Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnLine 49-1 (2016) 077–082

Estimation of network connectivity strengths in linear causal dynamic systems

Sudhakar Kathari * **Arun K. Tangirala** *

* *Indian Institute of Technology Madras, Chennai-600 036, INDIA (e-mail: arunkt@iitm.ac.in)*

Abstract: Identification of network structure and quantifying the connectivity strengths in multivariate systems is an important problem in many scientific areas. Data-driven approach to network reconstruction based on causality measures is an emerging field of research in this respect. Among several recently introduced data-driven causality measures, the partial directed coherence (PDC) and direct power transfer (DPT) have been shown to be very effective for linear systems. While the PDC is useful in reconstructing the network, DPT has been proved to be effective in both identifying the network structure as well as quantifying the strength of connectivity. In this work, we study the problem of obtaining efficient estimates of network connectivity strengths, which has hitherto not been addressed in the literature. To this end, we study two different estimation methods for network connectivity strengths and demonstrate that the goodness of estimates depends on nature of the data generating process (DGP). In order to characterize the multivariate DGP, we introduce two statistics, namely, the vector-valued autocorrelation function (VACF) and the vector-valued partial autocorrelation function (VPACF), and estimators of the same. Our studies show that the parametric models used in estimating connectivity strengths should be commensurate with the dynamics of the process as characterized by the newly introduced VACF and VPACF. Simulation studies are presented under different scenarios to support our findings and the newly introduced measures.

Keywords: VACF, VPACF, causality, direct power transfer, strength of connectivity

1. INTRODUCTION

Identification of network structures in multivariate systems from measurements is an important problem in many areas such as engineering, systems biology, econometrics, statistics, sociology and climatology etc (Granger, 1969; Baccala and Sameshima, 2001; Gigi and Tangirala, 2012). The main objective in network reconstruction is to identify causal interactions between the variables from given time series data. The problem of identifying the causal relationships from the measurements was initially addressed by Wiener (1956). Granger (1969) proposed the definition of causality, based on Wiener's ideas, which is known today as Granger causality. The concept of Granger causality is based on prediction. A variety of data-driven causality measures that work either in the time- or frequency-domain have been proposed for over four decades now (Hlavackova-Schindler et al., 2007; Winterhalder et al., 2005). A majority of these measures rely on the concept of Granger causality. Among the *frequency-domain measures*, the partial directed coherence (PDC) and direct power transfer (DPT) are well-suited for structure determination as they measure direct influences between the variables (Baccala and Sameshima, 2001; Gigi and Tangirala, 2010). Both the methods use parametric time-series models, namely, the *vector auto-regressive* (VAR) models (see Appendix A) as the primary vehicles, regardless of the underlying process. The choice of this structure is motivated primarily by the ease of estimation. VAR models give rise to linear predictors thereby admitting least squares estimators to provide unique solutions. On the other hand, vector moving average (VMA) models yield predictors that are non-linear functions of parameters and are therefore, more complicated to estimate (Lutkepohl, 2005).

An additional problem of interest in network reconstruction is the determination of connectivity strengths. The knowledge of network connectivity strengths is valuable in several applications. A common use of the connectivity strength is in determining the strongest and weakest links, which finds applications in fault diagnosis, control of networks, etc. Despite its importance, relatively little effort has gone into defining and estimating strengths of connectivities. An ad hoc definition and computation is provided by Baccala and Sameshima (2001) based on the PDC. The definition therein lacks a transparent connection with any statistical relationship between the variables. On the other hand, the DPT-based definition introduced by Gigi and Tangirala (2010) directly quantifies the "amount" of transfer of power (or variability) from the source to the sink variable. The key step is the decomposition of the total power into direct, indirect and interference terms at each frequency (Gigi and Tangirala, 2010). Subsequently, the connectivity strength is derived as the normalized DPT between two variables (see §2.2). However, obtaining efficient estimates of this strength of connectivity has neither been addressed nor studied in the literature.

The objective of this work is to develop / identify a suitable method for obtaining unbiased and efficient estimates of connectivity strengths in linear, causal dynamical systems,

10.1016/j.ifacol.2016.03.032

or jointly stationary linear processes, based on the definition given in Gigi and Tangirala (2010). The specific questions of interest are:

(1) Does blindly adapting a VAR structure, without paying attention to the process characteristics, lead to inefficient estimates of connectivity strengths?
(2) Is it possible to devise practically useful measures that provide insights into the underlying multivariate data generating process?
(3) Does estimation of strength of connectivity by fitting a model that is commensurate with the process characteristics always give rise to efficient estimates in all the cases?

