Prevention Science (2019) 20:442–451
https://doi.org/10.1007/s11121-018-0919-0

Granger Causality Testing with Intensive Longitudinal Data

Peter C. M. Molenaar[1]

Published online: 1 June 2018

Abstract
The availability of intensive longitudinal data obtained by means of ambulatory assessment opens up new prospects for prevention research in that it allows the derivation of subject-specific dynamic networks of interacting variables by means of vector autoregressive (VAR) modeling. The dynamic networks thus obtained can be subjected to Granger causality testing in order to identify causal relations among the observed time-dependent variables. VARs have two equivalent representations: standard and structural. Results obtained with Granger causality testing depend upon which representation is chosen, yet no criteria exist on which this important choice can be based. A new equivalent representation is introduced called hybrid VARs with which the best representation can be chosen in a data-driven way. Partial directed coherence, a frequency-domain statistic for Granger causality testing, is shown to perform optimally when based on hybrid VARs. An application to real data is provided.

Keywords Granger causality · Standard VAR · Structural VAR · Hybrid VAR · Partial directed coherence

The increased use of ambulatory assessment (Trull and Ebner-Priemer 2013) gives rise to new possibilities to carry out prevention research. Ambulatory assessment uses a wide range of methods to study people in their natural environment, including self-report, observational, and biological/physiological/behavioral. It is (a) idiographic in focus and allows for the examination of multiple individual processes (e.g., emotional, behavioral, psychophysiological) and (b) characterized by the collection of data in real-world environments, thus increasing the ecological validity of findings. It has been used to prevent relapse of major psychiatric diseases (Spaniel et al. 2008). Ambulatory assessment gives rise to intensive longitudinal data sets, which are the focus of this paper. Ecological momentary assessments are a prime example of ambulatory assessment (Gibbons 2016).

The availability of intensive longitudinal data allows for the application of time series analysis techniques with which the structure of intra-individual variation (IAV) can be analyzed. It is known that, for the same set of variables, the structure of IAV will in general differ from the structure of inter-individual variation (IEV). The latter applies at the level of populations and the former at the level of individuals (Molenaar 2004). Only if a process obeys strict criteria does there exist a relation between results obtained in analyses of IEV and IAV. Hence, if these strict criteria are violated, in order to obtain results which apply at the level of individual subjects, it is required to apply analysis of IAV (time series analysis of intensive longitudinal data; cf. Molenaar and Campbell 2009).

In what follows, Granger causality testing (GCT) of time series data will be discussed. It is based on the chapter by Molenaar and Lo (2016). GCT is a popular approach to determine causal relations among dynamic processes, originating in econometrics (Granger 1969), but now also having important applications in biophysics and neuroimaging (e.g., Goebel et al. 2003). GCT can be initially specified in the following way. Consider a bivariate process $\mathbf{z}(t)$ which is composed of univariate components x(t) and y(t). A time series model (which will be explained below) is fitted to the observations of $\mathbf{z}(t)$, including so-called lagged cross-relations between x(t) and earlier $y(t-u)$, $u>0$, as well as lagged cross-relations between y(t) and earlier $x(t-u)$, $u>0$. If the lagged cross-relations between x(t) and earlier $y(t-u)$ are substantial while the lagged cross-relations between y(t) and earlier $x(t-u)$ are not substantial, then y(t) is identified as a Granger cause of x(t).

If x(t) is a Granger cause of y(t) and direct preventive action on y(t) is difficult whereas such action on x(t) is directly applicable, then this preventive action can be focused on x(t) instead. Given that x(t) is a Granger cause of y(t), it is expected that changing x(t) will affect y(t) also.

✉ Peter C. M. Molenaar
pxm21@psu.edu

[1] The Pennsylvania State University, 415 BBH Building, University Park, PA 16802, USA

The simple operationalization of GCT described above will be generalized in some important ways. In particular, we will consider alternative ways to carry out GCT, using frequency domain representations (based on Fourier analysis) of time series models of p-variate processes where $p > 0$. However, the main focus will be on an important theoretical equivocality in current GCT for which no solution has been proposed in the literature yet. The details of this theoretical equivocality will be discussed at some length, and a new data-driven solution to resolve it will be introduced.

The approach to be presented yields unique candidate structural equation models for temporal data based on Granger causality testing of directed connections.[1] It is distinct from the approach based on pattern recognition to identify undirected links between pairs of variables which efficiently summarize the multivariate associations between nodes with respect to a small number of links. Within the field of prevention science, such models have been explored by Borsboom (2017) and others as parsimonious and more appropriate models applied to multivariate characterization of symptoms of psychopathology. It also is distinct from models, developed in large part from the Bayesian belief network perspective initially proposed by Pearl (2000) and others because this approach does not yield a unique candidate structural equation model.

Preliminaries

The following notational conventions will be used. We denote manifest variables by Roman letters and latent variables by Greek letters. We denote matrices by boldface uppercase letters and vectors by boldface lowercase letters. Vectors are always column vectors; if $\mathbf{y}$ is a (column) vector, then $\mathbf{y}'$, where the apostrophe denotes transposition, is the corresponding row vector.

Let $\mathbf{y}(t) = [y_1(t), y_2(t),\ldots, y_p(t)]'$ be a p-variate time series, $p \geq 1$. The mean of $\mathbf{y}(t)$ at each time point t is: $E[\mathbf{y}(t)] = \boldsymbol{\mu}(t)$, where E[.] denotes the expectation operator. Conceived of as function of t, $\boldsymbol{\mu}(t)$ stands for the p-variate mean function (trend) of $\mathbf{y}(t)$. If $\boldsymbol{\mu}(t) = \boldsymbol{\mu}$ (implying that the mean function is constant) then $\mathbf{y}(t)$ is referred to as having a stationary mean function. We define the sequential covariance of $\mathbf{y}(t)$ between each distinct pair of time points t_1 and t_2 as $\boldsymbol{\Sigma}(t_1, t_2) = \text{cov}[\mathbf{y}(t_1), \mathbf{y}(t_2)']$, where cov[.] denotes the covariance operator. Considered as function of two-dimensional time $\boldsymbol{\Sigma}(t_1, t_2)$ refers to the (p,p)-variate covariance function of $\mathbf{y}(t)$ at all times t_1 and t_2. If $\boldsymbol{\Sigma}(t_1, t_2)$ only is a function of the relative time difference, called the lag, $t_1 - t_2 = u$, that is, $\boldsymbol{\Sigma}(t_1, t_2) = \boldsymbol{\Sigma}(t_1 - t_2) = \boldsymbol{\Sigma}(u)$, for all integer values of u, then $\mathbf{y}(t)$ is said to have stationary covariance function. If both the mean and covariance functions of $\mathbf{y}(t)$ are stationary, then $\mathbf{y}(t)$ is referred to as a weakly stationary time series. Consequently the statistical specification of a weakly stationary p-variate series boils down to the specification of its p-variate mean level $\boldsymbol{\mu}$ and the sequence of (p,p)-dimensional covariance matrices $\boldsymbol{\Sigma}(u)$ describing the sequential dependencies at all lags u.

