

GRAPHICAL MODELS FOR NONSTATIONARY TIME SERIES

BY SUMANTA BASU^{1,a} AND SUHASINI SUBBA RAO^{2,b}

¹*Department of Statistics and Data Science, Cornell University, sumbose@cornell.edu*

²*Department of Statistics, Texas A&M University, suhasini@stat.tamu.edu*

We propose NonStGM, a general nonparametric graphical modeling framework, for studying dynamic associations among the components of a nonstationary multivariate time series. It builds on the framework of Gaussian graphical models (GGM) and stationary time series graphical models (StGM) and complements existing works on parametric graphical models based on change point vector autoregressions (VAR). Analogous to StGM, the proposed framework captures conditional noncorrelations (both intertemporal and contemporaneous) in the form of an undirected graph. In addition, to describe the more nuanced nonstationary relationships among the components of the time series, we introduce the new notion of conditional nonstationarity/stationarity and incorporate it within the graph. This can be used to search for small subnetworks that serve as the “source” of nonstationarity in a large system.

We explicitly connect conditional noncorrelation and stationarity between and within components of the multivariate time series to zero and Toeplitz embeddings of an infinite-dimensional inverse covariance operator. In the Fourier domain, conditional stationarity and noncorrelation relationships in the inverse covariance operator are encoded with a specific sparsity structure of its integral kernel operator. We show that these sparsity patterns can be recovered from finite-length time series by nodewise regression of discrete Fourier transforms (DFT) across different Fourier frequencies. We demonstrate the feasibility of learning NonStGM structure from data using simulation studies.

1. Introduction. Graphical modeling of multivariate time series has received considerable attention in the past decade as a tool to study dynamic relationships among the components of a large system observed over time. Key applications include, among others, analysis of brain networks in neuroscience [44] and understanding linkages among firms for measuring systemic risk buildup in financial markets [25].

The vast majority of graphical models for time series focuses on the stationary setting (see [3, 6, 7, 11, 14, 17, 18, 23, 28, 31, 38, 57, 61, 68], to name but a few). While the assumption of stationarity may be realistic in many situations, it is well known that *nonstationarity* arises in many applications. In neuroscience, for example, task based fMRI data sets are known to exhibit considerable nonstationarity in the network connections, a phenomenon known as *dynamic functional connectivity*; see [53]. A naive application of graphical modeling methods, designed for stationary processes, can lead to spurious network edges if the actual time series is nonstationary.

The limited body of work on graphical models for nonstationary time series has so far focused on a restricted class of nonstationary models, where the data generating process can be well approximated by a finite order change point vector autoregressive (VAR) model. Within this framework [65] and [58] have proposed methods for constructing a “dynamically changing” network at each of the estimated change points. However, these methods are designed

Received September 2021; revised March 2022.

MSC2020 subject classifications. Primary 62M10, 62M15; secondary 62J07.

Key words and phrases. Graphical models, locally stationary time series, partial covariance, spectral analysis.

for time series which are piecewise stationary and follow a finite order stationary VAR over each segment. For many data sets, these conditions can be too restrictive, for example, they do not allow for smoothly changing parameters.

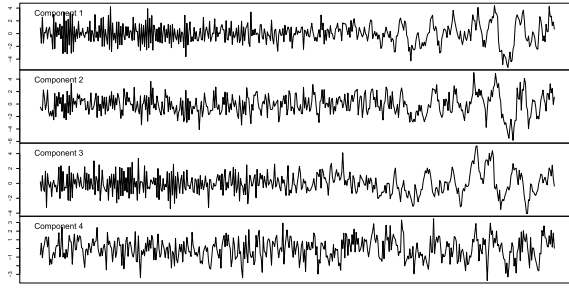
A naïve approximation of smoothly changing dynamics by a change point VAR could result in many short piecewise stationary segments. As a result, the estimated graph structure may be harder to interpret. Also, the model parameters within each short segment may be harder to estimate due to small sample size. Analogous to stationary time series where spectral methods allow for a nonparametric approach, it would be useful to define meaningful networks for nonstationary time series.

The objective of this paper is to move away from semiparametric models and propose a general framework for the graphical modeling of multivariate (say, p -dimensional) nonstationary time series. Our motivation comes from Gaussian graphical models (GGM), where the edges of a conditional dependence graph can distinguish between the direct and indirect nature of dependence in multivariate Gaussian random vectors. We argue that a general graphical model framework for nonstationary time series should have the capability to distinguish between two types of nonstationarity, the source of nonstationarity and one that inherits their nonstationarity by way of its connection with the source. This way of dimension reduction will be useful for modeling large systems where the nonstationarity arises only from a small subset of the process and then permeates through the entire system. Moreover, the identification of sources and propagation channels of nonstationarity may also be of scientific interest.

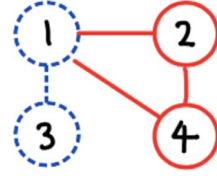
Analogous to GGM, in our framework the presence/absence of edges in the network encodes conditional correlation/noncorrelation relationships amongst the p components (nodes) of the time series. An additional attribute distinguishes between the types of nonstationarity. A graphical model is built using *conditional* relations. In this spirit we introduce the concept of *conditional stationarity* and *nonstationarity*. To the best of our knowledge, this is a new notion. A solid edge between two nodes in the network implies that their linear relationship, conditional on all the other nodes, does not change over time. In contrast, a dashed edge implies that their conditional relationship changes over time. We formalize these notions in Section 2. Nodes in our network also have *self-loops* to indicate whether the time series is nonstationary on its own or if it inherits nonstationarity from some other component in the system. The self-loops are denoted by a circle (solid or dashed) round the node.

The time-varying autoregressive model is often used to model nonstationarity. To illustrate the above ideas, in the following example we connect the parameters of a time-varying autoregressive model (tvVAR), which is a mixture of constant and time dependent parameters, to the concepts introduced above:

TOY EXAMPLE. Consider the trajectories of a *four*-dimensional time series given in Figure 1. The time series plots of all the components exhibit negative autocorrelation at the start of the time series that slowly changes to positive autocorrelation toward the end. Thus, the nonstationarity of each individual time series, at least from a visual inspection, is apparent. The data is generated from a time-varying vector autoregressive (tvVAR(1)) model (see Section 2 for details), where components 1 and 3 are the *sources of nonstationarity*, that is, they are affected by their own past through a (smoothly) time-varying parameter. In addition, component 3 affects component 1. Component 2 and 4 affect each other in a time-invariant way. Component 2 is also affected by 1, and components 1 and 4 are affected through 2. As a result, components 2 and 4 *inherit the nonstationarity from the sources* 1 and 3. As far as we are aware, there currently does not exist tools that adequately describe the nuanced differences in their dependencies and nonstationarity. Our aim in this paper is to capture these relationships in the form of the schematic diagram in Figure 1b. We note that the tvVAR model is a special case of our general framework, which does not make any explicit assumptions on the data generating process.



(a) Trajectories of a *four*-dimensional time series generated by a tvVAR(1) model. The multivariate process is jointly nonstationary. However, components 1 and 3 are the source of nonstationarity, while the other two components inherit the nonstationarity by means of their conditional dependence structure.



(b) The joint governing architecture is described in the graph. Dashed edges and self-loops represent conditional nonstationarity, while solid edges and self-loops represent conditional invariance and stationarity, notions of which we formalize in our new graphical modeling framework.

FIG. 1. Time series and conditional dependence graph of a time-varying VAR model.

It is interesting to contrast the networks constructed using the “dynamically changing” approach, developed in [65] and [58] for change point VAR models, with our approach. Both networks convey different information about the nonstationary time series. The “dynamically changing” network can be considered as local, in the sense that it identifies regions of stationarity and constructs a directed graph over each of the stationary periods. While the graph in our approach is undirected and yields global information about relationships between the nodes.

In order to connect the proposed framework to the current literature, we conclude this section by briefly reviewing the existing graphical modeling frameworks for Gaussian random vectors (GGM) and multivariate stationary time series (StGM). In Section 2 we lay the foundations for our nonstationary graphical models (NonStGM) approach. In particular, we formally define the notions of conditional noncorrelation and stationarity of nodes, edges, and subgraphs in terms of zero and Toeplitz embeddings of an infinite dimensional inverse covariance operator. We show that this framework offers a natural generalizations to existing notions of conditional noncorrelation in GGM and StGM. It should be emphasized that we do not assume that the underlying time series is Gaussian. All the relationships that we describe are in terms of the partial covariance and, therefore, apply to any multivariate time series whose covariance exists. In Section 3 we switch to the Fourier domain and show that the conditional noncorrelation and nonstationarity relationships are explicitly encoded in the sparsity pattern of the integral kernel of the inverse covariance operator. This connection opens the door to learning the graph structure from finite length time series data with the discrete Fourier transforms (DFT). In Section 4 we focus on locally stationary time series. We show that, by conducting nodewise regression of discrete Fourier transforms (DFT) of the multivariate time series across different Fourier frequencies, it is possible to learn the network. Section 5 describes how the proposed general framework looks in the special case of tvVAR models, where the notions of conditional noncorrelation and nonstationarity are transparent in the transition matrix. Some numerical results are presented in Section 6 to illustrate the methodology. All the proofs for the results in this paper can be found in the Supplementary Material [4].

Background. We outline some relevant works in graphical models and tests for stationarity that underpin the technical development of NonStGM:

Graphical Models. A graphical model describes the relationships among the components of a p -dimensional system in the form of a graph with a set of vertices $V = \{1, 2, \dots, p\}$ and an edge set $E \subseteq V \times V$ containing pairs of system components which exhibit strong association, even after conditioning on the other components.

The focus of GGM is on the conditional independence relationships in a p -dimensional (centered) Gaussian random vector $\underline{X} = (X^{(1)}, X^{(2)}, \dots, X^{(p)})^\top$. The nonzero partial correlations $\rho^{(a,b)}$, defined as $\text{Corr}(X^{(a)}, X^{(b)} | X^{-\{a,b\}})$ and also encoded in the sparsity structure of the *precision matrix* $\Theta = [\text{Var}(\underline{X})]^{-1}$, are used to define the edge set E . The task of *graphical model selection*, that is, learning the edge set E from finite sample, is accomplished by estimating Θ with a penalized likelihood estimator as in graphical Lasso ([32]) or by node-wise regression ([47]) where each component of the random vector is regressed on the other $(p - 1)$ components.

Switching to the time series setting, consider $\{\underline{X}_t = (X_t^{(1)}, \dots, X_t^{(a)}, \dots, X_t^{(p)})^\top\}_{t \in \mathbb{Z}}$, a p -dimensional time series with autocovariance function $\text{Cov}(\underline{X}_t, \underline{X}_\tau) = \mathbf{C}(t, \tau)$. Note that, in future, we usually use $\{\underline{X}_t\}$ to denote the sequence $\{\underline{X}_t\}_{t \in \mathbb{Z}}$. A direct adaptation of the GGM framework that estimates the *contemporaneous* precision matrix $\mathbf{C}^{-1}(0, 0)$ (see [57, 68]) does not provide conditional relationships between the *entire time series*. [7] and [14] laid the foundation of graphical models in stationary time series, where the conditional relationships between the *entire* time series $\{X_t^{(a)}\}$ and $\{X_t^{(b)}\}$ is captured. They show that the inverse of the *multivariate spectral density function* $\Sigma(\omega) := (1/2\pi) \sum_{\ell=-\infty}^{\infty} \mathbf{C}(\ell) \exp[-i\ell\omega]$ $\omega \in [0, \pi]$ explicitly encodes the conditional uncorrelated relationships. To be precise, $[\Sigma^{-1}(\omega)]_{a,b} = 0$ for all $\omega \in [0, \pi]$ if and only if $\{X_t^{(a)}\}$ and $\{X_t^{(b)}\}$ are conditionally uncorrelated, given *all the other time series*. The graphical model selection problem reduces to finding all pairs (a, b) , where $[\Sigma^{-1}(\omega)]_{a,b} \neq 0$ for some $\omega \in [0, \pi]$. For Gaussian time series, the graph is a conditional independence graph, while for non-Gaussian time series the graph encodes partial correlation information. For brevity, we refer to this approach as StGM (stationary time series graphical models). Estimation of $\Sigma^{-1}(\omega)$ is typically done using the discrete Fourier transform of the time series (see [29]). More recently, for relatively “large” p , penalized methods such as GLASSO [38] and CLIME [31] have been used to estimate $\Sigma^{-1}(\omega)$. This framework crucially relies on stationarity, in particular, the Toeplitz property of the autocovariance function $\mathbf{C}_{t,\tau} = \mathbf{C}(t - \tau)$ and is not immediately generalizable to the nonstationary case.

Testing for stationarity. There is a rich literature on testing for nonstationarity of a time series. Most methods are based on testing for invariance of the spectral density function or autocovariance function over time (see [48, 52, 56], to name but a few). An alternative approach is based on the fact that the discrete Fourier transform at certain frequencies is close to uncorrelated for stationary time series. [2, 27, 30, 37] use this property to test for nonzero correlation between DFTs at different frequencies to detect for departures from stationarity. The above mentioned tests focus on the “marginal” notion of nonstationarity, instead of the conditional notion defined in this paper. Tests for marginal nonstationarity are not equipped to delineate between direct and indirect nature of conditionally nonstationary relationships among the components of a multivariate time series. However, in this paper we show that, analogous to marginal tests, it is possible to utilize the Fourier domain to detect for different types of conditional (non)stationarity.

2. Graphical models and conditional stationarity. For a p -dimensional nonstationary time series $\{\underline{X}_t\}$, all the *pairwise covariance* information are contained in the infinite set of $p \times p$ autocovariance matrices $\mathbf{C}_{t,\tau} = \text{Cov}[\underline{X}_t, \underline{X}_\tau]$ for $t, \tau \in \mathbb{Z}$. We aggregate this information into an operator C and show that its inverse operator D captures meaningful *conditional (partial) covariance* relationships (Section 2.2). Leveraging this connection, we first define a

graphical model and conditional stationarity of its nodes, edges, and subgraphs in terms of the operator D (Section 2.3). Then we show that these notions can be viewed as natural generalizations of the GGM and StGM frameworks (Sections 2.4 and 2.5). We start by introducing some notation that will be used to formally define these structures (this can be skipped on first reading). Let $\mathbf{A} = (A_{a,b} : 1 \leq a \leq d_1, 1 \leq b \leq d_2)$ denote a $d_1 \times d_2$ -dimensional matrix; then we define $\|\mathbf{A}\|_2^2 = \sum_{a,b} |A_{a,b}|^2$, $\|\mathbf{A}\|_1 = \sum_{a,b} |A_{a,b}|$ and $\|\mathbf{A}\|_\infty = \sup_{a,b} |A_{a,b}|$.

2.1. Definitions and notation. We use ℓ_2 and $\ell_{2,p}$ to denote the sequence space $\{u = (\dots, u_{-1}, u_0, u_1, \dots)'; u_j \in \mathbb{C} \text{ and } \sum_j |u_j|^2 < \infty\}$ and the (column) sequence space $\{w = \text{vec}[u^{(1)}, \dots, u^{(p)}]; u^{(s)} \in \ell_2 \text{ for all } 1 \leq s \leq p\}$, respectively (vec denotes the vectorisation of a matrix). On the spaces ℓ_2 and $\ell_{2,p}$, we define the two inner products $\langle u, v \rangle = \sum_{j \in \mathbb{Z}} u_j v_j^*$ (where $*$ denotes the complex conjugate) for $u, v \in \ell_2$ and $\langle x, y \rangle = \sum_{s=1}^p \langle u^{(s)}, v^{(s)} \rangle$ for $x = (u^{(1)}, \dots, u^{(p)})'$, $y = (v^{(1)}, \dots, v^{(p)})' \in \ell_{2,p}$ such that ℓ_2 and $\ell_{2,p}$ are two Hilbert spaces. For $x \in \ell_{2,p}$ let $\|x\|_2 = \langle x, x \rangle$. For $s_1, s_2 \in \mathbb{Z}$, we use A_{s_1, s_2} to denote the (s_1, s_2) entry in the matrix $\mathbf{A}$, which can be infinite dimensional and involve negative indices.

We consider the p -dimensional real-valued time series $\{\underline{X}_t\}_{t \in \mathbb{Z}}$, $\underline{X}_t = (X_t^{(1)}, \dots, X_t^{(p)})'$, where the univariate random variables $X_t^{(a)}$, $a = 1, \dots, p$, are defined on the probability space $(\Omega, \mathcal{F}, P)$. We assume for all t that $E[\underline{X}_t] = 0$; this condition is not necessary in Sections 2 and 3, but it simplifies the exposition. Let $L^2(\Omega, \mathcal{F}, P)$ denote all univariate random variables X , where $\text{Var}[X] < \infty$; and for any $X, Y \in L^2(\Omega, \mathcal{F}, P)$, we define the inner product $\langle X, Y \rangle = \text{Cov}[X, Y]$. For every $t, \tau \in \mathbb{Z}$, we define the $p \times p$ covariance $\mathbf{C}_{t, \tau} = \text{Cov}[\underline{X}_t, \underline{X}_\tau]$ and assume $\sup_{t \in \mathbb{Z}} \|\mathbf{C}_{t, t}\|_\infty < \infty$. Under this assumption, for all $t \in \mathbb{Z}$ and $1 \leq c \leq p$ $X_t^{(c)} \in L^2(\Omega, \mathcal{F}, P)$, let $\mathcal{H} = \overline{\text{sp}}(X_t^{(c)}; t \in \mathbb{Z}, 1 \leq c \leq p) \subset L^2(\Omega, \mathcal{F}, P)$ be the closure of the space spanned by $(X_t^{(c)}; t \in \mathbb{Z}, 1 \leq c \leq p)$. Since $L^2(\Omega, \mathcal{F}, P)$ defines a Hilbert space, $\mathcal{H}$ is also a Hilbert space. Therefore, by the projection theorem and for any closed subspace $\mathcal{M}$ of $\mathcal{H}$, there is a unique projection of $Y \in \mathcal{H}$ onto $\mathcal{M}$, which minimises $E(Y - X)^2$ over all $X \in \mathcal{M}$ (see Theorem 2.3.1, [9]). We will use $P_{\mathcal{M}}(Y)$ to denote this projection. In this paper we will primarily use the following subspaces:

$$\mathcal{H} - X_t^{(a)} = \overline{\text{sp}}[X_s^{(c)}; s \in \mathbb{Z}, 1 \leq c \leq p, (s, c) \neq (t, a)],$$

$$\mathcal{H} - (X_t^{(a)}, X_\tau^{(b)}) = \overline{\text{sp}}[X_s^{(c)}; s \in \mathbb{Z}, 1 \leq c \leq p, (s, c) \notin \{(t, a), (\tau, b)\}],$$

$$\mathcal{H} - (X^{(c)}; c \in \mathcal{S}) = \overline{\text{sp}}[X_s^{(c)}; s \in \mathbb{Z}, c \in \mathcal{S}'],$$

where $\mathcal{S}'$ denotes the complement of $\mathcal{S}$.