The first question assumes prominence in light of the fact that VAR models are widely used in the network reconstruction of linear causal dynamical systems. While this choice of model structure might serve the purpose of reconstruction, it may not be appropriate for obtaining efficient estimates of connectivity strengths, particularly when the underlying mechanism of the DGP is in significant deviation from the VAR structure. The second issue of interest is a natural follow-up of the first one. Since it is suspected that VAR model structures can lead to biased and / or inefficient estimates of strengths, it is necessary to be equipped with measures that provide insights into the nature of the DGP. The measures that exist in the multivariate time-series literature, namely, the matrix auto-correlation function (ACF) and its partial counterpart (Wei, 2005) do not, unfortunately, present sufficient insights into the collective characteristics of the multivariate process (see Appendix B). They are, however, ideally suited for obtaining a knowledge of the individual channel characteristics. Finally, the third issue of interest, perhaps being the practically most relevant one, is also a question that demands broader and comprehensive study. Therefore, for present, we restrict our scope of study to the first two issues.

The connectivity strength, as defined in (3), is a non-linear rational function of the model coefficients. Therefore, they do not lend themselves easily to a theoretical analysis of the bias and variance, i.e., it is difficult to theoretically derive the distributional characteristics of the estimates. Therefore, we adopt a Monte-Carlo simulation approach in this work. In addition, we introduce two new statistics, namely, the vector-valued ACF and partial ACF, for characterizing the correlation structure in multivariate stationary processes. These measures are useful in providing a collective picture of the correlation and can be thought of as multivariate analogues of the ACF and partial ACF (PACF) for the univariate process (Shumway and Stoffer, 2000; Tangirala, 2014). They possess similar properties as that of the univariate versions and aid in determining the order of VMA and VAR models, respectively. Further, we provide expressions for estimating these functions and study their distributional characteristics through Monte-Carlo simulations.

One of the main findings of this work is that VAR models are only suited for efficient estimation of connectivity strengths when the DGP is also of the VAR type. Any deviation, for e.g., when the DGP is VMA type, the use of VAR models result in biased and / or inefficient estimates, even when the VAR model order is chosen appropriately, i.e., the resulting model passes all necessary diagnostic tests. This finding, while prima facie, may be unsurprising, is also interesting since the connectivity strengths are defined in terms of the VAR model. On the other hand, when the model structure is chosen in accordance with the process characteristics, as determined by the vector-valued ACF / PACF, one obtains unbiased and relatively efficient estimates. A broader message of this work is that it is advisable to choose a model that is commensurate with the data generating mechanism rather than always choosing a VAR model for network reconstruction, which is the general practice. Further, the vector-valued ACF and PACF is potentially useful in developing multivariate time-series models for other applications as well.

The paper is organized as follows. Section 2 reviews vector time series models and quantification of connectivity strengths along with the estimation techniques. In Section 3, we introduce the vector-valued ACF and PACF and their sample versions with three illustrative examples. Simulation case studies are presented in Section 4. The paper ends with a few concluding remarks in Section 5.

2. PRELIMINARIES

This section reviews the quantification of connectivity strengths and their estimation methods based on parametric vector time-series modeling, a brief overview of which is provided in Appendix A.

2.1 Quantification of power transfers

The quantification of direct and indirect influences in terms of power transfer is obtained by developing direct and indirect transfer functions based on the signal flow graph representation of the process (Gigi and Tangirala, 2010). The mathematical expression for direct power transfer function ($h_{D,ij}(\omega)$)from source x_j to sink x_i is given as,

$$h_{D,ij}(\omega) = \begin{cases} \frac{-\bar{a}_{ij}(\omega)\det(\bar{\mathbf{M}}_{ij})}{\det(\bar{\mathbf{A}}(\omega))}, & i \neq j \\ \frac{\det(\bar{\mathbf{M}}_{ij})}{\det(\bar{\mathbf{A}}(\omega))}, & i = j \end{cases} \tag{1}$$

where $\bar{\mathbf{M}}_{ij}(\omega)$ is the minor matrix of the matrix $\bar{\mathbf{A}}(\omega)$, which is obtained from $\bar{\mathbf{A}}(\omega)$ by eliminating both i^{th} and j^{th} row and column.

The squared magnitudes of direct power transfer (DPT) from source x_j to sink x_i is,

$$|\psi_{ij}(\omega)|^2 = |h_{D,ij}(\omega)|^2 \tag{2}$$

The DPT gives both the structural information and the strength of connectivities of the process.

2.2 Strength of connectivities

The strength of connectivity is quantified based on the direct power transfer between the variables (Gigi and Tangirala, 2010). For a link connecting source x_j to sink x_i, the normalized connectivity strength is defined as (Gigi and Tangirala, 2012),

$$\beta_{ij} = \frac{\int_0^{\pi} |\psi_{ij}(\omega)|^2 d\omega}{\int_0^{\pi} |\psi_{jj}(\omega)|^2 d\omega} \tag{3}$$