Standardly GCT is based on the class of linear time series models for weakly stationary dynamic processes called vector autoregressive (VAR) models. The VAR(a) model for a p-variate series $\mathbf{y}(t)$ is defined by the following:

$$\mathbf{y}(t) = \boldsymbol{\mu} + \Phi_1\mathbf{y}(t-1) + \ldots + \Phi_a\mathbf{y}(t-a) + \varepsilon(t) \quad (1)$$

in which a is the non-negative integer indicating the order of the VAR. In what follows, to ease the presentation, it is assumed that the mean function equals zero: $\boldsymbol{\mu} = \mathbf{0}$. The zero-mean p-variate $\varepsilon(t)$ process is white noise because it is lacking any sequential dependencies. Hence the covariance function of $\varepsilon(t)$ only is non-zero at lag $u = 0$: $\text{cov}[\varepsilon(t), \varepsilon(t)'] = \boldsymbol{\Sigma}_\varepsilon$. The sequence of (p,p)-dimensional matrices $\boldsymbol{\Phi}_k$, $k = 1,\ldots,a$, contains along the diagonals the lagged autoregressive coefficients; the off-diagonal elements are the cross-lagged regression coefficients. Using the so-called decomposition theorem of Wold, it can be shown that any weakly stationary process can be approximated to any degree by a VAR(a) if its order a is chosen to be sufficiently large.

If the backshift operator B defined by $B\mathbf{y}(t) = \mathbf{y}(t-1)$ is introduced, each VAR(a) can be represented as follows:

$$\begin{aligned} \mathbf{y}(t) &= \boldsymbol{\Phi}_1\mathbf{y}(t\text{-}1) + \ldots + \boldsymbol{\Phi}_a\mathbf{y}(t\text{-}a) + \varepsilon(t) \\ &= \boldsymbol{\Phi}_1 B\mathbf{y}(t) + \ldots + \boldsymbol{\Phi}_a B^a\mathbf{y}(t) + \varepsilon(t) \end{aligned}$$

Specifying the matrix polynomial in B defined by $\boldsymbol{\Phi}(B,a) = \mathbf{I} - \boldsymbol{\Phi}_1 B - \ldots - \boldsymbol{\Phi}_a B^a$, where $\mathbf{I}$ represents the (p,p)-dimensional identity matrix, the VAR(a) can now be written as follows: $\boldsymbol{\Phi}(B,a)\mathbf{y}(t) = \varepsilon(t)$.

To obtain the frequency domain equivalent (discrete Fourier transform) of the VAR(a), which we will need later on in GCT in the frequency domain, we substitute the complex exponential for the backshift operator in the matrix polynomial representation. Substituting $B \rightarrow \exp[-2\pi i\omega_k] = \cos[2\pi\omega_k] - i\sin[2\pi\omega_k]$, where $i = \sqrt{-1}$ is the imaginary unit and $k = 0, 1,\ldots$, the discrete Fourier transformed representation of a VAR(a) at each frequency ω_k becomes the following:

$$\boldsymbol{\Phi}(\omega_k)\mathbf{y}(\omega_k) = \varepsilon(\omega_k) \quad (2)$$

where $\boldsymbol{\Phi}(\exp[-2\pi i\omega_k],a)$ is concisely written as $\boldsymbol{\Phi}(\omega_k)$; $\mathbf{y}(\omega_k)$ and $\varepsilon(\omega_k)$ are, respectively, the discrete Fourier transforms of $\mathbf{y}(t)$ and $\varepsilon(t)$ (e.g., Brillinger 1975). For a finite stretch of time series data $\mathbf{y}(t)$, $t = 0,1,\ldots,T-1$, the frequency ω_k in the discrete Fourier transform is defined as follows: $\omega_k = k/T$, $k = 0,1,\ldots,T-1$. Then, the discrete Fourier transform of $\mathbf{y}(t)$ is defined as $\mathbf{y}(\omega_k) = T^{-1/2}\sum\mathbf{y}(t)\exp[-2\pi i t\omega_k]$,

[1] I thank an anonymous reviewer for pointing out the distinctions between networks.

where the sum is over $t = 0, 1, \ldots, T-1$ for each ω_k.

Expression (2) constitutes the starting point for the derivation of the spectral density matrices associated with a VAR. These are obtained by first taking the inverse of $\mathbf{\Phi}(\omega_k)$: $\mathbf{\Gamma}(\omega_k) = \mathbf{\Phi}(\omega_k)^{-1}$. Then, the spectral density matrix $\mathrm{cov}[\mathbf{y}(\omega_k), \mathbf{y}(\omega_k)^*] = \mathbf{\Psi}(\omega_k)$ is proportional to the following:

$$\psi(\omega_k) \propto \Gamma(\omega_k) \sum_{\varepsilon} \Gamma(\omega_k)^* \tag{3}$$

where * denotes the complex conjugated transpose. In Eq. (3), $\mathbf{\Sigma}_{\varepsilon}$ is the full covariance matrix of the white process noise $\boldsymbol{\varepsilon}(t)$. For each frequency, $\omega_k \neq 0, \frac{1}{2}$, it holds that the (p,p)-dimensional complex-valued spectral density matrix $\mathbf{\Psi}(\omega_k)$ has real values (the autospectra) along the diagonal and its off-diagonal elements are complex valued.

GCT in the Time Domain

The focus will be on weakly stationary time series. Also, we assume that a Granger cause precedes its effect by at least one time step. Therefore, contemporaneous Granger causality (Lütkepohl 2007, section 2.3.1) will not be considered.

In order to convey the main ideas, we consider the special case in which the simple VAR(1) holds for a bivariate series: $\mathbf{y}(t) = \mathbf{\Phi}_1 \mathbf{y}(t-1) + \boldsymbol{\varepsilon}(t)$, where $\mathbf{y}(t) = [y_1(t), y_2(t)]'$. This model can be written out as follows:

$$\begin{aligned} y_1(t) &= \phi(1)_{11} y1(t-1) + \phi(1)_{12} y2(t-1) + \varepsilon_1(t) \\ y_2(t) &= \phi(1)_{21} y1(t-1) + \phi(1)_{22} y2(t-1) + \varepsilon_2(t) \end{aligned} \tag{4}$$

in which $\phi(1)_{lm}$ denotes the (l,m)-th element of $\mathbf{\Phi}_1$. GCT in this simple model boils down to testing whether $\phi_{12}(1)$ differs significantly from zero while $\phi_{21}(1)$ does not, which would imply that $y_2(t)$ is a Granger cause of $y_1(t)$. Or, alternatively, testing whether $\phi_{21}(1)$ differs significantly from zero while $\phi_{12}(1)$ does not, which would imply that $y_1(t)$ is a Granger cause of $y_2(t)$ (cf. Lütkepohl 2007, section 2.3).

We can straightforwardly generalize from a VAR(1) to a VAR(a) for bivariate series:

$$\begin{aligned} y_1(t) &= \phi(1)_{11} y_1(t-1) + \phi(1)_{12} y_2(t-1) + \ldots + \phi(a)_{11} y_1(t-a) + \phi(a)_{12} y_2(t-a) + \varepsilon_1(t) \\ y_2(t) &= \phi(1)_{21} y_1(t-1) + \phi(1)_{22} y_2(t-1) + \ldots + \phi(a)_{21} y_1(t-a) + \phi(a)_{22} y_2(t-a) + \varepsilon_2(t) \end{aligned} \tag{5}$$

where $\phi(k)_{lm}$ denotes the (I,m)-th element of $\mathbf{\Phi}_k$, $k = 1, 2, \ldots, a$. It now is more feasible not to test whether each of $\phi(k)_{12}$ or $\phi(k)_{21}$, $k = 1, \ldots, a$, differs significantly from zero but uses the following approach. First, fit Eq. (5) to the bivariate series $\mathbf{y}(t)$. Then, fit univariate AR(q) and AR(r) models to, respectively, $y_1(t)$ and $y_2(t)$ in which the orders q and r may differ from the order a in the bivariate model (5):

$$\begin{aligned} y_1(t) &= \phi(1) y_1(t-1) + \ldots + \phi(q) y_1(t-q) + \zeta(t) \\ y_2(t) &= \gamma(1) y_2(t-1) + \ldots + \gamma(r) y_2(t-r) + \xi(t) \end{aligned} \tag{6}$$

