Inverse covariance operators of multivariate nonstationary time series

Jonas Krampe and Suhasini Subba Rao

October 8, 2022

Abstract

For multivariate stationary time series many important properties, such as partial correlation, graphical models and autoregressive representations are encoded in the inverse of its spectral density matrix. This is not true for nonstationary time series, where the pertinent information lies in the inverse infinite dimensional covariance matrix operator associated with the multivariate time series. This necessitates the study of the covariance of a multivariate nonstationary time series and its relationship to its inverse. We show that if the rows/columns of the infinite dimensional covariance matrix decay at a certain rate then the rate (up to a factor) transfers to the rows/columns of the inverse covariance matrix. This is used to obtain a nonstationary autoregressive representation of the time series and a Baxter-type bound between the parameters of the autoregressive infinite representation and the corresponding finite autoregressive projection. The aforementioned results lay the foundation for the subsequent analysis of locally stationary time series. In particular, we show that smoothness properties on the covariance matrix transfer to (i) the inverse covariance (ii) the parameters of the vector autoregressive representation and (iii) the partial covariances. All results are set up in such a way that the constants involved depend only on the eigenvalue of the covariance matrix and can be applied in the high-dimensional settings with non-diverging eigenvalues.

Keywords and phrases: Autoregressive parameters, Baxter's inequality, high dimensional time series, local stationarity and partial covariance.

1 Introduction

Several important properties in multivariate analysis are encrypted within the inverse covariance of the underlying random vector. For example, the partial correlation, regression parameters and the network corresponding to the (Gaussian) graphical model. For multivariate time series the covariance is now an infinite dimensional matrix. Nevertheless, analogous to classical multivariate analysis many interesting properties in time series are encoded in the inverse infinite dimensional variance matrix. They include (i) the partial covariance between different components of time series after conditioning on the other time series (ii) time series graphical models which takes into account the conditional relationships over the entire time series and (iii) vector autoregressive representations which yield information on Granger causality. For stationary time series, however, it is rare to directly deduce these relationships from the inverse covariance, as these quantities have an equivalent representation in terms of the finite dimensional inverse spectral density matrix corresponding to the autocovariance of the time series. For example, the partial covariance can be expressed in terms of the partial spectral coherence (which is a function of the inverse spectral density matrix; see, Priestley (1981), Chapter 9.2). The stationary time series graphical model can be deduced from the zero and non-zeroes of the inverse spectral density matrix (see, Dahlhaus (2000a)) and the vector autoregressive regressive representation can be deduced from the causal factorisation of the inverse spectral density matrix (see Wiener and Masani (1958)). However, once one moves away from stationarity, a rigorous understanding of the above properties can only be achieved by directly studying the inverse of the infinite dimensional covariance matrix (and its relationship to the corresponding covariance). This is the main objective of this paper, which we make precise below.

Let $\{X_t = (X_t^{(1)}, \ldots, X_t^{(p)})^\top; t \in \mathbb{Z}\}$ denote a p-dimensional multivariate time series with $p \times p$-dimensional covariance matrix $C_{t,\tau} = \mathbb{C}\mathrm{ov}[X_t, X_\tau]$ for all $t, \tau \in \mathbb{Z}$. Using $\{C_{t,\tau}\}_{t,\tau}$ we define the linear operator or, equivalently, infinite dimensional matrix $\boldsymbol{C} = (C_{t,\tau}; \tau, t \in \mathbb{Z})$. Under suitable conditions on $\boldsymbol{C}$, the inverse $\boldsymbol{D} = \boldsymbol{C}^{-1} = (D_{t,\tau}; t, \tau \in \mathbb{Z})$ exists. Basu and Subba Rao (2022), Section 2, show that a graphical model for nonstationary time series can be defined from the structure of $\boldsymbol{D}$ (based on zero, Toeplitz and non-Toeplitz submatrices in $\boldsymbol{D}$). This general framework does not impose any conditions on the nonstationary structure of the time series. However, in order to learn the network from data Basu and Subba Rao (2022) focus on locally stationary time series; by now a widely accepted and used class of nonstationary time series. Specifically, smoothness conditions are placed on the inverse covariance $\boldsymbol{D}$, and the subsequent analysis is done under these conditions. However, most locally stationary conditions are stated in terms of the covariance rather than the inverse covariance. This leads to the question "do smoothness conditions on $\boldsymbol{C}$ transfer to smoothness on $\boldsymbol{D}$?" and provided the initial motivation for this paper. It naturally lead to further questions on the "transfer" of smoothness on $\boldsymbol{C}$ to (a) vector autoregressive representations and (b) the partial covariance. Therefore, our aim is to develop a suite of tools that answer such questions. To the best of our knowledge there exists very few results in this area. One notable exception is the recent work of Ding and Zhou (2021), but the aims and results in their work are different to those of this paper. Ding and Zhou (2021) specifically focus on the univariate nonstationary time series $(X_1, \ldots, X_n)$ (with $n \to \infty$). They show that there exists an autoregressive representation of increasing order over the time points, whose coefficients decay at a certain rate. The results are used to test for correlation stationarity. In contrast, we work within the multivariate time series framework, and allow for both low and high dimensional time series. The latter case is important because often to make meaningful conditional statements about components in the time series (in terms of Granger causality and conditional covariance) the number of time series included in the analysis may need to be extremely large. We summarise the main results below.

