Edge Exclusion Tests for Graphical Model Selection: Complex Gaussian Vectors and Time Series

Jitendra K. Tugnait, *Life Fellow, IEEE*

***Abstract*—We consider the problem of inferring the conditional independence graph (CIG) of a stationary multivariate Gaussian time series. A p-variate Gaussian time series graphical model associated with an undirected graph with p vertices is defined as the family of time series that obey the conditional independence restrictions implied by the edge set of the graph. In some existing methods, partial coherence has been used as a test statistic for graphical model selection. To test inclusion/exclusion of a given edge in the graph, the test is applied at distinct frequencies, requiring multiple tests and leading to a loss in test power. The nonparametric method of Matsuda uses the Kullback-Leibler divergence measure to define a test statistic which does not need multiple testing to test between two competing models. In this paper we propose a generalized likelihood ratio test (GLRT) based edge exclusion test statistic that also does not need multiple testing. It is computationally significantly faster than the method of Matsuda, and simulations show that we achieve comparable power levels. Our proposed approach is based on a novel formulation of a GLRT based edge exclusion test for p-variate complex Gaussian graphical models (CGGMs); this result is also of independent interest. The computational complexity of the proposed statistic is $\mathcal{O}(p^3)$ compared to $\mathcal{O}(p^5)$ for the existing result. We also apply our time series graphical model selection method to a foreign exchange data set consisting of monthly trends of foreign exchange rates of 10 countries.**

***Index Terms*—Time series graphical models, complex Gaussian graphical models, undirected graph, multivariate time series, generalized likelihood ratio test.**

I. Introduction

GRAPHICAL interaction models ("graphical models," in short) are an important and useful tool for analyzing multivariate data [1]. A more recent emphasis stems from the big data perspective [2]. Graphical modeling is a form of multivariate analysis where one uses graphs to represent models. A central concept is that of conditional independence. Given a collection of random variables, one wishes to assess the relationship between two variables, conditioned on the remaining variables. In graphical models, graphs are used to display the conditional independence structure of the variables.

Manuscript received March 18, 2019; revised July 16, 2019; accepted August 10, 2019. Date of publication August 19, 2019; date of current version August 31, 2019. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Gwo Giun Lee. This work was supported by the National Science Foundation under Grant CCF-1617610. This paper was presented in part at the 52nd Annual Conference on Information Sciences and Systems, Princeton University, Princeton, NJ, USA, March 2018, and in part at the IEEE Statistical Signal Processing Workshop (SSP), Freiburg, Germany, June 2018.

The author is with the Departmen of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849 USA (e-mail: tugnajk@auburn.edu).

Digital Object Identifier 10.1109/TSP.2019.2935898

Consider a graph $\mathcal{G} = (V, \mathcal{E})$ with a set of p vertices (nodes) $V = \{1, 2, \ldots, p\} = [1, p]$, and a corresponding set of (undirected) edges $\mathcal{E} \subset V \times V$. Given a random vector $\mathbf{x} = [x_1\ x_2\ \cdots\ x_p]^\top$, in the corresponding graph $\mathcal{G}$, each variable x_i is represented by a node (i in V), and associations between variables x_i and x_j are represented by edges between nodes i and j of $\mathcal{G}$. In a conditional independence graph (CIG), there is no edge between nodes i and j if and only if (iff) x_i and x_j are conditionally independent given the remaining $p-2$ variables x_ℓ, $\ell \in [1, p]$, $\ell \neq i$, $\ell \neq j$. Thus, edge $\{i, j\} \in \mathcal{E}$ iff x_i and x_j are conditionally dependent, and edge $\{i, j\} \notin \mathcal{E}$ iff x_i and x_j are conditionally independent. Gaussian graphical models (GGM) are CIGs where $\mathbf{x}$ is multivariate Gaussian. Suppose $\mathbf{x}$ has positive-definite covariance matrix $\mathbf{\Sigma}$ with inverse covariance matrix (also known as precision matrix or concentration matrix) $\mathbf{\Omega} = \mathbf{\Sigma}^{-1}$. Then Ω_{ij}, the $(i, j)-$th element of $\mathbf{\Omega}$, is zero iff x_i and x_j are conditionally independent. If $\mathbf{x}$ is non-Gaussian, then Ω_{ij} encodes conditional uncorrelatedness.

Graphical models were originally developed for random vectors with multiple independent realizations [3, p. 234], i.e., for time series that is independent and identically distributed (i.i.d.): real-valued $p-$ dimensional $\mathbf{x}(t)$, $t = 0, 1, \ldots$, with $\mathbf{x}(t_1)$ independent of $\mathbf{x}(t_2)$ for $t_1 \neq t_2$, and $\mathbf{x}(t)$ identically distributed for any (integer) t. Such models have been extensively studied, and found to be useful in a wide variety of applications [2], [4]–[10]. For complex-valued $\mathbf{x}$, only the monograph [11] has studied such models. A significant application of graphical modeling of real-valued random vectors is for analysis of functional magnetic resonance imaging (fMRI) data [12], to provide insights into the functional connectivity of different brain regions [13]. However, as noted in [14], "Before the statistical analysis of the fMRI data, the phase portion of the data is generally discarded, despite physiologically useful information contained in the phase." It has been shown in [14], [15] that complex-valued fMRI data yields improved sensitivity in fMRI analysis. Thus, independent investigation of graphical modeling of complex-valued random vectors is well motivated, although in this paper, we consider it for use with complex-valued frequency-domain representation of real-valued time series.

Graphical modeling of real-valued time-dependent data (stationary time series) originated with [16], followed by [17]. Consider a stationary (real-valued), zero-mean, $p-$ dimensional multivariate Gaussian time series $\mathbf{x}(t)$, $t = 0, \pm 1, \pm 2, \ldots$, with correlation (covariance) matrix function $\mathbf{R}_{xx}(\tau) = \mathbb{E}\{\mathbf{x}(t+\tau)\mathbf{x}^\top(t)\}$, $\tau = 0, \pm 1, \ldots$. As in the random vector case, in the corresponding CIG (or GGM) $\mathcal{G}$, edge $\{i, j\} \in \mathcal{E}$ (there is an

1053-587X © 2019 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

edge between nodes i and j), iff time series components $x_i(t)$ and $x_j(t)$ are conditionally dependent, and edge $\{i,j\} \notin \mathcal{E}$ iff $x_i(t)$ and $x_j(t)$ are conditionally independent.

Time series graphical models have been applied to intensive care monitoring [18], financial time series [19], [20], air pollution data [17], [19], and analysis of EEG data [21]–[23], and fMRI data [12], [20] to provide insights into the functional connectivity of different brain regions [13].

A key insight in [17] was to transform the series correlation function in time domain to power spectral density (PSD) in the frequency domain, and express the graph relationships in the frequency domain. Denote the PSD matrix of $\{\mathbf{x}(t)\}$ by $\mathbf{S}_x(f)$, where $\mathbf{S}_x(f) = \sum_{\tau=-\infty}^{\infty} \mathbf{R}_{xx}(\tau) e^{-j2\pi f\tau}$, the discrete-time Fourier transform of $\mathbf{R}_{xx}(\tau)$. Here f is the normalized frequency, in Hz, in the interval [0,1) or $(-0.5, 0.5]$. In [16], [17] it was shown that conditional independence of two time series components given all other components of the time series, is encoded by zeros in the inverse PSD, that is, $\{i,j\} \notin \mathcal{E}$ iff the (i,j)-th element of $\mathbf{S}_x(f)$, $[\mathbf{S}_x^{-1}(f)]_{ij} = 0$ for every f. Thus, $\mathbf{S}_x^{-1}(f)$ plays the same role for a general time series as is done by the concentration matrix $\mathbf{\Omega} = \mathbf{\Sigma}^{-1}$ in the i.i.d. time series setting.

To test whether $\{i,j\} \in \mathcal{E}$, [17] suggested a test based on the maximum of nonparametrically estimated partial coherence over $f \in [0, 0.5]$. Exact asymptotic null distribution of Dahlhaus's statistic is not known. Partial coherence as a test statistic for edge exclusion has been widely used [18], [21], [22]. For a given test edge, the test is applied at distinct frequencies, leading to multiple tests. In order to control the significance level, the Holm procedure [24] and related methods have been used. In [22] several methods for p-value combining have been investigated.

Matsuda [25] proposed the use of nonparametrically estimated Kullback-Leibler (KL) divergence between two graphical models, for graphical modeling. This statistic is the result of integrating over $f \in [0, 0.5]$, hence does not need multiple testing to test between two competing models, unlike [18], [21], [22]. The asymptotic null distribution of Matsuda's test statistic is standard Gaussian, allowing for analytical calculation of the test threshold (critical region). Also, since there is one test per edge, there is little additional loss in test power which results when one resorts to multiple testing as in [18], [21], [22]. Recently, [23] used Matsuda's test statistic to devise a computationally efficient method for testing inclusion/exclusion of all edges in the selected graphical model.

In this paper, we propose a generalized likelihood ratio test (GLRT) based edge exclusion test statistic that also does not need multiple testing. It turns out to be computationally significantly faster than Matsuda's method, while simulations show that we achieve comparable test power levels. In order to address graphical model selection for real-valued time series, first we need to recall, and then refine and improve upon, some existing results from [11], by following some real-valued GGM results from [1]. Note that [1] deals with real-valued GGMs for real-valued Gaussian vectors (i.e., i.i.d. real time series), while [11] investigates complex-valued GGMs for complex-valued Gaussian vectors (i.e., i.i.d. complex time series). The complex-valued results of [11] are essential for our purposes since the discrete Fourier transform (DFT) of $\{\mathbf{x}(t)\}$ will, in general, be complex-valued (even if $\mathbf{x}(t)$ is real-valued).

Therefore, in this paper, we also investigate a GLRT based edge exclusion test statistic for complex Gaussian graphical models (CGGMs). A GLRT statistic for this problem is given in [11, Theorem 7.6], but it is not in the form which is usually given and exploited for real GGMs, as e.g., in [1, Eqn. (5.12), p. 130] and [1, p. 150], or [26]. We fill this gap. Furthermore, the computational complexity (number of flops (floating point operations)) of our proposed alternative statistic is $\mathcal{O}(p^3)$ versus $\mathcal{O}(p^5)$ for the expression of [11], when the test statistics for all edges need to be computed and tested. This is established in Sec. III-A. We also provide expressions for power of the test. Such tests form the basis of model selection in GGMs via multiple testing [27], [28]. We do not consider multiple testing issues in this paper.

Preliminary versions of parts of this paper appear in conference papers [29], [30]. CGGMs are addressed in [30] and time series graphical models are investigated in [29]. In this paper, we correct several errors in the theoretical results of [29], [30], and also correct errors in some simulation results reported therein. The material in Secs. IV-B, IV-E, IV-F and VI of this paper does not appear in [29], [30].

The rest of the paper is organized as follows. In Sec. II, we introduce notation, present the system model for time series graphical modeling, and discuss some alternative approaches. In Sec. III we address CGGMs and present our GLRT. These results are then utilized in Sec. IV to address times series GGMs by introducing a frequency-domain sufficient statistic based on the DFT of the time-domain data. This approach is novel. For a large class of stationary random processes, the DFT of a given stretch of measurements is complex, asymptotically jointly proper Gaussian, and independent at distinct frequencies on an appropriate frequency grid [31]. If the PSD of the measurements is constant over a selected frequency interval, the DFTs over this interval act as independent complex Gaussian measurements, allowing one to invoke the results of Sec. III. In Sec. IV-F we also discuss our version of Dahlhaus's test [17], for which exact asymptotic null distribution can be determined, unlike [17]. Computational comparisons with the approach of [23], [25] are provided in Sec. IV-E, which show that while our approach needs $\mathcal{O}(p^3 N)$ flops, $\mathcal{O}(p^5 N)$ flops are needed by [23], [25], for a sample size of N. Simulation results are presented in Sec. V, and some real data results are in Sec. VI.

II. Preliminaries and System Model

Here we first introduce notation in Sec. II-A, and then present the system model for time series graphical modeling in Sec. II-B. In Sec. II-C we briefly discuss and contrast some alternative approaches to this problem, such as parametric graphical models, and penalized log-likelihood based approaches in high-dimensional settings.

A. Notation and Terminology

We use $\mathbf{S} \succeq 0$ and $\mathbf{S} \succ 0$ to denote that Hermitian $\mathbf{S}$ is positive semi-definite and positive definite, respectively. For a square

matrix $\mathbf{A}$, $|\mathbf{A}|$ and $\mathrm{etr}(\mathbf{A})$ denote the determinant and the exponential of the trace of $\mathbf{A}$, respectively, i.e., $\mathrm{etr}(\mathbf{A}) = \exp(\mathrm{tr}(\mathbf{A}))$, $[\mathbf{B}_k]_{i:l,j:m}$ denotes the submatrix of the matrix $\mathbf{B}_k$ comprising its rows i through l and columns j through m, $[\mathbf{B}_k]_{ij}$ is its (i,j)-th element, B_{ij} is the (i,j)-th element of $\mathbf{B}$, x_i is the i-th component of $\mathbf{x} \in \mathbb{C}^p$, and $\mathbf{I}$ is the identity matrix. The superscripts $*$, $\top$ and H denote the complex conjugate, transpose and Hermitian (conjugate transpose) operations, respectively, and $\mathbb{E}$ denotes the expectation operation. The set of integers 1 through n is denoted by $[1,n]$. The sets of real and complex numbers are denoted by $\mathbb{R}$ and $\mathbb{C}$, respectively. The set $\mathbb{C}_+^{p\times p}$ denotes the set of $p \times p$ positive definite complex matrices, i.e., $\mathbb{C}_+^{p\times p} = \{\mathbf{S} \in \mathbb{C}^{p\times p} \,|\, \mathbf{S} \succ 0\}$. The notation $y = \mathcal{O}(g(x))$ means that there exists some finite real number $b > 0$ such that $\lim_{x\to\infty} |y/g(x)| \leq b$.

The notation χ_n^2 represents central chi-square distribution with n degrees of freedom. The notation $\mathbf{x} \sim \mathcal{N}_c(\mathbf{m}, \mathbf{\Sigma})$ denotes a random vector $\mathbf{x}$ that is proper complex Gaussian with mean $\mathbf{m}$ and covariance $\mathbf{\Sigma}$, and $\mathbf{x} \overset{a}{\sim} \mathcal{N}_c(\mathbf{m}, \mathbf{\Sigma})$ implies that $\mathbf{x}$ is asymptotically $\mathcal{N}_c(\mathbf{m}, \mathbf{\Sigma})$ as number of measurements tend to ∞. The notation $\overset{a}{\sim}$ applies to other distributions (χ^2, Wishart, etc) as well. Also, for real $\mathbf{x}$, $\mathbf{x} \sim \mathcal{N}_r(\mathbf{m}, \mathbf{\Sigma})$ denotes a real random vector $\mathbf{x}$ that is Gaussian with mean $\mathbf{m}$ and covariance $\mathbf{\Sigma}$. The abbreviations w.p.1. and i.i.d. stand for with probability one, and independent and identically distributed, respectively.

