Technical Note

Granger causality with signal-dependent noise

Qiang Luo [a,b], Tian Ge [b], Jianfeng Feng [b,c,*]

[a] Department of Mathematics and Department of Management, National University of Defense Technology, Hunan 410073, PR China
[b] Centre for Computational Systems Biology, School of Mathematical Sciences, Fudan University, Shanghai 200433, PR China
[c] Centre for Scientific Computing, University of Warwick, Coventry CV4 7AL, UK

ABSTRACT

It is generally believed that the noise variance in *in vivo* neuronal data exhibits time-varying volatility, particularly signal-dependent noise. Despite a widely used and powerful tool to detect causal influences in various data sources, Granger causality has not been well tailored for time-varying volatility models. In this technical note, a unified treatment of the causal influences in both mean and variance is naturally proposed on models with signal-dependent noise in both time and frequency domains. The approach is first systematically validated on toy models, and then applied to the physiological data collected from Parkinson patients, where a clear advantage over the classical Granger causality is demonstrated.

© 2011 Elsevier Inc. All rights reserved.

Introduction

The past few years have witnessed a significant growth in the application of Granger causality to various research areas, especially to physiological recordings and functional MRI data. Proposed by Granger (1969) and further extended to the frequency domain by Geweke (1982), Granger causality has become a powerful tool to detect causal influences and functional connectivity from the huge amount of temporal data easily accessible nowadays.

The classical Granger causal analysis is based on an autoregressive (AR) model (Friston, 2009a), which assumes that the current state of a process can be predicted by a linear function of its previous states plus a *white noise* or *innovation process*. A process is said to be the cause of another process if the inclusion of its past information can improve the prediction of the second process, *i.e.*, reduce the variance of the prediction error. In spite of its easy implementation, wide justification and successful applications, some of the working assumptions might be an over-simplification when performing time series analysis in some situations. One particular scenario is the violation of time-invariant noise, that is, the volatility changes over time. As a common phenomenon in financial market series, similar characteristics have been observed in many physiological recordings, such as the data from epilepsy patients, Parkinson patients and so on. Because of the close relationship of many series, it is obvious to conjecture that changes in the volatility of one series may have an impact on the mean activity or

volatility of another series and *vice versa*, which means that causal influences may happen on the second order statistics or there may be feedback between the first and second moments of the series. In these scenarios, a simple AR model obviously cannot capture these effects and a more detailed modeling of the volatility of the time series as well as the corresponding causality inference techniques are desired.

The development of specific models for changing volatility, or say, conditionally heteroskedastic data, has long been launched, primarily motivated by the research in the literature of economics. A natural development was boosted by Engle's invention of the autoregressive conditional heteroskedasticity (ARCH) model (Engle, 1982) which was then extended to generalized ARCH (GARCH) models (Bollerslev, 1986) as well as multivariate case (Engle and Kroner, 1995). Along with the extension of the models is the growing interest in testing for causality in variance. Two main approaches have been followed in the literature. On the one hand is the two-step methodology concentrating on the cross correlation function (CCF) of univariate residual estimates (Cheung and Ng, 1996). On the other hand is the (Quasi) Maximum-Likelihood (QML) method which utilizes a parametric specification of volatility dynamics (Hafner and Herwartz, 2004). However, few works provide a unified treatment of first and second order causality in the classical Granger sense. The significance of such an approach might be partly demonstrated in the 2003 Nobel Prize in Economics shared by Granger and Engle.

In this technical note, a novel unified approach is proposed to detect Granger causality in both mean and variance based on a single quantity in both time and frequency domains. The volatility model considered in this note shows time-dependent changes in the variance of the innovations. However, this variation is mediated entirely by dependency upon the states. In other words we only consider volatility models

that have state-dependent noise. Instead of allowing for innovations at both the level of the observations and on the variance parameters *per se*, the model considered has only one innovation and the variance (second order statistics) is coupled to the mean (first order statistics). This has proved a useful model in econometrics and has strong connections to neurobiology and motor control planning through things like Weber's Law where the noise scales with signals (Harris and Wolpert, 1998; Todorov, 2005; Selen et al., 2009; Harris and Wolpert, 1998). In these models, one time series can affect another by either influencing the mean activity (first order behavior) or by changing the variance of the innovations (second order behavior) in the other time series. The form of the model is also crucial to the subsequent causality inferences where a unified treatment of first and second order causality can be naturally introduced. Simple AR models will miss this second effect because it is not part of the model.

The proposed approach has been first systematically validated in toy models, compared with classical AR models, and then applied to the temporal recordings of a Parkinson patient. Our approach succeeded in detecting the feedback from the tremor to the brain (Smirnov et al., 2008) and spotting the maximum energy transported from the brain to the tremor in the frequency domain. Although here the application of our development is only demonstrated with electrophysiological and electromyography recordings, we envisage that, due to the signal dependent nature of noise in neuronal activity, our approach could be widely applied to infer functional connectivity in functional MRI (fMRI) or electroencephalogram (EEG) data. The *Matlab* program is available on request.

Methods

The model with signal-dependent noise

In this technical note, we consider the following multivariate time series model with time-varying volatility, particularly, signal-dependent noise:

$$Z_t = \sum_{i=1}^{p} A_i Z_{t-i} + r_t, \ r_t = H_t^{1/2}\varepsilon_t,$$
$$H_t = C'C + \sum_{i=1}^{q} B_i Z_{t-i} Z_{t-i}' B_i, \tag{1}$$

where p and q are the model orders, $A_i, i = 1, \cdots, p, B_i, i = 1, \cdots, q$ and C are the model coefficient matrices, $\{\varepsilon_t\}$ is a Gaussian white noise process. This is a modification of the Baba–Engle–Kraft–Kroner (BEKK) model (Engle and Kroner, 1995) in which the conditional covariance matrix H_t does not regress on the residual process r_t but depends on the activity of the signal. Hence, the model considered has only one innovation ε_t and the variance (second order statistics) is coupled to the mean (first order statistics). This form also ensures the positive definiteness of H_t. Clearly, when $B_i \equiv 0$ for all i, the conditional variance is constant and the model reduces to the well-known autoregressive regression (AR) model, *i.e.* the case of constant volatility. In light of these, we call our model (1) the AR–BEKK model.

