Caution on causality analysis of ERP data

Jikan Xue[1,a], Sanqing Hu[1,b], Jianhai Zhang[1], Wanzeng Kong[1]

[1]Colleage of Computer Science, Hangzhou Dianzi University, Hangzhou, China

[a]email:xuejikan@163.com, [b]email:sqhu@hdu.edu.cn

Keywords: Granger causality, New causality, Event-Related Potential.

Abstract. The event-related potential (ERP) technique is one of the most widely used methods in cognitive neuroscience research to study the physiological correlates of sensory, perceptual and cognitive activity associated with processing information [1,2]. An ERP eeg segment data usually involves three subsegments: pre-stimulus,stimulus and post-stimulus. Obviously, the mathematic models for pre-stimulus (or post-stimulus) and stimulus should be different. However, the general approach for ERP analysis in the literature always views the whole segment data as one model. In this paper we use Granger causality (GC) and recently proposed New causality (NC) methods to demonstrate that NC is more sensitive than GC to reveal true causality influence based on ERP data.

Introduction

In the literature Granger causality (GC) has been widely applied to detect the directional influence of system components in many different areas, such as economics, climate studies, genetics, and neuroscience.The basic idea of GC was originally conceived by Wiener in 1956 [3], and later formalized by Granger in 1969 [4] in the form of linear regression model. The idea can be briefly described as follows: If the historical information of time series Y significantly improves the prediction accuracy of the future of time series X in a multivariate autoregressive (MVAR) model, then GC from time series Y to X can be identified. Although GC has tremendous applications in many areas, this success has also been accompanied by criticism from different prospectives [5] and [6]. The criticism of GC has most been centered around the philosophical debate on the relationship between GC and true causality. In 2011, Hu et. al [7] proposed a new causality (NC) method which describes the proportion that Y occupies among all contributions to X.As demonstrated by a number of illustrative examples in [7] which include experimental EEG data, NC is much more sensitive than GC to reveal true causality (or trend of true causality). In this paper we will provide more evidence to show that NC method is better than GC method to reveal true causality.

GC and NC Methods

Consider two stochastic time series which are assumed to be jointly stationary. Individually, under fairly general conditions,each time series admits an autoregressive representation

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

and their joint representations are described a

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

where t = 0,1,⋯ ,N, the noise terms are uncorrelated over time,ε_i and η_i have zero means and variances of $\sigma^2(\varepsilon_i)$,and $\sigma^2(\eta_i)$, i = 1,2. The covariance between η_1 and η_2 is defined by $\sigma(\eta_1,\eta_2)=\mathrm{cov}(\eta_1,\eta_2)$[8]. For a practical system, a general approach for determining the order of the MVAR model is the AIC–Akaike Information Criterion [8], [9].

GC in Time Domain

Now consider the first equalities in (1) and (2), if $\sigma^2(\eta_1)$ is $\sigma^2(\varepsilon_1)$ less than in some suitable sense X_2 is said to have a causal influence on X_1. In this case, the first equality in (2) is more accurate than in (1) to estimate X_1.Otherwise, if $\sigma^2(\eta_1) = \sigma^2(\varepsilon_1)$,$X_2$ is said to have no causal influence on X_1. In this case,two equalities are same. Such kind of causal influence, called GC [10], [11], is defined by

$$F_{X_2 \to X_1} = \ln \frac{\sigma_{\varepsilon_1}^2}{\sigma_{\eta_1}^2} \tag{3}$$

Obviously, $F_{X2\to X1}=0$ when there is no causal influence from X_2 to X_1 and $F_{X2\to X1}>0$ when there is. Similarly, the causal influence from X1 to X2 is defined by

$$F_{X_1 \to X_2} = \ln \frac{\sigma_{\varepsilon_2}^2}{\sigma_{\eta_2}^2} \tag{4}$$

NC in Time Domain

Based on the first equality in (2), we can see contributions to $X_{1,t}$, which include $\sum_{j=1}^{m} a_{11,j} X_{1,t-j}$, $\sum_{j=1}^{m} a_{12,j} X_{2,t-j}$ and the noise term $\eta_{k,t}$ where the influence from $\sum_{j=1}^{m} a_{11,j} X_{1,t-j}$ is causality from X1's own past values. Each contribution plays an important role in determining $X_{1,t}$ .If $\sum_{j=1}^{m} a_{12,j} X_{2,t-j}$ occupies a larger portion among all those contributions, then X_2 has stronger causality on X_1, or vice versa. Thus, a good definition for causality from X_2 to X_1 in time domain should be able to describe what proportion X_2 occupies among all these contributions. So based on this general guideline NC from X_2 to X_1 is defined as [8]

$$n_{X_2 \to X_1} = \frac{\sum_{t=m}^{N} \left(\sum_{j=1}^{m} a_{12,j} X_{2,t-j} \right)^2}{\sum_{h=1}^{2} \sum_{t=m}^{N} \left(\sum_{j=1}^{m} a_{1h,j} X_{h,t-j} \right)^2 + \sum_{t=m}^{N} \eta_{1,t}^2} \tag{5}$$