In order to reconcile $\boldsymbol{C}$ and its inverse $\boldsymbol{D}$, in Section 2 we show if $\|C_{t,\tau}\|_2 \leq K|t-\tau|^{-\kappa}$ for $t \neq \tau$ and some $\kappa > 1$ ($\|\cdot\|_2$ denotes the induced ℓ_2/spectral norm), then $\|D_{t,\tau}\|_2 \leq \mathcal{K}(1+\log|t-\tau|)^{\kappa}(|t-\tau|)^{-\kappa+1}$. This leads to a nonstationary VAR(∞) representation of the time series $\{X_t\}_t$ where the corresponding VAR parameters decay at the same rate. We use this result to obtain a Baxter-type bound between the parameters of autoregressive infinite representation and the corresponding finite autoregressive projection. It is noteworthy that the constant $\mathcal{K}$ depends only on the eigenvalues of $\boldsymbol{C}$, but not on the dimension p. Hence, if the eigenvalues of $\boldsymbol{C}$ do not grow with dimension p, these results hold for arbitrary dimension.

The results in Section 2 are instrumental to proving the results in Section 3, where we focus on locally stationary time series. In terms of second order structure, a time series is called second order locally stationary if its covariance structure can locally be approximated by a smooth function $\boldsymbol{C}(u)$. We show in Section 3.2 that $\boldsymbol{C}(u)$ is an autocovariance of a stationary time series. In Section 3.3 we show that locally stationary conditions based on the covariance structure imply that its inverse covariance can locally be approximated by a smooth function $\boldsymbol{D}(u)$, which is the inverse autocovariance of a stationary time series i.e. $\boldsymbol{D}(u) = \boldsymbol{C}(u)^{-1}$. We use this result to show that the parameters of the vector autoregressive representation of the time series can be approximated by a smooth function. Finally, in Section 3.4, we show that the smoothness conditions on the nonstationary covariance transfer to smoothness conditions on the partial covariances. We use this result to justify using an estimator of the local spectral density function to estimate the local partial spectral coherence (as was done in Park et al. (2014)) and the local partial correlation. The proof of the results can be found in the Appendix.

2 Rate of decay of the inverse covariance

2.1 Notation and assumptions

In order to derive the results in this paper we need to define the space on which the operator $\boldsymbol{C}$ is acting. This requires the following notation.