Causality in the time domain

Suppose there are two multi-dimensional signals X and Y whose time observations are denoted by $X_t \in \mathbb{R}^{k_1}$ and $Y_t \in \mathbb{R}^{k_2}$, $t = 1, 2, \cdots, T$. As the Granger causality based on the AR model, consider the following two equations where X_t can either be predicted by the past information of its own or jointly represented by the past information of both X and Y:

$$X_t = \sum_{i=1}^{p} A_{xx,i} X_{t-i} + H_{xx,t}^{1/2} u_{xx,t},$$
$$H_{xx,t} = C'_{xx} C_{xx} + \sum_{j=1}^{q} B'_{xx,j} X_{t-j} X'_{t-j} B'_{xx,j}; \tag{2}$$

$$X_t = \sum_{i=1}^{p} A_{xy,i} X_{t-i} + \sum_{i=1}^{p} D_{xy,i} Y_{t-i} + H_{xy,t}^{1/2} u_{xy,t},$$
$$H_{xy,t} = C'_{xy} C_{xy} + \sum_{j=1}^{q} B'_{xy,j} \left[X'_{t-j}, Y'_{t-j} \right]' \left[X'_{t-j}, Y'_{t-j} \right] B_{xy,j}. \tag{3}$$

where p and q are model orders, $A_{xx,i}$, $A_{xy,i}$, $D_{xy,i}$, $i = 1, \cdots, p$, $B_{xx,j}$, $B_{xy,j}$, $j = 1, \cdots, q$ and C_{xx}, C_{xy} are all coefficient matrices. Recall the classical definition of Granger causality in the time domain. If the prediction of one time series can be improved, *i.e.*, the variance of the prediction error can be reduced, by including the past information of another time series, the causal influence exists. Note that in the signal-dependent volatility model considered here, the residuals are modeled in more detail compared with the AR model. Some part of it now explicitly depends on the activities of the signals, leaving only the terms C_{xx} and C_{xy} unexplained, which represent the levels of prediction errors for (2) and (3) respectively. Hence, the causal influence from Y to X can be defined as

$$F_{Y \to X} = \log \frac{\text{trace}[\sum_{xx}]}{\text{trace}\left[\sum_{xy}\right]}, \tag{4}$$

where $\sum_{xx} = C'_{xx} C_{xx}$ and $\sum_{xy} = C'_{xy} C_{xy}$. It is said that there is a causal influence from Y to X if $F_{Y \to X}$ is significantly different from zero. The causal influence from X to Y can be defined similarly. It can been seen that Y can help improve the prediction of X by impacting on either its mean activity through $D_{xy, i}$ or its variance through the coefficients in $B_{xy,j}$. These two cases correspond to the causality in mean and variance respectively. Hence, the causal influence in the time domain can be summarized into a single number as defined in (4). When the noise is not signal-dependent, *i.e.*, $B_{xx,j} \equiv 0$, $B_{xy,j} \equiv 0$, $j = 1, \cdots, q$, the model reduces to a multivariate autoregressive regression model and the definition of causality coincides with the conventional Granger causality in the time domain.

It is worth mentioning here that there is an ongoing debate on the selection of 'trace' or 'determinant' on multivariate models Barnett et al. (2009). Strictly speaking, when the time series do not show strong colinearity, the log ratio of determinants would be the optimum statistic with a lot of attracting properties especially in the definition of conventional Granger causality based on the simple AR model. For example, under very general conditions, the integration of Granger causality in the frequency domain equals the causality in the time domain Geweke (1982). However, with correlated signals, the use of 'determinant' might lead to numerical instability since the determinant of the covariance matrix might be close to zero. Hence, we recommend using 'trace' to ensure numerical stability as suggested in Ladroue et al. (2009), especially when dealing with fMRI data since signals from nearby brain regions are typically highly correlated.

The performances of these two definitions are similar when applied to the toy models. A detailed comparison between 'trace' and 'determinant' can be found in Appendix B. Crucially, we also numerically show in the Results part as well as in Appendix B, that in the AR–BEKK model, the integration of causality in the frequency domain is still approximately equal to the causality in the time domain under both definitions. (See below for the definition of causality in the frequency domain.) However, these conclusions on toy models should be generalized with caution and need more investigations in different data sets.

Causality in the frequency domain

One of the main advantages of the classical Granger causality over other approaches is its decomposition in the frequency domain. Here, the decomposition is extended to the proposed signal-dependent

volatility model. Suppose the joint representation of X and Y takes the form:

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \sum_{i=1}^{p} \begin{pmatrix} A_{xy,i} & D_{xy,i} \\ A_{yx,i} & D_{yx,i} \end{pmatrix} \begin{pmatrix} X_{t-i} \\ Y_{t-i} \end{pmatrix} + \begin{pmatrix} r_{xy,t} \\ r_{yx,t} \end{pmatrix}, \tag{5}$$

where

$$r_{xy,t} = H_{xy,t}^{1/2} u_{xy,t}, \ r_{yx,t} = H_{yx,t}^{1/2} v_{yx,t},$$

$$H_{xy,t} = C'_{xy} C_{xy} + \sum_{j=1}^{q} B'_{xy,j} \left[X'_{t-j}, Y'_{t-j} \right]' \left[X'_{t-j}, Y'_{t-j} \right] B_{xy,j},$$

$$H_{yx,t} = C'_{yx} C_{yx} + \sum_{j=1}^{q} B'_{yx,j} \left[X'_{t-j}, Y'_{t-j} \right]' \left[X'_{t-j}, Y'_{t-j} \right] B_{yx,j}.$$

Fourier transforming both sides of (5) leads to

$$\begin{pmatrix} \mathcal{A}_{11}(\omega) & \mathcal{A}_{12}(\omega) \\ \mathcal{A}_{21}(\omega) & \mathcal{A}_{22}(\omega) \end{pmatrix} \begin{pmatrix} X(\omega) \\ Y(\omega) \end{pmatrix} = \begin{pmatrix} r_{xy}(\omega) \\ r_{yx}(\omega) \end{pmatrix}, \tag{6}$$

where

$$\mathcal{A}_{11}(\omega) = I_{k_1} - \sum_{i=1}^{\infty} A_{xy,i} e^{-j\omega i}, \ \mathcal{A}_{12}(\omega) = -\sum_{i=1}^{\infty} D_{xy,i} e^{-j\omega i},$$

$$\mathcal{A}_{21}(\omega) = -\sum_{i=1}^{\infty} A_{yx,i} e^{-j\omega i}, \ \mathcal{A}_{22}(\omega) = I_{k_2} - \sum_{i=1}^{\infty} D_{yx,i} e^{-j\omega i},$$