Similarly, NC in time domain from X_1 to X_2 is defined by

$$n_{X_1 \to X_2} = \frac{\sum_{t=m}^{N} \left(\sum_{j=1}^{m} a_{21,j} X_{1,t-j} \right)^2}{\sum_{h=1}^{2} \sum_{t=m}^{N} \left(\sum_{j=1}^{m} a_{2h,j} X_{h,t-j} \right)^2 + \sum_{t=m}^{N} \eta_{2,t}^2} \tag{6}$$

An ERP eeg segment data usually involves three subsegments: pre-stimulus,stimulus and post-stimulus. Obviously, the mathematic models for pre-stimulus (or post-stimulus) and stimulus should be different. However, the general approach for ERP analysis in the literature always views the whole segment data as one model. Next, we will demonstrate that one should be caution when one views the whole data as one model to make a conclusion on causality.

Main Results

In this section we will provide two examples to discuss why we should be caution when we analyze causality of ERP data which may involve multiple segments of different models.

Example 1

We consider the following joint regression model:

$$\begin{cases} X_{1,t} = a_{11,1} X_{1,t} - 0.8 X_{2,t-1} + \eta_{1,t} \\ X_{2,t} = a_{21,1} X_{1,t} + 0.8 X_{2,t-1} + \eta_{2,t} \end{cases} \tag{7}$$

where t = 0,1,2,$\cdots$,10000, the noise terms $\eta_{i,t}$, i=1,2 are uncorrelated over time, have zero means and variances of $\sigma^2(n_1)=0.55$ and $\sigma^2(n_2)=1$.When t=0,1,2,$\cdots$,5000, $a_{11,1}=a_{21,1}=0$,when t=5001,5002,$\cdots$,1000, $a_{11,1}=a_{21,1}=0.5$.Thus,model (7) can be considered to be a regression model of time-variant coefficients for the whole time period, although it is a regression model of

time-invariant coefficients for each of two time periods:t = 0,1,2,⋯,5000 and t =5001,⋯ ,10000. That is, model (7) becomes

$$\begin{cases} X_{1,t} = -0.8X_{2,t-1} + \eta_{1,t} \\ X_{2,t} = 0.8X_{2,t-1} + \eta_{2,t} \end{cases} \quad (8)$$

when t = 0,1,2,⋯ ,5000, and model (7) becomes

$$\begin{cases} X_{1,t} = 0.5X_{1,t-1} - 0.8X_{2,t-1} + \eta_{1,t} \\ X_{2,t} = 0.5X_{1,t-1} + 0.8X_{2,t-1} + \eta_{2,t} \end{cases} \quad (9)$$

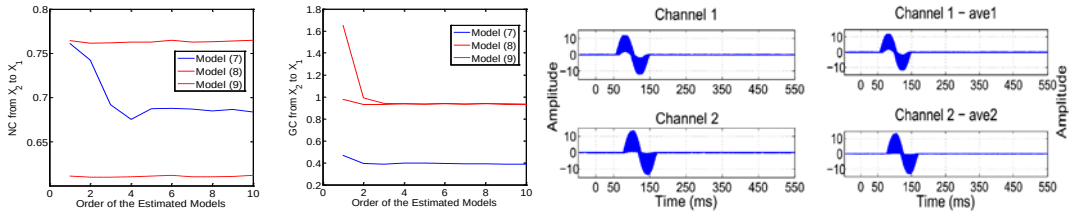
when t = 5001, ••• ,10000.