Using the covariance $\mathbf{C}_{t, \tau}$, we define the infinite dimensional matrix operator C as $C = (C_{a,b}; a, b \in \{1, \dots, p\})$, where $C_{a,b}$ denotes an infinite dimensional submatrix with entries $[C_{a,b}]_{t, \tau} = [\mathbf{C}_{t, \tau}]_{a, b}$ for all $t, \tau \in \mathbb{Z}$. For any $u \in \ell_2$, we define the (column) sequence $C_{a,b}u = \{[C_{a,b}u]_t; t \in \mathbb{Z}\}$, where $[C_{a,b}u]_t = \sum_{\tau \in \mathbb{Z}} [C_{a,b}]_{t, \tau} u_\tau$. For any $v = \text{vec}[u^{(1)}, \dots, u^{(p)}] \in \ell_{2,p}$, we define the (column) sequence Cv as

$$(2.1) \quad Cv = \begin{pmatrix} C_{1,1} & C_{1,2} & \dots & C_{1,p} \\ C_{2,1} & C_{2,2} & \dots & C_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ C_{p,1} & C_{p,2} & \dots & C_{p,p} \end{pmatrix} \begin{pmatrix} u^{(1)} \\ u^{(2)} \\ \vdots \\ u^{(p)} \end{pmatrix} = \begin{pmatrix} \sum_{s=1}^p C_{1,s} u^{(s)} \\ \sum_{s=1}^p C_{2,s} u^{(s)} \\ \vdots \\ \sum_{s=1}^p C_{p,s} u^{(s)} \end{pmatrix}.$$

An infinite dimensional matrix operator, B , is said to be zero, if all its entries are zero. An infinite dimensional matrix operator A is said to be Toeplitz if its entries satisfy $A_{t,\tau} = a_{t-\tau}$ for all $t, \tau \in \mathbb{Z}$ and for some sequence $\{a_r; r \in \mathbb{Z}\}$.

2.2. Covariance and inverse covariance operators. Within the nonstationary framework, we require the following assumptions on C to show that C is a mapping from $\ell_{2,p}$ to $\ell_{2,p}$ (and later that C^{-1} is a mapping from $\ell_{2,p}$ to $\ell_{2,p}$). For stationary time series, analogous assumptions are often made on the spectral density function (see Remark 2.1).

ASSUMPTION 2.1. Define $\lambda_{\sup} = \sup_{v \in \ell_{2,p}, \|v\|_2=1} \langle v, Cv \rangle$, $\lambda_{\inf} = \inf_{v \in \ell_{2,p}, \|v\|_2=1} \langle v, Cv \rangle$. Then

$$(2.2) \quad 0 < \lambda_{\inf} \leq \lambda_{\sup} < \infty.$$

Assumption 2.1 implies that $\sup_t \sup_a \sum_{\tau \in \mathbb{Z}} \sum_{b=1}^p [C_{t,\tau}]_{a,b}^2 < \infty$ and also the coefficients of the inverse are square summable. It can be shown that if $\sup_{t \in \mathbb{Z}} \sum_{\tau \in \mathbb{Z}} \|C_{t,\tau}\|_{\infty} < \infty$, then $\sup_{v \in \ell_{2,p}, \|v\|_2=1} \langle v, Cv \rangle < \infty$. This is analogous to a short memory condition for stationary time series. It is worth keeping in mind that cointegrated time series do not satisfy this condition. The theory developed in Sections 2 and 3 only require Assumption 2.1. However, to estimate the network stronger conditions on $\mathbf{D}_{t,\tau}$ are required, and these are stated in Section 4.

Under the above assumption $C : \ell_{2,p} \rightarrow \ell_{2,p}$, and since $C_{t,\tau} = C'_{\tau,t}$, $\langle v, Cu \rangle = \langle Cv, u \rangle$, thus C is a self-adjoint, bounded operator with $\|C\| = \lambda_{\sup}$, where $\|\cdot\|$ denotes the operator norm: $\|A\| = \sup_{u \in \ell_{2,p}, \|u\|_2=1} \|Au\|_2$.

REMARK 2.1. In the case of stationary time series, sufficient conditions for Assumption 2.1 to hold is that the eigenvalues of the spectral density matrix $\Sigma(\omega)$ are uniformly bounded away from zero and away from ∞ overall $\omega \in [0, \pi]$ (see, e.g., [9], Proposition 4.5.3).

The core theme of GGM is to learn conditional (partial) covariances between two variables after conditioning on a set of other variables. These conditional relationships can be derived from the inverse covariance matrix. Now, we will define a suitable inverse covariance operator $D = C^{-1}$ and show how its entries capture the conditional relationships. We will define these conditional relations in terms of *projections* with respect to the ℓ_2 -norm; this is equivalent to the least squares regression coefficients at the population level.

We consider the projection of $X_t^{(a)}$ onto $\mathcal{H} - X_t^{(a)}$, given by

$$(2.3) \quad P_{\mathcal{H}-X_t^{(a)}}(X_t^{(a)}) = \sum_{\tau \in \mathbb{Z}} \sum_{b=1}^p \beta_{(\tau,b) \rightarrow (t,a)} X_{\tau}^{(b)},$$

with $\beta_{(t,a) \rightarrow (t,a)} = 0$ (note the coefficients $\{\beta_{(\tau,b) \rightarrow (t,a)}\}$ are unique since C is nonsingular). Let $\sigma_{a,t}^2 = E[X_t^{(a)} - P_{\mathcal{H}-X_t^{(a)}}(X_t^{(a)})]^2$; it can be shown that $\sigma_{a,t}^2 \geq \lambda_{\min}$ (see Appendix A.1 of the Supplementary Material [4]). Analogous to finite dimensional covariance matrices, to obtain the entries of the inverse we use the coefficients of the projections of $X_t^{(a)}$ onto $\mathcal{H} - X_t^{(a)}$. For all $t, \tau \in \mathbb{Z}$, we define the $p \times p$ -dimensional matrices $\mathbf{D}_{t,\tau}$ as follows:

$$(2.4) \quad [\mathbf{D}_{t,\tau}]_{a,b} = \begin{cases} \frac{1}{\sigma_{a,t}^2} & a = b \text{ and } t = \tau, \\ -\frac{1}{\sigma_{a,t}^2} \beta_{(\tau,b) \rightarrow (t,a)} & \text{otherwise.} \end{cases}$$

Using $\mathbf{D}_{t,\tau}$, we define the infinite dimensional matrix

$$(2.5) \quad D_{a,b} = \{[D_{a,b}]_{t,\tau} = [\mathbf{D}_{t,\tau}]_{a,b}; t, \tau \in \mathbb{Z}\}.$$

Analogous to the definition of C , we define $D = (D_{a,b}; a, b \in \{1, \dots, p\})$.

Our next lemma shows that the operator D is indeed the inverse of the covariance operator C . We also state some upper bounds on its entries, which will be useful in our technical analysis.

LEMMA 2.1. *Suppose Assumption 2.1 holds. Let D be defined as in (2.4). Then $C^{-1} = D$ and $\|D\| = \lambda_{\inf}^{-1}$. Further, for all $a, b \in \{1, \dots, p\}$, $\|D_{a,b}\| \leq \lambda_{\inf}^{-1}$, $\|D_{a,a}^{-1}\| \leq \lambda_{\sup}$ and $\sup_t \sum_{\tau \in \mathbb{Z}} \|\mathbf{D}_{t,\tau}\|_2^2 \leq p\lambda_{\inf}^{-2}$.*

Proofs of results appear in the Supplementary Material [4].

2.3. Nonstationary graphical models (NonStGM). The operators C and D provide us with the objects needed to formally define the edges in our network and to connect them to the notions of conditional uncorrelatedness and conditional stationarity.

At this point we note an important distinction between edge construction in GGM and StGM, an issue that is crucial for generalizing graphical models to the nonstationarity case. In GGM conditional uncorrelatedness between two random variables is defined after conditioning on all the other random variables in the system. On the other hand, in StGM the conditional uncorrelatedness between two time series is defined after conditioning on all the other time series. This leads to two, potentially, different generalizations in the nonstationary setup. A direct generalization of the GGM framework would use the *partial covariances* $\text{Cov}(X_t^{(a)}, X_\tau^{(b)} | \mathcal{S}'_1)$, where $\mathcal{S}'_1 = \{X_s^{(c)} : (s, c) \notin \{(t, a), (\tau, b)\}\}$, while a generalization of the StGM framework, would suggest using *time series partial covariances* $\text{Cov}(X_t^{(a)}, X_\tau^{(b)} | \mathcal{S}'_2)$, where $\mathcal{S}'_2 = \{X_s^{(c)} : s \in \mathbb{Z}, c \notin \{a, b\}\}$.

To address this issue, we start by using the inverse covariance operator D to define edges that encode conditional uncorrelatedness and (non)stationarity. We show that, as expected, these notions are a direct generalization of the GGM framework. Then we present a surprising result (Theorem 2.2) that the encoding of the partial covariances in terms of the operator D remains unchanged, even if we adopt the StGM notion of partial covariance, that is, the conditionally uncorrelated and conditionally (non)stationary nodes, edges, subgraphs are preserved under the two frameworks.

We now define the network corresponding to the multivariate time series. Each edge in our network (V, E) will have an indicator to denote conditional invariance and conditional time-varying, a new notion we now introduce. The edge set E will contain all pairs (a, b) , where $\{X_t^{(a)}\}$ and $\{X_t^{(b)}\}$ are conditionally correlated. The edge set E will also contain self-loops that convey important information about the network. We start by formally defining the notions of conditional noncorrelation and (non)stationarity. This is stated in terms of the submatrices $\{D_{a,b}\}$ of D .

DEFINITION 2.1 (Nonstationary network). Conditional covariance and (non)stationarity of the components of a p -dimensional nonstationary time series are represented using a graph $G = (V, E)$, where $V = \{1, 2, \dots, p\}$ is the set of nodes, and $E \subseteq V \times V$ is a set of undirected edges $((a, b) \equiv (b, a))$ and includes *self-loops* of the form (a, a) :

- **Conditional noncorrelation:** The two time series $\{X_t^{(a)}\}$ and $\{X_t^{(b)}\}$ are conditionally uncorrelated if $D_{a,b} = 0$. As in GGM and StGM, this is represented by the absence of an edge between nodes a and b in the network, that is, $(a, b) \notin E$.

- *Conditionally stationary node*: The time series $\{X_t^{(a)}\}$ is conditionally stationary if $D_{a,a}$ is Toeplitz operator. We denote this using a solid self-loop (a, a) around the node a .
- *Conditionally time-invariant edge*: If $a \neq b$ and $D_{a,b}$ is a Toeplitz operator, then (a, b) is a conditionally time-invariant edge. We represent a conditionally time-invariant edge (a, b) in our network with a solid edge.
- *Conditionally stationary subgraph*: A subnetwork of nodes $\mathcal{S} \subset \{1, \dots, p\}$ is called a conditionally stationary subgraph if for all $a, b \in \mathcal{S}$, $D_{a,b}$ are Toeplitz operators, that is, $D_{\mathcal{S},\mathcal{S}}$ is a block Toeplitz operator.

As a special case of the above, we call a conditionally stationary subgraph of order two (consisting of the nodes $\{a, b\}$) a conditionally stationary pair if $D_{a,a}$, $D_{a,b}$, and $D_{b,b}$ are Toeplitz.

- *Conditionally nonstationary node/time-varying edge*: (i) If $D_{a,a}$ is not Toeplitz, then $\{X_t^{(a)}\}$ is conditionally nonstationary. (ii) For $a \neq b$, if $D_{a,b}$ is not Toeplitz, then (a, b) has a conditionally time-varying edge.

We represent conditional nonstationary nodes using a *dashed* self-loop and a conditionally time-varying edge with a dashed edge.

In Section 5 we show how the parameters of a general tvVAR model are related to the operator D and can be used to identify the network structure in NonStGM. As a concrete example, below we describe the network corresponding to the tvVAR(1) considered in the [Introduction](#).

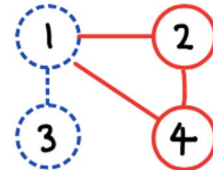
EXAMPLE 2.1. Consider the following tvVAR(1) model for a *four*-dimensional time series:

$$\begin{pmatrix} X_t^{(1)} \\ X_t^{(2)} \\ X_t^{(3)} \\ X_t^{(4)} \end{pmatrix} = \begin{pmatrix} \alpha(t) & 0 & \alpha_3 & 0 \\ \beta_1 & \beta_2 & 0 & \beta_4 \\ 0 & 0 & \gamma(t) & 0 \\ 0 & v_2 & 0 & v_4 \end{pmatrix} \begin{pmatrix} X_{t-1}^{(1)} \\ X_{t-1}^{(2)} \\ X_{t-1}^{(3)} \\ X_{t-1}^{(4)} \end{pmatrix} + \underline{\varepsilon}_t = A(t)\underline{X}_{t-1} + \underline{\varepsilon}_t,$$

where $\{\underline{\varepsilon}_t\}$ are independent random variables (i.i.d) with $\underline{\varepsilon}_t \sim N(0, I_4)$, and $\alpha(t)$, $\gamma(t)$ are smoothly varying functions of t . The four time series are marginally nonstationary; in the sense that for each $1 \leq a \leq 4$, the time series $\{X_t^{(a)}\}$ is second order nonstationary.

The inverse operator and network corresponding to $\{\underline{X}_t\}$ is given below and is deduced from the transition matrix $A(t)$ (the explicit connection between D and $\{A(t)\}$ is given in Section 5). Note that red and blue denote Toeplitz and non-Toeplitz matrix operators, respectively.

$$D = \begin{pmatrix} D_{1,1} & D_{1,2} & D_{1,3} & D_{1,4} \\ D_{2,1} & D_{2,2} & 0 & D_{2,4} \\ D_{3,1} & 0 & D_{3,3} & 0 \\ D_{4,1} & D_{4,2} & 0 & D_{4,4} \end{pmatrix}$$



Connecting the transition matrix to the network. The connections between the nodes is because node 1 is connected to node 3 (if $\alpha_3\alpha(t) \neq 0$ for some t), node 4 (if $\beta_1\beta_4 \neq 0$), and node 2 (if $\beta_1\beta_2 \neq 0$). By a similar argument, nodes 2 and 4 are connected (if $\beta_2\beta_4 \neq 0$ or $v_2v_4 \neq 0$).

The nonstationarity of the multivariate time series is due to the time-varying parameters $\alpha(t)$ and $\gamma(t)$. Specifically, the parameter $\alpha(t)$ is the reason that node 1 is nonstationary, and by a similar argument the time-varying parameter $\gamma(t)$ is the reason node 3 is nonstationary. Since the coefficients on the second and fourth columns are not time-varying, nodes 2

and 4 have “inherited” their nonstationarity from nodes 1 and 3. Thus, nodes 1 and 3 are conditionally stationary whereas nodes 2 and 4 are conditionally stationary. The connections between nodes 1 to 2 and 1 to 4 are time invariant because $\beta_1\beta_2$ and $\beta_1\beta_4$ are time invariant respectively.

REMARK 2.2 (Connection to GGM). Let $X^{(a)} = (X_t^{(a)}; t \in \mathbb{Z})$. It is clear that the density of the infinite dimensional vector $(X^{(a)}; 1 \leq a \leq p)$ is not well defined. However, we can informally view the joint density (at least in the Gaussian case) as “being proportional to”

$$\exp\left(-\frac{1}{2} \sum_{a=1}^p \langle X^{(a)}, D_{a,a} X^{(a)} \rangle - \frac{1}{2} \sum_{(a,b) \in E, a \neq b} \langle X^{(a)}, D_{a,b} X^{(b)} \rangle\right).$$

This is analogous to the representation of multivariate Gaussian vector in terms of its inverse covariance. Using the above representation, we conjecture that the above notions of conditional correlation/stationarity/nonstationarity can be generalized to time series which is not necessarily continuous valued, for example, binary valued time series.

