Graphical models for nonstationary time series

Sumanta Basu[*] and Suhasini Subba Rao[†]

March 20, 2022

Abstract

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 architecture. This allows one to distinguish between direct and indirect nonstationary relationships among system components, and can be used to search for small subnetworks that serve as the "source" of nonstationarity in a large system. Together, the two concepts of conditional noncorrelation and nonstationarity/stationarity provide a parsimonious description of the dependence structure of the time series.

In GGM, the graphical model structure is encoded in the sparsity pattern of the inverse covariance matrix. Analogously, 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 order to learn the graph, we move to the Fourier domain. We show that 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. Within the local stationary framework we show that these sparsity patterns can be recovered from finite-length time series by node-wise regression of discrete Fourier Transforms (DFT) across different Fourier frequencies. We illustrate the features of our general framework under the special case of time-varying Vector Autoregressive models. We demonstrate the feasibility of learning NonStGM structure from data using simulation studies.

Keywords and phrases: Graphical models, locally stationary time series, nonstationarity, partial covariance and spectral analysis.

[*]Department of Statistics and Data Science, Cornell University `sumbose@cornell.edu`
[†]Department of Statistics, Texas A&M University `suhasini@stat.tamu.edu`

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 Lurie et al. (2020) and understanding linkages among firms for measuring systemic risk buildup in financial markets Diebold and Yılmaz (2014).

The vast majority of graphical models for time series focuses on the stationary setting (see Brillinger (1996); Dahlhaus et al. (1997); Dahlhaus (2000b); Dahlhaus and Eichler (2003a); Eichler (2007); Böhm and von Sachs (2009); Jung et al. (2015); Basu and Michailidis (2015); Davis et al. (2016); Zhang and Wu (2017); Qiu et al. (2016); Sun et al. (2018); Fiecas et al. (2019); Chau and von Sachs (2020), to name but a few). While the assumption of stationarity may be realistic in many situations, it is well known that *non-stationarity* 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 Preti et al. (2017). A naive application of graphical modeling methods designed for stationary processes can lead to spurious network edges if the actual time series is nonstationary.

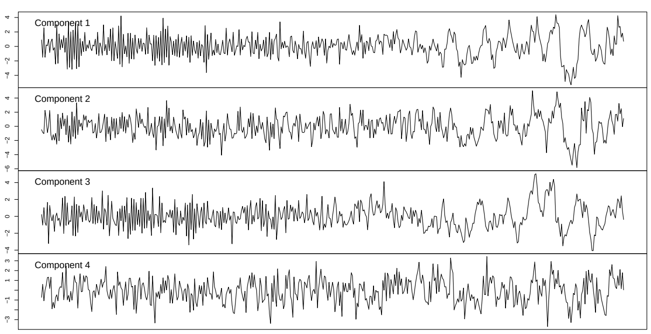
The limited body of work on graphical models for nonstationary time series has so far focused on a restricted class of nonstationary models, where the data generating process can be well approximated by a finite order change point vector autoregressive (VAR) model. Within this framework, Wang et al. (2019) and Safikhani and Shojaie (2020) have proposed methods for constructing a "dynamically changing" network at each of the estimated change points. However, these methods are designed for time series which are piece-wise 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. 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 semi-parametric models and propose a general framework for the graphical modeling of multivariate (say, p-dimensional) non-stationary 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 non-stationarity 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.

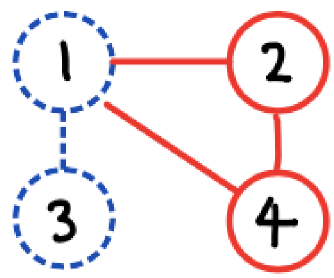
Analogous to GGM, in our framework, the presence/absence of edges in the network encodes conditional correlation/non-correlation relationships amongst the p components (nodes) of the time series. An additional attribute distinguishes between the types of nonstationarity. A graphical model is built using *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 4-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 towards 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



(a) Trajectories of a 4-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.

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

Section 2 for details), where components 1 and 3 are the *sources of nonstationarity*, i.e. 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 component 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.

It is interesting to contrast the networks constructed using the "dynamically changing" approach developed in Wang et al. (2019) and Safikhani and Shojaie (2020) 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 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 Appendix.

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 $V = \{1, 2, \ldots, p\}$, 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 $\underline{X} = (X^{(1)}, X^{(2)}, \ldots, X^{(p)})^\top$. The non-zero partial correlations $\rho^{(a,b)}$, defined as $\mathbb{C}\mathrm{orr}\left(X^{(a)}, X^{(b)} | X^{-\{a,b\}}\right)$ and also encoded in the sparsity structure of the *precision matrix* $\Theta = [\mathbb{V}\mathrm{ar}(\underline{X})]^{-1}$, are used to define the edge set E. The task of *graphical model selection*, i.e. learning the edge set E from finite sample, is accomplished by estimating Θ with a penalized likelihood estimator as in graphical Lasso (Friedman et al. (2008)), or by nodewise regression (Meinshausen and Bühlmann (2006)) where each component of the random vector is regressed on the other $(p-1)$ components.

Switching to the time series setting, consider $\{\underline{X}_t = (X_t^{(1)}, \ldots, X_t^{(a)}, \ldots, X_t^{(p)})^\top\}_{t\in\mathbb{Z}}$, a p-dimensional time series with autocovariance function $\mathbb{C}\text{ov}(\underline{X}_t, \underline{X}_\tau) = \mathbf{C}(t,\tau)$. 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 Zhang and Wu (2017); Qiu et al. (2016)) does not provide conditional relationships between the *entire time series*. Brillinger (1996) and Dahlhaus (2000b) 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 Eichler (2008)). More recently, for relatively "large" p, penalized methods such as GLASSO (Jung et al., 2015) and CLIME (Fiecas et al., 2019) 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 Priestley and Subba Rao (1969), Paparoditis (2009), Nason (2013) 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. Epharty et al. (2001), Dwivedi and Subba Rao (2011), Jentsch and Subba Rao (2015), Aue and van Delft (2020) use this property to test for nonzero correlation between DFTs of 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 $\mathbf{C}_{t,\tau} = \mathbb{C}\text{ov}[\underline{X}_t, \underline{X}_\tau]$, 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 $\mathbf{A} = (A_{a,b} : 1 \le a \le d_1, 1 \le b \le d_2)$ 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 $\{u = (\ldots, u_{-1}, u_0, u_1, \ldots)'; u_j \in \mathbb{C} \text{ and } \sum_j |u_j|^2 < \infty\}$ and the (column) sequence space $\{w = \text{vec}[u^{(1)}, \ldots, u^{(p)}]; u^{(s)} \in \ell_2 \text{ for all } 1 \le s \le p\}$ 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 $x = (u^{(1)}, \ldots, u^{(p)})'$, $y = (v^{(1)}, \ldots, v^{(p)}) \in \ell_{2,p}$, 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 A, which can be infinite dimensional and involve negative indices.

We consider the p-dimensional real-valued time series $\{\underline{X}_t\}_{t\in\mathbb{Z}}$, $\underline{X}_t = (X_t^{(1)}, \ldots, X_t^{(p)})'$, where the univariate random variables $X_t^{(a)}, a = 1, \ldots, p$, are defined on the probability space $(\Omega, \mathcal{F}, P)$. We assume for all t that $\mathrm{E}[\underline{X}_t] = 0$ 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 $\mathbb{V}\text{ar}[X] < \infty$, and for any $X, Y \in L^2(\Omega, \mathcal{F}, P)$ we define the inner product $\langle X, Y\rangle = \mathbb{C}\text{ov}[X, Y]$. For every $t, \tau \in \mathbb{Z}$, we define the $p \times p$ covariance $\mathbf{C}_{t,\tau} = \mathbb{C}\text{ov}[\underline{X}_t, \underline{X}_\tau]$ and assume $\sup_{t\in\mathbb{Z}} \|\mathbf{C}_{t,t}\|_\infty < \infty$. Under this assumption, for all $t \in \mathbb{Z}$ and $1 \le c \le p$ $X_t^{(c)} \in L^2(\Omega, \mathcal{F}, P)$. Let $\mathcal{H} = \overline{\text{sp}}(X_t^{(c)}; t \in \mathbb{Z}, 1 \le c \le p) \subset L^2(\Omega, \mathcal{F}, P)$ be the closure of the space spanned by $(X_t^{(c)}; t \in \mathbb{Z}, 1 \le c \le p)$. Since $L^2(\Omega, \mathcal{F}, P)$ defines a Hilbert space, $\mathcal{H}$ is also a Hilbert space. Therefore, by the projection theorem, for any closed subspace $\mathcal{M}$ of $\mathcal{H}$, there is a unique projection of $Y \in \mathcal{H}$ onto $\mathcal{M}$ which minimises $\mathrm{E}(Y - X)^2$ over all $X \in \mathcal{M}$ (see Theorem 2.3.1, Brockwell and Davis (2006)). We will use $P_{\mathcal{M}}(Y)$ to denote this projection. In this paper, we will primarily use the following subspaces

$$
\begin{aligned}
\mathcal{H} - X_t^{(a)} &= \overline{\text{sp}}[X_s^{(c)}; s \in \mathbb{Z}, 1 \le c \le p, (s, c) \ne (t, a)], \\
\mathcal{H} - (X_t^{(a)}, X_\tau^{(b)}) &= \overline{\text{sp}}[X_s^{(c)}; s \in \mathbb{Z}, 1 \le c \le p, (s, c) \notin \{(t, a), (\tau, b)\}], \\
\mathcal{H} - (X^{(c)}; c \in \mathcal{S}) &= \overline{\text{sp}}[X_s^{(c)}; s \in \mathbb{Z}, c \in \mathcal{S}'],
\end{aligned}
$$

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 $C = (C_{a,b}; a, b \in \{1, \ldots, p\})$ 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 $v = \text{vec}[u^{(1)}, \ldots, u^{(p)}] \in$

$\ell_{2,p}$ we define the (column) sequence Cv as

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

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*

$$0 < \lambda_{\inf} \leq \lambda_{\sup} < \infty. \tag{2}$$

Assumption 2.1 implies that $\sup_t \sup_a \sum_{\tau\in\mathbb{Z}} \sum_{b=1}^{p} [\mathbf{C}_{t,\tau}]_{a,b}^2 < \infty$ and also the coefficients of the inverse are square summable. It can be shown that if $\sup_{t\in\mathbb{Z}} \sum_{\tau\in\mathbb{Z}} \|\mathbf{C}_{t,\tau}\|_\infty < \infty$, 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 $C : \ell_{2,p} \to \ell_{2,p}$, and since $\mathbf{C}_{t,\tau} = \mathbf{C}'_{\tau,t}$, $\langle v, Cu\rangle = \langle Cv, u\rangle$, 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, for example, Brockwell and Davis (2006), 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

$$P_{\mathcal{H}-X_t^{(a)}}(X_t^{(a)}) = \sum_{\tau\in\mathbb{Z}} \sum_{b=1}^{p} \beta_{(\tau,b)\to(t,a)} X_\tau^{(b)}, \tag{3}$$

with $\beta_{(t,a)\to(t,a)} = 0$ (note the coefficients $\{\beta_{(\tau,b)\to(t,a)}\}$ are unique since C is non-singular). Let $\sigma_{a,t}^2 = \mathrm{E}[X_t^{(a)} - P_{\mathcal{H}-X_t^{(a)}}(X_t^{(a)})]^2$, it can be shown that $\sigma_{a,t}^2 \geq \lambda_{\min}$ (see Appendix A.1). 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 $1 \leq t, \tau \leq p$, we define the $p \times p$-dimensional matrices $\mathbf{D}_{t,\tau}$ as follows

$$[\mathbf{D}_{t,\tau}]_{a,b} = \begin{cases} \frac{1}{\sigma_{a,t}^2} & a = b \text{ and } t = \tau \\ -\frac{1}{\sigma_{a,t}^2}\beta_{(\tau,b)\to(t,a)} & \text{otherwise} . \end{cases} \tag{4}$$

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

$$D_{a,b} = \{[D_{a,b}]_{t,\tau} = [\mathbf{D}_{t,\tau}]_{a,b}; t, \tau \in \mathbb{Z}\} . \tag{5}$$

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

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 (4). Then $C^{-1} = D$ and $\|D\| = \lambda_{\inf}^{-1}$. Further, for all $a, b \in \{1, \ldots, p\}$, $\|D_{a,b}\| \leq \lambda_{\inf}^{-1}$, $\|D_{a,a}^{-1}\| \leq \lambda_{\sup}$ and $\sup_t \sum_{\tau\in\mathbb{Z}} \|\mathbf{D}_{t,\tau}\|_2^2 \leq p\lambda_{\inf}^{-2}$.*

PROOF In Appendix A.1. □

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 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* $\mathbb{C}\mathrm{ov}(X_t^{(a)}, X_\tau^{(b)}|\mathcal{S}_1')$, 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* $\mathbb{C}\mathrm{ov}(X_t^{(a)}, X_\tau^{(b)}|\mathcal{S}_2')$, 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, i.e. 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 $V = \{1, 2, \ldots, p\}$ 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, i.e. $(a, b) \notin E$.*

- **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 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 $\mathcal{S} \subset \{1, \ldots, p\}$ is a called a conditionally stationary subgraph if for all $a, b \in \mathcal{S}$, $D_{a,b}$ are Toeplitz operators i.e. $D_{\mathcal{S},\mathcal{S}}$ 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.

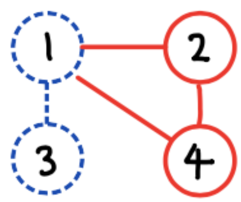
Example 2.1 *Consider the following tvVAR(1) model for a 4-dimensional time series*

$$\begin{pmatrix} X_t^{(1)} \\ X_t^{(2)} \\ X_t^{(3)} \\ X_t^{(4)} \end{pmatrix} = \begin{pmatrix} \alpha(t) & 0 & \alpha_3 & 0 \\ \beta_1 & \beta_2 & 0 & \beta_4 \\ 0 & 0 & \gamma(t) & 0 \\ 0 & \nu_2 & 0 & \nu_4 \end{pmatrix} \begin{pmatrix} X_{t-1}^{(1)} \\ X_{t-1}^{(2)} \\ X_{t-1}^{(3)} \\ X_{t-1}^{(4)} \end{pmatrix} + \underline{\varepsilon}_t = A(t)\underline{X}_{t-1} + \underline{\varepsilon}_t,$$

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.

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



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 $\beta_1\beta_2 \neq 0$) and node 2 (if $\beta_1\beta_4 \neq 0$). By a similar argument, nodes 2 and 4 are connected (if $\beta_1\beta_4 \neq 0$ or $\nu_2\nu_4 \neq 0$).

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 $\mathcal{H}$

$$\rho_{t,\tau}^{(a,b)} = \mathbb{C}\mathrm{ov}\left[X_t^{(a)} - P_{\mathcal{H}-(X_t^{(a)},X_\tau^{(b)})}(X_t^{(a)}), X_\tau^{(b)} - P_{\mathcal{H}-(X_t^{(a)},X_\tau^{(b)})}(X_\tau^{(b)})\right]. \tag{6}$$

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

$$\rho_{t,t}^{(a,a)} = \mathbb{V}\mathrm{ar}\left[X_t^{(a)} - P_{\mathcal{H}-X_t^{(a)}}(X_t^{(a)})\right] = \sigma_{a,t}^2. \tag{7}$$

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 (5). Then the entries of $D_{a,b}$ satisfy the identities*

$$\mathbb{C}\mathrm{orr}\left[X_t^{(a)} - P_{\mathcal{H}-(X_t^{(a)},X_\tau^{(b)})}(X_t^{(a)}), X_\tau^{(b)} - P_{\mathcal{H}-(X_t^{(a)},X_\tau^{(b)})}(X_\tau^{(b)})\right] = -\frac{[D_{a,b}]_{t,\tau}}{\sqrt{[D_{a,a}]_{t,t}[D_{b,b}]_{\tau,\tau}}} \tag{8}$$

and

$$\mathbb{V}\mathrm{ar}\left[\begin{pmatrix} X_t^{(a)} - P_{\mathcal{H}-(X_t^{(a)},X_\tau^{(b)})}(X_t^{(a)}) \\ X_\tau^{(b)} - P_{\mathcal{H}-(X_t^{(a)},X_\tau^{(b)})}(X_\tau^{(b)}) \end{pmatrix}\right] = \begin{pmatrix} [D_{a,a}]_{t,t} & [D_{a,b}]_{t,\tau} \\ [D_{b,a}]_{\tau,t} & [D_{b,b}]_{\tau,\tau} \end{pmatrix}^{-1}. \tag{9}$$

PROOF See Appendix A.2. $\square$

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, viz., 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 (6). Then*

- **Conditional Noncorrelation** $\rho_{t,\tau}^{(a,b)} = 0$ *for all t and τ (i.e. $\rho^{(a,b)} = 0$) iff $D_{a,b} = 0$*
- **Conditionally Stationary Node** $D_{a,a}$ *is Toeplitz iff for all t and τ*

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

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

- **Conditionally Stationary Pair** $D_{a,a}$, $D_{b,b}$ *and $D_{a,b}$ are Toeplitz iff for all t and τ*

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

  i.e. $\rho^{(a,a)}, \rho^{(b,b)}$ and $\rho^{(a,b)}$ are Toeplitz.

PROOF See Appendix A.2. □

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 $V\backslash\{a,b\}$. 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

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

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

$$\begin{pmatrix} \rho_{t,\tau}^{(a,a)|\text{-}\{a,b\}} & \rho_{t,\tau}^{(a,b)|\text{-}\{a,b\}} \\ \rho_{t,\tau}^{(b,a)|\text{-}\{a,b\}} & \rho_{t,\tau}^{(b,b)|\text{-}\{a,b\}} \end{pmatrix} := \mathbb{C}\text{ov}\left[\begin{pmatrix} X_t^{(a)|\text{-}\{a,b\}} \\ X_t^{(b)|\text{-}\{a,b\}} \end{pmatrix}, \begin{pmatrix} X_\tau^{(a)|\text{-}\{a,b\}} \\ X_\tau^{(b)|\text{-}\{a,b\}} \end{pmatrix}\right] \tag{10}$$

and node partial covariance

$$\rho_{t,\tau}^{(a,a)|\text{-}\{a\}} = \mathbb{C}\text{ov}[X_t^{(a)|\text{-}\{a\}}, X_\tau^{(a)|\text{-}\{a\}}]. \tag{11}$$

We will show that the partial covariance in (10) and (11) are closely related to the partial covariance in (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 $D = (D_{a,b}; a, b \in \{1, \ldots, p\})$. Before we present the equivalent definitions of our nonstationary networks in terms of the time series partial covariances $\rho_{t,\tau}^{(a,b)|\text{-}\mathcal{S}}$, we show that $\rho_{t,\tau}^{(a,b)|\text{-}\mathcal{S}}$ can be expressed in terms of the inverse covariance operator D.

Theorem 2.1 *Suppose Assumption 2.1 holds. Let* $\rho_{t,\tau}^{(a,a)|\text{-}\{a,b\}}$,$\rho_{t,\tau}^{(a,b)|\text{-}\{a,b\}}$ *and* $\rho_{t,\tau}^{(a,a)|\text{-}\{a\}}$ *be defined as in (10) and (11) respectively. Then*

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

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

$$\mathbb{V}\mathrm{ar}\left[X_t^{(c)|\text{-}\{a,b\}}; t \in \mathbb{Z}, c \in \{a, b\}\right] = \begin{pmatrix} D_{a,a} & D_{a,b} \\ D_{b,a} & D_{b,b} \end{pmatrix}^{-1} \quad (12)$$

with

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

PROOF See Appendix A.3. □

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* $\rho^{(a,b)|\text{-}\{a,b\}}$*] Suppose Assumption 2.1 holds. Let* $\rho_{t,\tau}^{(a,a)|\text{-}\{a,b\}}$, $\rho_{t,\tau}^{(a,b)|\text{-}\{a,b\}}$ *and* $\rho_{t,\tau}^{(a,a)|\text{-}\{a\}}$ *be defined as in (10) and (11) respectively. Then*

(i) **Conditional noncorrelation** $D_{a,b} = 0$ *iff* $\rho_{t,\tau}^{(a,a)|\text{-}\{a\}} = 0$ *for all* t *and* τ.

(ii) **Conditionally stationary node** $D_{a,a}$ *is a Toeplitz operator iff for all* t *and* τ, $\rho_{t,\tau}^{(a,a)|\text{-}\{a\}} = \rho_{0,t-\tau}^{(a,a)|\text{-}\{a\}}$.

(iii) **Conditionally stationary pair** $D_{a,a}$, $D_{b,b}$ *and* $D_{a,b}$ *are Toeplitz iff for all* t *and* τ, $\rho_{t,\tau}^{(a,a)|\text{-}\{a,b\}} = \rho_{0,t-\tau}^{(a,a)|\text{-}\{a,b\}}$, $\rho_{t,\tau}^{(b,b)|\text{-}\{a,b\}} = \rho_{0,t-\tau}^{(b,b)|\text{-}\{a,b\}}$ *and* $\rho_{t,\tau}^{(a,b)|\text{-}\{a,b\}} = \rho_{0,t-\tau}^{(a,a)|\text{-}\{a,b\}}$.

PROOF See Appendix A.3. $\square$

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 $\mathcal{S} = \{\alpha_1, \ldots, \alpha_r\}$ be a subset of $\{1, \ldots, p\}$ 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*

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

with

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

PROOF See Appendix A.3. $\square$

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 non-zeros". 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 $\underline{g}(\omega)' = (g_1(\omega), \ldots, g_p(\omega)) \in L_2[0, 2\pi)^p$ 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 $F : L_2[0,2\pi) \to \ell_2$ and $F^* : \ell_2 \to L_2[0,2\pi)$

$$[F(g)]_j = \frac{1}{2\pi}\int_0^{2\pi} \underline{g}(\lambda)\exp(ij\lambda)d\lambda \quad \text{and} \quad F^*(v)(\omega) = \sum_{j\in\mathbb{Z}} v_j \exp(-ij\omega). \tag{13}$$

It is well known that F and F^* are isomorphisms between ℓ_2 and $L_2[0,2\pi)$ (see, for example, Brockwell and Davis (2006), Section 2.9). For $d > 1$ the transformations $F(\underline{g}) = (F(g_1),\ldots,F(g_d))$ and $F^*v = (F^*v^{(1)},\ldots,F^*v^{(d)})$ where $v = (v^{(1)},\ldots,v^{(d)})$ are isomorphisms between $\ell_{2,d}$ and $L_2[0,2\pi)^d$. Often we use that $d = p$. These two isomorphims 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 $A = (A_{a,b}; a,b \in \{1,\ldots,d\})$, if $A : \ell_{2,d} \to \ell_{2,d}$ is a bounded operator, then standard results show that $F^*AF : L_2[0,2\pi)^d \to L_2[0,2\pi)^d$ is a bounded operator (see Conway (1990), Chapter II). F^*AF is an integral operator, such that for all $g \in L_2[0,2\pi)^d$

$$F^*AF(g)[\omega] = \frac{1}{2\pi}\int_0^{2\pi} \mathbf{A}(\omega,\lambda)\underline{g}(\lambda)d\lambda, \tag{14}$$

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

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

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*

$$\mathrm{Cov}[X_t, X_\tau] = C_{t,\tau} = \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi} G(\omega,\lambda)\exp(-it\omega + i\tau\lambda)d\omega d\lambda, \tag{15}$$

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 Gladyšev (1963), Lund et al. (1995), Lii and Rosenblatt (2002), Jensen and Colgin (2007), Hindberg and Olhede (2010), Olhede (2011), Olhede and Ombao (2013), Gorrostieta et al. (2019), Aston et al. (2019).

If $\{X_t\}$ were a second order stationary time series, then (15) reduces to Bochner's

Theorem

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

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.

$\mathbf{A}(\omega, \lambda)$ 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 $\mathbf{A}(\omega, \lambda)$ that we will exploit. This will become apparent in the following lemma due to Toeplitz (1911) (we state the result for the case $d = 1$).

Lemma 3.1 *Suppose A is an infinite dimensional bounded matrix operator $A : \ell_2 \to \ell_2$. 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.

PROOF See Appendix B.1 for details. □

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

PROOF In Appendix B.1. □

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

We use the operators $F : L_2[0, 2\pi)^p \to \ell_{2,p}$ and $F^* : \ell_{2,p} \to L_2[0, 2\pi)^p$ 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 (14) F^*CF is an integral operator with matrix kernel $\mathbf{C}(\omega, \lambda) = (C_{a,b}(\omega, \lambda); a, b \in \{1, \ldots, p\})$ 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 non-diagonal. 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 is used in (Epharty et al., 2001;

Dwivedi and Subba Rao, 2011; Jentsch and Subba Rao, 2015) 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 $\mathbf{K}(\omega,\lambda) = (K_{a,b}(\omega,\lambda); a,b \in \{1,\ldots,p\})$ where

$$K_{a,b}(\omega,\lambda) = \sum_{t\in\mathbb{Z}}\sum_{\tau\in\mathbb{Z}}[D_{a,b}]_{t,\tau}\exp(it\omega - i\tau\lambda) = \sum_{t\in\mathbb{Z}}\Gamma_t^{(a,b)}(\lambda)\exp(it(\omega-\lambda)) \tag{16}$$

and

$$\Gamma_t^{(a,b)}(\lambda) = \sum_{r\in\mathbb{Z}}[D_{a,b}]_{t,t+r}\exp(ir\lambda). \tag{17}$$

Note that under Assumption 2.1 $\|D\| < \infty$, this implies for all $a,b \in \{1,\ldots,p\}$ 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 $\Gamma_t^{(a,b)}(\lambda)$ haven 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 Priestley (1965), Martin and Flandrin (1985), Dahlhaus (1997), Birr et al. (2018)). 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 $\{\underline{X}_t\}$. 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.*

PROOF In Appendix B.2. □

These equivalences show that conditional noncorrelatedness and stationarity relationships in the graphical model, as defined by the D operator, are encoded in the object $K(.,.)$.

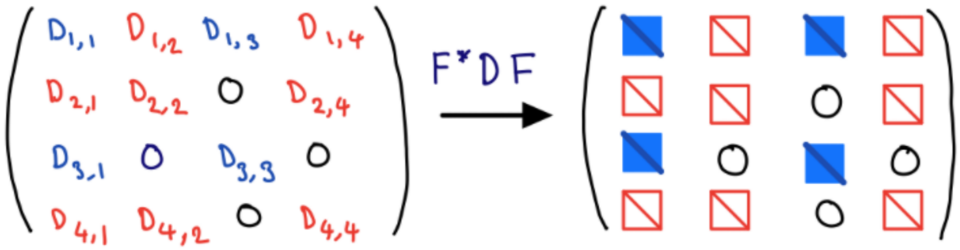


Figure 2: Illustration of the mapping of matrix D to the integral kernel corresponding to F^*DF. The diagonal red box indicates the mass of $F^*D_{a.b}F$ lies only on the diagonal (it corresponds to a Toeplitz matrix). The blue filled box indicates that the mass of $F^*D_{a.b}F$ lies both on the diagonal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

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*

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 (17), 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

$$\phi_{t,\tau}^{(a,b)} \quad = \quad \mathbb{C}\mathrm{orr}[X_t^{(a)} - P_{\mathcal{H}-(X_t^{(a)},X_\tau^{(b)})}(X_t^{(a)}), X_\tau^{(b)} - P_{\mathcal{H}-(X_t^{(a)},X_\tau^{(b)})}(X_\tau^{(b)})]. \qquad (18)$$

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 (6) and (18).*

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

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

(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

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

PROOF In Appendix B.2. □

For StGM, the partial spectral coherence is typically defined in terms of the Fourier transform of the partial time series covariances (see Priestley (1981), Section 9.3, and Dahlhaus (2000b)). 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

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

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

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

PROOF In Appendix B.2. □

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 Dahlhaus (2000b)). Analogously, by using Theorem 3.3(ii) the spectral partial coherence between the conditionally stationary pair (a,b) is

$$R_{a,b}(\omega) = -\frac{\Gamma^{(a,b)}(\omega)}{\sqrt{\Gamma^{(a,a)}(\omega)\Gamma^{(b,b)}(\omega)}}. \tag{19}$$

In Appendix F we show how this expression is related to the spectral partial coherence for stationary time series.