We use the following notation for covariance, variance and conditional variance

$$\mathrm{cov}(\mathbf{x}, \mathbf{y}) = \mathbb{E}\{[\mathbf{x} - \mathbb{E}\{\mathbf{x}\}][\mathbf{y} - \mathbb{E}\{\mathbf{y}\}]^H\} = \mathbf{\Sigma}_{\mathbf{x},\mathbf{y}} \tag{1}$$

$$\mathrm{var}(\mathbf{x}) = \mathrm{cov}(\mathbf{x}, \mathbf{x}) = \mathbf{\Sigma}_{\mathbf{x}} \tag{2}$$

$$\mathrm{var}(\mathbf{x} \,|\, \mathbf{y}) = \mathbb{E}\{[\mathbf{x} - \mathbb{E}\{\mathbf{x}\}][\mathbf{x} - \mathbb{E}\{\mathbf{x}\}]^H \,|\, \mathbf{y}\} = \mathbf{\Sigma}_{\mathbf{x}\,|\,\mathbf{y}}\,. \tag{3}$$

A simple undirected graph $\mathcal{G}$ is a pair $(V, \mathcal{E})$, where V is a finite set of elements called vertices (or nodes), and the set $\mathcal{E} \subset V \times V$ of unordered pairs (called edges) of distinct vertices. We denote the edge between vertices α and β of V as $\{\alpha, \beta\}$. The set $V \backslash A$ denotes the set of all elements in V that are not in A. A graph is said to be complete, or saturated, if any given vertex is connected to every other vertex.

B. System Model

A GGM associated with a simple undirected graph $\mathcal{G} = (V, \mathcal{E})$ is defined as the family of p-variate Gaussian random vectors $\mathbf{x} \in \mathbb{R}^p$, $p = |V|$ (= cardinality of V), that obey the conditional independence restrictions implied by the edge set $\mathcal{E}$. Similarly, a time series GGM associated with an undirected graph $\mathcal{G} = (V, \mathcal{E})$ is defined as the family of p-variate Gaussian time series $\mathbf{x}(t) \in \mathbb{R}^p, t = 0, \pm 1, \pm 2, \ldots, p = |V|$, that obey the conditional independence restrictions implied by the edge set $\mathcal{E}$, i.e., $\{i,j\} \notin \mathcal{E}$ iff $[\mathbf{S}_x^{-1}(f)]_{ij} = 0$ for every f, where $\mathbf{S}_x(f)$ is the $p \times p$ PSD matrix of (real-valued) stationary $\mathbf{x}(t)$.

Consider two undirected graphs $\mathcal{G} = (V, \mathcal{E})$ and $\mathcal{G}' = (V, \mathcal{E}')$ where $|V| = p =$ the number of nodes (vertices), and $\mathcal{G}'$ is a subgraph of $\mathcal{G}$, i.e., $\mathcal{E}'$ is a subset of $\mathcal{E}$. Let $\mathcal{E}$ be saturated, i.e., all edges belong to $\mathcal{E}$, meaning every node is connected to every other node in the graph; this is the complete graph. Now suppose that we remove one edge $\{i_0, j_0\}$ from $\mathcal{E}$ to obtain $\mathcal{E}' = \mathcal{E} \backslash \{i_0, j_0\}$. Note that since the graphs being considered are undirected, if $\{i,j\} \in \mathcal{E}$ then $\{j,i\} \in \mathcal{E}$, and if $\{i,j\} \notin \mathcal{E}$ then $\{j,i\} \notin \mathcal{E}$. Given p-variate Gaussian time series $\mathbf{x}(t) \in \mathbb{R}^p$, $t = 0, \pm 1, \pm 2, \ldots$, we wish to decide if the CIG associated with $\mathbf{x}(t)$ is $\mathcal{G}$ or $\mathcal{G}'$. In the case of random vectors, such tests have been called edge exclusion tests [28], [32].

For time series $\mathbf{x}(t)$, since the associated graph determines if $[\mathbf{S}_x^{-1}(f)]_{ij}$ is zero or nonzero, we will denote $\mathbf{S}_x(f)$ as $\mathbf{S}_x(f; \mathcal{G})$ to explicitly indicate this dependence. Therefore, the binary hypothesis problem of interest is as follows. Given observations of time series $\mathbf{x}(t)$, $t = 0, 1, \ldots, N-1$, with positive definite PSD matrix $\mathbf{S}(f)$, we need to distinguish between the null hypothesis $\mathcal{H}_0$ and the alternative $\mathcal{H}_1$:

$$\begin{aligned} \mathcal{H}_0 \,:\, & \mathbf{S}_x^{-1}(f) = \mathbf{S}_x^{-1}(f; \mathcal{G}) \in \mathbb{C}_+(\mathcal{G}') \;\forall\, f \in [0, 0.5], \\ & \mathcal{G}' = (V, \mathcal{E}'),\; \mathcal{E}' = \mathcal{E} \backslash \{i_0, j_0\} \\ \mathcal{H}_1 \,:\, & \mathbf{S}_x^{-1}(f) = \mathbf{S}_x^{-1}(f; \mathcal{G}) \in \mathbb{C}_+(\mathcal{G}) \;\forall\, f \in [0, 0.5], \\ & \mathcal{G} = (V, \mathcal{E}),\; \mathcal{E} \text{ is saturated} \end{aligned} \tag{4}$$

where

$$\begin{aligned} \mathbb{C}_+(\mathcal{G}) = \Big\{ \mathbf{\Omega} \in \mathbb{C}_+^{p\times p} \,\big|\, \forall i \neq j \in V \;:\; \{i,j\} \notin \mathcal{E} \\ \Rightarrow \mathbf{\Omega}_{ij} = 0 \Big\}. \end{aligned} \tag{5}$$

Thus, $\mathbb{C}_+(\mathcal{G})$ is the set of all $p \times p$ positive-definite Hermitian matrices containing a zero entry at the (i,j)-th element when $\{i,j\} \notin \mathcal{E}$ [11].

C. Alternative Approaches

This paper is focused on undirected graphs, non-sparse low-dimensional data regimes (i.e., $p \ll$ sample size), and nonparametric methods, similar to [17], [23], [25]. As an alternative to nonparametric modeling of time series, parametric graphical models utilizing a (Gaussian) vector autoregressive (VAR) process model of $\mathbf{x}(t)$, have been advocated in [3], [19], [33]. The VAR parameters are constrained by an associated graph, and there are $2^{p(p-2)/2}$ possible graphs for a p-dimensional time series. An information criterion can be used to select an appropriate model out of $2^{p(p-2)/2}$ possible models, but this makes the method suitable for only small values of p.

As an alternative to exhaustive search over various edges, a penalized maximum likelihood approach in conjunction with VAR models has been used in [20] where the penalty term incorporates sparsity constraints, making it suitable for high-dimensional setting (large p in comparison to N, number of time samples). For every pair of series components, the corresponding partial coherence is thresholded to decide if it is zero (exclude the edge), or nonzero (include the edge). No systematic (or principled) method is given for threshold selection. Nonparametric approaches for graphical modeling of time series in high-dimensional settings (p comparable to or less than sample size) have been formulated in [34], [35] using a neighborhood regression scheme, and in the form of penalized log-likelihood (graphical lasso) in [36]. The approaches of [34]–[36] require

the graph to be sparse ($|\mathcal{E}| \ll p(p-1)/2$), whereas, as in [17], [23], [25], such high degree of sparsity is not assumed in this paper. Furthermore, as in [20], no principled method is given for threshold selection in [34]–[36] to null out small entries of estimated inverse PSD matrix.

Our approach (and that of [17], [23], [25] and others) is designed for dependent, real time series (which includes independent, i.e., i.i.d., real times series), in contrast to [1] (and others) whose results depend crucially on the i.i.d. assumption. The GLRT of [1] for real random vectors (and that of [11] for complex random vectors) requires the time series to be i.i.d. (multiple realizations of the same random vector), otherwise the null distribution of the test statistic would not be true as given in [1]. This, in turn, makes the test of [1] inapplicable to our problem because the test threshold to yield a specified significance level cannot be computed. We address real-valued times series GGMs by introducing a frequency-domain sufficient statistic based on the DFT of the time-domain data. The DFT of a given stretch of measurements is complex, asymptotically jointly proper Gaussian, and independent at distinct frequencies on an appropriate frequency grid [31]. If the PSD of the measurements is constant over a selected frequency interval, the DFTs over this interval act as i.i.d. complex Gaussian measurements, allowing one to invoke the results of [11] and Sec. III pertaining to complex Gaussian random vectors and corresponding CGGMs. Such a formulation permits analytical characterization of the null distribution of the resulting GLRT for real time series GGMs.

III. GLRT for Edge Exclusion in CGGMs

Our objective is to derive the GLRT for the binary hypothesis testing problem specified in (4). But first we need to recall, and then refine and improve upon, some existing results from [11], by following some real-valued GGM results from [1]. Note that the pioneering work of [1] deals with real-valued GGMs for real-valued Gaussian vectors (i.e., i.i.d. real time series), while the pioneering work of [11] investigates complex-valued GGMs for complex-valued Gaussian vectors (i.e., i.i.d. complex time series). The complex-valued results of [11] are essential for our purposes since the DFT of $\{\mathbf{x}(t)\}$ will, in general, be complex-valued (even if $\mathbf{x}(t)$ is real-valued).

Let $\mathbf{y}(t) \in \mathbb{C}^p,\ t = 0, 1, \ldots, N-1$, be N i.i.d. complex, circularly symmetric (proper), Gaussian random vectors; to distinguish replicated random vectors (i.i.d. time series) from dependent time series, we use $\mathbf{y}(t)$ for the former, and $\mathbf{x}(t)$ for the latter. As in [11], define

$$\mathbf{Y} = [\mathbf{y}(0)\ \mathbf{y}(1)\ \cdots\ \mathbf{y}(N-1)]^H . \tag{6}$$

Suppose that $\mathbb{E}\{\mathbf{y}(t)\} = 0$ and $\mathbb{E}\{\mathbf{y}(t)\mathbf{y}^H(t)\} = \mathbf{\Sigma} \in \mathbb{C}^{p\times p}$. Since $\mathbf{y}(t)$ is circularly symmetric (proper), $\mathbb{E}\{\mathbf{y}(t)\mathbf{y}^\top(t)\} = 0$. We denote the distribution of complex-valued Gaussian $\mathbf{y}(t)$ as $\mathcal{N}_c(0, \mathbf{\Sigma})$, i.e., $\mathbf{y}(t) \sim \mathcal{N}_c(0, \mathbf{\Sigma})$, and we assume that $\mathbf{\Sigma}$ is positive definite. The joint probability density function (pdf) of $\mathbf{Y}$ is [11, p. 32]

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{Y}) &= \prod_{t=0}^{N-1} f_{\mathbf{y}(t)}(\mathbf{y}(t)) \\ &= \prod_{t=0}^{N-1} \frac{1}{\pi^p\,|\mathbf{\Sigma}|} \exp\left(-\mathbf{y}^H(t)\mathbf{\Sigma}^{-1}\mathbf{y}(t)\right) \\ &= \frac{1}{\pi^{Np}\,|\mathbf{\Sigma}|^N} \mathrm{etr}(-\mathbf{Y}^H\mathbf{Y}\mathbf{\Omega}) \end{aligned} \tag{7}$$

where $\mathbf{\Omega}$ is the concentration matrix $\mathbf{\Omega} = \mathbf{\Sigma}^{-1}$, $|\mathbf{\Sigma}|$ is the determinant of $\mathbf{\Sigma}$, tr is the trace operator, and we do not distinguish between a random vector/matrix and the values taken by them in our notation (for simplicity).

Consider two undirected graphs $\mathcal{G} = (V, \mathcal{E})$ and $\mathcal{G}' = (V, \mathcal{E}')$ where $|V| = p$, $\mathcal{E}$ is saturated, and $\mathcal{E}' = \mathcal{E}\backslash\{i_0, j_0\}$ for some $i_0 \neq j_0 \in V$. The binary hypothesis problem of interest is as follows. Given observations of i.i.d. $\mathbf{y}(t)$ for $t = 0, 1, \ldots, N-1$, with positive definite covariance matrix $\mathbf{\Sigma}$, we need to distinguish between the null hypothesis $\mathcal{H}_0$ and the alternative $\mathcal{H}_1$:

$$\begin{aligned} &\mathcal{H}_0 :\ \mathbf{\Omega} = \mathbf{\Omega}(\mathcal{G}) \in \mathbb{C}_+(\mathcal{G}'),\ \mathcal{G}' = (V, \mathcal{E}'),\ \mathcal{E}' = \mathcal{E}\backslash\{i_0, j_0\} \\ &\mathcal{H}_1 :\ \mathbf{\Omega} = \mathbf{\Omega}(\mathcal{G}) \in \mathbb{C}_+(\mathcal{G}),\ \mathcal{G} = (V, \mathcal{E}),\ \mathcal{E} \text{ is saturated} . \end{aligned} \tag{8}$$

Define the sample covariance matrix

$$\hat{\mathbf{\Sigma}} = \frac{1}{N}\mathbf{Y}^H\mathbf{Y} = \frac{1}{N}\sum_{t=0}^{N-1} \mathbf{y}(t)\mathbf{y}^H(t) . \tag{9}$$

For $\mathbf{y}(t)$ associated with $\mathcal{G}$ or $\mathcal{G}'$, for $\tilde{V} \subseteq V$, we define $\mathbf{y}_{\tilde{V}}(t) = [y_i(t)]_{i\in\tilde{V}}$ = a column vector of dimension $|\tilde{V}|$. Thus, $\mathbf{y}(t) = \mathbf{y}_V(t)$. The following result is a restatement of [11, Theorem 7.6] in the notation of this paper.