Let $\mathbb{R}$ denote the real numbers, $\mathbb{Z}$ all (positive and negative) integers and $\mathbb{N}$ strictly positive integers. For $u, v \in \mathbb{R}^p$ let $\langle u, v\rangle = u^\top v$ and $\|v\|_2$ denote the Euclidean distance. We use ℓ_2 and $\ell_{2,p}$ to denote the sequence spaces $\ell_2 = \{u = (\ldots, u_{-1}, u_0, u_1, \ldots); u_j \in \mathbb{R} \text{ and } \sum_{j\in\mathbb{Z}} u_j^2 < \infty\}$ and $\ell_{2,p} = \{v = (\ldots, v_{-1}, v_0, v_1, \ldots); v_j \in \mathbb{R}^p \text{ and } \sum_{j\in\mathbb{Z}} \|v_j\|_2^2 < \infty\}$. On the spaces ℓ_2 and $\ell_{2,p}$ we define the two inner products $\langle u, v\rangle = \sum_{j\in\mathbb{Z}} u_j v_j$ (for $u, v \in \ell_2$) and $\langle x, y\rangle = \sum_{j\in\mathbb{Z}}\langle x_j, y_j\rangle$ (for $x = (\ldots, x_{-1}, x_0, x_1, \ldots), y = (\ldots, y_{-1}, y_0, y_1, \ldots) \in \ell_{2,p}$). For $x \in \ell_{2,p}$, let $\|x\|_2 = \langle x, x\rangle$. Furthermore, for $x \in \ell_{2,p}$ and $s \in \mathbb{Z}, a \in 1, \ldots, p$, we use $x_s^{(a)}$ to denote the sth element of the ath (column) space. Suppose $\{A_{s_1,s_2}\}_{s_1,s_2}$ are $p_1 \times p_2$-dimensional matrices, using this we define the infinite dimensional matrix $\boldsymbol{A} = (A_{s_1,s_2}; s_1, s_2 \in \mathbb{Z})$. Under suitable conditions on $\boldsymbol{A}$, $\boldsymbol{A}$ is a linear operator $\boldsymbol{A} : \ell_{2,p_1} \to \ell_{2,p_2}$ in the sense that if $\boldsymbol{A}x = y$, then $y = (\ldots, y_{-1}, y_0, y_1, \ldots)$ where for all $t \in \mathbb{Z}$, $y_t \in \mathbb{R}^{p_2}$ and $y_t = \sum_{\tau\in\mathbb{Z}} A_{t,\tau} x_\tau$. Furthermore, we define $\|\boldsymbol{A}\|_2 = \sup_{\|x\|_2=1} \|\boldsymbol{A}x\|_2$. All operators are written in bold uppercase letters.

Assumption 2.1. *Let* $v(\cdot) = \max(1, |\cdot|)$.

(i) The covariance operator is positive definite with $\lambda_{\sup} = \sup_{v\in\ell_{2,p},\|v\|_2=1}\langle v, \boldsymbol{C}v\rangle < \infty$ *and* $0 < \lambda_{\inf} = \inf_{v\in\ell_{2,p},\|v\|_2=1}\langle v, \boldsymbol{C}v\rangle$.

(ii) There exists some $\kappa > 1$ *such that for all* $t \neq \tau$ *we have for the* $p \times p$*-dimensional sub-matrices*

$$\|C_{t,\tau}\|_2 \leq K v(t-\tau)^{-\kappa},$$

where $K < \infty$ is some positive constant.

Since $\boldsymbol{C}$ is positive definite, the inverse covariance operator exists with $\boldsymbol{D} = \boldsymbol{C}^{-1} = (D_{t,\tau}; t, \tau \in \mathbb{Z})$. We mention that the condition $\lambda_{\sup} < \infty$ is implied by Assumption 2.1(ii).

The results in this paper allow for both low and high dimensional multivariate time series and the assumptions used are specifically designed to allow for this. For high dimensional time series, the condition that the largest eigenvalue is bounded excludes time series with dynamic factors but allows for high dimensional sparse time series.[1] Popular examples include high dimensional sparse time series regression and vector autoregressive (VAR) models which have recently received considerable attention; see, for example, Basu and Michailidis (2015)[2], Krampe et al. (2021), Krampe and Paparoditis (2021), (in the context of stationary VAR models) and Ding et al. (2017) (for time-varying VAR models). The condition that $\lambda_{\inf} > 0$ omits co-linearity, where one component in the time series can be perfectly explained by other components. Assumption 2.1(ii) quantifies the pairwise dependencies between the components (over time) and is stated in terms of the (induced) ℓ_2-norm $\|\cdot\|_2$ of the $p \times p$ matrices. However, no conditions are placed on the ℓ_1-norm, which can grow with dimension p (as sparsity usually does in the sparse regression context). All results in this paper are derived in terms of the $\|\cdot\|_2$-norm. Thus we show that if the pairwise interactions are controlled in the ℓ_2 sense as p grows, then the conditional interactions are also controlled in the ℓ_2-sense.

Throughout this paper we use $\mathcal{K}$ to denote a generic constant that only depends on $\lambda_{\inf}, \lambda_{\sup}, K, \kappa$ and whose value may change from line to line. We define $v(\cdot) = \max(1, |\cdot|)$ and $\zeta(j)$ as follows; for $|j| \leq 1$ let $\zeta(j) = 1$ and for $|j| > 1$ let $\zeta(j) = \log|j|/|j|$.