Define $\mathbf{\Sigma}_{\varepsilon}$ to be the (2,2)-dimensional covariance matrix of the white process noise $\boldsymbol{\varepsilon}(t)$ in Eq. (5). Then, the total interdependence $F_{y1,y2}$ between $y_1(t)$ and $y_2(t)$ is determined as follows (cf. Wen et al. 2013):

$$F_{y1,y2} = \ln\{\mathrm{var}[\zeta]\mathrm{var}[\xi] / \det[\Sigma_{\varepsilon}]\} \tag{7}$$

where $\ln\{.\}$ is the natural logarithm, det[.] is the determinant and var[$\zeta(t)$], and var[$\xi(t)$] are, respectively, the variances of the white process noise $\zeta(t)$ and $\xi(t)$ in Eq. (6). Geweke (1982) proves that $F_{y1,y2}$ allows for the following decomposition:

$$F_{y1,y2} = F_{y1 \to y2} + F_{y2 \to y1} + F_{y1 * y2} \tag{8}$$

where $F_{y1 \to y2} = \ln\{\mathrm{var}[\zeta(t)] / \mathrm{var}[\varepsilon_1(t)]\}$, var[$\varepsilon_1(t)$] is the (1,1) element of $\mathbf{\Sigma}_{\varepsilon}$; $F_{y2 \to y1} = \ln\{\mathrm{var}[\xi(t)] / \mathrm{var}[\varepsilon_2(t)]\}$, var[$\varepsilon_2(t)$] is the (2,2) element of $\mathbf{\Sigma}_{\varepsilon}$; and $F_{y1 * y2} = \ln\{\mathrm{var}[\varepsilon_1(t)]\mathrm{var}[\varepsilon_2(t)] / \det[\mathbf{\Sigma}_{\varepsilon}]\}$. If $F_{y1 \to y2}$ is large positive and $F_{y2 \to y1}$ small, then $y_1(t)$ is a Granger cause of $y_2(t)$; alternatively, if $F_{y2 \to y1}$ is large positive and $F_{y1 \to y2}$ is small, then $y_2(t)$ is a Granger cause of $y_1(t)$.

To generalize this presentation to p-variate time series, where $p > 2$, is straightforward (see Wen et al. 2013 for details). Select two non-overlapping subseries from the p-variate time series $\mathbf{y}(t) = [y_1(t), \ldots, y_p(t)]'$ in which the p_1-variate series $\mathbf{y}_1(t)$ consists of p_1 univariate component series of $\mathbf{y}(t)$ and the p_2-variate series $\mathbf{y}_2(t)$ consists of p_2 different univariate component series of $\mathbf{y}(t)$. For instance, for $p = 7$, $p_1 = 2$, and $p_2 = 3$: $\mathbf{y}_1(t) = [y_1(t), y_3(t)]$ and $\mathbf{y}_2(t) = [y_2(t), y_5(t), y_7(t)]'$. Here, $\mathbf{y}_1(t)$ and $\mathbf{y}_2(t)$ are treated in the same way as the univariate series $y_1(t)$ and $y_2(t)$, that is, we determine appropriate multivariate generalizations of Eqs. (5)–(8). For instance, for the computation of $F_{y1 \to y2}$ and $F_{y2 \to y1}$ with multivariate subsets, we replace variances by determinants of (p_1,p_1)- and (p_2,p_2)-dimensional covariance matrices of the white process noise (see Barnett and Seth 2014).

The most important issue regarding GCT in the time domain based on comparison of $F_{y1 \to y2}$ and $F_{y2 \to y1}$, where $\mathbf{y}_1(t)$ and $\mathbf{y}_2(t)$ are, respectively, non-overlapping p_1- and p_2-variate subseries of a given p-variate series $\mathbf{y}(t)$, has to do with the consistency of results obtained in a sequence of such tests

obtained with different sets of subseries. For $p > 2$, several distinct subsets of non-overlapping p_1- and p_2-variate subseries $\mathbf{y}_1(t)$ and $\mathbf{y}_2(t)$ can be considered and due to sampling variability, one cannot guarantee that the results thus obtained converge to a consistent overall causal network. The frequency domain approach to GCT does not have that issue.

GCT in the Frequency Domain

A decomposition akin to Eq. (8) also can be derived in the frequency domain (cf. Wen et al. 2013). In the frequency domain, however, we instead focus on a set of measures derived from the spectrum $\mathbf{\Psi}(\omega_k)$ given by Eq. (3). There are several such measures; for an overview, see Schlögl and Supp (2006). In what follows, the focus is on a single measure, namely the partial directed coherence (PDC) which is defined below. But first, an important theoretical equivocality has to be addressed.

Two Equivalent Representations of a VAR(*a*) Until now, it was taken for granted that a VAR(*a*) for a weakly stationary p-variate series $\mathbf{y}(t)$ is given by the following:

$$\mathbf{y}(t) = \Phi_1\mathbf{y}(t-1) + \ldots + \Phi_a\mathbf{y}(t-a) + \varepsilon(t) \tag{9}$$

in which the white process noise $\varepsilon(t)$ has full covariance matrix $\mathbf{\Sigma}_\varepsilon$. However, there exists an equivalent VAR(*a*) representation defined by the following:

$$\mathbf{y}(t) = \Xi_0\mathbf{y}(t) + \Xi_1\mathbf{y}(t-1) + \ldots + \Xi_a\mathbf{y}(t-a) + \upsilon(t) \tag{10}$$

in which the white process noise $\upsilon(t)$ has *diagonal* covariance matrix $\mathbf{\Sigma}_\upsilon$. That is, the univariate component processes $\upsilon_k(t)$, $k = 1,\ldots, p$, are mutually uncorrelated at each time t. Notice that in Eq. (10), the contemporaneous relations among the component series of $\mathbf{y}(t)$ are given by unidirectional relations in $\mathbf{\Xi}_0$.

In agreement with the literature concerned, the model (9) will be referred to as the *standard VAR*, while the model (10) will be called a *structural VAR*. Each standard VAR (Eq. 9) can be transformed into an equivalent structural VAR (Eq. 10) by using the following decomposition of $\mathbf{\Sigma}_\varepsilon$:

$$\Sigma_\varepsilon = (\mathbf{I}-\Xi_0)^{-1}\Sigma_\upsilon(\mathbf{I}-\Xi_0)^{-T} \tag{11}$$

In which the superscript $^{-T}$ denotes inversion followed by transposition. We note that Eq. (11) constitutes a Cholesky decomposition. It then can be shown that the coefficient matrices associated with the lagged regression matrices in both equivalent models are given by the following (cf. Gates et al. 2011):

$$\Xi_k = (\mathbf{I}-\Xi_0)\Phi_k, \text{and } \Phi_k = (\mathbf{I}-\Xi_0)^{-1}\Xi_k, k = 1, \ldots, a. \tag{12}$$

It is clear from Eq. (12) that the choice of model (9) or (10), while statistically equivalent, can yield entirely different results in Granger causality testing if $\mathbf{\Xi}_0$ is not equal to the zero matrix. An example is given in Fig. 1, showing a standard VAR and an equivalent structural VAR for the same three-variate time series. The first component series $y_1(t)$ is a Granger cause for $y_3(t)$ in the standard VAR, but not so in the equivalent structural VAR.