3.4 Connection to node-wise 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 node-wise 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

$$P_{\mathcal{H}-(X^{(a)})}(X_t^{(a)}) = \sum_{b\neq a}\sum_{\tau\in\mathbb{Z}} \alpha_{(\tau,b)\to(t,a)} X_\tau^{(b)}. \tag{20}$$

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

$$[B_{b\to a}]_{t,\tau} = \alpha_{(\tau,b)\to(t,a)} \text{ for all } t,\tau \in \mathbb{Z}. \tag{21}$$

In the lemma below we connect the coefficients in the infinite dimensional matrix $B_{b\to a}$ 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 (5). Then for all $b \neq a$ we have*

$$D_{a,b} = -D_{a,a}B_{b\to a}. \tag{22}$$

PROOF See Appendix B.3. □

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 $B_{b\to a}$ be defined as in (21). Then*

(i) $B_{b\to a} = 0$ *iff* $D_{a,b} = 0$.

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

PROOF See Appendix B.3. □

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

Corollary 3.1 *Suppose Assumption 2.1 holds. Let $B_{b\to a}$ be defined as in (22). Let $K_{b\to a}(\omega, \lambda)$ denote the integral kernel associated with $B_{b\to a}$. Then*

(i) $B_{b\to a}$ is a bounded operator.

(ii) **Conditionally noncorrelated** $\{X_t^{(a)}, X_t^{(b)}\}_t$ *are conditionally noncorrelated iff* $K_{a \nrightarrow b}(\omega, \lambda) \equiv 0$.

(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* $K_{b \nrightarrow a}(\omega, \lambda)$ *are diagonal.*

PROOF In Appendix B.3. □

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

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

with $\mathrm{E}[X^{(a)} | \mathcal{H} - (X^{(a)})] = \sum_{b=1, b\neq a}^{p} B_{b \nrightarrow a} X^{(b)}$ and $\mathbb{V}\mathrm{ar}[X^{(a)} | \mathcal{H} - (X^{(a)})] = D_{aa}^{-1}$. Some interesting simplications 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

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

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 (13)) 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 $\mathbf{X}_n' = ((\underline{X}^{(1)})', \ldots, (\underline{X}^{(p)})')$, where $\underline{X}^{(a)} = (X_1^{(a)}, \ldots, X_n^{(a)})'$ for $a \in \{1, \ldots, p\}$. Then $F_n^* \mathbf{X}_n$ is a np-dimension vector where $(F_n^* \mathbf{X}_n)' = ((\mathcal{F}_n^* \underline{X}^{(1)})', \ldots, (\mathcal{F}_n^* \underline{X}^{(p)})')$ with

$$J_k^{(a)} = [\mathcal{F}_n^* \underline{X}^{(a)}]_k = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t^{(a)} \exp(it\omega_k) \quad k = 1, \ldots, n \text{ and } \omega_k = \frac{2\pi k}{n}. \tag{23}$$

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

$$\mathbf{K}_n = [\mathbb{V}\mathrm{ar}[F_n^* \mathbf{X}_n]]^{-1} = [F_n^* C_n F_n]^{-1} = [F_n]^{-1} C_n^{-1} [F_n^*]^{-1} = F_n^* \widetilde{D}_n F_n,$$

where $\widetilde{D}_n = C_n^{-1}$ and the above follows from the identity $F_n^{-1} = F_n^*$. Let $\mathbf{K}_n = ([\mathbf{K}_n]_{a,b}; a, b \in \{1, \ldots, p\})$ 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 $C_n = \mathbb{V}\mathrm{ar}[\mathbf{X}_n]$ is a submatrix of the infinite dimensional C. Unfortunately its inverse $\widetilde{D}_n = C_n^{-1}$ 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 $\widetilde{D}_n$ can be approximated by a finite dimensional submatrix of D. To do this we represent $\widetilde{D}_n$ as $p \times p$ submatrices each of dimension $n \times n$

$$\widetilde{D}_n = \left(\widetilde{D}_{a,b;n}; a, b \in \{1, \ldots, p\} \right). \tag{24}$$

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

$$D_n = (D_{a,b;n}; a, b \in \{1, \ldots, p\}) \tag{25}$$

where $D_{a,b;n} = \{[D_{a,b}]_{t,\tau}; t, \tau \in \{1, \ldots, n\}\}$. Below we show that under suitable conditions, $\widetilde{D}_n = C_n^{-1}$ 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 covariance $\mathbf{D}_{t,\tau}$ defined in (4) satisfy the condition* $\sup_t \sum_{j \neq 0} |j|^K \|\mathbf{D}_{t,t+j}\|_\infty < \infty$ *(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. Krampe and Subba Rao (2022) obtain an equivalence between the rate of decay on $\mathbf{D}_{t,\tau}$ and $\mathbf{C}_{t,\tau}$. In particular, Krampe and Subba Rao (2022) Theorem 2.1, show that under Assumption 2.1 and if for some $K > 7/2$ and all $|r| \neq 0$

we have that $\sup_t \|\mathbf{C}_{t,t+r}\| < K|r|^{-K}$ (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 $\widetilde{D}_n$ and D_n.

Theorem 4.1 *Suppose Assumptions 2.1 and 4.1 hold. Let $\widetilde{D}_n$ and D_n be defined as in (24) and (25). Then for all $1 \leq t \leq n$ we have*

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

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

PROOF See Appendix C.1. □

The theorem above shows that the further t lies from the two end boundaries of the sequence $\{1,2,\ldots,n\}$ the better the approximation between $[\widetilde{D}_n]_{(a-1)n+t,\cdot}$ and $[D_n]_{(a-1)n+t,\cdot}$. For example when $t = n/2$ (recall that p is fixed) $\|[\widetilde{D}_n]_{(a-1)n+t,\cdot} - [D_n]_{(a-1)n+t,\cdot}\|_1 = O(1/n^{K-1/2})$. Using Theorem 4.1 we replace $F_n^*\widetilde{D}_nF_n$ with $F_n^*D_nF_n$ 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 (17). Then*

$$\begin{aligned}
[\mathbf{K}_n(\omega_{k_1},\omega_{k_2})]_{a,b} &= \frac{1}{n}\sum_{t=1}^{n}\Gamma_t^{(a,b)}(\omega_{k_2})\exp(-it(\omega_{k_1}-\omega_{k_2})) + O\left(\frac{1}{n}\right) \\
&= \left[\frac{1}{n}\sum_{t=1}^{n}\Gamma_t^{(b,a)}(\omega_{k_1})\exp(-it(\omega_{k_2}-\omega_{k_1}))\right]^* + O\left(\frac{1}{n}\right) \qquad (26)
\end{aligned}$$

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

$$[\mathbf{K}_n(\omega_{k_1},\omega_{k_2})]_{a,b} = \begin{cases} \Gamma^{(a,b)}(\omega_k) + O(n^{-1}) & k_1 = k_2 (= k) \\ O(n^{-1}) & k_1 \neq k_2 \end{cases} \qquad (27)$$

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

PROOF See Appendix C.1. □

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 Priestley (1965), Dahlhaus and Giraitis (1998), Zhou and Wu (2009), Cardinali and Nason (2010), Kley et al. (2019), Dahlhaus et al. (2019), Sundararajan and Pourahmadi (2018), Ding and Zhou (2020), Ombao and Pinto (2021), 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.

The locally stationary process were formally proposed in Dahlhaus (1997). 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

$$\|\underline{X}_{t,n} - \underline{X}_t(u)\|_2 = O_p\left(\frac{1}{n} + \left|\frac{t}{n} - u\right|\right). \tag{28}$$

Thus for every t, $\underline{X}_{t,n} = (X_{t,n}^{(1)}, \ldots, X_{t,n}^{(p)})'$ can be closely approximated by an auxillary variable $\underline{X}_t(u)$ (where $u = t/n$); see (Dahlhaus and Subba Rao, 2006; Subba Rao, 2006; Dahlhaus, 2012; Dahlhaus et al., 2019). However, as the difference between t/n and u grows, the similarity between $X_{t,n}$ and the auxillary 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 (28) is a useful start for analysing nonstationary time series, analogous to Dahlhaus and Polonik (2006), we require additional local stationarity conditions on the moment structure. Dahlhaus (2000a) and Dahlhaus and Polonik (2006) state the conditions in terms of bounds between $\mathbb{C}\text{ov}[\underline{X}_{t,n}, \underline{X}_{\tau,n}]$ and $\mathbb{C}\text{ov}[\underline{X}_0(u), \underline{X}_{t-\tau}(u)]$. 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 $\mathbf{D}_{t-\tau} : \mathbb{R} \to \mathbb{R}^{p\times p}$ where*

$$\mathbf{D}_{t,\tau} = \mathbf{D}_{t-\tau}\left(\frac{t+\tau}{2n}\right) + O\left(\frac{1}{n\ell(t-\tau)}\right) \quad t, \tau \in \mathbb{Z}. \tag{29}$$

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) $\sup_u |\frac{d[\mathbf{D}_j(u)]_{a,b}}{du}| \leq \ell(j)^{-1}$, (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) $\sup_u |\frac{d[\mathbf{D}_j(u)]_{a,b}}{du}| \leq \ell(j)^{-1}$.

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 Krampe and Subba Rao (2022) 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 require that the entry wise derivative of matrix functions $\mathbf{D}_{t-\tau}(\cdot)$ exists. This technical condition 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 Dahlhaus and Polonik (2006).

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.

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 $\mathbf{\Sigma}(u;\omega)$. The time-varying spectral density matrix corresponding to $\{\underline{X}_{t,n}\}_t$ is $\{\mathbf{\Sigma}(t/n;\omega)\}_t$. In contrast, in this section our focus will be on the inverse $\mathbf{\Gamma}(u;\omega) = \mathbf{\Sigma}(u;\omega)^{-1}$, where by Lemma 3.2, $\mathbf{\Gamma}(u;\omega)$ is the Fourier transform of $\mathbf{D}_j(u)$ over the lags j i.e.

$$\mathbf{\Gamma}(u;\omega) = \sum_{j\in\mathbb{Z}} \mathbf{D}_j(u)\exp(ij\omega). \tag{30}$$

We note that $\mathbf{\Gamma}(u;\omega) = (\Gamma^{(a,b)}(u;\omega); 1 \leq a, b \leq p)$. We use Assumption 4.2 to relate $\Gamma^{(a,b)}(u;\omega)$ to $\Gamma_t^{(a,b)}(\omega)$ (defined in (17)). In particular, $\Gamma^{(a,b)}(t/n;\omega)$ is an approximation of $\Gamma_t^{(a,b)}(\omega)$ and

$$\left|\Gamma_t^{(a,b)}(\omega) - \Gamma^{(a,b)}(u;\omega)\right| \leq C\left(\left|\frac{t}{n} - u\right| + \frac{1}{n}\right). \tag{31}$$

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

Our aim is to relate $\mathbf{\Gamma}(u;\omega)$ to $\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})$. First we notice that $\mathbf{\Gamma}(u;\omega)$ is "local" in the sense that it is a time local approximation to the precision spectral density at time point $t = \lfloor un \rfloor$. 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 $\mathbf{\Gamma}(u;\omega)$, as it measures how $\mathbf{\Gamma}(u;\omega)$ 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

$$K_r^{(a,b)}(\omega) \;=\; \int_0^1 \exp(-2\pi iru)\Gamma^{(a,b)}(u;\omega)du. \tag{32}$$

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 (32). 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 conditionally time-invariant iff asymptotically* $K_r^{(a,b)}(\omega) \equiv 0$ *for all* $r \neq 0$ *and* $\omega \in [0, 2\pi]$.

PROOF in Appendix C.2. □

Note that the above result is asymptotic in rescaled time ($n \to \infty$). 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 (32). Then*

$$[\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})]_{a,b} = \frac{1}{n} \sum_{t=1}^{n} \exp\left(-\frac{i2\pi(k_1 - k_2)t}{n}\right) \Gamma^{(a,b)}\left(\frac{t}{n}; \omega_{k_2}\right) + O\left(\frac{1}{n}\right). \quad (33)$$

Further,

$$[\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})]_{a,b} = \begin{cases} K_{k_1-k_2}^{(a,b)}(\omega_{k_2}) + O\left(\frac{1}{n}\right) & \text{if } |k_1 - k_2| \leq n/2 \\ K_{k_1-k_2-n}^{(a,b)}(\omega_{k_2}) + O\left(\frac{1}{n}\right) & \text{if } n/2 < (k_1 - k_2) < n \\ K_{k_1-k_2+n}^{(a,b)}(\omega_{k_2}) + O\left(\frac{1}{n}\right) & \text{if } -n < (k_1 - k_2) < -n/2 \end{cases} \quad (34)$$

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 (34) 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.*

PROOF in Appendix C.2. □

Note we 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} lies at the boundaries of the interval $[0, 2\pi]$.

In Dwivedi and Subba Rao (2011) and Jentsch and Subba Rao (2015) we showed that the Fourier transform of the time-varying spectral density matrix $\boldsymbol{G}_r(\omega) = \int_0^1 e^{-2\pi i r u} \boldsymbol{\Sigma}(u; \omega) du$ decayed to zero as $|r| \to \infty$ 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*

$$\sup_{\omega} |K_r^{(a,b)}(\omega)| \to 0 \text{ as } r \to \infty \tag{35}$$

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

$$\left|K_r^{(a,b)}(\omega_1) - K_r^{(a,b)}(\omega_2)\right| \leq \begin{cases} C|\omega_1 - \omega_2| & r = 0 \\ C|r|^{-1}|\omega_1 - \omega_2| & r \neq 0 \end{cases} \tag{36}$$

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

PROOF in Appendix C.2. □

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 rth 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 $r\text{mod}(n)$ 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.

4.4 Node-wise 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 Dempster (1972) for the low-dimensional and Meinshausen and Bühlmann (2006) for the high-dimensional setting. In particular, the neighborhood selection approach of Meinshausen and Bühlmann (2006) regresses one component of a multivariate random vector on the other components with Lasso (Tibshirani, 1996), and uses non-zero regression coefficients to select its neighborhood, i.e. 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 (23). 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 $\mathcal{G}_n = \overline{\mathrm{sp}}(J_k^{(b)}; 1 \leq k \leq n, 1 \leq b \leq p)$ (note that the coefficients in this space can be complex). Then

$$P_{\mathcal{G}_n - J_k^{(a)}}(J_k^{(a)}) = \sum_{b=1}^{p}\sum_{s=1}^{n} B_{(b,s)\rightarrowtail(a,k)} J_s^{(b)}, \tag{37}$$

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

$$\Delta_k^{(a)} = \mathbb{V}\mathrm{ar}\left(J_k^{(a)} - P_{\mathcal{G}_n - J_k^{(a)}}(J_k^{(a)})\right). \tag{38}$$

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,

$$[\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})]_{a,b} = \begin{cases} \frac{1}{\Delta_{k_1}^{(a)}} & k_1 = k_2 \text{ and } a = b \\ -\frac{1}{\Delta_{k_1}^{(a)}} B_{(b,k_2)\rightarrowtail(a,k_1)} & \text{otherwise} \end{cases}. \tag{39}$$

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

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

where

$$B_{r,n}^{(b\rightarrowtail a)}(\omega_k) = \begin{cases} -K_0^{(a,a)}(\omega_k)^{-1} K_r^{(b,a)}(\omega_k)^* & \text{if } |r| \leq n/2, r \neq 0 \\ -K_0^{(a,a)}(\omega_k)^{-1} K_{r-n}^{(b,a)}(\omega_k)^* & \text{if } n/2 < r < n \\ -K_0^{(a,a)}(\omega_k)^{-1} K_{r+n}^{(b,a)}(\omega_k)^* & \text{if } -n < r < -n/2 \end{cases}. \tag{40}$$

Thus by using Proposition 4.4 we have

$$\left|B_{(b,k_1+r)\rightarrowtail(a,k_1)} - B_{(b,k_2+r)\rightarrowtail(a,k_2)}\right| \leq A|\omega_{k_1} - \omega_{k_2}| + O(n^{-1}), \tag{41}$$

where A is a finite constant. The benefit of these results is in the estimation of the coefficients $B_{(b,k+r)\rightarrowtail(a,k)}$. We recall (37) can be expressed as

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

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 (41) 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 $B_{(b,k+r)\rightarrowtail(a,k)}$ 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 $M << n$) 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 $|B_{(b,k+r)\to(a,k)}| \sim |r|^{-1}$ we can truncate the projection to a small number $(2\nu+1)$ of regressors about $\underline{J}_k$ to obtain the approximation

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

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 $\mathcal{Y}_k^{(a)} = (J_{k-M}^{(a)}, J_{k-M+1}^{(a)}, \ldots, J_k^{(a)}, J_{k+1}^{(a)}, \ldots, J_{k+M}^{(a)})'$, 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

$$\begin{aligned} \hat{B}_{(.,.)\to(a,k)} &= \Big(\hat{B}_{(1,k-\nu)\to(a,k)}, \ldots, \hat{B}_{(p,k-v)\to(a,k)}, \ldots, \hat{B}_{(1,k)\to(a,k)}, \ldots, \hat{B}_{(k-1,k)\to(a,k)}, \\ &\qquad \hat{B}_{(k+1,k)\to(a,k)}, \hat{B}_{(p,k)\to(a,k)}, \ldots, \hat{B}_{(1,k+v)\to(a,k)}, \ldots, \hat{B}_{(p,k+\nu)\to(a,k)}\Big)' \end{aligned}$$

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

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

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 (Maleki et al., 2013) 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 $\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})$ for different Fourier frequencies $\omega_{k_1}, \omega_{k_2}$. Furthermore, from equation (39) and (40) it is possible to learn the sparsity structure of $\mathbf{K}_n(\omega_{k_1}, \omega_{k_2})$ from the regression coefficients $B_{(b,k_1)\to(a,k_2)}$ (up to order $O(n^{-1})$). In particular, there is an edge $(a,b) \subseteq E$, i.e. the components a and b are conditionally correlated, if $B_{(b,k_1)\to(a,k_2)}$ is non-zero for some k_1, k_2 (within the locally stationary framework). Similarly, an edge between a and b is conditionally time-varying if $B_{(b,k_1)\to(a,k_2)}$ is non-zero 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

$$\hat{W}_{self} = \left(\left(\sum_{k} |\hat{B}_{(b,k) \rightarrow (a,k)}|^2 \right) \right)_{1 \leq a,b \leq p} \tag{42}$$

$$\hat{W}_{other} = \left(\left(\sum_{k_1 \neq k_2} |\hat{B}_{(b,k_1) \rightarrow (a,k_2)}|^2 \right) \right)_{1 \leq a,b \leq p} \tag{43}$$

for graphical model selection in NonStGM. Two components a and b are deemed conditionally noncorrelated if both the $(a,b)^{th}$ and the $(b,a)^{th}$ off-diagonal elements of $\hat{W}_{self} + \hat{W}_{other}$ are small. In contrast, a node a is deemed conditionally stationary if the $(a,a)^{th}$ element of $\hat{W}_{other}$ 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 $\hat{W}_{other}$ is small. Note that our node-wise regression approach does not ensure that the estimated weight matrices $\hat{W}$ are symmetric. However, following (Meinshausen and Bühlmann, 2006), 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 rigourous understanding of certain features in a tvVAR model.

The time-varying VAR (tvVAR) model is often used to model nonstationarity (see Subba Rao (1970), Dahlhaus (2000a), Dahlhaus and Polonik (2006), Zhang and Wu (2021), Safikhani and Shojaie (2020)). A time series is said to have a time-varying VAR(∞) representation if it can be expressed as

$$\underline{X}_t = \sum_{j=1}^{\infty} \mathbf{A}_j(t) \underline{X}_{t-j} + \underline{\varepsilon}_t \quad t \in \mathbb{Z} \tag{44}$$

where $\{\underline{\varepsilon}_t\}_t$ are i.i.d random vectors with $\mathbb{V}\mathrm{ar}[\underline{\varepsilon}_t] = \mathbf{\Sigma}$ and $\mathrm{E}[\underline{\varepsilon}_t] = 0$. For simplicity, we have centered the time series as the focus is on the second order structure of the time series. We assume that (44) 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 (44) 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 $\mathbf{C}_{t,\tau} = \mathbb{C}\text{ov}[\underline{X}_t, \underline{X}_\tau]$ and C denote the corresponding covariance operator as defined in (1). Let $\mathbf{H}$ denote the Cholesky decomposition of $\mathbf{\Sigma}^{-1}$ such that $\mathbf{\Sigma}^{-1} = \mathbf{H}'\mathbf{H}$ (where $\mathbf{H}'$ denotes the transpose of $\mathbf{H}$). To obtain D we use the Gram-Schmidt orthogonalisation. We define the following matrices;

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

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

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

By definition of (44) 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 rearranged version of C. We use this to deduce the inverse $D = C^{-1}$. We define $\mathbf{D}_{t,\tau}$ as

$$\mathbf{D}_{t,\tau} = \sum_{\ell=-\infty}^{\infty} \widetilde{\mathbf{A}}_\ell(t+\ell)'\Sigma^{-1}\widetilde{\mathbf{A}}_{(\tau-t)+\ell}(t+\ell). \tag{45}$$

The inverse of C is $D = (D_{a,b}; 1 \leq a, b \leq p)$, where $D_{a,b}$ is defined by substituting (45) into (5).

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

$$[D_{a,b}]_{t,t+r} = \begin{cases} \sum_{\ell=1}^{\infty}\langle[\mathbf{A}_\ell(t+\ell)]_{\cdot,a}[\mathbf{A}_{\ell+r}(t+\ell)]_{\cdot,b}\rangle - \langle[I_p]_{\cdot,a}, [\mathbf{A}_r(t+\ell)]_{\cdot,b}\rangle, & r \geq 0 \\ \sum_{\ell=1}^{\infty}\langle[\mathbf{A}_\ell(t+\ell)]_{\cdot,b}, [\mathbf{A}_{\ell-r}(t+\ell)]_{\cdot,a}\rangle - \langle[I_p]_{\cdot,b}, [\mathbf{A}_{-r}(t+\ell)]_{\cdot,a}\rangle, & r < 0 \end{cases}, \tag{46}$$

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 non-zero entries in the columns $[\widetilde{\mathbf{A}}_\ell(t)]_{\cdot,a}$ and $[\widetilde{\mathbf{A}}_\ell(t)]_{\cdot,b}$ 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 $[\widetilde{\mathbf{A}}_\ell(t)]_{\cdot,a}$ 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 $[\widetilde{\mathbf{A}}_\ell(t)]_{\cdot,a}$ and $[\widetilde{\mathbf{A}}_\ell(t)]_{\cdot,b}$ 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 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 Krampe and Subba Rao (2022), 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 i.e.*

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

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 $\mathrm{Var}[\varepsilon_t] = I_p$. 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 node-wise 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.

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 nonstationay and the edge $(1, 3)$ is conditionally time-varying. On the other hand, the nodes $2, 4$ are conditionally stationary and the edge $(2, 4)$, $(1, 2)$ and (1,4) are conditionally time-invariant. As Figure 1a shows, these nuanced relationships are not prominent from the four time series trajectories.

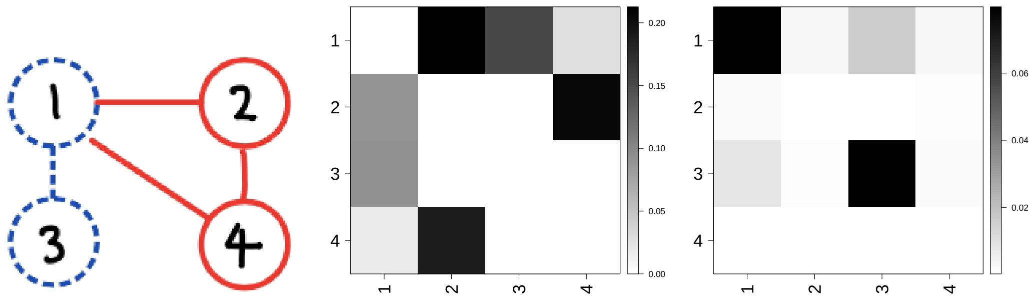


Figure 3: NonStGM selection with node-wise regression for a $p = 4$ dimensional system. [Left]: True graph structure. [Middle]: Heat map of $\hat{W}_{self}$ showing conditional noncorrelation between components $(1, 2)$, $(1, 3)$, $(1, 4)$ and $(2, 4)$. [Right]: Heat map of $\hat{W}_{other}$ showing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

We perform node-wise 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 $\hat{W}_{self}$ and $\hat{W}_{other}$. 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 $\hat{W}_{self}$ (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 $\hat{W}_{other}$ (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 $\varepsilon_t \overset{i.i.d.}{\sim} N(0, I_{10})$. Non-zero 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 $n = 15000$ observations from this system, and perform node-wise regression of DFTs with $M = \lceil \sqrt{n} \rceil$ and $\nu = 1$. The tuning parameters in the individual group

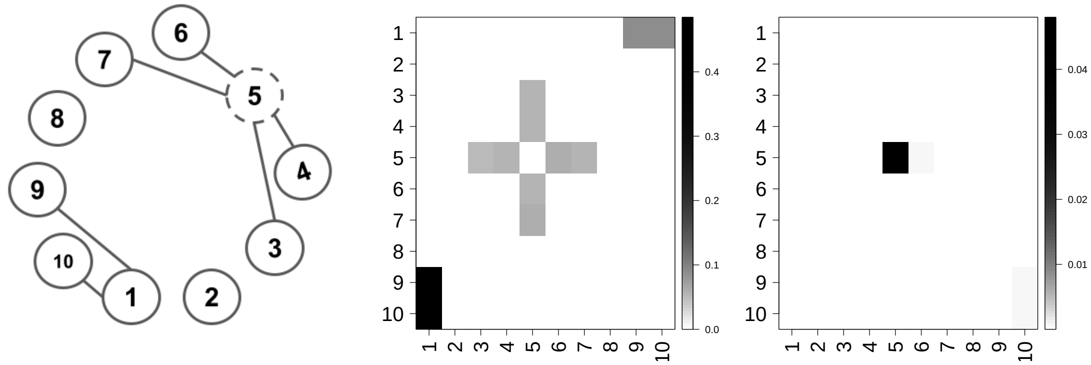


Figure 4: NonStGM selection with node-wise regression for a $p = 10$ dimensional system. [Left]: True graph structure. [Middle]: Heat map of $\hat{W}_{self}$ showing conditional noncorrelation captured by the edges. [Right]: Heat map of $\hat{W}_{other}$ showing conditional nonstationarity of node 5.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

lasso regressions were selected using cross-validation. The estimated regression coefficients $\hat{B}$ were used to construct the weight matrices $\hat{W}_{self}$ and $\hat{W}_{other}$. 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 $\hat{W}_{self}$ (middle). As expected the isolated nodes do not show up. The heat map of $\hat{W}_{other}$ (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. (Woodroofe and Van Ness, 1967; Brillinger, 2001; Wu and Zaffaroni, 2018; Rosuel et al., 2021)). To estimate the inverse of moderate to high-dimensional spectral density matrices, penalized estimation methods for detecting non-zero off-diagonal entries (Fiecas et al., 2019) 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 non-zero coefficients of node-wise DFT regression across different frequencies. In future work, we hope to develop a complete statistical theory for graphical model estimation and inference.

