A Note on Inferring Acyclic Network Structures Using Granger Causality Tests

Radhakrishnan Nagarajan, University of Arkansas for Medical Sciences

Recommended Citation:
DOI: 10.2202/1557-4679.1119
A Note on Inferring Acyclic Network Structures Using Granger Causality Tests

Radhakrishnan Nagarajan

Abstract

Granger causality (GC) and its extension have been used widely to infer causal relationships from multivariate time series generated from biological systems. GC is ideally suited for causal inference in bivariate vector autoregressive process (VAR). A zero magnitude of the upper or lower off-diagonal element(s) in a bivariate VAR is indicative of lack of causal relationship in that direction resulting in true acyclic structures. However, in experimental settings, statistical tests, such as F-test that rely on the ratio of the mean-squared forecast errors, are used to infer significant GC relationships. The present study investigates acyclic approximations within the context of bi-directional two-gene network motifs modeled as bivariate VAR. The fine interplay between the model parameters in the bivariate VAR, namely: (i) transcriptional noise variance, (ii) autoregulatory feedback, and (iii) transcriptional coupling strength that can give rise to discrepancies in the ratio of the mean-squared forecast errors is investigated. Subsequently, their impact on statistical power is investigated using Monte Carlo simulations. More importantly, it is shown that one can arrive at acyclic approximations even for bi-directional networks for suitable choice of process parameters, significance level and sample size. While the results are discussed within the framework of transcriptional network, the analytical treatment provided is generic and likely to have significant impact across distinct paradigms.

KEYWORDS: Granger Causality, VAR process, Gene Networks

Author Notes: Radhakrishnan Nagarajan was partially supported by 5R03LM008853-02 and NIH Grant # P20 RR-16460 from the IDeA Networks of Biomedical Research Excellence (INBRE) Program of the National Center for Research Resources. R. Nagarajan acknowledges Jeffrey Jonkman and the reviewer for useful suggestions.
1. INTRODUCTION

Granger’s seminal work (Granger, 1969) provided a way to infer linear causal relationships between the given processes. Granger causality (GC) relies on forecasting ability and ideally suited to investigate bivariate vector autoregressive processes (VARs) (Hamilton, 1994). Such bivariate VARs can also be represented as a two node network, Fig. 1. Recent studies have used Granger causality (Granger, 1969; Hamilton, 1994) and its extensions for modeling causal relationships from the observed multivariate time series including gene expression data (Brovelli et al, 2004; Seth et al 2006; Mukhopadhyay and Chatterjeee, 2007; Fujita et al., 2007; Guo et al., 2008). It is important to note that causal relationships inferred using GC can be either unidirectional (acyclic) or bi-directional (cyclic). Traditionally, within the context of bivariate VARs (Hamilton, 1994), unidirectional GC or acyclic structure is realized when the upper or lower off-diagonal elements are zero in the bivariate VAR. This ensures absence of forecasting ability in that direction rendering the resulting network acyclic. However, F-test (or $\chi^2$-test) that rely on the ratio of mean-squared forecast errors are used commonly used for establishing statistical significance of the causal relationship (Hamilton, 1994). The present study investigates the impact of the VAR process parameters that can give rise to discrepancies in the ratio of the mean-squared forecast errors along either direction. More importantly, it is shown that such discrepancies can give rise to marked difference in statistical power along either directions resulting in acyclic approximation even for bi-directional networks. While the results are discussed within the framework of transcriptional network motifs, the analytical treatment provided is generic and likely to have significant impact on multivariate time series analysis using GC across distinct paradigms.

Transcriptional network motifs represent recurring elementary network patterns within large complex networks (Milo et al., 2002; Alon, 2007). The present study considers two-gene transcriptional network motifs (Alon 2007; Chickarmane et al., 2006; Kim et al., 2008) Fig. 1, modeled as a stationary bivariate VAR process (Nagarajan and Upreti, 2008). The error terms are assumed to represent transcriptional noise variance. The diagonal terms and the off-diagonal terms are assumed to represent the auto-regulatory feedback strengths and transcriptional coupling strengths respectively. While the present study focuses only on stationary bivariate VARs (2) (Hamilton, 1994), the concerns presented are likely to be aggravated in the presence of non-stationarities. The motifs considered fall under two broad categories (i) bi-directionally coupled two-gene network without auto-regulatory feedback, Fig. 1a and (ii) bi-directionally coupled two-gene network with auto-regulatory feedback of a given order on only one of the nodes, Figs. 1b, 1c. The impact of VAR
process parameters namely the \((i)\) variance of i.i.d. error terms \((ii)\) diagonal terms and \((iii)\) off-diagonal terms on the mean-squared forecast error is investigated in a systematic manner. Subsequently, Monte Carlo simulations are presented that argue in support of discrepancies in power for certain choice of parameters resulting in acyclic approximations of bi-directional networks.

2. METHODS AND RESULTS

Prior to a detailed discussion of Granger causality of the multivariate time series generated from two-node transcriptional motifs, we briefly review classical linear least squares regression by considering a simple example. Consider the regression of \(t_y\) on \((x_t, z_t)\) versus the regression of \(y_t\) on \(z_t\). The model corresponding to the former is given by 

\[
y_t = \beta_0 + \beta_2 z_t + \beta_3 x_t + \epsilon_t, \quad t = 1 \ldots N, \]

where \(\epsilon_t\) is normally distributed. The null hypothesis that \(x_t\) is not a predictor/cause of the response \(y_t\) is given by

\[
H_0 : \beta_2 = 0
\]

The alternative hypothesis is

\[
H_1 : \beta_2 \neq 0
\]

Statistical testing begins by calculating the sum of squared residuals by \((i) s_0 : \) obtained by regression of \(y_t\) with respect to \((x_t, z_t)\) and \((ii) s_1 : \) obtained by regression \(y_t\) with respect to \(z_t\) alone. Under the null hypothesis and normality of the error term \(\epsilon_t\), the ratio

\[
F_{\text{estim}} = \frac{(s_0 - s_1)/1}{s_1 / (N - 3)}
\]

follows a F-distribution. The null hypothesis \(H_0 : \beta_2 = 0\) can be rejected at a given significance level \((\alpha)\) if \(F_{\text{estim}} > F(1, N-3)\). The statistical power under \(H_1\) can be evaluated from the corresponding non-central F-distribution (Rencher, 2000). Granger causality analysis follows a similar regression procedure as above and ideally suited to investigate cause and effect relationship in a bivariate VAR process. However, unlike the above case, the variables of interest can exhibit significant auto-correlation, Sec. 2.2.
2.1 Two-gene transcriptional network motif modeled as a bivariate VAR