The difference between the two equivalent models is that in the structural VAR (Eq. 10) the contemporaneous relations among the univariate component series $y_k(t)$, $k = 1,\ldots, p$, are represented by explicit unidirectional regressions with coefficients in $\mathbf{\Xi}_0$, whereas in the standard VAR (Eq. 9), the contemporaneous relations are derived from the contemporaneous associations among the univariate component process noise series $\varepsilon_k(t)$ constituting off-diagonal elements in $\mathbf{\Sigma}_\varepsilon$, $k = 1,\ldots,p$. One way to interpret this difference is as follows: in the structural VAR given by Eq. (10), contemporaneous relations are generated within the system composed of the $\mathbf{y}(t)$ process (endogenous) while in the standard VAR given by Eq. (9), contemporaneous relations are generated outside of the system composed of the $\mathbf{y}(t)$ process (exogenous).

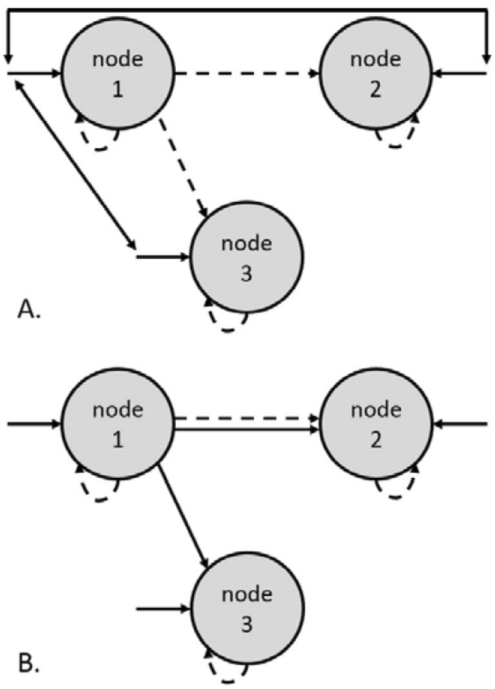

Fig. 1 Equivalent standard VAR (**a**) and structural VAR (**b**) for a given three-variate time series. The univariate component series are depicted by nodes. Contemporaneous relations are depicted by continuous links, lagged relations of order 1 by broken links. Weights of the links are not shown. One-way arrows show directed links; two-way arrows show undirected links. In both the standard and structural VARs, node 1 is a Granger cause of node 2. But only in the standard VAR that node 1 is a Granger cause of node 3

PDC as a Frequency Domain Index of Granger Causality

First, the standard VAR(a) (Eq. 9) for a p-variate observed time series $\mathbf{y}(t)$ is considered. Discrete Fourier transformation of the associated polynomial in the backshift operator B, $\mathbf{\Phi}(B,a) = \mathbf{I} - \mathbf{\Phi}_1 B - \ldots - \mathbf{\Phi}_a B^a$, yields at each frequency ω_k a (p,p)-dimensional complex-valued matrix $\mathbf{\Phi}(\omega_k)$ specified by Eq. (2). Let $\mathbf{\Delta}_\varepsilon$ be the diagonal (p,p)-dimensional matrix with the inverse diagonal elements of $\mathbf{\Sigma}_\varepsilon$ along the diagonal and zero off-diagonal elements. The m-th diagonal element of $\mathbf{\Delta}_\varepsilon$ is $1/\mathrm{var}[\varepsilon_m(t)]$, $m = 1,\ldots,p$. Then, the generalized partial directed coherence (gPDC) $\pi_{im}(\omega_k)$ between univariate component series $y_i(t)$ and univariate component series $y_m(t)$ is defined as follows (Faes and Nollo 2010):

$$\pi_{im}(\omega_k) = \left\{\Phi_{im}(\omega_k)/\mathrm{var}\left[\varepsilon_i(t)\right]\right\}/\mathrm{sqrt}\left[\phi_{:m}(\omega_k)^{*}\Delta_\varepsilon\phi_{:m}(\omega_k)\right] \quad (13)$$

where sqrt[.] is the square root, and $\mathbf{\phi}_{:m}(\omega_k)$ denotes the m-th column of $\mathbf{\Phi}(\omega_k)$. Because $\pi_{im}(\omega_k)$ is complex-valued, it is customary to take the absolute value $|\pi_{im}(\omega_k)|$, where because of the normalization it always holds that $1 \geq |\pi_{im}(\omega_k)| \geq 0$.

GCT based on the gPDC now is straightforward. If for univariate component series $y_i(t)$ and $y_m(t)$, it is found that $|\pi_{im}(\omega_k)|$is large whereas if $|\pi_{im}(\omega_k)|$ is about zero, then $y_m(t)$ is a Granger cause of $y_i(t)$ at frequency ω_k. For a p-variate series, $\mathbf{y}(t)$ application of the gPDC-based GCT thus involves p(p-1)/2 of such pairwise comparisons at each frequency.

Next, we consider the equivalent structural VAR(a) given by Eq. (10). We define the polynomial in the backshift operator B: $\mathbf{\Psi}[B,a) = \mathbf{I} - \mathbf{\Xi}_1 B - \ldots - \mathbf{\Xi}_a B^a$. Notice that $\mathbf{\Xi}_0$ is replaced by the identity matrix $\mathbf{I}$ in this definition. The discrete Fourier transform at each frequency ω_k of $\mathbf{\Psi}[B,a)$ is given by the (p,p)-dimensional complex-valued matrix $\mathbf{\Psi}(\omega_k)$. We let $\mathbf{\Delta}_\upsilon$ be the inverse of the diagonal covariance matrix $\mathbf{\Sigma}_\upsilon$. Then, the instantaneous partial directed coherence (iPDC) $\chi_{im}(\omega_k)$ between univariate component series $y_i(t)$ and univariate component series $y_m(t)$ at each frequency ω_k is given by Faes and Nollo (2010):

$$\chi_{im}(\omega_k) = \left\{\Psi_{im}(\omega_k)/\mathrm{var}\left[\upsilon_i(t)\right]\right\}/\mathrm{sqrt}\left[\psi_{:m}(\omega_k)^{*}\Delta_u\psi_{:m}(\omega_k)\right] (14)$$

in which $\mathbf{\psi}_{:m}(\omega_k)$ is the m-th column of $\mathbf{\Psi}(\omega_k)$. Because $\chi_{im}(\omega_k)$ is complex-valued, the absolute value $|\chi_{im}(\omega_k)|$ is taken, where because of the normalization it always holds that $1 \geq |\chi_{im}(\omega_k)| \geq 0$.

Note that Eq. (14) is based on the following polynomial in the backshift operator B: $\mathbf{\Psi}[B,a) = \mathbf{I} - \mathbf{\Xi}_1 B - \ldots - \mathbf{\Xi}_a B^a$. This polynomial leaves out of consideration $\mathbf{\Xi}_0$ which contains the directed contemporaneous connections. Hence, the iPDC is an indicator of purely lagged Granger causality.

A New Data-Driven Solution to Granger Causality Testing

Faes and Nollo (2010) present results obtained with time series generated according to a structural VAR, showing that the iPDC outperforms the gPDC. The iPDC is based on fitting the structural VAR (Eq. 10) to the data, almost always by means of a two-step procedure. In the first step, a standard VAR(a) (Eq. 9) is fitted to the data. Then, the estimated full covariance matrix of the white process noise in Eq. (9) is subjected to a Cholesky decomposition given by Eq. (11), which yields the coefficient matrix $\mathbf{\Xi}_0$ of the equivalent structural VAR together with the diagonal covariance matrix $\mathbf{\Sigma}_\upsilon$ of its white noise process. The remaining coefficient matrices $\mathbf{\Xi}_k$, $k = 1,\ldots,a$, in the equivalent structural VAR then are derived using Eq. (12).