$r_{xy}(\omega)$ and $r_{yx}(\omega)$ are the Fourier transform of the residuals, I_{k_1} and I_{k_2} are the identity matrices with k_1 and k_2 dimensions respectively. The covariance of the residuals $R = (r_{xy}, r_{yx})$ is

$$\phi_R = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}.$$

As suggested by Geweke, both sides of (6) are left multiplied by the normalizer

$$P = \begin{pmatrix} I_{k_1} & 0 \\ -\phi_{21}\phi_{11}^{-1} & I_{k_2} \end{pmatrix},$$

from which it follows that

$$\begin{pmatrix} \mathcal{A}_{11}(\omega) & \mathcal{A}_{12}(\omega) \\ \tilde{\mathcal{A}}_{21}(\omega) & \tilde{\mathcal{A}}_{22}(\omega) \end{pmatrix} \begin{pmatrix} X(\omega) \\ Y(\omega) \end{pmatrix} = \begin{pmatrix} r_{xy}(\omega) \\ \tilde{r}_{yx}(\omega) \end{pmatrix},$$

where $\tilde{\mathcal{A}}_{21} = \mathcal{A}_{21} - \phi_{21}\phi_{11}^{-1}\mathcal{A}_{11}$, $\tilde{\mathcal{A}}_{22} = \mathcal{A}_{22} - \phi_{21}\phi_{11}^{-1}\tilde{\mathcal{A}}_{12}$, $\tilde{r}_{yx} = r_{yx} - \phi_{21}\phi_{11}^{-1}r_{xy}$. The normalized model residual processes r_{xy} and $\tilde{r}_{yx}$ are now uncorrelated, with the covariance of $\tilde{r}_{yx}$ being $\phi_{22} - \phi_{21}\phi_{11}^{-1}\phi_{12}$. Reformulating this equation into the transfer function format results in

$$\begin{pmatrix} X(\omega) \\ Y(\omega) \end{pmatrix} = \begin{pmatrix} \mathcal{H}_{xx}(\omega) & \mathcal{H}_{xy}(\omega) \\ \mathcal{H}_{yx}(\omega) & \mathcal{H}_{yy}(\omega) \end{pmatrix} \begin{pmatrix} r_{xy}(\omega) \\ r_{yx}(\omega) \end{pmatrix},$$

where the transfer function $\mathcal{H}$ is the inverse of the coefficient matrix $\tilde{\mathcal{A}}$ in Eq. (6), i.e., $\mathcal{H}(\omega) = \tilde{\mathcal{A}}^{-1}(\omega)$ with the elements

$$H_{xx} = \left(\mathcal{A}_{11} - \mathcal{A}_{12}\tilde{\mathcal{A}}_{22}^{-1}\tilde{\mathcal{A}}_{21} \right)^{-1},$$

$$H_{xy} = -H_{xx}\mathcal{A}_{12}\tilde{\mathcal{A}}_{22}^{-1},$$

$$H_{yx} = -\tilde{\mathcal{A}}_{22}^{-1}\tilde{A}_{21}H_{xx},$$

$$H_{yy} = \tilde{\mathcal{A}}_{22}^{-1}\tilde{A}_{21}H_{xx}\tilde{\mathcal{A}}_{22}\tilde{\mathcal{A}}_{22}^{-1} + \tilde{\mathcal{A}}_{22}^{-1},$$

and its spectral matrix can be given by ensemble averaging as

$$\mathcal{S}(\omega) = \mathcal{H}(\omega) \, \tilde{\phi}_R(\omega) \mathcal{H}^*(\omega),$$

where * means the complex conjugate and matrix transpose. The auto-spectra of X is

$$S_{xx} = H_{xx}\phi_{11}H_{xx}^* + H_{xy}\left(\phi_{22} - \phi_{21}\phi_{11}^{-1}\phi_{12} \right) H_{xy}^*. \tag{7}$$

Up to date, the transformations are exactly the same as the classical formulation of causality in the frequency domain. Recall that in the case of simple AR model, the first term in (7) can be interpreted as the intrinsic power while the second term as the causal power. The classical measure can therefore be defined as the log ratio between the auto-spectra S_{xx} and the intrinsic power $H_{xx}\phi_{11}H_{xx}^*$. However, due to the more detailed modeling of the signal-dependent residual, part of ϕ_{11}, i.e., the variance of the residual r_{xy}, can now be explained by the signals. To see this, we have

$$\phi_{11} = \mathrm{E}\left[r_{xy}r'_{xy} \right] = \mathrm{E}\left[H_{xy,t}^{1/2} u_{xy,t} u'_{xy,t} \left(H_{xy,t}^{1/2} \right)' \right]$$

$$= C'_{xy}C_{xy} + \sum_{j=1}^{q} B'_{xy,j} \left[X'_{t-j}, Y'_{t-j} \right]' \left[X'_{t-j}, Y'_{t-j} \right] B_{xy,j}$$

$$= \sum_{xy} + \phi_{xx} + \phi_{xy},$$

where

$$\sum_{xy} = C'_{xy}C_{xy},$$

$$\phi_{xx} = \sum_{j=1}^{q} \left[B_{xy,j} \right]'_{11} \mathrm{E}\left[X_t X'_t \right] \left[B_{xy,j} \right]_{11},$$

$$\phi_{xy} = \sum_{j=1}^{q} \left[B_{xy,j} \right]'_{12} \mathrm{E}\left[Y_t X'_t \right] \left[B_{xy,j} \right]_{11}$$

$$+ \sum_{j=1}^{q} \left[B_{xy,j} \right]'_{11} \mathrm{E}[X_t Y'_t] \left[B_{xy,j} \right]_{12} + \sum_{j=1}^{q} \left[B_{xy,j} \right]'_{12} \mathrm{E}[Y_t Y'_t] \left[B_{xy,j} \right]_{12}.$$

Note that ϕ_{xx} is the term that can be explained by the signal X alone, ϕ_{xy} comprises of those terms having interactions with Y while $\sum_{xy}$ remains to be unexplained. Hence, in the signal-dependent volatility model, the causal influence from Y to X at the frequency ω can be defined as

$$f_{Y \to X}(\omega) = \log \frac{\mathrm{trace}[S_{xx}(\omega) - H_{xx}(\omega)\phi_{xx}H_{xx}^*(\omega)]}{\mathrm{trace}\left[H_{xx}(\omega) \sum_{xy} H_{xx}^*(\omega) \right]}. \tag{8}$$

This is a non-negative quantity and when there is no contribution from Y, i.e., all the coefficients before the terms with Y vanish, $f_{Y \to X}(\omega) = 0$. Therefore, it is said that there is a causal influence from Y to X at the frequency ω when $f_{Y \to X}(\omega) > 0$. The causal influence from X to Y can be established in a similar way. When the noise is not signal-dependent, this definition coincides with the conventional Granger causality in the frequency domain.