A two-gene network motif consisting of \((x_t, y_t)\), Fig.1, can be modeled by a \(p\)th order bivariate VAR given by

\[
\omega_t = k + \Phi_1 \omega_{t-1} + \Phi_2 \omega_{t-2} + ... + \Phi_p \omega_{t-p} + \epsilon_t, \quad \text{..................... (1)}
\]

In (1), the expression of the genes \(\omega_t = [x_t\ y_t]^T, t = 1 \ldots n + p\) is given as linear combination of their past \(\omega_{t-i}, i = 1 \ldots p\). The relative contributions of the past expression values is represented by the \(2 \times 2\) matrices \(\Phi_i, i = 1 \ldots p\). While the diagonal elements of \(\Phi_i, i = 1 \ldots p\) can be thought of as auto-regulatory feedback between the two genes, the off-diagonal elements represent transcriptional coupling strengths. The zero-mean uncorrelated i.i.d Gaussian noise \(\epsilon_t\) with finite non-zero variance represent transcriptional noise, attributed to inherent stochastic mechanisms in transcription (McAdams and Arkin, 1999; Elowitz et al., 2002; Kaern et al., 2005). The extent to which expression of a gene is regulated by the other is modeled by the transcriptional coupling strength, reflected by the off-diagonal terms of \(\Phi_i\) in the bivariate VAR (1). The extent to which a gene regulates its own expression is modeled by the diagonal terms of \(\Phi_i\) in the bivariate VAR (1). For \(\omega_t\) to be covariance-stationary (Hamilton, 1994), the roots \(z_i\) of the reverse-characteristic polynomial given by

\[
|I_{2 \times 2} - \Phi_1 z - \Phi_2 z^2 - ... - \Phi_p z^p| = 0 \quad \text{..................... (2)}
\]

should lie outside the unit circle.

The process parameters \(\Phi_i, i = 1 \ldots p\), in the present study are chosen so as to satisfy the covariance-stationary constraint (2). As noted earlier, the concerns presented in the present study for covariance-stationary process are likely to be aggravated for non-stationary processes.

2.2 Granger Causality

Granger causality represents the forecasting ability between the given processes and may imply true causal relationship between them. In order to determine whether \(y_t\ Granger causes\ x_t\) (i.e. \(y_t \rightarrow x_t\)), the mean-squared forecast error
(s₀) obtained by regression of xᵢ with its own past, i.e. \( xᵢ = k_1 + \sum_{i=1}^{P} aᵢ x_{t-i} + \varepsilonᵢ \) given by

\[
s₀ = \sum_{m=1}^{n} \varepsilon_m^2 \quad \text{..........................} \quad (3)
\]

is statistically compared to the mean-squared forecast error (s₁) obtained by regression of xᵢ with its own past as well as the past of yᵢ, i.e. \( xᵢ = k_2 + \sum_{i=1}^{P} bᵢ x_{t-i} + \sum_{j=1}^{Q} cᵢ y_{t-j} + \etaᵢ \) given by

\[
s₁ = \sum_{m=1}^{n} \eta_m^2 \quad \text{..........................} \quad (4)
\]

Following (Hamilton, 1994, Chapter. 11.2), process yᵢ fails to Granger cause xᵢ if \( s₀ = s₁ \).

In the above expressions (p) represents the auto-regressive lag-length (Hamilton, 1994). A similar approach is used to determine whether xᵢ Granger causes yᵢ. In practical settings, it is common to use F-test (or χ² test) to infer statistically significant Granger causal relationships (Hamilton, 1994). The working principle is enclosed below for completeness.

**F test:**

Determine \( F_{estim} = \frac{(s₀ - s₁)/τ}{s₁/(N-2τ-1)} \) ............................................. (5)

The null \( yᵢ \) does not GC \( xᵢ \) is rejected when

\( F_{estim} > F_{crit} \), where \( F_{crit} = F(τ, N-2τ-1) \) ............................................. (6)

determined from the F-distribution at a given significance level \( α \). It is important to note that expression (5) implicitly depends on the ratio of the forecast error.
errors \( \gamma = \frac{S_0}{S_1} \) which in turn are governed by the process parameters. Discrepancies in \( \gamma = \frac{S_0}{S_1} \) along either direction, can give rise to discrepancies in the magnitude of p-values inferred from F-test (5, 6). This in turn can affect the Type II error, hence the statistical power of the test for a given choice of parameters, sample size and significance level. The present study investigates discrepancies in statistical power resulting in acyclic approximation of bi-directional networks using Monte Carlo simulations. The variation in power as a function of the sample-size and significance level for particular choice of parameters in support of our claims is also presented.

2.3. Acyclic approximations inferred from GC analysis

Acyclic approximations in two-gene network motifs modeled as bivariate VAR is investigated in the present section. The analytical expressions for each case are followed by simulation results.

Case 1: The impact of unequal noise variance in a bivariate VAR with no auto-regulatory feedback, Fig. 1a, is investigated. Analytical expression establishing the role of noise variance that can contribute to discrepancies in the ratio of the mean-squared forecast errors of the bi-directional network (Fig. 1a) is provided. Variation in statistical power as a function of significance level and sample-size for a particular choice of the process parameters that can result in acyclic approximations of the bi-directional network is also discussed.

Case 2: The interplay between noise-variance and auto-regulatory feedback strength in a bivariate VAR, Fig. 1b, is investigated. Analytical expression on the role of the process parameters that can give rise to discrepancies in the ratio of the mean-squared forecast errors of the bi-directional network (Fig. 1b) is provided. Variation in statistical power as a function of significance level and sample-size for particular choice of the process parameters that can result in acyclic approximations of the bi-directional network is also discussed.