Acknowledgements

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

References

Steen A Andersson, David Madigan, and Michael D Perlman. Alternative markov properties for chain graphs. *Scandinavian journal of statistics*, 28(1):33–85, 2001.

J. Aston, D. Dehay, J-M. Dudek, A. Freyermuth, D. Szucs, and L. Colling. (dual-frequency)-dependent dynamic functional connectivity analysis of vidual working memory capacity. *Preprint: Hal-021335535*, 2019.

A. Aue and A. van Delft. Testing for stationarity of functional time series in the frequency domain. *Ann. Statist.*, 48:2505–2547, 2020.

S. Basu and G. Michailidis. Regularized estimation in sparse high-dimensional time series models. *Ann. Statist.*, 43(4):1535–1567, 2015.

G. Berkolaiko and P. Kuchment. Spectral shift via literal pertubation. *arXiv preprint arXiv:2011.11142*, 2020.

S. Birr, H. Dette, M. Hallin, T. Kley, and S. Volgushev. On WignerVille spectra and the uniqueness of time-varying copula-based spectral densities. *J. Time Series Anal.*, 39: 242–250, 2018.

H. Böhm and R. von Sachs. Shrinkage estimation in the frequency domain of multivariate time series. *J. Multivar. Anal.*, 100(5):913–935, 2009.

Albrecht Böttcher and Sergei M. Grudsky. *Toeplitz matrices, asymptotic linear algebra, and functional analysis*. Birkhäuser Verlag, Basel, 2000.

D. R. Brillinger. Remarks concerning graphical models for time series and point processes. *R de Econometrica*, 16:1–23, 1996.

David R. Brillinger. *Time series: Data Analysis and theory*, volume 36 of *Classics Appl. Math.* SIAM, Philadelphia, PA, 2001.

Peter J. Brockwell and Richard A. Davis. *Time series: theory and methods*. Springer Series in Statistics. Springer, New York, 2006. Reprint of the second (1991) edition.

A. Cardinali and G.P. Nason. Costationarity of locally stationary time series. *J. Time Ser. Econom.*, 2(2):Art. 1, 33, 2010.

J. Chau and R. von Sachs. Intrinsic wavelet regression for surfaces of Hermitian positive definite matrices. *J. Am. Stat. Assoc.*, 2020.

J. B. Conway. *A course in functional analysis*. Springer, New York, 1990.

R. Dahlhaus. Fitting time series models to nonstationary processes. *Ann. Statist.*, 25(1): 1–37, 1997.

R. Dahlhaus. A likelihood approximation for locally stationary processes. *Ann. Statist.*, 28(6):1762–1794, 2000a.

R. Dahlhaus. Graphical interaction models for multivariate time series. *Metrika*, 51(2): 157–172, 2000b.

R. Dahlhaus. *Handbook of Statistics, volume 30*, chapter Locally Stationary processes, pages 351–413. Elsevier, 2012.

R. Dahlhaus and M. Eichler. Causality and graphical models in time series analysis. In *Highly structured stochastic systems*, volume 27 of *Oxford Statist. Sci. Ser.*, pages 115–144. 2003a.

R. Dahlhaus and L. Giraitis. On the optimal segment length for parameter estimates for locally stationary time series. *J. Time Series Anal.*, 19:629–655, 1998.

R. Dahlhaus and W. Polonik. Nonparametric quasi-maximum likelihood estimation for Gaussian locally stationary processes. *Ann. Statist.*, 34(6):2790–2824, 2006.

R. Dahlhaus and S. Subba Rao. Statistical inference of time varying ARCH processes. *Ann. Statistics*, 34:1074–1114, 2006.

R. Dahlhaus, M. Eichler, and J. Sandühler. *J. Neuroscience Methods*, 77:93–107, 1997.

R. Dahlhaus, S. Richter, and W. B. Wu. Towards a general theory for nonlinear locally stationary processes. *Bernoulli*, 25(2):1013–1044, 2019.

Rainer Dahlhaus and Michael Eichler. Causality and graphical models in time series analysis. *Oxford Statistical Science Series*, pages 115–137, 2003b.

R.A. Davis, P. Zang, and T. Zheng. Sparse vector autoregressive modeling. *J. of Computat. Graph. Statist.*, 25(4):1077–1096, 2016.

A. P. Dempster. Covariance selection. *Biometrics*, pages 157–175, 1972.

F.X. Diebold and K. Yılmaz. On the network topology of variance decompositions: Measuring the connectedness of financial firms. *J. Econometrics*, 182(1):119–134, 2014.

X. Ding and Z. Zhou. Estimation and inference for precision matrices of non-stationary time series. *Ann. Statist.*, 48:2455–2477, 2020.

Y. Dwivedi and S. Subba Rao. A test for second order stationarity based on the Discrete Fourier Transform. *J. Time Series Anal.*, 32:68–91, 2011.

M. Eichler. Granger causality and path diagrams for multivariate time series. *J. Econometrics*, 137(2):334–353, 2007.

M. Eichler. Testing nonparametric and semiparametric hypotheses in vector stationary processes. *J. Multivariate Anal.*, 99:968–1009, 2008.

A. Epharty, J. Tabrikian, and H. Messer. Underwater source detection using a spatial stationary test. *Journal of the Acoustical Society of America*, 109:1053–1063, 2001.

M. Fiecas, C. Leng, W. Liu, and Y Yu. Spectral analysis of high-dimensional time series. *Electron. J. Stat.*, 13(2):4079–4101, 2019.

J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

E. G. Gladyšev. Periodically and semi-periodically correlated random processes with continuous time. *Teor. Verojatnost. i Primenen.*, 8:184–189, 1963.

C. Gorrostieta, H. Ombao, and R. von Sachs. Time-dependent dual-frequency coherence in multivariate non-stationary time series. *J. Time Series Anal.*, 40(1):3–22, 2019.

H. Hindberg and S. C. Olhede. Estimation of ambiguity functions with limited spread. *IEEE Transactions on Signal Processing*, 58:2383–2388, 2010.

O. Jensen and L. Colgin. Cross-frequency coupling between neuronal oscillations. *Trends in Cognitive Sciences*, 11:267–269, 2007.

C. Jentsch and S. Subba Rao. A test for second order stationarity of a multivariate time series. *J. Econometrics*, 185, 2015.

A. Jung, G. Hannak, and N. Goertz. Graphical lasso based model selection for time series. *IEEE Signal Processing Letters*, 22(10):1781–1785, 2015.

T. Kley, P. Preuss, and P. Fryzlewicz. Predictive, finite-sample model choice for time series under stationarity and nonstationarity. *Electron. J. Stat.*, 13:3710–3774, 2019.

J. Krampe and S. Subba Rao. Inverse covariance operators of nonstationary multivariate time series. *https://arxiv.org/abs/2202.00933*, 2022.

H.P. Künsch. A note on causal solutions for locally stationary AR-processes. Technical report, ETH, 1995.

K-S. Lii and M. Rosenblatt. Spectral analysis for harmonizable processes. *Ann. Statist.*, 30(1):258–297, 2002.

R. Lund, H. Hurd, P. Bloomfield, and R. Smith. Climatological time series with periodic correlation. *Journal of Climate*, 11:2787–2809, 1995.

D.J. Lurie, D. Kessler, D.S Bassett, R.F. Betzel, M. Breakspear, S. Kheilholz, A. Kucyi, R. Liégeois, M.A. Lindquist, and A. R. McIntosh. Questions and controversies in the study of time-varying functional connectivity in resting fmri. *Network Neuroscience*, 4 (1):30–69, 2020.

A. Maleki, L. Anitori, Z. Yang, and R. G. Baraniuk. Asymptotic analysis of complex lasso via complex approximate message passing (camp). *IEEE Trans. Inf. Theory*, 59 (7):4290–4308, 2013.

W. Martin and P. Flandrin. Wigner-Ville spectral analysis of nonstationary processes. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33:1461–1470, 1985.

N. Meinshausen and P. Bühlmann. High-dimensional graphs and variable selection with the lasso. *Ann. Statist.*, 34(3):1436–1462, 2006.

M. Meyer, C. Jentsch, and J.-P. Kreiss. Baxter's inequality and sieve bootstrap for random fields. *Bernoulli*, 23(4B):2988–3020, 2017.

G.P. Nason. A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series. *J. Roy. Statist. Soc. (B)*, 75, 2013.

S. C. Olhede. Ambiguity sparse processes. *arXiv preprint arXiv:1103.3932v2.pdf*, 2011.

S. C. Olhede and H. Ombao. Covariance of replicated modulated cyclical time series. *IEEE Transactions on Signal Processing*, 61:1944–1957, 2013.

H. Ombao and M. Pinto. Spectral dependence. *https://arxiv.org/abs/2103.17240*, 2021.

E. Paparoditis. Testing temporal constancy of the spectral structure of a time series. *Bernoulli*, 15:1190–1221, 2009.

M. G. Preti, T. AW Bolton, and D. Van De Ville. The dynamic functional connectome: State-of-the-art and perspectives. *Neuroimage*, 160:41–54, 2017.

M. B. Priestley. Evolutionary spectra and non-stationary processes. *J. Roy. Statist. Soc. (B)*, 27:204–37, 1965.

M. B. Priestley. *Spectral analysis and time series. Vol. 2.* Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London-New York, 1981. Multivariate series, prediction and control, Probability and Mathematical Statistics.

M. B. Priestley and T. Subba Rao. A test for non-stationarity of a time series. *J. Roy. Statist. Soc. (B)*, 31:140–149, 1969.

H. Qiu, F. Han, H. Liu, and B. Caffo. Joint estimation of multiple graphical models from high dimensional time series. *J. R. Statist. Soc. (B)*, 78(2):487, 2016.

Alexis Rosuel, Philippe Loubaton, and Pascal Vallet. On the asymptotic distribution of the maximum sample spectral coherence of gaussian time series in the high dimensional regime. *arXiv preprint arXiv:2107.02891*, 2021.

A. Safikhani and A. Shojaie. Joint structural break detection and parameter estimation in high-dimensional non-stationary var models. *J. Am. Stat. Assoc.*, (just-accepted): 1–26, 2020.

S. Subba Rao. On some nonstationary, nonlinear random processes and their stationary approximations. *Adv. Appl. Probab.*, 38(4):1155–1172, 2006.

T. Subba Rao. The fitting of non-stationary time-series models with time-dependent parameters. *J. Roy. Statist. Soc.* B, **32**:312–22, 1970.

Y. Sun, Y. Li, A. Kuceyeski, and S. Basu. Large spectral density matrix estimation by thresholding. *arXiv preprint arXiv:1812.00532*, 2018.

R. R. Sundararajan and M. Pourahmadi. Stationary subspace analysis of nonstationary processes. *J. Time Series Anal.*, pages 338–355, 2018.

R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. (B)*, 58(1):267–288, 1996.

O. Toeplitz. Zur Theorie der quadratischen und bilinearen Formen von unendlichvielen Veränderlichen. *Math. Ann.*, 70(3):351–376, 1911.

Christiane Tretter. *Spectral theory of block operator matrices and applications*. Imperial College Press, 2008.

D. Wang, Y. Yu, A. Rinaldo, and R. Willett. Localizing changes in high-dimensional vector autoregressive processes. *arXiv preprint arXiv:1909.06359*, 2019.

Michael B Woodroofe and John W Van Ness. The maximum deviation of sample spectral densities. *The Annals of Mathematical Statistics*, pages 1558–1569, 1967.

Wei Biao Wu and Paolo Zaffaroni. Asymptotic theory for spectral density estimates of general multivariate time series. *Econometric Theory*, 34(1):1–22, 2018.

D. Zhang and W. B. Wu. Gaussian approximation for high dimensional time series. *Ann. Statist.*, 45:1895–1919, 2017.

D. Zhang and W. B. Wu. Second-order asymptotics for high dimensional locally stationary processes. *Ann. Statist.*, 49:233–254, 2021.

Z. Zhou and W. B. Wu. Local linear quantile estimation for nonstationary time series. *Ann. Statist.*, 37:2696–2729, 2009.

A Proofs for Section 2

A.1 Proofs of results in Section 2.2

We first show that $\sigma_{a,t}^2 \geq \lambda_{\inf}$. This ensures that $\sigma_{a,t}^2 > 0$ and the operator D is well defined (see (4)). We recall that

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

where $\beta_{(t,a)\to(t,a)} = 0$. For all (τ, b) except (t, a) let $v_{(\tau,b)\to(t,a)} = -\beta_{(\tau,b)\to(t,a)}$ and let $v_{(t,a)\to(t,a)} = 1$. For every $b \in \{1, \dots, p\}$ define $v^{(b)} = (v_{(\tau,b)\to(t,a)}; \tau \in \mathbb{Z})$ and $v = \mathrm{vec}[v^{(1)}, \dots, v^{(p)}]$. It is easily seen that

$$\sigma_{a,t}^2 = \mathrm{E}\left[X_t^{(a)} - P_{\mathcal{H}-X_t^{(a)}}(X_t^{(a)})\right]^2 = \langle v, Cv\rangle. \tag{47}$$

Since $\|v\|_2 \geq 1$, by Assumption 2.1 we have

$$\sigma_{a,t}^2 = \langle v, Cv\rangle \geq \lambda_{\inf}. \tag{48}$$

We use this result and the notation above to prove Lemma 2.1.

PROOF of Lemma 2.1 For $1 \leq b \leq p$ we define the column vectors $X^{(b)} = (\dots, X_{-1}^{(b)}, X_0^{(b)}, X_1^{(b)}, \dots)'$ and $X = \mathrm{vec}[X^{(1)}, \dots, X^{(p)}]$. Using the notation introduced at the start of this section we have

$$X_t^{(a)} - P_{\mathcal{H}-X_t^{(a)}}(X_t^{(a)}) = X_t^{(a)} - \sum_{\tau\in\mathbb{Z}}\sum_{b=1}^{p} \beta_{(\tau,b)\to(t,a)} X_\tau^{(b)} = \langle v, X\rangle.$$

Since $P_{\mathcal{H}-X_t^{(a)}}(X_t^{(a)})$ minimises the mean squared error $\mathrm{E}[X_t^{(a)} - Z]^2$ over all $Z \in \mathcal{H} - X_t^{(a)}$ and $\mathrm{E}[X_t^{(a)} - P_{\mathcal{H}-X_t^{(a)}}(X_t^{(a)})]^2 = \sigma_{a,t}^2$, this gives rise to the normal equations

$$\mathbb{C}\mathrm{ov}[\langle v, X\rangle, X_s^{(c)}] = \begin{cases} 0 & (c,s) \neq (a,t) \\ \sigma_{a,t}^2 & (c,s) = (a,t) \end{cases}.$$

Comparing the above with D, we observe that this proves $DC = I$ and $CD = I$, thus $D = C^{-1}$. To prove that $\|D\| = \lambda_{\inf}^{-1}$, we note that under Assumption 2.1, $0 < \lambda_{\inf} = \inf_{\|v\|_2=1, v\in\ell_{2,p}}\langle v, Cv\rangle \leq \sup_{\|v\|_2=1, v\in\ell_{2,p}}\langle v, Cv\rangle = \lambda_{\sup} < \infty$. Since C is a self-adjoint operator, $\|C\| = \lambda_{\sup}$ and $\|D\| = \lambda_{\inf}^{-1}$.

To prove that $\|D_{a,b}\| \leq \lambda_{\inf}^{-1}$ we first focus on the case $a = b$. Since $D_{a,a}$ are submatrices on the diagonal of D and $0 < \lambda_{\sup}^{-1} = \inf_{\|v\|_2=1, v\in\ell_{2,p}}\langle v, Dv\rangle \leq \sup_{\|v\|_2=1, v\in\ell_{2,p}}\langle v, Dv\rangle = \lambda_{\inf}^{-1} < \infty$, then it immediately follows that
$\lambda_{\sup}^{-1} \leq \inf_{\|v\|_2=1, v\in\ell_2}\langle v, D_{a,a}v\rangle \leq \sup_{\|v\|_2=1, v\in\ell_2}\langle v, D_{a,a}v\rangle = \lambda_{\inf}^{-1}$. Therefore, since $D_{a,a}$ is self-adjoint (symmetric) we have $\|D_{a,a}\| \leq \lambda_{\inf}^{-1}$. By a similar argument $\|D_{a,a}^{-1}\| \leq \lambda_{\sup}$.

To prove the result for $a \neq b$ we focus on the sub-matrix

$$D_{\{a,b\}} = \left(\begin{array}{cc} D_{a,a} & D^*_{a,b} \\ D_{a,b} & D_{b,b} \end{array} \right).$$

Using the same argument to prove that $\|D_{a,a}\| \leq \lambda_{\text{inf}}^{-1}$ it can be shown that $\|D_{\{a,b\}}\| \leq \lambda_{\text{inf}}^{-1}$. Thus for all $v' = (u^{(1)}, u^{(2)})' \in \ell_2^2$ we have $\|D_{\{a,b\}}v\| \leq \lambda_{\text{inf}}^{-1}\|v\|_2$. We use this bound below. We recall that an operator (matrix) B is bounded if there exists a finite constant K where for all $u \in \ell_2$ we have $\|Bu\| \leq K\|u\|$, it follows that $\|B\| \leq K$. Returning to $D_{a,b}$, we will show that $\|D_{a,b}u^{(1)}\| \leq \lambda_{\text{inf}}^{-1}\|u^{(1)}\|_2$. For all $u^{(1)} \in \ell_2$ we have

$$\|D_{a,b}u^{(1)}\|_2 \leq \sqrt{\|D_{a,a}u^{(1)}\|_2^2 + \|D_{a,b}u^{(1)}\|_2^2} = \|D_{\{a,b\}}v\|_2 \leq \|D_{\{a,b\}}\|\|v\|_2 = \|D_{\{a,b\}}\|\|u^{(1)}\|_2$$

where $v' = (u^{(1)}, 0)$. Thus $\|D_{a,b}\| \leq \lambda_{\text{inf}}^{-1}$, as required.

Finally, to prove that $\sup_t \sum_{\tau\in\mathbb{Z}} \|\mathbf{D}_{t,\tau}\|_2^2 \leq p\lambda_{\text{inf}}^{-2}$, we first prove that for every $t_0 \in \mathbb{Z}$, we have $\sum_{a=1}^{p}\sum_{\tau\in\mathbb{Z}}[\mathbf{D}_{\tau,t_0}]_{a,1}^2 \leq \lambda_{\text{inf}}^{-2}$. Define the sequence $v = \text{vec}[u^{(1)}, u^{(2)}, \ldots, u^{(p)}] \in \ell_{2,p}$ where we set $[u^{(1)}]_{t_0} = 1$ and $[u^{(1)}]_s = 0$ for $s \neq t_0$ and $u^{(a)} = 0$ (zero sequence) for all $a \neq 1$. Then by definition of v (which mainly consists of zeros except for one non-zero entry) we have

$$Dv = \left(\begin{array}{c} D_{1,1}u^{(1)} \\ \vdots \\ D_{p,1}u^{(1)} \end{array} \right) = \left(\begin{array}{c} ([D_{1,1}]_{\tau,t_0}; \tau \in \mathbb{Z}) \\ \vdots \\ ([D_{p,1}]_{\tau,t_0}; \tau \in \mathbb{Z}) \end{array} \right).$$

Thus for every $t_0 \in \mathbb{Z}$ we have $\|Dv\|_2^2 = \sum_{a=1}^{p}\sum_{\tau\in\mathbb{Z}}[D_{a,1}]_{\tau,t_0}^2 = \sum_{a=1}^{p}\sum_{\tau\in\mathbb{Z}}[\mathbf{D}_{\tau,t_0}]_{a,1}^2 \leq \|D\|^2\|v\|_2^2 \leq \lambda_{\text{inf}}^{-2}$. By the same argument for any $b \in \{1, \ldots, p\}$ and $t \in \mathbb{Z}$ we have $\sum_{a=1}^{p}\sum_{\tau\in\mathbb{Z}}[\mathbf{D}_{\tau,t}]_{a,b}^2 \leq \|D\|^2 \leq \lambda_{\text{inf}}^{-2}$, this gives $\sum_{\tau\in\mathbb{Z}}\|\mathbf{D}_{\tau,t}\|_2^2 \leq p\lambda_{\text{inf}}^{-2}$. This proves the claim. $\square$

Many of the results in this section use the block operator inversion identity (see Tretter (2008), page 35, and Berkolaiko and Kuchment (2020), Section 2.3)). For completeness we give the identity below. As we are working with covariance matrix operators we focus on symmetric/self-adjoint matrices. Suppose

$$G = \left(\begin{array}{cc} A & B \\ B^* & C \end{array} \right),$$

and G^{-1} exists, then

$$G^{-1} = \left(\begin{array}{cc} P^{-1} & -P^{-1}BC^{-1} \\ -C^{-1}B^*P^{-1} & (C - B^*A^{-1}B)^{-1} \end{array} \right) \tag{49}$$

where $P = A - BC^{-1}B^*$. We mention that P is the Schur complement of C of the matrix G.

A.2 Proof of results in Section 2.4

There are different methods for proving the results in Section 2.4. One method is to use the properties of projections the other is to use decompositions of the infinite dimensional matrix C. In this section we take the matrix decomposition route, as similar matrix decompositions form the core of the proofs in Section 2.5.

As the results in Section 2.4 concern the partial covariance between $X_t^{(a)}$ and $X_\tau^{(b)}$ given *all* the other random variables, we will consider a permuted version of C and its inverse, where we bring the covariance structure of $(X_t^{(a)}, X_\tau^{(b)})$ to the top left hand corner of the matrix. To avoid introducing new notation we label these permuted matrix operators as C and D. The variance of $(X_t^{(a)}, X_\tau^{(b)})$ is

$$\mathbb{V}\mathrm{ar}\begin{bmatrix} X_t^{(a)} \\ X_\tau^{(b)} \end{bmatrix} = \widetilde{C}_{1,1} = \begin{pmatrix} [C_{a,a}]_{t,t} & [C_{a,b}]_{t,\tau} \\ [C_{a,b}]_{t,\tau} & [C_{b,b}]_{\tau,\tau} \end{pmatrix}.$$

This embeds in the top left hand side of the operator C, where

$$C = \begin{pmatrix} \widetilde{C}_{1,1} & \widetilde{C}_{1,2} \\ \widetilde{C}_{2,1} & \widetilde{C}_{2,2} \end{pmatrix} \tag{50}$$

with $\widetilde{C}_{1,2} = \{[C_{c,e}]_{u,v}; (c,u) \in \{(a,t),(b,\tau)\}, (e,v) \notin \{(a,t),(b,\tau)\}\}$, $\widetilde{C}_{2,1} = \widetilde{C}_{1,2}^*$ and $\widetilde{C}_{2,2} = \{[C_{c,e}]_{u,v}; (c,u),(e,v) \notin \{(a,t),(b,\tau)\}\}$ (we have used the tilde notation in $\widetilde{C}_{i,j}$ to distinguish it from $C_{a,b}$). It is well known that the Schur complement encodes the partial covariance. Applying this to $(X_t^{(a)}, X_\tau^{(b)})$, the Schur complement of $\widetilde{C}_{1,1}$ in C is

$$\mathbb{V}\mathrm{ar}\left[\begin{pmatrix} X_t^{(a)} - P_{\mathcal{H}-(X_t^{(a)},X_\tau^{(b)})}(X_t^{(a)}) \\ X_\tau^{(b)} - P_{\mathcal{H}-(X_t^{(a)},X_\tau^{(b)})}(X_\tau^{(b)}) \end{pmatrix}\right] = \widetilde{C}_{1,1} - \widetilde{C}_{1,2}\widetilde{C}_{2,2}^{-1}\widetilde{C}_{2,1} = P. \tag{51}$$

The above matrix (which we label as P) forms an important part of all the proofs in this section. As the entries in the variance matrix on the left hand side of (51) is quite long we replace it with some shorter notation for the conditional variances and covariances. Comparing (51) with (6) we observe that the off-diagonal is $\rho_{t,\tau}^{(a,b)} = \mathbb{C}\mathrm{ov}[X_t^{(a)} - P_{\mathcal{H}-(X_t^{(a)},X_\tau^{(b)})}(X_t^{(a)}), X_\tau^{(b)} - P_{\mathcal{H}-(X_t^{(a)},X_\tau^{(b)})}(X_\tau^{(b)})]$. However, the diagonal of P has not been defined in Section 2.4. As this is the partial variance of $X_t^{(a)}$ after conditioning on everything *but* $X_t^{(a)}$ and $X_\tau^{(b)}$ we use the notation

$$\rho_{t,t}^{(a,a)|\text{-}\{(a,t),(b,\tau)\}} = \mathbb{V}\mathrm{ar}[X_t^{(a)} - P_{\mathcal{H}-(X_t^{(a)},X_\tau^{(b)})}(X_t^{(a)})]. \tag{52}$$

To avoid confusion, we mention that this is different to the time series partial covariance defined in Section 2.5, where $\rho_{t,t}^{(a,a)|\text{-}\{a,b\}}$ is the partial variance of $X_t^{(a)}$ after conditioning on all the other time series but time series $X^{(a)}$ and $X^{(b)}$. Using the new notation we have

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left(\begin{array}{c} X_t^{(a)} - P_{\mathcal{H}-(X_t^{(a)},X_\tau^{(b)})}(X_t^{(a)}) \\ X_\tau^{(b)} - P_{\mathcal{H}-(X_t^{(a)},X_\tau^{(b)})}(X_\tau^{(b)}) \end{array}\right) &= \left(\begin{array}{cc} \rho_{t,t}^{(a,a)|\text{-}\{(a,t),(b,\tau)\}} & \rho_{t,\tau}^{(a,b)} \\ \rho_{\tau,t}^{(b,a)} & \rho_{\tau,\tau}^{(b,b)|\text{-}\{(a,t),(b,\tau)\}} \end{array}\right) \\ &= \widetilde{C}_{1,1} - \widetilde{C}_{1,2}\widetilde{C}_{2,2}^{-1}\widetilde{C}_{2,1} = P, \qquad (53)\end{aligned}$$

where $\rho_{t,\tau}^{(a,b)}$ and $\rho_{t,t}^{(a,a)|\text{-}\{(a,t),(b,\tau)\}}$ are defined in (6) and (52) respectively.

Next we relate P to the inverse C^{-1}. Using the block operator inversion (see (49)) we have

$$D = \left(\begin{array}{cc} P^{-1} & -P^{-1}\widetilde{C}_{1,2}\widetilde{C}_{2,2}^{-1} \\ -\widetilde{C}_{2,2}^{-1}\widetilde{C}_{2,1}P^{-1} & (\widetilde{C}_{2,2} - \widetilde{C}_{2,1}\widetilde{C}_{1,1}^{-1}\widetilde{C}_{1,2})^{-1} \end{array}\right) \qquad (54)$$

where P is defined in (51). Comparing P^{-1} with the upper left block of D, we connect the conditional variance of $(X_t^{(a)}, X_\tau^{(b)})$ to the entries of D. In particular

$$\left(\begin{array}{cc} \rho_{t,t}^{(a,a)|\text{-}\{(a,t),(b,\tau)\}} & \rho_{t,\tau}^{(a,b)} \\ \rho_{\tau,t}^{(b,a)} & \rho_{\tau,\tau}^{(b,b)|\text{-}\{(a,t),(b,\tau)\}} \end{array}\right) = \left(\begin{array}{cc} [D_{a,a}]_{t,t} & [D_{a,b}]_{t,\tau} \\ [D_{a,b}]_{t,\tau} & [D_{b,b}]_{\tau,\tau} \end{array}\right)^{-1}. \qquad (55)$$

The identity (55) forms an important component in the proofs below.