2.2 The inverse covariance

In the following theorem we obtain a bound on the rate of decay of the matrices $D_{t,\tau}$ that make up the inverse covariance $\boldsymbol{D} = \boldsymbol{C}^{-1}$. $\boldsymbol{C}$ is a bi-infinite matrix in the sense that the entries $C_{t,\tau}$ span $t, \tau \in \mathbb{Z}$. We will also consider the one-sided infinite dimensional matrix $\boldsymbol{C}(-\infty, T) = (C_{t,\tau}; t, \tau \leq T)$. As will be clear later in the paper, the inverse of $\boldsymbol{C}(-\infty, T)$

[1]By dynamic factors we refer to the common component described in the representation given in Forni et al. (2000). The common component contains (if any) the diverging eigenvalues of the process. The bounded eigenvalues define in this decomposition the so-called idiosyncratic component. Hence, if eigenvalues of $\boldsymbol{C}$ diverge with p, the results can be applied to the idiosyncratic component of this decomposition.

[2]Note that the finite sample error bounds derived in Basu and Michailidis (2015) for the Lasso express the dependence of the processes also in terms of $\lambda_{\inf}$ and $\lambda_{\sup}$.

contains (up to a factor) the AR prediction coefficients and the following result will be used to obtain a bound on its rate of decay.

Theorem 2.1. *Under Assumption 2.1, for all $t, \tau \in \mathbb{Z}$ we have*

$$\|D_{t,\tau}\|_2 \leq \mathcal{K}\zeta(t-\tau)^{\kappa-1}, \tag{1}$$

where $\mathcal{K}$ is a constant depending on $K, \kappa, \lambda_{\inf}$, and $\lambda_{\sup}$ only and $\zeta(j) = v(\log[v(j)])/v(j)$. For $t, \tau \leq T$

$$\|[\boldsymbol{C}(-\infty; T)]^{-1}]_{t,\tau}\|_2 \leq \mathcal{K}\zeta(t-\tau)^{\kappa-1}. \tag{2}$$

Proof. The key ingredient in the proof is Lemma B.1 (in Appendix B), which bounds the entries of the inverse of a banded matrix operator (and is a generalisation of Proposition 1 in Demko et al. (1984)). The details of the proof are in Appendix B. □

The above result shows that if the pairwise interaction between the components is bounded with a certain rate in the ℓ_2-sense then the conditional interactions are also bounded with a certain rate in the ℓ_2-sense, see Remark 3.1 for a discussion on the role of the dimension p.

Remark 2.1. *In the case entries in $\boldsymbol{C}$ decay geometrically or are banded, then the entries of $\boldsymbol{D}$ decay at a geometric rate.*

Remark 2.2 (An alternative representation of the covariance $\boldsymbol{C}$ and its inverse)**.** *We recall that we defined $\boldsymbol{C}$ as $\boldsymbol{C} = (C_{t,\tau}; t, \tau \in \mathbb{Z})$, where $C_{t,\tau}$ are $p \times p$-dimensional matrices. An alternative method for defining $\boldsymbol{C}$ is to group the covariances according to component i.e. $\widetilde{\boldsymbol{C}} = (C^{(a,b)}; 1 \leq a, b \leq p)$ where $[C^{(a,b)}]_{t,\tau} = C_{t,\tau}^{(a,b)} = \mathbb{C}\text{ov}[X_t^{(a)}, X_\tau^{(b)}]$. $\widetilde{\boldsymbol{C}}$ is simply a permutation of $\boldsymbol{C}$, thus $\widetilde{\boldsymbol{D}} = \widetilde{\boldsymbol{C}}^{-1}$ is a permutation of $\boldsymbol{D}$. In certain applications, such as nonstationary graphical models or condition covariance between two components of a time series, the representations $\widetilde{\boldsymbol{C}}$ and $\widetilde{\boldsymbol{D}}$ may be more useful in the analysis than $\boldsymbol{C}$ and $\boldsymbol{D}$ (see, for example, Basu and Subba Rao (2022)).*

