EXPLORATIONS OF TEMPORAL CAUSALITY USING PARTIAL COHERENCE

Louis Scharf [1] *and Yuan Wang* [2]

[1]Departments of Mathematics and Statistics, Colorado State University, Fort Collins, CO, USA
[2]Department of Mathematics and Statistics, Washington State University, Pullman, WA, USA
scharf@colostate.edu, ywang@math.wsu.edu

ABSTRACT

In this paper we explore partial coherence as a tool for evaluating the causal, anti-causal, or mixed-causal dependence of one time series on another. The key idea is to establish a connection between questions of causality and partial coherence. Once this connection is established, then a scale-invariant partial coherence statistic is used to resolve the question of temporal causality. This coherence statistic is shown to be a likelihood ratio. It may be computed from a composite covariance matrix or from its inverse, the information matrix. Numerical experiments demonstrate the application of partial coherence to the resolution of temporal causality.

Index Terms— partial correlation, partial coherence, information matrix, causality, time series

1. INTRODUCTION

Granger causality [1] has been widely used for addressing questions of causality. In general, a time series $\{y_n\}$ is causally dependent on another time series $\{x_n\}$ if the past of $\{x_n\}$ contains unique information about $\{y_n\}$. Geweke [3] has studied the linear dependence and feedback between two time series and extended in [4] to the conditional linear dependent under presence of a third time series. Barnett et al. [5, 6] have adapted Granger's point of view to the study of causality in state-space models. In practice, causality can be evaluated with respect to the available information. Let $\{J_n\}$ be the set of available information at time n. Then, the time series $\{x_n\}$ is said to be a prima facie cause of $\{y_n\}$ with respect to $\{J_n\}$ if the knowledge about $\{y_n\}$ is increased by adding $\{x_n\}$ into the information set $\{J_n\}$.

In this paper, we are interested in assessing temporal causality between two scalar time series, using partial correlation or partial coherence. Partial correlation coefficients measure correlation between two estimator errors for random variables that have been regressed onto a third set of variables. The partial correlation between random variables x and y, given $\mathbf{z}$, is the correlation between the residuals for x regressed on $\mathbf{z}$ and residuals for y regressed on $\mathbf{z}$, where $\mathbf{z}$ is a third set of random variables. A zero partial correlation indicates that x is conditionally independent of y given $\mathbf{z}$. Thus, x does not contain unique information about y in addition to $\mathbf{z}$. In this paper, we will define the prima facie evidence $\mathbf{z}$ in such a way that the question of causality can be resolved from an analysis of partial coherence. We show that a partial coherence statistic is a scale-invariant measure of causality that has a connection to likelihood testing of hypotheses.

In section 2, we introduce partial coherence and establish its connection to likelihood. Moreover, we show in Section 2.2 assessment of causal relation between two time series using the partial coherence statistics. In section 3, we conduct several experiments to explore the causal, anti-causal, and mixed-causal relationship between time series. Section 4 discusses the connection between this paper and prior work of Geweke [3, 4] and by Barnett, et al. [5, 6]

Yuan: We need to reference Geweke's likelihood result, if it is the same or similar to ours.

Notation: The superscripts $(\cdot)^T$ and $(\cdot)^H$ denote transpose and Hermitian, respectively. $\mathbf{I}_M$ is the identity matrix of dimensions $M \times M$, and $\mathbf{0}$ denotes either a column vector with M zeros, or the zero matrix of appropriate dimensions (the difference should be clear from the context). Bold upper case letters denote matrices, boldface lower case letters denote column vectors, and italics denote scalars. We use $\mathbf{A}^{1/2}$ ($\mathbf{A}^{-1/2}$) to denote the square root matrix of the Hermitian matrix $\mathbf{A}$ ($\mathbf{A}^{-1}$); $\text{blkdiag}[\mathbf{A}_1, \ldots, \mathbf{A}_k)]$ is a block-diagonal matrix whose diagonal blocks are $\mathbf{A}_1, \ldots, \mathbf{A}_k$. $\mathbf{x} \sim \mathcal{CN}_M(\mathbf{0}, \mathbf{R})$ indicates that $\mathbf{x}$ is an M-dimensional complex circular Gaussian random vector of zero mean and covariance $\mathbf{R}$.