Before we state the next lemma, we require the following notation for the partial correlation between $X_t^{(a)}$ and $X_\tau^{(b)}$

$$\phi_{t,\tau}^{(a,b)} = \mathbb{C}\mathrm{orr}\left[X_t^{(a)} - P_{\mathcal{H}-(X_t^{(a)},X_\tau^{(b)})}(X_t^{(a)}), X_t^{(b)} - P_{\mathcal{H}-(X_t^{(a)},X_\tau^{(b)})}(X_\tau^{(b)})\right]. \qquad (56)$$

Lemma A.1 *Let $\beta_{(\tau,b)\to(t,a)}$, $\sigma_{t,a}^2$, $\rho_{t,t}^{(a,a)|\text{-}\{(a,t),(b,\tau)\}}$ and $\phi_{t,\tau}^{(a,b)}$ be defined as in (3), (52) and (56). Suppose Assumption 2.1 holds. Then*

$$\frac{\rho_{t,t}^{(a,a)|\text{-}\{(a,t),(b,\tau)\}}}{\rho_{\tau,\tau}^{(b,b)|\text{-}\{(a,t),(b,\tau)\}}} = \frac{\sigma_{t,a}^2}{\sigma_{\tau,b}^2} \qquad (57)$$

and

$$\beta_{(\tau,b)\to(t,a)} = \phi_{t,\tau}^{(a,b)} \times \frac{\sigma_{a,t}}{\sigma_{b,\tau}}. \qquad (58)$$

PROOF. The proof of (57) is based on comparing $[D_{a,a}]_{t,t}/[D_{b,b}]_{\tau,\tau}$ and the ratio of the diagonal entries of the conditional variance in (55)

$$\begin{aligned}
&\begin{pmatrix} \rho_{t,t}^{(a,a)|\text{-}\{(a,t),(b,\tau)\}} & \rho_{t,\tau}^{(a,b)} \\ \rho_{\tau,t}^{(b,a)} & \rho_{\tau,\tau}^{(b,b)|\text{-}\{(a,t),(b,\tau)\}} \end{pmatrix} \\
= \ & \frac{1}{[D_{a,a}]_{t,t}[D_{b,b}]_{\tau,\tau} - [D_{a,b}]_{t,\tau}^2} \begin{pmatrix} [D_{b,b}]_{\tau,\tau} & -[D_{a,b}]_{t,\tau} \\ -[D_{a,b}]_{t,\tau} & [D_{a,a}]_{t,t} \end{pmatrix}. \qquad (59)
\end{aligned}$$

We recall from (4) that

$$[D_{a,a}]_{t,t} = \frac{1}{\sigma_{a,t}^2} \text{ and } [D_{b,b}]_{\tau,\tau} = \frac{1}{\sigma_{b,\tau}^2} \quad \Rightarrow \frac{\sigma_{a,t}^2}{\sigma_{b,\tau}^2} = \frac{[D_{b,b}]_{\tau,\tau}}{[D_{a,a}]_{t,t}}. \qquad (60)$$

Furthermore, by comparing the entries in (59) we have

$$\begin{aligned}
\rho_{t,t}^{(a,a)|\text{-}\{(a,t),(b,\tau)\}} &= \frac{1}{[D_{a,a}]_{t,t}[D_{b,b}]_{\tau,\tau} - [D_{a,b}]_{t,\tau}^2}[D_{b,b}]_{\tau,\tau} \\
\text{and } \rho_{\tau,\tau}^{(b,b)|\text{-}\{(a,t),(b,\tau)\}} &= \frac{1}{[D_{a,a}]_{t,t}[D_{b,b}]_{\tau,\tau} - [D_{a,b}]_{t,\tau}^2}[D_{a,a}]_{t,t}.
\end{aligned}$$

Thus evaluating ratio of the above gives

$$\frac{\rho_{t,t}^{(a,a)|\text{-}\{(a,t),(b,\tau)\}}}{\rho_{\tau,\tau}^{(b,b)|\text{-}\{(a,t),(b,\tau)\}}} = \frac{[D_{b,b}]_{\tau,\tau}}{[D_{a,a}]_{t,t}}. \qquad (61)$$

Comparing (60) and (61) gives (57).

To prove (58) we decompose the projection of $P_{\mathcal{H}-X_t^{(a)}}(X_t^{(a)})$ in terms of its projections onto the two space $\mathcal{H} - (X_t^{(a)}, X_\tau^{(b)})$ and $\overline{\text{sp}}(X_\tau^{(b)} - P_{\mathcal{H}-(X_t^{(a)},X_\tau^{(b)})}(X_\tau^{(b)}))$. These two spaces are orthogonal and lead to a simple expression for the coefficient $\beta_{(\tau,b)\to(t,a)}$;

$$\begin{aligned}
P_{\mathcal{H}-X_t^{(a)}}(X_t^{(a)}) &= \sum_{\tau\in\mathbb{Z}}\sum_{b=1}^{p} \beta_{(\tau,b)\to(t,a)} X_\tau^{(b)} \\
&= P_{\mathcal{H}-(X_t^{(a)},X_\tau^{(b)})}(X_t^{(a)}) + \beta_{(\tau,b)\to(t,a)}\left[X_\tau^{(b)} - P_{\mathcal{H}-(X_t^{(a)},X_\tau^{(b)})}(X_\tau^{(b)})\right]. \qquad (62)
\end{aligned}$$

Using the orthogonality of the two projections we have

$$\beta_{(\tau,b)\to(t,a)} = \frac{\mathbb{C}\text{ov}\left[X_t^{(a)}, X_\tau^{(b)} - P_{\mathcal{H}-(X_t^{(a)},X_\tau^{(b)})}(X_\tau^{(b)})\right]}{\mathbb{V}\text{ar}(X_\tau^{(b)} - P_{\mathcal{H}-(X_t^{(a)},X_\tau^{(b)})}(X_\tau^{(b)}))} = \frac{\rho_{t,\tau}^{(a,b)}}{\rho_{t,t}^{(a,a)|\text{-}\{(a,t),(b,\tau)\}}}.$$

Replacing the covariance $\rho_{t,\tau}^{(a,b)}$ in $\beta_{(\tau,b)\to(t,a)}$ with its correlation $\phi_{t,\tau}^{(a,b)}$ gives

$$\beta_{(\tau,b)\to(t,a)} \quad = \quad \phi_{t,\tau}^{(a,b)}\sqrt{\frac{\rho_{t,t}^{(a,a)|-\{(a,t),(b,\tau)\}}}{\rho_{\tau,\tau}^{(b,b)|-\{(a,t),(b,\tau)\}}}}. \tag{63}$$

This links the partial correlation to the projection coefficients. Finally, we substitute the identity (57) into (63) to give

$$\beta_{(\tau,b)\to(t,a)} \quad = \quad \phi_{t,\tau}^{(a,b)} \times \frac{\sigma_{a,t}}{\sigma_{b,\tau}}.$$

This proves (58). $\square$

We use the above to prove Lemma 2.2.
PROOF of Lemma 2.2 By using (58) we connect $\phi_{t,\tau}^{(a,b)}$ to the precision matrix. Since

$$\beta_{(\tau,b)\to(t,a)} \quad = \quad \phi_{t,\tau}^{(a,b)} \times \frac{\sigma_{a,t}}{\sigma_{\tau,b}}$$

and by definition of $D_{a,b}$ in (5) we have

$$\rho_{t,\tau}^{(a,b)} = -\frac{[D_{a,b}]_{t,\tau}}{\sqrt{[D_{a,a}]_{t,t}[D_{b,b}]_{\tau,\tau}}}.$$

This proves (8). The proof of (9) immediately follows from (55). $\square$

PROOF of Proposition 2.1 The proof hinges on the identity in (55) for the separate cases $a = b$ and $a \neq b$. For the case $a \neq b$ and using (55) it is clear that $\rho_{t,\tau}^{(a,b)} = 0$ iff $[D_{a,b}]_{t,\tau} = 0$. Thus $D_{a,b} = 0$ iff for all t and τ, $\rho_{t,\tau}^{(a,b)} = 0$, this proves (i).

To prove (ii), we use (55) with $a = b$ and compare the entries of

$$\begin{aligned}&\begin{pmatrix} \rho_{t,t}^{(a,a)|-\{(a,t),(a,\tau)\}} & \rho_{t,\tau}^{(a,a)} \\ \rho_{\tau,t}^{(a,a)} & \rho_{\tau,\tau}^{(a,a)|-\{(a,t),(a,\tau)\}} \end{pmatrix} \\ = \quad & \frac{1}{[D_{a,a}]_{t,t}[D_{a,a}]_{\tau,\tau} - [D_{a,a}]_{t,\tau}^2} \begin{pmatrix} [D_{a,a}]_{\tau,\tau} & -[D_{a,a}]_{t,\tau} \\ -[D_{a,a}]_{t,\tau} & [D_{a,a}]_{t,t} \end{pmatrix}.\end{aligned} \tag{64}$$

We first show that if $D_{a,a}$ is a symmetric, Toeplitz matrix, then $\rho_{t,\tau}^{(a,a)}$ is shift invariant (depends only on $t - \tau$). If $D_{a,a}$ is a symmetric, Toeplitz matrix, using that $\rho_{t,t}^{(a,a)} = 1/[D_{a,a}]_{t,t} = 1/[D_{a,a}]_{0,0}$ it is clear that $\rho_{t,t}^{(a,a)}$ does not depend on t. We now study $\rho_{t,\tau}^{(a,a)}$ when $t \neq \tau$. To show that $\rho_{t,\tau}^{(a,a)}$ only depends on $|t-\tau|$ we use that $[D_{a,a}]_{t,\tau} = [D_{a,a}]_{0,\tau-t} = [D_{a,a}]_{0,t-\tau}$ (due to $D_{a,a}$ being Toeplitz). Comparing the off-diagonal entries on the left and right hand side of (64) it follows that for all t and τ $\rho_{t,\tau}^{(a,a)} = \rho_{0,t-\tau}^{(a,a)} = \rho_{0,\tau-t}^{(a,a)}$.

Next we show the converse, that is if for all t and τ; $\rho_{t,\tau}^{(a,a)} = \rho_{0,t-\tau}^{(a,a)} = \rho_{0,\tau-t}^{(a,a)}$ then $D_{a,a}$ is a symmetric, Toeplitz matrix. First the diagonal, since $[D_{a,a}]_{t,t} = 1/\rho_{t,t}^{(a,a)} = 1/\rho_{0,0}^{(a,a)}$ it is clear that the diagonal $D_{a,a}$ does not depend on t. Next we show that if for all t and

τ; $\rho^{(a,a)}_{t,\tau} = \rho^{(a,a)}_{0,t-\tau} = \rho^{(a,a)}_{0,\tau-t}$, then

(a) $[D_{a,a}]_{t,\tau}$ only depends on $|t-\tau|$.

(b) The conditional variance $\rho^{(a,a)|\text{-}\{(a,t),(a,\tau)\}}_{t,t}$ only depends on $|t-\tau|$. Note that this is not in the statement of the theorem, but is a useful by product of the proof.

Comparing the entries of the matrices in (64) we have

$$\begin{aligned} \rho^{(a,a)|\text{-}\{(a,t),(a,\tau)\}}_{t,t} &= \frac{[D_{a,a}]_{\tau,\tau}}{[D_{a,a}]_{t,t}[D_{a,a}]_{\tau,\tau} - [D_{a,a}]^2_{t,\tau}}, && (65)\\ \rho^{(a,a)}_{t,\tau} &= \frac{-[D_{a,a}]_{t,\tau}}{[D_{a,a}]_{t,t}[D_{a,a}]_{\tau,\tau} - [D_{a,a}]^2_{t,\tau}}, && (66)\\ \text{and } 0 &< [D_{a,a}]_{t,t}[D_{a,a}]_{\tau,\tau} - [D_{a,a}]^2_{t,\tau} \text{ (since this is the determinant).} && (67) \end{aligned}$$

We first show that $[D_{a,a}]_{t,\tau}$ only depends on $|t-\tau|$. To reduce notation, we set the entries on the diagonal of $D_{a,a}$ to $\theta = [D_{a,a}]_{t,t}$ and let $\theta_{t-\tau} = \rho^{(a,a)}_{t,\tau}$. Substituting this into (66) gives

$$\theta_{t-\tau}(\theta^2 - [D_{a,a}]^2_{t,\tau}) \;=\; -[D_{a,a}]_{t,\tau}.$$

The above is quadratic equation in $[D_{a,a}]_{t,\tau}$. Thus we can express $[D_{a,a}]_{t,\tau}$ in terms of θ and $\theta_{t-\tau}$;

$$[D_{a,a}]_{t,\tau} = \frac{-1+\sqrt{1+4\theta^2_{t-\tau}\theta^2}}{2\theta_{t-\tau}}.$$

Note that $-1+\sqrt{1+4\theta^2_{t-\tau}\theta^2}$ is part of the solution and not $-1-\sqrt{1+4\theta^2_{t-\tau}\theta^2}$ due to the positivity condition in (67). This proves that $D_{a,a}$ is a symmetric, Toeplitz matrix. This proves (a) and (ii) in the lemma. To prove (b), we use (65) and observe that the right hand side depends only on $|t-\tau|$, thus proving that $\rho^{(a,a)|\text{-}\{(a,t),(a,\tau)\}}_{t,t}$ only depends on $|t-\tau|$.

To prove (iii) we use (55) with $a \neq b$. From (55) it immediately follows that if $D_{a,a}$, $D_{b,b}$ and $D_{a,b}$ are Toeplitz, then $\rho^{(a,b)}_{t,\tau}$ only depends on the lag $(t-\tau)$.

Conversely, to prove that $D_{a,b}$ is Toeplitz given that for all t and τ; $\rho^{(a,a)}_{t,\tau} = \rho^{(a,a)}_{0,t-\tau} = \rho^{(a,a)}_{0,\tau-t}$. $\rho^{(b,b)}_{t,\tau} = \rho^{(b,b)}_{0,t-\tau} = \rho^{(b,b)}_{0,\tau-t}$, $\rho^{(a,b)}_{t,\tau} = \rho^{(a,b)}_{0,t-\tau}$, we use the same strategy used to prove (ii). This yields the solution

$$[D_{a,b}]_{t,\tau} \;=\; \frac{-1+\sqrt{1+4(\rho^{(a,b)}_{0,t-\tau})^2\sigma_a^{-2}\sigma_b^{-2}}}{2\rho^{(a,b)}_{0,t-\tau}},$$

which proves that $D_{a,b}$ is Toeplitz. Thus proving the result. $\square$

A.3 Proof of results in Section 2.5

To prove the results in Section 2.5 we follow a similar strategy to the proofs of Section 2.4, but permute the submatrices $\{C_{a,b}\}$ in C rather than the individual entries in C. The proofs in this section are less technical than those in Section 2.4.

Define the two non-intersecting sets $\mathcal{S} = \{\alpha_1, \ldots, \alpha_r\}$ and its complement $\mathcal{S}' = \{\beta_1, \ldots, \beta_s\}$ where $\mathcal{S} \cup \mathcal{S}' = \{1, 2, \ldots, p\}$. We now obtain an expression for the covariance of $\{X_t^{(c)}; c \in \mathcal{S}\}$ after removing their linear dependence on $\{X_s^{(c)}; s \in \mathbb{Z}, c \in \mathcal{S}'\}$. To do so, we define the submatrix $C_{\mathcal{S},\mathcal{S}} = (C_{a_1,b_1}; a_1, b_1 \in \mathcal{S})$ where we note that

$$\mathbb{V}\mathrm{ar}[X_t^{(a)}; t \in \mathbb{Z}, a \in \mathcal{S}] = C_{\mathcal{S},\mathcal{S}}.$$

A block permuted version of C with $C_{\mathcal{S},\mathcal{S}}$ in the top left hand corner is

$$C = \begin{pmatrix} C_{\mathcal{S},\mathcal{S}} & C_{\mathcal{S},\mathcal{S}'} \\ C^*_{\mathcal{S},\mathcal{S}'} & C_{\mathcal{S}',\mathcal{S}'} \end{pmatrix},$$

where $C_{\mathcal{S},\mathcal{S}'} = (C_{a_1,b_2}; a_1 \in \mathcal{S}$ and $b_2 \in \mathcal{S}')$, $C_{\mathcal{S}',\mathcal{S}'} = (C_{a_2,b_2}; a_2 \in \mathcal{S}'$ and $b_2 \in \mathcal{S}')$. By using standard results, the conditional variance of $(X_t^{(a)}; t \in \mathbb{Z}, a \in \mathcal{S})$ given $\{X^{(b)}; b \in \mathcal{S}'\}$ is the Schur complement of $C_{\mathcal{S}',\mathcal{S}'}$ of C:

$$\mathbb{V}\mathrm{ar}\left[X_t^{(a)} - P_{\mathcal{H}-(X^{(c)};c\in\mathcal{S}')}(X_t^{(a)}); t \in \mathbb{Z}, a \in \mathcal{S}\right] = C_{\mathcal{S},\mathcal{S}} - C_{\mathcal{S},\mathcal{S}'}C^{-1}_{\mathcal{S}',\mathcal{S}'}C_{\mathcal{S}',\mathcal{S}} = P. \tag{68}$$

Using the above, entrywise for all $a, b \in \mathcal{S}$ and $t, \tau \in \mathbb{Z}$, we have

$$\mathbb{C}\mathrm{ov}\left[X_t^{(a)} - P_{\mathcal{H}-(X^{(c)};c\in\mathcal{S}')}(X_t^{(a)}), X_\tau^{(b)} - P_{\mathcal{H}-(X^{(c)};c\in\mathcal{S}')}(X_\tau^{(b)})\right] = [P_{a,b}]_{t,\tau}.$$

We now relate the conditional variance P (defined in (68)) to the matrix D. Using the block operator inversion identity in (49) we have

$$D = \begin{pmatrix} P^{-1} & -P^{-1}C_{\mathcal{S},\mathcal{S}'}C^{-1}_{\mathcal{S}',\mathcal{S}'} \\ -C^{-1}_{\mathcal{S}',\mathcal{S}'}C_{\mathcal{S}',\mathcal{S}}P^{-1} & (C_{\mathcal{S}',\mathcal{S}'} - C_{\mathcal{S}',\mathcal{S}}C^{-1}_{\mathcal{S},\mathcal{S}}C_{\mathcal{S},\mathcal{S}'})^{-1} \end{pmatrix}, \tag{69}$$

where P is defined in (68).

We use (69) to prove the results in Section 2.5.

PROOF of Theorem 2.1 To prove the result we use (68), where we set $\mathcal{S} = \{a\}$ and $\mathcal{S} = \{a, b\}$.

To prove (i) we let $\mathcal{S} = \{a\}$. By using (68) we have

$$\mathbb{V}\mathrm{ar}\left[X_t^{(a)} - P_{\mathcal{H}-X^{(a)}}(X_t^{(a)}); t \in \mathbb{Z}\right] = D^{-1}_{a,a}. \tag{70}$$

Thus entrywise by definition we have $\rho_{t,\tau}^{(a,a)|-\{a\}} = [D^{-1}_{a,a}]_{t,\tau}$, this proves (i).

To prove (ii) we let $\mathcal{S} = \{a, b\}$. By using (68) and (49) we have

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

where the above follows from (49). Comparing entries in the above matrix gives

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

This proves (ii). □

PROOF of Theorem 2.2 Before we prove the result, we note the following invariance properties of (infinite dimension) Toeplitz operators. If A and B are bounded Toeplitz operators then (a) AB is Toeplitz (b) if A is Toeplitz and has a bounded inverse, then A^{-1} is Toeplitz; these results are a consequence of Toeplitz Theorem, Toeplitz (1911). It is important to mention that these results only hold if the Toeplitz operators are bi-infinite in the sense the entries of A are $A_{t,\tau} = A_{t-\tau}$ for all $t, \tau \in \mathbb{Z}$. The same results do not hold if the Toeplitz operators are semi-infinite where A is defined as $A_{t,\tau} = A_{t-\tau}$ for all $t, \tau \in \mathbb{Z}^+$.

We recall from the proof of Theorem 2.1 that

$$\mathbb{V}\mathrm{ar}\left[X_t^{(a)|\text{-}\{a\}}; t \in \mathbb{Z}, c \in \{a\}\right] = D_{a,a}^{-1} \tag{71}$$

and

$$\begin{aligned}&\mathbb{V}\mathrm{ar}\left[X_t^{(a)|\text{-}\{a,b\}}; t \in \mathbb{Z}, c \in \{a,b\}\right] \\ &= \begin{pmatrix} (D_{a,a} - D_{a,b}D_{b,b}^{-1}D_{b,a})^{-1} & -(D_{a,a} - D_{a,b}D_{b,b}^{-1}D_{b,a})^{-1}D_{a,b}D_{b,b}^{-1} \\ -D_{b,b}^{-1}D_{b,a}(D_{a,a} - D_{a,b}D_{b,b}^{-1}D_{b,a})^{-1} & (D_{b,b} - D_{b,a}D_{a,a}^{-1}D_{a,b})^{-1} \end{pmatrix}\end{aligned} \tag{72}$$

we use this to prove the result.

We first prove (i). If $\{X_t^{(a)|\text{-}\{a,b\}}, X_t^{(b)|\text{-}\{a,b\}}\}_t$ is conditionally noncorrelated then $D_{a,b} = D_{b,a} = 0$. From (72) we have

$$\mathbb{V}\mathrm{ar}\left[X_t^{(a)|\text{-}\{a,b\}}; t \in \mathbb{Z}, c \in \{a,b\}\right] = \begin{pmatrix} D_{a,a}^{-1} & 0 \\ 0 & D_{b,b}^{-1} \end{pmatrix}.$$

Thus $\rho_{t,\tau}^{(a,b)|\text{-}\{a,b\}} = \mathbb{C}\mathrm{ov}[X_t^{(a)|\text{-}\{a,b\}}, X_\tau^{(b)|\text{-}\{a,b\}}] = 0$ for all t and τ. Conversely, if $\rho_{t,\tau}^{(a,b)|\text{-}\{a,b\}} = \mathbb{C}\mathrm{ov}[X_t^{(a)|\text{-}\{a,b\}}, X_\tau^{(b)|\text{-}\{a,b\}}] = 0$ for all t and τ, then using (72) we have $D_{a,b} = 0$. This proves (i).

To prove (ii) we use (71). If $D_{a,a}$ is Toeplitz, then $D_{a,a}^{-1}$ is Toeplitz and $\rho_{t,\tau}^{(a,a)|\text{-}\{a\}} = [D_{a,a}^{-1}]_{t,\tau} = \rho_{0,t-\tau}^{(a,a)|\text{-}\{a\}}$ for t and τ (thus $\rho_{t,\tau}^{(a,a)|\text{-}\{a\}}$ is shift invariant). Conversely, if for all t and τ, there exists a sequence $\{\rho_{r}^{(a,a)|\text{-}\{a\}}\}_{r}$ where $\rho_{t-\tau}^{(a,a)|\text{-}\{a\}} = \rho_{t,\tau}^{(a,a)|\text{-}\{a\}}$, then since $\rho_{t,\tau}^{(a,a)|\text{-}\{a\}} = [D_{a,a}^{-1}]_{t,\tau}$ this implies $D_{a,a}^{-1}$ is Toeplitz. Thus $D_{a,a}$ is Toeplitz. This proves (ii).

To prove (iii) we use (72). If $D_{a,a}, D_{a,b}$ and $D_{b,b}$ are Toeplitz, then by the inverse properties of Toeplitz operators (described at the start of the proof) $(D_{a,a} - D_{a,b}D_{b,b}D_{ba})^{-1}$, $-(D_{a,a} - D_{a,b}D_{b,b}D_{ba})^{-1}D_{a,b}D_{b,b}^{-1}$ and $(D_{b,b} - D_{b,a}D_{a,a}D_{a,b})^{-1}$ are Toeplitz. Thus the conditional covariances $\{\rho_{t,\tau}^{(a,a)|\text{-}\{a,b\}}\}$, $\{\rho_{t,\tau}^{(a,b)|\text{-}\{a,b\}}\}$ and $\{\rho_{t,\tau}^{(a,b)|\text{-}\{a,b\}}\}$ are shift invariant. Conversely, suppose

$$\mathbb{V}\mathrm{ar}\left[X_t^{(c)|\text{-}\{a,b\}}; t \in \mathbb{Z}, c \in \{a,b\}\right] = \begin{pmatrix} E_{a,a} & E_{a,b} \\ E_{a,b}^{*} & E_{b,b} \end{pmatrix}$$

where $E_{a,a}$, $E_{a,b}$ and $E_{b,b}$ are Toeplitz. Then by using the relation

$$\begin{pmatrix} E_{a,a} & E_{a,b} \\ E_{a,b}^{*} & E_{b,b} \end{pmatrix}^{-1} = \begin{pmatrix} D_{a,a} & D_{a,b} \\ D_{a,b}^{*} & D_{b,b} \end{pmatrix},$$

and (49), we have that $D_{a,a}, D_{a,b}$ and $D_{b,b}$ are Toeplitz. This proves (iii). □

PROOF of Corollary 2.1 The result follows immediately from (68) where

$$\mathbb{V}\mathrm{ar}\left[X_t^{(a)} - P_{\mathcal{H}-(X^{(c)};c\in\mathcal{S}')}(X_t^{(a)}); t \in \mathbb{Z}, a \in \mathcal{S}\right]$$
$$= \begin{pmatrix} D_{\alpha_1,\alpha_1} & D_{\alpha_1,\alpha_2} & \dots & D_{\alpha_1,\alpha_r} \\ D_{\alpha_2,\alpha_1} & D_{\alpha_2,\alpha_2} & \dots & D_{\alpha_2,\alpha_r} \\ \vdots & \vdots & \ddots & \vdots \\ D_{\alpha_r,\alpha_1} & D_{\alpha_r,\alpha_2} & \dots & D_{\alpha_r,\alpha_r} \end{pmatrix}^{-1}.$$

Thus proving the result. □

B Proofs for Section 3

B.1 Proof of results in Section 3.1

We start by reviewing some of the relationships between the bounded matrix operator $A : \ell_2 \to \ell_2$ (where $A = (A_{t,\tau}; t, \tau \in \mathbb{Z})$) and the corresponding integral kernel of F^*AF, which is $\sum_{t\in\mathbb{Z}}\sum_{\tau\in\mathbb{Z}} A_{t,\tau}e^{it\omega - i\tau\lambda}$. We mention that if the entries of A were the covariance of a time series and $\sum_{t,\tau\in\mathbb{Z}} A_{t,\tau}^2 < \infty$, then $A(\omega,\lambda) = \sum_{t\in\mathbb{Z}}\sum_{\tau\in\mathbb{Z}} A_{t,\tau}e^{it\omega - i\tau\lambda}$ (the Loeve dual-frequency spectrum) is a well defined function in $L_2[0,2\pi)^2$ (see, for example, Gorrostieta et al. (2019) and Aston et al. (2019)).