2.4. NonStGM as a generalization of GGM. We start by defining partial covariances in the spirit of the definition used in GGM but for infinite dimensional random variables. This is defined by removing *two random variables* from the spanning set of $\mathcal{H}$

$$(2.6) \quad \rho_{t,\tau}^{(a,b)} = \mathbb{Cov}[X_t^{(a)} - P_{\mathcal{H}-(X_t^{(a)}, X_\tau^{(b)})}(X_t^{(a)}), X_\tau^{(b)} - P_{\mathcal{H}-(X_t^{(a)}, X_\tau^{(b)})}(X_\tau^{(b)})].$$

Note that for the case $t = \tau$ and $a = b$, the above reduces to

$$(2.7) \quad \rho_{t,t}^{(a,a)} = \mathbb{Var}[X_t^{(a)} - P_{\mathcal{H}-X_t^{(a)}}(X_t^{(a)})] = \sigma_{a,t}^2.$$

In the discussion below, we refer to the infinite dimensional conditional covariance matrices $\rho^{(a,b)} = (\rho_{t,\tau}^{(a,b)}; t, \tau \in \mathbb{Z})$ and $\rho^{(a,a)} = (\rho_{t,\tau}^{(a,a)}; t, \tau \in \mathbb{Z})$. In GGM the partial covariances are encoded in the precision matrix. In a similar spirit, we show that $\rho_{t,\tau}^{(a,b)}$ is encoded in the inverse covariance operator D .

LEMMA 2.2. Suppose Assumption 2.1 holds. Let $D_{a,b}$ be defined as in (2.5). Then the entries of $D_{a,b}$ satisfy the identities

$$(2.8) \quad \mathbb{Corr}[X_t^{(a)} - P_{\mathcal{H}-(X_t^{(a)}, X_\tau^{(b)})}(X_t^{(a)}), X_\tau^{(b)} - P_{\mathcal{H}-(X_t^{(a)}, X_\tau^{(b)})}(X_\tau^{(b)})] = -\frac{[D_{a,b}]_{t,\tau}}{\sqrt{[D_{a,a}]_{t,t}[D_{b,b}]_{\tau,\tau}}}$$

and

$$(2.9) \quad \mathbb{Var}\left[\begin{pmatrix} X_t^{(a)} - P_{\mathcal{H}-(X_t^{(a)}, X_\tau^{(b)})}(X_t^{(a)}) \\ X_\tau^{(b)} - P_{\mathcal{H}-(X_t^{(a)}, X_\tau^{(b)})}(X_\tau^{(b)}) \end{pmatrix}\right] = \begin{pmatrix} [D_{a,a}]_{t,t} & [D_{a,b}]_{t,\tau} \\ [D_{b,a}]_{\tau,t} & [D_{b,b}]_{\tau,\tau} \end{pmatrix}^{-1}.$$

Proofs of results appear in the Supplementary Material [4].

An immediate consequence of Lemma 2.2 is that the notions of conditional noncorrelation and conditional stationarity can be equivalently defined in terms of the properties of the partial covariances $\rho^{(a,b)}$. In particular, conditional noncorrelation between the two series a and b translates to zero $\rho^{(a,b)}$, while conditional stationarity of the pair (a, b) translates to Toeplitz structures on $\rho^{(a,a)}$, $\rho^{(b,b)}$ and $\rho^{(a,b)}$. It is worth noting that the Toeplitz structure of $\rho^{(a,a)}$ (the partial covariance of a) captured in our framework is an important property, namely, the conditional (non)stationarity of a node. A similar role on the diagonal entries of the precision or spectral precision matrices ($\Theta_{a,a}$ or $[\Sigma^{-1}(\omega)]_{a,a}$) is absent in both the classical GGM and StGM frameworks.

PROPOSITION 2.1 (NonStGM in terms of $\rho_{t,\tau}^{(a,b)}$). Suppose Assumption 2.1 holds. Let $\rho_{t,\tau}^{(a,b)}$ be defined as in (2.6). Then:

- (i) Conditional noncorrelation: $\rho_{t,\tau}^{(a,b)} = 0$ for all t and τ (i.e., $\rho^{(a,b)} = 0$) iff $D_{a,b} = 0$.
- (ii) Conditionally stationary node: $D_{a,a}$ is Toeplitz iff for all t and τ

$$\rho_{t,\tau}^{(a,a)} = \rho_{0,t-\tau}^{(a,a)},$$

that is, $\rho^{(a,a)}$ is Toeplitz.

- (iii) Conditionally stationary pair: $D_{a,a}$, $D_{b,b}$ and $D_{a,b}$ are Toeplitz iff for all t and τ

$$\mathbb{V}\text{ar} \left[\begin{pmatrix} X_t^{(a)} - P_{\mathcal{H}-(X_t^{(a)}, X_\tau^{(b)})}(X_t^{(a)}) \\ X_\tau^{(b)} - P_{\mathcal{H}-(X_t^{(a)}, X_\tau^{(b)})}(X_\tau^{(b)}) \end{pmatrix} \right] = \begin{pmatrix} \rho_{0,0}^{(a,a)|-\{(a,0),(b,\tau-\tau)\}} & \rho_{0,t-\tau}^{(a,b)} \\ \rho_{0,t-\tau}^{(a,b)} & \rho_{0,0}^{(b,b)|-\{(a,t-\tau),(b,0)\}} \end{pmatrix}$$

(where $\rho_{t,t}^{(a,a)|-\{(a,t),(b,\tau)\}} = \mathbb{V}\text{ar}[X_t^{(a)} - P_{\mathcal{H}-(X_t^{(a)}, X_\tau^{(b)})}(X_t^{(a)})]$), that is, $\rho^{(a,a)}$, $\rho^{(b,b)}$ and $\rho^{(a,b)}$ are Toeplitz.

Proofs of results appear in the Supplementary Material [4].

2.5. *NonStGM as a generalization of StGM.* Now, we define the time series partial covariance analogous to that used in StGM. We recall that the classical time series definition of partial covariance in a multivariate time series evaluates the covariance between two random variables $X_t^{(a)}$ and $X_\tau^{(b)}$ after conditioning on all random variables in the $(p-2)$ component series $V \setminus \{a, b\}$. In other words, we exclude the *entire* time series a and b from the conditioning set.

Formally, for any $\mathcal{S} \subseteq V$, we define the residual of $X_t^{(a)}$ after projecting on $\overline{\text{sp}}(X_s^{(c)}; s \in \mathbb{Z}, c \notin \mathcal{S}) = \mathcal{H} - (X^{(c)}; c \in \mathcal{S})$ as

$$X_t^{(a)|-\mathcal{S}} := X_t^{(a)} - P_{\mathcal{H}-(X^{(c)}; c \in \mathcal{S})}(X_t^{(a)}) \quad \text{for } t \in \mathbb{Z}.$$

In the definitions below, we focus on the two sets $\mathcal{S} = \{a, b\}$ and $\mathcal{S} = \{a\}$. We mention that the set $\mathcal{S} = \{a\}$ is not considered in StGMM but plays an important role in NonStGM. Using the above, we define the edge partial covariance

$$(2.10) \quad \begin{pmatrix} \rho_{t,\tau}^{(a,a)|-\{a,b\}} & \rho_{t,\tau}^{(a,b)|-\{a,b\}} \\ \rho_{t,\tau}^{(b,a)|-\{a,b\}} & \rho_{t,\tau}^{(b,b)|-\{a,b\}} \end{pmatrix} := \mathbb{C}\text{ov} \left[\begin{pmatrix} X_t^{(a)|-\{a,b\}} \\ X_\tau^{(b)|-\{a,b\}} \end{pmatrix}, \begin{pmatrix} X_\tau^{(a)|-\{a,b\}} \\ X_t^{(b)|-\{a,b\}} \end{pmatrix} \right]$$

and node partial covariance

$$(2.11) \quad \rho_{t,\tau}^{(a,a)|-\{a\}} = \mathbb{C}\text{ov}[X_t^{(a)|-\{a\}}, X_\tau^{(a)|-\{a\}}].$$

We will show that the partial covariance in (2.10) and (2.11) are closely related to the partial covariance in (2.6). In Lemma 2.2 we have shown that the partial correlations $\rho_{t,\tau}^{(a,b)}$ define the entries of the operator D . We now connect the time series definition of a partial covariance to the operator $D = (D_{a,b}; a, b \in \{1, \dots, p\})$. Before we present the equivalent definitions of our nonstationary networks in terms of the time series partial covariances $\rho_{t,\tau}^{(a,b)|-\mathcal{S}}$, we show that $\rho_{t,\tau}^{(a,b)|-\mathcal{S}}$ can be expressed in terms of the inverse covariance operator D .

THEOREM 2.1. Suppose Assumption 2.1 holds. Let $\rho_{t,\tau}^{(a,a)|-\{a,b\}}$, $\rho_{t,\tau}^{(a,b)|-\{a,b\}}$ and $\rho_{t,\tau}^{(a,a)|-\{a\}}$ be defined as in (2.10) and (2.11), respectively. Then:

- (i) $\rho_{t,\tau}^{(a,a)|-\{a\}} = [D_{a,a}^{-1}]_{t,\tau}$.

(ii) If $a \neq b$, then

$$(2.12) \quad \mathbb{V}\text{ar}[X_t^{(c)|-\{a,b\}}; t \in \mathbb{Z}, c \in \{a, b\}] = \begin{pmatrix} D_{a,a} & D_{a,b} \\ D_{b,a} & D_{b,b} \end{pmatrix}^{-1}$$

with

$$\begin{aligned} \rho_{t,\tau}^{(a,a)|-\{a,b\}} &= [-(D_{a,a} - D_{a,b}D_{b,b}^{-1}D_{b,a})^{-1}D_{a,b}D_{b,b}^{-1}]_{t,\tau}, \\ \rho_{t,\tau}^{(a,b)|-\{a,b\}} &= [(D_{a,a} - D_{a,b}D_{b,b}^{-1}D_{b,a})^{-1}]_{t,\tau}, \\ \rho_{t,\tau}^{(b,b)|-\{a,b\}} &= [(D_{b,b} - D_{b,a}D_{a,a}^{-1}D_{a,b})^{-1}]_{t,\tau}. \end{aligned}$$

Proofs of results appear in the Supplementary Material [4].

A careful examination of the expressions for the GGM covariance $\rho_{t,\tau}^{(a,b)}$, given in Lemma 2.2, with the StGM covariance, given in $\rho_{t,\tau}^{(a,b)|-\{a,b\}}$, shows they are very different quantities. Therefore, it is suprising that despite these stark differences they preserve the same structures. More precisely, in Proposition 2.1 we showed that Definition 2.1 had a clear interpretation in terms of $\rho_{t,\tau}^{(a,b)}$. We show below that the network definition, given in Definition 2.1, can be interpreted in terms of the conditional dependence (or residuals) of the time series. The fact that two very different conditional covariance definitions lead to the same conditional graph is due to the property that infinite dimensional Toeplitz operators remain Toeplitz, even after inversion and multiplication with other Toeplitz operators.

THEOREM 2.2. [NonStGM in terms of $\rho^{(a,b)|-\{a,b\}}$] Suppose Assumption 2.1 holds. Let $\rho_{t,\tau}^{(a,a)|-\{a,b\}}$, $\rho_{t,\tau}^{(a,b)|-\{a,b\}}$ and $\rho_{t,\tau}^{(b,b)|-\{a,b\}}$ be defined as in (2.10) and (2.11), respectively. Then:

- (i) Conditional noncorrelation: $D_{a,b} = 0$ iff $\rho_{t,\tau}^{(a,a)|-\{a,b\}} = 0$ for all t and τ .
- (ii) Conditionally stationary node: $D_{a,a}$ is a Toeplitz operator iff for all t and τ , $\rho_{t,\tau}^{(a,a)|-\{a,b\}} = \rho_{0,t-\tau}^{(a,a)|-\{a,b\}}$.
- (iii) Conditionally stationary pair: $D_{a,a}$, $D_{b,b}$ and $D_{a,b}$ are Toeplitz iff for all t and τ , $\rho_{t,\tau}^{(a,a)|-\{a,b\}} = \rho_{0,t-\tau}^{(a,a)|-\{a,b\}}$, $\rho_{t,\tau}^{(b,b)|-\{a,b\}} = \rho_{0,t-\tau}^{(b,b)|-\{a,b\}}$, and $\rho_{t,\tau}^{(a,b)|-\{a,b\}} = \rho_{0,t-\tau}^{(a,b)|-\{a,b\}}$.

Proofs of results appear in the Supplementary Material [4].

We show in the following result that the *time series partial covariances* can be used to define conditional stationarity of a subgraph containing three or more nodes.

COROLLARY 2.1 (Conditionally stationary subgraph). Let $\mathcal{S} = \{\alpha_1, \dots, \alpha_r\}$ be a subset of $\{1, \dots, p\}$ and $\mathcal{S}'$ denote the complement of $\mathcal{S}$. Suppose for all $a, b \in \mathcal{S}$, $D_{a,b}$ are Toeplitz (including the case $a = b$). Then $\{X_t^{(a)}; t \in \mathbb{Z}, a \in \mathcal{S}\}$ is a conditionally stationary subgraph, where

$$\mathbb{V}\text{ar}[X_t^{(a)} - P_{\mathcal{H}-(X^{(c)}; c \in \mathcal{S}')} (X_t^{(a)}); t \in \mathbb{Z}, a \in \mathcal{S}] = P$$

with

$$P^{-1} = \begin{pmatrix} D_{\alpha_1, \alpha_1} & D_{\alpha_1, \alpha_2} & \cdots & D_{\alpha_1, \alpha_r} \\ D_{\alpha_2, \alpha_1} & D_{\alpha_2, \alpha_2} & \cdots & D_{\alpha_2, \alpha_r} \\ \vdots & \vdots & \ddots & \vdots \\ D_{\alpha_r, \alpha_1} & D_{\alpha_r, \alpha_2} & \cdots & D_{\alpha_r, \alpha_r} \end{pmatrix}.$$

Proofs of results appear in the Supplementary Material [4].

3. Sparse characterisations within the Fourier domain. For general nonstationary processes, it is infeasible to estimate the operator D and learn its network within the time domain. The problem is akin to StGM, where it is difficult to learn the graph structure in the time domain by studying all the autocovariance matrices. Estimation is typically carried out in the Fourier domain by detecting conditional independence from the zeros of $\Sigma^{-1}(\omega)$. Following the same route, we will switch to the Fourier domain and construct a quantity that can be used to “detect zeros and nonzeros.” In addition, within the Fourier domain we will define meaningful notions of weights/strengths of conditionally stationary nodes and pairs that are analogous to well-known partial spectral coherence measures used in StGM.

Notation. We first summarize some of the notation we will use in this section. We define the function space of square integrable functions $L_2[0, 2\pi)$ as all complex functions, where $g \in L_2[0, 2\pi)$ if $\int_0^{2\pi} |g(\omega)|^2 d\omega < \infty$. We define the function space of all square summable vector complex functions $L_2[0, 2\pi)^p$, where $\underline{g}(\omega)' = (g_1(\omega), \dots, g_p(\omega)) \in L_2[0, 2\pi)^p$ if for all $1 \leq j \leq p$ $g_j \in L_2[0, 2\pi)$. For all $\underline{g}, \underline{h} \in L_2[0, 2\pi)^p$, we define the inner-product $\langle \underline{g}, \underline{h} \rangle = \sum_{j=1}^p \langle g_j, h_j \rangle$, where $\langle g_j, h_j \rangle = \int_0^{2\pi} g_j(\omega) h_j(\omega)^* d\omega$. Note that $L_2[0, 2\pi)^p$ is a Hilbert space. We use $\delta_{\omega, \lambda}$ to denote the Dirac delta function and set $i = \sqrt{-1}$.

3.1. Transformation to the Fourier domain. In this section we summarize results which are pivotal to the development in the subsequent sections. This section can be skipped on first reading.

To connect the time and Fourier domain, we define a transformation between the sequence and function space. We define the functions $F : L_2[0, 2\pi) \rightarrow \ell_2$ and $F^* : \ell_2 \rightarrow L_2[0, 2\pi)$, where

$$(3.1) \quad [F(g)]_j = \frac{1}{2\pi} \int_0^{2\pi} \underline{g}(\lambda) \exp(ij\lambda) d\lambda \quad \text{and} \quad F^*(v)(\omega) = \sum_{j \in \mathbb{Z}} v_j \exp(-ij\omega).$$

It is well known that F and F^* are isomorphisms between ℓ_2 and $L_2[0, 2\pi)$ (see, e.g., [9], Section 2.9). For $d > 1$, the transformations $F(\underline{g}) = (F(g_1), \dots, F(g_d))$ and $F^*v = (F^*v^{(1)}, \dots, F^*v^{(d)})$, where $v = (v^{(1)}, \dots, v^{(d)})$ are isomorphisms between $\ell_{2,d}$ and $L_2[0, 2\pi)^d$. Often, we use that $d = p$. These two isomorphisms will provide a link between the infinite dimensional matrix operators D defined in the time domain to an equivalent operator in the Fourier domain.

Let $A = (A_{a,b}; a, b \in \{1, \dots, d\})$; if $A : \ell_{2,d} \rightarrow \ell_{2,d}$ is a bounded operator, then standard results show that $F^*AF : L_2[0, 2\pi)^d \rightarrow L_2[0, 2\pi)^d$ is a bounded operator (see [12], Chapter II). F^*AF is an integral operator, such that, for all $g \in L_2[0, 2\pi)^d$,

$$(3.2) \quad F^*AF(g)[\omega] = \frac{1}{2\pi} \int_0^{2\pi} \mathbf{A}(\omega, \lambda) \underline{g}(\lambda) d\lambda,$$

and $\mathbf{A}$ is the $d \times d$ -dimensional matrix integral kernel, where

$$\mathbf{A}(\omega, \lambda) = \left(\sum_{t \in \mathbb{Z}} \sum_{\tau \in \mathbb{Z}} [A_{a,b}]_{t,\tau} \exp(it\omega - i\tau\lambda); a, b \in \{1, \dots, d\} \right).$$

To understand how A and $\mathbf{A}(\omega, \lambda)$ are related, we focus on the case $d = 1$ and note that the (t, τ) entry of the infinite dimensional matrix A is

$$A_{t,\tau} = \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^{2\pi} \mathbf{A}(\omega, \lambda) \exp(-it\omega + i\tau\lambda) d\omega d\lambda \quad \text{for all } t, \tau \in \mathbb{Z}.$$

REMARK 3.1 (Connection with covariances and stationary time series). We note if C were a covariance operator of a univariate time series $\{X_t\}$ with integral kernel G , then

$$(3.3) \quad \mathbb{Cov}[X_t, X_\tau] = C_{t,\tau} = \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^{2\pi} G(\omega, \lambda) \exp(-it\omega + i\tau\lambda) d\omega d\lambda,$$

where $G(\omega, \lambda)$ is the Loève dual frequency spectrum. The Loève dual frequency spectrum is used to describe nonstationary features in a time series and has been extensively studied in [1, 33–36, 41, 43, 49, 50].