2. PARTIAL COHERENCE AND CAUSALITY

2.1. Partial Coherence

Begin with the random vectors $(x, y, \mathbf{z})$. Define $\mathbf{u} = (x, y)$. The most elementary filtering question we shall ask is to estimate each of x and y from $\mathbf{z}$. The linear least squares estimate of x is $\hat{x} = \mathbf{W}_{x|\mathbf{z}}\mathbf{z}$, where $\mathbf{W}_{x|\mathbf{z}} = \mathbf{R}_{x\mathbf{z}}\mathbf{R}_{\mathbf{zz}}^{-1}$, and the linear least squares estimator of y is $\hat{y} = \mathbf{W}_{y|\mathbf{z}}\mathbf{z}$, where $\mathbf{W}_{y|\mathbf{z}} = \mathbf{R}_{y\mathbf{z}}\mathbf{R}_{\mathbf{zz}}^{-1}$. The composite error covariance matrix

This work was supported in part by the National Science Foundation under contract CCF-1712788.

for these estimators is

$$\mathbf{Q}_{\mathbf{uu}|\mathbf{z}} = E\begin{bmatrix} x-\hat{x}(\mathbf{z}) \\ y-\hat{y}(\mathbf{z}) \end{bmatrix}\begin{bmatrix} (x-\hat{x}(\mathbf{z}))^H & (y-\hat{y}(\mathbf{z}))^H \end{bmatrix}$$
$$= \begin{bmatrix} Q_{xx|\mathbf{z}} & S_{xy|\mathbf{z}} \\ S^H_{xy|\mathbf{z}} & Q_{yy|\mathbf{z}} \end{bmatrix} \quad (1)$$

where $Q_{xx|\mathbf{z}}$ is the eror covariance matrix for estimating x from $\mathbf{z}$ only, and $Q_{yy|\mathbf{z}}$ is the error covariance matrix for estimating y for $\mathbf{z}$ only; the *partial correlation* $S_{xy|\mathbf{z}} = E\left[(x-\hat{x})(y-\hat{y})^H\right]$ is the cross covariance between the two errors. The partial correlation is zero if x and y are conditionally linear independent, i.e., x does not provide any new information about y in addition to $\mathbf{z}$.

The normalized error covariance matrix is called the partial coherence matrix:

$$\mathbf{C}_{xy|\mathbf{z}} = \text{blkdiag}\,[Q^{-1/2}_{xx|\mathbf{z}}, Q^{-1/2}_{yy|\mathbf{z}}]\mathbf{Q}_{\mathbf{uu}|\mathbf{z}}\,\text{blkdiag}[Q^{-1/2}_{xx|\mathbf{z}}, Q^{-1/2}_{yy|\mathbf{z}}]$$
$$= \begin{bmatrix} 1 & Q^{-1/2}_{xx|\mathbf{z}} S_{xy|\mathbf{z}} Q^{-1/2}_{yy|\mathbf{z}} \\ Q^{-1/2}_{xx|\mathbf{z}} S^H_{xy|\mathbf{z}} Q^{-1/2}_{yy|\mathbf{z}} & 1 \end{bmatrix}$$

The NE term $Q^{-1/2}_{xx|\mathbf{z}} S_{xy|\mathbf{z}} Q^{-1/2}_{yy|\mathbf{z}}$ is the partial coherence between x and y given $\mathbf{z}$. The determinant of the partial coherence matrix $\mathbf{C}_{xy|\mathbf{z}}$ is a Hadamard ratio, and it determines the coherence statistic we propose for testing the hypothesis that $S_{xy|\mathbf{z}}$ is zero:

$$\rho^2 = 1 - \det[\mathbf{C}_{xy|\mathbf{z}}]$$

Yuan: Ck this.

This statistic is invariant to scaling of x and y and to nonsingular transformation of $\mathbf{z}$. Its null distribution is known, making it suitable for testing. If the covariance $\mathbf{R}$ are computed from M spherically-contoured realizations of the random variables x, y, and $\mathbf{z}$, then the null distribution of ρ^2 is Beta $[2, 2(M-N+1)]$ where N is the dimension of vector $\mathbf{z}$.