The jth row of A can be extracted from A using $A'u_j$, where $u_j \in \ell_2$ with $u_j = (\dots, 0, 1, 0, 0, \dots)$ with 1 at the jth entry. It is clear that $A'u_j = (A_{j,\cdot})'$ (the jth row of A) and $\{A'u_j\}_{j\in\mathbb{Z}}$ reproduces all the rows of A. We now find the parallel to $A'u_j$ for

F^*AF. Since F is an isomorphism from ℓ_2 to $L_2[0,2\pi)$ the equivalent of u_j in $L_2[0,2\pi)$ is $F^*u_j = \exp(-ij\omega)$ (inverting back gives $[F\exp(-ij\cdot)]_t = [u_j]_t$, the tth entry in the vector u_j). Therefore, if $E = F^*AF$ has integral kernel $A(\omega,\lambda)$, then

$$[EF^*u_j](\lambda) = \int_0^{\pi} A(\omega,\lambda)\exp(-ij\omega)d\omega = A_j(\lambda),$$

where $A_j(\lambda) = \sum_{\tau\in\mathbb{Z}} A_{j,\tau}\exp(-i\tau\lambda) \in L_2[0,2\pi)$ and forms the building blocks of $A(\omega,\lambda)$ (since $A(\omega,\lambda) = \sum_{t\in\mathbb{Z}} A_t(\lambda)\exp(it\omega)$). $(FEF^*)u_j$ yields the jth row of the infinite dimensional matrix (FEF^*) and the $(j,s)th$ entry of $A = (FEF^*)$ is

$$[(FEF^*)u_j]_s = \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi} A(\omega,\lambda)\exp(-ij\omega)\exp(is\lambda)d\omega d\lambda. \quad (73)$$

The above gives the relationship between $A(\omega,\lambda)$ and A.

The proof of Lemma 3.1 follows from Toeplitz (1911) (see Böttcher and Grudsky (2000), Theorem 1.1). However, for completeness and to explicitly connect the result to $A(\omega,\lambda)$ we give a proof below (it is based on the discussion above).

PROOF of Lemma 3.1 We prove that the infinite dimensional Toeplitz matrix A leads to a diagonal kernel of the form $\delta_{\omega,\lambda}A(\omega)$. Suppose that A is a bounded operator that is a Toeplitz matrix with entries $\{a_j\}_j$. Then the integral kernel is

$$\begin{aligned} A(\omega,\lambda) &= \sum_{t\in\mathbb{Z}}\sum_{\tau\in\mathbb{Z}} a_{t-\tau}\exp(it\omega - i\tau\lambda) \\ &= \sum_{\tau\in\mathbb{Z}}\exp(-i\tau(\lambda-\omega))\sum_{r\in\mathbb{Z}} a_r\exp(ir\omega) = \delta_{\omega,\lambda}A(\omega) \end{aligned}$$

where $A(\omega) = \sum_{r\in\mathbb{Z}} a_r\exp(ir\omega)$. Since $\|A\| < \infty$, defining the infinite sequence $v = \{v_j\}$ where $v_j = 0$ for all $j \neq 0$ and $v_0 = 1$ we have $\sum_{j\in\mathbb{Z}} a_j^2 = \|Av\|_2 \leq \|A\|\|v\|_2 \leq \|A\|$, thus $A(\cdot) \in L_2[0,2\pi)$.

We now use (73) to prove the converse. Substituting $A(\omega,\lambda) = \delta_{\omega,\lambda}A(\omega)$ into (73) gives

$$[(FEF^*)u_j]_s = \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi} A(\omega)\delta_{\omega,\lambda}\exp(-ij\omega)\exp(is\lambda)d\omega d\lambda = a_{j-s}$$

where $a_r = (2\pi)^{-1}\int_0^{2\pi} A(\omega)\exp(-ir\omega)d\omega$. Thus the jth column of FEF^* is $\{a_{s-j}\}_{s\in\mathbb{Z}}$, which proves that the matrix defined by FEF^* is Toeplitz. □

PROOF of Lemma 3.2 Since A is block Toeplitz it follows from Lemma 3.1 that $\mathbf{A}(\omega,\lambda) = \mathbf{A}(\omega)\delta_{\omega,\lambda}$.

To derive an expression for the inverse, we first consider the case that $d = 1$. By definition $AA^{-1} = I$ (where I denotes the infinite dimension identity matrix), thus $F^*F = (F^*AF)(F^*A^{-1}F)$. By Lemma 3.1, the kernel operator of F^*AF is $A(\omega)\delta_{\omega,\lambda}$ and the kernel operator of $F^*A^{-1}F$ (since A^{-1} is Toeplitz) is $B(\omega)\delta_{\omega,\lambda}$. Since for all $g \in L_2[0,2\pi)$ we have

$$
\begin{aligned}
g(\omega) &= [F^*F(g)](\omega) = \frac{1}{(2\pi)^2}\int_0^{2\pi} B(\omega)\delta_{\omega,u}\int_0^{2\pi} A(u)\delta_{u,\lambda}g(\lambda)d\lambda du \\
&= \frac{1}{(2\pi)}\int_0^{2\pi} B(\omega)A(u)g(u)\delta_{\omega,u}du = B(\omega)A(\omega)g(\omega),
\end{aligned}
$$

then $B(\omega) = A(\omega)^{-1}$. This proves the result for all $d = 1$.

The proof for $d > 1$ uses the following invariance properties. If A and B are bounded Toeplitz matrix operators with kernels $A(\omega)\delta_{\omega,\lambda}$ and $B(\omega)\delta_{\omega,\lambda}$ respectively, then $A + B$ and AB are Toeplitz with kernels $[A(\omega)+B(\omega)]\delta_{\omega,\lambda}$ and $A(\omega)B(\omega)\delta_{\omega,\lambda}$ respectively. Using these properties together with the block operator inversion identity (in (49)) we will show, below, that the Lemma 3.2 holds for $d \geq 2$. We focus on $d = 2$ (the proof for $d > 2$ follows by induction). Let

$$
G = \begin{pmatrix} A & B \\ B^* & C \end{pmatrix}
$$

where G is a bounded operator and A, B and C are Toeplitz operators on ℓ_2, with integral kernels $A(\omega)\delta_{\omega,\lambda}$, $B(\omega)\delta_{\omega,\lambda}$ and $C(\omega)\delta_{\omega,\lambda}$. Then by (49)

$$
FG^{-1}F^* = \begin{pmatrix} FP^{-1}F^* & -FP^{-1}BC^{-1}F^* \\ -FC^{-1}B^*P^{-1}F^* & F(C - B^*A^{-1}B)^{-1}F^* \end{pmatrix},
$$

where $P = A - BC^{-1}B^*$. By the Toeplitz invariance properties described above, the integral kernel of FPF^* is $P(\omega)\delta_{\omega,\lambda}$ where

$$
P(\omega) = [A(\omega) - |B(\omega)|^2 C(\omega)].
$$

Thus by the proof for $d = 1$, the integral kernel of $FP^{-1}F^*$ (the top left hand side of $FG^{-1}F^*$) is $P(\omega)^{-1}\delta_{\omega,\lambda}$. A similar result holds for the other entries in $FG^{-1}F^*$. Therefore, the integral kernel of $FG^{-1}F^*$ is

$$
\begin{aligned}
&\begin{pmatrix} P(\omega)^{-1} & P(\omega)^{-1}B(\omega)C(\omega)^{-1} \\ C(\omega)^{-1}B(\omega)^*P(\omega)^{-1} & (C(\omega) - |B(\omega)|^2 A(\omega)^{-1})^{-1} \end{pmatrix}\delta_{\omega,\lambda} \\
= &\begin{pmatrix} A(\omega) & B(\omega) \\ B(\omega)^* & C(\omega) \end{pmatrix}^{-1}\delta_{\omega,\lambda}.
\end{aligned}
$$

This proves the result for $d = 2$. By induction the result can be proved for $d > 2$. $\square$

B.2 Proof of results in Sections 3.2 and 3.3

PROOF of Theorem 3.1 Under Assumption 2.1 and by using Lemma 2.1 for all $1 \leq a, b \leq p$, $D_{a,b}$ are bounded operators. Thus the proof is a straightforward application of Lemma 3.1. We summarize the main points below.

To prove (i) we note that $D_{a,b} = 0$ is a special case of Toeplitz matrix, thus $F^*D_{a,b}F =$

$0 \cdot \delta_{\omega,\lambda} = 0$. Conversely, if $F^{*}D_{a,b}F = 0$, then $D_{a,b} = 0$.

The proof of (ii) and (iii) immediately follow from Definition 2.1 and Lemma 3.1. $\square$

We now prove the results in Section 3.3. We first consider the Fourier transform of the rows of $D_{a,b}$ and $D_{a,b}$ in the case a node or edge is conditionally stationary.

PROOF of Theorem 3.2 We first prove (i). If the node a is conditionally stationary then $D_{a,a}$ is Toeplitz and its entries are determined by the row $\{[D_{a,a}]_{0,r}\}_{r}$. By using Lemma 3.1 we have $\Gamma_{t}^{(a,a)}(\omega) = \Gamma^{(a,a)}(\omega) = \sum_{r=-\infty}^{\infty}[D_{a,a}]_{0,r}\exp(ir\omega)$. To understand the meaning of this quantity, we note that from Proposition 2.1 for all $t \neq \tau$

$$\phi_{0,t-\tau}^{(a,a)} = \frac{\rho_{0,t-\tau}^{(a,a)}}{\sqrt{\rho_{0,0}^{(a,a)|\text{-}\{(a,0),(a,t-\tau)\}}\rho_{0,0}^{(a,a)|\text{-}\{(a,0),(a,t-\tau)\}}}} = -\frac{[D_{a,a}]_{t,\tau}}{\sqrt{[D_{a,a}]_{t,t}[D_{a,a}]_{\tau,\tau}}},$$

where $\rho_{0,t-\tau}^{(a,a)}$ and $\rho_{0,0}^{(a,a)|\text{-}\{(a,0),(a,t-\tau)\}}$ is defined in (6) and (52) respectively. Thus, we have $[D_{a,a}]_{t,\tau} = \sqrt{[D_{a,a}]_{t,t}[D_{a,a}]_{\tau,\tau}}\phi_{0,t-\tau}^{(a,a)}$. Further, we know that $[D_{a,a}]_{t,t} = 1/\sigma_{a}^{2} = 1/\rho_{0,0}^{(a,a)}$. Together this gives

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

This proves (i).

The proof of (ii) is identical to (i), thus we omit the details. $\square$

PROOF of Theorem 3.3 We first prove (i). By using Lemma 2.1, the integral kernel of $F^{*}D_{a,a}^{-1}F$ is $\delta_{\omega,\lambda}[\Gamma^{(a,a)}(\omega)]^{-1}$. We recall that $D_{a,a}^{-1}$ contains the time series partial covariances and by conditional stationarity and Theorem 2.1 we have $[D_{a,a}^{-1}]_{t,\tau} = \rho_{0,t-\tau}^{(a,a)|\text{-}\{a\}}$. Using this it is easily seen that the partial spectrum for the nodal time series partial covariance is

$$\Gamma^{(a,a)}(\omega)^{-1} = \sum_{r\in\mathbb{Z}}\mathbb{C}\mathrm{ov}[X_{0}^{(a,a)|\text{-}\{a\}}, X_{r}^{(a,a)|\text{-}\{a\}}]\exp(ir\omega) = \sum_{r\in\mathbb{Z}}\rho_{r}^{(a,a)|\text{-}\{a\}}\exp(ir\omega).$$

To prove (ii) we use that (a,b) is a conditionally stationary edge and define the suboperator block Toeplitz matrix $D_{\{a,b\}} : \ell_{2,2} \to \ell_{2,2}$, where $D_{\{a,b\}} = (D_{e,f}; e,f \in \{a,b\})$. The integral kernel of $F^{*}D_{\{a,b\}}F$ is $\Gamma_{\{a,b\}}(\omega)\delta_{\omega,\lambda}$ where

$$\Gamma_{\{a,b\}}(\omega) = \begin{pmatrix} \Gamma^{(a,a)}(\omega) & \Gamma^{(a,b)}(\omega) \\ \Gamma^{(a,b)}(\omega)^{*} & \Gamma^{(b,b)}(\omega) \end{pmatrix}. \quad (74)$$

The time series partial covariances are contained within the inverse $[D_{\{a,b\}}]^{-1}$ (which is block Toeplitz) (see Theorem 2.1). By using Lemma 3.2 the kernel of $F^{*}[D_{\{a,b\}}]^{-1}F$ is

$\Gamma_{\{a,b\}}(\omega)^{-1}\delta_{\omega,\lambda}$. Using this together with equation (72) we have

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

This proves the result. □

B.3 Proof of the results in Section 3.4

PROOF of Proposition 3.1 To connect the regression coefficients to entries in D we use the identity in (69) where we set $\mathcal{S} = \{a\}$ and $\mathcal{S}' = \{1,\ldots,p\}\backslash\{a\}$. This gives

$$
D = \begin{pmatrix} D_{a,a} & -D_{a,a}H_a' G_a^{-1} \\ -G_a^{-1}H_a D_{a,a} & (G_a - H_a^{*}C_{a,a}^{-1}H_a)^{-1} \end{pmatrix} \tag{75}
$$

where $H_a = C_{a,\mathcal{S}'}$ and $G_a = C_{\mathcal{S}',\mathcal{S}'}$. We recall that $D = (D_{e,f}, e, f \in \{1,\ldots,p\})$. Therefore, comparing the blocks on the left and right hand side of (75) gives the block vector

$$
(D_{a,b}; b \neq a) = -D_{a,a}H_a^{*}G_a^{-1}. \tag{76}
$$

Furthermore, by comparision, it is clear that the prediction coefficients $B_{b\mapsto a}$ satisfy

$$
H_a^{*}G_a^{-1} = (B_{b\mapsto a}; b \neq a)\,. \tag{77}
$$

Comparing (76) and (77) for $b \neq a$ we have $-D_{a,a}B_{b\mapsto a} = D_{a,b}$. Using that $D_{a,a}$ has an inverse yields the identity

$$
B_{b\mapsto a} = -D_{a,a}^{-1}D_{a,b}.
$$

This gives the result. □

PROOF of Theorem 3.4 To prove (i) and (ii) we use Lemma 3.1 where

$$
B_{b\mapsto a} = -D_{a,a}^{-1}D_{a,b}.
$$

To prove (i) we note that under Assumption 2.1, the null space of $D_{a,a}^{-1}$ is 0. Therefore, $D_{a,b} = 0$ iff $B_{b\mapsto a} = 0$. This proves (i).

To prove (ii) we note by the invariance properties of infinite Toeplitz matrix operators if $D_{a,b}$ and $B_{b\mapsto a}$ are Toeplitz, then $D_{a,b} = -D_{a,b}B_{b\mapsto a}$ is Toeplitz. Conversely, if $D_{a,b}$ and $D_{b,a}$ are Toeplitz, then $B_{b\mapsto a} = -D_{a,a}^{-1}D_{a,b}$ is Toeplitz. This proves (ii). □

PROOF of Corollary 3.1 To prove (i) we use that (a) under Assumption 2.1 that $\|D_{a,b}^{-1}\| \leq \lambda_{\sup}$ and (b) from Lemma 2.1, $D_{a,b}$ is a bounded operator. Thus, since $B_{b\mapsto a} = D_{a,b}^{-1}D_{a,b}$, we have $\|B_{b\mapsto a}\| \leq \|D_{a,b}^{-1}\|\|D_{a,b}\| < \infty$, thus proving (i).

The proofs of (ii) and (iii) are similar to the proof of Theorem 3.1, thus we omit the details. □

C Proof of Section 4

C.1 Proof of results in Section 4.1

We break the proof of Theorem 4.1 into a few steps. To bound the difference between the rows of $\widetilde{D}_n = C_n^{-1}$ and D_n (the submatrix of D), we use that the entries of $\widetilde{D}_n$ and D_n are the entries of coefficients in a regression. This allows us to use the Baxter inequality methods developed in Meyer et al. (2017) to bound the difference between projections on finite dimensional spaces and infinite dimensional spaces. The infinite and finite dimensional spaces we will use are $\mathcal{H} = \overline{\text{sp}}(X_t^{(c)}; t \in \mathbb{Z}, 1 \leq c \leq p)$ and $\mathcal{H}_n = \overline{\text{sp}}(X_\tau^{(b)}; 1 \leq \tau \leq n, 1 \leq b \leq p)$.

We recall from (3) that

$$P_{\mathcal{H}-X_t^{(a)}}(X_t^{(a)}) = \sum_{\tau \in \mathbb{Z}} \sum_{b=1}^{p} \beta_{(\tau,b) \to (t,a)} X_\tau^{(b)} \tag{78}$$

with $\beta_{(t,a)\to(t,a)} = 0$. Similarly, projecting $X_t^{(a)}$ onto the finite dimensional space $\mathcal{H}_n$ is

$$P_{\mathcal{H}_n-X_t^{(a)}}(X_t^{(a)}) = \sum_{\tau=1}^{n} \sum_{b=1}^{p} \theta_{(\tau,b)\to(t,a),n} X_\tau^{(b)}, \tag{79}$$

with $\theta_{(t,a)\to(t,a),n} = 0$. Let

$$\begin{aligned} \sigma_{a,t}^2 &= \mathrm{E}[X_t^{(a)} - P_{\mathcal{H}-X_t^{(a)}}(X_t^{(a)})]^2 \\ \text{and } \widetilde{\sigma}_{a,t,n}^2 &= \mathrm{E}[X_t^{(a)} - P_{\mathcal{H}_n-X_t^{(a)}}(X_t^{(a)})]^2. \end{aligned} \tag{80}$$

Define the $(np-1)$-dimensional vectors

$$\begin{aligned} \underline{B}_n^{(a,t)} &= \{\beta_{(\tau,b)\to(t,a)}; 1 \leq \tau \leq n \text{ and } 1 \leq b \leq p, \text{ not } (\tau,b) = (t,a)\} \\ \text{and } \underline{\Theta}_n^{(a,t)} &= \{\theta_{(\tau,b)\to(t,a),n}; 1 \leq \tau \leq n \text{ and } 1 \leq b \leq p, \text{ not } (\tau,b) = (t,a)\}. \end{aligned} \tag{81}$$

To minimise notation we will drop the n, and let $\theta_{(\tau,b)\to(t,a)} = \theta_{(\tau,b)\to(t,a),n}$ and $\widetilde{\sigma}_{a,t}^2 = \widetilde{\sigma}_{a,t,n}^2$. But we should keep in mind that both θ and σ depend on n. Since the coefficients of a precision matrix are closely related to the coefficients in a regression it is clear that the tth "row" of the matrix $\widetilde{D}_n = C_n^{-1}$ at node a (which is the $((a-1)n+t)$th row of $\widetilde{D}_n$) is the rearranged vector

$$\underline{\widetilde{\Theta}}_n^{(a,t)} = \frac{1}{\widetilde{\sigma}_{a,t}^2}[1, -\underline{\Theta}_n^{(a,t)}]. \tag{82}$$

The tth row of matrix D_n at node a is the similarly rearranged vector

$$\underline{\widetilde{B}}_n^{(a,t)} = \frac{1}{\sigma_{a,t}^2}[1, -\underline{B}_n^{(a,t)}].$$

Thus the difference between $\underline{\widetilde{\Theta}}_n^{(a,t)}$ and $\underline{\widetilde{B}}_n^{(a,t)}$ is

$$\begin{aligned} & \underline{\widetilde{\Theta}}_n^{(a,t)} - \underline{\widetilde{B}}_n^{(a,t)} \\ = & \left[\frac{1}{\widetilde{\sigma}_{a,t}^2} - \frac{1}{\sigma_{a,t}^2}\right]\left(1, -\underline{\Theta}_n^{(a,t)}\right) + \frac{1}{\sigma_{a,t}^2}\left[0, \left(\underline{B}_n^{(a,t)} - \underline{\Theta}_n^{(a,t)}\right)\right]. \end{aligned}$$

Since both $\underline{\widetilde{\Theta}}_n^{(a,t)}$ and $\underline{\widetilde{B}}_n^{(a,t)}$ are the (same) rearranged rows of $[\widetilde{D}_n]_{(a-1)n+t,\cdot}$ and $[D_n]_{(a-1)n+t,\cdot}$, The ℓ_1-difference between the $((a-1)n+t)$th row of D_n and $\widetilde{D}_n$ is

$$\begin{aligned} & \left\|[D_n]_{(a-1)n+t,\cdot} - [\widetilde{D}_n]_{(a-1)n+t,\cdot}\right\|_1 = \left\|\underline{\widetilde{\Theta}}_n^{(a,t)} - \underline{\widetilde{B}}_n^{(a,t)}\right\|_1 \\ \leq & \frac{|\sigma_{a,t}^2 - \widetilde{\sigma}_{a,t}^2|}{\widetilde{\sigma}_{a,t}^2\sigma_{a,t}^2}\left(1 + \|\underline{\Theta}_n^{(a,t)}\|_1\right) + \frac{1}{\sigma_{a,t}^2}\left\|\underline{B}_n^{(a,t)} - \underline{\Theta}_n^{(a,t)}\right\|_1. \end{aligned} \tag{83}$$

In the two lemmas below we obtain a bound for the differences $|\sigma_{a,t}^2 - \widetilde{\sigma}_{a,t}^2|$ and $\|\underline{B}_n^{(a,t)} - \underline{\Theta}_n^{(a,t)}\|_1$. These two bounds will prove Theorem 4.1.

Lemma C.1 *Suppose Assumptions 2.1 and 4.1 hold. Let $\underline{B}_n^{(a,t)}$ and $\underline{\Theta}_n^{(a,t)}$ be defined as in (81). Then*

$$\left\|\underline{B}_n^{(a,t)} - \underline{\Theta}_n^{(a,t)}\right\|_2 \leq \lambda_{\inf}^{-1}\lambda_{\sup}\sum_{b=1}^{p}\sum_{\tau\notin\{1,\ldots,n\}}|\beta_{(\tau,b)\to(t,a)}|.$$

PROOF The proof is based on the innovative technique developed in Meyer et al. (2017) (who used the method to obtain Baxter bounds for stationary spatial processes). We start by deriving the normal equations corresponding to (78) and (79) for $1 \leq s \leq n$ and $c = 1, \ldots, p$ (excluding $(c,s) = (a,t)$). For equation (78) this gives the normal equations

$$\begin{aligned} \mathbb{C}\mathrm{ov}(X_t^{(a)}, X_s^{(c)}) = & \sum_{b=1}^{p}\sum_{\tau=1}^{n}\beta_{(\tau,b)\to(t,a)}\,\mathbb{C}\mathrm{ov}(X_\tau^{(b)}, X_s^{(c)}) + \\ & \sum_{b=1}^{p}\sum_{\tau\notin\{1,\ldots,n\}}\beta_{(\tau,b)\to(t,a)}\,\mathbb{C}\mathrm{ov}(X_\tau^{(b)}, X_s^{(c)}) \end{aligned} \tag{84}$$

and for (79) this gives

$$\mathbb{C}\mathrm{ov}(X_t^{(a)}, X_s^{(c)}) = \sum_{b=1}^{p}\sum_{\tau=1}^{n}\theta_{(\tau,b)\to(t,a)}\,\mathbb{C}\mathrm{ov}(X_\tau^{(b)}, X_s^{(c)}). \tag{85}$$

Taking the difference between (84) and (85) we have

$$\sum_{b=1}^{p}\sum_{\tau=1}^{n}\left[\beta_{(\tau,b)\to(t,a)}-\theta_{(\tau,b)\to(t,a)}\right]\mathbb{C}\mathrm{ov}(X_{\tau}^{(b)},X_{s}^{(c)}) \;\;=\;\; -\sum_{b=1}^{p}\sum_{\tau\notin\{1,\ldots,n\}}\beta_{(\tau,b)\to(t,a)}\,\mathbb{C}\mathrm{ov}(X_{\tau}^{(b)},X_{s}^{(c)}).$$

As the above holds for all $1\leq s\leq n$ and $1\leq c\leq p$ (excluding $X_{t}^{(a)}$) we can write the above as a vector equation

$$\begin{aligned}&\sum_{b=1}^{p}\sum_{\tau\in\{1,\ldots,n\}}\left[\beta_{(\tau,b)\to(t,a)}-\theta_{(\tau,b)\to(t,a)}\right]\mathbb{C}\mathrm{ov}(X_{\tau}^{(b)},\underline{Y}_{n})\\ =\;& -\sum_{b=1}^{p}\sum_{\tau\notin\{1,\ldots,n\}}\beta_{(\tau,b)\to(t,a)}\,\mathbb{C}\mathrm{ov}(X_{\tau}^{(b)},\underline{Y}_{n}),\end{aligned} \tag{86}$$

where $\underline{Y}_{n}=(X_{s}^{(c)};1\leq s\leq n,1\leq c\leq p,(c,s)\neq(a,t)\})$. We observe that the LHS of the above can be expressed as

$$\begin{aligned}&\sum_{b=1}^{p}\sum_{\tau\in\{1,\ldots,n\}}\left[\beta_{(\tau,b)\to(t,a)}-\theta_{(\tau,b)\to(t,a)}\right]\mathbb{C}\mathrm{ov}(X_{\tau}^{(b)},\underline{Y}_{n})\\ =\;& \mathbb{C}\mathrm{ov}\left(\sum_{b=1}^{p}\sum_{\tau\in\{1,\ldots,n\}}\left[\beta_{(\tau,b)\to(t,a)}-\theta_{(\tau,b)\to(t,a)}\right]X_{\tau}^{(b)},\underline{Y}_{n}\right)\\ =\;& \mathbb{C}\mathrm{ov}\left(\left[\underline{B}_{n}^{(a,t)}-\underline{\Theta}_{n}^{(a,t)}\right]'\underline{Y}_{n},\underline{Y}_{n}\right),\end{aligned} \tag{87}$$

where the last line of the above is due to

$$\sum_{b=1}^{p}\sum_{\tau\in\{1,\ldots,n\}}\left[\beta_{(\tau,b)\to(t,a)}-\theta_{(\tau,b)\to(t,a)}\right]X_{\tau}^{(b)}=\left[\underline{B}_{n}^{(a,t)}-\underline{\Theta}_{n}^{(a,t)}\right]'\underline{Y}_{n}.$$

Substituting (87) into the LHS of (86) gives the vector equation

$$(\mathbb{V}\mathrm{ar}[\underline{Y}_{n}])\,[\underline{B}_{n}^{(a,t)}-\underline{\Theta}_{n}^{(a,t)}]=-\sum_{b=1}^{p}\sum_{\tau\notin\{1,\ldots,n\}}\beta_{(\tau,b)\to(t,a)}\,\mathbb{C}\mathrm{ov}(X_{\tau}^{(b)},\underline{Y}_{n}).$$

Therefore

$$\left[\underline{B}_{n}^{(a,t)}-\underline{\Theta}_{n}^{(a,t)}\right]=-\,\mathbb{V}\mathrm{ar}[\underline{Y}_{n}]^{-1}\sum_{b=1}^{p}\sum_{\tau\notin\{1,\ldots,n\}}\beta_{(\tau,b)\to(t,a)}\,\mathbb{C}\mathrm{ov}(X_{\tau}^{(b)},\underline{Y}_{n}).$$

Now taking the ℓ_2-norm of the above we have

$$\begin{aligned}
\left\|\underline{B}_n^{(a,t)} - \underline{\Theta}_n^{(a,t)}\right\|_2 &\leq \left\|\mathbb{V}\mathrm{ar}[\underline{Y}_n]^{-1}\right\| \sum_{b=1}^{p} \sum_{\tau \notin \{1,\ldots,n\}} |\beta_{(\tau,b)\to(t,a)}| \left\|\mathbb{C}\mathrm{ov}(X_\tau^{(b)}, \underline{Y}_n)\right\|_2 \\
&\leq \left\|\mathbb{V}\mathrm{ar}[\underline{Y}_n]^{-1}\right\| \sum_{b=1}^{p} \sum_{\tau \notin \{1,\ldots,n\}} |\beta_{(\tau,b)\to(t,a)}| \left\|\mathbb{C}\mathrm{ov}(X_\tau^{(b)}, \underline{Y}_n)\right\|_2 \\
&\leq \left\|\mathbb{V}\mathrm{ar}[\underline{Y}_n]^{-1}\right\| \left(\sup_{\tau,b}[\sum_{c=1}^{p}\sum_{s=-\infty}^{\infty} \mathbb{C}\mathrm{ov}(X_\tau^{(b)}, X_s^{(c)})^2]^{1/2}\right) \sum_{b=1}^{p} \sum_{\tau \notin \{1,\ldots,n\}} |\beta_{(\tau,b)\to(t,a)}|,
\end{aligned}$$

where $\|\cdot\|$ denotes the (spectral) matrix norm. By using Assumption 2.1 we have $\lambda_{\min}(\mathbb{V}\mathrm{ar}[\underline{Y}_n]) \geq \lambda_{\inf}$, thus $\|\mathbb{V}\mathrm{ar}[\underline{Y}_n]^{-1}\| \leq \lambda_{\inf}^{-1}$. Again by Assumption 2.1 we have $\sup_{\tau,b}[\sum_{c=1}^{p}\sum_{s=-\infty}^{\infty} \mathbb{C}\mathrm{ov}(X_\tau^{(b)}, X_s^{(c)})^2]^{1/2} \leq \lambda_{\sup}$. Substituting these two bounds into the above, gives

$$\left\|\underline{B}_n^{(a,t)} - \underline{\Theta}_n^{(a,t)}\right\|_2 \leq \lambda_{\inf}^{-1}\lambda_{\sup} \sum_{b=1}^{p} \sum_{\tau \notin \{1,\ldots,n\}} |\beta_{(\tau,b)\to(t,a)}|.$$

This proves the result. $\square$

The above gives a bound for the ℓ_2-norm. To obtain a bound on the ℓ_1-norm we use the Cauchy-Schwarz inequality to give

$$\left\|\underline{B}_n^{(a,t)} - \underline{\Theta}_n^{(a,t)}\right\|_1 \leq (np)^{1/2}\lambda_{\inf}^{-1}\lambda_{\sup} \sum_{b=1}^{p} \sum_{\tau \notin \{1,\ldots,n\}} |\beta_{(\tau,b)\to(t,a)}|. \quad (88)$$

Next we bound the difference between $\widetilde{\sigma}_{a,t}^2$ and $\sigma_{a,t}^2$.

