot large-sample distributions for $N \times n_{X_2 \to X_1}$ and $N \times F_{X_2 \to X_1}$ in Fig. 7a and b, respectively. One can see probability density functions for both new causality and GC follow $\chi^2(1)$-distribution. In general, as for whether or not $N \times n_{X_2 \to X_1}$ follows a χ^2-distribution under the null hypothesis of zero causality keeps unknown and is left for further study in near future.

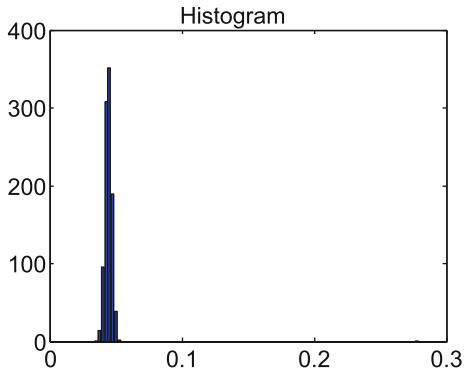

Fig. 6 Histogram of new causality values resulting from 1,000 surrogate data sets

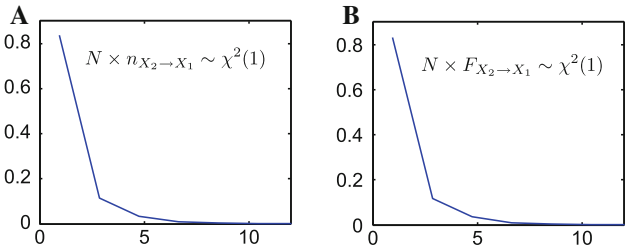

Fig. 7 Probability distribution functions (PDF). **a** PDF for new causality metric. **b** PDF for GC metric

Conclusions

In this paper, we further make more discussions for GC or Granger-alike causality and new causality metrics by following our previous paper (Hu et al. 2011) and focus on time domain. In addition to the previous motivation for introducing new causality metric in time domain, we provide additional strong motivation (or theoretical basis) to the proposed metric for multivariate time series. After introducing one property for conditional GC, we point out the inherent shortcomings/limitations of conditional GC and demonstrate that it cannot reveal directed causal influence from one time series to the other one among three time series. We also point out the shortcomings/limitations of Granger-alike causality, i.e., DC or normalized DC metric and show that it cannot disclose directed causality influence at all for multivariate time series. Furthermore, all these shortcomings/limitations demonstrate reasonability and advantages of new causality metric. Therefore, researchers must be caution in drawing any conclusion based on DC (or normalized DC) value and conditional GC value. By calculating GC and new causality values for an example we show the influence of one of the time series on the other is linearly increased as the coupling strength is linearly increased. Finally for new causality metric in time domain we analyze statistical test for significance and show significance level for a simulated data, meanwhile we conduct analysis for asymptotic distribution property by an examples in which both GC and new causality metrics follow $\chi^2(1)$-distribution. However, in general, as for whether or not new causality metric follows a χ^2-distribution under the null hypothesis of zero causality keeps unknown and is left for further study in near future. Thus our proposed new causality metrics in time and frequency domains may have wide applications in economics and neuroscience.

Acknowledgments This work was supported in part by the National Natural Science Foundation of China under Grant 61070127, and the International Cooperation Project of Zhejiang Province, China, under Grant 2009C14013, Qianjiang Project of Zhejiang Province, China, under Grant 2011R10063, U.S. National Science Foundation (NSF) under Grant 0821820, and Tennessee Higher Education Commission, the State of Tennessee, USA.

References

Atmanspacher H, Rotter S (2008) Interpreting neurodynamics: concepts and facts. Cogn Neurodyn 2:297–318

Baccal LA, Sameshima K (2001) Partial directed coherence: a new concept in neural structure determination. Biol Cybern 84(6): 463–474

Bressler SL, Anil KS (2010) Wiener-Granger causality: a well established methodology. NeuroImage (in press)

Cadotte AJ, Mareci TH, DeMarse TB et al (2009) Temporal lobe epilepsy: anatomical and effective connectivity. IEEE Trans Neural Syst Rehabil Eng 17(3):214–223

Ding M, Chen Y, Bressler SL Granger causality: basic theory and applications to neuroscience. In: Schelter B, Winterhalder M, Timmer J (eds) Handbook of time series analysis. pp 437–460 Wiley, Weinheim, (2006)

Freiwald WA, Valdes P, Bosch J et al (1999) Testing non-linearity and directedness of interactions between neural groups in the macaque inferotemporal cortex. J Neurosci Methods 94(1): 105–119

Geweke J (1982) Measurement of linear-dependence and feedback between multiple time series. J Am Stat Assoc 77(378):304–313

Geweke J (1984) Measures of conditional linear dependence and feedback between time series. J Am Stat Assoc 79(388):907–915

Granger CWJ (1969) Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37(4):424–438

Gow DW, Segawa JA, Alfhors S et al (2008) Lexical influences on speech perception: a Granger causality analysis of MEG and EEG source estimates. Neuroimage 43(3):614–623

Gow DW, Keller CJ, Eskandar E et al (2009) Parallel versus serial processing dependencies in the perisylvian speech network: a Granger analysis of intracranial EEG data. Brain Lang 110(1):43–48

Hesse W, Oller M, Arnold M et al (2003) The use of time-variant EEG Granger causality for inspecting directed interdependencies of neural assemblies. J Neurosci Methods 124(1):27–44

Hu S, Dai G, Worrell GA et al (2011) Causality analysis of neural connectivity: critical examination of existing methods and advances of new methods. IEEE Trans on Neural Netw 22(6):829–844

Kaminski M, Ding M, Truccolo-Filho W et al (2001) Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. Biol Cybern 85:145–157

Oya H, Poon PWF, Brugge JF et al (2007) Functional connections between auditory cortical fields in humans revealed by Granger causality analysis of intra-cranial evoked potentials to sounds: comparison of two methods. Biosystems 89:198–207

Rajagovindan R, Ding M (2008) Decomposing neural synchrony: toward an explanation for near-zero phase-lag in cortical oscillatory networks. Plos one 3(11):e3649

Roebroeck A, Formisano E, Goebel R (2005) Mapping directed influence over the brain using Granger causality and fMRI. Neuroimage 25(1):230–242

Seth AK (2008) Causal networks in simulated neural systems. Cogn Neurodyn 2:49–64

Wiener N The theory of prediction. In: Beckenbach EF (eds) Modern mathematics for engineers, Chap. 8.. McGraw-Hill, New York (1956)

Wang X, Chen Y, Ding M (2008) Estimating Granger causality after stimulus onset: a cautionary note. Neuroimage 41(3):767–776

Wang X, Chen Y, Bressler SL et al (2007) Granger causality between multiple interdependent neurobiological time series: blockwise versus pairwise methods. Int J Neural Syst 17(2):71–78

Zhang L, Zhong G, Wu Y et al (2010) Using granger-geweke causality model to evaluate the effective connectivity of primary motor cortex, supplementary motor area and cerebellum. J Biomed Sci Eng 3:848–860

Yamashita O, Sadato N, Okada T et al (2005) Evaluating frequency-wise directed connectivity of BOLD signals applying relative power contribution with the linear multivariate time-series models. Neuroimage 25(2):478–490