If $\{X_t\}$ were a second order stationary time series, then (3.3) reduces to Bochner's Theorem

$$\mathbb{Cov}[X_t, X_\tau] = C_{0,t-\tau} = \frac{1}{(2\pi)} \int_0^{2\pi} f(\omega) \exp(-i(t-\tau)\omega) d\omega.$$

The relationship between the spectral density function $f(\omega)$ and the Loève dual frequency spectrum $G(\omega, \lambda)$ is made apparent in Lemma 3.1 below.

$A(\omega, \lambda)$ is a formal representation, and typically, it will not be a well-defined function over $[0, 2\pi)^2$, as it is likely to have singularities. Despite this, it has a very specific sparsity structure when the operator A is Toeplitz. For the identification of nodes and edges in the nonstationary networks, it is the location of zeros in $A(\omega, \lambda)$ that we will exploit. This will become apparent in the following lemma due to [64] (we state the result for the case $d = 1$).

LEMMA 3.1. *Suppose A is an infinite dimensional bounded matrix operator $A : \ell_2 \rightarrow \ell_2$. The matrix operator A is Toeplitz iff the integral kernel associated with F^*AF has the form*

$$A(\omega, \lambda) = \delta_{\omega, \lambda} A(\omega),$$

where $A(\omega) \in L_2[0, 2\pi)$ and $\delta_{\omega, \lambda}$ is the Dirac delta function.

Proofs of results appear in the Supplementary Material [4].

The crucial observation in the above lemma is that $A(\omega, \lambda) = 0$ for $\lambda \neq \omega$ iff A is a Toeplitz matrix. Below we generalize the above to the case that A (and its inverse) is a block Toeplitz matrix operator.

LEMMA 3.2. *Suppose that A is an infinite dimensional, symmetric, block matrix operator $A : \ell_{2,d} \rightarrow \ell_{2,d}$, where $0 < \inf_{\|v\|_2=1} \langle v, Av \rangle \leq \sup_{\|v\|_2=1} \langle v, Av \rangle < \infty$ with $A = (A_{a,b}; a, b \in \{1, \dots, d\})$ and $A_{a,b}$ is Toeplitz. Then the integral kernel associated with F^*AF is $A(\omega, \lambda) = A(\omega) \delta_{\omega, \lambda}$, where $A(\omega)$ is a $d \times d$ matrix with entries $[A(\omega)]_{a,b} = \sum_{r \in \mathbb{Z}} [A_{a,b}]_{0,r} \exp(ir\omega)$. Further, the integral kernel associated with $F^*A^{-1}F$ is $A(\omega)^{-1} \delta_{\omega, \lambda}$.*

Proofs of results appear in the Supplementary Material [4].

From now on we say that the kernel $A(\omega, \lambda)$ is diagonal if it can be represented as $\delta_{\omega, \lambda} A(\omega)$.

We use the operators $F : L_2[0, 2\pi)^p \rightarrow \ell_{2,p}$ and $F^* : \ell_{2,p} \rightarrow L_2[0, 2\pi)^p$ to recast the covariance and inverse covariance operators of a multivariate time series within the Fourier domain. We recall that C is the covariance operator of the time series $\{\underline{X}_t\}$, and by using (3.2), F^*CF is an integral operator with matrix kernel $C(\omega, \lambda) = (C_{a,b}(\omega, \lambda); a, b \in \{1, \dots, p\})$, where $C_{a,b}(\omega, \lambda) = \sum_{t \in \mathbb{Z}} \sum_{\tau \in \mathbb{Z}} [C_{a,b}]_{t,\tau} \exp(it\omega - i\tau\lambda)$.

In the case that $\{\underline{X}_t\}$ is second order stationary, then $C_{a,b}(\omega, \lambda) = [\Sigma(\omega)]_{a,b} \delta_{\omega, \lambda}$ where $\Sigma(\cdot)$ is the spectral density matrix of $\{\underline{X}_t\}$. However, if $\{\underline{X}_t\}$ is second order nonstationary,

then by Lemma 3.1 at least one of the kernels $C_{a,b}(\omega, \lambda)$ will be nondiagonal. The dichotomy that the mass of $\mathbf{C}(\omega, \lambda)$ lies on the diagonal $\omega = \lambda$ if and only if the underlying process is multivariate second order stationary and is used in [27, 30, 37] to test for second order stationarity.

3.2. The nonstationary inverse covariance in the Fourier domain. The covariance operator C and corresponding integral kernel $\mathbf{C}(\omega, \lambda)$ does not distinguish between *direct* and *indirect* nonstationary relationships. We have shown in Section 2 that conditional relationships are encoded in the inverse covariance D . Therefore, in this section we study the properties of the integral kernel corresponding to F^*DF . Under Assumption 2.1, $D = C^{-1}$ is a bounded operator; thus, F^*DF is a bounded operator defined by the matrix kernel $\mathbf{K}(\omega, \lambda) = (K_{a,b}(\omega, \lambda); a, b \in \{1, \dots, p\})$, where

$$(3.4) \quad K_{a,b}(\omega, \lambda) = \sum_{t \in \mathbb{Z}} \sum_{\tau \in \mathbb{Z}} [D_{a,b}]_{t,\tau} \exp(it\omega - i\tau\lambda) = \sum_{t \in \mathbb{Z}} \Gamma_t^{(a,b)}(\lambda) \exp(it(\omega - \lambda))$$

and

$$(3.5) \quad \Gamma_t^{(a,b)}(\lambda) = \sum_{r \in \mathbb{Z}} [D_{a,b}]_{t,t+r} \exp(ir\lambda).$$

Note that, under Assumption 2.1 $\|D\| < \infty$, this implies for all $a, b \in \{1, \dots, p\}$ that the sequence $\{[D_{a,b}]_{t,t+r}\}_r \in \ell_2$, thus $\Gamma_t^{(a,b)}(\cdot) \in L_2[0, 2\pi]$. As far as we are aware, neither $K_{a,b}(\omega, \lambda)$ nor $\Gamma_t^{(a,b)}(\lambda)$ have been studied previously. But $\Gamma_t^{(a,b)}(\lambda)$ can be viewed as the inverse covariance version of the time-varying spectrum that is commonly used to analyze nonstationary covariances (see [5, 13, 46, 54]). We observe that $D_{a,b}$ is Toeplitz if and only if $\Gamma_t^{(a,b)}(\lambda)$ does not depend on t .

In the following theorem, we show that $\mathbf{K}(\omega, \lambda)$ defines a very clear sparsity pattern depending on the conditional properties of $\{X_t\}$. This will allow us to discriminate between different types of edges in a network. In particular, zero matrices $D_{a,b}$ map to zero kernels and Toeplitz matrices $D_{a,b}$ map to diagonal kernels.

THEOREM 3.1. *Suppose Assumption 2.1 holds. Then:*

- (i) *Conditionally noncorrelated: $\{X_t^{(a)}, X_t^{(b)}\}_t$ are conditionally noncorrelated iff $K_{a,b}(\omega, \lambda) \equiv 0$ for all $\omega, \lambda \in [0, 2\pi]$.*
- (ii) *Conditionally stationary node: $\{X_t^{(a)}\}_t$ is conditionally stationary iff the integral kernel $K_{a,a}(\omega, \lambda)$ is diagonal.*
- (iii) *Conditionally time-invariant edge: The edge (a, b) is conditionally time invariant iff the integral kernel $K_{a,b}(\omega, \lambda)$ is diagonal.*

Proofs of results appear in the Supplementary Material [4].

These equivalences show that conditional noncorrelatedness and stationarity relationships in the graphical model, as defined by the D operator, are encoded in the object $K(\cdot, \cdot)$. This provides the foundation for an alternate route to learning the graph structure in the frequency domain.

EXAMPLE 3.1. We return to tvAR(1) model described in Example 2.1. In Figure 2 we give a schematic illustration of the matrix D in the frequency domain.

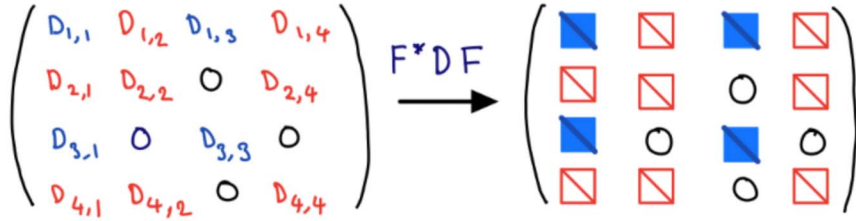


FIG. 2. Illustration of the mapping of matrix D to the integral kernel corresponding to F^*DF . The diagonal red box indicates the mass of $F^*D_{a,b}F$ lies only on the diagonal (it corresponds to a Toeplitz matrix). The blue filled box indicates that the mass of $F^*D_{a,b}F$ lies both on the diagonal and elsewhere (it corresponds to a non-Toeplitz matrix).

3.3. Partial spectrum for conditionally stationary time series. So far we have considered the construction of an undirected, unweighted network which encodes the conditional uncorrelation and nonstationarity properties of time series components. In practice, we would be interested in assigning weights to network edges that represent the strength or magnitude of these conditional relationships. This will also be useful for learning the graph structure from finite samples. In GGM partial correlation values are used to define edge weights. In StGM the partial spectral coherence (the frequency domain analogue of partial correlation) is used to define suitable edge weights.

We now define the notion of partial spectral coherence for conditionally stationary time series. We start by interpreting $\Gamma_t^{(a,a)}(\omega)$ and $\Gamma_t^{(a,b)}(\omega)$, defined in (3.5), in the case that the node or edge is conditionally stationary. In the following proposition, we relate these quantities to the partial covariance $\rho_{t,\tau}^{(a,b)}$. Analogous to the definition of $\rho_{a,b}^{(t,\tau)}$, we define the partial correlation

$$(3.6) \quad \phi_{t,\tau}^{(a,b)} = \mathbb{C}\text{orr}[X_t^{(a)} - P_{\mathcal{H}-(X_t^{(a)}, X_\tau^{(b)})}(X_t^{(a)}), X_\tau^{(b)} - P_{\mathcal{H}-(X_t^{(a)}, X_\tau^{(b)})}(X_\tau^{(b)})].$$

Using the above, we obtain an expression for $\Gamma_t^{(a,a)}(\omega)$ and $\Gamma_t^{(a,b)}(\omega)$ in the case that an edge or a node is conditionally stationary.

THEOREM 3.2. Suppose Assumption 2.1 holds. Let $\rho_{t,\tau}^{(a,b)}$ and $\phi_{t,\tau}^{(a,b)}$ be defined as in (2.6) and (3.6):

(i) If the node a is conditionally stationary, then $\Gamma_t^{(a,a)}(\omega) = \Gamma^{(a,a)}(\omega)$ for all t , where

$$\Gamma^{(a,a)}(\omega) = \sum_{r=-\infty}^{\infty} [D_{a,a}]_{(0,r)} \exp(ir\omega) = \frac{1}{\rho_{0,0}^{(a,a)}} \left[1 - \sum_{r \in \mathbb{Z} \setminus \{0\}} \phi_{0,r}^{(a,a)} \exp(ir\omega) \right].$$

(ii) If (a, b) is a conditionally stationary pair, then expressions for $\Gamma^{(a,a)}(\omega)$ and $\Gamma^{(b,b)}(\omega)$ are given in (i) and $\Gamma_t^{(a,b)}(\omega) = \Gamma^{(a,b)}(\omega)$ for all t , where

$$\Gamma^{(a,b)}(\omega) = \sum_{r=-\infty}^{\infty} [D_{a,b}]_{(0,r)} \exp(ir\omega) = -\frac{1}{(\rho_{0,0}^{(a,a)} \rho_{0,0}^{(b,b)})^{1/2}} \sum_{r \in \mathbb{Z}} \phi_{0,r}^{(a,b)} \exp(ir\omega).$$

Proofs of results appear in the Supplementary Material [4].

For StGM the partial spectral coherence is typically defined in terms of the Fourier transform of the partial time series covariances (see [55], Section 9.3, and [14]). We now show that an analogous result holds in the case of conditional stationarity.

THEOREM 3.3. Suppose Assumption 2.1 holds:

(i) If the node a is conditionally stationary, then

$$\sum_{r \in \mathbb{Z}} \rho_{0,r}^{(a,a)|-\{a\}} \exp(ir\omega) = \Gamma^{(a,a)}(\omega)^{-1}.$$

(ii) If (a, b) is a conditionally stationary pair, then

$$\sum_{r \in \mathbb{Z}} \begin{pmatrix} \rho_{0,r}^{(a,a)|-\{a,b\}} & \rho_{0,r}^{(a,b)|-\{a,b\}} \\ \rho_{0,r}^{(b,a)|-\{a,b\}} & \rho_{0,r}^{(b,b)|-\{a,b\}} \end{pmatrix} \exp(ir\omega) = \begin{pmatrix} \Gamma^{(a,a)}(\omega) & \Gamma^{(a,b)}(\omega) \\ \Gamma^{(a,b)}(\omega)^* & \Gamma^{(b,b)}(\omega) \end{pmatrix}^{-1}.$$

Proofs of results appear in the Supplementary Material [4].

The above allows us to define the notion of spectral partial coherence in the case that underlying time series is nonstationary. We recall that the spectral partial coherence between $\{X_t^{(a)}\}_t$ and $\{X_t^{(b)}\}_t$ for stationary time series is the standardized spectral conditional covariance (see [14]). Analogously, by using Theorem 3.3(ii) the spectral partial coherence between the conditionally stationary pair (a, b) is

$$(3.7) \quad R_{a,b}(\omega) = -\frac{\Gamma^{(a,b)}(\omega)}{\sqrt{\Gamma^{(a,a)}(\omega)\Gamma^{(b,b)}(\omega)}}.$$

In Appendix F of the Supplementary Material [4], we show how this expression is related to the spectral partial coherence for stationary time series.

3.4. Connection to nodewise regression. In Lemma 2.1 we connected the coefficients of D to the coefficients in a linear regression. The regressors are in the spanning set of $\mathcal{H} - X_t^{(a)}$. In contrast, in nodewise regression each node is regressed on all of the *other* nodes (the coefficients in this regression can also be connected to the precision matrix). We now derive an analogous result for multivariate time series. In particular, we regress the time series at node a ($\{X_t^{(a)}\}_t$) onto all the other time series (excluding node a , i.e., the spanning set of $\mathcal{H} - (X^{(a)})$) and connect these to the matrix D . These results can be used to encode conditions for a conditionally stationary edge in terms of the regression coefficients. Furthermore, they allow us to deduce the time series at node a , conditioned on all the other nodes (if the time series is Gaussian).

The best linear predictor of $X_t^{(a)}$, given the “other” time series $\{X_s^{(b)}; s \in \mathbb{Z}, b \neq a\}$, is

$$(3.8) \quad P_{\mathcal{H}-(X^{(a)})}(X_t^{(a)}) = \sum_{b \neq a} \sum_{\tau \in \mathbb{Z}} \alpha_{(\tau,b) \rightarrow (t,a)} X_\tau^{(b)}.$$

We group the coefficients according to time series and define the infinite dimensional matrix $B_{b \rightarrow a}$ with entries

$$(3.9) \quad [B_{b \rightarrow a}]_{t,\tau} = \alpha_{(\tau,b) \rightarrow (t,a)} \quad \text{for all } t, \tau \in \mathbb{Z}.$$

In the lemma below, we connect the coefficients in the infinite dimensional matrix $B_{b \rightarrow a}$ to $D_{a,b}$

PROPOSITION 3.1. *Suppose Assumption 2.1 holds. Let $(D_{a,b}; 1 \leq a, b \leq p)$ be defined as in (2.5). Then, for all $b \neq a$, we have*

$$(3.10) \quad D_{a,b} = -D_{a,a} B_{b \rightarrow a}.$$

Proofs of results appear in the Supplementary Material [4].

In the following theorem, we rewrite the conditions for conditional noncorrelation and conditional time-invariant edge in terms of node-regression coefficients.

THEOREM 3.4. *Suppose Assumption 2.1 holds. Let $B_{b \rightarrow a}$ be defined as in (3.9). Then:*

- (i) $B_{b \rightarrow a} = 0$ iff $D_{a,b} = 0$.
- (ii) If $D_{a,a}$ and $B_{b \rightarrow a}$ are Toeplitz, then $D_{a,b} = -D_{a,a}B_{b \rightarrow a}$ is Toeplitz.

Proofs of results appear in the Supplementary Material [4].

Below we show that the integral kernel associated with $B_{b \rightarrow a}$ has a clear sparsity structure.