Lemma C.2 *Suppose Assumptions 2.1 and 4.1 hold. Let $\sigma_{a,t}^2$ and $\widetilde{\sigma}_{a,t,n}^2$ be defined as in (80). Then*

$$0 \leq \widetilde{\sigma}_{a,t,n}^2 - \sigma_{a,t}^2 \leq \left[\lambda_{\inf}^{-1}\lambda_{\sup}^2 + \lambda_{\sup}\right] \sum_{b=1}^{p} \sum_{\tau \notin \{1,\ldots,n\}} |\beta_{(\tau,b)\to(t,a)}|.$$

PROOF First we note that since $[\mathcal{H}_n - X_t^a] \subseteq [\mathcal{H} - X_t^{(a)}]$, then $\widetilde{\sigma}_{a,t,n}^2 \geq \sigma_{a,t}^2$ and $0 \leq \widetilde{\sigma}_{a,t,n}^2 - \sigma_{a,t}^2$. To prove the result, we recall if $P_{\mathcal{G}}(Y)$ is the projection of Y onto $\mathcal{G}$, then

$$\mathbb{V}\mathrm{ar}[Y - P_{\mathcal{G}}(Y)] = \mathbb{V}\mathrm{ar}[Y] - \mathbb{C}\mathrm{ov}[Y, P_{\mathcal{G}}(Y)].$$

Using the above, with $\mathcal{G}_n = \mathcal{H}_n - X_t^{(a)}$ and $\mathcal{G} = \mathcal{H} - X_t^{(a)}$ and taking differences gives

$$\begin{aligned}
\widetilde{\sigma}_{a,t,n}^2 - \sigma_{a,t}^2 &= \mathbb{V}\mathrm{ar}[X_t^{(a)} - P_{\mathcal{H}_n - X_t^{(a)}}(X_t^{(a)})] - \mathbb{V}\mathrm{ar}[X_t^{(a)} - P_{\mathcal{H} - X_t^{(a)}}(X_t^{(a)})] \\
&= \mathbb{C}\mathrm{ov}\left[X_t^{(a)}, P_{\mathcal{H} - X_t^{(a)}}(X_t^{(a)}) - P_{\mathcal{H}_n - X_t^a}(X_t^{(a)})\right].
\end{aligned}$$

Substituting the expressions for $P_{\mathcal{H}-X_t^{(a)}}(X_t^{(a)})$ and $P_{\mathcal{H}_n-X_t^{(a)}}(X_t^{(a)})$ in (78) and (79) into the above gives

$$
\begin{aligned}
\widetilde{\sigma}^2_{a,t,n} - \sigma^2_{a,t} &= \mathbb{C}\mathrm{ov}\left[X_t^{(a)}, \sum_{b=1}^{p}\sum_{\tau=1}^{n}\left(\beta_{(\tau,b)\to(t,a)} - \theta_{(\tau,b)\to(t,a)}\right)X_\tau^{(b)}\right] + \\
& \quad \mathbb{C}\mathrm{ov}\left[X_t^{(a)}, \sum_{b=1}^{p}\sum_{\tau\notin\{1,\ldots,n\}}\beta_{(\tau,b)\to(t,a)}X_\tau^{(b)}\right] \\
&= \sum_{b=1}^{p}\sum_{\tau=1}^{n}\left(\beta_{(\tau,b)\to(t,a)} - \theta_{(\tau,b)\to(t,a)}\right)\mathbb{C}\mathrm{ov}[X_\tau^{(b)}, X_t^{(a)}] + \\
& \quad \sum_{b=1}^{p}\sum_{\tau\notin\{1,\ldots,n\}}\beta_{(\tau,b)\to(t,a)}\,\mathbb{C}\mathrm{ov}[X_\tau^{(b)}, X_t^{(a)}].
\end{aligned}
$$

Applying the Cauchy-Schwarz inequality to the above gives

$$
\begin{aligned}
\widetilde{\sigma}^2_{a,t,n} - \sigma^2_{a,t} &\leq \left[\sum_{b=1}^{p}\sum_{\tau=1}^{n}\left(\beta_{(\tau,b)\to(t,a)} - \theta_{(\tau,b)\to(t,a)}\right)^2\right]^{1/2}\left[\sum_{b=1}^{p}\sum_{\tau=1}^{n}\mathbb{C}\mathrm{ov}[X_\tau^{(b)}, X_t^{(a)}]^2\right]^{1/2} \\
& \quad + \left[\sum_{b=1}^{p}\sum_{\tau\notin\{1,\ldots,n\}}\beta^2_{(\tau,b)\to(t,a)}\right]^{1/2}\left[\sum_{b=1}^{p}\sum_{\tau\notin\{1,\ldots,n\}}\mathbb{C}\mathrm{ov}[X_\tau^{(b)}, X_t^{(a)}]^2\right]^{1/2}.
\end{aligned}
$$

Applying the bound in Lemma C.1 to the first term on the RHS and using that the sum of the covariances squared are bounded by $\lambda_{\sup}$ we have

$$
\begin{aligned}
\widetilde{\sigma}^2_{a,t,n} - \sigma^2_{a,t} &\leq \lambda_{\inf}^{-1}\lambda_{\sup}^2\sum_{b=1}^{p}\sum_{\tau\notin\{1,\ldots,n\}}|\beta_{(\tau,b)\to(t,a)}| + \lambda_{\sup}\left[\sum_{b=1}^{p}\sum_{\tau\notin\{1,\ldots,n\}}\beta^2_{(\tau,b)\to(t,a)}\right]^{1/2} \\
&\leq \left[\lambda_{\inf}^{-1}\lambda_{\sup}^2 + \lambda_{\sup}\right]\sum_{b=1}^{p}\sum_{\tau\notin\{1,\ldots,n\}}|\beta_{(\tau,b)\to(t,a)}|,
\end{aligned}
$$

where the above follows from the fact that the ℓ_2-norm of a vector is bounded from above by the ℓ_1-norm. This gives the required result. □

We use Lemmas C.1, C.2 and equation (88) to prove Theorem 4.1.

PROOF of Theorem 4.1 We first obtain a bound for the sum of the regression coefficients

$$
\sum_{b=1}^{p}\sum_{\tau\notin\{1,\ldots,n\}}|\beta_{(\tau,b)\to(t,a)}| = \sum_{b=1}^{p}\left[\sum_{\tau=n+1}^{\infty} + \sum_{\tau=-\infty}^{0}\right]|\beta_{(\tau,b)\to(t,a)}|.
$$

Under Assumption 4.1 we have

$$
\begin{aligned}
\sum_{b=1}^{p}\sum_{\tau\notin\{1,\ldots,n\}}|\beta_{(\tau,b)\to(t,a)}| &\leq \lambda_{\inf}^{-1}\sum_{b=1}^{p}\left[\sum_{\tau=n+1}^{\infty}+\sum_{\tau=-\infty}^{0}\right]\frac{1}{\ell(\tau-t)} \\
&= \lambda_{\inf}^{-1}\sum_{b=1}^{p}\sum_{\tau=n+1}^{\infty}\frac{1}{\ell(\tau-t)}+\sum_{b=1}^{p}\sum_{\tau=-\infty}^{0}\frac{1}{\ell(\tau-t)} \\
&= \lambda_{\inf}^{-1}\sum_{b=1}^{p}\sum_{j=n+1-t}^{\infty}\frac{1}{\ell(j)}+\sum_{b=1}^{p}\sum_{\tau=-\infty}^{-t}\frac{1}{\ell(j)}.
\end{aligned}
$$

Under Assumption 4.1 we have for $r>0$ $\lambda_{\inf}^{-1}\sum_{j=r}^{\infty}\ell(j) \leq \lambda_{\inf}^{-1}r^{-K}\sum_{j=r}^{\infty}j^{K}\ell(j)^{-1} \leq C_{\ell}r^{-K}$, where $C_{\ell}=\lambda_{\inf}^{-1}\sum_{j\in\mathbb{Z}}\ell(j)^{-1}$. And by a similar argument for $r<0$, $\lambda_{\inf}^{-1}\sum_{j=-\infty}^{r}\ell(j)\leq C_{\ell}|r|^{-K}$. Applying these two bounds to the above we have

$$
\sum_{b=1}^{p}\sum_{\tau\notin\{1,\ldots,n\}}|\beta_{(\tau,b)\to(t,a)}| \leq 2pC_{\ell}\min(|n+1-t|,|t|)^{-K} \tag{89}
$$

We use this inequality to prove the result.

We return to (83) which gives the bound

$$
\left\|[\widetilde{D}_n]_{(a-1)n+t,\cdot}-[D_n]_{(a-1)n+t,\cdot}\right\|_1 \leq \frac{|\sigma_{a,t}^2-\widetilde{\sigma}_{a,t}^2|}{\widetilde{\sigma}_{a,t}^2\sigma_{a,t}^2}\left(1+\|\underline{\Theta}_n^{(a,t)}\|_1\right)+\frac{1}{\sigma_{a,t}^2}\left\|\underline{B}_n^{(a,t)}-\underline{\Theta}_n^{(a,t)}\right\|_1.
$$

Substituting the bounds in (88) and Lemma C.2 into the above gives

$$
\begin{aligned}
&\left\|[\widetilde{D}_n]_{(a-1)n+t,\cdot}-[D_n]_{(a-1)n+t,\cdot}\right\|_1 \\
\leq\ & \frac{1}{\widetilde{\sigma}_{a,t}^2\sigma_{a,t}^2}\left[\lambda_{\inf}^{-1}\lambda_{\sup}^2+\lambda_{\sup}\right]\left(1+\|\underline{\Theta}_n^{(a,t)}\|_1\right)\sum_{b=1}^{p}\sum_{\tau\notin\{1,\ldots,n\}}|\beta_{(\tau,b)\to(t,a)}| \\
&+\frac{1}{\sigma_{a,t}^2}(np)^{1/2}\lambda_{\inf}^{-1}\lambda_{\sup}\sum_{b=1}^{p}\sum_{\tau\notin\{1,\ldots,n\}}|\beta_{(\tau,b)\to(t,a)}| \\
\leq\ & \left(\frac{1}{\widetilde{\sigma}_{a,t}^2\sigma_{a,t}^2}\left[\lambda_{\inf}^{-1}\lambda_{\sup}^2+\lambda_{\sup}\right]\left(1+\|\underline{\Theta}_n^{(a,t)}\|_1\right)+\frac{1}{\sigma_{a,t}^2}(np)^{1/2}\lambda_{\inf}^{-1}\lambda_{\sup}\right)\sum_{b=1}^{p}\sum_{\tau\notin\{1,\ldots,n\}}|\beta_{(\tau,b)\to(t,a)}|.
\end{aligned}
$$

We now bound $\|\underline{\Theta}_n^{(a,t)}\|_1$, $\widetilde{\sigma}_{a,t}^{-2}$ and $\sigma_{a,t}^{-2}$ in terms of the eigenvalues of C. By using (82) we have $\underline{\widetilde{\Theta}}_n^{(a,t)}=\frac{1}{\widetilde{\sigma}_{a,t}^2}[1,-\underline{\Theta}_n^{(a,t)}]$, this gives the inequality

$$
\|\underline{\Theta}_n^{(a,t)}\|_1 \leq \widetilde{\sigma}_{a,t}^2\|\underline{\widetilde{\Theta}}_n^{(a,t)}\|_1.
$$

Since $\underline{\widetilde{\Theta}}_n^{(a,t)}$ are the (rearranged) rows of $\widetilde{D}_n=C_n^{-1}$ and the smallest eigenvalue of C_n is bounded from below by $\lambda_{\inf}$ we have that

$$\|\underline{\Theta}_n^{(a,t)}\|_1 \leq \widetilde{\sigma}_{a,t}^2 \lambda_{\inf}^{-1}.$$

Since $\widetilde{\sigma}_{a,t}^2 \leq \mathbb{V}\mathrm{ar}[X_t^{(a)}] = [\mathbf{C}_{t,t}]_{a,a}$, and $[\mathbf{C}_{t,t}]_{a,a} \leq \sum_\tau \|[\mathbf{C}_{t,\tau}]_{a,\cdot}\|_2^2 \leq \lambda_{\sup}$ then $\widetilde{\sigma}_{a,t}^2 \leq \lambda_{\sup}$, thus $\|\underline{\Theta}_n^{(a,t)}\|_1 \leq \lambda_{\inf}^{-1}\lambda_{\sup}$.

By using (48) (from the start of Appendix A.1) we have $\sigma_{a,t}^{-2} \leq \lambda_{\inf}^{-1}$. Furthermore, by using the same arguments used to show that $\sigma_{a,t}^{-2} \leq \lambda_{\inf}^{-1}$ we can also show $\widetilde{\sigma}_{a,t}^{-2} \leq \lambda_{\inf}^{-1}$.

Altogether, these bounds with (89) give

$$\begin{aligned}
& \left\|[\widetilde{D}_n]_{(a-1)n+t,\cdot} - [D_n]_{(a-1)n+t,\cdot}\right\|_1 \\
\leq\ & \left(\frac{1}{\lambda_{\inf}^2}\left[\lambda_{\inf}^{-1}\lambda_{\sup}^2 + \lambda_{\sup}\right]\left(1+\lambda_{\inf}^{-1}\lambda_{\sup}\right) + \frac{1}{\lambda_{\inf}}(np)^{1/2}\lambda_{\inf}^{-1}\lambda_{\sup}\right) 2pC_\ell \min(|n+1-t|,|t|)^{-K} \\
=\ & O\left((np)^{1/2}\min(|n+1-t|,|t|)^{-K}\right),
\end{aligned}$$

where the constants above only depend on $\lambda_{\inf}$, $\lambda_{\sup}$, p and $C_\ell = \lambda_{\inf}^{-1}\sum_{j\in\mathbb{Z}}\ell(j)^{-1}$. Thus proving the result. $\square$

PROOF of Proposition 4.1 By definition we have

$$[\mathbf{K}_n(\omega_{k_1},\omega_{k_2})]_{a,b} = [F_n^*\widetilde{D}_{a,b;n}F_n]_{k_1,k_2} = \frac{1}{n}\sum_{\tau,t=1}^{n}[\widetilde{D}_{a,b;n}]_{t,\tau}\exp(-it\omega_{k_1}+i\tau\omega_{k_2})$$

Replacing $\widetilde{D}_{a,b;n}$ with $D_{a,b;n}$ and using Theorem 4.1 gives

$$\begin{aligned}
|([F_n^*(\widetilde{D}_n - D_n)F_n]_{k_1,k_2})_{a,b}| &\leq \frac{1}{n}\sum_{t=1}^{n}\sum_{\tau=1}^{n}\left|[\widetilde{D}_{a,b;n} - D_{a,b;n}]_{t,\tau}\right| \\
&\leq \frac{1}{n}\sum_{t=1}^{n}\frac{(np)^{1/2}}{\min(|t-n+1|,|t|)^K} = O\left(\frac{(np)^{1/2}}{n^K}\right),
\end{aligned}$$

where the above holds for $K>1$. This gives

$$\begin{aligned}
[\mathbf{K}_n(\omega_{k_1},\omega_{k_2})]_{a,b} &= [F_n^*D_{a,b;n}F_n]_{k_1,k_2} + O\left(\frac{(np)^{1/2}}{n^K}\right) \\
&= \frac{1}{n}\sum_{t=1}^{n}\sum_{\tau=1}^{n}[D_{a,b}]_{t,\tau}\exp(-it\omega_{k_1}+i\tau\omega_{k_2}) + O\left(\frac{(np)^{1/2}}{n^K}\right). \quad (90)
\end{aligned}$$

Now we obtain an expression for the leading term in the RHS of the above in terms of

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

$$
\begin{aligned}
& \frac{1}{n}\sum_{t=1}^{n}\sum_{\tau=1}^{n}[D_{a,b}]_{t,\tau}\exp(-it\omega_{k_1}+i\tau\omega_{k_2}) \\
= & \frac{1}{n}\sum_{t=1}^{n}\sum_{\tau=1}^{n}[D_{a,b}]_{t,\tau}\exp(-it(\omega_{k_1}-\omega_{k_2}))\exp(i(\tau-t)\omega_{k_2}) \quad \text{let } r=\tau-t \\
= & \frac{1}{n}\sum_{t=1}^{n}\exp(-it(\omega_{k_1}-\omega_{k_2}))\sum_{r=1-t}^{n-t}[D_{a,b}]_{t,t+r}\exp(ir\omega_{k_2}) \\
= & \frac{1}{n}\sum_{t=1}^{n}\exp(-it(\omega_{k_1}-\omega_{k_2}))\sum_{r=-\infty}^{\infty}[D_{a,b}]_{t,t+r}\exp(ir\omega_{k_2}) + O\left(\frac{1}{n}\sum_{t=1}^{n}\left[\sum_{r=-\infty}^{1-t}+\sum_{r=n-t+1}^{\infty}\right]\frac{1}{\ell(r)}\right) \\
= & \frac{1}{n}\sum_{t=1}^{n}\exp(-it(\omega_{k_1}-\omega_{k_2}))\sum_{r=-\infty}^{\infty}[D_{a,b}]_{t,t+r}\exp(ir\omega_{k_2}) + O\left(\frac{1}{n}\sum_{r\in\mathbb{Z}}\frac{|r|}{\ell(r)}\right) \\
= & \frac{1}{n}\sum_{t=1}^{n}\exp(-i(k_1-k_2)\omega_t)\Gamma_t^{(a,b)}(\omega_{k_2}) + O\left(n^{-1}\right).
\end{aligned}
$$

By a similar argument we can show that

$$
\begin{aligned}
& \frac{1}{n}\sum_{t=1}^{n}\sum_{\tau=1}^{n}[D_{a,b}]_{t,\tau}\exp(-it\omega_{k_1}+i\tau\omega_{k_2}) \\
= & \frac{1}{n}\sum_{\tau=1}^{n}\exp(i(k_2-k_1)\omega_\tau)\Gamma_\tau^{(b,a)}(\omega_{k_1})^{*} + O\left(n^{-1}\right) \\
= & \left[\frac{1}{n}\sum_{\tau=1}^{n}\exp(-i(k_2-k_1)\omega_\tau)\Gamma_\tau^{(b,a)}(\omega_{k_1})\right]^{*} + O\left(n^{-1}\right)
\end{aligned}
$$

Therefore, since $O((np)^{1/2}/n^K) = O(1/n)$ when $K \geq 3/2$ and p is fixed, substituting the above into (90) we have

$$
[\mathbf{K}_n(\omega_{k_1},\omega_{k_2})]_{a,b} = \frac{1}{n}\sum_{t=1}^{n}\exp(-i(k_1-k_2)\omega_t)\Gamma_t^{(a,b)}(\omega_{k_2}) + O\left(\frac{1}{n}\right)
$$

and

$$
[\mathbf{K}_n(\omega_{k_1},\omega_{k_2})]_{a,b} = \left[\frac{1}{n}\sum_{t=1}^{n}\exp(-i(k_2-k_1)\omega_t)\Gamma_t^{(b,a)}(\omega_{k_1})\right]^{*} + O\left(\frac{1}{n}\right)
$$

this proves (26).

To prove (27) (under conditional stationarity) we use that $\Gamma_t^{(a,b)}(\omega) = \Gamma^{(a,b)}(\omega)$ for all

t. Substituting this into (26) gives

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

Now by using that

$$\frac{1}{n}\sum_{t=1}^{n}\exp\left(-it\omega_{k_1-k_2,n}\right) = \begin{cases} 0 & k_1-k_2 \notin n\mathbb{Z} \\ 1 & k_1-k_2 \in n\mathbb{Z} \end{cases}$$

immediately proves (27). □

C.2 Proofs for Section 4.3

PROOF of Proposition 4.2 The proof follows from the definition of $K_r^{(a,b)}(\omega)$

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

PROOF of Proposition 4.3 To prove the result we use (26) in Proposition 4.1 to give

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

We replace $\Gamma_t^{(a,b)}(\omega_{k_2})$ with $\Gamma^{(a,b)}(t/n,\omega_{k_2})$. Using the locally stationary approximation bound in (31) we have

$$[\mathbf{K}_n(\omega_{k_1},\omega_{k_2})]_{a,b} \quad = \quad \frac{1}{n}\sum_{t=1}^{n}\exp(-i(k_1-k_2)\omega_t)\Gamma^{(a,b)}\left(\frac{t}{n},\omega_{k_2}\right) + O\left(\frac{1}{n}\right). \qquad (91)$$

This proves (33).

In order to prove (34) we study the smoothness of $\Gamma^{(a,b)}(u,\omega)$ over u and its corresponding Fourier coefficients (keeping ω fixed). We first observe that under Assumption 4.2 we have that

$$\frac{\partial\Gamma^{(a,b)}(u,\omega)}{\partial u} = \sum_{j\in\mathbb{Z}}\frac{d[\mathbf{D}_j(u)]_{a,b}}{du}\exp(ij\omega).$$

This leads to the bound

$$\sup_{u,\omega}\left|\frac{\partial\Gamma^{(a,b)}(u,\omega)}{\partial u}\right| = \sup_{u}\sum_{j\in\mathbb{Z}}\left|\frac{d[\mathbf{D}_j(u)]_{a,b}}{du}\right| \leq \sum_{j\in\mathbb{Z}}\ell(j)^{-1} < \infty. \qquad (92)$$

We use this bound below. To simplify notation, we drop the (a,b) and ω in $\Gamma^{(a,b)}(u,\omega)$

(as they do not play a role in the bound). In order to understand the rate of decay of the Fourier coefficients of $\Gamma(\cdot)$ we note that Γ is a piecewise continuous 1-periodic function (where $\Gamma(u) = \Gamma(u+n)$ for all $n \in \mathbb{Z}$). Define the Fourier coefficient

$$K_r = \int_0^1 \Gamma(u)\exp(-ir2\pi u)du.$$

By using (92) the derivative of $\Gamma(\cdot)$ is bounded on the interior $(0,1)$ (it is unlikely to exist at 0 and 1 since typically $\Gamma(0) \neq \Gamma(1)$). Thus by integration by parts we have the bound

$$|K_r| \leq C|r|^{-1} \text{ for } r \neq 0. \tag{93}$$

We now obtain the limit for

$$\frac{1}{n}\sum_{k=1}^{n}\Gamma(k/n)\exp(-ir\omega_k) \qquad |r| \leq n/2.$$

In particular, we show that

$$\sup_{|r|\leq n/2}\left|\frac{1}{n}\sum_{k=1}^{n}\Gamma(k/n)\exp\left(-ir\frac{2\pi k}{n}\right) - \int_0^1 \Gamma(u)\exp(-i2\pi ur)du\right| = O\left(\frac{1}{n}\right).$$

Using the mean value theorem a crude bound for the above is $O((|r|+1)/n)$. To obtain a uniform $O(1/n)$ bound over $|r| \leq n/2$ requires a more subtle techique which we describe below.

Taking difference between the sum and integral gives

$$\begin{aligned}
& \sum_{k=1}^{n}\int_{(k-1)/n}^{k/n}\left[\Gamma(k/n)\exp\left(-ir\omega_k\right) - \Gamma(u)\left(-ir2\pi u\right)\right]du \\
= & \sum_{k=1}^{n}\int_{(k-1)/n}^{k/n}\left[\Gamma(k/n) - \Gamma(u)\right]\exp\left(-ir\omega_k\right)du + \\
& \sum_{k=1}^{n}\int_{(k-1)/n}^{k/n}\Gamma(u)\left[\exp\left(-ir\omega_k\right) - \exp\left(-ir2\pi u\right)\right]du \\
= & I_1 + I_2.
\end{aligned}$$

It is clear by the Lipschitz continuity of Γ and $|\exp(i2\pi u)| \leq 1$ that $I_1 = O(1/n)$ uniformly over all r. To obtain a similar bound for the second term we exploit the symmetries of the cos and sin functions that make up $\exp(-ir\omega_k)$.