Importantly, iff $\rho^2 = 0$, the linear minimum mean-squared error estimator of x from y and $\mathbf{z}$, does not use y. In this case, we say x is conditionally linearly independent of y. If y and $\mathbf{z}$ are constructed properly, this fact may be used to test for causality between time series, space series or nodes in a network.

Connection to likelihood There is a connection to likelihood. Assume measurements are multivariate normal with distribution $\mathcal{CN}_L[\mathbf{0}, \mathbf{R}]$, where L is the $L = p+q+r$, with p, q, r the respective dimensions of $x, y, \mathbf{z}$. The covariance matrix $\mathbf{R}$ is the composite covariance matrix of the three vector-value variables. The data matrix $\mathbf{V} \in \mathcal{C}^{L\times M}, M \geq L$ is a random draw from the composite vector $(x, y, \mathbf{z}^T)^T$. Under the null hypothesis H_0 that x and y are conditionally independent given $\mathbf{z}$, the MLE for the covariance matrix is denoted $\mathbf{R}_0$, and under the alternative hypothesis H_1, the MLE for the covariance matrix is denoted $\mathbf{R}_1$. The likelihood ratio for testing H_1 vs H_0 is

$$\Lambda = \frac{\det[\mathbf{R}_1]}{\det[\mathbf{R}_0]} \quad (2)$$

The determinants of the $\mathbf{R}_i$ may be written as

$$\det[\mathbf{R}_1] = \det\left[\mathbf{Q}^1_{\mathbf{uu}|\mathbf{z}}\right]\det[\mathbf{R}_{\mathbf{zz}}]$$
$$\det[\mathbf{R}_0] = \det\left[\mathbf{Q}^0_{\mathbf{uu}|\mathbf{z}}\right]\det[\mathbf{R}_{\mathbf{zz}}]$$

where $\mathbf{Q}^i_{\mathbf{uu}|\mathbf{z}}$ denotes the error covariance matrix for estimating the pair $\mathbf{u} = (x, y)^T$ from $\mathbf{z}$, under hypothesis H_i. This error covariance matrix is

$$\mathbf{Q}_{\mathbf{uu}|\mathbf{z}} = E\left[(\mathbf{u}-\hat{\mathbf{u}})(\mathbf{u}-\hat{\mathbf{u}})^H\right] = \begin{bmatrix} Q_{xx|\mathbf{z}} & S_{xy|\mathbf{z}} \\ S^H_{xy|\mathbf{z}} & Q_{yy|\mathbf{z}} \end{bmatrix}.$$

So under the null hypothesis that x and y are conditionally independent, the NE matrix $S_{xy|\mathbf{z}}$ is zero. Therefore, $\det[\mathbf{Q}^0_{\mathbf{uu}|\mathbf{z}}] = \det[Q_{xx|\mathbf{z}}]\det[Q_{yy|\mathbf{z}}]$. Under the alternative hypothesis, $\det[\mathbf{Q}^1_{\mathbf{uu}|\mathbf{z}}] = \det[\mathbf{Q}_{\mathbf{uu}|\mathbf{z}}]$ given in (1). Thus,

$$\Lambda = \frac{\det[\mathbf{Q}^1_{\mathbf{uu}|\mathbf{z}}]}{\det[\mathbf{Q}^0_{\mathbf{uu}|\mathbf{z}}]} = \det[\mathbf{C}_{xy|\mathbf{z}}] = 1-\rho^2$$

With no further modeling of the covariance matrices $\mathbf{R}_i$, other than these patterns, the generalized likelihood ratio for testing H_1 vs H_0 replaces these error covariance matrices by their sample-data estimates.