Case 3: The interplay between noise-variance, auto-regulatory feedback strength and memory of auto-regulatory feedback in a bivariate VAR, Fig. 1b, is investigated. Analytical expression on the role of the process parameters that can give rise to discrepancies in the ratio of the mean-squared forecast errors of the bi-directional network (Fig. 1b) is provided. Variation in statistical power as a function of significance level and sample-size for particular choice of the process
parameters that can result in acyclic approximations of the bi-directional network is also discussed.

It is important to note cases (1, 2 and 3) are designed in hierarchical order of increasing complexity for convenience, i.e. Case 1 can be realized from Case 2, which in turn can be realized from Case 3.

**Case 1:** *Impact of noise-variance on discrepancies in the forecast errors*

Consider a two-gene network motif with no auto-regulatory feedback, Fig.1a, modeled by the bivariate VAR,

\[
\begin{bmatrix}
x_t \\
y_t
\end{bmatrix} = \begin{bmatrix}
0 & \beta_1 \\
\beta_2 & 0
\end{bmatrix} \begin{bmatrix}
x_{t-1} \\
y_{t-1}
\end{bmatrix} + \begin{bmatrix}
\epsilon_t \\
\eta_t
\end{bmatrix}
\]

\[\text{.........................} \quad \text{(7)}\]

where the magnitude of the coupling strengths \( \beta_1, \beta_2 \neq 0 \) are chosen so as to satisfy the covariance stationarity constraint (2) (Hamilton, 1994). The model represents the case where gene \( t_x \) and \( t_y \) regulate each others transcription, Fig. 1a. It is important to re-iterate that (7) represents a bi-directionally coupled motif by definition, i.e. \( \beta_1, \beta_2 \neq 0 \), Fig. 1a. However, GC analysis can give rise to acyclic approximations whose direction is governed by the magnitude of the transcriptional noise variance as discussed below.

**Claim 1:** For equal transcriptional coupling strengths \( \beta_1 = \beta_2 \), the ratio of the mean-squared forecast error for the bi-directional network (7) is higher for the gene that exhibits greater transcriptional noise variance.

**Proof:** In order to determine whether \( y_t \rightarrow x_t \), we estimate the ratio of the mean-squared forecast errors \( s_0, s_1 \) as follows:

From (7) we get,

\[ x_t = \beta_1(\beta_2 x_{t-2} + \eta_{t-1}) + \epsilon_t \]  
\[ \text{.................................} \quad \text{(8)} \]

Mean-squared forecast error obtained by regressing \( x_t \) with its own past is given by

\[ s_0 = E(\beta_1 \eta_{t-1} + \epsilon_t)^2 = \beta_1^2 E(\eta_{t-1}^2) + E(\epsilon_t^2) = \beta_1^2 \sigma_\eta^2 + \sigma_\epsilon^2 \]  
\[ \text{.........................} \quad \text{(9)} \]

Mean-squared forecast error obtained by regressing \( x_t \) with its own past and the past of \( y_t \) is given by
\[ s_1 = E(\varepsilon_i^2) = \sigma_e^2 \] ................................................................. (10)

The ratio of the mean-squared forecast errors corresponding to \( y_t \rightarrow x_t \) from (9 and 10) is

\[ \gamma_{y \rightarrow x} = \frac{s_0}{s_1} = \frac{\beta_1^2 \sigma_{\eta}^2 + \sigma_e^2}{\sigma_e^2} \] ................................................................. (11)

A similar analysis to determine whether \( x_t \rightarrow y_t \) yields

\[ \gamma_{x \rightarrow y} = \frac{s_0}{s_1} = \frac{\beta_2^2 \sigma_{\eta}^2 + \sigma_e^2}{\sigma_{\eta}^2} \] ................................................................. (12)

Acyclic approximation in favor of \( x_t \rightarrow y_t \) can be realized when

\[ \gamma_{x \rightarrow y} >> \gamma_{y \rightarrow x} \]

i.e.

\[ \frac{\beta_2^2 \sigma_e^2 + \sigma_{\eta}^2}{\sigma_{\eta}^2} >> \frac{\beta_1^2 \sigma_{\eta}^2 + \sigma_e^2}{\sigma_e^2} \] ................................................................. (13)

Simplifying (13) results in the inequality

\[ \frac{\beta_2^2}{\beta_1^2} >> \frac{\sigma_{\eta}^4}{\sigma_e^4} \] ................................................................. (14)

For equal transcriptional coupling strengths i.e. \( \beta_1 = \beta_2 \), inequality (14) is dictated solely by the transcriptional noise variances. Hence the proof.

**Remark 1:**
Statistical inference using F-test relies on the ratio of the mean-squared forecast errors (5). Since \( \beta_1, \beta_2 \neq 0 \) (7), the null hypothesis should ideally be rejected in either direction for the bi-directional network. However, for equal transcriptional coupling strengths (\( \beta_1 = \beta_2 \)) ratio of the mean-squared forecast errors is implicitly greater for the gene which has greater transcriptional noise variance. Thus the rejection frequencies, hence the statistical power for suitable choice of process parameters, sample size and significance level can be greater for the gene with greater noise variance. This gene in turn acts as the cause in the resulting
acyclic approximation. The result (14) is also immune to the sign of the process parameters.

**Result 1:** Statistical power estimated as the successful rejection frequency across 1000 independent realizations of (7) with parameters \((\beta_1 = 0.05, \beta_2 = 0.05, \sigma_x = 1.5, \sigma_\eta = 1)\), sample-size \((N = 16, 64, 256, 1024, 4096)\) and significance level \((\alpha = 0.05, 0.01, 0.001)\) is shown in Figure 2a-2c respectively. The auto-regressive lag-length was fixed as \((p = 1)\), since (7) is a first-order bivariate VAR. It should be noted that the power for \((x_i \rightarrow y_i)\) is relatively higher than that of \((y_i \rightarrow x_i)\) irrespective of the choice of the sample size and significance level. These results are in support of our Claim 1 and suggest that the statistical power can be greater across the gene with higher transcriptional variance when the transcriptional coupling strength is held constant. Of interest, is to note Fig. 2(a) where statistical power of \((x_i \rightarrow y_i)\) is consistently higher (> 0.8) whereas those of \((y_i \rightarrow x_i)\) are consistently (< 0.8) for \((N = 16, 64, 256, 1024, 4096)\).