COROLLARY 3.1. *Suppose Assumption 2.1 holds. Let $B_{b \rightarrow a}$ be defined as in (3.10). Let $K_{b \rightarrow a}(\omega, \lambda)$ denote the integral kernel associated with $B_{b \rightarrow a}$. Then:*

- (i) $B_{b \rightarrow a}$ is a bounded operator.
- (ii) Conditionally noncorrelated: $\{X_t^{(a)}, X_t^{(b)}\}_t$ are conditionally noncorrelated iff $K_{a \rightarrow b}(\omega, \lambda) \equiv 0$.
- (iii) Conditionally stationary pair: $\{X_t^{(a)}, X_t^{(b)}\}_t$ are conditionally jointly stationary iff the kernels $K_{a,a}(\omega, \lambda)$, $K_{b,b}(\omega, \lambda)$ and $K_{b \rightarrow a}(\omega, \lambda)$ are diagonal.

Proofs of results appear in the Supplementary Material [4].

We use the results above to deduce the conditional distribution of $X^{(a)}$ under the assumption that the time series $\{\underline{X}_t\}$ is jointly Gaussian. The conditional distribution of $X^{(a)}$, given $\mathcal{H} - (X^{(a)})$, is Gaussian, where

$$X^{(a)} | \mathcal{H} - (X^{(a)}) \sim N \left(\sum_{b=1, b \neq a}^p B_{b \rightarrow a} X^{(b)}, D_{aa}^{-1} \right)$$

with $E[X^{(a)} | \mathcal{H} - (X^{(a)})] = \sum_{b=1, b \neq a}^p B_{b \rightarrow a} X^{(b)}$ and $\text{Var}[X^{(a)} | \mathcal{H} - (X^{(a)})] = D_{aa}^{-1}$. Some interesting simplifications can be made if the nodes and corresponding edges are conditionally stationary and time invariant. If $X^{(a)}$ has a conditionally stationary node, then by Theorem 2.2(ii) the conditional variance will be stationary (Toeplitz). If, in addition, the conditionally stationary node a is connected to the set of nodes $\mathcal{S}_a$ and all the edge connections are conditionally time invariant, then by Theorem 3.4 the coefficients in the conditional expectation are shift invariant, where

$$E[X_t^{(a)} | \mathcal{H} - (X^{(a)})] = \sum_{b \in \mathcal{S}_a} \sum_{j \in \mathbb{Z}} \alpha_j^{(b \rightarrow a)} X_{t-j}^{(b)}.$$

Therefore, if the node a is conditionally stationary and all its connecting edges are conditionally time invariant, then the conditional distribution $X^{(a)} | (\mathcal{H} - (X^{(a)}))$ is stationary.

4. Learning the network from finite length time series. The network structure of $\{\underline{X}_t\}_t$ is succinctly described in terms of $\mathbf{K}(\omega, \lambda)$. However, for the purpose of estimation, there are three problems. The first is that $\mathbf{K}(\omega, \lambda)$ is a singular kernel making direct estimation impossible. The second is that, for conditional nonstationary time series, the structure of $[\mathbf{K}(\omega, \lambda)]_{a,b}$ is not well defined. Finally, in practice, we only observe a finite length sample $\{\underline{X}_t\}_{t=1}^n$. Thus, our object of interest changes from $\mathbf{K}(\omega, \lambda)$ to its finite dimensional counterpart (which we define below). For the purpose of network identification, we show that the finite dimensional version of $\mathbf{K}(\omega, \lambda)$ inherits the sparse properties of $\mathbf{K}(\omega, \lambda)$. Moreover, in a useful twist, whereas $\mathbf{K}(\omega, \lambda)$ is a singular kernel, its finite dimensional counterpart is a well-defined matrix, making estimation possible.

4.1. Finite dimensional approximation. To obtain the finite dimensional version of $\mathbf{K}(\omega, \lambda)$, we recall that the discrete Fourier transform (DFT) can be viewed as the analogous version of the Fourier operator F (defined in (3.1)) in finite dimensions. Let F_n denote the $(np \times np)$ -dimension DFT transformation matrix. It comprises of p^2 identical $(n \times n)$ -dimension DFT matrices, which we denote as $\mathcal{F}_n$. Define the concatenated np -dimension vector $\mathbf{X}'_n = ((\underline{X}^{(1)})', \dots, (\underline{X}^{(p)})')$, where $\underline{X}^{(a)} = (X_1^{(a)}, \dots, X_n^{(a)})'$ for $a \in \{1, \dots, p\}$. Then $F_n^* \mathbf{X}_n$ is a np -dimension vector, where $(F_n^* \mathbf{X}_n)' = ((\mathcal{F}_n^* \underline{X}^{(1)})', \dots, (\mathcal{F}_n^* \underline{X}^{(p)})')$ with

$$(4.1) \quad J_k^{(a)} = [\mathcal{F}_n^* \underline{X}^{(a)}]_k = \frac{1}{\sqrt{n}} \sum_{t=1}^n X_t^{(a)} \exp(it\omega_k) \quad k = 1, \dots, n \text{ and } \omega_k = \frac{2\pi k}{n}.$$

Let $\text{Var}[\mathbf{X}_n] = C_n$, then $\text{Var}[F_n^* \mathbf{X}_n] = F_n^* C_n F_n$. Our focus will be on the $(np \times np)$ -dimensional inverse matrix

$$\mathbf{K}_n = [\text{Var}[F_n^* \mathbf{X}_n]]^{-1} = [F_n^* C_n F_n]^{-1} = [F_n]^{-1} C_n^{-1} [F_n^*]^{-1} = F_n^* \tilde{D}_n F_n,$$

where $\tilde{D}_n = C_n^{-1}$ and the above follows from the identity $F_n^{-1} = F_n^*$. Let $\mathbf{K}_n = ([\mathbf{K}_n]_{a,b}; a, b \in \{1, \dots, p\})$, where $[\mathbf{K}_n]_{a,b}$ denotes the $(n \times n)$ -dimensional sub-matrix of $\mathbf{K}_n$ and $[\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})]_{a,b}$ denotes the (k_1, k_2) th entry in the submatrix matrix $[\mathbf{K}_n]_{a,b}$. For future reference we define the $(p \times p)$ -dimensional matrix $\mathbf{K}_n(\omega_{k_1}, \omega_{k_2}) = ([\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})]_{a,b}; 1 \leq a, b \leq p)$. We show below that $[\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})]_{a,b}$ can be viewed as the finite dimensional version of $K_{a,b}(\omega, \lambda)$.

The covariance matrix $C_n = \text{Var}[\mathbf{X}_n]$ is a submatrix of the infinite dimensional C . Unfortunately, its inverse $\tilde{D}_n = C_n^{-1}$ is not a submatrix of D . As our aim is to show that the properties of the inverse covariance map to those in finite dimensions, we will show that, under suitable conditions, $\tilde{D}_n$ can be approximated by a finite dimensional submatrix of D . To do this, we represent $\tilde{D}_n$ as $p \times p$ submatrices each of dimension $n \times n$

$$(4.2) \quad \tilde{D}_n = (\tilde{D}_{a,b;n}; a, b \in \{1, \dots, p\}).$$

Analogously, we define $p \times p$ submatrices of D each of dimension $n \times n$

$$(4.3) \quad D_n = (D_{a,b;n}; a, b \in \{1, \dots, p\}),$$

where $D_{a,b;n} = \{[D_{a,b}]_{t,\tau}; t, \tau \in \{1, \dots, n\}\}$. Below we show that, under suitable conditions, $\tilde{D}_n = C_n^{-1}$ can be approximated well by D_n . This result requires the following conditions on the rate of decay of the inverse covariances $\mathbf{D}_{t,\tau}$, which is stronger than the conditions in Assumption 2.1.

ASSUMPTION 4.1. The inverse covariances $\mathbf{D}_{t,\tau}$, defined in (2.4), satisfy the condition $\sup_t \|\mathbf{D}_{t,t+j}\|_\infty \leq \ell(j)^{-1}$, where $\{\ell(|j|)^{-1}\}_j$ is some monotonically decreasing sequence with $\sup_t \sum_j |j|^K \ell(j)^{-1} < \infty$ (for some $K \geq 3/2$).

The conditions in Assumption 4.1 are analogous to those used in the analysis of stationary time series, where certain conditions on the rate of decay of the autocovariances coefficients are often used. [40] obtain an equivalence between the rate of decay on $\mathbf{D}_{t,\tau}$ and $\mathbf{C}_{t,\tau}$. In particular, [40] Theorem 2.1 show that if Assumption 2.1 holds and for all $|r| \neq 0$ we have $\sup_t \|\mathbf{C}_{t,t+r}\| < K|r|^{-K}$ for some $K > 7/2$ (where $\|\cdot\|$ denotes the spectral norm), then $\sup_t \|\mathbf{D}_{t,t+r}\| < K((1 + \log|r|)/|r|)^{K-1}$. Thus, Assumption 4.1 holds.

In the lemma below, we obtain a bound between the rows of $\tilde{D}_n$ and D_n .

THEOREM 4.1. *Suppose Assumptions 2.1 and 4.1 hold. Let $\tilde{D}_n$ and D_n be defined as in (4.2) and (4.3). Then for all $1 \leq t \leq n$, we have*

$$\sup_{1 \leq a \leq p} \|[\tilde{D}_n]_{(a-1)n+t, \cdot} - [D_n]_{(a-1)n+t, \cdot}\|_1 = O\left(\frac{(np)^{1/2}}{\min(|n+1-t|, |t|)^K}\right),$$

where $A_{(a-1)n+t, \cdot}$ denotes the $((a-1)n+t)$ th row of the matrix A or, equivalently, the t th row along the a th block of A .

Proofs of results appear in the Supplementary Material [4].

The theorem above shows that the further t lies from the two end boundaries of the sequence $\{1, 2, \dots, n\}$ the better the approximation between $[\tilde{D}_n]_{(a-1)n+t, \cdot}$ and $[D_n]_{(a-1)n+t, \cdot}$. For example, when $t = n/2$ (recall that p is fixed), $\|[\tilde{D}_n]_{(a-1)n+t, \cdot} - [D_n]_{(a-1)n+t, \cdot}\|_1 = O(1/n^{K-1/2})$. Using Theorem 4.1, we replace $F_n^* \tilde{D}_n F_n$ with $F_n^* D_n F_n$ to obtain the following approximation.

PROPOSITION 4.1. *Suppose Assumptions 2.1 and 4.1 hold. Let $\Gamma_t^{(a,b)}(\omega)$ be defined as in (3.5). Then*

$$\begin{aligned} [\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})]_{a,b} &= \frac{1}{n} \sum_{t=1}^n \Gamma_t^{(a,b)}(\omega_{k_2}) \exp(-it(\omega_{k_1} - \omega_{k_2})) + O\left(\frac{1}{n}\right) \\ (4.4) \quad &= \left[\frac{1}{n} \sum_{t=1}^n \Gamma_t^{(b,a)}(\omega_{k_1}) \exp(-it(\omega_{k_2} - \omega_{k_1})) \right]^* + O\left(\frac{1}{n}\right). \end{aligned}$$

Further, if $\{X_t^{(a)}\}_t$ and $\{X_t^{(b)}\}_t$ are conditionally stationary, then

$$(4.5) \quad [\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})]_{a,b} = \begin{cases} \Gamma^{(a,b)}(\omega_k) + O(n^{-1}) & k_1 = k_2 (= k), \\ O(n^{-1}) & k_1 \neq k_2, \end{cases}$$

where $\Gamma^{(a,b)}(\omega) = \sum_{r=-\infty}^{\infty} [D_{a,b}]_{(0,r)} \exp(ir\omega)$.

Proofs of results appear in the Supplementary Material [4].

4.2. Locally stationary time series. We showed in Proposition 4.1 that in the case the node or edge (a, b) is conditional stationary or conditionally time-variant $[\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})]_{a,b}$ has a well-defined structure; the diagonal dominates the off-diagonal terms (which are of order $O(n^{-1})$). However, in the case of conditional nonstationary node/time-varying edge the precise structure of $[\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})]_{a,b}$ is not apparent; this makes detection of conditional nonstationarity difficult. In this section we impose some structure on the form of the nonstationarity. We will work under the canopy of local stationarity. It formalizes the notion that the “nonstationarity” in a time series evolves “slowly” through time. It is arguably one of the most popular methods for describing nonstationary behaviour and describes a wide class of nonstationarity; various applications are discussed in [10, 19, 21, 26, 39, 51, 54, 62, 70], to name but a few. We show below that for locally stationary time series $[\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})]_{a,b}$ has a distinct structure that can be detected.

Locally stationary processes were formally proposed in [13]. In the locally stationary framework, the asymptotics hinge on the rescaling device n , which is linked to the sample size. It measures how close the nonstationary time series is to an auxillary (latent) process

$\{\underline{X}_t(u)\}_t$, which for a fixed u is stationary over t . More precisely, a time series $\{\underline{X}_{t,n}\}_t$ is said to be locally stationary if there exists a stationary time series $\{\underline{X}_t(u)\}_t$, where

$$(4.6) \quad \|\underline{X}_{t,n} - \underline{X}_t(u)\|_2 = O_p\left(\frac{1}{n} + \left|\frac{t}{n} - u\right|\right).$$

Thus, for every t , $\underline{X}_{t,n} = (X_{t,n}^{(1)}, \dots, X_{t,n}^{(p)})'$ can be closely approximated by an auxiliary variable $\underline{X}_t(u)$ (where $u = t/n$); see [16, 21, 22, 59]. However, as the difference between t/n and u grows, the similarity between $X_{t,n}$ and the auxiliary stationary process $X_t(u)$ decreases. This asymptotic device allows one to obtain well-defined limits for nonstationary time series, which otherwise would not be possible within classical real time asymptotics. Though the formulation in (4.6) is a useful start for analysing nonstationary time series, analogous to [20], we require additional local stationarity conditions on the moment structure. [15] and [20] state the conditions in terms of bounds between $\text{Cov}[\underline{X}_{t,n}, \underline{X}_{\tau,n}]$ and $\text{Cov}[\underline{X}_0(u), \underline{X}_{t-\tau}(u)]$. Below we state similar conditions in terms of the inverse covariances $\mathbf{D}_{t,\tau}$ and its stationary approximation counterpart.

ASSUMPTION 4.2. There exists a sequence $\{\ell(j)\}_j$ such that $\sum_{j \in \mathbb{Z}} j \ell(j)^{-1} < \infty$ and matrix function $\mathbf{D}_{t-\tau} : \mathbb{R} \rightarrow \mathbb{R}^{p \times p}$, where

$$(4.7) \quad \mathbf{D}_{t,\tau} = \mathbf{D}_{t-\tau}\left(\frac{t+\tau}{2n}\right) + O\left(\frac{1}{n\ell(t-\tau)}\right) \quad t, \tau \in \mathbb{Z}.$$

Further, the matrix function $\mathbf{D}_j(\cdot)$ is such that: (i) $\sup_u \sum_{j \in \mathbb{Z}} \|j \mathbf{D}_j(u)\|_1 < \infty$, (ii) $\sup_u \left| \frac{d[\mathbf{D}_j(u)]_{a,b}}{du} \right| \leq \ell(j)^{-1}$, (iii) for all $u, v \in \mathbb{R}$ $\|\mathbf{D}_j(u) - \mathbf{D}_j(v)\|_1 \leq |u - v| \ell(j)^{-1}$, and (iv) $\sup_u \left| \frac{d[\mathbf{D}_j(u)]_{a,b}}{du} \right| \leq \ell(j)^{-1}$.

Standard within the locally stationary paradigm $\mathbf{D}_{t,\tau}$ should be indexed by n (but to simplify notation we have dropped the n).

Theorem 3.3 in [40] shows that Assumption 4.2 is fulfilled by a large class of locally stationary time series under certain smoothness conditions on their covariance.

The above assumptions allow for two important types of behaviour: (i) conditionally stationary nodes and time-invariant edges, where $[\mathbf{D}_j(u)]_{a,b} = [\mathbf{D}_j]_{(a,b)}$, and (ii) conditional nonstationarity where the partial covariance between $X_t^{(a)}$ and $X_{t+j}^{(b)}$ (for fixed lag j) evolves “nearly” smoothly over t . The technical condition that the entrywise derivative of the matrix functions $\mathbf{D}_{t-\tau}(\cdot)$ exist can be relaxed to include matrix functions $\mathbf{D}_j(\cdot)$ of bounded variation (which would allow for change point models as a special case) similar to [20].

4.3. Properties of $\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})$ under local stationarity. Typically, the second order analysis of locally stationary time series is conducted through its time-varying spectral density matrix. This is the spectral density matrix corresponding to the locally stationary approximation $\{\underline{X}_t(u)\}_t$, which we denote as $\boldsymbol{\Sigma}(u; \omega)$. The time-varying spectral density matrix corresponding to $\{\underline{X}_{t,n}\}_t$ is $\{\boldsymbol{\Sigma}(t/n; \omega)\}_t$. In contrast, in this section our focus will be on the inverse $\boldsymbol{\Gamma}(u; \omega) = \boldsymbol{\Sigma}(u; \omega)^{-1}$, where by Lemma 3.2, $\boldsymbol{\Gamma}(u; \omega)$ is the Fourier transform of $\mathbf{D}_j(u)$ over the lags j , that is,

$$(4.8) \quad \boldsymbol{\Gamma}(u; \omega) = \sum_{j \in \mathbb{Z}} \mathbf{D}_j(u) \exp(ij\omega).$$

We note that $\boldsymbol{\Gamma}(u; \omega) = (\Gamma^{(a,b)}(u; \omega); 1 \leq a, b \leq p)$. We use Assumption 4.2 to relate $\Gamma^{(a,b)}(u; \omega)$ to $\Gamma_t^{(a,b)}(\omega)$ (defined in (3.5)). In particular, $\Gamma^{(a,b)}(t/n; \omega)$ is an approximation

of $\Gamma_t^{(a,b)}(\omega)$ and

$$(4.9) \quad |\Gamma_t^{(a,b)}(\omega) - \Gamma^{(a,b)}(u; \omega)| \leq C \left(\left| \frac{t}{n} - u \right| + \frac{1}{n} \right).$$

Thus, the time-varying spectral *precision* matrix corresponding to $\{\underline{X}_{t,n}\}_t$ is $\{\Gamma(t/n; \omega)\}_t$.