Parameter estimation and model order selection

According to (Ling and McAleer, 2003), the conditional log likelihood function of model (1) can be given by the following quasi Gaussian form

$$l_\theta(t) = -\frac{1}{2}\ln|H_t| - \frac{1}{2}r'_t H_t^{-1} r_t, \ t = p \vee q, \cdots, T \tag{9}$$

where the model parameters are $\theta = \{A_i, i = 1, \cdots, p; B_j, j = 1, \cdots, q; C\}$. Therefore, with the observation data, the maximum likelihood estimation (MLE) of the model parameters can be established by maximizing conditional log likelihood function on the observation data (See Appendix A for details):

$$\hat{\theta} = \arg\max_{\theta} \sum_{t = p \vee q}^{T} l_{\theta}(t). \tag{10}$$

With this log likelihood function, the Akaike information criteria (AIC) can be employed to determine the optimal model orders, p and q, for the model in k dimensions by minimizing

$$AIC(p,q) = -2 \sum_{t = p \vee q}^{T} l_{\theta}(t) + 2\left[k^2 p + k(k+1)/2 + k^2 q\right]. \tag{11}$$

The extra terms in the AIC depend upon the number of free parameters and constitute a complexity cost or penalty.

Results

Toy models

An illustrative example

Considering a trivariate model which takes the form as (5) where X has 2 dimensions and Y is a scalar. The coefficient matrices employed in this model are listed in Appendix C, from which it can be seen that there is a causal influence from Y to X through impacting on the variance of X with the coefficient matrix B_{xy} and no causal influence from X to Y exists. The causal structure of this toy model is depicted in Fig. 1a. The model is simulated for 3000 data points and the time traces are shown in Fig. 1b.

The signal-dependent volatility phenomenon is clear by direct observation which can also be captured by the auto-correlation functions of the squared signals shown in Fig. 1c. The power spectral density of the three components is presented in Fig. 1d.

Both the classical Granger causality and the proposed measure based on the AR–BEKK model are employed to evaluate the causal influences in both directions. A bootstrap method is carried out to construct the confidence intervals. The results of causal analysis along with their standard deviations obtained from bootstrapping are presented in Table 1. It can be seen that the AR–BEKK model can clearly capture the true causal influence from Y to X while the classical AR model fails to detect this. This is due to the assumption of the time-invariant residual variance in the AR model which cannot detect the impact through signal-dependent noise. To further illustrate this, the auto-correlation functions of the squared residual $r_{x_1}^2(t)$ for the variable X_1 before and after being normalized by its estimated variance $H_{x_1,y}(t)$, are shown in Fig. 2. The auto-correlation of the normalized squared residual almost vanishes when both the past information of X and Y are included for the estimation of the variance, while some auto-correlations with particular time lags remain appearing with only the information from X. This indicates that the residual of the signal does depend on the signal Y, from which a causal influence is introduced through the variance.

To see the relationship between the causality in time and frequency domains, 300 bootstrap samples based on AR–BEKK models in both domains, after removing the outliers by a 3σ criterion, are shown in Fig. 3b. It can be seen that the averaged causality in the time domain and the averaged values of the integrations in the frequency domain are close to each other. In fact, the Kolmogorov–Smirnov test shows no significant difference between these two results at the 5% significance level which indicates that the integration of the causality in the
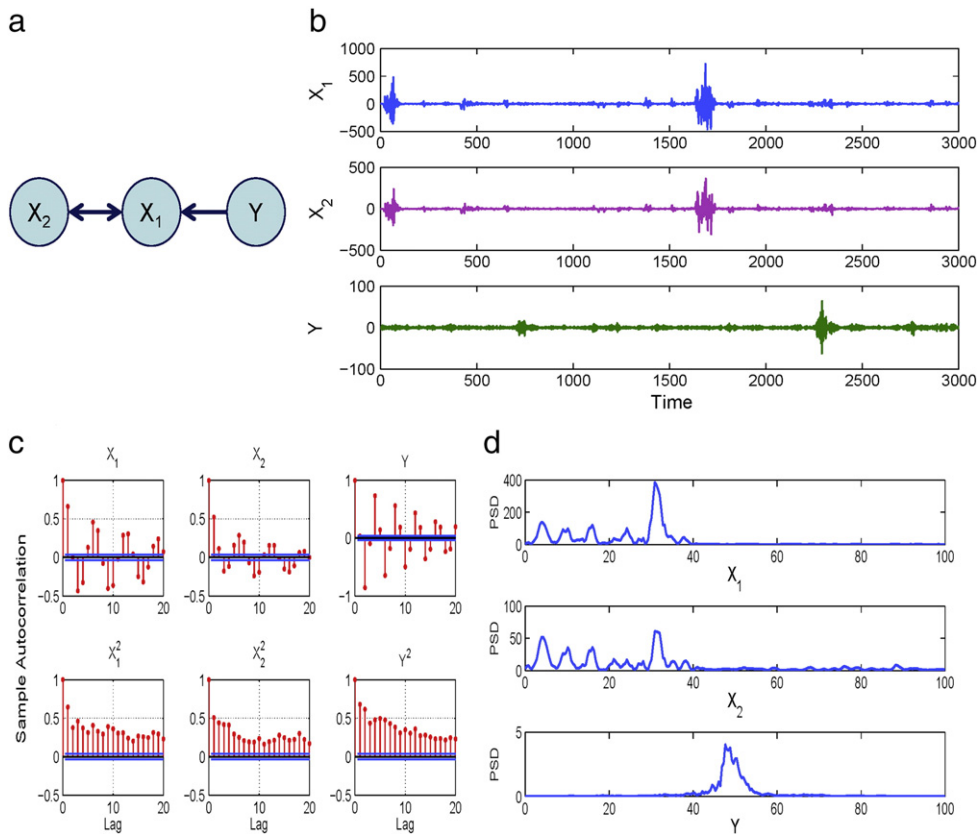

Fig. 1. Simulation of the illustrative model. a. The causal structure of the illustrative model; b. time traces of the simulated data with 3000 data points; c. the auto-correlation functions of the simulated signals; d. the power spectral density of each time series.

Table 1
Results of the causal analysis.

Causality	AR model	AR–BEKK model
$Y \rightarrow X$	0.0005 (0.0012)	1.1307 (0.1685)
$X \rightarrow Y$	0.0004 (0.0054)	0.0029 (0.0259)

frequency domain is approximately equal to the causality in the time domain. Note that it's unclear whether this property stands in theory and the results on toy models should be generated with caution.