To calculate causality from $X_{2,t}$ to $X_{1,t}$ in models (7)~(9),for each specific model we generate a data set of 200 realizations. For each realization, we estimate models (autoregressive representations model (1) and joint representations model (2)) with the order of 8 by using the least-squares method and calculate GC and NC. Then we obtain the average value across all realizations and get GC and NC from X_2 to X_1 where the order 8 fits well (see Figs. 1(a) and 1(b) from which one can see GC and NC keep steady when the order of the estimated model is greater than 8). From Figure 1 one can clearly see that i) GC from X_2 to X_1 in model (8) is the same as that in model (9), that is, GC from X_2 to X_1 has nothing to do with parameters $a_{11,1}$ and $a_{21,1}$ in two models. This is clearly pointed out in (ii) of Property 1 [8].Due to the fact that model (7) is a combination of model (8) and model (9) in terms of two different time periods, if the true causality from X_2 to X_1 in model (8) is the same as that in model (9), then the true causality from X_2 to X_1 in model (7) should be equal to that in model (8) or model (9). However, GC (= 0.39) from X_2 to X_1 for model (7) is much smaller than that (= 0.94) in model (8) or model (9).This fact actually once again strongly demonstrates that GC value does not reveal true causality at all. ii) Note that $X_{1,t}$ in (2) includes three parts: $\sum_{j=1}^{m} a_{11,j} X_{1,t-j}$, $\sum_{j=1}^{m} a_{12,j} X_{2,t-j}$ and the noise term $\eta_{1,t}$.If $\sum_{j=1}^{m} a_{12,j} X_{2,t-j}$ occupies larger portion among all these three parts, then X_2 has stronger true causality on X_1,or vice versa. Since $X_{1,t}$ in model (8) includes two parts:$-0.8X_{2,t-1}$ and the noise term $\eta_{1,t}$,and $X_{1,t}$ in model (9) includes three parts:$0.5X_{1,t-1}$, $-0.8X_{2,t-1}$ and the noise term $\eta_{1,t}$, the portion which $-0.8X_{2,t-1}$ occupies in two parts:$-0.8X_{2,t-1}$ and the noise term $\eta_{1,t}$ obviously is larger than that portion which $-0.8X_{2,t-1}$ occupies among three parts:$0.5X_{1,t-1}$,$0.8X_{2,t-1}$ and the noise term $\eta_{1,t}$. Then one can see that the true causality from X_2 to X_1 in model (8) is surely larger than that in model (9). Since model (7) is a combination of model (8) and model (9) in terms of two different time periods, then the true causality from X_2 to X_1 in model (7) should be in between the true causality from X_2 to X_1 in model (8) and the true causality from X_2 to X_1 in model (9). From Figure 1(a) NC (= 0.76) in model (8) is indeed larger than that (= 0.61) in model (9). Moreover, NC (= 0.68) in model (7) is indeed in the interval [0.61,0.76]. Therefore, in this example, NC better reveals the underlying true causality than GC.

Example 2

We consider the simulation example of two univariate time series described in the first paragraph of simulation section [12]. According to the description the two channels have event-related potentials (ERPs) produced by one cycle of 12Hz sinusoidal waves which are combined with ongoing activities.The single-trial amplitudes (A_i) of the sinusoidal wave for Channel 1 (X_1) are chosen independently in the interval. The single-trial amplitude (Bi) for Channel 2 (X_2) is the amplitude (A_i) of Channel 1 plus standard Gaussian noise (α_i). The single-trial latency shifts (τ_{i1} ,τ_{i2}) for the ERP components of two channels are also considered and uniformly distributed between 0ms and 10ms. For analysis convenience, we further assume $\tau_{i1} = \tau_{i2} = \tau_i$ in this paper. 500 trials (realizations), each with 120 data points at a sampling rate of 200Hz, are generated. Each trial is 600ms long, 100ms of which occurred prior to stimulus onset (0ms). The ERP for Channel 1 starts about 50ms after the stimulus onset, while the ERP from Channel 2 is delayed by about 20ms. The ongoing activity for both channels is Gaussian white noise processes with zero mean and 0.05 standard deviation (η_1,η_2). These two noise processes are uncorrelated with each other. Figure 1(c) shows the 500 simulated realizations (trials) for both channels.

For two ERP components we can obtain NC value from X_1 to X_2.Then by averaging all these NC values for 500 trials we get one average NC value $n_{X1\to X2}$=0.995.That means ERP component X_2 is mainly determined by ERP component X_1. The true causality from ERP component X_1 to X_2 is close to 1 (i.e.,the strongest causality). This result is real. By estimating the auto-regression model of X_2 whose order is chosen as m = 3 based on AIC criteria we can obtain GC values from X_1 to X_2 for 500 trials and then get the average GC $F_{X1\to X2}$=3.9641 which is far away from $+\infty$(i.e., the strongest causality by GC definition). So, GC value of 3.9641cannot really reveal the strongest causality. For other time points which do not belong to ERP (X_2), $X_{1,t}$ and $X_{2,t}$ are tow noise terms from which one can easily get $n_{X1\to X2}$=$F_{X1\to X2}$=0.Note that the relation between X_1 and X_2 has two different models in the whole time points. However, the widely used approach in ERP analysis in the literature is to estimate one model in the whole time points.The two models definitely cannot be represented by one estimated model. Thus the true causality from X_1 to X_2 in the whole time points may be completely different from the causality calculated based on the estimated model.Since two different models are involved in the whole time points, we need to develop new method to discuss the true causality from X_1 to X_2 which is left for further study in our future work. For NC method, the causality from X_1 to X_2 in the time points set which only include ERP (X_2) is 0.9995 and the causality from X_1 to X_2 in the other time points set which do not belong to ERP (X_2) is 0, so NC value from X_1 to X_2 in the whole time points set should be in (0,0.9995). In fact after estimating the joint regression model for X_2 with the order of 4 based on AIC criteria we can get $n_{X1\to X2}$=0.9792 which indeed belongs to (0,0.9995). For GC method, the causality from X_1 to X_2 in the time points set which only include ERP (X_2) is 3.9641 and the causality from X_1 to X_2 in the other time points set which do not belong to ERP (X_2) is 0, so GC value from X_1 to X_2 in the whole time points set should be in (0,3.9641). In fact after estimating the joint regression model with the order of 4 and the auto-regression model for X_2 with the order of 4 based on AIC criteria we can get $F_{X1\to X2}$=1.2716 which indeed belongs to (0,3.9641). Therefore, no matter which method (GC or NC) is used, the obtained causality from X_1 to X_2 in the whole time points set is in the range between the causality (0) obtained in the time points set which does not belong to ERP (X_2) where ongoing activity are involved and the causality obtained in the time points set which belongs to ERP (X_2) where ERP components are involved. This may be why researchers in literature always conduct causality analysis on ERP data by using the whole time segment data which includes not only ongoing activity but also ERP component data. GC may not correctly reveal the true causality underlying in the whole time segment data.