The information matrix. The inverse of the covariance matrix may be written

$$\mathbf{P} = \mathbf{R}^{-1} = \begin{bmatrix} \mathbf{Q}^{-1}_{\mathbf{uu}|\mathbf{z}} & \mathbf{Q}^{-1}_{\mathbf{uu}|\mathbf{z}}\mathbf{W}_{\mathbf{u}|\mathbf{z}} \\ \mathbf{W}^H_{\mathbf{u}|\mathbf{z}}\mathbf{Q}^{-1}_{\mathbf{uu}|\mathbf{z}} & \mathbf{R}^{-1}_{\mathbf{zz}} + \mathbf{W}^H_{\mathbf{u}|\mathbf{z}}\mathbf{Q}^{-1}_{\mathbf{uu}|\mathbf{z}}\mathbf{W}_{\mathbf{u}|\mathbf{z}} \end{bmatrix}$$

So from the inverse of the sample covariance matrix, the matrix $\mathbf{Q}^{-1}_{\mathbf{uu}|\mathbf{z}}$ may be read out of the NW corner, and used to construct the statistic ρ^2.

2.2. Causality and Partial Coherence

Let's suppose $\{x_n\}$ and $\{y_n\}$ are two time series. We are interested in if $\{y_n\}$ is causally dependent on $\{x_n\}$. Granger Causality is popularly used for the causal relation between stochastic processes. A time series $\{x_n\}$ is said to be the prima facie cause for $\{y_n\}$ with respect to $\mathbf{z}$ if

$$F(y_{n+1} \mid \mathbf{z}, x_n, x_{n-1}, \ldots,) \neq F(y_{n+1} \mid \mathbf{z}). \quad (3)$$

Here $F(\cdot)$ is the probabilistic distribution. The inequality means the conditional distribution of y_{n+1} is dependent on the past of $\{x_n\}$ in addition to $\mathbf{z}$. Here $\mathbf{z}$ is the set of information from outside $\{x_n\}$ available at time n.

The definition (3) is broadly applicable to linear or nonlinear systems. The causal effect is often estimated through a linear/non-linear multivariate regression model involving a large number of coefficients. By fitting two models with and without $\{x_n\}$, a conditional Granger causality index (CGCI) is used to quantify the Granger causality [7]. Model selection or dimension reduction methods have to be implemented for choosing a reasonable regression model involving a small number of lagged variables of the time series.

Our trick will be to define a third channel $\{\mathbf{z}_n\}$ of prima facie evidence in such a way that the question of causality can be resolved from an analysis of partial coherence. The idea is illustrated in Fig. 1. Let s and t be two time points, we define the random variables $y = y[s]$ and $x = x[t]$. The random vector $\mathbf{z}$ is defined as the available information set at time s. For example, $\mathbf{z}_s = \{x_n, n \leq s, n \neq t\}$, i.e., the past of the time series $\{x_n\}$ at time s. Note that we have $x[t]$ excluded by the prima facie evidence $\mathbf{z}$ if $t < s$ such that x remains random given $\mathbf{z}$. With $x, y, \mathbf{z}$ defined, we can compute the partial coherence statistic $\rho^2(s,t)$ as introduced in Section 2.1. The values of $\rho^2(s,t)$ vs the pair (s,t) reveals causality or non-causality.

The partial coherence $\rho^2(s,t)$ measures the amount of unique information about y_s that contained by x_t. When the time series $\{y_n\}$ is not causally dependent on $\{x_n\}$, the partial coherence is zero for any pair of (s,t) with $t < s$. We also compute the value of $\rho^2(s,t)$ for $t \geq s$ to help with investigating other structures such as anti-causality and mixed structure.

This approach is applicable to the multivariate time series. When $x_t \in \mathbb{C}^p$ and $y_s \in \mathbb{C}^q$, we can derive the $p \times q$ partial coherence matrix where the $(i,j)th$ element give the partial coherence between the ith component of x_t and the jth component of y_s.

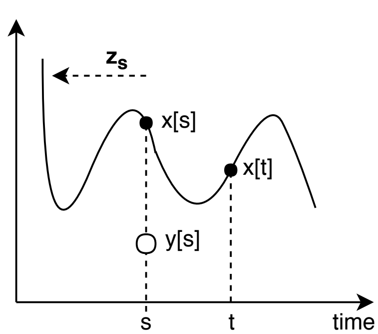

Fig. 1. The time series

3. EXPERIMENTS

In this section, we will conduct several experiment and use the partial coherence to assess the causality between two time series. We consider the following linear system, as shown in Fig. 2,

