Journal of Neuroscience Methods 214 (2013) 80–90

Contents lists available at SciVerse ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Influence analysis for high-dimensional time series with an application to epileptic seizure onset zone detection

Christoph Flamm[a,*], Andreas Graef[a], Susanne Pirker[b], Christoph Baumgartner[b], Manfred Deistler[a]

[a] Institute for Mathematical Methods in Economics, Vienna University of Technology, Vienna, Austria
[b] General Hospital Hietzing with Neurological Center Rosenhügel and Karl Landsteiner Institute for Clinical Epilepsy Research and Cognitive Neurology, Vienna, Austria

HIGHLIGHTS

- ▶ We present a novel method for the causal analysis of high-dimensional time series.
- ▶ This method combines factor models and Granger causal analysis.
- ▶ An application is the detection of epileptic seizure onset zone based on ECoG data.

ARTICLE INFO

Article history:
Received 9 August 2012
Received in revised form
27 December 2012
Accepted 29 December 2012

Keywords:
Granger causality
Factor model
Principal Component Analysis
EEG
Epilepsy
Onset zone detection

ABSTRACT

Granger causality is a useful concept for studying causal relations in networks. However, numerical problems occur when applying the corresponding methodology to high-dimensional time series showing co-movement, e.g. EEG recordings or economic data. In order to deal with these shortcomings, we propose a novel method for the causal analysis of such multivariate time series based on Granger causality and factor models. We present the theoretical background, successfully assess our methodology with the help of simulated data and show a potential application in EEG analysis of epileptic seizures.

1. Introduction

1.1. Motivation

In many cases the problem of identification of the dependence structure in multivariate time series arises. This is important, for example, in biology and economics, and in particular for neuroscience data, e.g. electroencephalographic (EEG) data, where the connections between brain regions are analyzed.

Investigations of this kind have been conducted in several ways, which include (Formisano et al., 2008; Astolfi et al., 2005; Pereda et al., 2005; Möller et al., 2001; Cassidy and Brown, 2002; Gates et al., 2010).

The focus of this paper will be on the detection of Granger causality in multivariate time series which show strong co-movement, i.e. high correlation between the component-series. In Granger (1969) the causality between two time series is analyzed. The idea of this causality concept is based on predictability: if the knowledge of the past of one time series improves the prediction of a second one, the first is said to be Granger causal for the second. Note that this specific definition is just one possible way among many of defining causality. We refer the interested reader to Bressler and Seth (2011) for background information on Granger causality and to Pearl (2000) for a historical overview of causality.

Multivariate extensions of this causality concept have been developed, for recent references see e.g. Lütkepohl (2007) or Eichler (2007). Besides, various topics of Granger causality have been discussed, see e.g. Sims (1972), Geweke (1982), Dhamala et al. (2008), Barnett and Seth (2011) and Marinazzo et al. (2011). For recent applications in neuroscience see e.g. Guo et al. (2008), Liao et al. (2010), Sommerlade et al. (2012) and Flamm et al. (2012).

In practice we often encounter high-dimensional time series, which show co-movement. As this co-movement normally generates numerical problems, the question arises how to investigate the causality structure of these data.

* Corresponding author at: Institute for Mathematical Methods in Economics, Vienna University of Technology, Argentinierstraße 8/2/105-2, 1040 Vienna, Austria. Tel.: +43 1 58801 105273.
E-mail address: christoph.flamm@tuwien.ac.at (C. Flamm).

0165-0270
http://dx.doi.org/10.1016/j.jneumeth.2012.12.025

For the analysis of highly correlated data, such as EEG data, factor models are a useful tool, see e.g. Molenaar and Nesselroade (2001) and Molenaar (1985). The idea behind factor models is the separation of the observations into latent variables (describing the co-movement) and noise. The latent variables are described by a small number of factors. In this modeling approach, up to now causality was not considered.

In this paper we propose a methodology for the causal analysis of high-dimensional co-moving data, by combining factor models and Granger causality analysis. We will present the theoretical background as well as an application to simulated data and to EEG recordings.

This paper is structured as follows: in order to get a grasp of Granger causality and factor models, we give a short introduction to both topics in the remainder of this section. In Section 2 we apply Granger causality to factor models and discuss the challenges arising. In Section 3 we propose a methodology for this kind of analysis. We apply this methodology to simulated and EEG data and present the results in Section 4. This paper is concluded in Section 5.

1.2. Mathematical background and notation

In this paper we distinguish between two different types of processes. In the classic Granger causal analysis we investigate n-dimensional stochastic processes $(y(t))_{t\in\mathbb{Z}}$ generated by n components, as discussed in this section. In the factor model case we analyze n-dimensional stochastic processes $(x(t))_{t\in\mathbb{Z}}$, whose latent variable process is generated by a small number of components, as discussed in Section 1.5. For notational purposes we simply write y when referring to the whole stochastic process $(y(t))_{t\in\mathbb{Z}}$, this also applies for all other processes.

For the classic Granger causal analysis, we consider an n-dimensional stochastic process $(y(t))_{t\in\mathbb{Z}}$, $y(t):\Omega\to\mathbb{R}^n$, which is weakly stationary with mean zero. We refer to Hannan and Deistler (2012) and Brockwell and Davis (1991) for treatment of time series.