Theorem 1: [11, Theorem 7.6]: The GLRT for the test (8) is given by

$$\mathcal{L}(\mathbf{Y}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau \tag{10}$$

where

$$\begin{aligned} \mathcal{L}(\mathbf{Y}) &= \frac{\sup_{\mathbf{\Omega}(\mathcal{G})\in\mathbb{C}_+(\mathcal{G})}\ f_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}|\mathcal{H}_1)}{\sup_{\mathbf{\Omega}(\mathcal{G})\in\mathbb{C}_+(\mathcal{G}')}\ f_{\mathbf{Y}|\mathcal{H}_0}(\mathbf{Y}|\mathcal{H}_0)} \\ &= \left[\frac{|\hat{\mathbf{\Sigma}}_{V\backslash\{i_0,j_0\}}|\,|\hat{\mathbf{\Sigma}}_V|}{|\hat{\mathbf{\Sigma}}_{V\backslash\{i_0\}}|\,|\hat{\mathbf{\Sigma}}_{V\backslash\{j_0\}}|}\right]^{-N}, \end{aligned} \tag{11}$$

$$\hat{\mathbf{\Sigma}}_V = \frac{1}{N}\sum_{t=0}^{N-1} \mathbf{y}_V(t)\mathbf{y}_V^H(t) = \hat{\mathbf{\Sigma}} \tag{12}$$

$$\hat{\mathbf{\Sigma}}_{V\backslash\{i_0,j_0\}} = \frac{1}{N}\sum_{t=0}^{N-1} \mathbf{y}_{V\backslash\{i_0,j_0\}}(t)\mathbf{y}_{V\backslash\{i_0,j_0\}}^H(t) \tag{13}$$

$$\hat{\mathbf{\Sigma}}_{V\backslash\{i_0\}} = \frac{1}{N}\sum_{t=0}^{N-1} \mathbf{y}_{V\backslash\{i_0\}}(t)\mathbf{y}_{V\backslash\{i_0\}}^H(t) \tag{14}$$

$$\hat{\mathbf{\Sigma}}_{V\backslash\{j_0\}} = \frac{1}{N}\sum_{t=0}^{N-1} \mathbf{y}_{V\backslash\{j_0\}}(t)\mathbf{y}_{V\backslash\{j_0\}}^H(t) . \tag{15}$$

Under $\mathcal{H}_0$, $\mathcal{L}^{-1/N}(\mathbf{Y}) \sim \mathrm{beta}(N-p+1, 1)$, where if $Z \sim \mathrm{beta}(d, 1)$, then its pdf is $f_Z(z) = dz^{d-1}$ for $0 \leq z \leq 1$, and is zero otherwise •

The following results (Lemmas 1 and 2) are novel, and are complex-valued counterparts of the corresponding real-valued GGM results found in [1]. Lemma 1 is the counterpart of [1, Eqn. (5.12), p. 130], and Lemma 2 is the counterpart of [1, Eqn. on p. 150].

Define the expectations

$$\mathbf{\Sigma}_V = \mathbb{E}\{\mathbf{y}_V(t)\mathbf{y}_V^H(t)\} = \mathbf{\Sigma} = \mathbf{\Omega}^{-1} \tag{16}$$

$$\mathbf{\Sigma}_{V\backslash\{i_0,j_0\}} = \mathbb{E}\{\mathbf{y}_{V\backslash\{i_0,j_0\}}(t)\mathbf{y}_{V\backslash\{i_0,j_0\}}^H(t)\} \tag{17}$$

$$\mathbf{\Sigma}_{V\backslash\{i_0\}} = \mathbb{E}\{\mathbf{y}_{V\backslash\{i_0\}}(t)\mathbf{y}_{V\backslash\{i_0\}}^H(t)\} \tag{18}$$

$$\mathbf{\Sigma}_{V\backslash\{j_0\}} = \mathbb{E}\{\mathbf{y}_{V\backslash\{j_0\}}(t)\mathbf{y}_{V\backslash\{j_0\}}^H(t)\}\,. \tag{19}$$

Lemma 1: For the CGGMs under consideration in Theorem 1, we have

$$\frac{|\mathbf{\Sigma}_{V\backslash\{i_0,j_0\}}|\,|\mathbf{\Sigma}_V|}{|\mathbf{\Sigma}_{V\backslash\{i_0\}}|\,|\mathbf{\Sigma}_{V\backslash\{j_0\}}|} = 1 - \frac{|\Omega_{i_0j_0}|^2}{\Omega_{i_0i_0}\Omega_{j_0j_0}} = 1 - \left|\rho_{i_0j_0\,|\,V\backslash\{i_0,j_0\}}\right|^2 \tag{20}$$

where Ω_{ij} is the $(i,j)-$th element of the concentration matrix $\mathbf{\Omega}$,

$$\rho_{i_0j_0\,|\,V\backslash\{i_0,j_0\}} = \frac{-\Omega_{i_0j_0}}{\sqrt{\Omega_{i_0i_0}\Omega_{j_0j_0}}}\,, \tag{21}$$

$$\begin{bmatrix} \Omega_{i_0i_0} & \Omega_{i_0j_0} \\ \Omega_{j_0i_0} & \Omega_{j_0j_0} \end{bmatrix} = \left(\mathrm{var}\left(\begin{bmatrix} x_{i_0}(t) \\ x_{j_0}(t) \end{bmatrix} \,\middle|\, \mathbf{x}_{V\backslash\{i_0,j_0\}}(t)\right)\right)^{-1} \tag{22}$$

and $\rho_{i_0j_0\,|\,V\backslash\{i_0,j_0\}}$ is the partial correlation coefficient, i.e., correlation coefficient between $x_{i_0}(t)$ and $x_{j_0}(t)$ conditioned on $\mathbf{x}_{V\backslash\{i_0,j_0\}}(t)$ •

Proof: Consider the correspondence of the following conformably partitioned matrices

$$\mathbf{\Omega} = \begin{bmatrix} \Omega_{i_0i_0} & \vdots & * \\ \cdots & \cdots & \cdots \\ & \vdots & * \end{bmatrix} = \mathbf{\Sigma}^{-1} = \begin{bmatrix} \Sigma_{i_0} & \vdots & * \\ \cdots & \cdots & \cdots \\ & \vdots & * \end{bmatrix}^{-1}$$

It follows from [11, Theorem A.11] that ($\mathbf{\Sigma}_{\{i_0\}} = \Sigma_{i_0}$, etc.)

$$\Omega_{i_0i_0} = \left[\Sigma_{i_0} - \mathbf{\Sigma}_{i_0,V\backslash\{i_0\}}\mathbf{\Sigma}_{V\backslash\{i_0\}}^{-1}\mathbf{\Sigma}_{V\backslash\{i_0\},i_0}\right]^{-1}$$

By [11, Theorem 7.1]

$$\Sigma_{i_0\,|\,V\backslash\{i_0\}} = \Sigma_{i_0} - \mathbf{\Sigma}_{i_0,V\backslash\{i_0\}}\mathbf{\Sigma}_{V\backslash\{i_0\}}^{-1}\mathbf{\Sigma}_{V\backslash\{i_0\},i_0},$$

therefore

$$\Omega_{i_0i_0} = \Sigma_{i_0\,|\,V\backslash\{i_0\}}^{-1} = \left[\mathrm{var}(x_{i_0}(t)\,|\mathbf{x}_{V\backslash\{i_0\}}(t))\right]^{-1} \tag{23}$$

Define

$$\tilde{\mathbf{\Omega}} = \begin{bmatrix} \Omega_{i_0i_0} & \Omega_{i_0j_0} \\ \Omega_{j_0i_0} & \Omega_{j_0j_0} \end{bmatrix}$$

and consider

$$\mathbf{\Omega} = \begin{bmatrix} \tilde{\mathbf{\Omega}} & \vdots & * \\ \cdots & \cdots & \cdots \\ & \vdots & * \end{bmatrix} = \begin{bmatrix} \Sigma_{i_0} & \Sigma_{i_0j_0} & \vdots & * \\ \Sigma_{j_0i_0} & \Sigma_{j_0} & & \\ & \cdots & \cdots & \cdots \\ & & \vdots & * \end{bmatrix}^{-1}$$

As for (23), it follows from [11, Theorems 7.1, A.11] that

$$\tilde{\mathbf{\Omega}} = \mathbf{\Sigma}_{\{i_0,j_0\}\,|\,V\backslash\{i_0,j_0\}}^{-1}$$

leading to ($\tilde{V} := V\backslash\{i_0,j_0\}$)

$$\mathbf{\Sigma}_{\{i_0,j_0\}\,|\,V\backslash\{i_0,j_0\}} = \begin{bmatrix} \Sigma_{i_0\,|\,\tilde{V}} & \Sigma_{i_0j_0\,|\,\tilde{V}} \\ \Sigma_{j_0i_0\,|\,\tilde{V}} & \Sigma_{j_0\,|\,\tilde{V}} \end{bmatrix} \tag{24}$$

$$= \frac{1}{|\tilde{\mathbf{\Omega}}|}\begin{bmatrix} \Omega_{j_0j_0} & -\Omega_{i_0j_0} \\ -\Omega_{j_0i_0} & \Omega_{i_0i_0} \end{bmatrix}. \tag{25}$$

By [11, Theorems A.11] concerning determinant of partitioned matrices, we have

$$|\mathbf{\Sigma}_V| = |\mathbf{\Sigma}_{V\backslash\{i_0\}}|\,\Sigma_{i_0\,|\,V\backslash\{i_0\}}. \tag{26}$$

Similarly, we have

$$|\mathbf{\Sigma}_{V\backslash\{j_0\}}| = |\mathbf{\Sigma}_{V\backslash\{i_0,j_0\}}|\,\Sigma_{i_0\,|\,V\backslash\{i_0,j_0\}}. \tag{27}$$

From (24) and (25), it follows that

$$\Sigma_{i_0\,|\,V\backslash\{i_0,j_0\}} = \frac{\Omega_{j_0j_0}}{|\tilde{\mathbf{\Omega}}|} = \frac{\Omega_{j_0j_0}}{\Omega_{i_0i_0}\Omega_{j_0j_0} - \Omega_{i_0j_0}\Omega_{j_0i_0}} \tag{28}$$

Therefore, using (27) and (28), we have

$$\frac{|\mathbf{\Sigma}_{V\backslash\{i_0,j_0\}}|\,|\mathbf{\Sigma}_V|}{|\mathbf{\Sigma}_{V\backslash\{i_0\}}|\,|\mathbf{\Sigma}_{V\backslash\{j_0\}}|} = \frac{|\mathbf{\Sigma}_V|\,|\tilde{\mathbf{\Omega}}|}{|\mathbf{\Sigma}_{V\backslash\{i_0\}}|\,\Omega_{j_0j_0}} = \frac{|\tilde{\mathbf{\Omega}}|}{\Omega_{i_0i_0}\Omega_{j_0j_0}} \tag{29}$$

$$= 1 - \frac{|\Omega_{i_0j_0}|^2}{\Omega_{i_0i_0}\Omega_{j_0j_0}} \tag{30}$$

where, in the second equality in (29), we have used (23) and (26), and in the equality in (30), we note that since $\mathbf{\Omega}$ is Hermitian, $\Omega_{i_0j_0} = \Omega_{j_0i_0}^*$. Finally, (21) follows from (24)–(25). This proves the desired result ■

Assume that $N \geq p$, so that $\hat{\mathbf{\Sigma}}^{-1}$ exists w.p.1. Lemma 2 then follows just as Lemma 1, by manipulations of $\hat{\mathbf{\Sigma}}_V^{-1}$ and submatrices of $\hat{\mathbf{\Sigma}}_V$.

Lemma 2: For the CGGMs under consideration in Lemma 1, we have

$$\frac{|\hat{\mathbf{\Sigma}}_{V\backslash\{i_0,j_0\}}|\,|\hat{\mathbf{\Sigma}}_V|}{|\hat{\mathbf{\Sigma}}_{V\backslash\{i_0\}}|\,|\hat{\mathbf{\Sigma}}_{V\backslash\{j_0\}}|} = 1 - \left|\hat{\rho}_{i_0j_0\,|\,V\backslash\{i_0,j_0\}}\right|^2 \tag{31}$$

where $\hat{\rho}_{i_0j_0\,|\,V\backslash\{i_0,j_0\}}$, the empirical partial correlation coefficient, is given by

$$\hat{\rho}_{i_0j_0\,|\,V\backslash\{i_0,j_0\}} = \frac{-\hat{\Omega}_{i_0j_0}}{\sqrt{\hat{\Omega}_{i_0i_0}\hat{\Omega}_{j_0j_0}}}\,, \tag{32}$$

and $\hat{\mathbf{\Omega}} = \hat{\mathbf{\Sigma}}^{-1}$ •

From Theorem 1 and Lemma 2, one has the log-GLRT statistic as

$$\ln(\mathcal{L}(\mathbf{Y})) = -N \ln\left(1 - \left|\hat{\rho}_{i_0 j_0 \,|\, V\setminus\{j_0\}}\right|^2\right) \sim \exp\left(\frac{N-p+1}{N}\right) \text{ under } \mathcal{H}_0\,, \tag{33}$$

where we have used the fact that if $Z \sim \text{beta}(\alpha, 1)$, then $-\ln(Z) \sim \exp(\alpha)$, and $-N\ln(Z) \sim \exp(\frac{\alpha}{N})$. If $Z \sim \exp(\lambda)$, then $f_Z(z) = \lambda e^{-\lambda z}$, $z \geq 0$.

This leads to the following GLRT counterpart to the real GGM GLRT found in [1, Eqn. (5.12), p. 130] or [26]:

$$-N \ln\left(1 - \left|\hat{\rho}_{i_0 j_0 \,|\, V\setminus\{i_0,j_0\}}\right|^2\right) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau_1 \tag{34}$$

where $\tau_1 = -(N/(N-p+1))\ln(P_{fa})$ to achieve the probability of false alarm (significance level) P_{fa}. An equivalent GLRT is given by

$$\left|\hat{\rho}_{i_0 j_0 \,|\, V\setminus\{i_0,j_0\}}\right|^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau_2 = 1 - \exp(-\tau_1/N). \tag{35}$$

This is the CGGM counterpart of real GGM test of [27]. We summarize the above results in Theorem 2.

Theorem 2: The GLRT for the test (8) can be expressed as (34) or (35), where, under $\mathcal{H}_0$,

$$-N \ln\left(1 - \left|\hat{\rho}_{i_0 j_0 \,|\, V\setminus\{i_0,j_0\}}\right|^2\right) \sim \exp\left(\frac{N-p+1}{N}\right)$$

and

$$\left|\hat{\rho}_{i_0 j_0 \,|\, V\setminus\{i_0,j_0\}}\right|^2 \sim \text{beta}(1, N-p+1) \quad \bullet$$

A. Flop Count Comparison Between Test Statistics of Theorems 1 and 2

We will compare computation of the left-side of (31) with its right-side. Suppose that we need to carry out exclusion tests for all $p(p-1)/2$ edges in $\mathcal{E}$. The determinant $|\hat{\mathbf{\Sigma}}_V|$ is computed just once and needs $\mathcal{O}(p^3)$ flops, whereas, while $|\hat{\mathbf{\Sigma}}_{V\setminus\{i_0,j_0\}}|$ needs $\mathcal{O}((p-2)^3)$ flops, it has be computed for each of the $p(p-1)/2$ distinct edges, leading to a total of $\mathcal{O}(p^5)$ flops. To compute the right-side of (31) via (32), $\hat{\mathbf{\Sigma}}$ needs to be inverted just once, with complexity $\mathcal{O}(p^3)$, and then it remains to compute $|\hat{\rho}_{i_0 j_0 \,|\, V\setminus\{i_0,j_0\}}|^2$ for each of the distinct edges, with complexity $\mathcal{O}(p^2)$. Hence, the computational complexity of our proposed alternative statistic is $\mathcal{O}(p^3)$ versus $\mathcal{O}(p^5)$ for the expression of [11].

