INDIAN INSTITUTE OF TECHNOLOGY MADRAS

Department of Chemical Engineering

CH5350 Applied Time Series Analysis

Project

Due: Thursday, November 30, 2017

---

1. A broad class of modelling problems is primarily concerned with building a model between "causal" and "response" variables or effects. A more advanced problem, however, is that of determining from among a given set of variables which acts as causes and/or effects (it is possible for a variable to be both a cause and an effect, especially in bi-directional relations). In this respect, statistical analysts from different fields have proposed several theoretical and empirical measures of **causality**. A widely used empirical measure is *Granger causality*, which stems from econometrics, but finds wide-applicability in several scientific areas. The underlying idea in determining if a signal z_j is potentially a cause of another signal z_i is to compare the one-step ahead forecasts of z_i with and without the past of z_j. The roles are reversed while testing for causality in the other direction. Granger recommends for this purpose, the use of what are known as *vector auto-regressive (VAR) models*, which are an extension of the univariate AR representation to the vector case:

$$\mathbf{z}[k] = \sum_{r=1}^{P} \mathbf{A}_r \mathbf{z}[k-r] + \mathbf{e}[k] \tag{1}$$

where $\mathbf{z}[k] = \begin{bmatrix} z_1[k] & z_2[k] & \cdots & z_M[k] \end{bmatrix}^T$, $\mathbf{A}_r,\ r = 1, \cdots, P$ are $M \times M$ coefficient matrices, and $\mathbf{e}[k] \sim \text{GWN}(\mathbf{0}, \Sigma_{\mathbf{e}})$. The Granger-causal (essentially a lagged causal) relationship does not exist from variable $z_j \rightarrow z_i$ if and only if

$$a_{ij}(r) = 0, \quad \forall r \tag{2}$$

and vice versa when testing for causality in the direction $z_i \rightarrow z_j$. In practice, the estimated VAR model coefficient matrices have to be used; in which case, the existence of Granger-causal relationship from $z_j \rightarrow z_i$ can be tested by conducting a hypothesis test using the statistic,

$$S_{ij} = N \hat{\mathbf{a}}_{ij}^T \hat{\mathbf{V}}_{ij}^{-1} \hat{\mathbf{a}}_{ij} \tag{3}$$

where N is the number of observations, $\hat{\mathbf{a}}_{ij} = (a_{ij}(1), a_{ij}(2), \cdots, a_{ij}(P))^T$ is the AR coefficient estimates for the pair of variables (z_i, z_j) at all lags, and $\hat{\mathbf{V}}_{ij}$ is a $P \times P$ matrix for the pair of variables obtained from the estimate of covariance matrix, $\hat{\Sigma}_{\mathbf{A}}$ of VAR model coefficients (the size of $\hat{\Sigma}_{\mathbf{A}}$ is $M^2P \times M^2P$). Under the null hypothesis that z_j does not Granger-cause z_i, the test statistic S_{ij} is asymptotically χ^2-distributed with P degrees of freedom.

There also exist frequency-domain implementations of Granger causality, of which the direct pathway function (DPF) is an effective measure. The DPF, denoted by $\psi_{ij}(\omega)$, directionally connects the innovations e_j to the variable z_i at each frequency. The following expressions

are necessary to compute the DPF from a VAR model of the process.

$$\psi_{ij}(\omega) = \frac{h_{D,ij}(\omega)}{\sqrt{\sum_{i=1}^{M} |h_{D,ij}(\omega)|^2}} \tag{4a}$$

$$h_{D,ij}(\omega) = \begin{cases} \dfrac{-\bar{a}_{ij}(\omega)\det(\bar{\mathbf{M}}_{ij})}{\det(\bar{\mathbf{A}}(\omega))}, & i \neq j \\ \dfrac{\det(\mathbf{M}_{ij})}{\det(\bar{\mathbf{A}}(\omega))}, & i = j \end{cases} \tag{4b}$$

$$\bar{\mathbf{A}}(\omega) = \mathbf{I} - \sum_{r=1}^{p} \mathbf{A}_r e^{-jr\omega} \tag{4c}$$

where $h_{D,ij}(\omega)$ is the frequency response function of the direct pathway from e_j to z_i. The quantity $\bar{\mathbf{M}}_{ij}(\omega)$ is the minor matrix obtained from $\bar{\mathbf{A}}(\omega)$ by eliminating both i^{th} and j^{th} row and column, while $\mathbf{M}_{ij}(\omega)$ is the standard minor matrix obtained by eliminating the i^{th} row and j^{th} column, respectively from $\bar{\mathbf{A}}(\omega)$.