Our aim is to relate $\Gamma(u; \omega)$ to $\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})$. First, we notice that $\Gamma(u; \omega)$ is “local,” in the sense that it is a time local approximation to the precision spectral density at time point $t = \lfloor un \rfloor$ (where $\lfloor y \rfloor$ denotes the largest integer less than y). On the other hand, $\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})$ is “global,” in the sense that it is based on the entire observed time series. However, we show below that $\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})$ is connected to $\Gamma(u; \omega)$, as it measures how $\Gamma(u; \omega)$ evolves over time. These insights allow us to deduce the network structure from $\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})$.

In the following lemma, we show that the entries of the matrix $\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})$ can be approximated by the Fourier coefficients of $\Gamma^{(a,b)}(\cdot; \omega)$, where

$$(4.10) \quad K_r^{(a,b)}(\omega) = \int_0^1 \exp(-2\pi i r u) \Gamma^{(a,b)}(u; \omega) du.$$

The Fourier coefficients $K_r^{(a,b)}(\omega)$ fully determine the function $\Gamma^{(a,b)}(u; \omega)$. In particular: (i) if all the Fourier coefficients are zero, then $\Gamma^{(a,b)}(u; \omega) = 0$; (ii) if all the Fourier coefficients are zero, except $r = 0$, then $\Gamma^{(a,b)}(u; \omega)$ does not depend on u . Using this, it is clear the coefficients $K_r^{(a,b)}(\omega)$ hold information on the network. We summarize these properties in the following proposition.

PROPOSITION 4.2. *Suppose Assumptions 2.1, 4.1, and 4.2 hold. Let $K_r^{(a,b)}(\cdot)$ be defined as in (4.10). Then:*

- (i) $\{X_{t,n}^{(a)}\}_{t=1}^n$ and $\{X_{t,n}^{(b)}\}_{t=1}^n$ is a (asymptotically) conditionally noncorrelated edge iff $K_r^{(a,b)}(\omega) \equiv 0$ for all $r \in \mathbb{Z}$ and $\omega \in [0, 2\pi]$.
- (ii) $\{X_{t,n}^{(a)}\}_{t=1}^n$ is a (asymptotically) conditionally stationary node iff $K_r^{(a,a)}(\omega) \equiv 0$ for all $r \neq 0$ and $\omega \in [0, 2\pi]$.
- (iii) The edge (a, b) is (asymptotically) conditionally time invariant iff $K_r^{(a,b)}(\omega) \equiv 0$ for all $r \neq 0$ and $\omega \in [0, 2\pi]$.

Proofs of results appear in the Supplementary Material [4].

Note that the above result is asymptotic in rescaled time ($n \rightarrow \infty$). We make this precise in the following proposition, where we show that $[\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})]_{a,b}$ closely approximates the Fourier coefficients $K_{k_1-k_2}^{(a,b)}(\omega_{k_2})$.

PROPOSITION 4.3. *Suppose Assumptions 2.1, 4.1, and 4.2 hold. Let $K_r^{(a,b)}(\cdot)$ be defined as in (4.10). Then*

$$(4.11) \quad [\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})]_{a,b} = \frac{1}{n} \sum_{t=1}^n \exp\left(-\frac{i2\pi(k_1 - k_2)t}{n}\right) \Gamma^{(a,b)}\left(\frac{t}{n}; \omega_{k_2}\right) + O\left(\frac{1}{n}\right).$$

Further,

$$(4.12) \quad [\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})]_{a,b} = \begin{cases} K_{k_1-k_2}^{(a,b)}(\omega_{k_2}) + O\left(\frac{1}{n}\right) & \text{if } |k_1 - k_2| \leq n/2, \\ K_{k_1-k_2-n}^{(a,b)}(\omega_{k_2}) + O\left(\frac{1}{n}\right) & \text{if } n/2 < (k_1 - k_2) < n, \\ K_{k_1-k_2+n}^{(a,b)}(\omega_{k_2}) + O\left(\frac{1}{n}\right) & \text{if } -n < (k_1 - k_2) < -n/2, \end{cases}$$

where the $O(n^{-1})$ bound is uniform over all $1 \leq r \leq n$ (and n is in rescaled time).

Since $[\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})]_{a,b} = [\mathbf{K}_n(\omega_{k_2}, \omega_{k_1})]_{b,a}^*$, then (4.12) can be replaced with $K_{k_2-k_1}^{(b,a)}(\omega_{k_1})^*$, $K_{k_2-k_1+n}^{(b,a)}(\omega_{k_1})^*$ and $K_{k_2-k_1-n}^{(b,a)}(\omega_{k_1})^*$, respectively.

Proofs of results appear in the Supplementary Material [4].

In the proposition above, we have split $[\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})]_{a,b}$ into three separate cases due to the circular wrapping of the DFT, which is most pronounced when ω_{k_1} is close to the boundaries of the interval $[0, 2\pi]$.

In [27] and [37], we showed that the Fourier transform of the time-varying spectral density matrix $\mathbf{G}_r(\omega) = \int_0^1 e^{-2\pi i r u} \boldsymbol{\Sigma}(u; \omega) du$ decayed to zero as $|r| \rightarrow \infty$ and was smooth over ω . In the following lemma, we show that a similar result holds for the Fourier transform of the inverse spectral density matrix.

PROPOSITION 4.4 (Properties of $K_r^{(a,b)}(\omega)$). *Suppose Assumption 4.2 holds. Then, for all $1 \leq a, b \leq p$, we have*

$$(4.13) \quad \sup_{\omega} |K_r^{(a,b)}(\omega)| \rightarrow 0 \quad \text{as } r \rightarrow \infty$$

and $\sup_{\omega} |K_r^{(a,b)}(\omega)| \sim |r|^{-1}$. Furthermore, for all $\omega_1, \omega_2 \in [0, \pi]$ and $r \in \mathbb{Z}$,

$$(4.14) \quad |K_r^{(a,b)}(\omega_1) - K_r^{(a,b)}(\omega_2)| \leq \begin{cases} C|\omega_1 - \omega_2| & r = 0, \\ C|r|^{-1}|\omega_1 - \omega_2| & r \neq 0, \end{cases}$$

where C is a finite constant that does not depend on r or ω .

Proofs of results appear in the Supplementary Material [4].

The above results describe two important features in $\mathbf{K}_{a,b}$:

1. For a given subdiagonal r , $[\mathbf{K}]_{a,b}^{(r)}$ changes smoothly along the subdiagonal, where $\mathbf{K}_{a,b}^{(r)}$ denotes the r th subdiagonal $-(n-1) \leq r \leq (n-1)$. Analogous to locally smoothing the periodogram, to estimate the entries of $[\mathbf{K}]_{a,b}$ from the DFTs we use the smoothness property and frequencies in a local neighbourhood to obtain multiple “near replicates.”
2. For a given row k , $[\mathbf{K}(\omega_k, \omega_{k+r})]_{a,b}$ is large when $r \bmod(n)$ is close to zero and decays the further it is from zero.

These observations motivate the regression method that we describe below for learning the nonstationary network structure from data.

4.4. Nodewise regression of the DFTs. In this section we propose a method for estimating the entries of $F_n^* D_n F_n$. The problem of learning the network structure from finite sample time series is akin to the graphical model selection problem in GGM, addressed by [24] for the low-dimensional and [47] for the high-dimensional setting. In particular, the neighborhood selection approach of [47] regresses one component of a multivariate random vector on the other components with Lasso [63] and uses nonzero regression coefficients to select its neighborhood, that is, the nodes which are conditionally noncorrelated with the given component.

Assuming the multivariate time series is locally stationary and satisfies Assumption 4.2, we show that the nonstationary network learning problem can be formulated in terms of a regression of DFTs at a specific Fourier frequency on neighboring DFTs. Let $J_k^{(a)}$ denote the DFT of the time series $\{X_t^{(a)}\}_t$ at Fourier frequency ω_k , as defined in (4.1). We denote the

p -dimensional vector of DFTs at ω_k by $\underline{J}_k$ and use $\underline{J}_k^{-(a)}$ to denote the $(p-1)$ -dimensional vector consisting of all the coordinates of $\underline{J}_k$, except $J_k^{(a)}$.

We define the space $\mathcal{G}_n = \overline{\text{sp}}(J_k^{(b)}; 1 \leq k \leq n, 1 \leq b \leq p)$ (note that the coefficients in this space can be complex). Then

$$(4.15) \quad P_{\mathcal{G}_n - J_k^{(a)}}(J_k^{(a)}) = \sum_{b=1}^p \sum_{s=1}^n B_{(b,s) \rightarrow (a,k)} J_s^{(b)},$$

where we set $B_{(a,k) \rightarrow (a,k)} = 0$. Let

$$(4.16) \quad \Delta_k^{(a)} = \mathbb{V}\text{ar}(J_k^{(a)} - P_{\mathcal{G}_n - J_k^{(a)}}(J_k^{(a)})).$$

The above allows us to rewrite the entries of $[\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})]_{a,b}$ in terms of regression coefficients. In particular,

$$(4.17) \quad [\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})]_{a,b} = \begin{cases} \frac{1}{\Delta_{k_1}^{(a)}} & k_1 = k_2 \text{ and } a = b, \\ -\frac{1}{\Delta_{k_1}^{(a)}} B_{(b,k_2) \rightarrow (a,k_1)} & \text{otherwise.} \end{cases}$$

Comparing the above with Proposition 4.3 for $(a, k_1) \neq (b, k_2)$, we have

$$B_{(b,k_2) \rightarrow (a,k_1)} = B_{k_2-k_1, n}^{(b \rightarrow a)}(\omega_{k_1}) + O(n^{-1}) \quad \text{and} \quad \Delta_k^{(a)} = [K_0^{(a,a)}(\omega_k)]^{-1} + O(n^{-1}),$$

where

$$(4.18) \quad B_{r,n}^{(b \rightarrow a)}(\omega_k) = \begin{cases} -K_0^{(a,a)}(\omega_k)^{-1} K_r^{(b,a)}(\omega_k)^* & \text{if } |r| \leq n/2, r \neq 0, \\ -K_0^{(a,a)}(\omega_k)^{-1} K_{r-n}^{(b,a)}(\omega_k)^* & \text{if } n/2 < r < n, \\ -K_0^{(a,a)}(\omega_k)^{-1} K_{r+n}^{(b,a)}(\omega_k)^* & \text{if } -n < r < -n/2. \end{cases}$$

Thus, by using Proposition 4.4 we have

$$(4.19) \quad |B_{(b,k_1+r) \rightarrow (a,k_1)} - B_{(b,k_2+r) \rightarrow (a,k_2)}| \leq A|\omega_{k_1} - \omega_{k_2}| + O(n^{-1}),$$

where A is a finite constant. The benefit of these results is in the estimation of the coefficients $B_{(b,k+r) \rightarrow (a,k)}$. We recall (4.15) can be expressed as

$$P_{\mathcal{G}_n - J_k^{(a)}}(J_k^{(a)}) = \sum_{b=1}^p \sum_{r=-k+1}^{n-k} B_{(b,k+r) \rightarrow (a,k)} J_{k+r}^{(b)},$$

where the above is due to the periodic nature of $J_k^{(a)}$, which allows us to extend the definition to frequencies outside $[0, 2\pi]$. By using the near Lipschitz condition in (4.19) if k_1 and k_2 are “close,” then the coefficients of the projections $P_{\mathcal{G}_n - J_{k_1}^{(a)}}(J_{k_1}^{(a)})$ and $P_{\mathcal{G}_n - J_{k_2}^{(a)}}(J_{k_2}^{(a)})$ will be similar. This observation will allow us to estimate $B_{(b,k+r) \rightarrow (a,k)}$, using the DFTs whose frequencies all lie in the M -neighbourhood of k (analogous to smoothing the periodogram of stationary time series). We note that, with these quasi replicates, the estimation would involve $(2M+1)$ (where $M \ll n$) response variables and $pn-1$ regressors. Even with the aid of sparse estimation methods, this is a large number of regressors. However, Proposition 4.4 allows us to reduce the number of regressors in the regression. Since $|B_{(b,k+r) \rightarrow (a,k)}| \sim |r|^{-1}$, we can truncate the projection to a small number $(2\nu+1)$ of regressors about $\underline{J}_k$ to obtain the approximation

$$P_{\mathcal{G}_n - J_k^{(a)}}(J_k^{(a)}) \approx \sum_{b=1}^p \sum_{r=-\nu}^{\nu} B_{(b,k+r) \rightarrow (a,k)} J_{k+r}^{(b)}.$$

Thus, smoothness together with near sparsity of the coefficients make estimation of the entries in the high-dimensional precision matrix $F_n^* D_n F_n$ feasible.

For a given choice of M and ν and every value of a, k , we define the $(2M+1)$ -dimensional complex response vector $\mathcal{Y}_k^{(a)} = (J_{k-M}^{(a)}, J_{k-M+1}^{(a)}, \dots, J_k^{(a)}, J_{k+1}^{(a)}, \dots, J_{k+M}^{(a)})'$, and the $(2M+1) \times ((2\nu+1)p-1)$ dimensional complex design matrix

$$\mathcal{X}_k^{(a)} = \begin{bmatrix} \underline{J}_{k-M-\nu}' & \cdots & \underline{J}_{k-M-1}' & (\underline{J}_{k-M}^{(a)})' & \underline{J}_{k-M+1}' & \cdots & \underline{J}_{k-M+\nu}' \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \underline{J}_{k-\nu}' & \cdots & \underline{J}_{k-1}' & (\underline{J}_k^{(a)})' & \underline{J}_{k+1}' & \cdots & \underline{J}_{k+\nu}' \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \underline{J}_{k+M-\nu}' & \cdots & \underline{J}_{k+M-1}' & (\underline{J}_{k+M}^{(a)})' & \underline{J}_{k+M+1}' & \cdots & \underline{J}_{k+M+\nu}' \end{bmatrix}.$$

Then the estimator

$$\hat{B}_{(\cdot, \cdot) \rightarrow (a, k)} = (\hat{B}_{(1, k-\nu) \rightarrow (a, k)}, \dots, \hat{B}_{(p, k-\nu) \rightarrow (a, k)}, \dots, \hat{B}_{(1, k) \rightarrow (a, k)}, \dots, \hat{B}_{(k-1, k) \rightarrow (a, k)}, \\ \hat{B}_{(k+1, k) \rightarrow (a, k)}, \hat{B}_{(p, k) \rightarrow (a, k)}, \dots, \hat{B}_{(1, k+\nu) \rightarrow (a, k)}, \dots, \hat{B}_{(p, k+\nu) \rightarrow (a, k)})'$$

of $\{B_{(b, k+r) \rightarrow (a, k)}; 1 \leq b \leq p, -\nu \leq r \leq \nu\}$ is obtained by solving the complex lasso optimization problem

$$\min_{\beta \in \mathbb{C}^{(2\nu+1)p-1}} \left[\frac{1}{2M+1} \|\mathcal{Y}_k^{(a)} - \mathcal{X}_k^{(a)} \beta\|_2^2 + \lambda \|\beta\|_1 \right],$$

where $\|\beta\|_1 := \sum_j |\beta_j|$, the sum of moduli of all the (complex) coordinates, and λ is a (real positive) tuning parameter controlling the degree of regularization. It is well known [45] that the above optimization problem can be equivalently expressed as a group lasso optimization over real variables and can be solved using existing software. We use this property to compute the estimators in our numerical experiments.

From Proposition 4.2 we observe that the problem of graphical model selection reduces to learning the locations of large entries of $\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})$ at different Fourier frequencies $\omega_{k_1}, \omega_{k_2}$. Furthermore, from equations (4.17) and (4.18) it is possible to learn the sparsity structure of $\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})$ from the regression coefficients $B_{(b, k_1) \rightarrow (a, k_2)}$ (up to order $O(n^{-1})$). In particular, there is an edge $(a, b) \subseteq E$; that is, the components a and b are conditionally correlated, if $B_{(b, k_1) \rightarrow (a, k_2)}$ is nonzero for some k_1, k_2 (within the locally stationary framework). Similarly, an edge between a and b is conditionally time varying if $B_{(b, k_1) \rightarrow (a, k_2)}$ is nonzero for some $k_1 \neq k_2$. In the above we have ignored the $O(n^{-1})$ terms.

In view of these connections, we define two quantities involving the estimated regression coefficients whose sparsity patterns encode information on the graph structure. In particular, we aggregate the estimated regression coefficients across different Fourier frequencies into two $p \times p$ weight matrices

$$(4.20) \quad \hat{W}_{\text{self}} = \left(\left(\sum_k |\hat{B}_{(b, k) \rightarrow (a, k)}|^2 \right) \right)_{1 \leq a, b \leq p},$$

$$(4.21) \quad \hat{W}_{\text{other}} = \left(\left(\sum_{k_1 \neq k_2} |\hat{B}_{(b, k_1) \rightarrow (a, k_2)}|^2 \right) \right)_{1 \leq a, b \leq p}$$

for graphical model selection in NonStGM. Two components a and b are deemed conditionally uncorrelated if both the (a, b) th and the (b, a) th off-diagonal elements of $\hat{W}_{\text{self}} + \hat{W}_{\text{other}}$ are small. In contrast, a node a is deemed conditionally stationary if the (a, a) th element of $\hat{W}_{\text{other}}$ is small. Similarly, an edge between a and b is deemed conditionally time invariant if both the (a, b) th and the (b, a) th elements of $\hat{W}_{\text{other}}$ is small. Note that our nodewise

regression approach does not ensure that the estimated weight matrices $\hat{W}$ are symmetric. However, following [47], one can formulate suitable “and” (or “or”) rule to construct an undirected graph, where an edge (a, b) is present if the (a, b) th and (b, a) th entries are both large (or at least one of them is large).