A systematic investigation

It has been shown that, with the signal-dependent noise, one time series can have causal influence on the other by either impacting on the mean or the variance. To have a systematic investigation of the advantages of the proposed AR–BEKK model over the classical AR model under different situations, a series of trivariate toy models as the illustrative model above, taking the general form (3) but with different causal structures, are constructed and both approaches are employed into causal inferences. The coefficient matrices used in these models are listed in Appendix C and the results are summarized in Fig. 4. Five different causal structures are included for this comparison with the diagrams shown in the first column. The first four models, from a to d, have a detailed modeling of the signal-dependent noise. Hence, the time series Y may have a causal influence on the mean of X, the variance of X, or both. The last one is simply a classical AR model, *i.e.*, the data generated have time-invariant volatility. All these models are simulated to generate time series with 3000 points. Both the proposed AR–BEKK model and the classical AR model are then employed to detect the causal influence from Y to X. The orders of the models are selected by AIC introduced in the Methods part. A red tick indicates a significant causal influence is detected from Y to X after confidence intervals are constructed using a bootstrap approach stated in the previous section while a red circle means no significant influence exists. It can be seen that, as long as the data exhibit signal-dependent volatility, classical Granger causal analysis based on AR model cannot detect the causal influence while the AR–BEKK model succeeds in inferring all the true relationships. On the other hand, if the data are generated from an AR model with constant variance, both approaches turn out to give correct results. We also generate three purely random processes to test both approaches. In this case, denoted as f, both models successfully detect that no causal structure exists.

Although it seems the AR–BEKK model works well in all the above cases, one would prefer the simple AR model in the case of time-invariant volatility since fewer parameters will be estimated which
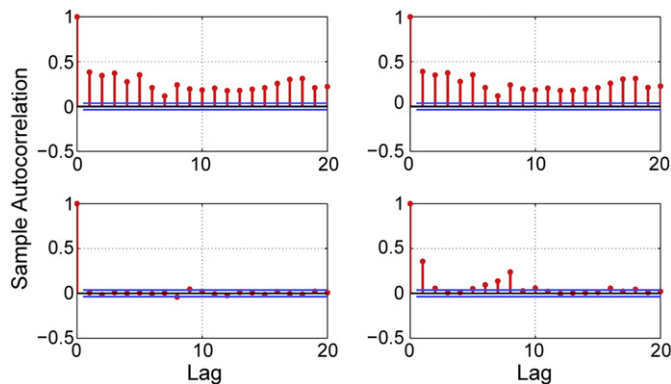

Fig. 3. Causality in time and frequency domains. a. Comparison of the causalities established by AR and AR–BEKK models; b. comparison of the causalities in both time domain and frequency domains using AR–BEKK model.

releases the burden of computational efforts. Hence, a criterion for model selection will be very helpful. We find that, in all the simulations under different causal structures, the proper model coincides with the one which has a smaller AIC value as shown in Table 2. This provides an instruction on the model selection.

Fig. 2. Autocorrelation functions of the normalized squared residual processes for the variable X_1 given by different models. Upper figures: auto-correlation functions of the squared residual $r_{x_1}^2(t)$ for the variable X_1; lower figures: auto-correlation functions of the squared residual $r_{x_1}^2(t)$ for the variable X_1 after being normalized by its variance $H_{x_1y}(t)$ estimated by the past information of X alone (left) and by both X and Y (right).
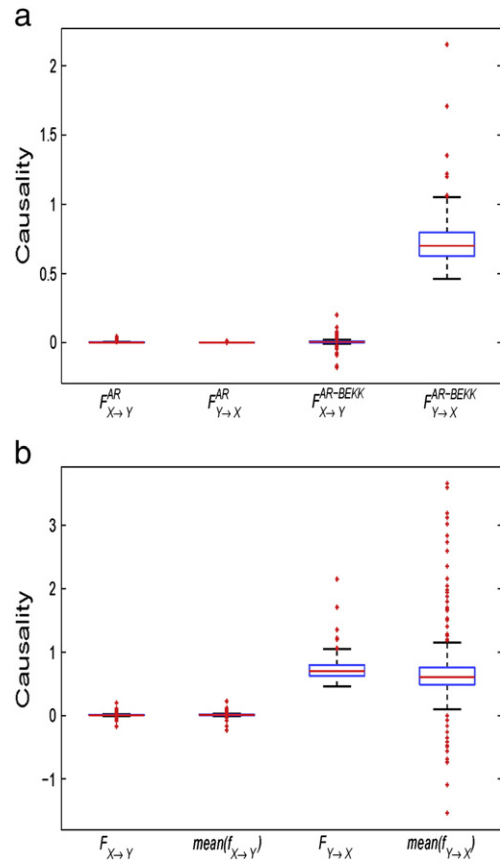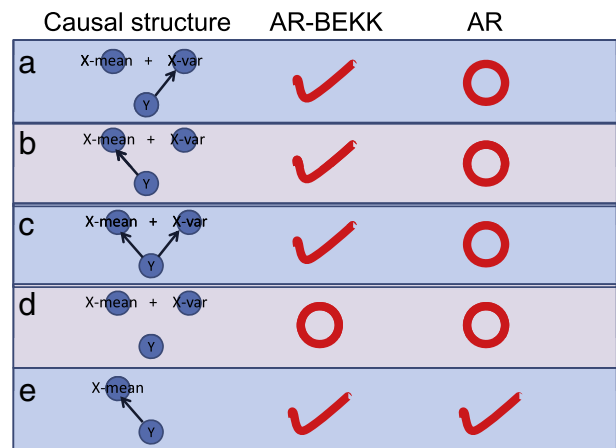

Fig. 4. Comparison of the AR–BEKK model and the AR model on different causal structures. Five different causal structures, from a to e, are included for comparison. The causal diagrams are shown in the first column. Each arrow indicates a causal influence of the time series Y on the mean or variance of X. A red tick indicates a significant causal influence is detected from Y to X using the corresponding model while a red circle means no significant influence exists.

Table 2
AIC values for different models under different causal structures (unit: 10^4).

Models	a	b	c	d	e	f
AR	6.2462	5.4302	6.2475	4.5332	2.5449	2.5453
AR–BEKK	3.7339	3.5192	3.6855	3.4038	2.5456	2.5463

Table 3
Results of the causal analysis for Parkinson patients.

Method	LFPs→EMGs	EMGs→LPFs
AR	0.0257 (0.0109)	0.0136 (0.0145)
AR–BEKK	0.3145 (0.1037)	0.0868 (0.0403)

Application to Parkinson data

The recordings of the local field potentials (LFPs) from subthalamic nucleus (STN) and surface eletromyograms (EMGs) simultaneously from the forearm muscles of a Parkinson patient have been analyzed in Wu et al.(2008). LFPs were recorded by the macro-electrodes with a common electrode placed on the surface of mastoid, and EMGs were recorded using surface electrodes placed over the tremulous forearm extensor and flexors (Nandi, 2002). The time series recordings are available for 63 s at the sampling frequency of 500 Hz. We only use the recordings from 11 s to 19 s, since the tremor started at around 12 s on the arm. The data have been passed through a low-pass filter and have been down sampled to 100 Hz before causal analysis as shown in Fig. 5.