There is, however, a major problem associated with this two-step procedure, namely the results obtained in the Cholesky decomposition (Eq. 11) depend on the ordering of the p univariate component series $y_k(t)$, $k = 1,\ldots,p$ (cf. Lütkepohl 2007, pp. 61–62). If this ordering is permuted, the results obtained in the Cholesky decomposition of the associated permuted covariance matrix of the white process noise also change in that a different $\mathbf{\Xi}_0$ is obtained. Because the ordering of the univariate component series in a vector-valued observed time series should be conceived of as being arbitrary, this dependence of the coefficient matrices in the structural VAR obtained by means of the two-step procedure on the particular ordering is problematic.

To avoid the two-step procedure with its attendant ordering problem, we developed an alternative approach to fit structural VARs directly to the data. This direct approach, described in Gates et al. (2011), consists of rewriting the structural VAR as a structural equation model (SEM), called the unified SEM (uSEM), and using standard SEM software to fit the model to the data. The results of a large-scale simulation study validating this alternative approach can be found in Gates and Molenaar (2012). Henceforth, we will refer to this alternative approach as fitting a uSEM. The computer program implementing the fit of a uSEM can be freely obtained at http://CRAN.R-project.org/package=gimme (Lane et al. 2014)

The availability of an alternative approach that avoids the problems with the usual two-step approach to fit structural VARs is also important because it opens up a possibility to solve the problem of equivalent VAR representations in GCT. This problem was described above as involving that the standard VAR given by Eq. (9) and the structural VAR given by Eq. (10) are equivalent, but that the results of Granger causality testing based on either Eq. (9) or (10) can be different (see Fig. 1 above). For empirical data, the true dynamic model almost always is unknown. Our new approach to fit a uSEM directly to the data can, however, be generalized to determine in a data-driven way what the appropriate representation is.

The Fit of a uSEM To describe the main steps in fitting a uSEM, the structural VAR(1) is taken as the most elementary example (see Gates et al. 2011 for a complete description). Equation (10) reduces for a VAR(1) to $\mathbf{y}(t) = \Xi_0\mathbf{y}(t) + \Xi_1\mathbf{y}(t-1) + \upsilon(t)$. To fit this model to the data, the model is expanded and considered for two consecutive time points t and t + 1 in the following way:

$$\begin{aligned} \mathbf{y}(t) &= \upsilon^{\circ}(t) \\ \mathbf{y}(t+1) &= \Xi_0\mathbf{y}(t+1) + \Xi_1\mathbf{y}(t) + \upsilon(t+1) \end{aligned} \tag{15}$$

The first equation in Eq. (15) acts as an initial condition, where $\upsilon^{\circ}(t)$ denotes p-variate pseudo-noise equal to $\mathbf{y}(t)$. The (2p,2p)-dimensional covariance matrix implied by Eq. (15) is called a block-Toeplitz matrix (Molenaar 1985), the estimate of which serves as input to the SEM software. To fit a uSEM, the following steps are carried out:

a. Fit Eq. (15) to the block-Toeplitz input covariance matrix while fixing all coefficients in Ξ_0 and Ξ_1 at zero. Only the diagonal elements of the (p,p)-dimensional subcovariance matrix of $\upsilon(t+1)$ are freed up. Also, free up all non-redundant elements of the (p,p)-dimensional subcovariance matrix of $\upsilon^{\circ}(t)$. The goodness-of-fit of this model will in general be bad. If so, go to step b. If not, stop.
b. Determine the Lagrange multiplier tests of all fixed parameters in Ξ_0 and Ξ_1; each value of this test has asymptotically a chi-square distributed with one degree of freedom. Select the fixed parameter having the largest value of the Lagrange multiplier test.
c. If this value is significant, free up this parameter. Refit the model thus extended and go to b. If not, stop.

Extension of the Fit of a uSEM In the procedure described in the previous section, the (p,p)-dimensional subcovariance matrix of $\upsilon(t+1)$ is not subject to change. It is a diagonal covariance matrix during all steps of the procedure. Consequently, the result of the above procedure will be a uSEM.

However, step b in the above procedure can be extended in such a way that the data decide whether a structural VAR, a standard VAR, or a hybrid combination of structural and standard VAR (hybrid VAR) is the appropriate representation. To accomplish this, step b is changed as follows:

> b′. Perform the Lagrange multiplier tests of all fixed parameters in Ξ_0, Ξ_1 and the (p,p)-dimensional subcovariance matrix of $\upsilon(t+1)$. Each such value has asymptotically a chi-square distribution with one degree of freedom. Select the fixed parameter having the largest value of the Lagrange multiplier test.

It is claimed that with the stepwise procedures a, b′, and c, the best-fitting representation is obtained, whether it is a standard VAR, a structural VAR, or a hybrid VAR. The new procedure will be denoted as fitting a hybrid VAR and is depicted in Fig. 2. If in this extended approach only parameters in Ξ_0 and Ξ_1 happen to be freed up, the best-fitting representation is a uSEM. If only parameters in Ξ_1 and the (p,p)-dimension subcovariance matrix of $\upsilon(t+1)$ happen to be freed up, the best-fitting representation is a standard VAR. If parameters in Ξ_0, Ξ_1 and the (p,p)-dimensional subcovariance matrix of $\upsilon(t+1)$ happen to be freed up, the best-fitting representation is a hybrid VAR.

Application of the Hybrid VAR Fit to Simulated Data We use the following hybrid VAR(1) to simulate data:

$$\begin{aligned} y_1(t) &= 0.7y_1(t-1) + \eta_1(t) \\ y_2(t) &= 0.7y_1(t) + 0.7y_2(t-1) + \eta_2(t) \\ y_3(t) &= 0.7y_3(t-1) + \eta_3(t) \\ y_4(t) &= 0.7y_4(t-1) + \eta_4(t) \end{aligned} \tag{16}$$

where $\mathrm{var}[\eta_1(t)] = \mathrm{var}[\eta_2(t)] = \mathrm{var}[\eta_3(t)] = \mathrm{var}[\eta_4(t)] = 1$. Unit variances are used to simplify the application. All Gaussian white process noise component series are uncorrelated, save for $\eta_3(t)$ and $\eta_4(t)$: $\mathrm{cov}[\eta_3(t), \eta_4(t)] = 0.7$. Consequently, the simulation model is a hybrid VAR(1) including a directed contemporaneous relation from $y_1(t)$ to $y_2(t)$ as well as a non-zero off-diagonal element of the (4,4)-dimensional covariance matrix of $\xi(t)$ representing a contemporaneous association between $\eta_3(t)$ and $\eta_4(t)$. Notice that there are no cross-lagged relations in Eq. (16); hence, lagged Granger causality is absent.

Application of the hybrid VAR fit procedure to a time series of length $T = 900$ simulated using this model yields the true hybrid model structure, where all estimated parameters are within 95% confidence intervals about their true values. For instance, the estimate of the directed contemporaneous relation from $y_1(t)$ to $y_2(t)$ (true value 0.7) is 0.7 with standard error s.e. = 0.03. The estimate of the covariance between $\eta_3(t)$ and $\eta_4(t)$ (true value 0.7) is 0.66 (s.e. = 0.04). Figure 3 shows the recovered hybrid VAR model and the iPDCs showing that the series lack lagged Granger causality.