In this paper we only consider linearly regular processes, which admit a Wold representation, see Rozanov (1967) and Hannan (1970). The covariance function of y is given by $\gamma(s)=\mathbb{E}\,y(t+s)y(t)'$. For the remainder of the paper we assume, that $\sum\|\gamma(s)\|<\infty$ holds and that the *spectral density*

$$f(\lambda)=\frac{1}{2\pi}\sum_{s=-\infty}^{\infty}\gamma(s)e^{-i\lambda s} \tag{1}$$

is uniformly bounded above and below, i.e. there exists a real constant c such that[1]

$$c^{-1}I_n\le f_{yy}(\lambda)\le cI_n \quad \text{for all} \quad \lambda\in[-\pi,\pi] \tag{2}$$

holds. According to Wiener and Masani (1957), this assumption yields that $y(t)$ has an $AR(\infty)$ *representation*

$$\sum_{m=0}^{\infty}A(m)y(t-m)=\varepsilon(t) \tag{3}$$

where $A(m)\in\mathbb{R}^{n\times n}$, $\sum_{m=0}^{\infty}\|A(m)\|^2<\infty$ and $A(0)=I_n$ holds. The right-hand side $\varepsilon(t)$ is white noise, i.e. $\mathbb{E}\,\varepsilon(t)=0$, $\mathbb{E}\,\varepsilon(s)\varepsilon(t)^*=\delta_{st}\Sigma$, and Σ denotes its positive definite covariance matrix. We additionally assume that even $\sum_{m=0}^{\infty}\|A(m)\|<\infty$ holds.

We use z to denote the backshift operator on $\mathbb{Z}$: $z(y(t)|t\in\mathbb{Z})=(y(t-1)|t\in\mathbb{Z})$, as well as a complex variable. Using this notation we may rewrite Eq. (3) as

$$a(z)y(t)=\varepsilon(t), \tag{4}$$

where $a(z)=\sum_{m=0}^{\infty}A(m)z^m$ exists inside and on the unit circle.

We additionally assume that the *stability condition* $\det a(z)\neq 0$ for $|z|\le 1$ holds.

By using $\tilde{a}(z)=-\sum_{m=1}^{\infty}A(m)z^m$ we rewrite (4) as

$$y(t)=\tilde{a}(z)y(t)+\varepsilon(t). \tag{5}$$

The *transfer function* $k(z)=a^{-1}(z)=\sum_{m=0}^{\infty}K(m)z^m$ exists inside and on the unit circle. There is a unique weakly stationary solution of (3) of the form

$$y(t)=\sum_{m=0}^{\infty}K(m)\varepsilon(t-m)=k(z)\varepsilon(t) \tag{6}$$

This solution (6) of the system (3) is called an *autoregressive* (∞) *process*. It corresponds to the Wold representation. For the sake of simplicity of notation we will skip the (∞) sign henceforth.

For a stationary process z, let $\overline{z(t)}=closure(span(z(s)|s\le t))$ denote the space spanned by the past and present of z in the Hilbert space of all square integrable random variables. Time t represents the present unless noted otherwise.

Note that, if (2) holds for the whole process, it also holds for all sub-processes, and therefore all subprocesses have AR representations.

Due to the nature of our application, we will often refer to their components as *channels* in this paper.

1.3. Granger causality

There have been long and thorough discussions about causality throughout the last decades, a brief summary can be found in Pearl (2000). As already stated various ideas exist how to formalize causality. The definition we will use for our causal investigation is Granger causality, as introduced in Granger (1969), based on a suggestion in Wiener (1956).

According to the original definition in Granger (1969), we say a time series y_1 is causing another time series y_2, denoted by $y_1\to y_2$, if we are able to predict y_2 better using all available information in the universe than using all information apart from y_1.

Granger's definition is based on the decrease of the variance of the (linear least squares) prediction error. For a better understanding we present an equivalent definition of Granger causality based on the autoregressive coefficients for the bivariate case.

1.3.1. Definition of bivariate Granger non-causality

Let $y(t)=(y_1(t), y_2(t))'$ satisfy the assumptions of Section 1.2, then we consider the joint AR representation (5) at time point $t+1$.

$$\begin{pmatrix} y_1(t+1) \\ y_2(t+1) \end{pmatrix} = \underbrace{\tilde{a}(z)\begin{pmatrix} y_1(t+1) \\ y_2(t+1) \end{pmatrix}} + \begin{pmatrix} \varepsilon_1(t+1) \\ \varepsilon_2(t+1) \end{pmatrix} \tag{7}$$

$$\begin{pmatrix} \sum_{m=1}^{\infty}A_{11}(m)y_1(t+1-m)+\sum_{m=1}^{\infty}A_{12}(m)y_2(t+1-m) \\ \sum_{m=1}^{\infty}A_{21}(m)y_1(t+1-m)+\sum_{m=1}^{\infty}A_{22}(m)y_2(t+1-m) \end{pmatrix}.$$

We say that y_1 *is Granger non-causal for* y_2 if $A_{21}(m)=0\;\forall\, m$ (i.e. $\tilde{a}_{21}(z)=0$). In other words $\overline{y_1(t)}$ does not influence the prediction of $y_2(t+1)$.

Otherwise we say y_1 *is Granger causal for* y_2. In this case the knowledge of the present and past of y_1 improves the prediction of $y_2(t+1)$, i.e. the variance of the prediction error is smaller when using the past and present of both y_1 and y_2 compared to using only the past and present of y_2 itself.

[1] In this context $A<B$ means that $B-A$ is a positive definite matrix.