The AR model with 6 time lags and the AR–BEKK model with 4 lags in the mean, and 2 lags in the variance are employed to detect the causal influences. The AIC score for the AR–BEKK model is 8.0893×10^3 which is smaller than that for the AR model, namely 8.2120×10^3, indicating a preference to the AR–BEKK model. The results of causalities are reported in Table 3 and Fig. 6. When the tremor occurs, both AR and AR–BEKK models identify the causal influence from brain to tremor. However, the AR–BEKK model also detects a feedback from tremor to brain which has been broadly reported in the literatures (Smirnov et al., 2008; Tass et al., 2010). A significant influence from brain to tremor at around 4 Hz is spotted which is also consistent with the previous findings (Wu et al., 2008).

Discussion

A novel and natural approach has been proposed to unify the work by Granger and Engle who shared the Nobel prize in Economics in 2003. Causality in both mean and variance can be tested simultaneously based on a model with time-varying volatility, or more specifically, signal-dependent noise. The approach is first systematically tested in toy models and then applied to recordings from Parkinson patients. We would expect that the example included in this technical note is only an example among many other potential applications to high throughput data currently available in various areas since signal-dependent volatility is a common phenomenon in real data from different sources.

Granger causality with signal-dependent noise and its frequency domain decomposition

In the literature, the directional influence can be detected by dynamic Bayesian network (Zou and Feng, 2009), transfer information (Barnett et al., 2009), directed transfer function (Kaminski et al., 2001), Granger causality (Ding et al., 2006), and many other methods. Actually, Granger causality has been widely used to explore intrinsic relationships in temporal data, such as neurophysiological recordings (Ge et al., 2009), financial data (Hong et al., 2009), gene expression data (Lozano et al., 2009), *etc*. Recently, many new insights and developments of Granger
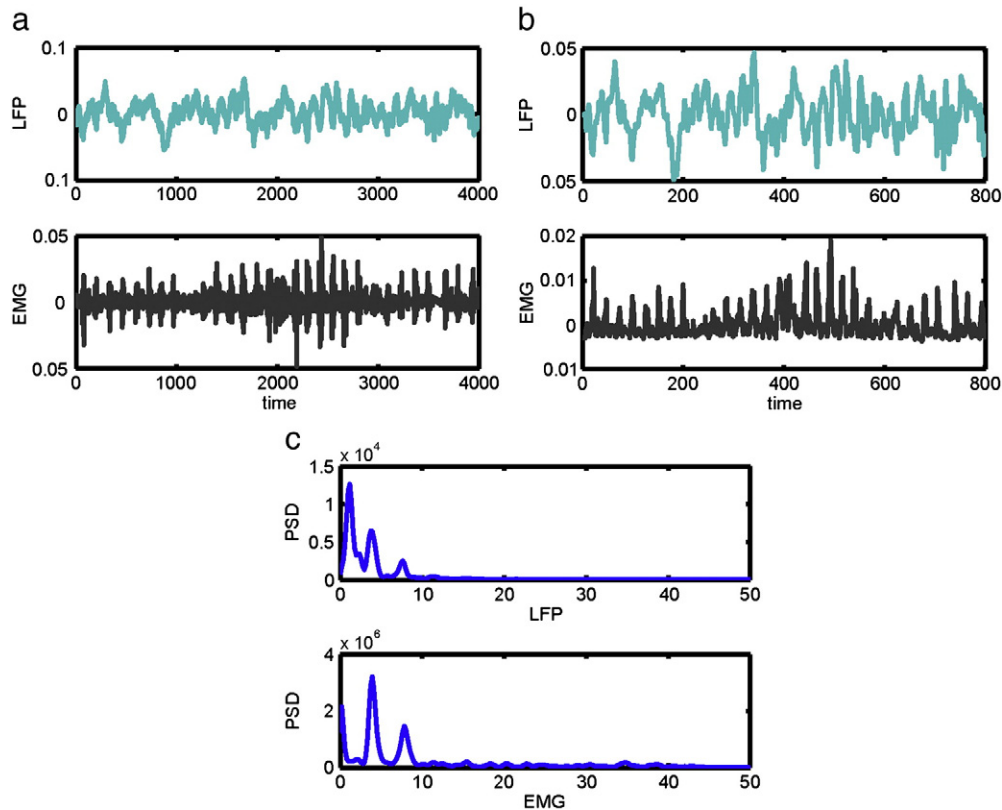

Fig. 5. LFP and EMG recordings from Parkinson patients. a. Original recordings from 11 s to 19 s for LFP and EMG; b. time series after low-pass filtering; c. power spectral density of the recordings.
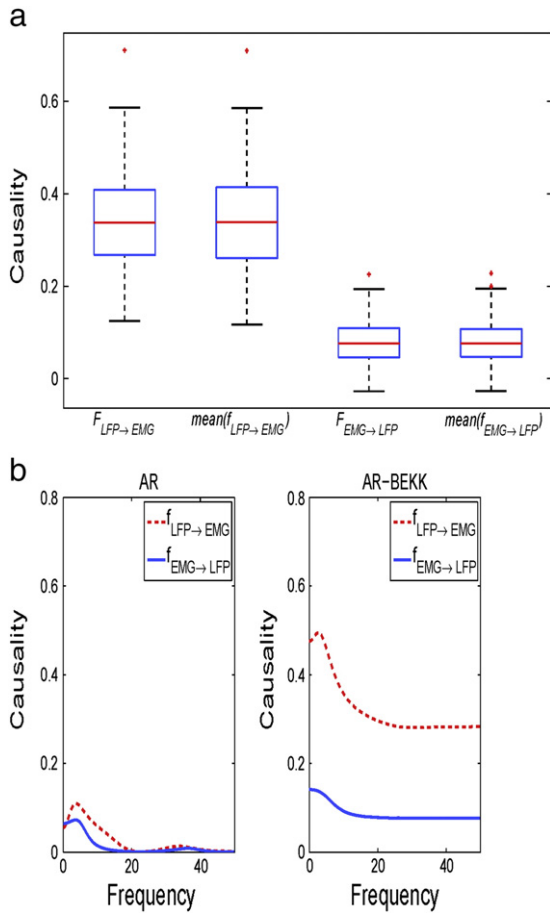
Fig. 6. Result for the Parkinson data. a. Causality obtained from AR–BEKK model in both time and frequency domains; b. causality decomposition on the frequency domain.

causality have also been reported (Barnett et al., 2009; Dhamala et al., 2008). In the case of time-invariant volatility, the inferences of the Granger causality have been formalized in terms of an AR model by Granger in 1969 (Granger, 1969) and extended to the frequency domain by Geweke (1982). Uncovering functional connectivity at different frequencies has become one of the biggest advantages of Granger causality over many other approaches (Wu et al., 2008; Guo et al., 2008a,b), since the frequency-specific influences have particular interest in neuroscience. This technical note introduces Granger causality to model with a more detailed modeling of the signal-dependent noise, which is a common scenario in various recordings in neuroscience. This may endow the Granger causality with power to infer influences in variance. Frequency decomposition is also proposed which may shed some light on frequency-specific interactions.