$$x_n = \sum_{k=-\infty}^{+\infty} h_k u_{n-k}$$
$$y_n = \sum_{k=-\infty}^{+\infty} g_k v_{n-k} + \sum_{k=-\infty}^{+\infty} f_k x_{n-k},$$

where u, v are unit variance white noises. It is easy to show the covariance coefficients are determined by the filter coefficients as

$$r_m^{xx} = E[x_n x_{n+m}] = \sum_k h_k h_{k+m}$$
$$r_m^{yy} = E[y_n y_{n+m}] = \sum_{k,l} f_k f_l r_{m+k-l}^{xx} + \sum_k g_k g_{k+m}$$
$$r_m^{xy} = E[x_n y_{n+m}] = \sum_k f_k r_{m-k}^{xx}$$

As a result, the partial coherence can be calculated analytically as introduced in Section 2.1. We set the filter coefficients for the noises as $h_k = h_0 a^k$ and $g_k = g_0 b^k$ for $k \geq 0$ where $h_0 = 0.8$, $a = 0.9$, $g_0 = 0.7$, and $b = 0.95$. The coefficients filtering $\{x_n\}$, i.e., $\{g_k\}$, is the key component for the relationship between $\{x_n\}$ and $\{y_n\}$. We consider three different scenarios.

Case 1: $\{y_n\}$ is causally dependent on $\{x_n\}$:

$$y_n = x_n + 0.5x_{n-1} + 0.4x_{n-2} + 0.3x_{n-3} + 0.2x_{n-4} + 0.2x_{n-5} + \sum_{k=-\infty}^{+\infty} g_k v_{n-k}$$

Case 2: $\{y_n\}$ is anti-causally dependent on $\{x_n\}$:

$$y_n = x_n + 0.5x_{n+1} + 0.4x_{n+2} + 0.3x_{n+3} + 0.2x_{n+4} + 0.2x_{n+5} + \sum_{k=-\infty}^{+\infty} g_k v_{n-k}$$

Case 3: the mixed case:

$$y_n = x_n + 0.5x_{n-1} + 0.4x_{n-2} + 0.5x_{n+1} + 0.4x_{n+2} + \sum_{k=-\infty}^{+\infty} g_k v_{n-k}$$

The resulted partial coherence maps are illustrated in Fig. 2. For any given time points s and t, we show that value of the partial coherence statistic $\rho^2(s,t)$ for y_s and x_t. Here $\mathbf{z} := \{x_s, x_{s-1}, \ldots,\}$, i.e., the past of $\{x_n\}$ at time s. The top panel illustrates the partial coherence when the time series $\{y_n\}$ is causally dependent on $\{x_n\}$. It can be seen that $\rho^2(s,t) \neq 0$ for $t < s$, meaning that x_t contains some unique information about y_s in addition to $\mathbf{z}$. Moreover, $\rho^2(s,t) = 0$ for any $t > s$, which suggests that the future value of x does not contain unique information about y_s given $\mathbf{z}$. The middle panel of Fig. 2 is the results for Case 2. In this case, $\rho^2(s,t) = 0$ for any $t < s$ suggesting the past of x contains no information about y under this anti-causal model. On the other hand, $\rho^2(s,t)$ is nonzero for $t \geq s$. Case 3, shown in the bottom panel, is the mixed case. The nonzero pattern of $\rho^2(s,t)$ for both $t > s$ and $t < s$ demonstrates the mixed relation between the two series.