B. Power of GLRT

The pdf of $\left|\hat{\rho}_{i_0 j_0 \,|\, V\setminus\{j_0\}}\right|^2$ under $\mathcal{H}_1$ can be specified as in Theorem 3, following [31, p. 295]. First recall the definition of a hypergeometric function

$${}_2F_1(a,b;c,x) = \sum_{j=0}^{\infty} \frac{\Gamma(a+j)}{\Gamma(a)} \frac{\Gamma(b+j)}{\Gamma(b)} \frac{\Gamma(c)}{\Gamma(c+j)} \frac{x^j}{j!}\,. \tag{36}$$

Theorem 3: For the CGGMs under consideration in Theorem 1, let $\rho^{(*)}_{i_0 j_0 \,|\, V\setminus\{i_0,j_0\}}$ (>0), denote the true partial correlation coefficient under $\mathcal{H}_1$. Then the pdf of $z := \left|\hat{\rho}_{i_0 j_0 \,|\, V\setminus\{i_0,j_0\}}\right|^2$ under $\mathcal{H}_1$ is given by $(0 \leq z \leq 1)$

$$f_z(z) = (N-p+1)\left(1 - |\rho^{(*)}_{i_0 j_0 \,|\, V\setminus\{i_0,j_0\}}|^2\right)^{N-p+2} (1-z)^{N-p} \times {}_2F_1\left(N-p+2, N-p+2; 1; z\,|\rho^{(*)}_{i_0 j_0 \,|\, V\setminus\{i_0,j_0\}}|^2\right). \tag{37}$$

The probability of detection (power of the GLRT) is then given by $P_D = \int_{\tau_2}^1 f_z(z)\,dz$ where τ_2 is as in (35) corresponding to a specified P_{fa} $\bullet$

Proof: The pdf (37) follows from [31, p. 295] (see discussion following (8.2.59) therein). The expression for P_D follows from the equivalent test (35) ■

Using (36), (37), and the fact that $\Gamma(N+1) = N!$ for integer N, $f_z(z)$ can be expressed as

$$f_z(z) = (N-p+1)(1 - |\rho^{(*)}|^2)^{N-p+2}(1-z)^{N-p} \times \sum_{j=0}^{\infty} \left(\frac{(N-p+1+j)!}{(N-p+1)!\,j!}\right)^2 \left(z|\rho^{(*)}|^2\right)^j$$

where we have used the shorthand notation $\rho^{(*)}$ for $\rho^{(*)}_{i_0 j_0 \,|\, V\setminus\{i_0,j_0\}}$. There is no closed-form solution for P_D, but we can calculate it numerically.

The log-GLRT statistic (33) for complex GGM hypothesis test problem (8) proves to be useful in solving the time series GGM hypothesis test problem (4), discussed next.

IV. GLRT for Edge Exclusion in Time Series GGMs

We now return to the hypothesis testing problem (4) for time series, given $\mathbf{x}(t)$ for $t = 0, 1, 2, \ldots, N-1$. In Sec. IV-A we introduce a frequency-domain sufficient statistic based on the DFT of the time-domain data. This approach is novel. For a large class of stationary random processes, the DFT of a given stretch of measurements is complex, asymptotically jointly proper Gaussian, and independent at distinct frequencies on an appropriate frequency grid (Sec. IV-C). If the PSD of the measurements is constant over a selected frequency interval, the DFTs over this interval act as independent complex Gaussian measurements, allowing one to invoke the results of Sec. III.

A. Frequency-Domain Sufficient Statistic

Consider the (normalized) DFT $\mathbf{d}_x(f_n)$ of real-valued $\mathbf{x}(t)$, $t = 1, 2, \ldots, N-1$, given by

$$\mathbf{d}_x(f_n) := \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} \mathbf{x}(t) e^{-j2\pi f_n t} \tag{38}$$

where $f_n = n/N$, $n = 0, 1, \ldots, N-1$. Since $\{\mathbf{x}(t)\}$ is Gaussian, so is $\mathbf{d}_x(f_n)$. Note that $\mathbf{d}_x(f_n)$ is periodic in n with period N, and is periodic in normalized frequency f_n with period 1. Since $\mathbf{x}(t)$ is real-valued, we have $\mathbf{d}_x^*(f_n) = \mathbf{d}_x(-f_n) = \mathbf{d}_x(1-f_n)$, so $\mathbf{d}_x(f_n)$ for $n = 0, 1, \ldots, (N/2)$, (N even), completely determines $\mathbf{d}_x(f_n)$ for all integers n.

As proved in [37, p. 280, Sec. 6.2], for any statistical inference problem, the complete sample is a sufficient statistic, and so is any one-to-one function of a sufficient statistic. Since the inverse

DFT yields (one-to-one transformation)

$$\mathbf{x}(t) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \mathbf{d}_x(f_n) e^{j2\pi f_n t}, \tag{39}$$

the set $\{\mathbf{d}_x(f_n)\}_{n=0}^{N-1}$ is a sufficient statistic, which can be further reduced to $\{\mathbf{d}_x(f_n)\}_{n=0}^{N/2}$ since $\mathbf{x}(t)$ is real-valued, inducing symmetries $\mathbf{d}_x^*(f_n) = \mathbf{d}_x(-f_n) = \mathbf{d}_x(1-f_n)$. Thus, the set of complex-valued random vectors $\{\mathbf{d}_x(f_n)\}_{n=0}^{N/2}$ is a sufficient statistic for our binary hypothesis testing problem (4).

B. Model Assumptions

Here we formally state assumptions on $\mathbf{x}(t)$ used later in the paper.

A1) (Specialization of **Assumption 2.6.1** of [31] to Gaussian sequences). Zero-mean, real-valued Gaussian random sequence $\{\mathbf{x}(t)\} \in \mathbb{R}^p$ is stationary with components $\mathbf{x}_a(t)$, $a = 1, 2, \ldots, p$, such that $\mathbb{E}\{|\mathbf{x}_a(t)|^2\} < \infty$, satisfying

$$\sum_{\tau=-\infty}^{\infty} |\mathbb{E}\{\mathbf{x}_{a_1}(t)\mathbf{x}_{a_2}(t-\tau)\}| < \infty \tag{40}$$

for $a_1, a_2 = 1, 2, \ldots, p$.

A2) The PSD matrix of $\mathbf{x}(t)$, $\mathbf{S}_x(f) \succ 0$ for any $0 \le f \le 1$.

A3) The PSD matrix $\mathbf{S}_x(f_n)$ is locally smooth such that it is constant over K ($\ge p$) consecutive frequency points, where $f_n = n/N$, $n \in [0, N-1]$.

Assumption (A1) is needed so that some asymptotic results from [31] regarding the DFT $\mathbf{d}_x(f_n)$ and PSD estimators can be invoked. Regarding assumption (A2), one can always add artificial proper white Gaussian noise to $\mathbf{x}(t)$ to achieve $\mathbf{S}_x(f) \succ 0$. Assumption (A3) is a standard assumption in the PSD estimation literature [31].

C. Asymptotic Distribution of Sufficient Statistic

Now we consider asymptotic distribution (as $N \to \infty$) of $\{\mathbf{d}_x(f_n)\}_{n=0}^{N/2}$, based upon [31, Chap. 4]. Under the assumption that $\{\mathbf{x}(t)\}$ satisfies Assumption 2.6.1 of [31] which specializes to assumption (A1) of Sec. IV-B, it follows from [31, Theorem 4.4.1] that asymptotically, $\mathbf{d}_x(f_n)$, $n = 1, 2, \ldots, (N/2)-1$, ($N$ even), are independent proper (i.e., circularly symmetric), complex Gaussian $\mathcal{N}_c(\mathbf{0}, \mathbf{S}_x(f_n))$ random vectors, respectively. Also, asymptotically, $\mathbf{d}_x(f_0)$ and $\mathbf{d}_x(f_{N/2})$, (N even), are independent real Gaussian $\mathcal{N}_r(\mathbf{0}, \mathbf{S}_x(f_0))$ and $\mathcal{N}_r(\mathbf{0}, \mathbf{S}_x(f_{N/2}))$ random vectors, respectively, independent of $\mathbf{d}_x(f_n)$, $n \in [1, (N/2)-1]$. In our derivation of the GLRT in the next section, we will ignore these two frequency points.

D. GLRT

We now derive the GLRT for the hypothesis testing problem (4). We will use the frequency-domain sufficient statistic $\{\mathbf{d}_y(f_n)\}_{n=0}^{N/2}$ after discarding the end frequency points f_0 and $f_{N/2}$, as the asymptotic distribution of the latter is real Gaussian, whereas the DFT at other frequencies is proper complex Gaussian.

Define

$$\mathbf{D} = \left[\mathbf{d}_x(f_1) \ \cdots \ \mathbf{d}_x(f_{(N/2)-1})\right]^H \in \mathbb{C}^{((N/2)-1)\times p}. \tag{41}$$

The asymptotic joint pdf of $\mathbf{D}$ is given by

$$f_{\mathbf{D}}(\mathbf{D}) = \prod_{n=1}^{(N/2)-1} \frac{\exp\left(-\mathbf{d}_x^H(f_n)\mathbf{S}_x^{-1}(f_n)\mathbf{d}_x(f_n)\right)}{\pi^p |\mathbf{S}_x(f_n)|}. \tag{42}$$

Now assume that $\mathbf{S}_x(f_n)$ is locally smooth (assumption (A3) of Sec. IV-B), so that $\mathbf{S}_x(f_n)$ is (approximately) constant over $K = 2m_t + 1 \ge p$ consecutive frequency points. Pick

$$\tilde{f}_k = \frac{(k-1)K + m_t + 1}{N}, \quad k = 1, 2, \ldots, M, \tag{43}$$

$$M = \left\lfloor \frac{\frac{N}{2} - m_t - 1}{K} \right\rfloor, \tag{44}$$

leading to M equally spaced frequencies $\tilde{f}_k$ in the interval (0, 0.5), at intervals of K/N. It is assumed that for each $\tilde{f}_k$ (local smoothness),

$$\mathbf{S}_x(\tilde{f}_{k,\ell}) = \mathbf{S}_x(\tilde{f}_k) \ \text{ for } \ell = -m_t, -m_t+1, \ldots, m_t, \tag{45}$$

where

$$\tilde{f}_{k,\ell} = \frac{(k-1)K + m_t + 1 + \ell}{N}. \tag{46}$$

Using (45) in (42), we have

$$f_{\mathbf{D}}(\mathbf{D}) = \prod_{k=1}^{M} \left[\prod_{\ell=-m_t}^{m_t} \frac{1}{\pi^p |\mathbf{S}_x(\tilde{f}_k)|} \times \exp\left(-\mathbf{d}_x^H(\tilde{f}_{k,\ell})\mathbf{S}_x^{-1}(\tilde{f}_k)\mathbf{d}_x(\tilde{f}_{k,\ell})\right) \right] \tag{47}$$

$$= \prod_{k=1}^{M} \frac{\mathrm{etr}(-\tilde{\mathbf{D}}(\tilde{f}_k)\mathbf{S}_x^{-1}(\tilde{f}_k))}{\pi^{Kp} |\mathbf{S}_x(\tilde{f}_k)|^K}$$

$$= \prod_{k=1}^{M} f_{\check{\mathbf{D}}(\tilde{f}_k)}(\check{\mathbf{D}}(\tilde{f}_k)) \tag{48}$$

where $K \times p$ $\check{\mathbf{D}}(\tilde{f}_k)$ is

$$\check{\mathbf{D}}(\tilde{f}_k) = \left[\mathbf{d}_x(\tilde{f}_{k,-m_t}) \ \mathbf{d}_x(\tilde{f}_{k,-m_t+1}) \ \cdots \ \mathbf{d}_x(\tilde{f}_{k,m_t})\right]^H, \tag{49}$$

and $p \times p$ $\tilde{\mathbf{D}}(\tilde{f}_k)$ is

$$\tilde{\mathbf{D}}(\tilde{f}_k) = \sum_{\ell=-m_t}^{m_t} \mathbf{d}_x(\tilde{f}_{k,\ell})\mathbf{d}_x^H(\tilde{f}_{k,\ell}). \tag{50}$$

Explicitly indicating the dependence on the underlying hypothesis $\mathcal{H}_i$, $i = 0, 1$, the pdf of $\mathbf{D}$ under $\mathcal{H}_i$ is given by

$$f_{\mathbf{D}|\mathcal{H}_i}(\mathbf{D}|\mathcal{H}_i) = \prod_{k=1}^{M} f_{\check{\mathbf{D}}(\tilde{f}_k)|\mathcal{H}_i}(\check{\mathbf{D}}(\tilde{f}_k)|\mathcal{H}_i) \tag{51}$$

where

$$f_{\check{\mathbf{D}}(\tilde{f}_k)|\mathcal{H}_i}(\check{\mathbf{D}}(\tilde{f}_k)|\mathcal{H}_i) = \frac{\mathrm{etr}(-\tilde{\mathbf{D}}(\tilde{f}_k)\mathbf{S}_x^{-1}(\tilde{f}_k))}{\pi^{Kp} |\mathbf{S}_x(\tilde{f}_k)|^K}. \tag{52}$$

Under local smoothness (assumption (A3) of Sec. IV-B) of $\mathbf{S}_x(f_n)$, the hypothesis testing problem (4) is reformulated as

$$\begin{aligned}\mathcal{H}_0 \,:\, \mathbf{S}_x^{-1}(\tilde{f}_k) &= \mathbf{S}_x^{-1}(\tilde{f}_k;\mathcal{G}) \in \mathbb{C}_+(\mathcal{G}')\ \forall\, k \in [1,M],\\ \mathcal{G}' &= (V,\mathcal{E}'),\ \mathcal{E}' = \mathcal{E}\backslash\{i_0,j_0\}\\ \mathcal{H}_1 \,:\, \mathbf{S}_x^{-1}(\tilde{f}_k) &= \mathbf{S}_x^{-1}(\tilde{f}_k;\mathcal{G}) \in \mathbb{C}_+(\mathcal{G})\ \forall\, k \in [1,M],\\ \mathcal{G} &= (V,\mathcal{E}),\ \mathcal{E} \text{ is saturated}\end{aligned} \tag{53}$$