The fit of a standard VAR to the same data yields a goodness-of-fit that is of the same quality as that of a hybrid VAR, but the set of estimated parameters contains two spurious elements. All autoregressive parameters are correctly recovered with values within 95% of the confidence intervals about their true values, but there now also is a significant cross-lagged relationship according to which $y_1(t-1)$ influences $y_2(t)$ with estimated value 0.5 (s.e. = 0.04). This significant cross-lagged relationship incorrectly implies that $y_1(t)$ is a lagged Granger cause of

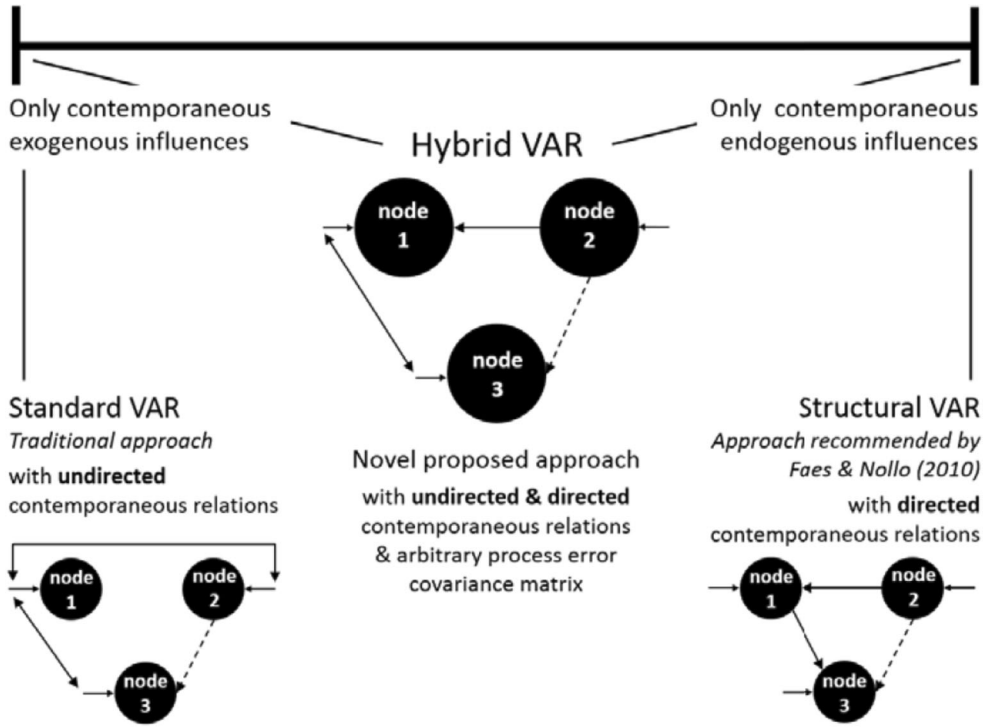

Fig. 2 Hybrid, standard, and structural VARs for a given three-variate time series of which the univariate component series are depicted by nodes. Links are depicted as in Fig. 1. The three models are not equivalent but show their structural differences. The hybrid VAR can have all possible configurations of exogenous and endogenous random environmental influences (process noise). The exogenously generated contemporaneous relations are manifested by the covariance matrix of the process noise having non-zero off-diagonal elements depicted by continuous two-way arrows; the endogenously generated contemporaneous influences are manifested by directed continuous arrows among the nodes. Limiting cases of the hybrid VAR are the standard VAR in which all contemporaneous relations are generated exogenously and the structural VAR in which all contemporaneous relations are generated endogenously

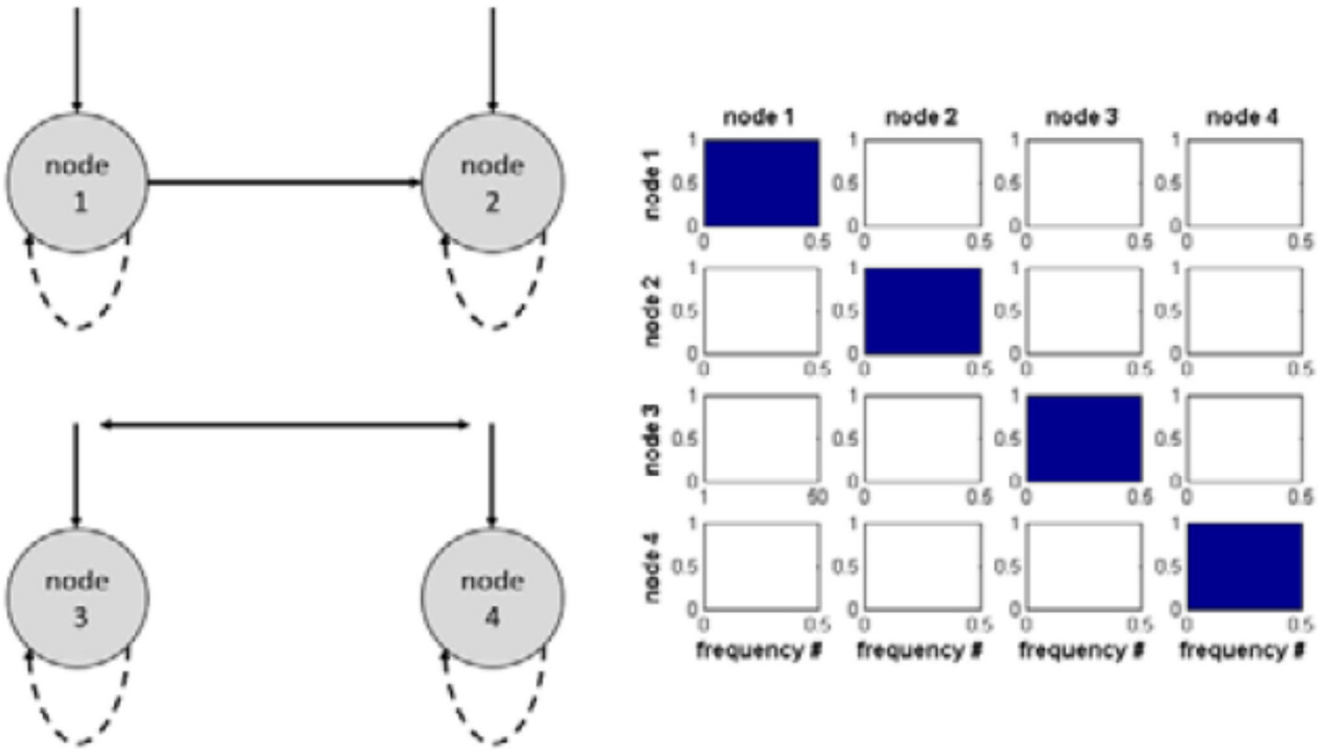

Fig. 3 Estimated hybrid VAR which corresponds to the true model with which the four-variate time series was generated. On the right-hand side are shown the iPDCs, indicating that there is no lagged Granger causality among the four-component series