We now compare Theorem 2.1 with the classical result for stationary time series. For this, suppose $\boldsymbol{C} = (C_{t-\tau}; t, \tau \in \mathbb{Z})$ is a block Toeplitz operator from $\ell_{2,p}$ to $\ell_{2,p}$, where $\boldsymbol{C}$ satisfies the rate and positive definiteness conditions in Assumption 2.1. Then $\boldsymbol{D} = \boldsymbol{C}^{-1} = (D_{t-\tau}; t, \tau \in \mathbb{Z})$ exists and is also a block Toeplitz operator. For block Toeplitz operators Cheng and Pourahmadi (1993); Meyer and Kreiss (2015) work with a global condition on the sequence $(C_s)_{s\in\mathbb{Z}}$ instead of the individual one used in this paper. They showed that if the global condition $\sum_{s\in\mathbb{Z}}(1+|s|^{\kappa})\|C_s\|_2 < \infty$ holds, then $\sum_{s\in\mathbb{Z}}(1+|s|^{\kappa})\|D_s\|_2 < \infty$. The global condition implies for all $t,\tau \in \mathbb{Z}$ that $\|C_{t-\tau}\|_2 \leq Kv(t-\tau)^{-\kappa}$ and $\|D_{t-\tau}\|_2 \leq Kv(t-\tau)^{-\kappa}$. Conversely, the individual condition that yields this global condition is $\|C_{t-\tau}\|_2 \leq Kv(t-\tau)^{-\kappa-1-\varepsilon}$. In summary, even for block Toeplitz matrices, at the individual level if $\|C_{t-\tau}\|_2 \leq Kv(t-\tau)^{-\kappa-\varepsilon}$ then the above arguments yield

$$\|D_{t-\tau}\|_2 \leq Kv(t-\tau)^{-\kappa+1}, \tag{3}$$

which is (without the log-factor) same as the rate derived in Theorem 2.1. To the best of our knowledge, it is an open question if this rate at the individual level for the inverse can be improved for stationary as well general nonstationary time series.

2.3 Vector Autoregressive representation and Baxter's inequality

It is well known that for stationary time series the entries of $\boldsymbol{C}(-\infty,T)$ are closely related to vector autoregressive (VAR(∞)) parameters of the underlying time series. The same is true for nonstationary time series. Precisely, under Assumption 2.1 and by using the projection theorem the bottom row of $\boldsymbol{C}(-\infty,T)^{-1}$ contains the VAR(∞) coefficients in the linear projection of X_T onto the space spanned by $\overline{\text{sp}}(X_{T-1},X_{T-2},\ldots)$ i.e.,

$$X_T = \sum_{j=1}^{\infty}\Phi_{T,j}X_{T-j} + \varepsilon_T, \text{ where } \Phi_{T,j} = -([\boldsymbol{C}(-\infty,T)^{-1}]_{T,T})^{-1}[\boldsymbol{C}(-\infty,T)^{-1}]_{T,T-j}, \tag{4}$$

where ε_T is uncorrelated with $\{X_{T-j}\}_{j=1}^{\infty}$. Substituting the bound in Theorem 2.1 into (4) gives

$$\|\Phi_{T,j}\|_2 \leq \mathcal{K}\zeta(t-\tau)^{\kappa-1}. \tag{5}$$

In practice, it is often not possible to estimate the infinite number of AR parameters from a finite data set. Therefore one often estimates the parameters of the projection of X_T

onto the finite past $\overline{\text{sp}}(X_{T-1}, \ldots, X_{T-d})$ i.e.,

$$X_T = \sum_{j=1}^{d} \Phi_{T,d,j} X_{T-j} + \varepsilon_{T,d}. \tag{6}$$

The above is analogous to the best fitting VAR(d) parameters for stationary time series. In stationary time series the difference between the finite past projection and the corresponding infinite past projection is called the Baxter inequality; see Section 6 in Hannan and Deistler (1988), Cheng and Pourahmadi (1993), and Meyer and Kreiss (2015). In the same spirit, we now obtain a Baxter-type inequality for nonstationary multivariate time series, between the VAR(∞) coefficients $\{\Phi_{T,j}\}_j$ and the finite prediction coefficients $\{\Phi_{T,d,j}\}_j$.

The coefficients $\{\Phi_{T,d,j}\}_j$ are embedded in the bottom row of the finite dimensional matrix $\boldsymbol{C}(T-d,T)^{-1}$ where $\boldsymbol{C}(T-d,T) = (C_{t,\tau}; T-d+1 \leq t,\tau \leq T)$. Thus the coefficients $\{\Phi_{T,j}\}_j$ and $\{\Phi_{T,d,j}\}_j$ are connected through $\boldsymbol{C}(T-d,T)$ and $\boldsymbol{C}(-\infty,T)$ and their inverses. Due to this connection we use Theorem 2.1 and the block operator inverse identity (see equation (30) in Appendix A.2) to prove the result below.