Unified mean and variance causality testing

In the case of time-varying volatility, most of the existing approaches attempt to test causality in variance through a two-step procedure. Building upon tests on causality in mean, an additional AR model is established to infer cross sectional dependencies of the squared residuals after the first-step regression. It is difficult to prove, in these approaches, whether the accuracy of the model fitting for mean will not be affected by the causal influence in variance. The situation becomes even tougher when the mean and variance of the signal are coupled, *i.e.*, a signal-dependent noise exists. Furthermore, since the classical Granger causality *per se* is a second order statistics, it is more natural to test causality, in the framework of Granger, with a direct measure. The approach proposed in this technical note presents

such a unified treatment to summarize causality in both mean and variance into a single number.

Feedback from tremor to brain for Parkinson's disease

Identifying directional couplings is a key step for us to explore the information flow between two systems. As reported in Smirnov et al. (2008), the generation of Parkinsonian tremor is probably due to a bidirectional coupling between the tremor and the brain. Therefore, the detection of causal influence from the tremor to the brain indicates that the proposed approach is a promising tool for causal inferences. Furthermore, the decomposition of the causality in the frequency spots the maximum energy transportation from the brain to the tremor occurs at 2–4 Hz, which is the main frequency band observed for the tremor as reported in the literature (Wu et al., 2008).

Acknowledgments

Appendix A

Based on the observations, the parameters are estimated by maximizing the objective function, *i.e.*, the log likelihood function (9). The standard Quasi-Newton method with a cubic line search procedure has been employed for this optimization. This approach can be carried out by the function '*fminunc*' in the optimization toolbox of *Matlab*. The initial values of the parameters for this iterative procedure are specified as follows:

First, an AR model is assumed and the initial coefficient matrices A_i, $i = 1, \cdots, p$ can be given by the least square estimation of this AR model:

$$Z_t = \sum_{i=1}^{p} A_i Z_{t-i} + r_t.$$

Second, for the parameters C and $B_i, i = 1, \cdots, q$ in the variance, a linear system is formulated by assuming C and B_i are diagonal matrices. For each dimension l, the following model is obtained:

$$(r_t)_l^2 = C_l^2 + \sum_{i=1}^{q} (B_i)_{ll}^2 (Z_{t-i})_l^2.$$

The least square estimations of these parameters are used as the initial values of the corresponding diagonal elements of C and B_i. All the off-diagonal elements are set to zero initially.

Appendix B

If 'determinant' is employed in the definitions of causality in both the time and frequency domains instead of 'trace', the proposed method has been tested on the same simulation data generated by the simple illustrative model in the Results part. The established causalities in the time domain are listed in Table 4.

The bootstrap method described in the main text has also been carried out for this version of Granger causality. Fig. 7a shows that, the

Table 4
Comparison of the results using 'trace' or 'determinant' in the definition.

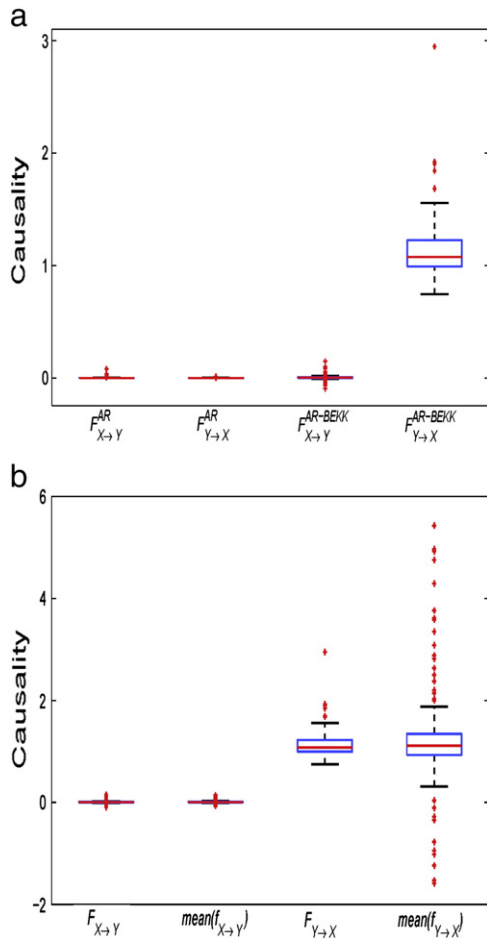| Causality	Determinant		Trace	
	AR	AR–BEKK	AR	AR–BEKK
$Y \to X$	0.0008 (0.0021)	1.6512 (0.2227)	0.0005 (0.0012)	1.1307 (0.1685)
$X \to Y$	0.0004 (0.0057)	0.0020 (0.0162)	0.0004 (0.0054)	0.0029 (0.0259)

Fig. 7. Causality in time and frequency domains using 'determinant' in the definition. a. Comparison of the causalities established by AR and AR–BEKK models; b. Comparison of the causalities in both time domain and frequency domains using AR–BEKK model.

AR–BEKK model can detect the true causal influence in variance from Y to X while the AR model fails. Fig. 7b shows that, using 'determinant' in the definitions, the integration of Granger causality in the frequency domain is approximately equal to the causality in the time domain. These results indicate that these two versions of Granger causality give similar results, as far as the illustrative example is concerned.