**Case 2:** Impact of auto-regulatory feedback on discrepancies in the forecast errors

Consider a two-gene network motif with auto-regulatory feedback, Fig. 1b, modeled by the bivariate VAR.

\[
\begin{bmatrix}
  x_i \\
  y_i \\
\end{bmatrix} =
\begin{bmatrix}
  \alpha_1 & \beta_1 \\
  \beta_2 & 0 \\
\end{bmatrix}
\begin{bmatrix}
  x_{i-1} \\
  y_{i-1} \\
\end{bmatrix} +
\begin{bmatrix}
  \epsilon_i \\
  \eta_i \\
\end{bmatrix}
\]

(15)

In (15) the magnitude of the coupling strength \(\beta_1, \beta_2 \neq 0\) and auto-regulatory feedback strength \(\alpha_i \neq 0\) are chosen so as to satisfy the covariance stationarity constraint (2) (Hamilton, 1994). This case represents a two-gene network motif, Fig. 1b, where genes \(x_i\) and \(y_i\) regulate each others transcription. In addition, gene \(x_i\) regulates its own transcription. This has to be contrasted with Case 1 (Fig.1a) where none of the genes exhibited auto-regulatory feedback. It is important to re-iterate that (15) represents a bi-directionally coupled motif by definition, i.e. \(\beta_1, \beta_2 \neq 0\), Fig. 1b. However, we show that GC analysis can give rise to acyclic approximations, where the gene exhibiting auto-regulatory feedback acts as the cause.
Claim 2: For equal transcriptional coupling strengths ($\beta_1 = \beta_2$) and noise variance ($\sigma^2_e = \sigma^2_\eta$), the ratio of the mean-squared forecast error for the bi-directional network (15) is greater along the direction where the gene exhibiting auto-regulatory feedback acts as the cause.

Proof: In order to determine whether $y_t \rightarrow x_t$, we estimate the ratio of the mean-squared forecast errors $s_0, s_1$ as follows:

From (15) we get,

$$x_t = \alpha_t x_{t-1} + \beta_1 (\beta_2 x_{t-2} + \eta_{t-1}) + \epsilon_t \quad \text{............................................... (16)}$$

Mean-squared forecast error on regression of $x_t$ with its own past is given by

$$s_0 = E(\beta_1 \eta_{t-1} + \epsilon_t)^2 = \beta^2_1 E(\eta^2_t) + E(\epsilon^2_t) = \beta^2_1 \sigma^2_\eta + \sigma^2_e \quad \text{......................... (17)}$$

Mean-squared forecast error on regression of $x_t$ with its own past and the past of $y_t$ is given by

$$s_1 = E(\epsilon^2_t) = \sigma^2_e \quad \text{............................................... (18)}$$

The ratio of the mean-squared forecast errors corresponding to $y_t \rightarrow x_t$ from (17 and 18) is

$$\gamma_{y \rightarrow x} = \frac{s_0}{s_1} = \frac{\beta^2_1 \sigma^2_\eta}{\sigma^2_e} + 1 \quad \text{............................................... (19)}$$

In order to determine whether $x_t \rightarrow y_t$, we estimate the ratio of the mean-squared forecast errors $s_0, s_1$ as follows:

From (15) we have

$$x_t = \alpha_t x_{t-1} + \beta_1 y_{t-1} + \epsilon_t \quad \text{. This in turn can be written as}$$

$$x_t = \frac{\beta_1 y_{t-1} + \epsilon_t}{(1-\alpha_t B)} \quad \text{............................................... (20)}$$

where $B$ represents the back-shift operator, i.e. $Bx_t = x_{t-1}$.
Substituting for \( x_t \) in terms of \( y_t \) from (20) in (15) results in

\[
y_t = \beta_2 \left( \frac{\beta_1 y_{t-2} + \epsilon_{t-1}}{1 - \alpha_1 B} \right) + \eta_t
\]

Binomial expansion of \((1 - \alpha_1 B)^{-1}\) with \(|\alpha_1| < 1\) yields

\[
y_t = \beta_2 \left[ (1 + \alpha_1 B + \alpha_1 B^2 + \ldots) (\beta_1 y_{t-2} + \epsilon_{t-1}) \right] + \eta_t \quad \text{.................. (21)}
\]

Mean-squared forecast error on regression of \( y_t \) with its own past is given by

\[
s_0 = \beta_2^2 \left[ E(\epsilon_{t-1}^2) + \alpha_1^2 E(\epsilon_{t-2}^2) + \alpha_1^4 E(\epsilon_{t-3}^2) + \ldots \right] + E(\eta_t^2) \quad \text{............... (22)}
\]

\[
= \beta_2^2 \sigma_e^2 \sum_{i=0}^{\infty} (\alpha_1^2)^i + \sigma_\eta^2 = \frac{\beta_2^2 \sigma_e^2}{(1 - \alpha_1^2)} + \sigma_\eta^2
\]

Mean-squared forecast error on regression of \( y_t \) with its own past and the past of \( x_t \) is given by

\[
s_1 = E(\eta_t^2) = \sigma_\eta^2 \quad \text{................................................... (23)}
\]

The ratio of the mean-squared forecast errors corresponding to \( x_t \to y_t \) from (22 and 23) is

\[
\gamma_{x\to y} = \frac{s_0}{s_1} = \frac{\beta_2^2 \sigma_e^2}{(1 - \alpha_1^2) \sigma_\eta^2} + 1 \quad \text{................................................... (24)}
\]

Acyclic approximation in favor of \( x_t \to y_t \) can be realized when

\[
\gamma_{x\to y} \gg \gamma_{y\to x}
\]

i.e.

\[
\frac{\beta_2^2 \sigma_e^2}{(1 - \alpha_1^2) \sigma_\eta^2} + 1 \gg \frac{\beta_1^2 \sigma_\eta^2}{\sigma_e^2} + 1 \quad \text{................................................... (25)}
\]
Simplifying (25) results in the inequality

$$\frac{\beta_2^2}{\beta_1^2} \gg \frac{(1 - \alpha_2^2)\sigma_n^4}{\sigma_e^4}$$

........................................................................................................................................ (26)

For equal transcriptional coupling strength and noise variance (i.e. $\beta_1 = \beta_2$ and $\sigma_e^2 = \sigma_n^2$), (26) reduces to

$$(1 - \alpha_1^2) << 1$$

which is always true, since $|\alpha_1| < 1$ (21). Hence the proof.