Theorem 2.2 (Baxter type inequality)**.** *Suppose Assumption 2.1 holds with $\kappa > 3/2$. Let $\{\Phi_{T,j}\}_j$ and $\{\Phi_{T,d,j}\}$ be defined as in (4) and (6) respectively. Then for $d \in \mathbb{N}, j = 1,\ldots,d$ we have*

$$\sup_T \|\Phi_{T,d,j} - \Phi_{T,j}\|_2 \leq \mathcal{K}\zeta(d)^{\kappa-3/2}\zeta(d-j)^{\kappa-3/2}. \tag{7}$$

Furthermore, if Assumption 2.1 holds with $\kappa > 5/2$ we have

$$\sup_T \sum_{j=1}^{d} \|\Phi_{T,d,j} - \Phi_{T,j}\|_2 \leq \mathcal{K}\zeta(d)^{\kappa-3/2}. \tag{8}$$

Proof. In Appendix B. □

Inequality (5) and Theorem 2.2 are related to Theorem 2.4 in Ding and Zhou (2021), who obtain autoregressive approximations for nonstationary univariate time series. However, it is important to note that there are some differences in the autoregressive representations derived in both papers. The autoregressive representation derived in (Ding and Zhou, 2021) is based on the finite vector $(X_1, \ldots, X_n)$ and their aim is to build an autoregressive representation of increasing order over the time points of the data vector, i.e., X_i is represented as an AR($i-1$) model. In contrast, we derive an autoregressive representation of a time series $\{X_t; t \in \mathbb{Z}\}$ where each time point has a $VAR(\infty)$ representation. In the stationary context, building an autoregressive representation of an increasing order relates to the Cholesky decomposition of $\mathbb{V}\mathrm{ar}(X_1, \ldots, X_n)^{-1}$ where the ith model is given by the ith line. The AR(∞) model using the entire time series can be considered as a limit of this, see Section 2 in Krampe and McMurry (2021) for further discussion. With this difference in mind, we now compare the rates in Section 2.2 with the results in Theorem 2.4 in Ding and Zhou (2021). Their decay rate for the autoregressive coefficients matches with that derived in (5). In terms of Baxter's inequality, they show $\max_{T>b} \max_{1\leq j\leq b} |\Phi_{T,T-1,j} - \Phi_{T,b,j}| \leq C(\log b)^{\kappa-1} b^{-\kappa+3}$. Using Theorem 2.2 we compare the coefficients of the two finite AR models (order $T-1$ and order b), and obtain tighter bounds for their result. To be precise

$$\begin{aligned}\max_T \|\Phi_{T,T-1,j} - \Phi_{T,b,j}\|_2 &\leq \max_T(\|\Phi_{T,T-1,j} - \Phi_{T,j}\|_2 + \|\Phi_{T,j} - \Phi_{T,b,j}\|_2) \\ &\leq \mathcal{K}\left(\zeta(T-j)^{\kappa-3/2} + \zeta(b-j)^{\kappa-3/2}\right)(\log(b)/b)^{\kappa-3/2}.\end{aligned}$$

The above leads to the bound $\max_{T>b}\max_{1\leq j\leq b} |\Phi_{T,T-1,j} - \Phi_{T,b,j}| = O(b^{-\kappa+3/2}\log^{\kappa-3/2} b)$ instead of $O(b^{-\kappa+3}\log^{\kappa-1} b)$ (given in Ding and Zhou (2021)).

We now compare Theorem 2.2 to the stationary set-up. Meyer and Kreiss (2015) showed that under the following global condition on the vector autoregressive parameters $\sum_{s\in\mathbb{Z}}(1+|s|^{\kappa})\|\Phi_s\|_2 < \infty$, that

$$\sum_{j=1}^{d}(1+j)^{\kappa}\|\Phi_{d,j} - \Phi_j\|_2 \leq \mathcal{K}\sum_{j=d+1}^{\infty}(1+j)^{\kappa}\|\Phi_j\|_2, \tag{9}$$

noting that we have dropped T as it is not necessary under stationarity. (9) implies $\sum_{j=1}^{d}\|\Phi_{d,j} - \Phi_j\|_2 \leq \mathcal{K}d^{-\kappa}$. Based on the discussion at the end of Section 2.2, at the individual level this means if $\|C_s\|_2 \leq Kv(s)^{-\kappa-\varepsilon}$, then $\sum_{j=1}^{d}\|\Phi_{d,j} - \Phi_j\|_2 \leq \mathcal{K}d^{-\kappa+1}$, whereas Theorem 2.2 gives $\sum_{j=1}^{d}\|\Phi_{d,j} - \Phi_j\|_2 \leq \mathcal{K}d^{-\kappa+3/2}$. Thus stationarity of the time series yields a better approximation bound between the finite and infinite AR parameters than the bound in Theorem 2.2.