Based on (51), (52) and (53), the generalized likelihood ratio $\mathcal{L}(\mathbf{D})$ is given by

$$\begin{aligned}\mathcal{L}(\mathbf{D}) &= \frac{\sup_{\mathbf{S}_x^{-1}(\tilde{f}_k;\mathcal{G})\in\mathbb{C}_+(\mathcal{G}),\, k=1,2,\ldots,M}\ f_{\mathbf{D}|\mathcal{H}_1}(\mathbf{D}|\mathcal{H}_1)}{\sup_{\mathbf{S}_x^{-1}(\tilde{f}_k;\mathcal{G})\in\mathbb{C}_+(\mathcal{G}'),\, k=1,2,\ldots,M}\ f_{\mathbf{D}|\mathcal{H}_0}(\mathbf{D}|\mathcal{H}_0)}\\ &= \prod_{k=1}^{M} \frac{\sup_{\mathbf{S}_x^{-1}(\tilde{f}_k;\mathcal{G})\in\mathbb{C}_+(\mathcal{G})}\ f_{\check{\mathbf{D}}(\tilde{f}_k)|\mathcal{H}_1}(\check{\mathbf{D}}(\tilde{f}_k)|\mathcal{H}_1)}{\sup_{\mathbf{S}_x^{-1}(\tilde{f}_k;\mathcal{G})\in\mathbb{C}_+(\mathcal{G}')}\ f_{\check{\mathbf{D}}(\tilde{f}_k)|\mathcal{H}_0}(\check{\mathbf{D}}(\tilde{f}_k)|\mathcal{H}_0)}\,.\end{aligned} \tag{54}$$

Comparing (52) with (7) and noting the correspondences $K \leftrightarrow N, \tilde{\mathbf{D}}(\tilde{f}_k) \leftrightarrow \mathbf{Y}^H\mathbf{Y}$ and $\mathbf{S}_x^{-1}(\tilde{f}_k)) \leftrightarrow \mathbf{\Omega} = \mathbf{\Sigma}^{-1}$, it follows from Theorem 1 and Lemma 2 that

$$\begin{aligned}&\frac{\sup_{\mathbf{S}_x^{-1}(\tilde{f}_k;\mathcal{G})\in\mathbb{C}_+(\mathcal{G})}\ f_{\check{\mathbf{D}}(\tilde{f}_k)|\mathcal{H}_1}(\check{\mathbf{D}}(\tilde{f}_k)|\mathcal{H}_1)}{\sup_{\mathbf{S}_x^{-1}(\tilde{f}_k;\mathcal{G})\in\mathbb{C}_+(\mathcal{G}')}\ f_{\check{\mathbf{D}}(\tilde{f}_k)|\mathcal{H}_0}(\check{\mathbf{D}}(\tilde{f}_k)|\mathcal{H}_0)}\\ &= \left[1 - \left|\hat{\rho}^{(s)k}_{i_0j_0\,|\,V\backslash\{i_0,j_0\}}\right|^2\right]^{-K}\end{aligned} \tag{55}$$

where empirical partial coherence $\left|\hat{\rho}^{(s)k}_{i_0j_0\,|\,V\backslash\{i_0,j_0\}}\right|^2$ is given by

$$\left|\hat{\rho}^{(s)k}_{i_0j_0\,|\,V\backslash\{i_0,j_0\}}\right|^2 = \frac{\left|[\hat{\mathbf{S}}_x^{-1}(\tilde{f}_k)]_{i_0j_0}\right|^2}{[\hat{\mathbf{S}}_x^{-1}(\tilde{f}_k)]_{i_0i_0}[\hat{\mathbf{S}}_x^{-1}(\tilde{f}_k)]_{j_0j_0}}\,, \tag{56}$$

and $\hat{\mathbf{S}}_x(\tilde{f}_k)$ is given by

$$\hat{\mathbf{S}}_x(\tilde{f}_k) = \frac{1}{K}\sum_{\ell=-m_t}^{m_t} \mathbf{d}_x(\tilde{f}_{k,\ell})\mathbf{d}_x^H(\tilde{f}_{k,\ell})\,. \tag{57}$$

Note that $\hat{\mathbf{S}}_x(\tilde{f}_k)$ in (57) is an estimator of the PSD of $\mathbf{x}(t)$ at frequency $\tilde{f}_k$, based on unweighted smoothing in frequency-domain (also known as the Daniell method), as given in [31, Eqn. (7.3.2)].

Using (33), Theorem 1, (54) and (55), we obtain Theorem 4.

Theorem 4: The GLRT for the test (53) is given by

$$\begin{aligned}\ln(\mathcal{L}(\mathbf{D})) = &-K\sum_{k=1}^{M}\ln\left(1 - \left|\hat{\rho}^{(s)k}_{i_0j_0\,|\,V\backslash\{i_0,j_0\}}\right|^2\right)\\ &\underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau\end{aligned} \tag{58}$$

where $\left|\hat{\rho}^{(s)k}_{i_0j_0\,|\,V\backslash\{i_0,j_0\}}\right|^2$ is given by (56). Under $\mathcal{H}_0$,

$$\ln(\mathcal{L}(\mathbf{D})) \sim \text{Gamma}\left(M, \frac{K-p+1}{K}\right) \tag{59}$$

where if $Z \sim \text{Gamma}(M,\lambda)$, then $f_Z(z) = \lambda e^{-\lambda z}(\lambda z)^{M-1}/(M-1)!$, $z \geq 0$, and is zero otherwise, resulting from sum of i.i.d. exponential random variables $Z = \sum_{i=1}^{M} Z_i$, $Z_i \sim \exp(\lambda)$ •

E. Flop Count Comparison Between Test Statistics of Matsuda [25] (as Used in [23]) and Theorem 4

1) Matsuda's Test Statistic: Matsuda [25] proposed an edge exclusion test statistic based on the Kullback-Leibler divergence between two graphical models, for graphical modeling of time series. It is used in [25] to devise an iterative procedure for model selection where, in each step, the null hypothesis that a subgraph with one fewer edge is correct, is tested. It is shown in [23] that for model selection, it is sufficient to consider tests that compare the saturated graph (alternative hypothesis) with graphs that have exactly one edge missing (null hypothesis). For edge exclusion tests, [23] uses Matsuda's statistic, but it follows a different iterative model identification method (sequence of edge exclusion tests). Our proposed GLRT for edge exclusion can be used in the model selection method of [23], replacing Matsuda's test statistic. One needs $p(p-1)/2$ edge exclusion tests to test all edges in an undirected graph in [23], whereas Matsuda's graphical model selection method needs $\mathcal{O}(p^4)$ number of edge exclusion tests [23].

We now describe Matsuda's test statistic as used in [23]. Given $\mathbf{x}(t)$, $t = 0, 1, \ldots, N-1$.

- Calculate $\mathbf{d}_x(f_n)$ for $n \in [0, N-1]$ as in (38). Estimate $\mathbf{S}_x(f_n)$ as

$$\hat{\mathbf{S}}_x(f_n) = \sum_{k=-M_s}^{M_s} w_k \mathbf{d}_x(f_{n-k})\mathbf{d}_x^H(f_{n-k})\,,\quad f_n = n/N, \tag{60}$$

  where $w_k = w_{-k} > 0, \sum_{k=-M_s}^{M_s} w_k = 1$, and $2M_s + 1 \geq p$.
- Given saturated graph $\mathcal{G} = (V,\mathcal{E})$ and subgraph under null hypothesis $\mathcal{G}' = (V,\mathcal{E}')$ with one missing edge, calculate $\mathbf{T}(f_n)$ and $\mathbf{T}'(f_n)$ in $\mathbb{C}_+^{p\times p}$ such that $\mathbf{T}(f_n) = \hat{\mathbf{S}}_x(f_n)$, and

$$[\mathbf{T}'(f_n)]_{jk} = \begin{cases}[\hat{\mathbf{S}}_x(f_n)]_{jk} & \text{if } \{j,k\} \in \mathcal{E}'\\ 0 & \text{if } \{j,k\} \notin \mathcal{E}'\,.\end{cases} \tag{61}$$

- The Kullback-Leibler divergence is then estimated as

$$\begin{aligned}eKL(\mathbf{T},\mathbf{T}') = \frac{1}{N}\sum_{n=1}^{N/2}\Big[&\text{tr}\left(\mathbf{T}(f_n)(\mathbf{T}'(f_n))^{-1}\right)\\ &- \ln\left|\mathbf{T}(f_n)(\mathbf{T}'(f_n))^{-1}\right| - p\Big]\,.\end{aligned} \tag{62}$$

- The edge exclusion test statistic $Z_N(\mathbf{T},\mathbf{T}')$ is given by

$$Z_N(\mathbf{T},\mathbf{T}') = \sqrt{\frac{2M_sN}{D_u}}\left[eKL(\mathbf{T},\mathbf{T}') - \frac{C_u}{2M_s}\right] \tag{63}$$

  where D_u and C_u are constants depending upon $w_k = u(k/(2M_s)), k \in [-M_s, M_s]$. Both [25] and [23] use $w_k = \cos(\pi k/(2M_s))$, leading to $D_u = 0.446$ and $C_u = 0.617$. As $N \to \infty$, $Z_N(\mathbf{T},\mathbf{T}') \sim \mathcal{N}_r(0,1)$ under $\mathcal{H}_0$.

2) Flop Count: Here we discuss computational complexity comparisons in terms of flop count between the test statistics (58) and (63); the count for the latter is also partially available in [23, Sec. V]. Computation of $\mathbf{d}_x(f_n)$ via fast

Fourier transform (FFT) needs $pN\log_2 N$ flops, and that of $\mathbf{d}_x(f_n)\mathbf{d}_x^H(f_n)$ needs $p^2N/2$ flops for $n \in [1, (N/2)-1]$. Since $M = \mathcal{O}(N/(2K)) = \mathcal{O}(N/(4m_t))$ and inversion $\hat{\mathbf{S}}_x^{-1}(\tilde{f}_k)$ needs $\mathcal{O}(p^3)$ flops, one needs number of flops of the order of $p^3N/(4m_t)$ to compute PSD inverses over M frequencies. These computations are independent of the edge $\{i_0, j_0\}$ being tested. Finally, there are $\mathcal{O}(p^2/2)$ edges to be tested and each test needs 2 flops for partial coherence per frequency point, leading to $p^2M = p^2N/(4m_t)$ flops to test all edges. Thus, total number of flops needed to test all edges using the proposed GLRT of Theorem 3 is of the order of

$$F_{GLRT} = pN\log_2 N + p^2\frac{N}{2} + p^3\frac{N}{4m_t} + p^2\frac{N}{4m_t} \tag{64}$$

$$= \mathcal{O}\left(p^3\frac{N}{m_t} + p^2N\right) = \mathcal{O}\left(p^3N\right) \tag{65}$$

Similarly, to calculate $\mathbf{d}_x(f_{n-k})\mathbf{d}_x^H(f_{n-k}) \in \mathbb{C}^{p\times p}$ in (60) for all f_n s, one needs $pN\log_2 N + p^2\frac{N}{2}$ flops, and to compute the product $w_k\mathbf{d}_x(f_{n-k})\mathbf{d}_x^H(f_{n-k})\,\forall k$, one needs another $2M_sNp^2$ flops. These computations are independent of the edge $\{i_0, j_0\}$ being tested. The matrix $\mathbf{T}'(f_n)$ in (61) is edge-dependent. Each trace in (62) needs $\mathcal{O}(p^2)$ flops, and since $\left|\mathbf{T}(f_n)(\mathbf{T}'(f_n))^{-1}\right| = |\mathbf{T}(f_n)| \times |\mathbf{T}'(f_n)|^{-1}$, and $\mathbf{T}(f_n)$ is the same for all edges (so we ignore it), each determinant in (62) needs $\mathcal{O}(p^3)$ flops. Combining this with number of frequencies $N/2$ and number of edges $\mathcal{O}(p^2/2)$, one needs $(N/2)(p^2/2)[\mathcal{O}(p^2) + \mathcal{O}(p^3)] = Np^4/4 + Np^5/4$ flops for (62). Thus, total number of flops needed to test all edges using Matsuda's statistic (62) is of the order of

$$F_{Matsuda} = pN\log_2 N + p^2\frac{N}{2} + 2M_sNp^2 + p^5\frac{N}{2} + p^4\frac{N}{4} \tag{66}$$

$$= \mathcal{O}\left(p^5N + M_sNp^2\right) = \mathcal{O}\left(p^5N\right). \tag{67}$$

In [23, Sec. V], the computational complexity of Matsuda's statistic is calculated as $\mathcal{O}\left(p^4N\right)$. It is stated therein that "It can be shown by considering the steps in the construction process that computation time for each statistic is $\sim 2NM_s + p^2N\ell_k$" where $\ell_k = 1$ for the Matsuda method. Combining this with $\mathcal{O}(p^2/2)$ edges, [23, Sec. V] arrives at complexity $\mathcal{O}\left(p^4N\right)$. More details are available in [38, Prop. 17, p. 48], where complexity of computing $\left|\mathbf{T}(f_n)(\mathbf{T}'(f_n))^{-1}\right|$ has been ignored. It is correctly stated therein ([38, Remark 35, p. 48]) that "In reality, the calculation of the eKL summand is $\mathcal{O}(p^3)$ with the determinant being the dominant function in terms of complexity (calculated using LU decomposition)." However, $\mathcal{O}(p^3)$ needs to be scaled by $N/2$ (number of frequencies) since the matrices involved are all different at different frequencies, and further scaled by $\mathcal{O}(p^2/2)$ edges to be tested. Taking these two omitted factors into account, we see that the computational complexity is indeed $\mathcal{O}\left(p^5N\right)$.

F. On the Approach of Dahlhaus [17]

Dahlhaus [17] estimates partial coherence using nonparametric PSD estimates (obtained by first applying tapers to $\mathbf{x}(t)$ before computing its DFT, then calculating the periodogram, and finally smoothing the periodogram in the frequency domain). Under the null hypothesis of a given edge missing, the partial coherence is zero, else greater than zero. Each edge is individually tested based on the estimated partial coherence. The test statistic used in [17] is the supremum of estimated partial coherence over all frequencies, which, in the notation of this paper, is

$$\sup_{\tilde{f}_k} \left|\hat{\rho}^{(s)k}_{i_0j_0\,|\,V\setminus\{i_0,j_0\}}\right|^2. \tag{68}$$

To design a constant false-alarm rate test, one needs distribution of the estimated partial coherence when its true value is zero. In [17], it is approximated as a χ_2^2 random variable at any given frequency, based upon some asymptotic considerations in [31, Sec. 8.8]. Taking different frequencies $\tilde{f}_k$ over (0, 0.5) to be sufficiently apart so that the partial coherences at distinct frequencies are independent, the distribution of the test statistic is that of the supremum of Q independent χ_2^2 random variables, where Q is number of selected frequencies over (0, 0.5). It is noted in [17] (and echoed in [23], [25]) that "It is very difficult to determine the exact asymptotic distribution ..." of this supremum.