Remark 2:

(i) Statistical inference using F-test relies on the ratio of the mean-squared forecast errors (5). For equal transcriptional coupling strength $\beta_1 = \beta_2 \neq 0$ and noise variance $\sigma_e^2 = \sigma_n^2 > 0$, $\gamma_{y \rightarrow x}$ is implicitly lesser than $\gamma_{x \rightarrow y}$. This in turn can considerably affect the rejection frequency along either direction, hence the statistical power. More importantly, the power is likely to be greater along the direction where the gene with auto-regulatory feedback acts as the cause. This in turn results in acyclic approximation of bi-directional network. The result (26) is also immune to the sign of the process parameters.

(ii) While $\gamma_{y \rightarrow x}$ is unchanged across Cases 1 and 2, $\gamma_{x \rightarrow y}$ under Case (2) is implicitly greater than that of Case (1) for the same choice of the process parameters and $|\alpha_1| < 1$.

Result 2: Statistical power estimated as the rejection frequency across 1000 independent realizations of (15) with parameters $(\beta_1 = 0.05, \beta_2 = 0.05, \alpha_1 = 0.6, \sigma_n = 1, \sigma_e = 1)$ and their variation with sample-size ($N = 16, 64, 256, 1024, 4096$) and significance level ($\alpha = 0.05, 0.01, 0.001$) is shown in Figure 2d-2f respectively. The auto-regressive lag-length was fixed as ($p = 1$), since (15) is a first-order bivariate VAR. It should be noted that the power for $(x_i \rightarrow y_i)$ is relatively higher than that of $(y_i \rightarrow x_i)$ irrespective of the choice of the sample size and significance level. These results suggest that suitable choice of parameters under Claim 2 can give rise to discrepancies in power resulting in acyclic approximations of bi-directional networks where the gene exhibiting auto-regulatory feedback acts as the cause.
**Case 3: Impact of the memory of the auto-regulatory feedback on discrepancies in forecast errors**

Consider a two-gene network motif, Fig. 1b, where one of the genes exhibits auto-regulatory feedback modeled by the bivariate VAR.

\[
\begin{bmatrix}
    x_t \\
    y_t
\end{bmatrix} =
\begin{bmatrix}
    \alpha_1 & \beta_1 \\
    \beta_2 & 0
\end{bmatrix}
\begin{bmatrix}
    x_{t-1} \\
    y_{t-1}
\end{bmatrix} +
\begin{bmatrix}
    \alpha_2 & 0 \\
    0 & 0
\end{bmatrix}
\begin{bmatrix}
    x_{t-2} \\
    y_{t-2}
\end{bmatrix} +
\begin{bmatrix}
    \varepsilon_t \\
    \eta_t
\end{bmatrix}
\]

\[\text{…………………………….. (27)}\]

In (27) the magnitude of the coupling strength \(\beta_1, \beta_2 \neq 0\) and auto-regulatory feedback strength \(\alpha_1, \alpha_2 \neq 0\) are chosen so as to satisfy the covariance stationarity constraint (2) (Hamilton, 1994). This case represents a two-gene network motif, Fig. 1b, where genes \(x_t\) and \(y_t\) regulate each other's transcription. In addition gene \(x_t\) regulates its own transcription. Unlike Case 2, auto-regulatory feedback is assumed to be of order two in Case 3, Fig. 1b. It is important to re-iterate that (27) is bi-directionally coupled by definition, i.e. \(\beta_1, \beta_2 \neq 0\), Fig. 1b. However, we show that GC analysis can give rise to acyclic approximations in favor of the gene exhibiting auto-regulatory feedback even under the assumption of equal transcriptional noise variance and coupling strengths. The impact of the order of the auto-regulatory feedback contributing to the acyclic approximation is also investigated.

**Claim 3:** For equal transcriptional coupling strengths (\(\beta_1 = \beta_2\)) and noise variance (\(\sigma^2_e = \sigma^2_\eta\)), the ratio of the mean-squared forecast error for the bi-directional network (27) is greater along the direction where the gene exhibiting auto-regulatory feedback acts as the cause.

**Proof:** In order to determine whether \(y_t \rightarrow x_t\) we estimate the ratio of the mean-squared forecast errors \(s_0, s_1\) as follows:

From (27) we get,

\[
x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \beta_1 (\beta_2 x_{t-2} + \eta_{t-1}) + \varepsilon_t \]

\[\text{…………………………….. (28)}\]

Mean-squared forecast error on regression of \(x_t\) with its own past is given by

\[
s_0 = E((\beta_1 \eta_{t-1} + \varepsilon_t)^2) = \beta_1^2 E(\eta_{t-1}^2) + E(\varepsilon_t^2) = \beta_1^2 \sigma_\eta^2 + \sigma_e^2 \]

\[\text{…………………………….. (29)}\]

Mean-squared forecast error on regression of \(x_t\) with its own past and the past of \(y_t\) is given by

\[\text{…………………………….. (28)}\]

DOI: 10.2202/1557-4679.1119
\( s_1 = E(\varepsilon_1^2) = \sigma_\varepsilon^2 \) …………………………………………………. (30)

From (29 and 30) the ratio of the mean-squared forecast errors corresponding to \( y_t \rightarrow x_t \) is

\[
\gamma_{y \rightarrow x} = \frac{s_0}{s_1} = \frac{\beta_1^2 \sigma_\eta^2}{\sigma_\varepsilon^2} + 1 \] …………………………………………………. (31)

In order to determine whether \( x_t \rightarrow y_t \) we estimate the ratio of the mean-squared forecast errors \( s_0, s_1 \) as follows:

From (27) we have \( x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \beta_1 y_{t-1} + \varepsilon_t \). This in turn can be written as

\[
x_t = \frac{\beta_1 y_{t-1} + \varepsilon_t}{(1 - \alpha_1 B - \alpha_2 B^2)} \] …………………………………………………. (32)

where \( B \) represents the back-shift operator, i.e. \( B x_t = x_{t-1} \).