(a) (b) (c) (d)

Figure 1. *(a) NC from X_2 to X_1 as a function of the order of the estimated models for models (7) ~(9). (b) GC from X_2 to X_1 as a function of the order of the estimated models for models (7) ~(9).(c) 500 realizations (trials) of simulated data for two channels. Both latency variability and amplitude variability are considered. (d) Residuals after subtracting their own average values (of each single trial) from each single trial for both channels.*

Conclusions

In this paper, on one hand, we pointed out that NC better reveals the underlying true causality than GC when one analyzed segment data involves two or more subsegments of different models. On the other hand, we explained why most researchers in the literature may conduct causality analysis for the whole analyzed ERP data as one estimated model although it may include two segments: ongoing activities segment and ERP components segment which have two different models. The above two examples may demonstrate that NC measure may be better to reveal true causality than GC method.

Acknowledgement

This work was supported in part by the national Natural Science Foundation of China under Grant No.61473110, Natural Science Foundation of Zhejiang Province, China, under Grant No.Z13F030002, International Science & Technology Cooperation Program of China under Grant No.2014DFG12570,Key Laboratory of Complex Systems Modeling and Simulation,Ministry of Ed ucation, China.

References

[1] P. Cheolsoo, L. David, R. Naveed, A. Alireza, and P. M. Danlio,"Classification of Motor Imagery BCI Using Multivariate Empirical Mode Decomposition", Neural Systems and Rehabilitatin Engineering,vol. 21, no. 1, pp. 10–22, Jan. 2013.

[2] S.J. Luck, An Introduction to the Event-Related Potential Technique,Second edition. Cambridge, Mass.: The MIT Press, 2014.

[3] N. Wiener, The Theory of Prediction. In E.F. Beckenbach, editor,Modern Mathematics for Engineers, Chap. 8. McGraw-Hill, New York,1956.

[4] C. W. J. Granger, "Investigating Causal Relations by Econometric Models and Cross-spectral Methods", Econometrica, vol. 37, no. 4, pp.424–438, Aug. 1969.

[5] J. Pearl, Causality: Models, Reasning and Inference, Cambridge University Press, 2009.

[6] P. Spirtes, C. Glymour, and R. Scheines. Causation, Prediction, and Search, Second Edition, The MIT Press, 2001.

[7] S. Hu, G. Dai, G. Worrell, Q. Dai, and H. Liang, "Causality analysis of neural connectivity: Critical examination of existing methods and advances of new methods", IEEE Trans Neural Networks, vol. 22, no.6, pp. 829–844, 2011.

[8] H. Akaike, "A new look at the statisticalmodel identification," IEEE Transactions on Automatic Control, vol. 19, no. 6, pp. 716–723, Dec.1974.

[9] A. Seghouane, "Model selection criteria for image restoration," IEEE Trans. on Neural Networks, vol. 20, no. 8, pp. 1357–1363, Aug. 2009.

[10] M. Ding, Y. Chen, and S. L. Bressler, "Granger causality: Basic theory and applications to neuroscience", in Handbook of Time Series Analysis,B. Schelter, M. Winterhalder, and J. Timmer Eds. Weinheim, Germany:Wiley-VCH, pp. 437–460, 2006.

[11] J. Geweke, "Measurement of linear dependence and feedback between multiple time series", J. Amer. Stat. Assoc., vol. 77, no. 378, pp. 304–313, Jun. 1982.

[12] X. Wang, Yong. Chen, and M. Ding, "Estimating Granger causality after stimulus onset: A cautionary note," Neuroimage, vol. 41, pp. 767–778, 2008.