We will use our formulation of frequency-domain sufficient statistic, computation of $\sup_{\tilde{f}_k} |\hat{\rho}^{(s)k}_{i_0j_0\,|\,V\setminus\{i_0,j_0\}}|^2$ and choice of $\tilde{f}_k$ s, to determine the null distribution more accurately. Define our version of Dahlhaus's test statistic as

$$\mathcal{T}_D = \sup_{k\in[1,M]} \left|\hat{\rho}^{(s)k}_{i_0j_0\,|\,V\setminus\{i_0,j_0\}}\right|^2 \tag{69}$$

where $|\hat{\rho}^{(s)k}_{i_0j_0\,|\,V\setminus\{i_0,j_0\}}|^2$ satisfies (56). By Theorem 1 and Lemma 1, under $\mathcal{H}_0$, $1 - |\hat{\rho}^{(s)k}_{i_0j_0\,|\,V\setminus\{i_0,j_0\}}|^2$ is distributed as a $\mathrm{beta}(K-p+1, 1)$ random variable. It then follows that $|\hat{\rho}^{(s)k}_{i_0j_0\,|\,V\setminus\{i_0,j_0\}}|^2$ is a $\mathrm{beta}(1, K-p+1)$ random variable under $\mathcal{H}_0$; denote its cumulative distribution function (CDF) as $F_{\hat{\rho}}(x) = P\{|\hat{\rho}^{(s)k}_{i_0j_0\,|\,V\setminus\{i_0,j_0\}}|^2 \le x\}$. Then the CDF $F_{\mathcal{T}_D}(x)$ of $\mathcal{T}_D$ is

$$F_{\mathcal{T}_D}(x) = \left(F_{\hat{\rho}}(x)\right)^M = \left(1 - (1-x)^{K-p+1}\right)^M. \tag{70}$$

Therefore, our version of Dahlhaus's test

$$\mathcal{T}_D \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau \tag{71}$$

has a significance level P_{fa} if we pick

$$\tau = 1 - \left(1 - (1-P_{fa})^{1/M}\right)^{1/(K-p+1)} \tag{72}$$

in (71).

1) Flop Count: The flop count for the Dahlhaus test statistic (69) is of the same order (in fact, almost the same) as that for the proposed test statistic (58). In both cases we first calculate

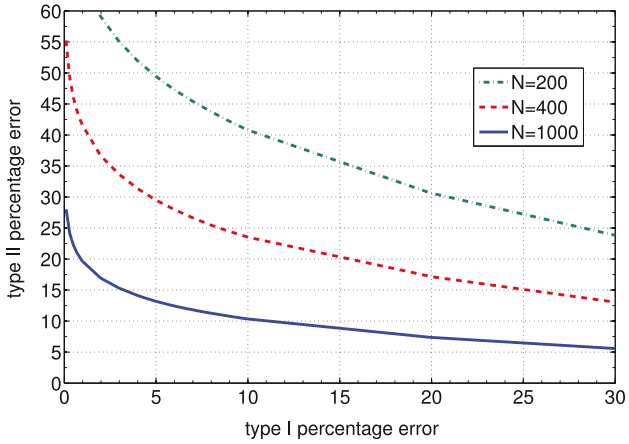

Fig. 1. CGGMs: Type I and II percentage errors for $p = 40$ and $N = 200$, 400 and 1000, as P_{fa} is varied. Results based on 1,000 Monte Carlo runs.

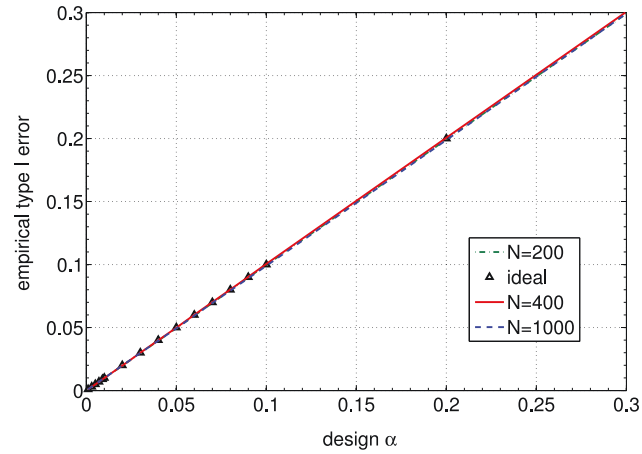

Fig. 2. CGGMs: Type I error for $p = 40$ and $N = 200$, 400 and 1000, as design $P_{fa} = \alpha$ is varied. Results based on 1,000 Monte Carlo runs.

partial coherence for all distinct edges at M frequency points. The only difference between the two is that in (69), we compute the maximum of partial coherence for a given edge over M frequency points (complexity $\mathcal{O}(Mp^2)$ over all edges), whereas in (58), we first compute the log of one minus partial coherence for a given edge at M frequencies and then we sum them over these frequencies, resulting in complexity $\mathcal{O}(Mp^2)$ over all edges. Therefore, the flop count (65) is valid for the Dahlhaus test statistic as well, leading to $F_{Dahlhaus} = F_{GLRT}$.

V. Numerical Experiments

Given $\mathbf{x}(t)$ or $\mathbf{y}(t) \in \mathbb{C}^p$, there are $p(p-1)/2$ unordered pair of vertices in the associated CIG that may or may not be connected. So we have to perform at least $p(p-1)/2$ binary hypothesis tests (a given edge is missing from the graph is the null, and the complete graph is the alternative hypothesis). Thus, we have a multiple testing problem where the main issue is how to control the overall significance level. Instead, as in [23, Sec. VII] (in the context of time series), we will use trade-off between average type I (false-alarm rate) and type II (miss probability) errors (over all edges) as a performance measure.

A. CGGMs

Consider i.i.d. $\mathbf{a}_i \in \mathcal{C}^p$, $\mathbf{a}_i \sim \mathcal{N}_c(0, \mathbf{I})$, $p \times p$ $\mathbf{A} = [\mathbf{a}_1 \cdots \mathbf{a}_p]^H$, and set $\check{\mathbf{\Omega}} = \mathbf{A}^H\mathbf{A}$. With probability q and independently, we set off-diagonal elements in the upper triangle of $\check{\mathbf{\Omega}}$ to zero (taking care to set the corresponding elements in lower triangle also to zero so that the resulting matrix $\bar{\mathbf{\Omega}}$ is Hermitian). Now set $\mathbf{\Omega} = \bar{\mathbf{\Omega}} + \beta\mathbf{I}$ with β picked to make $\mathbf{\Omega}$ positive definite. Then approximately $100q\%$ entries of $\mathbf{\Omega}$ are null, and $\mathbf{\Omega}$ is a valid positive definite concentration matrix. With $\mathbf{\Phi}\mathbf{\Phi}^H = \mathbf{\Omega}^{-1}$, we generate $\mathbf{x} = \mathbf{\Phi}\mathbf{w}$ with $\mathbf{w} \sim \mathcal{N}_c(0, \mathbf{I})$, to obtain N i.i.d. observations from $\mathcal{N}_c(0, \mathbf{\Omega}^{-1})$. We set $q = 0.6$ to have approximately 60% entries of $\mathbf{\Omega}$ as null.

We pick the test threshold in (34) for a specified significance level of P_{fa}, and then we carry out the tests for all edges and compute the types I and II percentage errors. The results averaged over 1,000 runs are shown in Figs. 1 and 2 for $p = 40$ and $N = 200, 400$ or 1000, using a different randomly generated model in each run. The type I and II percentage errors are defined, respectively, as

$$\frac{\#\text{ of edges accepted when missing}}{\#\text{ of edges missing}} \times 100\,,$$

$$\frac{\#\text{ of edges rejected when present in true graph}}{\#\text{ of edges present}} \times 100\,.$$

It is seen from Fig. 1 that performance improves with increasing N and P_{fa} (type I error), as expected. It is seen from Fig. 2 that empirical P_{fa} tracks the design P_{fa} quite well, and empirical P_{fa} changes little with N.

B. Time Series

We now present some simulation results to illustrate the proposed method (Theorem 4). We use an example from [23] whose approach, a version of [25], we also use as a benchmark. As explained in Sec. II-C, since we are dealing with dependent real time series, the methods designed for i.i.d. real times series (such as [1]) do not apply and therefore, were not chosen for comparison (except in Sec. V-C). Dahlhaus [17] was the first to systematically address this problem, therefore, our implementation of [17], as in Sec. IV-F, is chosen in this paper to compare against our GLRT. The partial coherence based approaches of [18], [21], [22] are all applicable to our problem but, as discussed in Sec. I, for a given test edge, their tests are applied at distinct frequencies, requiring multiple testing, inevitably leading to a loss in test power. In contrast, the method of Matsuda [25] needs just one test per edge, and so does our proposed GLRT. This motivates our choice of [25], as used in [23], to compare against.

Consider the following random, VAR model for $\mathbf{x}(t) \in \mathbb{R}^p$

$$\mathbf{x}(t) = \mathbf{\Phi}\mathbf{x}(t-1) + \mathbf{w}(t) \tag{73}$$

where $\mathbf{w}(t) \sim \mathcal{N}_r(\mathbf{0}, \mathbf{I})$ is i.i.d. real Gaussian with identity covariance matrix. Following [23, Appendix], initially let $\mathbf{\Phi} \in \mathbb{R}^{p\times p}$ have all zero entries. Then, for a fixed k, elements in position (i, j) for which $(i + j)_{mod\,k} = 1$, are randomly sampled from $\mathcal{N}_r(0, 1)$ (standard Gaussian) distribution. Carry out

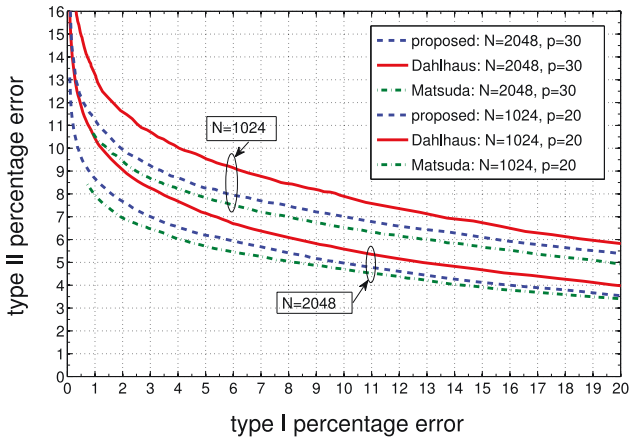

Fig. 3. Time Series: Type I and II percentage errors for $(N, p) = (1024, 20)$ and (2048, 30), as α is varied. The label "Matsuda" refers to the method of [25] as used in [23]. The label "Dahlhaus" is our implementation of the method of [17], as discussed in Sec. IV-F. Results based on 500 Monte Carlo runs.

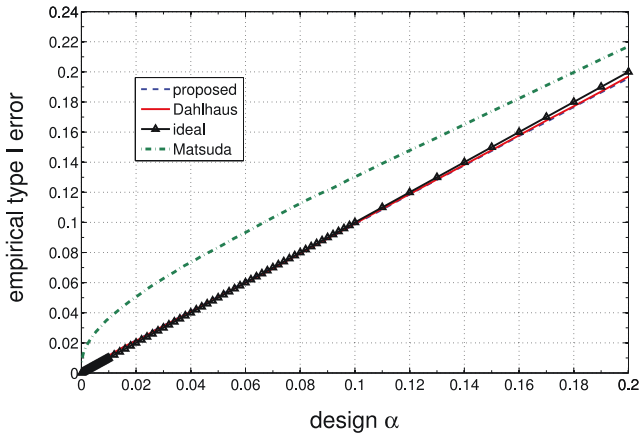

Fig. 4. Time Series: Type I error for $(N, p) = (1024, 20)$ and (2048, 30), as α is varied. The label "Matsuda" refers to the method of [25] as used in [23]. The label "Dahlhaus" is our implementation of the method of [17], as discussed in Sec. IV-F. Results based on 500 Monte Carlo runs.

the eigen-decomposition of $\mathbf{\Phi}$, and replace any eigenvalue with magnitude greater than one with its reciprocal. So reconstituted $\mathbf{\Phi}$ leads to a stable AR model with stationary sequence. The choice of k controls the sparsity: choosing $k = 10$ makes approximately 80% entries of the $\mathbf{\Phi}$ matrix zero for $10 \leq p \leq 50$. The above VAR model was used for our data generation. Then we have $\mathbf{S}_x^{-1}(f) = [\mathbf{I} - \mathbf{\Phi}^\top e^{j2\pi f}][\mathbf{I} - \mathbf{\Phi} e^{-j2\pi f}]$ with similar sparsity.

We pick the test threshold of GLRT in Theorem 4 for a specified significance level of α, and then we carry out the tests for all edges and compute the types I and II percentage errors. The results averaged over 500 runs are shown in Figs. 3 and 4 for $(N, p) = (1024, 20)$ and (2048, 30), using a different randomly generated VAR(p) model in each run. We also show the results for the approaches of [25] (as used in [23], and labeled "Matsuda" in the figures) and [17] (as modified in Sec. IV-F via (71), and labeled "Dahlhaus" in the figures). The smoothing parameters, m_t $(= (K-1)/2)$ in (45) and (57), and

TABLE I
COMPUTATIONAL COMPLEXITY. THE LABEL "MATSUDA" REFERS TO THE METHOD OF [25] AS USED IN [23]. THE LABEL "DAHLHAUS" IS OUR IMPLEMENTATION OF THE METHOD OF [17], AS DISCUSSED IN SEC. IV-F.

Approach	Proposed	Matsuda	Dahlhaus
N=1,024; p=20; m_t=82=M_s			
time (s)	0.0122	22.191	0.0115
flop ratio	1	4,008	1
time ratio	1	1,930	0.958
N=2,048; p=30; m_t=164=M_s			
time (s)	0.0414	232.375	0.0399
flop ratio	1	13,716	1
time ratio	1	5,613	0.964

M_s in (60), were selected as $m_t = M_s = 82$ for $N = 1024$, and $m_t = M_s = 164$ for $N = 2048$; the choices for M_s are similar to what is given in [23] for comparable values of N. It is seen from Fig. 3 that for a given value of type I error, the test statistic of [23], [25] yields type II error that is about 0.5 percentage points lower (better) than our proposed test statistic, whereas Dahlhaus statistic (71) is about 1 to 1.5 percentage points higher (worse). But as we shall see shortly (Table I), our method is at least three orders of magnitude faster than [23], [25], and takes about the same time as modified [17].

Fig. 4 shows the empirical type I error as a function of the design significance level α. It is seen that our proposed test statistic and (71), both yield empirical P_{fa} that tracks the design P_{fa} quite well. However, Matsuda's statistic yields significantly biased results.

We also carried out computational complexity comparisons in terms of flop count, as discussed in Sec. IV-E, and computation time (per run, averaged over 500 runs) as measured by tic-toc functions of MATLAB, on a Windows 8 system with 8GB RAM and Intel Core i3-4130 processor at 3.4GHz. We use flop ratio defined as flop count $F_{Matsuda}$ (66) for [23], [25], divided by the flop count F_{GLRT} (64) for our method, and similarly $F_{Dahlhaus}/F_{GLRT}$. These results are shown in Table I and Fig. 5. It is seen from Table I that our method is at least three orders of magnitude faster than [23], [25], with a small loss in performance (Fig. 3), and is as fast as modified Dahlhaus, but with superior performance (Fig. 3).