We separate I_2 into its sin and cosine transforms

$$I_2 = I_{2,C} - iI_{2,S}$$

where

$$\begin{aligned}
I_{2,C} &= \sum_{k=1}^{n-1} \int_{(k-1)/n}^{k/n} \Gamma(u) \left[\cos\left(r\frac{2\pi k}{n}\right) - \cos\left(r 2\pi u\right)\right] du \\
I_{2,S} &= \sum_{k=1}^{n-1} \int_{(k-1)/n}^{k/n} \Gamma(u) \left[\sin\left(r\frac{2\pi k}{n}\right) - \sin\left(r 2\pi u\right)\right] du.
\end{aligned}$$

We focus on the cosine transform

$$\begin{aligned}
I_{2,C} &= \sum_{k=1}^{n-1} \int_{(k-1)/n}^{k/n} \Gamma(u) \left[\cos\left(r\frac{2\pi k}{n}\right) - \cos\left(r 2\pi u\right)\right] du \\
&= \sum_{k=1}^{n} \int_{0}^{1/n} \Gamma\left(u + \frac{k}{n}\right) \left[\cos\left(2\pi r\frac{k}{n}\right) - \cos\left(2\pi r\left[u + \frac{k}{n}\right]\right)\right] du.
\end{aligned}$$

Applying the mean value theorem to the above term would give the bound $O(|r|/n)$. Instead we turn the above integral into the differences of cosines *and* Γ. We show that the resulting product of differences cancel the unwanted $|r|$ term. We split the sum $\sum_{k=1}^{n} f_k$ into a double sum $\sum_{j=0}^{r-1}\sum_{k=1}^{n/(2r)} f_{jn/(2r)+k} + \sum_{j=0}^{r-1}\sum_{k=1}^{n/(2r)} f_{jn/(2r)+n/(2r)+k}$. This gives the double sum

$$\begin{aligned}
I_{2,C} = & \sum_{j=0}^{r-1}\sum_{k=1}^{n/(2r)} \int_0^{1/n} \Gamma\left(u + \frac{k + jn/r}{n}\right) \times \\
& \left\{\cos\left(2\pi r\frac{k + jn/r}{n}\right) - \cos\left[2\pi r\left(u + \frac{k + jn/r}{n}\right)\right]\right\} du \\
& + \sum_{j=0}^{r-1}\sum_{k=1}^{n/(2r)} \int_0^{1/n} \Gamma\left(u + \frac{k + jn/r + n/(2r)}{n}\right) \times \\
& \left[\cos\left(2\pi r\left[\frac{k + jn/r + n/(2r)}{n}\right]\right) - \cos\left(2\pi r\left[u + \frac{k + jn/r + n/(2r)}{n}\right]\right)\right] du
\end{aligned}$$

Now we use that

$$\begin{aligned}
& \cos\left(2\pi r\frac{k + jn/r + n/(2r)}{n}\right) - \cos\left(2\pi r[u + \frac{k + jn/r + n/(2r)}{n}]\right) \\
= & -\left[\cos\left(2\pi r\frac{k + jn/r}{n}\right) - \cos\left(2\pi r[u + \frac{k + jn/r}{n}]\right)\right]
\end{aligned}$$

and substitute this into the above to give

$$I_{2,C} = \sum_{j=0}^{r-1}\sum_{k=1}^{n/(2r)}\int_0^{1/n}\left[\Gamma\left(u+\frac{k+jn/r}{n}\right)-\Gamma\left(u+\frac{k+jn/r}{n}+\frac{n/(2r)}{n}\right)\right]\times \left[\cos\left(2\pi r\frac{k}{n}\right)-\cos\left(2\pi r\left[u+\frac{k}{n}\right]\right)\right]du.$$

Observe that $I_{2,C}$ is expressed as a double difference. We bound both these differences using the Lipschitz continuity of $\Gamma(\cdot)$ and $\cos(r\cdot)$; $|\Gamma(u)-\Gamma(v)|\leq \sup|\Gamma'(u)|\cdot|u-v|$ and $|\cos(ru)-\cos(rv)|\leq r|u-v|$. This yields the bound

$$I_{2,C} \leq \sup_u|\Gamma'(u)|\sum_{j=0}^{r-1}\sum_{k=1}^{n/(2r)}\frac{1}{r}\times\frac{r}{n}\times\frac{1}{n} = \sup_u|\Gamma'(u)|n^{-1}$$

which is a uniform bound for all $|r|\leq n/2$. The same bound holds for the sin transform $I_{2,S}$. Altogether, the bounds for I_1, $I_{2,C}$ and $I_{2,S}$ give

$$\sup_{\omega}\sup_{|r|\leq n/2}\left|\frac{1}{n}\sum_{k=1}^{n}\Gamma^{(a,b)}(k/n,\omega)\exp\left(ir\frac{2\pi k}{n}\right)-K_r^{(a,b)}(\omega)\right| = O(n^{-1}). \tag{94}$$

Thus for $|k_1-k_2|\leq n/2$ we have

$$[\mathbf{K}_n(\omega_{k_1},\omega_{k_2})]_{a,b} = K_{k_1-k_2}^{(a,b)}(\omega_{k_2})+O(n^{-1}).$$

For $n/2<k_1-k_2<n$ we return to (91)

$$\begin{aligned}[\mathbf{K}_n(\omega_{k_1},\omega_{k_2})]_{a,b} &= \frac{1}{n}\sum_{t=1}^{n}\exp(i(k_1-k_2)\omega_t)\Gamma^{(a,b)}\left(\frac{t}{n},\omega_{k_2}\right)+O(n^{-1}) \\ &= \frac{1}{n}\sum_{t=1}^{n}\exp(i(k_1-k_2-n)\omega_t)\Gamma^{(a,b)}\left(\frac{t}{n},\omega_{k_2}\right)+O(n^{-1}) \\ &= K_{k_1-k_2-n}^{(a,b)}(\omega_{k_2})+O(n^{-1})\end{aligned}$$

where we use that $|k_1-k_2-n|<n/2$ and (94). By a similar argument for $-n<k_1-k_2<n/2$ we have

$$[\mathbf{K}_n(\omega_{k_1},\omega_{k_2})]_{a,b} = K_{k_1-k_2+n}^{(a,b)}(\omega_{k_2})+O(n^{-1}),$$

this proves (34). $\square$

PROOF of Proposition 4.4 To prove (35) we use that under Assumption 4.2 $\Gamma^{(a,b)}(\cdot;\omega)\in L_2[0,1]$. Thus $\sum_r|K_r^{(a,b)}(\omega)|^2<\infty$, this immediately gives (35). The bound $\sup_\omega|K_r^{(a,b)}(\omega)|\leq C|r|^{-1}$ follows immediately from (93).

To prove (36) we use the mean value theorem

$$|K_r^{(a,b)}(\omega_1) - K_r^{(a,b)}(\omega_2)| \leq \sup_{\omega} |dK_r^{(a,b)}(\omega)/d\omega| \cdot |\omega_1 - \omega_2|.$$

To bound $\sup_{\omega} |dK_r^{(a,b)}(\omega)/d\omega|$ we use that

$$\left|\frac{d}{d\omega}K_r^{(a,b)}(\omega)\right| \leq \sum_{j\in\mathbb{Z}}(1+|j|)\left|\int_0^1 e^{-2\pi iru}[\mathbf{D}_j(u)]_{a,b}du\right|. \quad (95)$$

To bound the integral in the above we use integration by parts, this together with Assumption 4.2 gives

$$\left|\int_0^1 e^{-2\pi iru}D_j^{(a,b)}(u)du\right| \leq \left\{ \begin{array}{cc} C\ell(j)^{-1} & r=0 \\ C|r|^{-1}\ell(j)^{-1} & r\neq 0 \end{array}\right. .$$

Substituting this into (96) gives

$$\left|\frac{d}{d\omega}K_r^{(a,b)}(\omega)\right| = \left\{ \begin{array}{cc} C\sum_{j\in\mathbb{Z}}(1+|j|)\ell(j)^{-1} & r=0 \\ C|r|^{-1}\sum_{j\in\mathbb{Z}}(1+|j|)\ell(j)^{-1} & r\neq 0 \end{array}\right. , \quad (96)$$

this immediately leads to the required result. □

D Assumptions 2.1, 4.1 and 4.2 and the tvVAR process

We show that under certain conditions the tvVAR process satisfies Assumptions 2.1, 4.1 and 4.2. Then in Appendix D.2 we consider the inverse time-varying spectral density of a tvVAR(1) model.

D.1 Assumptions and the tvVAR

tvVAR and Assumption 2.1

We first show that Assumption 2.1 holds for the model $\underline{X}_t = \mathbf{A}(t)\underline{X}_{t-1} + \underline{\varepsilon}_t$ where $\sup_t \|\mathbf{A}(t)\| < 1-\delta$. We will show that both the largest eigenvalues of C and D are finite (which proves Assumption 2.1). We prove the result by showing the absolute sum of each row of $C_{a,b}$ and $D_{a,b}$ is bounded for each $1 \leq a,b \leq p$.

We first obtain a bound for the largest eigenvalue of C in terms of the covariances. Since $\sup_t \|\mathbf{A}(t)\| < 1-\delta$, $\underline{X}_t$ almost surely has the causal solution $\underline{X}_t = \sum_{\ell=0}^{\infty}[\prod_{j=0}^{\ell}\mathbf{A}(t-j)]\underline{\varepsilon}_{t-\ell}$ Using this expansion and $\sup_t \|\mathbf{A}(t)\| < 1-\delta$ it is easily shown that $|\,\mathbb{C}\mathrm{ov}[X_t^{(a)}, X_\tau^{(b)}]| \leq K(1-\delta)^{|t-\tau|}$ for some finite constant K. Thus by using Gerschgorin Circle Theorem we have

$$\lambda_{\sup}(C) \leq Kp\sum_{r\in\mathbb{Z}}(1-\delta)^{|r|}.$$

Next we show that $\lambda_{\sup}(D) < \infty$. Under $\sup_t \|\mathbf{A}(t)\| < 1-\delta$, the rows of $\mathbf{A}(t)$ are such that

$$\sup_{t,a} \|[\mathbf{A}(t)]_{a\cdot}\|_1 \leq p^{1/2} \sup_{t,a} \|[\mathbf{A}(t)]_{a\cdot}\|_2 \leq p^{1/2}(1-\delta).$$

Therefore, by using the above, the representation of $D_{a,b}$ in (46) together with Gerschgorin Circle Theorem we have $\lambda_{\sup}(D) < \infty$. Thus Assumptions 2.1 is satisfied.

We mention that $\sup_t \|\mathbf{A}(t)\| < 1-\delta$ is a sufficient condition. It can be relaxed to allow for a contraction on the spectral radius of $\mathbf{A}(t)$ and smoothness conditions on $\mathbf{A}(t)$ (see Künsch (1995)). The above result can be extended to finite order tvVAR(d) models, by rewriting the p-dimensional tvVAR(d) model as a pd-dimension tvVAR(1) model and placing similar conditions on the corresponding tvVAR(1) matrix.

tvVAR and Assumption 4.1

Suppose that $\{\underline{X}_t\}$ has a tvVAR(∞) representation where $\sup_t \|\mathbf{A}_j(t)\|_2 \leq \ell(j)^{-1}$ and $\{\ell(j)\}$ is a monotonically increasing sequence as $|j| \to \infty$. If $\sum_{j\in\mathbb{Z}} |j|^{K+1}\ell(j)^{-1} < \infty$ (for some $K \geq 1$), then we show below that Assumption 4.1 is satisfied.

To show this we require the following lemma.

Lemma D.1 *Suppose $\ell(j)^{-1}$ is monotonically decreasing as $|j| \to \infty$ with $\sum_{j\in\mathbb{Z}} |j|^K \ell(j)^{-1} < \infty$ (for $K \geq 2$). Then for all $r \in \mathbb{Z}$*

$$\sum_{s=-\infty}^{\infty} \frac{1}{\ell(s)\ell(s+r)} \leq \frac{1}{\widetilde{\ell}(r)} \text{ where } \widetilde{\ell}(j) = \left[3\frac{1}{\ell(\lfloor |j|/2 \rfloor)} \sum_{s\in\mathbb{Z}} \frac{1}{\ell(s)}\right]^{-1}$$

and $\sum_{j\in\mathbb{Z}} |j|^K \widetilde{\ell}(j)^{-1} < \infty$.

PROOF. Without loss of generality we prove the result for $r \geq 0$. We partition the sum $\sum_{s=-\infty}^{\infty}$ into three terms

$$\begin{aligned}
\sum_{s=-\infty}^{\infty} \frac{1}{\ell(s)\ell(s+r)} &= \sum_{s=0}^{\infty} \frac{1}{\ell(s)\ell(s+r)} + \sum_{s=-\infty}^{-r} \frac{1}{\ell(s)\ell(s+r)} + \sum_{s=-r+1}^{-1} \frac{1}{\ell(s)\ell(s+r)} \\
&= II_1 + II_2 + II_3.
\end{aligned}$$

Using that $1/\ell(j)$ is monotonically decreasing as $j \to \infty$ is is easily seen that

$$II_1 \leq \frac{1}{\ell(r)} \sum_{s=0}^{\infty} \frac{1}{\ell(s)} \text{ and } II_2 \leq \frac{1}{\ell(r)} \sum_{s=-\infty}^{0} \frac{1}{\ell(s)}. \tag{97}$$

To bound II_3 we use that for $-r/2 \leq s \leq -1$ that $\ell(s) \leq \ell(\lfloor r/2 \rfloor)$ and for $-r+1 \leq s \leq$

$-r/2$ then $\ell(s+r) \leq \ell(\lfloor r/2 \rfloor)$. Altogether this gives the bound

$$II_3 \leq \left[\sum_{s=-r/2}^{-1} + \sum_{s=-r+1}^{-r/2}\right] \frac{1}{\ell(s)} \frac{1}{\ell(s+r)} \leq \frac{1}{\ell(\lfloor r/2 \rfloor)} \sum_{s \in \mathbb{Z}} \frac{1}{\ell(s)}.$$

The above bound together with (97) (noting that $\ell(r) > \ell(\lfloor r/2 \rfloor)$) gives

$$\sum_{s=-\infty}^{\infty} \frac{1}{\ell(s)\ell(s+r)} \leq 3 \frac{1}{\ell(\lfloor (r/2 \rfloor)} \sum_{s \in \mathbb{Z}} \frac{1}{\ell(s)}. \tag{98}$$

For all j define

$$\widetilde{\ell}(j) = \left[3 \frac{1}{\ell(\lfloor |j|/2 \rfloor)} \sum_{s \in \mathbb{Z}} \frac{1}{\ell(s)}\right]^{-1}.$$

Then from (98) we have the bound

$$\sum_{s=-\infty}^{\infty} \frac{1}{\ell(s)\ell(s+r)} \leq \frac{1}{\widetilde{\ell}(r)}.$$

Since by assumption $\sum_{j \in \mathbb{Z}} |j|^K \ell(|j|)^{-1} < \infty$, it is immediately clear from the definition of $\widetilde{\ell}(j)$ that $\sum_{j \in \mathbb{Z}} |j|^K \widetilde{\ell}(j)^{-1} < \infty$. This proves the result. $\square$

Lemma D.2 *Suppose $\underline{X}_t$ has a tvVAR(∞) representation that satisfies (44) and $\mathbf{D}_{t,\tau}$ be defined as in (45). If the time-varying AR matrices satisfy $\sup_t \|\mathbf{A}_j(t)\|_2 \leq \ell(j)^{-1}$ where $\ell(j)^{-1}$ is monotonically decreasing as $|j| \to \infty$ and $\sum_{j \in \mathbb{Z}} |j|^K \ell(j)^{-1} < \infty$, then $\sup_t \|\mathbf{D}_{t,t+j}\|_1 \leq \widetilde{\ell}(j)^{-1}$ where $\sup_t \sum_{j \neq 0} |j|^K \widetilde{\ell}(j)^{-1} < \infty$.*

PROOF By using (45) we have

$$\begin{aligned} \|\mathbf{D}_{t,t+j}\|_1 &\leq \sup_t \sum_{s=-\infty}^{\infty} \left\| \widetilde{\mathbf{A}}_s(t+s)' \widetilde{\mathbf{A}}_{j+s}(t+s) \right\|_1 \\ &\leq \sum_{s=-\infty}^{\infty} \left\| \widetilde{\mathbf{A}}_s(t+s) \right\|_2 \left\| \widetilde{\mathbf{A}}_{j+s}(t+s) \right\|_2 \leq \sum_{s=-\infty}^{\infty} \frac{1}{\ell(s)\ell(s+j)}. \end{aligned}$$

Finally, from the above and by using Lemma D.1 we have

$$\|\mathbf{D}_{t,t+j}\|_1 \leq \widetilde{\ell}(j)^{-1},$$

this proves the result. $\square$

tvVAR and Assumption 4.2

We now show that under certain conditions on $\{\mathbf{A}_j(t)\}$ the tvVAR process satisfies Assumption 4.2. Define the matrices $\mathbf{A}_j : [0,1] \to \mathbb{R}^{p\times p}$, which are Lipschitz in the sense that

$$\|\mathbf{A}_j(u) - \mathbf{A}_j(v)\|_1 \leq \frac{1}{\ell(j)}|u-v| \tag{99}$$

where $\ell(|j|)^{-1}$ is monotonically decreasing as $|j| \to \infty$ with $\sum_{j\in\mathbb{Z}} j^2 \ell(j)^{-1} < \infty$ and

$$\sup_u \|\mathbf{A}_j(u)\|_1 \leq \ell(j)^{-1}. \tag{100}$$

Following Dahlhaus (2000a) we define the locally stationary tvVAR model as

$$\underline{X}_{t,n} = \sum_{j=1}^{\infty} \mathbf{A}_j\left(\frac{t}{n}\right)\underline{X}_{t-j,n} + \underline{\varepsilon}_t, \tag{101}$$

where $\{\underline{\varepsilon}_t\}_t$ are i.i.d random variables with $\mathbb{V}\mathrm{ar}[\underline{\varepsilon}_t] = \mathbf{\Sigma}$ $(0 < \lambda_{\min}(\mathbf{\Sigma}) \leq \lambda_{\max}(\mathbf{\Sigma}) < \infty)$. To define the suitable $\mathbf{D}_j(u)$ (as given in Assumption 4.2), we first define the auxillary, stationary process corresponding to $\underline{X}_{t,n}$;

$$\underline{X}_t(u) = \sum_{j=1}^{\infty} \mathbf{A}_j(u)\underline{X}_{t-j}(u) + \underline{\varepsilon}_t. \tag{102}$$

The inverse covariance of $\{\underline{X}_t(u)\}_t$ is $D(u) = (D_{a,b}(u); 1 \leq a,b \leq p)$ with $[D_{a,b}(u)]_{t,\tau} = [\mathbf{D}_{t-\tau}(u)]_{a,b}$ and

$$\mathbf{D}_{t-\tau}(u) = \sum_{\ell=-\infty}^{\infty} \widetilde{\mathbf{A}}_\ell(u)' \, \Sigma^{-1} \widetilde{\mathbf{A}}_{\ell+(\tau-t)}(u). \tag{103}$$

The spectral density matrix corresponding to $\{\underline{X}_t(u)\}_t$ is

$$\mathbf{\Sigma}(u;\omega) = [I_p - \sum_{j=1}^{\infty} \mathbf{A}_j(u)\exp(-ij\omega)]^{-1}\mathbf{\Sigma}([I_p - \sum_{j=1}^{\infty} \mathbf{A}_j(u)\exp(-ij\omega)]^{-1})^*. \tag{104}$$

Thus the time-varying spectral precision matrix associated with $\{X_{t,n}\}_t$ is $\mathbf{\Gamma}(t/n;\omega) = \mathbf{\Sigma}(t/n;\omega)^{-1}$ (see Section 4.3). In the following lemma we show that the time series $\{\underline{X}_{t,n}\}_t$ satisfies Assumption 4.2.

Lemma D.3 *Suppose that the time series $\{\underline{X}_{t,n}\}_t$ has the representation in (101), where the tvVAR matrices satisfy conditions (99) and (100). Let $\mathbf{D}_{t,\tau}$ and $\mathbf{D}_{t-\tau}(u)$ be defined* *as in (45) and (103). Then*

$$\left\|\mathbf{D}_{t,\tau}-\mathbf{D}_{t-\tau}\left(\frac{t+\tau}{2n}\right)\right\|_1 \leq \frac{|t-\tau|+1}{n\widetilde{\ell}(t-\tau)} \text{ and } \|\mathbf{D}_{t-\tau}(u)-\mathbf{D}_{t-\tau}(v)\|_1 \leq \frac{|u-v|}{n\widetilde{\ell}(t-\tau)} \quad (105)$$

and $\sup_u \sum_{j\in\mathbb{Z}} j^2\|\mathbf{D}_j(u)\|_1 < \infty$*, where* $\widetilde{\ell}(|j|)$ *is monotonically increasing as* $|j| \to \infty$ *and* $\sum_{j\in\mathbb{Z}} j^2 \cdot \widetilde{\ell}(|j|)^{-1} < \infty$*.*

Further, if $\sup_u \|d\mathbf{A}_j(u)/du\|_1 \leq 1/\ell(j)$ *then*

$$\sup_u \left\|\frac{d\mathbf{D}_{t-\tau}(u)}{du}\right\|_1 = \frac{1}{\widetilde{\ell}(t-\tau)}. \quad (106)$$

PROOF We first prove (105). We recall that

$$\mathbf{D}_{t,\tau} = \sum_{\ell=-\infty}^{\infty} \widetilde{\mathbf{A}}_{\ell}\,(t+s)'\,\Sigma^{-1}\widetilde{\mathbf{A}}_{(\tau-t)+\ell}(t+s),$$

and

$$\|\mathbf{A}_j(u)-\mathbf{A}_j(v)\|_1 \;\; \leq \;\; \frac{1}{\ell(j)}|u-v| \text{ and } \sup_u \|\mathbf{A}_j(u)\|_1 \leq \ell(j)^{-1}. \quad (107)$$

To simplify the notation (the proof does not change), we set $\mathbf{H} = I_p$. Using the above and evaluating the difference gives

$$\mathbf{D}_{t,\tau}-\mathbf{D}_{t-\tau}(u) \;\; = \;\; \sum_{s=-\infty}^{\infty}\Big(\widetilde{\mathbf{A}}_s((t+s)/n)'\widetilde{\mathbf{A}}_{s+(\tau-t)}((t+s)/n)-\widetilde{\mathbf{A}}_s(u)'\widetilde{\mathbf{A}}_{s+(\tau-t)}(u)\Big) = I_1+I_2,$$

where

$$I_1 \;\; = \;\; \sum_{s=-\infty}^{\infty}\Big(\widetilde{\mathbf{A}}_s((t+s)/n)'\widetilde{\mathbf{A}}_{s+(\tau-t)}((t+s)/n)-\widetilde{\mathbf{A}}_s((t+s)/n)'\widetilde{\mathbf{A}}_{s+(\tau-t)}(u)\Big)$$

$$I_2 \;\; = \;\; \sum_{s=-\infty}^{\infty}\Big(\widetilde{\mathbf{A}}_s((t+s)/n)'\widetilde{\mathbf{A}}_{s+(\tau-t)}(u)-\widetilde{\mathbf{A}}_s(u)'\widetilde{\mathbf{A}}_{s+(\tau-t)}(u)\Big).$$

The two bounds are very similar, we focus on obtaining a bound for I_1. By the Cauchy-Schwarz inequality and that $\|\cdot\|_2 \leq \|\cdot\|_1$ we have

$$\begin{aligned}
\|I_1\|_1 &\leq \sum_{s=-\infty}^{\infty} \left\| \widetilde{\mathbf{A}}_s((t+s)/n)' \left[\widetilde{\mathbf{A}}_{s+(\tau-t)}((t+s)/n) - \widetilde{\mathbf{A}}_{s+(\tau-t)}(u) \right] \right\|_1 \\
&\leq \sum_{s=-\infty}^{\infty} \left\| \widetilde{\mathbf{A}}_s(t/n) \right\|_2 \left\| \widetilde{\mathbf{A}}_{s+(\tau-t)}((t+s)/n) - \widetilde{\mathbf{A}}_{s+(\tau-t)}(u) \right\|_2 \\
&\leq \sum_{s=-\infty}^{\infty} \left\| \widetilde{\mathbf{A}}_s(t/n) \right\|_1 \left\| \widetilde{\mathbf{A}}_{s+(\tau-t)}((t+s)/n) - \widetilde{\mathbf{A}}_{s+(\tau-t)}(u) \right\|_1
\end{aligned}$$

Substituting the bounds for $\mathbf{A}_j(u)$ given in (107) into the above we have

$$\|I_1\|_1 \leq \sum_{s=-\infty}^{\infty} \frac{1}{\ell(s)} \frac{1}{\ell(s+\tau-t)} \left| \frac{t+s}{n} - u \right|.$$

By the same argument we have

$$\|I_2\|_1 \leq \sum_{s=-\infty}^{\infty} \frac{1}{\ell(s)} \frac{1}{\ell(s+\tau-t)} \left| \frac{t+s}{n} - u \right|.$$

Thus

$$\|\mathbf{D}_{t,\tau} - \mathbf{D}_{t-\tau}(u)\|_1 \leq 2 \sum_{s=-\infty}^{\infty} \frac{1}{\ell(s)} \frac{1}{\ell(s+\tau-t)} \cdot \left|\frac{t+s}{n} - u\right|.$$

Setting $u = (t+\tau)/(2n)$ gives

$$\begin{aligned}
\left\| \mathbf{D}_{t,\tau} - \mathbf{D}_{t-\tau}\left(\frac{t+\tau}{2n}\right) \right\|_1 &\leq \sum_{s=-\infty}^{\infty} \frac{1}{\ell(s)} \frac{1}{\ell(s+\tau-t)} \left(\frac{|\tau-t|}{2n} + \frac{|s|}{n} \right) \\
&= \frac{|t-\tau|}{2n} \sum_{s=-\infty}^{\infty} \frac{1}{\ell(s)} \frac{1}{\ell(s+\tau-t)} + \frac{1}{n} \sum_{s=-\infty}^{\infty} \frac{|s|}{\ell(s)} \frac{1}{\ell(s+\tau-t)}
\end{aligned}$$

Now by using Lemma D.1 we have we have the bound

$$\left\| \mathbf{D}_{t,\tau} - \mathbf{D}_{t-\tau}\left(\frac{t+\tau}{2n}\right) \right\|_1 \leq \frac{|t-\tau|}{2n\widetilde{\ell}(t-\tau)} + \frac{1}{n\widetilde{\ell}(t-\tau)}. \tag{108}$$

Since by assumption $\sum_{j\in\mathbb{Z}}(j^2+1)/\ell(|j|) < \infty$, it is immediately clear from the definition of $\widetilde{\ell}(j)$ that $\sum_{j\in\mathbb{Z}}(j^2+1)\widetilde{\ell}(j)^{-1} < \infty$. Under the stated assumptions in (107) and using

Lemma D.1 we can show

$$\|\mathbf{D}_{t-\tau}(u) - \mathbf{D}_{t-\tau}(v)\|_1 \leq \frac{|u-v|}{n\widetilde{\ell}(t-\tau)} \tag{109}$$

and $\sup_u \sum_{j\in\mathbb{Z}} j^2 \|\mathbf{D}_j(u)\|_1 < \infty$. (108) and (109) together prove (105).

We now prove (106). The elementwise derivative of $\mathbf{D}_j(u)$ is

$$\frac{d}{du}\mathbf{D}_j(u) = \sum_{s=-\infty}^{\infty} \frac{d}{du}\widetilde{\mathbf{A}}_s(u)' \widetilde{\mathbf{A}}_{(\tau-t)+s}(u) + \sum_{s=-\infty}^{\infty} \widetilde{\mathbf{A}}_s(u)' \frac{d}{du}\widetilde{\mathbf{A}}_{(\tau-t)+s}(u).$$

Using the conditions in (107), $\sup_u \|d\mathbf{A}_j(u)/du\|_1 \leq 1/\ell(j)$ and Lemma D.1 we can show that

$$\left\|\frac{d}{du}\mathbf{D}_j(u)\right\|_1 \leq 2\sum_{s\in\mathbb{Z}} \frac{1}{\ell(s)\ell(s+j)},$$

this gives (106). $\square$

D.2 Example: Locally stationary time-varying VAR(1)

We consider the locally stationary time-varying VAR(1) model

$$\underline{X}_{t,n} = \mathbf{A}\left(\frac{t}{n}\right)\underline{X}_{t-1,n} + \underline{\varepsilon}_t \quad t \in \mathbb{Z},$$

where $\{\underline{\varepsilon}_t\}$ are i.i.d. with $\mathbb{V}\mathrm{ar}[\underline{\varepsilon}_t] = I_p$. We assume that the matrix $\mathbf{A}(u)$ satisfies (99) and (100). Using (104) the time-varying spectral precision matrix corresponding to $\{\underline{X}_{t,n}\}$ is $\mathbf{\Gamma}(u;\omega) = [I_p - \mathbf{A}(u)\exp(-i\omega)]^*[I_p - \mathbf{A}(u)\exp(-i\omega)]$. We partition $\mathbf{\Gamma}(u;\omega)$ into the conditional stationary and nonstationary matrices.