A variable z_j does not Granger-cause z_i if and only if

$$|\hat{\psi}_{ij}(\omega)|^2 = 0, \quad \forall\omega \tag{5}$$

Using the above information, do the following.

(a) Write an `R` function to implement the time-domain Granger-causality test given an estimated VAR model using the statistic S_{ij} in (3) (use `Vhat` function posted on the course website and provide the required input arguments, i.e., multivariate data and order to get $\hat{\mathbf{V}}_{ij}$).

(b) Write an `R` function to compute the DPF (ψ_{ij}) given the VAR model and optionally a frequency band. The function should also have an option of plotting the squared-magnitude of DPF for all channels.

(c) A closed-form expression for the distribution of $|\hat{\psi}_{ij}(\omega)|^2$ is difficult to derive. Therefore, the test of zero Granger causality is determined through surrogate data analysis as described below.

The basic idea is to generate *artificial* realizations of the data that are commensurate with the *null hypothesis* (of zero Granger-causality) conditions while preserving certain properties / features of the original time-series. In the approach below, a frequency-domain method is used where the phase is randomized while preserving the auto-correlation of original series. The phase randomization severs all causal couplings with other variables. The critical value for hypothesis test is determined as follows. For each realization the DPF is estimated. The estimates for all realizations are collected and the $100(10\alpha)\%$ quantile (where $\alpha\%$ significance level) is determined. The step-wise procedure is described below.

i. Compute the DFT of given series.

ii. Randomize the phase by replacing the phase with a randomly generated sequence over the interval $[0, 2\pi]$.

iii. Obtain the surrogate via inverse DFT of the modified Fourier transform.

iv. Repeat steps (i) to (iii) R times so as to generate R surrogate data sets.

The test of zero Granger-causality is then implemented as follows:

i. Estimate the DPF at every frequency point ω for each surrogate realization.

ii. In each $i-j$ $(i \neq j)$ channel, at each frequency, determine the 95% quantile of the DPF estimates across R realizations. Denote this $\zeta_{ij,\alpha}(\omega)$.

iii. Infer that e_j (and hence z_j) does not Granger-cause z_i if $|\psi_{ij}(\omega)|^2 < \zeta_{ij,\alpha}(\omega),\ \forall \omega$.

(d) Write a script to implement the frequency-domain Granger non-causality test based on the above methodology.

(e) Test your scripts on $(N = 2000)$ observations generated from the following process (use `VARMAsim` from the `MTS` package).

$$\mathbf{z}[k] = \begin{pmatrix} 0.3 & 0 & 0 \\ 0.6 & 0.4 & 0 \\ 0 & 0.4 & 0.5 \end{pmatrix} \mathbf{z}[k-1] + \begin{pmatrix} 0.2 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0.3 & 0.4 \end{pmatrix} \mathbf{z}[k-2] + \mathbf{e}[k], \quad \mathbf{e}[k] \sim \mathcal{N}(0, \mathbf{I})$$

The VAR model can be estimated using the `VAR` routine in the `MTS` package or the `vars` routine in the `VARS` package in `R`. Use the SIC (or the SC) option in the latter to determine the "best" VAR model (do not assume known order).

(f) Do both the time- and frequency-domain methods yield the correct result for Granger causality (and non-causality). Do you expect your answer remain invariant with respect to realization?

2. Test the implemented measures in part (1a) and part (1b) on the data given in `proj2017.Rdata` and report the identified causal relationships.