C. Applying An Existing Method of [1] Under Time-Domain I.I.D. Model

As discussed in Sec. II-C, existing time-domain methods for real-valued time series GGMs, such as those in [3], [19], [20], [33], and frequency-domain methods of [34]–[36], are not applicable in our context. How about application of the edge exclusion test of [1, Sec. 5.3.3] for real GGMs (on which Theorem 1 and Lemma 1 of Sec. III pertaining to CGGMs are patterned) to real-valued time series GGMs? A theoretical drawback of the approach of [1, Sec. 5.3.3] is that it is based on an i.i.d. time series model.

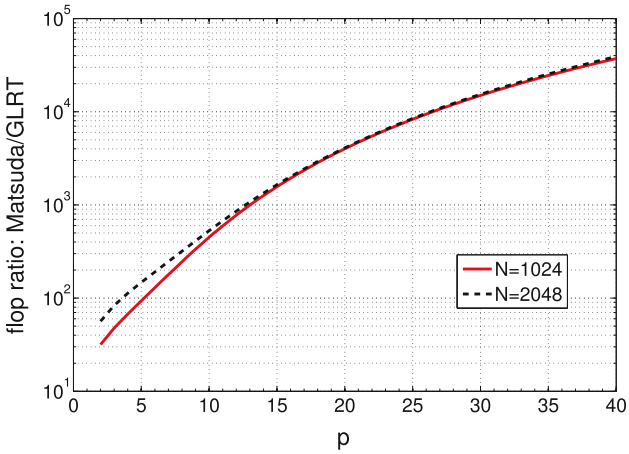

Fig. 5. The ratio of flop count (number of complex multiplications or divisions) for the method of [25] as used in [23], divided by that for the proposed GLRT approach, versus time series dimension p. Flop ratio $= F_{Matsuda}/F_{GLRT}$ (see (64) and (66)).

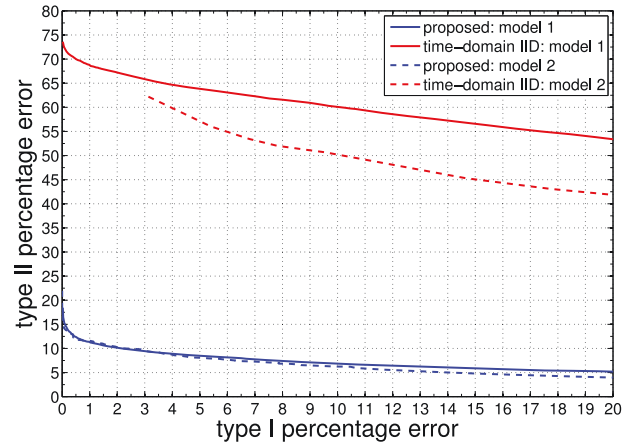

Fig. 6. Time Series: Type I and II percentage errors for $(N, p) = (1024, 20)$ for Model 1 ((73) of Sec. V-B), and $(N, p) = (1024, 5)$ for Model 2 ((73) with Φ specified by (76)), as α is varied. The label "time-domain IID" refers to test (74). Results based on 500 Monte Carlo runs.

Consider a real GGM with $\mathbf{x} \sim \mathcal{N}_r(\mathbf{0}, \mathbf{\Sigma})$, $\mathbf{x} \in \mathbb{R}^p$, associated with graph $\mathcal{G} = (V, \mathcal{E})$, where $\{i, j\} \notin \mathcal{E}$ iff $\Omega_{ij} = 0$. Here $\mathbf{\Omega} = \mathbf{\Sigma}^{-1}$ and $\mathbf{\Sigma}$ is real, symmetric, positive-definite. The GLRT for testing if the edge $\{i_0, j_0\} \in \mathcal{E}$ is as in Theorem 1 and Lemma 1 (see [1, Sec. 5.3.3]) except that now $\mathbf{\Omega}$ and $\mathbf{\Sigma}$ are real. Using the equivalent test (35), we have

$$\mathcal{L}_R = \frac{\hat{\Omega}^2_{i_0 j_0}}{\hat{\Omega}_{i_0 i_0} \hat{\Omega}_{j_0 j_0}} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau \tag{74}$$

where

$$\hat{\mathbf{\Sigma}} = \frac{1}{N} \sum_{t=0}^{N-1} \mathbf{x}(t)\mathbf{x}^\top(t) \ , \quad \hat{\mathbf{\Omega}} = \hat{\mathbf{\Sigma}}^{-1} \ . \tag{75}$$

It follows from [1, Sec. 5.3.3] (see also [1, Prop. 5.12]) that under $\mathcal{H}_0$, $\mathcal{L}_R \sim \text{beta}(1/2, (N - p + 1)/2)$ if $\{\mathbf{x}(t)\}$ is an i.i.d. sequence. If $\{\mathbf{x}(t)\}$ turns out to be non-i.i.d. (but still stationary), the distribution of $\mathcal{L}_R$ is unknown under $\mathcal{H}_0$.

We will apply test (74) as an edge exclusion test to time series GGMs with the threshold τ selected based on $\mathcal{L}_R \sim \text{beta}(1/2, (N - p + 1)/2)$ under $\mathcal{H}_0$. We compare it with our proposed test (Theorem 4) based on CGGMs in the frequency-domain, not requiring the assumption of i.i.d. time series. We use synthetic data (where ground truth is known) based on two different models. The VAR model of (73) of Sec. V-B with $p = 20$ and $k = 10$ (used in Figs. 3–5) is labeled as Model 1. We also consider model A of [23] which uses (73) with $p = 5$ and

$$\Phi = \begin{bmatrix} 0.2 & 0 & -0.1 & 0 & -0.5 \\ 0.4 & -0.2 & 0 & 0.2 & 0 \\ -0.2 & 0 & 0.3 & 0 & 0.1 \\ 0.3 & 0.1 & 0 & 0.3 & 0 \\ 0 & 0 & 0 & 0.5 & 0.2 \end{bmatrix} . \tag{76}$$

This model results in 3 edges missing out of total 10 edges. This model is labeled as Model 2.

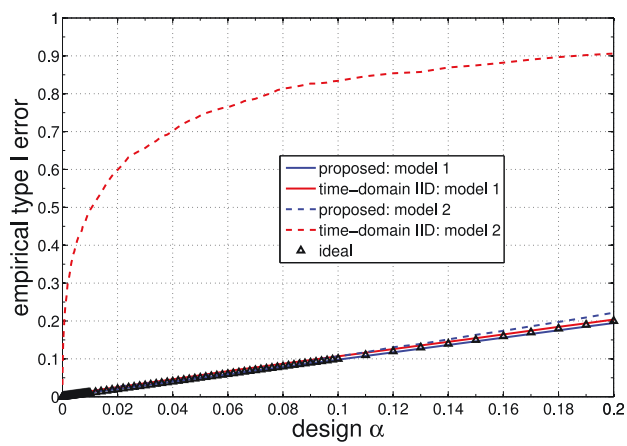

Fig. 7. Time Series: Type I error for $(N, p) = (1024, 20)$ for Model 1 ((73) of Sec. V-B), and $(N, p) = (1024, 5)$ for Model 2 ((73) with Φ specified by (76)), as α is varied. The label "time-domain IID" refers to test (74). Results based on 500 Monte Carlo runs.

Simulation results based on 500 runs are shown in Figs. 6 and 7, for $(N, p) = (1024, 20)$ for Model 1 and $(N, p) = (1024, 5)$ for Model 2. For the proposed method we used $m_t = 82$ for both models. The test (74) is labeled as "time-domain IID" in Figs. 6 and 7. It is seen from Fig. 7 that while the proposed approach yields empirical P_{fa} that tracks the design P_{fa} quite well for both Models 1 and 2, the time-domain IID test (74) yields empirical P_{fa} that tracks the design P_{fa} quite well only for Model 1, but is significantly biased for Model 2. The results of Fig. 7 imply that for Model 2, the time-domain IID test (74) when applied for a design significance level α, yields an empirical P_{fa} that is several times higher, i.e., quite a few missing edges are declared as connected. It is seen from Fig. 6 that the type II errors are much larger for both models under the time-domain IID test (74), when compared to the proposed approach. This implies that the time-domain IID test (74) misses many more true connected edges compared to the proposed approach, provided one can

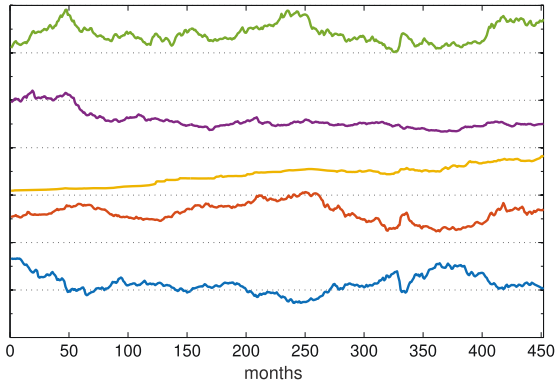

Fig. 8. Foreign exchange time series for five countries (scaled and offset for display).

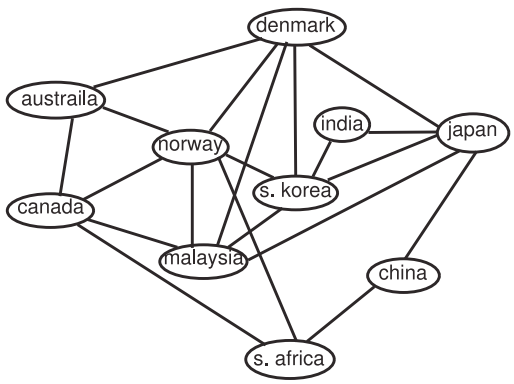

Fig. 9. Graph estimate for foreign exchange data set under the proposed approach. Number of connected edges =20 out of 45.

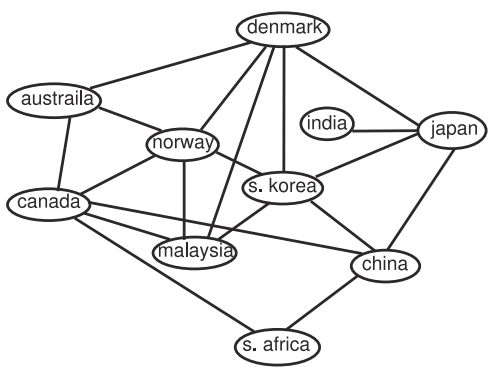

Fig. 10. Graph estimate for foreign exchange data set under our implementation of the method of Dahlhaus [17]. Number of connected edges =19 out of 45.

calculate the test threshold to yield the desired type I error. For Model 1, the latter is possible as seen from Fig. 7, but not so for Model 2 where the computed test threshold yields much higher empirical P_{fa}. Thus, while the performance of the proposed approach to real-valued time series GGMs is "consistent" for both Models 1 and 2 in that in both cases the empirical P_{fa} tracks the design P_{fa} quite well, this is not the case for the time-domain IID test (74) which may or may not yield an empirical P_{fa} that tracks the design P_{fa}. Therefore, in practice, when using the time-domain IID test (74) for a specified design significance

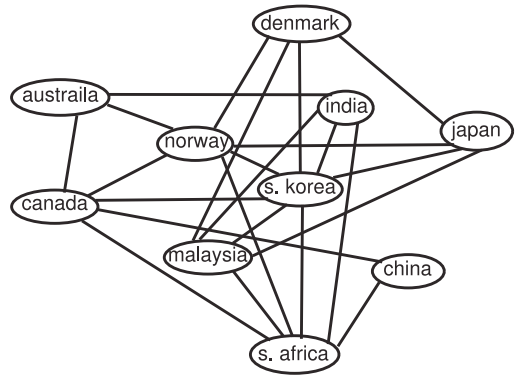

Fig. 11. Graph estimate for foreign exchange data set under the method of Matsuda [25] as used in [23]. Number of connected edges = 23 out of 45.

level on a non-i.i.d. time series, the test may yield quite a few extra connected edges because the empirical P_{fa} is much larger than the design P_{fa} (or may miss a significant number of connected edges because the type II error is much larger than our proposed approach, for a given value of empirical P_{fa}).

VI. Application to Foreign Exchange Data

We now apply our time series graphical model selection method to a foreign exchange data set. We consider a multivariate time series of monthly trends of foreign exchange rates[1] of currencies of ten countries (Australia, Canada, China, Denmark, India, Japan, Malaysia, Norway, South Africa and South Korea) with respect to the US dollar from April 1, 1981 to November 1, 2018. This results in a 10-dimensional time series of length 452. Similar times series consisting of foreign exchange rates of 19 countries from October 1, 1983 to February 1, 2012, resulting in a 19-dimensional times series of length 341, was analyzed in [10] in a high-dimensional setting, based on i.i.d. times series modeling.

The times series for 5 countries (Australia, Canada, India, Japan and Norway) is shown in Fig. 8 where the series for each country is differently scaled and offset for ease of display. The figure shows that times series components may have linear trends. Therefore, before applying the various methods for graphical model selection, the 10-dimensional time series was first detrended (i.e., remove the best straight-line fit linear trend from each component series using the MATLAB function detrend). This is to conform to our modeling assumption of zero-mean stationary time series.

In order to estimate the conditional independence graph for the foreign exchange data set, we applied the proposed GLRT, and the approaches of Matsuda [25] (as used in [23]) and Dahlhaus [17] (as modified in Sec. IV-F via (71)). We selected test significance level (per edge) $\alpha = 0.01$, and the smoothing parameters, m_t $(= (K-1)/2)$ in (45) and (57), and M_s in (60), as $m_t = M_s = 21$. The resulting conditional independence graph estimates are shown in Figs. 9, 10 and 11 for the proposed

[1]Link *http://research.stlouisfed.org/fred2/categories/15/downloaddata* allows one to download foreign exchange rate data sets from the Federal Reserve Bank of St. Louis website.