Appendix C

The trivariate example of the model (5) in Results used the following coefficient matrices:

$$A_1 \equiv \begin{bmatrix} A_{xy,1} & D_{xy,1} \\ A_{yx,1} & D_{yx,1} \end{bmatrix} = \begin{bmatrix} 0.95\sqrt{2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0.1\sqrt{2} \end{bmatrix},$$

$$A_2 \equiv \begin{bmatrix} A_{xy,2} & D_{xy,2} \\ A_{yx,2} & D_{yx,2} \end{bmatrix} = \begin{bmatrix} -0.9025 & 0.5 & 0 \\ 0.5 & 0 & 0 \\ 0 & 0 & -0.6\sqrt{2} \end{bmatrix};$$

$$C'_{xy}C_{xy} = \begin{bmatrix} 1 & 0.1 \\ 0.1 & 1 \end{bmatrix}; C'_{yx}C_{yx} = 1; \ B_{xy} = \begin{bmatrix} \sqrt{0.2} & 0.1 \\ 0.1 & \sqrt{0.2} \\ \sqrt{0.2} & 0 \end{bmatrix}, B_{yx} = \begin{bmatrix} 0 \\ 0 \\ \sqrt{0.2} \end{bmatrix}.$$

The orders of the model are set to their true values $p = 2$ and $q = 1$ before fitting the data.

In the systematic investigation of the AR model and AR–BEKK model, the coefficient matrices used are listed as follows.

In case **a**, all the coefficient matrices are the same as in the illustrative model. In case **b**, all the coefficient matrices are the same as in the illustrative model except A_1 is replaced by $\tilde{A}_1$ and B_{xy} is replaced by $\tilde{B}_{xy}$:

$$\tilde{A}_1 = \begin{bmatrix} 0.95\sqrt{2} & 0 & 0.4 \\ 0 & 0 & 0.1 \\ 0 & 0 & 0.1\sqrt{2} \end{bmatrix}, \ \tilde{B}_{xy} = \begin{bmatrix} \sqrt{0.2} & 0.1 \\ 0.1 & \sqrt{0.2} \\ 0 & 0 \end{bmatrix}.$$

In case **c**, all the coefficient matrices are the same as in the illustrative model except A_1 is replaced by $\tilde{A}_1$. In case **d**, all the coefficient matrices are the same as in the illustrative model except B_{xy} is replaced by $\tilde{B}_{xy}$. In case e, only an AR model is used with coefficient matrices $\tilde{A}_1$ and A_2.

References

Barnett, L., Barrett, A.B., Seth, A.K., 2009. Granger causality and transfer entropy are equivalent for Gaussian variables. Phys. Rev. Lett. 103, 238701.
Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Econ. 31 (3), 307–327.
Cheung, Y.-W., Ng, L.K., 1996. A causality-in-variance test and its application to financial market prices. J. Econ. 72 (1–2), 33–48.
Dhamala, M., Rangarajan, G., Ding, M., 2008. Estimating granger causality from fourier and wavelet transforms of time series data. Phys. Rev. Lett. 100 (1), 018701.
Ding, M., Chen, Y., Bressler, S.L., 2006. Granger causality: basic theory and application to neuroscience. In: Björn Schelter, M.W. (Ed.), Handbook of Time Series Analysis: Recent Theoretical Developments and Applications. WILEY-VCH Verlag GmbH & Co. KGaA, pp. 437–460.
Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50 (4), 987–1007.
Engle, R.F., Kroner, K.F., 1995. Multivariate simultaneous generalized arch. Econ. Theory 11 (01), 122–150.
Friston, K., 2009a. Causal modelling and brain connectivity in functional magnetic resonance imaging. PLoS Biol. 7 (2), e1000033+ (February).
Ge, T., Kendrick, K.M., Feng, J., 2009. A novel extended granger causal model approach demonstrates brain hemispheric differences during face recognition learning. PLoS Comput. Biol. 5 (11), e1000570.
Geweke, J., 1982. Measurement of linear dependence and feedback between multiple time series. J. Am. Stat. Assoc. 77 (378), 304–313.
Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37 (3), 424–438.
Guo, S., Wu, J., Ding, M., Feng, J., 2008a. Uncovering interactions in the frequency domain. PLoS Comput. Biol. 4 (5), e1000087+.
Guo, S., Seth, A.K., Kendrick, K.M., Zhou, C., Feng, J., 2008b. Partial granger causality—eliminating exogenous inputs and latent variables. J. Neurosci. Methods 172, 79–93.
Hafner, C.M., Herwartz, H., 2004. Testing for causality in variance using multivariate garch models. Economics Working Papers 2004, 03. Christian-Albrechts-University of Kiel, Department of Economics.
Harris, C.M., Wolpert, D.M., 1998. Signal-dependent noise determines motor planning. Nature 394 (6695), 780–784.
Hong, Y., Liu, Y., Wang, S., 2009. Granger causality in risk and detection of extreme risk spillover between financial markets. J. Econ. 150 (2), 271–287 recent Development in Financial Econometrics.
Kaminski, M., Ding, M., Tryccolo, W., Bressler, S., 2001. Evaluating causal relations in neural system: Granger causality, directed transfer function and statistical assessment of significance. Biol. Cybern. 145–157.
Ladroue, C., Guo, S., Kendrick, K., Feng, J., 2009. Beyond element-wise interactions: identifying complex interactions in biological processes. PLoS One 4 (9), e6899.
Ling, S., McAleer, M., 2003. Asymptotic theory for a vector arma-garch model. Econ. Theory 19 (2), 280–310.
Lozano, A.C., Abe, N., Liu, Y., Rosset, S., 2009. Grouped graphical Granger modeling for gene expression regulatory networks discovery. Bioinformatics 25 (12), i110–i118 (June).
Nandi, D., 2002. Deep brain stimulation of the pedunculopontine region in the normal non-human primate. J. Clin. Neurosci. 9 (2), 170–174 (March).
Selen, L.P.J., Franklin, D.W., Wolpert, D.M., 2009. Impedance control reduces instability that arises from motor noise. J. Neurosci. 29 (40), 12606–12616.
Smirnov, D.A., Barnikol, U.B., Barnikol, T.T., Bezruchko, B.P., Hauptmann, C., Bührle, C., Maarouf, M., Sturm, V., Freund, H.J., Tass, P.A., 2008. The generation of parkinsonian tremor as revealed by directional coupling analysis. EPL (Europhys. Lett.) 83 (2), 20003+.
Tass, P., Smirnov, D., Karavaev, A., Barnikol, U., Barnikol, T., Adamchic, I., Hauptmann, C., Pawelcyzk, N., Maarouf, M., Sturm, V., Freund, H.J., Bezruchko, B., 2010. The causal relationship between subcortical local field potential oscillations and parkinsonian resting tremor. J. Neural Eng. 016009.
Todorov, E., 2005. Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system. Neural Comput. 17, 1084–1108.
Wu, J., Liu, X., Feng, J., 2008. Detecting causality between different frequencies. J. Neurosci. Methods 167 (2), 367–375.
Zou, C., Feng, J., 2009. Granger causality vs. dynamic Bayesian network inference: a comparative study. BMC Bioinforma. 10, 122.