Define the set $\mathcal{S} \subseteq \{1,\ldots,p\}$ where for all $a \in \mathcal{S}$ the columns $[\mathbf{A}(t)]_{\cdot,a}$ do not depend on t. Then $\{X_t^{(a)}; a \in \mathcal{S}\}$ is a conditionally stationary subgraph and $D_{\mathcal{S}} = \{D_{a,b}; a,b \in \mathcal{S}\}$ is a block Toeplitz matrix (see Corollary 2.1). By Lemma 3.1 the integral kernel associated with $D_{\mathcal{S}}$ is $\mathbf{\Gamma}_{\mathcal{S},\mathcal{S}}(\omega)\delta_{\omega,\lambda}$. We obtain an expression for $\mathbf{\Gamma}_{\mathcal{S},\mathcal{S}}(\omega)$ below.

We denote the set $\mathcal{S}$ as $\mathcal{S} = \{a_1,\ldots,a_{|\mathcal{S}|}\}$, where $|\mathcal{S}|$ denotes the cardinality of $\mathcal{S}$. Define the $p\times|\mathcal{S}|$ matrix $\mathbf{A}_{\mathcal{S}}$ where $\mathbf{A}_{\mathcal{S}} = ([\mathbf{A}_{\mathcal{S}}]_{\cdot,r} = [\mathbf{A}(0)]_{\cdot,a_r}; a_r \in \mathcal{S}, 1 \leq r \leq |\mathcal{S}|)$, let $I_{p,|\mathcal{S}|}$ denote the $p\times|\mathcal{S}|$ "indicator" matrix which is comprised of zeros except at the entries $\{(r,a_r); a_r \in \mathcal{S}\}$ where $[I_{p,|\mathcal{S}|}]_{r,a_r} = 1$. Then

$$\mathbf{\Gamma}_{\mathcal{S},\mathcal{S}}(\omega) = \left[I_{p,|\mathcal{S}|} - \mathbf{A}_{\mathcal{S}}\exp(-i\omega)\right]^* \left[I_{p,|\mathcal{S}|} - \mathbf{A}_{\mathcal{S}}\exp(-i\omega)\right]. \tag{110}$$

Using $\mathbf{\Gamma}_{\mathcal{S},\mathcal{S}}(\omega)$ we can deduce the the partial spectral coherence for conditionally stationary nodes and pairs (see Lemma 3.3).

We now obtain the nonstationary submatrices in $\mathbf{\Gamma}(u;\omega)$. Let $\mathcal{S}' = \{b_1,\ldots,b_{|\mathcal{S}'|}\}$ denote the complement of $\mathcal{S}$. Analogous to $\mathbf{A}_{\mathcal{S}}$ and $I_{p,|\mathcal{S}|}$, we define the $p\times|\mathcal{S}'|$ dimensional matrices $\mathbf{A}_{\mathcal{S}'}(u) = ([\mathbf{A}_{\mathcal{S}'}(u)]_{\cdot,r} = [\mathbf{A}(u)]_{\cdot,b_r}; b_r \in \mathcal{S}', 1 \leq r \leq |\mathcal{S}_r|)$ and $I_{p\times|\mathcal{S}'|}$ which is comprised of zeros except at the entries $\{(r,b_r); b_r \in \mathcal{S}'\}$ where $[I_{p,|\mathcal{S}'|}]_{r,b_r} = 1$. A rearranged version of $\boldsymbol{\Gamma}(u;\omega)$ (which for simplicity we call $\boldsymbol{\Gamma}(u;\omega)$) is

$$\boldsymbol{\Gamma}(u;\omega) = \begin{pmatrix} \Gamma_{\mathcal{S},\mathcal{S}}(\omega) & \Gamma_{\mathcal{S},\mathcal{S}'}(u;\omega) \\ \Gamma_{\mathcal{S},\mathcal{S}'}(u;\omega)^* & \Gamma_{\mathcal{S}',\mathcal{S}'}(u;\omega) \end{pmatrix},$$

where

$$\begin{aligned} \Gamma_{\mathcal{S},\mathcal{S}}(\omega) &= \left[I_{p,|\mathcal{S}|} - \mathbf{A}_{\mathcal{S}}\exp(-i\omega)\right]^* \left[I_{p,|\mathcal{S}|} - \mathbf{A}_{\mathcal{S}}\exp(-i\omega)\right], \\ \Gamma_{\mathcal{S},\mathcal{S}'}(u;\omega) &= \left[I_{p,|\mathcal{S}|} - \mathbf{A}_{\mathcal{S}}\exp(-i\omega)\right]^* \left[I_{p,|\mathcal{S}'|} - \mathbf{A}_{\mathcal{S}'}(u)\exp(-i\omega)\right] \\ \text{and } \Gamma_{\mathcal{S}',\mathcal{S}'}(u;\omega) &= \left[I_{p,|\mathcal{S}'|} - \mathbf{A}_{\mathcal{S}'}(u)\exp(-i\omega)\right]^* \left[I_{p,|\mathcal{S}'|} - \mathbf{A}_{\mathcal{S}'}(u)\exp(-i\omega)\right]. \end{aligned}$$

Using the above we can deduce $K_r^{(a,b)}(\omega)$ and thus approximations to the entries of $\mathbf{K}_n(\omega_{k_1},\omega_{k_2})$. The system for Example 2.1 is described in detail in Appendix E.

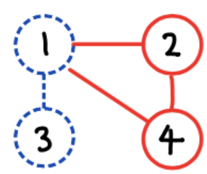
E Examples

In the following two sections we study the running time-varying AR(1) example (introduced in Example 2.1). In Appendix E.3 we compare the nonstationary graph of a piecewise stationary time series with its piecewise stationary gaphs.

E.1 The tvVAR model and corresponding local spectral precision matrix

For all $1 \leq t \leq n$ the model is defined as

$$\begin{pmatrix} X_t^{(1)} \\ X_t^{(2)} \\ X_t^{(3)} \\ X_t^{(4)} \end{pmatrix} = \begin{pmatrix} \alpha_1(t/n) & 0 & \alpha_3 & 0 \\ \beta_1 & \beta_2 & 0 & \beta_4 \\ 0 & 0 & \gamma_3(t/n) & 0 \\ 0 & \nu_2 & 0 & \nu_4 \end{pmatrix} \begin{pmatrix} X_{t-1}^{(1)} \\ X_{t-1}^{(2)} \\ X_{t-1}^{(3)} \\ X_{t-1}^{(4)} \end{pmatrix} + \begin{pmatrix} \varepsilon_t^{(1)} \\ \varepsilon_t^{(2)} \\ \varepsilon_t^{(3)} \\ \varepsilon_t^{(4)} \end{pmatrix}$$



and $\{\underline{\varepsilon}_t\}_t$ are iid random vectors with $\mathbb{V}\mathrm{ar}[\underline{\varepsilon}_t]$ and $\alpha_1(\cdot), \gamma_3(\cdot) \in L_2[0,1]$ and are Lipschitz continuous. By using the results in Section 5 we obtain the network on the right.
We now obtain the time-varying conditional spectral density $\boldsymbol{\Gamma}(u;\omega)$. Let

$$\begin{aligned} [I_{2,4} - \mathbf{A}_{\mathcal{S}}\exp(-i\omega)] &= \begin{pmatrix} 0 & 0 \\ 1-\beta_2 e^{-i\omega} & -\beta_4 e^{-i\omega} \\ 0 & 0 \\ -\nu_2 e^{-i\omega} & 1-\nu_4 e^{-i\omega} \end{pmatrix} \\ \text{and } [I_{1,3} - \mathbf{A}_{\mathcal{S}'}(u)\exp(-i\omega)] &= \begin{pmatrix} 1-\alpha_1(u)e^{-i\omega} & -\alpha_3 e^{-i\omega} \\ -\beta_1 e^{-i\omega} & 0 \\ 0 & 1-\gamma_3(u)e^{-i\omega} \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

Then

$$\mathbf{\Gamma}(u;\omega) = \begin{pmatrix} \Gamma_{(2,4),(2,4)}(\omega) & \Gamma_{(2,4),(1,3)}(u;\omega) \\ \Gamma_{(2,4),(1,3)}(u;\omega)^{*} & \Gamma_{(1,3),(1,3)}(u;\omega) \end{pmatrix}$$

where

$$\begin{aligned}
\Gamma_{(2,4),(2,4)}(\omega) &= [I_{2,4} - \mathbf{A}_{\mathcal{S}} \exp(-i\omega)]^{*} [I_{2,4} - \mathbf{A}_{\mathcal{S}} \exp(-i\omega)] \\
&= \begin{pmatrix} |1-\beta_2 e^{-i\omega}|^2 + \nu_2^2 & -\beta_4 e^{i\omega}(1-\beta_2 e^{-i\omega}) - \nu_2 e^{-i\omega}(1-\nu_4 e^{i\omega}) \\ -\beta_4 e^{-i\omega}(1-\beta_2 e^{i\omega}) - \nu_2 e^{i\omega}(1-\nu_4 e^{-i\omega}) & |1-\nu_4 e^{-i\omega}|^2 + \beta_4^2 \end{pmatrix} \\
\Gamma_{(2,4),(1,3)}(u;\omega) &= [I_{2,4} - \mathbf{A}_{\mathcal{S}} \exp(-i\omega)]^{*} [I_{1,3} - \mathbf{B}_{\mathcal{S}'}(u) \exp(-i\omega)] \\
&= \begin{pmatrix} -\beta_1 e^{i\omega}[1-\beta_2 e^{-i\omega}] & 0 \\ \beta_1\beta_4 & 0 \end{pmatrix} \\
\Gamma_{(1,3),(1,3)}(u;\omega) &= [I_{1,3} - \mathbf{A}_{\mathcal{S}'}(u) \exp(-i\omega)]^{*} [I_{1,3} - \mathbf{A}_{\mathcal{S}'}(u) \exp(-i\omega)] \\
&= \begin{pmatrix} |1-\alpha_1(u) e^{-i\omega}|^2 + \beta_1^2 & -\alpha_3 e^{i\omega}[1-\alpha_1(u) e^{-i\omega}] \\ -\alpha_3 e^{-i\omega}[1-\alpha_1(u) e^{i\omega}] & |1-\gamma_3(u) e^{-i\omega}|^2 + \alpha_3^2 \end{pmatrix}.
\end{aligned}$$

E.2 The partial spectral coherence

Based on the results in Section 3.3, we use $\Gamma_{(2,4),(2,4)}(\omega)$ to define the partial spectral coherence for the conditionally stationary nodes and edges (2 and 4). We observe that $\Gamma_{(2,4),(2,4)}(\omega)$ resembles the spectral density matrix of a stationary vector moving average model of order one (or equivalently the inverse of a vector autoregressive of order one). Using $\Gamma_{(2,4),(2,4)}(\omega)$, the partial spectra for the conditionally stationary nodes 2 and 4 are

$$\Gamma^{(2,2)}(\omega)^{-1} = \frac{1}{|1-\beta_2 e^{-i\omega}|^2 + \nu_2^2} \text{ and } \Gamma^{(4,4)}(\omega)^{-1} = \frac{1}{|1-\nu_4 e^{-i\omega}|^2 + \beta_4^2}.$$

Furthermore, by using (19), the partial spectral coherence for the conditionally stationary edge $(2,4)$ is

$$R_{2,4}(\omega) = -\frac{-\beta_4 e^{i\omega}(1-\beta_2 e^{-i\omega}) - \nu_2 e^{-i\omega}(1-\nu_4 e^{i\omega})}{\sqrt{\Gamma^{(2,2)}(\omega)\Gamma^{(4,4)}(\omega)}}$$

E.3 A comparision of networks in a piecewise stationary VAR(1) model

Below we consider both the directed and undirected graphs for a piece-wise stationary VAR(1) model where $p = 4$. We suppose that $\underline{X}_t$ is piece-wise stationary in the sense that

$$\underline{X}_t = \begin{cases} A_1 \underline{X}_{t-1} + \underline{\varepsilon}_t & 1 \leq t \leq n/2 \\ A_2 \underline{X}_{t-1} + \underline{\varepsilon}_t & n/2+1 \leq t \leq n \end{cases}$$

where $\mathbb{V}\mathrm{ar}[\underline{\varepsilon}_t] = I_3$ and $\{\underline{\varepsilon}_t\}_t$ are random vectors,

$$A_1 = \begin{pmatrix} \alpha_1 & 0 & 0 & 0 \\ \beta_1 & \beta_2 & 0 & \beta_4 \\ 0 & 0 & \gamma_3 & 0 \\ 0 & \nu_2 & 0 & \nu_4 \end{pmatrix} \quad \text{and} \quad A_2 = \begin{pmatrix} \widetilde{\alpha}_1 & 0 & \alpha_3 & 0 \\ \beta_1 & \beta_2 & 0 & \beta_4 \\ 0 & 0 & \widetilde{\gamma}_3 & 0 \\ 0 & \nu_2 & 0 & \nu_4 \end{pmatrix}.$$

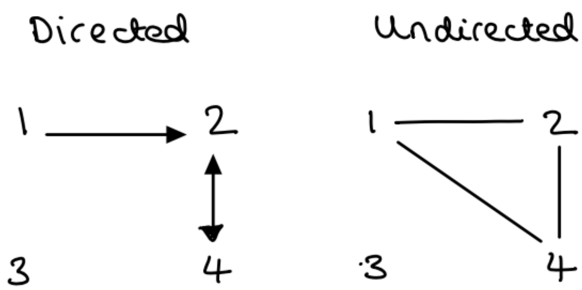
note that $\alpha_1 \neq \widetilde{\alpha}_1$ and $\gamma_3 \neq \widetilde{\gamma}_3$. For $1 \leq t \leq n$ the above model can be written as the nonstationary model

$$\begin{pmatrix} X_t^{(1)} \\ X_t^{(2)} \\ X_t^{(3)} \\ X_t^{(4)} \end{pmatrix} = \begin{pmatrix} \alpha(t) & 0 & \alpha_3(t) & 0 \\ \beta_1 & \beta_2 & 0 & \beta_4 \\ 0 & 0 & \gamma(t) & 0 \\ 0 & \nu_2 & 0 & \nu_4 \end{pmatrix} \begin{pmatrix} X_{t-1}^{(1)} \\ X_{t-1}^{(2)} \\ X_{t-1}^{(3)} \\ X_{t-1}^{(4)} \end{pmatrix} + \underline{\varepsilon}_t = A(t)\underline{X}_{t-1} + \underline{\varepsilon}_t$$

Below we will give the directed and undirected graphs for each component (this is in the stationary and can be deduced from standard results and definitions in the literature; see the definition of moralized k-complex, Figures 2 and 3 in Andersson et al. (2001), and section 3.3 in Dahlhaus and Eichler (2003b)), then we give the combined nonstationary graph.

For $t \in \{1, \dots, n/2\}$ the graph is based on

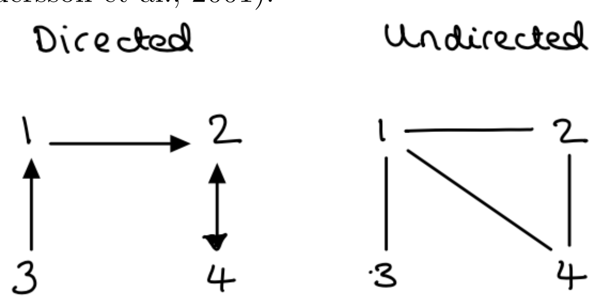
$$A_1 = \begin{pmatrix} \alpha_1 & 0 & 0 & 0 \\ \beta_1 & \beta_2 & 0 & \beta_4 \\ 0 & 0 & \gamma_3 & 0 \\ 0 & \nu_2 & 0 & \nu_4 \end{pmatrix}$$



Observe that the undirected graph adds an additional edge (1,4), and converts the arrows in the directed graphs to lines. This is the process of *moralizing* a k-complex ($k = 1$) in the original directed graph (Andersson et al., 2001).

For $t \in \{n/2 + 1, \dots, n\}$ the graph is based on

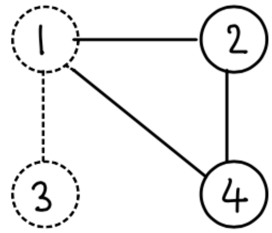
$$A_2 = \begin{pmatrix} \widetilde{\alpha}_1 & 0 & \alpha_3 & 0 \\ \beta_1 & \beta_2 & 0 & \beta_4 \\ 0 & 0 & \widetilde{\gamma}_3 & 0 \\ 0 & \nu_2 & 0 & \nu_4 \end{pmatrix}$$



On the other hand the nonstationary undirected graph is

For $t \in \{1, \ldots, n/2\}$ the graph is based on

$$A(t) = \begin{pmatrix} \alpha(t) & 0 & \alpha_3(t) & 0 \\ \beta_1 & \beta_2 & 0 & \beta_4 \\ 0 & 0 & \gamma(t) & 0 \\ 0 & \nu_2 & 0 & \nu_4 \end{pmatrix}$$



The two stationary graphs are what we mean by local information. Comparing the two stationary undirected graphs with the nonstationary directed graphs we observe that the nonstationary directed graph is summarising all the information in the two stationary undirected graphs. This is what we mean by global information, it tells us that something is changing in nodes (1) and (3) (which correspond to columns (1) and (3) in the transition matrix).

F Connection to graphical models for stationary time series

We now apply the results above to stationary multivariate time series. This gives an alternative derivation for the partial spectral coherency of stationary time series (see Brillinger (1996) and Dahlhaus (2000b)) which is usually based on the Wiener filter.

Suppose that $\{\underline{X}_t\}_t$ is a p-dimension second order stationary time series, with spectral density matrix $\mathbf{\Sigma}(\omega) = \sum_{r\in\mathbb{Z}} \mathbf{C}_r \exp(-ir\omega)$. By Lemmas 3.1 and 2.1 $\mathbf{C}(\omega,\lambda)$ and $\mathbf{K}(\omega,\lambda)$ are diagonal kernels where $\mathbf{C}(\omega,\lambda) = \mathbf{\Sigma}(\omega)\delta_{\omega,\lambda}$ and $\mathbf{K}(\omega,\lambda) = \mathbf{\Gamma}(\omega)\delta_{\omega,\lambda}$ where $\mathbf{\Gamma}(\omega) = \mathbf{\Sigma}(\omega)^{-1}$. Let $\Gamma^{(a,b)}(\omega)$ denote the (a,b)th entry of $\mathbf{\Gamma}(\omega)$. Our aim is to interprete the entries of $\mathbf{\Gamma}(\omega)$ in terms of the partial correlation and partial spectral coherence. We keep in mind that since $\{\underline{X}_t\}_t$ is second order stationary time series all the nodes and edges of its corresponding network are conditionally stationary.

We first interprete $\Gamma^{(a,b)}(\omega)$. Under stationarity for all t and τ and (a,b) we have

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

Let $\sigma_a^2 = \mathbb{V}\mathrm{ar}[X_0^{(a)} - P_{\mathcal{H}-(X_0^{(a)})}(X_0^{(a)})]$. By using Lemma 2.2 for all $(t,a) \neq (\tau,b)$

$$[\mathbf{D}_{a,b}]_{t,\tau} = \begin{cases} \frac{1}{\sigma_a^2} & t = \tau \text{ and } a = b \\ -\frac{1}{\sigma_a\sigma_b}\phi_{t-\tau}^{(a,b)} & \text{otherwise} \end{cases}$$

By using the above, we have

$$\Gamma^{(a,b)}(\omega) = \begin{cases} -\frac{1}{\sigma_a\sigma_b}\sum_{r\in\mathbb{Z}}\phi_r^{(a,b)}\exp(ir\omega) & a \neq b \\ \frac{1}{\sigma_a^2}\left(1 - \sum_{r\neq 0}\phi_r^{(a,a)}\exp(ir\omega)\right) & a = b \end{cases}.$$

Thus the entries of $\Gamma(\omega)$ are the Fourier transforms of the partial correlations. Let

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

By stationarity we have

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

By using Lemma 3.3 we have

$$[\Gamma^{(a,a)}(\omega)]^{-1} = \sum_{r\in\mathbb{Z}} \rho_r^{(a,a)|\text{-}\{a\}} \exp(ir\omega).$$

We now use the methods laid out in this paper to derive the partial spectral coherence. For $a \neq b$ we define

$$X_t^{(a)|\text{-}\{a,b\}} \;=\; X_t^{(a)} - P_{\mathcal{H}-(X^{(a)},X^{(b)})}(X_t^{(a)}) \text{ and } X_\tau^{(b)|\text{-}\{a,b\}} = X_\tau^{(b)} - P_{\mathcal{H}-(X^{(a)},X^{(b)})}(X_\tau^{(b)}).$$

Since the time series is stationary we define the time series partial covariance as

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

Thus by using Lemma 3.3 we have

$$\begin{aligned} &\sum_{r\in\mathbb{Z}} \begin{pmatrix} \rho_r^{(a,a)|\text{-}\{a,b\}} & \rho_r^{(a,b)|\text{-}\{a,b\}} \\ \rho_r^{(b,a)|\text{-}\{a,b\}} & \rho_r^{(b,b)|\text{-}\{a,b\}} \end{pmatrix} \exp(ir\omega) \\ =\;& \frac{1}{\Gamma^{(a,b)}(\omega)\Gamma^{(b,b)}(\omega) - |\Gamma^{(a,b)}(\omega)|^2} \begin{pmatrix} \Gamma^{(b,b)}(\omega) & -\Gamma^{(a,b)}(\omega) \\ -\Gamma^{(a,b)}(\omega)^* & \Gamma^{(a,a)}(\omega) \end{pmatrix}. \end{aligned}$$

Therefore for $c \in \{a, b\}$ we have

$$\sum_{r\in\mathbb{Z}} \rho_r^{(c,c)|\text{-}\{a,b\}} \exp(ir\omega) = \frac{\Gamma^{(c,c)}(\omega)}{\Gamma^{(a,a)}(\omega)\Gamma^{(b,b)}(\omega) - |\Gamma^{(a,b)}(\omega)|^2} = H^{(c,c)}(\omega)$$

and

$$\sum_{r\in\mathbb{Z}} \rho_r^{(a,b)|\text{-}\{a,b\}} \exp(ir\omega) = -\frac{\Gamma^{(a,b)}(\omega)}{\Gamma^{(a,a)}(\omega)\Gamma^{(b,b)}(\omega) - |\Gamma^{(a,b)}(\omega)|^2} = H^{(a,b)}(\omega).$$

Thus the partial spectral coherence between edge (a, b) is

$$R_{ab}(\omega) = \frac{H^{(a,b)}(\omega)}{\sqrt{H^{(a,a)}(\omega)H^{(b,b)}(\omega)}} = -\frac{\Gamma^{(a,b)}(\omega)}{\sqrt{\Gamma^{(a,a)}(\omega)\Gamma^{(b,b)}(\omega)}}. \tag{111}$$

This coincides with with the partial spectral coherence given in Dahlhaus (2000b), equation (2.2) who shows that the partial spectral coherence is

$$R_{ab}(\omega) = \frac{g_{a,b}(\omega)}{\sqrt{g_{a,a}(\omega)g_{b,b}(\omega)}} \tag{112}$$

where

$$g_{c,d}(\omega) = \Sigma_{c,d}(\omega) - \Sigma_{c,-(a,b)}(\omega)\Sigma_{-(a,b)}(\omega)^{-1}\Sigma_{d,-(a,b)}(\omega)^{*} \quad c,d \in \{a,b\}$$

and $\Sigma_{a,-(a,b)}$ denotes the spectral cross correlation between $\{X_t^{(a)}\}_t$ and $\{X_t^{(c)}; c \notin \{a,b\}\}$, $\Sigma_{b,-(a,b)}$ denotes the spectral cross correlation between $\{X_t^{(b)}\}_t$ and $\{X_t^{(c)}; c \neq \{a,b\}\}$ and $\Sigma_{-(a,b)}$ denotes the spectral cross correlation of $\{X_t^{(c)}; c \notin \{a,b\}\}_t$ i.e.

$$\mathbf{\Sigma}(\omega) = \begin{pmatrix} \Sigma_{a,a}(\omega) & \Sigma_{a,b}(\omega) & \Sigma_{a,-(a,b)}(\omega) \\ \Sigma_{b,a}(\omega) & \Sigma_{b,b}(\omega) & \Sigma_{b,-(a,b)}(\omega) \\ \Sigma_{a,-(a,b)}(\omega)^{*} & \Sigma_{b,-(a,b)}(\omega)^{*} & \Sigma_{-(a,b)}(\omega) \end{pmatrix}.$$

Dahlhaus (2000b), Theorem 2.4 shows that (111) and (112) are equivalent. For completeness we give the proof using the block matrix inversion identity. The Schur complement of the $(p-2)\times(p-2)$ matrix $\Sigma_{-(a,b)}(\omega)$ in $\mathbf{\Sigma}(\omega)$ is

$$\begin{aligned} & P^{(a,b)}(\omega) \\ = & \begin{pmatrix} \Sigma_{aa}(\omega) - \Sigma_{a,-(a,b)}(\omega)\Sigma_{-(a,b)}(\omega)^{-1}\Sigma_{-a,(a,b)}(\omega)^{*} & \Sigma_{ba}(\omega) - \Sigma_{a,-(a,b)}(\omega)\Sigma_{-(a,b)}(\omega)^{-1}\Sigma_{-b,(a,b)}(\omega)^{*} \\ \Sigma_{ab}(\omega) - \Sigma_{b,-(a,b)}(\omega)\Sigma_{-(a,b)}(\omega)^{-1}\Sigma_{a,-(a,b)}(\omega)^{*} & \Sigma_{bb}(\omega) - \Sigma_{b,-(a,b)}(\omega)\Sigma_{-(a,b)}(\omega)^{-1}\Sigma_{-b,(a,b)}(\omega)^{*} \end{pmatrix}. \end{aligned}$$

Using the block inverse identity we recall that $P^{(a,b)}(\omega)^{-1}$ is the top left hand matrix in $\mathbf{\Sigma}(\omega)^{-1} = \mathbf{\Gamma}(\omega)$. Thus

$$P^{(a,b)}(\omega)^{-1} = \begin{pmatrix} \Gamma^{(a,a)}(\omega) & \Gamma^{(a,b)}(\omega) \\ \Gamma^{(b,a)}(\omega) & \Gamma^{(b,b)}(\omega) \end{pmatrix}.$$

Therefore from the above we have

$$\begin{aligned} & \begin{pmatrix} \Sigma_{aa}(\omega) - \Sigma_{a,-(a,b)}(\omega)\Sigma_{-(a,b)}(\omega)^{-1}\Sigma_{-a,(a,b)}(\omega)^{*} & \Sigma_{ba}(\omega) - \Sigma_{a,-(a,b)}(\omega)\Sigma_{-(a,b)}(\omega)^{-1}\Sigma_{-b,(a,b)}(\omega)^{*} \\ \Sigma_{ab}(\omega) - \Sigma_{b,-(a,b)}(\omega)\Sigma_{-(a,b)}(\omega)^{-1}\Sigma_{a,-(a,b)}(\omega)^{*} & \Sigma_{bb}(\omega) - \Sigma_{b,-(a,b)}(\omega)\Sigma_{-(a,b)}(\omega)^{-1}\Sigma_{-b,(a,b)}(\omega)^{*} \end{pmatrix} \\ = & \frac{1}{\Gamma^{(a,b)}(\omega)\Gamma^{(b,b)}(\omega) - |\Gamma^{(a,b)}(\omega)|^{2}} \begin{pmatrix} \Gamma^{(b,b)}(\omega) & -\Gamma^{(a,b)}(\omega) \\ -\Gamma^{(b,a)}(\omega) & \Gamma^{(a,a)}(\omega) \end{pmatrix}. \end{aligned}$$

Comparing the entries in the above it is immediately clear that

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

Thus proving that (111) and (112) are equivalent expression for multivariate stationary time series.