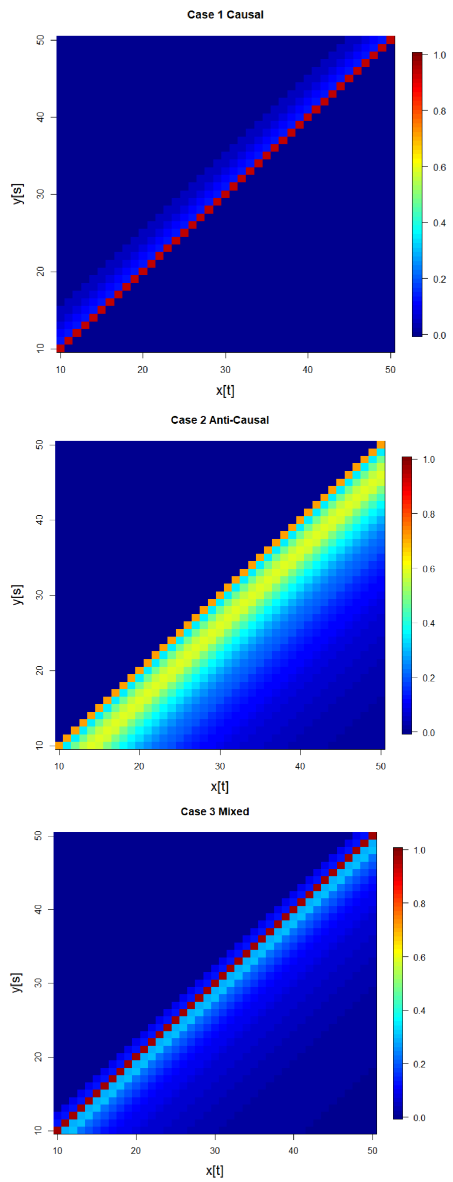

Fig. 2. The partial coherence matrices for the three cases: causal (top), anti-causal (middle), and mixed (bottom).

4. RAPPROCHEMENT WITH THE WORK OF GEWEKE AND BARRETT, ET AL.

There is flexibility in the choice of variables in the vectors $(x, y, \mathbf{z})$. To demonstrate, let us consider these two-channel problems consisting of the time series $x = \{x_t\}$ and the time series $y = \{y_t\}$.

Example 1: $(x = x_t, t > s, y, \mathbf{z} = \{x_u, u \leq s\})$**.** The time series value $y = y_s, t < s$, is to be predicted from its own past, $\mathbf{z} = \{y_u, u \leq s\}$, and the question is whether this prediction can improved with the use of the past of an ancillary time series x, namely $\{x_u, u \leq s\}$. This is the prediction question considered by Geweke [3, 4] and by Barnett, et al. [5, 6]. In this problem the prima facie evidence is the past of x and the past of y. We have compared the results carefully and it turns out that their test statistic

$$\log \frac{\det(\mathbf{\Sigma}_{11}^R)}{\det(\mathbf{\Sigma}_{11})} = \log(1 - \rho^2).$$

Example 2. The value $y = y_s$ of the y-series is to be predicted by the past of the $\mathbf{z}$-series, namely $\{z_u, u \leq s\}$, and the question is whether $x = z_t, t > s$ improves the estimate. This is the estimation problem considered in this paper. The prima facie evidence is the past of $\mathbf{z}$ and a future value of it.

5. DISCUSSION

In this paper, our objective has been to explore the use of partial coherence as a scale-invariant statistic to assess the temporal causality between two time series. Of course these time series might well be space series, in which case the idea of anti- or mixed-causality is a physically meaningful question. The partial coherence statistic we argue for is grounded in likelihood theory, and it may be computed from the information matrix. The key idea in the application of partial coherence to questions of causality is to define three channels of measurements so that they code for the causality question of interest. We have demonstrated this for one particular question of temporal causality.

6. REFERENCES

[1] C. W. J. Granger, "Investigating Causal Relations by Econometric Models and Cross-spectral Methods," *Econometrica*, 37 (3): 424–438, 1969.

[2] C. W. J. Granger, "Testing for Causality a Personal View," *Journal of Economic Dynamics and Control*, 2: 329–352, 1980.

[3] J. Geweke " Measurement of linear dependence and feedback between multiple time series," *J. Am. Stat. Assoc.* 77(378):304–313, 1982.

[4] J. Geweke "Measures of conditional linear dependence and feedback between time series," *J. Am. Stat. Assoc.*,79(388):907–915, 1984.

[5] L. Barnett and A. K. Seth, "Granger causality for state-space models," *Physical Review E* 91:040101(R), 2015

[6]

[7] E. Siggiridou and D. Kugiumtzis, "Granger Causality in Multivariate Time Series Using a Time-Ordered Restricted Vector Autoregressive Model," *IEEE Transactions on Signal Processing*, 64(7):1759–1773, 2016.

[8] J. Whittaker, "Graphical Models in Applied Multivariate Statistics," *Wiley*, 2009.