3 Locally stationary time series

The first rigorous treatment of locally stationary time series was given in (Dahlhaus, 1997, 2000b). This was done by representing $\{X_{t,T}\}_{t=1}^{T}$ in terms of a Cramér representation $X_{t,T} = \int_0^{2\pi} A_{t,T}(\omega)dZ(\omega)$, where $\{Z(\omega); \omega \in [0, 2\pi]\}$ is an orthogonal increment process and the time-varying transfer function $A_{t,T}(\omega)$ can locally be approximated by the Lipschitz smooth function $A(\omega; \cdot)$ i.e. $\|A_{t,T}(\omega) - A(\omega; u)\|_2 \leq K(|t/N - u| + 1/N)$. This definition immediately leads to certain smoothness properties on the covariance structure of the time series. More recently, several authors have extended this definition to nonlinear time series cf. (Dahlhaus and Subba Rao, 2006; Subba Rao, 2006; Zhou and Wu, 2009; Vogt, 2012; Truquet, 2019; Dahlhaus et al., 2019; Karmakar et al., 2021). In this section, we return, in some sense, to the original formulation of local stationarity and focus on the locally stationary second order structure. However, unlike (Dahlhaus, 1997, 2000b), we work within the time domain and not the frequency domain. We start by introducing the locally stationary setting, i.e., we impose certain smoothness conditions on the nonstationary time series. In Section 3.2 we obtain bounds on the eigenvalues of the underlying covariance. Using Theorem 2.1, in Section 3.3 we show that smoothness conditions placed on the covariance structure transfer over to the inverse covariance and the parameters in the nonstationary AR(∞) representation. Finally, in Section 3.4, we apply these results to show that the smoothness conditions also transfer to the partial covariances.

3.1 Assumptions

We start by defining an infinite array, where for each $N \in \mathbb{N}$ we associate a (non)stationary multivariate time series $\{X_{t,N}; t \in \mathbb{Z}\}$ and covariance $C_{t,\tau}^{(N)} = \mathbb{C}\mathrm{ov}[X_{t,N}, X_{\tau,N}]$ (for all $t, \tau \in \mathbb{Z}$). For each N we define the infinite dimensional covariance matrix $\boldsymbol{C}^{(N)} = (C_{t,\tau}^{(N)}; t, \tau \in \mathbb{Z})$. In the assumptions below we explicitly connect the sequence of infinite dimensional covariance matrices $\{\boldsymbol{C}^{(N)}\}_{N \in \mathbb{N}}$ through N, which plays the role of a smoothing parameter. We mention that it is standard practice in the locally stationary literature to define $X_{t,N}$ on a triangular array i.e. $\{X_{t,N}\}_{t=1}^{N}$. However, to avoid confusion, we do not link N to sample size. It is also worth pointing out that we use $N \in \mathbb{N}$ to simplify the exposition, we could, without loss of generality, allow N to be a non-integer and define it on $N \in [\alpha, \infty)$ (for some $\alpha > 0$).

Assumption 3.1. *(i) Eigenvalue condition: There exists some $N_0 \geq 1$ where*

$$0 < \lambda_{\inf} \leq \inf_{N \geq N_0} \lambda_{\inf}(\mathbf{C}^{(N)}) \leq \sup_{N \geq N_0} \lambda_{\sup}(\mathbf{C}^{(N)}) \leq \lambda_{\sup} < \infty.$$

(ii) Covariance decay condition: For all N, t and τ $\|\boldsymbol{C}_{t,\tau}^{(N)}\|_2 \leq \frac{K}{v(t-\tau)^\kappa}$.