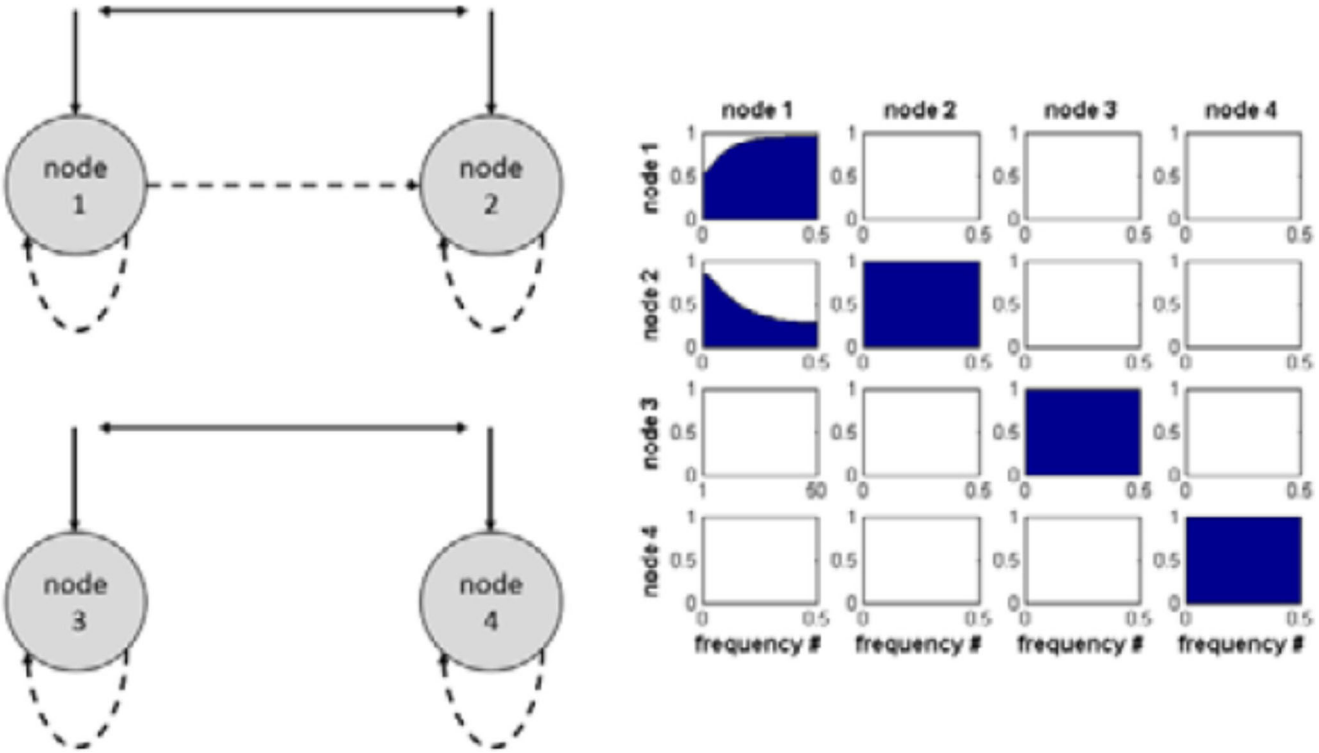

Fig. 4 Standard VAR obtained with the same procedure as developed for fitting uSEMs described in the text. The standard VAR was fitted to the data generated by a hybrid VAR without lagged Granger causality. The gPDCs on the right-hand side indicate that $y_1(t)$ is a lagged Granger cause of $y_2(t)$

$y_2(t)$. Also, the correct recovery of the covariance between $\eta_3(t)$ and $\eta_4(t)$ with estimated value 0.66 (s.e. = 0.04) is accompanied by a significant covariance between $\eta_2(t)$ and $\eta_1(t)$ with estimated value 0.7 (s.e. = 0.05). Figure 4 shows the estimated standard VAR and the gPDCs showing that $y_1(t)$ is a (spurious) Granger cause of $y_2(t)$.

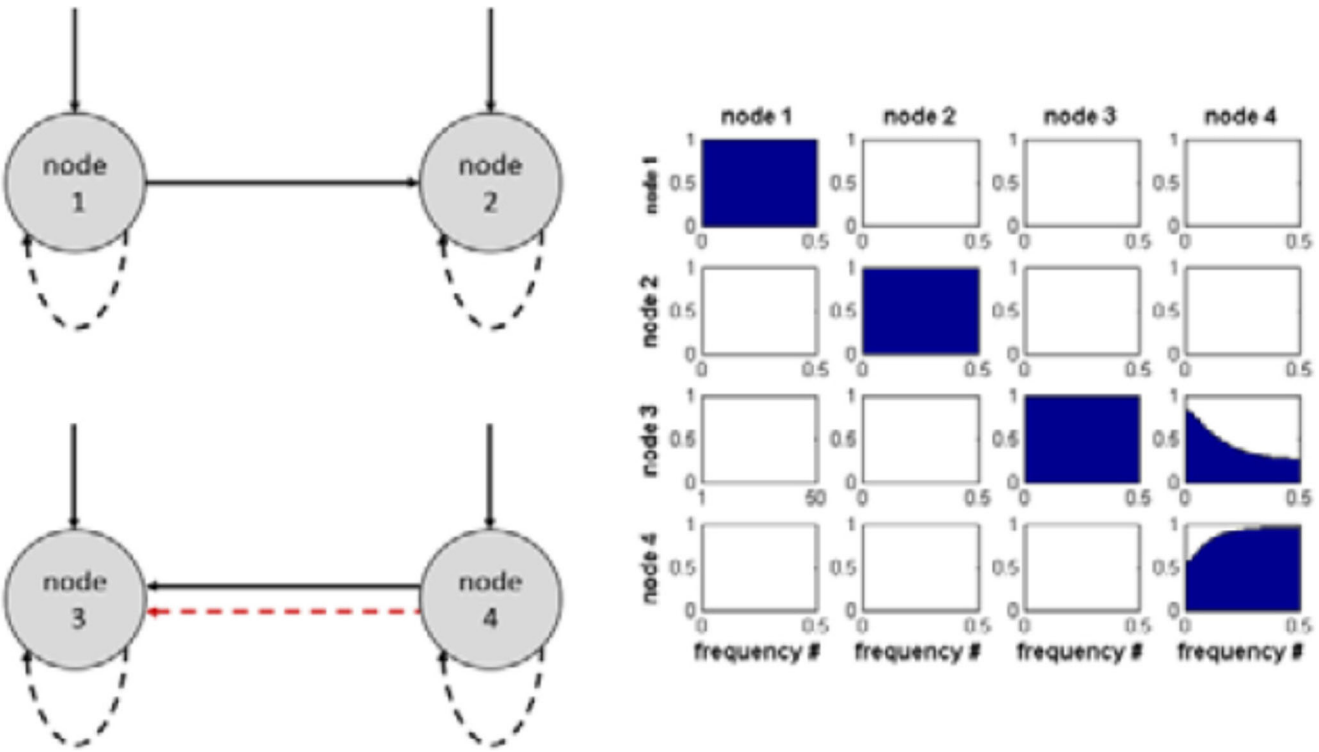

Fig. 5 Structural VAR fitted to the data generated with a hybrid VAR without lagged Granger causality. The iPDCs on the right-hand side show that $y_4(t)$ is a lagged Granger cause of $y_3(t)$

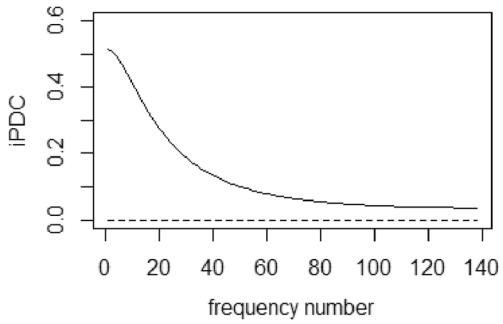

Fig. 6 iPDC of the Graanger causal influence of child on therapist (continuous line) and of therapist on child (broken line). Abcis is k index in frequency ω_k in Eq. (14)