Substituting for \( y_t \) in terms of \( x_t \), (32) in (15) we get

\[
y_t = \beta_2 \left[ \frac{\beta_1 y_{t-2} + \varepsilon_{t-1}}{(1 - \alpha_1 B - \alpha_2 B^2)} \right] + \eta_t
\]

\[
= \frac{1}{(1 - \alpha_1 B - \alpha_2 B^2)} (\beta_2 \beta_1 y_{t-2} + \beta_2 \varepsilon_{t-1}) + \eta_t
\]

Representing the quadratic expression in the denominator as partial fractions

\[
= \frac{1}{(\theta_2 - \theta_1)} \left[ \frac{\theta_2}{(1 - \theta_1 B)} - \frac{\theta_1}{(1 - \theta_1 B)} \right] (\beta_2 \beta_1 y_{t-2} + \beta_2 \varepsilon_{t-1}) + \eta_t \] …………. (33)

where \( \theta_1 + \theta_2 = \alpha_1 \) and \( \theta_1 \theta_2 = -\alpha_2 \).

Binomial expansion of \( (1 - \theta_1 B)^{-1} \) and \( (1 - \theta_2 B)^{-1} \) under the constraints \( |\theta_1| < 1 \) and \( |\theta_2| < 1 \) yields

\[
y_t = \frac{1}{(\theta_2 - \theta_1)} \left[ \theta_2 (1 + \theta_2 B + ...) - \theta_1 (1 - \theta_1 B + ...) \right] (\beta_2 \beta_1 y_{t-2} + \beta_2 \varepsilon_{t-1}) + \eta_t
\]

…………………………….. (34)
From (33), it should be noted that \((\theta^2 - \theta_1)^2 = (\theta_1 + \theta_2)^2 - 4\theta_1\theta_2 = \alpha_1^2 + 4\alpha_2 \neq 0\). Therefore, (34) exists.

Mean-squared forecast error on regression of \(y\) with its own past is given by

\[
s_0 = \frac{\beta^2}{(\theta_2 - \theta_1)^2}[(\theta_2 - \theta_1)^2.E(e_2^2) + (\theta_2^2 - \theta_1^2)^2.E(e_2^2) + (\theta_2^3 - \theta_1^3)^2.E(e_2^2) + \ldots] + E(\eta^2) \]

\[
= \frac{\beta^2\sigma^2}{(\theta_2 - \theta_1)^2}[\sum_{i=1}^{\infty}(\theta_i^2 - \theta_1^2)^2] + \sigma^2 \tag{35}
\]

Simplifying (35) by summing the geometric series results in

\[
s_0 = \beta^2\sigma^2\left[\frac{1 + \theta_1\theta_2}{(1 - \theta_1^2)(1 - \theta_2^2)(1 - \theta_1\theta_2)}\right] + \sigma^2 \tag{36}
\]

From (34) we get \(|\theta_1| < 1\) and \(|\theta_2| < 1\). This in turn implies \(|\theta_1\theta_2| = |\theta_1| \cdot |\theta_2| < 1\).

Incorporating these constraints in (36) results in

\[
\beta^2\sigma^2\left[\frac{1 + \theta_1\theta_2}{(1 - \theta_1^2)(1 - \theta_2^2)(1 - \theta_1\theta_2)}\right] > 0 \tag{37}
\]

Substituting for \(\theta_1\) and \(\theta_2\) in terms of the process parameters from (33) in (37) results in

\[
\beta^2\sigma^2\left[\frac{1 - \alpha_2}{(1 - \alpha_2^2 - \alpha_1^2)(1 + \alpha_2)}\right] > 0 \tag{38}
\]

For (38) to hold the process parameters should satisfy the constraint \((1 - \alpha_2^2 - \alpha_1^2) > 0\), since \(|\alpha_2| < 1\).

Mean-squared forecast error on regression of \(y\) with its own past and the past of \(x\) is given by

\[
s_1 = E(\eta^2) = \sigma^2 \tag{39}
\]
From (38 and 39) the ratio of the mean-squared forecast errors corresponding to $x_i \rightarrow y_j$ is

$$\gamma_{x \rightarrow y} = \frac{s_0}{s_1} = \frac{\beta_2 \sigma_e^2}{\sigma_y^2} \left[ \frac{1 - \alpha_2}{(1 - \alpha_2)^2 - \alpha_1^2} \right] + 1 \quad \ldots \ldots \quad (40)$$

Acyclic approximation in favor of $x_i \rightarrow y_j$ can be realized when

$$\gamma_{x \rightarrow y} >> \gamma_{y \rightarrow x}$$

i.e. \[ \frac{\beta_2 \sigma_e^2}{\sigma_y^2} \left[ \frac{1 - \alpha_2}{(1 - \alpha_2)^2 - \alpha_1^2} \right] + \sigma_y^2 \sigma_i^2 + \sigma_e^2 \quad \ldots \ldots \quad (41) \]

Simplifying (41) results in the inequality

$$\frac{\beta_2^2}{\beta_1^2} \left[ \frac{1 - \alpha_2}{(1 - \alpha_2)^2 - \alpha_1^2} \right] >> \frac{\sigma_y^4}{\sigma_e^4} \quad \ldots \ldots \quad (42)$$

For equal transcriptional coupling strengths ($\beta_1 = \beta_2$) and noise variances ($\sigma_e^2 = \sigma_i^2$) expression (42) results in the inequality

$$\frac{1 - \alpha_2}{(1 - \alpha_2)^2 - \alpha_1^2} >> 1$$

Simplifying the above expression results in

$$\alpha_2^2(1 - \alpha_2) + \alpha_1^2(1 + \alpha_2) >> 0$$

which is always true, since $|\alpha_2| = |\theta_2 \theta_1| < 1$ (38). Hence the proof.