5. Time-varying vector autoregressive models. In this section we link the structure of the coefficients of the time-varying vector autoregressive (tvVAR) process with the notion of conditional noncorrelation and conditional stationarity. This gives a rigorous understanding of certain features in a tvVAR model.

The time-varying VAR (tvVAR) model is often used to model nonstationarity (see [15, 20, 58, 60, 69]). A time series is said to have a time-varying VAR(∞) representation if it can be expressed as

$$(5.1) \quad \underline{X}_t = \sum_{j=1}^{\infty} \mathbf{A}_j(t) \underline{X}_{t-j} + \underline{\varepsilon}_t \quad t \in \mathbb{Z},$$

where $\{\underline{\varepsilon}_t\}_t$ are i.i.d random vectors with $\mathbb{V}\text{ar}[\underline{\varepsilon}_t] = \Sigma$ and $\mathbb{E}[\underline{\varepsilon}_t] = 0$. For simplicity, we have centered the time series as the focus is on the second order structure of the time series. We assume that (5.1) has a well-defined time-varying moving average representation as its solution (we show below that this allows the inverse covariance to be expressed in terms of $\{\mathbf{A}_j(t)\}$). We show below that the inverse covariance matrix operator, corresponding to (5.1), has a simple form that can easily be deduced from the VAR parameters.

5.1. The tvVAR model and the nonstationary network. In this section we obtain an expression for D in terms of the tvVAR parameters.

Let $\mathbf{C}_{t,\tau} = \text{Cov}[\underline{X}_t, \underline{X}_\tau]$ and C denote the corresponding covariance operator, as defined in (2.1). Let $\mathbf{H}$ denote the Cholesky decomposition of Σ^{-1} such that $\Sigma^{-1} = \mathbf{H}'\mathbf{H}$ (where $\mathbf{H}'$ denotes the transpose of $\mathbf{H}$). To obtain D , we use the Gram–Schmidt orthogonalisation. We define the following matrices:

$$\tilde{\mathbf{A}}_\ell(t) = \begin{cases} I_p & \ell = 0, \\ -\mathbf{A}_\ell(t) & \ell > 0, \\ 0 & \ell < 0. \end{cases}$$

Using $\{\tilde{\mathbf{A}}_\ell(t)\}$, we define the infinite dimensional, block, lower triangular matrix L , where the (t, τ) th block of L is defined as $L_{t,\tau} = \mathbf{H}\tilde{\mathbf{A}}_{t-\tau}(t)$ for all $t, \tau \in \mathbb{Z}$. Define $X = (\dots, \underline{X}_{-1}, \underline{X}_0, \underline{X}_1, \dots)$, then LX is defined as

$$(LX)_t = \mathbf{H} \sum_{\ell=0}^{\infty} \tilde{\mathbf{A}}_\ell(t) \underline{X}_{t-\ell} = \mathbf{H} \left(\underline{X}_t - \sum_{\ell=1}^{\infty} \mathbf{A}_\ell(t) \underline{X}_{t-\ell} \right) = \mathbf{H}\underline{\varepsilon}_t \quad t \in \mathbb{Z}.$$

By definition of (5.1), it can be seen that $\{(LX)_t\}_t$ are uncorrelated random vectors with $\mathbb{V}\text{ar}[(LX)_t] = I_p$. From this it is clear that $L'L$ is the inverse of a permuted version of C . We use this to deduce the inverse $D = C^{-1}$. We define $\mathbf{D}_{t,\tau}$ as

$$(5.2) \quad \mathbf{D}_{t,\tau} = \sum_{\ell=-\infty}^{\infty} \tilde{\mathbf{A}}_\ell(t+\ell)' \Sigma^{-1} \tilde{\mathbf{A}}_{(\tau-t)+\ell}(t+\ell).$$

The inverse of C is $D = (D_{a,b}; 1 \leq a, b \leq p)$, where $D_{a,b}$ is defined by substituting (5.2) into (2.5).

We now focus on the case $\Sigma = I_p$ and derive conditions for conditional noncorrelation and stationarity. In this case the suboperators $D_{a,b}$ have the entries

(5.3)

$$[D_{a,b}]_{t,t+r} = \begin{cases} \sum_{\ell=1}^{\infty} \langle [\mathbf{A}_\ell(t+\ell)]_{\cdot,a} [\mathbf{A}_{\ell+r}(t+\ell)]_{\cdot,b} \rangle - \langle [I_p]_{\cdot,a} [\mathbf{A}_r(t+\ell)]_{\cdot,b} \rangle & r \geq 0, \\ \sum_{\ell=1}^{\infty} \langle [\mathbf{A}_\ell(t+\ell)]_{\cdot,b} [\mathbf{A}_{\ell-r}(t+\ell)]_{\cdot,a} \rangle - \langle [I_p]_{\cdot,b} [\mathbf{A}_{-r}(t+\ell)]_{\cdot,a} \rangle & r < 0, \end{cases}$$

where $\mathbf{A}_{\cdot,a}$ denotes the a^{th} column of the matrix $\mathbf{A}$ and $\langle \cdot, \cdot \rangle$ the standard dot product on $\mathbb{R}^p$. Using the above expression for $D_{a,b}$, the parameters of the tvVAR model can be connected to conditional noncorrelation and conditional stationarity:

- (i) *Conditional noncorrelation*: If for all $\ell \in \mathbb{R}$ the nonzero entries in the columns $[\tilde{\mathbf{A}}_\ell(t)]_{\cdot,a}$ and $[\tilde{\mathbf{A}}_\ell(t)]_{\cdot,b}$ do not coincide, then $\{X_t^{(a)}\}$ and $\{X_t^{(b)}\}$ are conditionally noncorrelated.
- (ii) *Conditionally stationary node*: If for all $\ell \in \mathbb{Z}$ the columns $[\tilde{\mathbf{A}}_\ell(t)]_{\cdot,a}$ do not depend on t , then the node a is conditionally stationary and the submatrix $D_{a,a}$ simplifies to

$$[D_{a,a}]_{t,t+r} = \sum_{\ell=1}^{\infty} \langle [\mathbf{A}_\ell(0)]_{\cdot,a} [\mathbf{A}_{\ell+|r|}(0)]_{\cdot,a} \rangle - \langle [I_p]_{\cdot,a} [\mathbf{A}_{|r|}(0)]_{\cdot,a} \rangle \quad \text{for all } r, t \in \mathbb{Z}.$$

- (iii) *Conditionally time-invariant edge*: If for all ℓ and r the dot products $\langle [\mathbf{A}_\ell(t)]_{\cdot,b}, [\mathbf{A}_{\ell-r}(t)]_{\cdot,a} \rangle$ do not depend on t and $[\mathbf{A}_r(t)]_{a,b}$ and $[\mathbf{A}_r(t)]_{b,a}$ does not depend on t , then $D_{a,b}$ is Toeplitz, where

$$[D_{a,b}]_{t,t+r} = \begin{cases} \sum_{\ell=1}^{\infty} \langle [\mathbf{A}_\ell(0)]_{\cdot,a} [\mathbf{A}_{\ell+r}(0)]_{\cdot,b} \rangle - \langle [I_p]_{\cdot,a} [\mathbf{A}_r(0)]_{\cdot,b} \rangle & r \geq 0, \\ \sum_{\ell=1}^{\infty} \langle [\mathbf{A}_\ell(0)]_{\cdot,b} [\mathbf{A}_{\ell-r}(0)]_{\cdot,a} \rangle - \langle [I_p]_{\cdot,b} [\mathbf{A}_{-r}(0)]_{\cdot,a} \rangle & r < 0, \end{cases}$$

for all $t \in \mathbb{Z}$.

There can arise situations where some $[\tilde{\mathbf{A}}_\ell(t)]_{\cdot,a}$ and $[\tilde{\mathbf{A}}_\ell(t)]_{\cdot,b}$ depend on t , but the corresponding node or edge is conditionally stationary or time invariant. This happens when there is a cancellation in the entries of $\mathbf{A}_\ell(t)$. However, these cases are quite exceptional.

In Appendix D of the Supplementary Material [4], we state conditions on the tvVAR process such that Assumptions 2.1, 4.1, and 4.2 are satisfied.

REMARK 5.1 (The time-varying AR approximation of locally stationary time series). In [40] Theorem 3.3, it is shown that if a multivariate nonstationary time series satisfies certain second order locally stationary conditions, then the time series has a tvAR(∞) representation with nearly smooth VAR parameters, that is,

$$X_t - \sum_{j=1}^{\infty} \Phi_j(t/n) X_{t-j} \approx H(t/n) \varepsilon_t,$$

where $H(\cdot)$ is a lower triangular matrix, $H(\cdot)$ and $\Phi_j(\cdot)$ are Lipschitz continuous, and $\{\varepsilon_t\}_t$ are uncorrelated random variables with $\text{Var}[\varepsilon_t] = I_p$. Using $\{\Phi_j(\cdot)\}_j$ and $H(\cdot)$, it would be possible to determine the approximate network of a nonstationary time series based on the conditions (i), (ii), (iii) stated above.

6. Numerical experiments. We demonstrate the applicability of nodewise regression in selecting NonStGM on two systems of multivariate time series, a small ($p = 4$) dimensional tvVAR(1) process described in Example 2.1, and a large ($p = 10$) dimensional tvVAR(1) process.

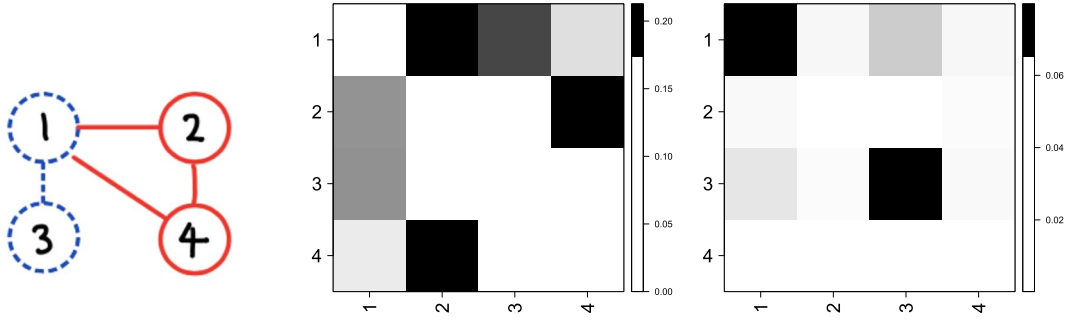


FIG. 3. NonStGM selection with nodewise regression for a $p = 4$ dimensional system. [Left]: True graph structure. [Middle]: Heat map of $\hat{W}_{\text{self}}$ showing conditional noncorrelation between components (1, 2), (1, 3), (1, 4), and (2, 4). [Right]: Heat map of $\hat{W}_{\text{other}}$ showing conditional nonstationarity of nodes 1 and 3 and the conditionally time-varying edge (1, 3). Results are aggregated over 20 replicates.

6.1. Small system. We simulate the $p = 4$ dimensional tvVAR(1) system, described in Example 2.1, where all the time-invariant parameters set to 0.4 and with $n = 5000$ observations. The two time-varying parameters $\alpha(t)$ and $\gamma(t)$ change from -0.8 to 0.8 , as t varies from 1 to n according to the function $f(t) = -0.8 + 1.6 \times e^{-5+10(t-1)/(n-1)}/(1 + e^{-5+10(t-1)/(n-1)})$. Using the results from Section 5.1, nodes 1, 3 are conditionally nonstationary, and the edge (1, 3) is conditionally time varying. On the other hand, the nodes 2, 4 are conditionally stationary, and the edges (2, 4), (1, 2), and (1, 4) are conditionally time invariant. As Figure 1a shows, these nuanced relationships are not prominent from the four time series trajectories.

We perform nodewise regression of DFTs with $M = \lceil \sqrt{n} \rceil$ and $\nu = 1$. The tuning parameters in the individual group lasso regressions were selected using cross-validation. The estimated regression coefficients $\hat{B}$ were used to construct the weight matrices $\hat{W}_{\text{self}}$ and $\hat{W}_{\text{other}}$. The heat maps of these weight matrices, aggregated over 20 replicates, are displayed in Figure 3.

The true graph structure (left) has two conditionally nonstationary nodes 1, 3 and two stationary nodes 2, 4. A heat map of $\hat{W}_{\text{self}}$ (middle) clearly shows the edges (1, 3), (1, 2), (2, 4), and (1, 4) capturing conditional noncorrelation in the true graph structure. The heat map of $\hat{W}_{\text{other}}$ (right) shows the conditionally nonstationary nodes 1 and 3 on the diagonal. The conditionally time-varying edge (1, 3) is also clearly visible on this heat map.

6.2. Large system. We now consider a larger system of $p = 10$. The data generating process is tvVAR(1) $X_t = A(t)X_{t-1} + \varepsilon_t$. Here $\varepsilon_t \stackrel{i.i.d.}{\sim} N(0, I_{10})$, nonzero time-invariant entries of the transition matrix $A(t)$ are constant functions as follows: $A_{j,j}(t) = 0.5$ for all $j \neq 5$, $A_{9,1}(t) = A_{10,1}(t) = A_{3,5}(t) = A_{4,5}(t) = A_{6,5}(t) = A_{7,5}(t) = 0.3$. The only time-varying entry is $A_{5,5}(t) = \alpha(t)$, where $\alpha(t)$ decays exponentially from 0.7 to -0.7 as t varies from 1 to n , according to the function $f(t) = 0.7 - 1.4 \times e^{-5+10(t-1)/(n-1)}/(1 + e^{-5+10(t-1)/(n-1)})$. As we can see from the structure of $A(t)$ (and the true graph structure in the left panel of Figure 4), this network has two connected components and two isolated nodes (2 and 8). These two nodes are independent of the other nodes and are treated as the “control.” The component consisting of (1, 9, 10) is stationary (due to time invariant AR parameters). On the other hand, the component (3, 4, 5, 6, 7) is nonstationary. However, the source of nonstationarity is node 5, which permeates through to nodes 3, 4, 6 and 7. Thus, the four nodes 3, 4, 6, and 7 are conditionally stationary (due to time-invariant parameters).

We simulate $n = 15,000$ observations from this system and perform nodewise regression of DFTs with $M = \lceil \sqrt{n} \rceil$ and $\nu = 1$. The tuning parameters in the individual group lasso regressions were selected using cross-validation. The estimated regression coefficients $\hat{B}$ were

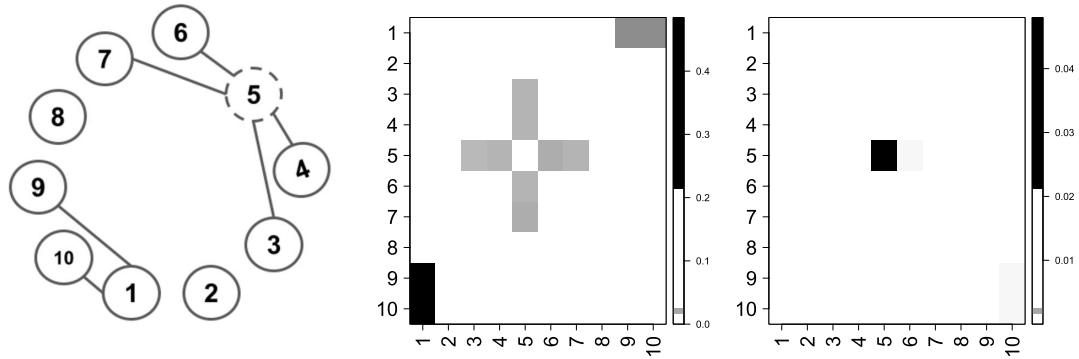


FIG. 4. *NonStGM selection with nodewise regression for a $p = 10$ dimensional system. [Left]: True graph structure. [Middle]: Heat map of $\hat{W}_{\text{self}}$ showing conditional noncorrelation captured by the edges. [Right]: Heat map of $\hat{W}_{\text{other}}$ showing conditional nonstationarity of node 5. Results are aggregated over 20 replicates.*

used to construct the weight matrices $\hat{W}_{\text{self}}$ and $\hat{W}_{\text{other}}$. The heat maps of these weight matrices, aggregated over 20 replicates, are displayed in Figure 4.

We observe that the edges for both components (1, 9, 10) and (3, 4, 5, 6, 7) are visible in the heat map of $\hat{W}_{\text{self}}$ (middle). As expected, the isolated nodes do not show up. The heat map of $\hat{W}_{\text{other}}$ (right) correctly identifies node 5 as conditionally nonstationary.

Conclusion. We introduced a general graphical modeling framework for describing conditional relationships among the components of a multivariate nonstationary time series using an undirected network. In this network absence of an edge corresponds to conditional noncorrelation relationships, as is common in GGM and StGM. An additional node or edge attribute (dashed or solid) further describes a newly introduced notion of *conditional nonstationarity*, which can be used to provide a parsimonious description of nonstationarity inherent in the overall system. We showed that this framework is a natural generalization of the existing GGM and StGM network. Under the locally stationary framework, we proposed methods to learn the nonstationary graph structure from finite-length time series in the Fourier domain. Numerical experiments on simulated data demonstrate the feasibility of our proposed method.