To close with, the fit of a structural VAR (uSEM) to the same data again yields a goodness-of-fit that is of the same quality as that of a hybrid VAR, yet the set of estimated parameters contains two spurious elements. All autoregressive parameters as well as the directed contemporaneous relation from $y_1(t)$ to $y_2(t)$ are correctly retrieved with their estimated values within 95% confidence interval about the true values. In addition, a contemporaneous directed relation from $y_4(t)$ to $y_3(t)$ with estimated value 0.69 (s.e. = 0.02) is also estimated. It is clear that this contemporaneous directed relation captures the contemporaneous covariance between $\eta_3(t)$ and $\eta_4(t)$ in the simulation model. However, a significant directed lagged cross-relation from $y_4(t-1)$ to $y_3(t)$ with estimated value −0.48 (s.e. = 0.03) also is freed up. The latter spurious directed lagged relation implies that $y_4(t)$ is a lagged Granger cause of $y_3(t)$. The estimated structural VAR is shown in Fig. 5; the iPDCs show that $y_4(t)$ is a (spurious) Granger cause of $y_3(t)$.

It is noted that the illustration is limited in important ways (one replication, one simulation model, etc.). What is needed is a large-scale simulation study in order to thoroughly investigate its statistical properties. This will be carried out in the near future. The results of a successful validation of the fit of uSEMs using the sequential procedure based on modification indices with large-scale simulated data are reported in Gates and Molenaar (2012).

Empirical Application In Liu and Molenaar (2016), an application of Granger causality testing based on fitting standard VARs to a bivariate time series of electro-dermal activity (EDA) is reported.[2] The two component series are EDA of a child with sensory processing disorder and the EDA of his therapist during interaction in occupational therapy. Figure 6 depicts the iPDCs thus obtained: it clearly shows the Granger causal influence of the child's EDA on the therapist EDA but not vice versa. This result corroborates the result reported in Liu and Molenaar (2016).

[2] Thanks are due to Dr. Matthew Goodwin, Northeastern University, for allowing to use this data.

Conclusion

With the advent of ambulatory and ecological momentary assessment methods, intensive longitudinal data become increasingly available in prevention science. The multivariate time series data thus obtained can be subjected to GCT in order to determine important causal relations among the univariate component series. These causal relations can be exploited when considering interventions. It was explained that the VAR models in which GCT is based come in at least two equivalent variants which each can lead to different conclusions about causal relations. A new data-driven approach based on a hybrid VAR was described with which it can be determined which version should be selected to validly base GCT on.

In closing, it is remarked that to conclude that Granger causality is the case in the way as presented in this paper requires that all potential causes are part of the data to which the test is applied. This rules out consideration of subsets of variables to carry out GCT. GCT is comparable in this respect to functional approaches to causality testing like Pearl's (2000) approach (for an interesting comparison between GCT and Pearl's approach, the reader is referred to White et al. 2013).

Funding Funding of the research presented in this paper was partially provided by NSF 1157220 (PI PCM Molenaar).

Compliance with Ethical Standards

Conflict of Interest The author declares that there is no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by the author.

Informed Consent Informed consent was not required for this study.

References

Barnett, L., & Seth, A.K. (2014). The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inferemve. *Journal of Neuroscience Methods, 223,* 50–68.

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry, 16*, 5–13. https://doi.org/10.1002/wps.20375.

Brillinger, D. R. (1975). *Time series: Data analysis and theory.* New York: Holt, Rinehart & Winston.

Faes, L., & Nollo, G. (2010). Extended causal modeling to assess partial directed coherence in multiple time series with significant instantaneous interactions. *Biological Cybernetics, 103*, 387–400. https://doi.org/10.1007/s00422-010-0406-6.

Gates, K. M., & Molenaar, P. C. M. (2012). Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage, 63*, 310–319. https://doi.org/10.1016/j.neuroimage.2012.06.026.

Gates, K.M., Molenaar, P.C.M., Hillary, F.G., & Slobounov, S. (2011). Extended unified SEM approach for modeling event-related fMRI data. *NeuroImage, 54,* 1151–1158.

Geweke, J. (1982). Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association, 77*, 304–313. https://doi.org/10.1080/01621459.1982.10477803.

Gibbons C.J. (2016) Turning the page on pen-and-paper questionnaires: Combining ecological momentary assessment and computer adaptive tests to transform psychological assessment in the 21st century. Frontiers in Psychology, 7 DOI: https://doi.org/10.3389/fpsyg.2016.01933.

Goebel, R., Roebroeck, A., Kim, D. S., & Formisano, E. (2003). Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magnetic Resonance Imaging, 21*, 1251–1261. https://doi.org/10.1016/j.mri.2003.08.026.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica, 37*, 424–438. https://doi.org/10.1017/ccol052179207x.002.

Lane, S., Gates, K.M., & Molenaar, P. C. M. (2014). *Gimme: Group iterative multiple model estimation*. R package version 0.1–2. http://CRAN.R-project.org/package=gimme.

Liu, S., & Molenaar, P. C. M. (2016). Testing for Granger causality in the frequency domain: A phase resampling method. *Multivariate Behavioral Research, 51*, 53–66. https://doi.org/10.1080/00273171.2015.1100528.

Lütkepohl, H. (2007). *New introduction to multiple time series analysis*. Berlin: Springer-Verlag.

Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika, 50*, 181–202. https://doi.org/10.1007/BF02294246.

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research and Perspectives, 2*, 201–218. https://doi.org/10.1207/s15366359mea0204_1.

Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychology, 18*, 112–117. https://doi.org/10.1111/j.1467-8721.2009.01619.x.

Molenaar, P. C. M., & Lo, L. L. (2016). Alternative forms of Granger causality, heterogeneity and nonstationarity. In W. Wiedermann & A. von Eye (Eds.), *Statistics and causality: Methods for applied empirical research* (pp. 205–229). Hoboken: Wiley.

Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge: Cambridge University Press.

Schlögl, A., & Supp, G. (2006). Analyzing event-related EEG data with multivariate autoregressive parameters. In C. Neuper & W. Klimesch (Eds.), *Event-related dynamics of brain oscillations. Progress in brain research 159* (pp. 135–147). Amsterdam: Elsevier.

Spaniel, F., Vohlidka, P., Hrdlicka, J., Kozeny, J., Novak, T., Motlova, L., Cermak, J., Bednarik, J., Novak, D., & Höschl, C. (2008). ITAREPS: Information technology aided relapse prevention programme in schizophrenia. *Schizophrenia Research, 98*, 312–317. https://doi.org/10.1016/j.schres.2007.09.005.

Trull, T. J., & Ebner-Priemer, U. (2013). Ambulatory assessment. *Annual Review of Clinical Psychology, 9*, 151–176. https://doi.org/10.1146/annurev-clinpsy-050212-185510.

Wen, X., Rangarajan, G., & Ding, M. (2013). Multivariate Granger causality: An estimation framework based on factorization of the spectral density matrix. *Philosophical Transactions of the Royal Society A, 371*, 20110610. https://doi.org/10.1098/rsta.2011.0610.

White, H., Chalak, K., & Lu, X. (2013). Linking Granger causality and the Pearl causal model with settable systems. In F. Popescu & I. Guyon (Eds.), *Causality in time series. Challenges in machine learning 5* (pp. 107–137). Brookline: Microtome Publishing.