Remark 3:

(i) Statistical inference using F-test relies on the ratio of the mean-squared forecast errors (5). By very definition (15), $\beta_1, \beta_2 \neq 0$, the null hypothesis should be rejected in either direction for the bi-directional network. However, for equal
transcriptional coupling strength and noise variance $\gamma_{y\rightarrow x}$ is implicitly lesser than $\gamma_{x\rightarrow y}$. This in turn can considerably affect the rejection frequency along either direction, hence the statistical power. For suitable choice of process parameters, sample size and significance level the power is likely to be greater along the direction where the gene with auto-regulatory feedback acts as the cause i.e. $\gamma_{x\rightarrow y}$.

(ii) $\gamma_{y\rightarrow x}$ is unchanged across Cases 1, 2 and 3. $\gamma_{x\rightarrow y}$ under Case (3) is implicitly greater than that of Case (1) for the same choice of parameters. $\gamma_{x\rightarrow y}$ under Case (3) is implicitly greater than that of Case (2) for the same choice of parameters and $\alpha_2 > 0$.

**Result 3:** Statistical power estimated as the rejection frequency across 1000 independent realizations of (27) with parameters ($\beta_1 = 0.05, \beta_2 = 0.05, \alpha_1 = 0.6, \alpha_2 = 0.6, \sigma_y = 1, \sigma_\epsilon = 1$) and their variation with sample-size (N = 16, 64, 256, 1024, 4096) and significance level ($\alpha = 0.05, 0.01, 0.001$) is shown in Figure 2g-2i respectively. The auto-regressive lag-length was fixed as ($p = 2$), since (15) is a second-order bivariate VAR. It should be noted that the power for ($x_t \rightarrow y_t$) is relatively higher than that of ($y_t \rightarrow x_t$) irrespective of the choice of the sample size and significance level. These results are in support of our Claim 3 and suggest that the statistical power can be greater across the gene with auto-regulatory feedback when the transcriptional variance and coupling strength are held constant. Of interest, is to note Figs. 2h and 2i where statistical power of ($x_t \rightarrow y_t$) is consistently ($> 0.8$) whereas those of ($y_t \rightarrow x_t$) are consistently ($< 0.8$) for (N = 16, 64, 256, 1024, 4096). These in turn can result in acyclic approximations of bi-directional networks where the gene with auto-regulatory feedback acts as the cause. It is also of interest to note that the power corresponding to ($x_t \rightarrow y_t$), Figs. 2g-2i seems to be higher than those obtained under Case (2), Figs. 2d-2f. This can be expected since the expression for $\gamma_{x\rightarrow y}$ (40) is greater than that of (24) for the same transcriptional strength, noise variance and $\alpha_2 = 0.6$.

While GC analysis is ideally suited to investigate possible linear causal relationships in bivariate processes, recent studies have used GC analysis to infer causal relationships from multivariate processes (Brovelli et al, 2004; Seth et al 2006; Mukhopadhyay and Chatterjeee, 2007; Fujita et al., 2007; Guo et al., 2008). Interpretation of GC results across multivariate processes can especially be challenging when there are unobserved variables. For completeness we consider a simple example of a bivariate VAR process in the presence of an unknown
common cause, also known as *fork* (Pearl, 2000) (see Appendix). The off-diagonal elements in the bivariate VAR are set to zero. Thus, there exists no direct relationship between $x_t$ and $y_t$. However, it is shown that an unobserved common cause $z_t$ can result in spurious conclusion of direct causal relationship between the processes $x_t$ and $y_t$. Within the context of transcriptional networks, a common cause can be attributed an upstream gene that regulates the expression of downstream genes after finite delays. The magnitude of the delays plays a crucial role in determining direction of the cause and effect relationship when $z_t$ is unobserved. Conditioning the processes $x_t$ and $y_t$ on the common cause $z_t$ can prevent such spurious conclusions. A similar behavior can be observed for mediators (*chains*) (Pearl, 2000) where the transcriptional regulation between genes $x_t$ and $y_t$ is mediated through $z_t$ (see Appendix). A more rigorous treatment of causal inference across multivariate data can be found elsewhere (Robins et al., 1999; Pearl, 2000; Robins, 2003).

3. DISCUSSION

GC is ideally suited for inferring possible causal relationships in stationary bivariate VAR processes. Acyclic structures are ideally accompanied by setting the upper or lower off-diagonal element(s) to zero in the bivariate VAR. This in turn results in absence of forecasting ability in that direction. However, in experimental settings statistical tests that rely on the ratio of the mean-squared errors are used infer causal relationships. The present study investigates two-gene network motifs of increasing complexity modeled as stationary bivariate VARs. The impact of auto-regulatory feedback strength and memory of the auto-regulatory feedback that can give rise to significant discrepancies in the ratio of the mean-squared forecast errors for bi-directional networks with equal transcriptional coupling strength is elucidated. Subsequently, Monte Carlo simulations were used to establish discrepancies in statistical power as a function of sample size and significance level for the case studies. More importantly, it is shown that one can arrive at acyclic approximations even for a bi-directional network for suitable choice of process parameters, significance level and sample size. The concerns presented in the present study can only be aggravated in the presence of nonstationarities and unobserved processes (Appendix). Apriori knowledge in conjunction with experimental manipulations may be necessary to overcome some of these inherent limitations and strengthen the conclusions arrived using GC analysis.
Figure 1 Two-gene transcriptional network motifs with (b) and without (a) autoregulatory feedback consisting of genes $x$ and $y$. In (a) and (b) parameters ($\beta_1$ and $\beta_2$), ($\epsilon_t$ and $\eta_t$) and ($\alpha_1, \alpha_2$) can be thought of as the transcriptional coupling strengths, uncorrelated transcriptional noise, and autoregulatory feedback strengths respectively. Each of the arrows in these motifs can represent either activation or inhibition.
Figure 2 Power corresponding to \((x_i \rightarrow y_i, \text{ black line})\) and \((y_i \rightarrow x_i, \text{ red line})\) for Case (1) with the parameters \((\beta_1 = 0.05, \beta_2 = 0.05, \sigma_\eta = 1, \sigma_e = 1.5)\), significance levels \((\alpha = 0.05, 0.01 \text{ and } 0.001)\) and sample size \((N = 16, 64, 256, 1024, 4096)\) estimated from 1000 independent realizations is shown in (a), (b) and (c) respectively. Power from a similar analysis corresponding to \((x_i \rightarrow y_i, \text{ black line})\) and \((y_i \rightarrow x_i, \text{ red line})\) for Case (2) with the parameters \((\beta_1 = 0.05, \beta_2 = 0.05, \alpha_1 = 0.6, \alpha_2 = 0, \sigma_\eta = \sigma_e = 1)\) is shown in (d), (e) and (f) respectively. Those for Case (3) with parameters \((\beta_1 = 0.05, \beta_2 = 0.05, \alpha_1 = 0.6, \alpha_2 = 0.3, \sigma_\eta = \sigma_e = 1)\) are shown in (g), (h) and (i) respectively. The dotted line in the figure corresponds to power (0.8) and included as a reference.
APPENDIX