TABLE II
FOREIGN EXCHANGE DATA SET: p-VALUES OF EDGE EXCLUSION TEST STATISTICS FOR VARIOUS EDGES, $N = 452$, $p = 10$, $m_t = 21 = M_s$. THE LABEL "MATSUDA" REFERS TO THE METHOD OF [25] AS USED IN [23]. THE LABEL "DAHLHAUS" IS OUR IMPLEMENTATION OF THE METHOD OF [17], AS DISCUSSED IN SEC. IV-F

Proposed Test Statistic p-Value									
	Canada	China	Denmark	India	Japan	Malaysia	Norway	S. Africa	S. Korea
Australia	0.000	0.021	0.000	0.510	0.394	0.242	0.000	0.741	0.311
Canada		0.024	0.339	0.030	0.136	0.001	0.000	0.000	0.017
China			0.818	0.347	0.005	0.301	0.061	0.000	0.067
Denmark				0.376	0.000	0.000	0.000	0.305	0.001
India					0.007	0.097	0.472	0.032	0.001
Japan						0.006	0.175	0.270	0.000
Malaysia							0.001	0.057	0.000
Norway								0.010	0.001
S. Africa									0.029

Dahlhaus Test Statistic p-Value									
	Canada	China	Denmark	India	Japan	Malaysia	Norway	S. Africa	S. Korea
Australia	0.000	0.011	0.000	0.335	0.476	0.191	0.000	0.729	0.359
Canada		0.007	0.089	0.014	0.478	0.000	0.000	0.001	0.035
China			0.475	0.065	0.001	0.580	0.012	0.000	0.007
Denmark				0.081	0.001	0.000	0.000	0.213	0.000
India					0.001	0.117	0.221	0.182	0.021
Japan						0.047	0.290	0.084	0.001
Malaysia							0.000	0.224	0.000
Norway								0.101	0.001
S. Africa									0.192

Matsuda Test Statistic p-Value									
	Canada	China	Denmark	India	Japan	Malaysia	Norway	S. Africa	S. Korea
Australia	0.000	0.031	0.028	0.000	0.259	0.498	0.000	0.744	0.236
Canada		0.000	0.898	0.208	0.074	0.269	0.000	0.000	0.002
China			0.675	0.684	0.052	0.075	0.035	0.000	0.511
Denmark				0.143	0.000	0.001	0.000	0.106	0.001
India					0.377	0.000	0.190	0.000	0.000
Japan						0.001	0.000	0.645	0.001
Malaysia							0.017	0.000	0.000
Norway								0.000	0.000
S. Africa									0.000

GLRT, Dahlhaus and Matsuda methods, respectively, with the number of connected edges as 20, 19 and 23, respectively, out of total 45 edges. All three graphs have 12 edges in common.

Table II shows the p-values for the 45 edges and each test statistic, proposed GLRT (58), Dahlhaus (71) and Matsuda (62). The p-values can be interpreted as providing a measure of strength of the presence/absence of edges. Independent of the chosen α (which is 0.01 for the graphs in Figs. 9, 10 and 11), a "low" p-value indicates a "definite" edge (rejection of the null hypothesis that the edge is absent), while a "high" p-value indicates acceptance of the null hypothesis that the edge is absent. The graphs shown in Figs. 9, 10 and 11 result when we apply the threshold of $\alpha = 0.01$, if p-value is less than 0.01, there is an edge, else no edge. Consider the absent edge between Australia and China in all three graphs. If we had picked $\alpha = 0.05$ (as is done in [17] for a different data set), our results would place an edge between Australia and China in all three graphs. Expert domain knowledge needs to be a consideration in such decisions,

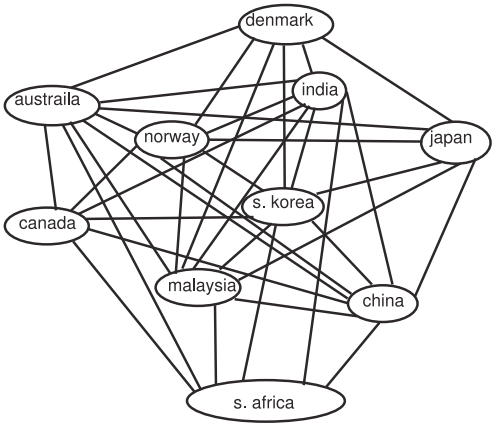

Fig. 12. Graph estimate for foreign exchange data set using the time-domain IID test (74). Number of connected edges =36 out of 45.

guided by graph modeling results, and vice versa. On the other hand, both the proposed GLRT and Dahlhaus methods definitely place an edge between Australia and Denmark (p-values equal to 0 to four decimal digits), but the Matsuda method with p-value of 0.028 is less reliable since our synthetic data based results of Fig. 4 show that this method yields an empirical α that is much higher than design α, i.e., its asymptotic null distribution can be quite inaccurate for finite length data, and our p-values are based on the asymptotic null distribution. Thus, the Matsuda method may yield more edges than the true value, and this may account for our results of 23 edges for Matsuda versus 20 for the proposed approach. Finally, we note that our synthetic data based results of Fig. 3 show that the test power of the Dahlhaus method is smaller than that of the proposed GLRT, and this may account for our results of 20 edges for the proposed approach versus 19 for Dahlhaus.

We also applied the time-domain IID test (74) to financial data for the design test significance level (per edge) $\alpha = 0.01$. The resulting conditional independence graph estimate is shown in Fig. 12, with the number of connected edges as 36 out of total 45 edges. Thus, the test (74) yields a much denser graph compared to the time series GGM approaches (proposed, Dahlhaus and Matsuda). The ground truth is unknown, but the results pertaining to synthetic data Model 2 (Sec. V-C) shown in Fig. 7 suggest that time-domain IID test (74) can yield empirical α that is much higher than design α (much more so than is the case for the Matsuda method). Thus, the test (74) may yield significantly more edges than the true value, and this may account for our results of 36 edges for time-domain IID test (74) versus 20 for the proposed approach.

VII. Conclusion

We considered the problem of inferring the conditional independence graph of a stationary multivariate Gaussian time series. Existing nonparametric methods of [25] and [23] use the Kullback-Leibler divergence measure to define an edge exclusion test statistic. In this paper, we proposed an alternative GLRT-based edge exclusion test statistic which is based on the asymptotic distribution of the DFT of the time series, and a frequency-domain sufficient statistic. It is computationally significantly faster than the methods of [25] and [23], and our simulations show that we achieve comparable test power levels. In particular, our method is at least three orders of magnitude faster than [23]. Our proposed approach is also based on a novel formulation of a GLRT based edge exclusion test for p-variate CGGMs; this result is also of independent interest. The proposed test statistic for CGGMs is expressed in an alternative form compared to an existing result, where the alternative expression is in a form usually given and exploited for real GGMs. The computational complexity of the proposed statistic is $\mathcal{O}(p^3)$ compared to $\mathcal{O}(p^5)$ for the existing result. We also applied our time series graphical model selection method to a foreign exchange data set consisting of monthly trends of foreign exchange rates of 10 countries.

References

[1] S. L. Lauritzen, *Graphical Models*. Oxford, U.K.: Oxford Univ. Press, 1996.
[2] A. O. Hero, III and B. Rajaratnam, "Foundational principles for large-scale inference: Illustrations through correlation mining," in *Proc. IEEE*, vol. 64, no. 104, pp. 93–110, Jan. 2016.
[3] M. Eichler, "Graphical modelling of multivariate time series," *Probability Theory Related Fields*, vol. 153, no. 1/2, pp. 233–268, Jun. 2012.
[4] P. Danaher, P. Wang, and D. M. Witten, "The joint graphical lasso for inverse covariance estimation across multiple classes," *J. Roy. Statist. Soc., Ser. B (Methodol.)*, vol. 76, pp. 373–397, 2014.
[5] N. Friedman, "Inferring cellular networks using probabilistic graphical models," *Science*, vol 303, pp. 799–805, 2004.
[6] S. L. Lauritzen and N. A. Sheehan, "Graphical models for genetic analyses," *Statistic. Sci.*, vol. 18, pp. 489–514, 2003.
[7] N. Meinshausen and P. Bühlmann, "High-dimensional graphs and variable selection with the Lasso," *Ann. Statist.*, vol. 34, no. 3, pp. 1436–1462, 2006.
[8] K. Mohan, M. J. Y. Chung, S. Han, D. Witten, S. I. lee, and M. Fazel, "Structured learning of Gaussian graphical models," in *Proc. Adv. Neural Inf. Process. Syst.*, Lake Tahoe, NV, USA, Dec. 2012, pp. 620–628.
[9] K. Mohan, P. London, M. Fazel, D. Witten, and S. I. Lee, "Node-based learning of multiple Gaussian graphical models," *J. Mach. Learn. Res.*, vol. 15, pp. 445–488, 2014.
[10] A. Anandkumar, V. Tan, F. Huang, and A. Willsky, "High-dimensional Gaussian graphical model selection: Walk summability and local separation criterion," *J. Mach. Learn. Res.*, vol. 13, pp. 2293–2337, 2012.
[11] H. H. Andersen, M. Hojbjerre, D. Sorensen, and P. S. Eriksen, *Linear and Graphical Models for the Multivariate Complex Normal Distribution* (Lecture Notes in Statistics), vol. 101. New York, NY USA: Springer, 1995.
[12] G. Marrelec *et al.*, "Partial correlation for functional brain interactivity investigation in functional MRI," *NeuroImage*, vol. 32, no. 1, pp. 228–237, 2006.
[13] O. Sporns, *Networks of the Brain*. Cambridge, MA, USA: MIT Press, 2010.
[14] M. C. Kociuba and D. B. Rowe, "Complex-valued time-series correlation increases sensitivity in fMRI analyis," *Magn. Reson. Imag.*, vol. 34, pp. 765–770, 2016.
[15] D. B. Rowe, "Magnitude and phase signal detection in complex-valued fMRI data," *Magn. Reson. Medicine*, vol. 62, pp. 1356–1357, 2009.
[16] D. R. Brillinger, "Remarks concerning graphical models of times series and point processes," (in Portuguese) *Revista de Econometria (Brazilian Rev. Econometr.)*, vol. 16, pp. 1–23, 1996.
[17] R. Dahlhaus, "Graphical interaction models for multivariate time series," *Metrika*, vol. 51, pp. 157–172, 2000.
[18] U. Gather, M. Imhoff, and R. Fried, "Graphical models for multivariate time series from intensive care monitoring," *Statist. Medicine*, vol. 21, no. 18, pp. 2685–2701, 2002.
[19] J. Songsiri, J. Dahl, and L. Vandenberghe, "Graphical models of autoregressive processes," in *Convex Optimization in Signal Processing and Communications*, Y. Eldar and D. Palomar Eds. Cambridge, U.K.: Cambridge Univ. Press, 2009, pp. 89–116.

[20] J. Songsiri and L. Vandenberghe, "Topology selection in graphical models of autoregressive processes," *J. Mach. Learn. Res.*, vol. 11, pp. 2671–2705, 2010.
[21] T. Medkour, A. T. Walden, A. P. Burgess, and V. B. Strelets, "Brain connectivity in positive and negative syndrome schizophrenia," *Neuroscience*, vol. 169, no. 4, pp. 1779–1788, 2010.
[22] D. Schneider-Luftman, "p-Value combiners for graphical modelling of EEG data in the frequency domain," *J. Neurosci. Methods*, vol. 271, pp. 92–106, 2016.
[23] R. J. Wolstenholme and A. T. Walden, "An efficient approach to graphical modeling of time series," *IEEE Trans. Signal Process.*, vol. 64, no. 12, pp. 3266–3276, Jun. 2015.
[24] S. Holm, "A simple sequentially rejective multiple test procedure," *Scand. J. Statist.*, vol. 6, pp. 65–70, 1979.
[25] Y. Matsuda, "A test statistic for graphical modelling of multivariate time series," *Biometrika*, vol. 93, no. 2, pp. 399–409, 2006.
[26] P. W. F. Smith and J. Whittaker, "Edge exclusion tests for graphical Gaussian models," in *Learning in Graphical Models*, M. I. Jordan, Ed. Cambridge, MA, USA: MIT Press, 1998, pp. 555–574.
[27] M. Drton and M. D. Perlman, "Model selection for Gaussian concentration graphs," *Biometrika*, vol. 91, pp. 591–602, 2004.
[28] M. Drton and M. D. Perlman, "Multiple testing and error control in Gaussian graphical model selection," *Statist. Sci.*, vol. 22, pp. 430–439, 2007.
[29] J. K. Tugnait, "An edge exclusion test for graphical modeling of multivariate time series," in *Proc. 52nd Annu. Conf. Inf. Sci. Syst. (CISS)*, Princeton Univ., Princeton, NJ, USA, Mar. 21–23, 2018, pp. 1–6.
[30] J. K Tugnait, "An edge exclusion test for complex Gaussian graphical model selection," in *Proc. IEEE Statist. Signal Process. Workshop*, Freiburg, Germany, Jun. 10–13, 2018, pp. 678–682.
[31] D. R. Brillinger, *Time Series: Data Analysis and Theory*, Expanded ed. New York, NY, USA: McGraw-Hill, 1981.
[32] M. F. Salgueiro, P. W. F. Smith, and J. W. McDonald, "Power of edge exclusion tests in graphical Gaussian models," *Biometrika*, vol. 92, pp. 173–182, Mar. 2005.
[33] M. Eichler, "Graphical modelling of dynamic relationships in multivariate time series," in *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*, B. Schelter, M. Winterhalder, and J. Timmer, Eds. New York, NY, USA: Wiley, 2006, pp. 335–372.
[34] A. Jung, R. Heckel, H. Bölcskei, and F. Hlawatsch, "Compressive nonparametric graphical model selection for time series," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Florence, Italy, May 2014, pp. 769–773.
[35] A. Jung, "Learning the conditional independence structure of stationary time series: A multitask learning approach," *IEEE Trans. Signal Process.*, vol. 63, no. 21, pp. 5677–5690, Nov. 2015.
[36] A. Jung, G. Hannak, and N. Goertz, "Graphical LASSO based model selection for time series," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1781–1785, Oct. 2015.
[37] G. Casella and R. L. Berger, *Statistical Inference*, 2nd ed. Pacific Grove, CA, USA: Duxbury, 2002.
[38] R. Wolstenholme, "Clustering time series data by analysing graphical models of connectivity and the application to diagnosis of brain disorders," Ph.D. dissertation, Dept. Math., Imperial College, London, U.K., Dec. 2016.

Jitendra K. Tugnait (M'79–SM'93–F'94–LF'16) received the B.Sc. (Hons.) degree in electronics and electrical communication engineering from Punjab Engineering College, Chandigarh, India, in 1971, the M.S. and E.E. degrees from Syracuse University, Syracuse, NY, USA, and the Ph.D. degree from the University of Illinois, Urbana-Champaign, Champaign, IL, USA, in 1973, 1974, and 1978, respectively, all in electrical engineering.

From 1978 to 1982, he was an Assistant Professor in electrical and computer engineering with the University of Iowa, Iowa City, IA, USA. From June 1982 to September 1989, he was with the Long Range Research Division, Exxon Production Research Company, Houston, TX, USA. In September 1989, he joined as a Professor the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL, USA, where he is currently the James B. Davis Professor. His current research interests are in statistical signal processing, wireless physical and secure communications, and multiple target tracking.

Dr. Tugnait was an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE SIGNAL PROCESSING LETTERS, and IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and a Senior Area Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and a Senior Editor for IEEE WIRELESS COMMUNICATIONS LETTERS.