For stationary time series, there is well-established asymptotic theory for spectral density matrix estimators (see, e.g., [8, 42, 66, 67]). To estimate the inverse of moderate to high-dimensional spectral density matrices, penalized estimation methods for detecting nonzero off-diagonal entries [31] have shown promise. These methods are based on learning the conditional correlation structure of the DFTs at different nodes at the same frequency. Using the results in Section 4.4, we conjecture that the nonstationary network can be estimated by learning the nonzero coefficients of nodewise DFT regression across different frequencies. In future work we hope to develop a complete statistical theory for graphical model estimation and inference.

Acknowledgments. The authors thank Gregory Berkolaiko for several useful suggestions and Jonas Krampe for careful reading. The authors thank the Associate Editor and two anonymous referees for their thoughtful comments and suggestions which substantially improved the paper.

Funding. SB and SSR acknowledge the partial support of the National Science Foundation (grants DMS-1812054, DMS-1812128, DMS-2210726, DMS-2210675, and DMS-2239102). In addition, SB acknowledges partial support from the National Institute of Health (grants R01GM135926 and R21NS120227).

SUPPLEMENTARY MATERIAL

Proofs of technical results (DOI: [10.1214/22-AOS2205SUPP](https://doi.org/10.1214/22-AOS2205SUPP); .pdf). Proofs of all technical lemmas, propositions and theorems.

REFERENCES

- [1] ASTON, J., DEHAY, D., DUDEK, A., FREYERMUTH, J.-M., SZUCS, D. and COLLING, L. (2019). (dual-frequency)-dependent dynamic functional connectivity analysis of vidual working memory capacity. Preprint: Hal-021335535.
- [2] AUE, A. and VAN DELFT, A. (2020). Testing for stationarity of functional time series in the frequency domain. *Ann. Statist.* **48** 2505–2547. [MR4152111](https://doi.org/10.1214/19-AOS1895) <https://doi.org/10.1214/19-AOS1895>
- [3] BASU, S. and MICHAILIDIS, G. (2015). Regularized estimation in sparse high-dimensional time series models. *Ann. Statist.* **43** 1535–1567. [MR3357870](https://doi.org/10.1214/15-AOS1315) <https://doi.org/10.1214/15-AOS1315>
- [4] BASU, S. and SUBBA RAO, S. (2023). Supplement to “Graphical models for nonstationary time series.” <https://doi.org/10.1214/22-AOS2205SUPP>
- [5] BIRR, S., DETTE, H., HALLIN, M., KLEY, T. and VOLGUSHEV, S. (2018). On Wigner–Ville spectra and the uniqueness of time-varying copula-based spectral densities. *J. Time Series Anal.* **39** 242–250. [MR3796518](https://doi.org/10.1111/jtsa.12252) <https://doi.org/10.1111/jtsa.12252>
- [6] BÖHM, H. and VON SACHS, R. (2009). Shrinkage estimation in the frequency domain of multivariate time series. *J. Multivariate Anal.* **100** 913–935. [MR2498723](https://doi.org/10.1016/j.jmva.2008.09.009) <https://doi.org/10.1016/j.jmva.2008.09.009>
- [7] BRILLINGER, D. R. (1996). Remarks concerning graphical models for time series and point processes. *Braz. Rev. Econom.* **16** 1–23.
- [8] BRILLINGER, D. R. (2001). *Time Series: Data Analysis and Theory. Classics in Applied Mathematics* **36**. SIAM, Philadelphia, PA. Reprint of the 1981 edition. [MR1853554](https://doi.org/10.1137/1.9780898719246) <https://doi.org/10.1137/1.9780898719246>
- [9] BROCKWELL, P. J. and DAVIS, R. A. (2006). *Time Series: Theory and Methods. Springer Series in Statistics*. Springer, New York. Reprint of the second (1991) edition. [MR2839251](https://doi.org/10.1007/978-1-4939-9830-5)
- [10] CARDINALI, A. and NASON, G. P. (2010). Costationarity of locally stationary time series. *J. Time Ser. Econom.* **2** Art. 1, 33. [MR2915633](https://doi.org/10.2202/1941-1928.1074) <https://doi.org/10.2202/1941-1928.1074>
- [11] CHAU, J. and VON SACHS, R. (2021). Intrinsic wavelet regression for curves of Hermitian positive definite matrices. *J. Amer. Statist. Assoc.* **116** 819–832. [MR4270027](https://doi.org/10.1080/01621459.2019.1700129) <https://doi.org/10.1080/01621459.2019.1700129>
- [12] CONWAY, J. B. (1990). *A Course in Functional Analysis*, 2nd ed. *Graduate Texts in Mathematics* **96**. Springer, New York. [MR1070713](https://doi.org/10.1007/978-1-4939-9830-5)
- [13] DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25** 1–37. [MR1429916](https://doi.org/10.1214/aos/1034276620) <https://doi.org/10.1214/aos/1034276620>
- [14] DAHLHAUS, R. (2000). Graphical interaction models for multivariate time series. *Metrika* **51** 157–172. [MR1790930](https://doi.org/10.1007/s001840000055) <https://doi.org/10.1007/s001840000055>
- [15] DAHLHAUS, R. (2000). A likelihood approximation for locally stationary processes. *Ann. Statist.* **28** 1762–1794. [MR1835040](https://doi.org/10.1214/aos/1015957480) <https://doi.org/10.1214/aos/1015957480>
- [16] DAHLHAUS, R. (2012). Locally stationary processes. In *Handbook of Statistics* **30** 351–413. Elsevier, Amsterdam.
- [17] DAHLHAUS, R. and EICHLER, M. (2003). Causality and graphical models in time series analysis. In *Highly Structured Stochastic Systems. Oxford Statist. Sci. Ser.* **27** 115–144. Oxford Univ. Press, Oxford. [MR2082408](https://doi.org/10.1017/9780521876223.008)
- [18] DAHLHAUS, R., EICHLER, M. and SANDÜHLER, J. (1997). *J. Neurosci. Methods* **77** 93–107.
- [19] DAHLHAUS, R. and GIRAITIS, L. (1998). On the optimal segment length for parameter estimates for locally stationary time series. *J. Time Series Anal.* **19** 629–655. [MR1665941](https://doi.org/10.1111/1467-9892.00114) <https://doi.org/10.1111/1467-9892.00114>
- [20] DAHLHAUS, R. and POLONIK, W. (2006). Nonparametric quasi-maximum likelihood estimation for Gaussian locally stationary processes. *Ann. Statist.* **34** 2790–2824. [MR2329468](https://doi.org/10.1214/009053606000000867) <https://doi.org/10.1214/009053606000000867>
- [21] DAHLHAUS, R., RICHTER, S. and WU, W. B. (2019). Towards a general theory for nonlinear locally stationary processes. *Bernoulli* **25** 1013–1044. [MR3920364](https://doi.org/10.3150/17-bej1011) <https://doi.org/10.3150/17-bej1011>
- [22] DAHLHAUS, R. and SUBBA RAO, S. (2006). Statistical inference for time-varying ARCH processes. *Ann. Statist.* **34** 1075–1114. [MR2278352](https://doi.org/10.1214/009053606000000227) <https://doi.org/10.1214/009053606000000227>
- [23] DAVIS, R. A., ZANG, P. and ZHENG, T. (2016). Sparse vector autoregressive modeling. *J. Comput. Graph. Statist.* **25** 1077–1096. [MR3572029](https://doi.org/10.1080/10618600.2015.1092978) <https://doi.org/10.1080/10618600.2015.1092978>

- [24] DEMPSTER, A. P. (1972). Covariance selection. *Biometrics* **28** 157–175. [MR3931974](#)
- [25] DIEBOLD, F. X. and YILMAZ, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *J. Econometrics* **182** 119–134. [MR3212765](#) <https://doi.org/10.1016/j.jeconom.2014.04.012>
- [26] DING, X. and ZHOU, Z. (2020). Estimation and inference for precision matrices of nonstationary time series. *Ann. Statist.* **48** 2455–2477. [MR4134802](#) <https://doi.org/10.1214/19-AOS1894>
- [27] DWIVEDI, Y. and SUBBA RAO, S. (2011). A test for second-order stationarity of a time series based on the discrete Fourier transform. *J. Time Series Anal.* **32** 68–91. [MR2790673](#) <https://doi.org/10.1111/j.1467-9892.2010.00685.x>
- [28] EICHLER, M. (2007). Granger causality and path diagrams for multivariate time series. *J. Econometrics* **137** 334–353. [MR2354948](#) <https://doi.org/10.1016/j.jeconom.2005.06.032>
- [29] EICHLER, M. (2008). Testing nonparametric and semiparametric hypotheses in vector stationary processes. *J. Multivariate Anal.* **99** 968–1009. [MR2405101](#) <https://doi.org/10.1016/j.jmva.2007.06.003>
- [30] EPHARTY, A., TABRIKIAN, J. and MESSER, H. (2001). Underwater source detection using a spatial stationary test. *J. Acoust. Soc. Am.* **109** 1053–1063.
- [31] FIECAS, M., LENG, C., LIU, W. and YU, Y. (2019). Spectral analysis of high-dimensional time series. *Electron. J. Stat.* **13** 4079–4101. [MR4017528](#) <https://doi.org/10.1214/19-EJS1621>
- [32] FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441. <https://doi.org/10.1093/biostatistics/kxm045>
- [33] GLADYŠEV, E. G. (1963). Periodically and semi-periodically correlated random processes with continuous time. *Teor. Veroyatn. Primen.* **8** 184–189. [MR0152005](#)
- [34] GORROSTIETA, C., OMBAO, H. and VON SACHS, R. (2019). Time-dependent dual-frequency coherence in multivariate non-stationary time series. *J. Time Series Anal.* **40** 3–22. [MR3891477](#) <https://doi.org/10.1111/jtsa.12408>
- [35] HINDBERG, H. and OLHEDE, S. C. (2010). Estimation of ambiguity functions with limited spread. *IEEE Trans. Signal Process.* **58** 2383–2388. [MR2752078](#) <https://doi.org/10.1109/TSP.2009.2037663>
- [36] JENSEN, O. and COLGIN, L. L. (2007). Cross-frequency coupling between neuronal oscillations. *Trends Cogn. Sci.* **11** 267–269. <https://doi.org/10.1016/j.tics.2007.05.003>
- [37] JENTSCH, C. and SUBBA RAO, S. (2015). A test for second order stationarity of a multivariate time series. *J. Econometrics* **185** 124–161. [MR3300340](#) <https://doi.org/10.1016/j.jeconom.2014.09.010>
- [38] JUNG, A., HANNAK, G. and GOERTZ, N. (2015). Graphical lasso based model selection for time series. *IEEE Signal Process. Lett.* **22** 1781–1785.
- [39] KLEY, T., PREUSS, P. and FRYZLEWICZ, P. (2019). Predictive, finite-sample model choice for time series under stationarity and non-stationarity. *Electron. J. Stat.* **13** 3710–3774. [MR4013749](#) <https://doi.org/10.1214/19-ejs1606>
- [40] KRAMPE, J. and SUBBA RAO, S. (2022). Inverse covariance operators of nonstationary multivariate time series. Available at [arXiv:2202.00933](https://arxiv.org/abs/2202.00933).
- [41] LII, K.-S. and ROSENBLATT, M. (2002). Spectral analysis for harmonizable processes. *Ann. Statist.* **30** 258–297. [MR1892664](#) <https://doi.org/10.1214/aos/1015362193>
- [42] LOUBATON, P., ROSUEL, A. and VALLET, P. (2023). On the asymptotic distribution of the maximum sample spectral coherence of Gaussian time series in the high dimensional regime. *J. Multivariate Anal.* **194** Paper No. 105124, 22. [MR4516398](#) <https://doi.org/10.1016/j.jmva.2022.105124>
- [43] LUND, R., HURD, H., BLOOMFIELD, P. and SMITH, R. (1995). Climatological time series with periodic correlation. *J. Climate* **11** 2787–2809.
- [44] LURIE, D. J., KESSLER, D., BASSETT, D. S., BETZEL, R. F., BREAKSPEAR, M., KHEILHOLZ, S., KUCYI, A., LIÉGEOS, R., LINDQUIST, M. A. et al. (2020). Questions and controversies in the study of time-varying functional connectivity in resting fMRI. *Netw. Neurosci.* **4** 30–69.
- [45] MALEKI, A., ANITORI, L., YANG, Z. and BARANIUK, R. G. (2013). Asymptotic analysis of complex LASSO via complex approximate message passing (CAMP). *IEEE Trans. Inf. Theory* **59** 4290–4308. [MR3071330](#) <https://doi.org/10.1109/TIT.2013.2252232>
- [46] MARTIN, W. and FLANDRIN, P. (1985). Wigner–Ville spectral analysis of nonstationary processes. *IEEE Trans. Acoust. Speech Signal Process.* **33** 1461–1470.
- [47] MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. [MR2278363](#) <https://doi.org/10.1214/009053606000000281>
- [48] NASON, G. (2013). A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 879–904. [MR3124795](#) <https://doi.org/10.1111/rssb.12015>
- [49] OLHEDE, S. C. (2011). Ambiguity sparse processes. arXiv preprint. Available at [arXiv:1103.3932v2](https://arxiv.org/abs/1103.3932v2).

- [50] OLHEDE, S. C. and OMBAO, H. (2013). Modeling and estimation of covariance of replicated modulated cyclical time series. *IEEE Trans. Signal Process.* **61** 1944–1957. [MR3038192](#) <https://doi.org/10.1109/TSP.2012.2237168>
- [51] OMBAO, H. and PINTO, M. (2021). Spectral dependence. Available at [arXiv:2103.17240](#).
- [52] PAPARODITIS, E. (2009). Testing temporal constancy of the spectral structure of a time series. *Bernoulli* **15** 1190–1221. [MR2597589](#) <https://doi.org/10.3150/08-BEJ179>
- [53] PRETI, M. G., BOLTON, T. A. and VILLE, D. V. D. (2017). The dynamic functional connectome: State-of-the-art and perspectives. *NeuroImage* **160** 41–54. <https://doi.org/10.1016/j.neuroimage.2016.12.061>
- [54] PRIESTLEY, M. B. (1965). Evolutionary spectra and non-stationary processes. *J. Roy. Statist. Soc. Ser. B* **27** 204–237. [MR0199886](#)
- [55] PRIESTLEY, M. B. (1981). *Spectral Analysis and Time Series: Multivariate Series, Prediction and Control, Vol. 2. Probability and Mathematical Statistics*. Academic Press, London. [MR0628736](#)
- [56] PRIESTLEY, M. B. and SUBBA RAO, T. (1969). A test for non-stationarity of time-series. *J. Roy. Statist. Soc. Ser. B* **31** 140–149. [MR0269062](#)
- [57] QIU, H., HAN, F., LIU, H. and CAFFO, B. (2016). Joint estimation of multiple graphical models from high dimensional time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 487–504. [MR3454206](#) <https://doi.org/10.1111/rssb.12123>
- [58] SAFIKHANI, A. and SHOJAIE, A. (2022). Joint structural break detection and parameter estimation in high-dimensional nonstationary VAR models. *J. Amer. Statist. Assoc.* **117** 251–264. [MR4399083](#) <https://doi.org/10.1080/01621459.2020.1770097>
- [59] SUBBA RAO, S. (2006). On some nonstationary, nonlinear random processes and their stationary approximations. *Adv. in Appl. Probab.* **38** 1155–1172. [MR2285698](#) <https://doi.org/10.1239/aap/1165414596>
- [60] SUBBA RAO, T. (1970). The fitting of non-stationary time-series models with time-dependent parameters. *J. Roy. Statist. Soc. Ser. B* **32** 312–322. [MR0269065](#)
- [61] SUN, Y., LI, Y., KUCEYESKI, A. and BASU, S. (2018). Large spectral density matrix estimation by thresholding. arXiv preprint. Available at [arXiv:1812.00532](#).
- [62] SUNDARARAJAN, R. R. and POURAHMADI, M. (2018). Stationary subspace analysis of nonstationary processes. *J. Time Series Anal.* **39** 338–355. [MR3796523](#) <https://doi.org/10.1111/jtsa.12274>
- [63] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. [MR1379242](#)
- [64] TOEPLITZ, O. (1911). Zur Theorie der quadratischen und bilinearen Formen von unendlichvielen Veränderlichen. *Math. Ann.* **70** 351–376. [MR1511625](#) <https://doi.org/10.1007/BF01564502>
- [65] WANG, D., YU, Y., RINALDO, A. and WILLETT, R. (2019). Localizing changes in high-dimensional vector autoregressive processes. arXiv preprint. Available at [arXiv:1909.06359](#).
- [66] WOODROOFE, M. B. and VAN NESS, J. W. (1967). The maximum deviation of sample spectral densities. *Ann. Math. Stat.* **38** 1558–1569. [MR0216717](#) <https://doi.org/10.1214/aoms/1177698710>
- [67] WU, W. B. and ZAFFARONI, P. (2018). Asymptotic theory for spectral density estimates of general multivariate time series. *Econometric Theory* **34** 1–22. [MR3739234](#) <https://doi.org/10.1017/S0266466617000068>
- [68] ZHANG, D. and WU, W. B. (2017). Gaussian approximation for high dimensional time series. *Ann. Statist.* **45** 1895–1919. [MR3718156](#) <https://doi.org/10.1214/16-AOS1512>
- [69] ZHANG, D. and WU, W. B. (2021). Convergence of covariance and spectral density estimates for high-dimensional locally stationary processes. *Ann. Statist.* **49** 233–254. [MR4206676](#) <https://doi.org/10.1214/20-AOS1954>
- [70] ZHOU, Z. and WU, W. B. (2009). Local linear quantile estimation for nonstationary time series. *Ann. Statist.* **37** 2696–2729. [MR2541444](#) <https://doi.org/10.1214/08-AOS636>