GC inference in bivariate VAR in the presence of an unobserved common cause

Consider a two-gene network motif represented by the bivariate VAR

\[
\begin{align*}
x_t &= \epsilon_t \\
y_t &= \beta_1 y_{t-1} + \eta_t
\end{align*}
\]  
(A1)

where \((\beta_1 \neq 0)\) and \((\epsilon_t, \eta_t)\) are finite-variance uncorrelated noise. In the above case, the off-diagonal elements of the bivariate VAR are zeros. Therefore, neither \(x_t \rightarrow y_t\) nor \(y_t \rightarrow x_t\).

**Fork:** Now consider (A1) in the presence of a common cause \(z_t\)

\[
\begin{align*}
x_t &= \alpha_1 z_{t-m} + \epsilon_t \\
y_t &= \alpha_2 z_{t-n} + \beta_1 y_{t-1} + \eta_t
\end{align*}
\]  
(A2)

where \((\alpha_1, \alpha_2, \beta_1 \neq 0)\) and \((\epsilon_t, \eta_t)\) are finite-variance uncorrelated noise. The above structure represents a fork (Pearl, 2000) where a common up-stream gene
regulates the expression of down-stream genes after finite delays \(0 < m < n\). It is important to note that the magnitude of the delays govern the temporal precedence, hence the direction of causality between \(x_1\) and \(y_1\), when \(z_1\) is unobserved. As in (A1) there is no direct relationship between \(x_1\) and \(y_1\). However, the unobserved common cause in conjunction with differential delays can introduce spurious causal relationship between \((x_1, y_1)\).

In order to determine whether \(x_1 \to y_1\) \((m < n)\) we estimate the ratio of the mean-squared forecast errors \(s_0, s_1\) as follows:

Substituting for \(z_{t-m}\) from (A2) into (A3) we get

\[
y_t = \left(\frac{\alpha_2}{\alpha_1}\right)\left[x_{t+m-a} - \epsilon_{t+m-a}\right] + \beta_1 y_{t-1} + \eta_t
\]

Mean-squared forecast error on regression of \(y_t\) with its own past is given by

\[
s_0 = \left(\frac{\alpha_2}{\alpha_1}\right)^2 \left[ E(x_t^2) + E(\epsilon_t^2) \right] + E(\eta_t^2) \quad \text{.......................... (A4)}
\]

Mean-squared forecast error on regression of \(y_t\) with its own past and the past of \(x_t\) is given by

\[
s_1 = \left(\frac{\alpha_2}{\alpha_1}\right)^2 E(\epsilon_t^2) + E(\eta_t^2) \quad \text{................................. (A5)}
\]

From (A4) and (A5) we get

\[
\gamma_{y \to x} = \frac{s_0}{s_1} = \frac{\left(\frac{\alpha_2}{\alpha_1}\right)^2 E(x_t^2) + \left(\frac{\alpha_2}{\alpha_1}\right)^2 E(\epsilon_t^2) + E(\eta_t^2)}{\left(\frac{\alpha_2}{\alpha_1}\right)^2 E(\epsilon_t^2) + E(\eta_t^2)}
\]
Substituting for $\sigma_x^2$ from (A2) we get

$$
\gamma_{y \rightarrow x} = \frac{\left(\frac{\alpha_z}{\alpha_1}\right)^2 (\sigma_z^2 + \sigma_e^2) + \left[ \left(\frac{\alpha_z}{\alpha_1}\right)^2 \sigma_e^2 + \sigma_\eta^2 \right]}{\left(\frac{\alpha_z}{\alpha_1}\right)^2 \sigma_e^2 + \sigma_\eta^2} \quad \text{........... (A6)}
$$

Unlike (A1), $s_0$ is implicitly greater than $s_1$ in (A6). This discrepancy can be large for suitable choice of the parameters, rendering $x \rightarrow y$, statistically significant. Therefore it is possible to conclude spurious causal relationship of the form $x \rightarrow y$, in the presence of an unobserved common cause. If $z_i$ is observed, then conditioning $(x_i, y_i)$ on $z_i$ prior to GC estimation can be useful in avoiding spurious conclusions.

**Chains:** In the above case, gene $z_i$ acts as the common cause of $x_i$ and $y_i$. However, there can be instances where a gene ($z_i$) mediates the expression between an up-stream $x_i$ and a down-stream gene $y_i$. Such structures are termed as **chains** (Pearl, 2000) e.g. $(x_i \rightarrow z_i \rightarrow y_i)$ and $(y_i \rightarrow z_i \rightarrow x_i)$. There are significant similarities in the behavior of chains and fork. (i) As in a fork, there is no direct causal relationship between $x_i$ and $y_i$ in chains. (ii) It is possible arrive at spurious causal relationship of the form $x_i \rightarrow y_i$ when $z_i$ is unobserved in $(x_i \rightarrow z_i \rightarrow y_i)$. (iii) If $z_i$ is observed, then conditioning the processes with respect to $z_i$ prior to GC estimation can be useful in avoiding spurious conclusions.

**REFERENCES**


DOI: 10.2202/1557-4679.1119