(iii) Smoothness condition: There exists a Lipschitz continuous matrix function $\{C_r(\cdot), r \in Z\}$ where (a) $C_r(u) = C_{-r}(u)^\top$, (b) for all $u, v \in \mathbb{R}, r \in \mathbb{Z}$ $\sup_u \|C_r(u)\|_2 \leq K/v(r)^\kappa$, and (c) $\|C_r(u) - C_r(v)\|_2 \leq \frac{K|u-v|}{v(r)^\kappa}$, such that for all N

$$\|C_{t,\tau}^{(N)} - C_{t-\tau}(t/N)\|_2 \leq \frac{K}{v(t-\tau)^{\kappa-1}} \min\left(\frac{1}{N}, \frac{2}{v(t-\tau)}\right). \tag{10}$$

We assume that $\kappa > 3$.

Note that the above assumptions imply that

$$\|C_{t,\tau}^{(N)} - C_{t-\tau}(u)\|_2 \leq \frac{K}{v(t-\tau)^{\kappa-1}} \min\left[\left(|u - \frac{t}{N}| + \frac{1}{N}\right), \frac{2}{v(t-\tau)}\right].$$

Furthermore, the sequence $\{C_r(\cdot), r \in Z\}$ defines the infinite dimensional matrix operator $\boldsymbol{C}(\cdot) = (C_{t-\tau}(\cdot); t, \tau \in \mathbb{Z})$ (from $\ell_{2,p}$ to $\ell_{2,p}$), where $\boldsymbol{C}(\cdot)$ is block Toeplitz.

Assumption 3.1(i) and (ii) can be viewed as Assumption 2.1 within the framework of an infinite array. Assumption 3.1(iii) places smoothness conditions on the covariance i.e., the (potentially) non-Toeplitz-operator $\boldsymbol{C}^{(N)}$ can locally be approximated by a block Toeplitz-operator $\boldsymbol{C}(\cdot)$, where the approximation error is determined by the smoothing parameter N. The use of min in Assumption 3.1(iii) is not standard within the locally stationary literature. This arises because the time series $\{X_{t,N}\}_t$ is defined on $t \in \mathbb{Z}$ and not $t = 1, \ldots, N$ (the typical locally stationary set-up). If $|t - \tau| < 2N$, then Assumption 3.1(iii) implies that $\|C_{t,\tau}^{(N)} - C_{t-\tau}(t/N)\|_2 \leq \frac{K}{Nv(t-\tau)^{\kappa-1}}$ (the classical locally stationary condition). On the other hand, if $|t - \tau| \geq 2N$, then the smoothing parameter N does not improve on the individual terms $C_{t,\tau}^{(N)}$ and $C_{t-\tau}(t/N)$ (which are extremely small) and we have $\|C_{t,\tau}^{(N)} - C_{t-\tau}(t/N)\|_2 \leq \frac{2K}{v(t-\tau)^\kappa}$. To distinguish these two cases all the relevant results will be stated with min.

Remark 3.1 (The role of dimension p). *In Assumption 3.1 we have not included the dimension p as an additional variable. This is to reduce cumbersome notation. However, it is possible to state Assumption 3.1 in terms of uniform bounds over a three dimensional* *array where the eigenvalues are uniformly bounded over both N and p (and $\boldsymbol{C}^{(N)}$ and $\boldsymbol{C}(u)$ are indexed with p too). If these assumptions hold, then the results in this section hold for high dimensional p too.*

Assumption 3.1 is satisfied by a wide range of locally stationary time series. In Example 3.1 (below) and 3.2 we define the time-varying Vector Moving Average (tv-VMA) model and show that this model satisfies Assumption 3.1.

Example 3.1 (The time-varying vector MA(∞)(tv-VMA) process)**.** *Consider the tv-VMA(∞)*

$$X_{t,N} = \sum_{j=0}^{\infty} \Psi_{t,j}^{(N)} \varepsilon_{t-j} = \sum_{j=1}^{\infty} \Psi_{t,j}^{(N)} \varepsilon_{t-j} + \Psi_{t,0}\varepsilon_t, \qquad t \in \mathbb{Z},$$

where $\{\varepsilon_t\}_t$ are uncorrelated random variables with zero mean and variance I_p. In order for the process to be well defined certain summability or decay conditions need to be imposed on the coefficients $\{\Psi_{t,j}\}$. We assume that $\sup_{N\in\mathbb{N}} \sup_{t\in\mathbb{Z}} \|\Psi_{t,j}^{(N)}\|_2 \leq Kv(j)^{-\kappa}$. With this, we have

$$C_{t,\tau}^{(N)} = \mathbb{C}\mathrm{ov}(\sum_{j=0}^{\infty} \Psi_{t,j}^{(N)} \varepsilon_{t-j}, \sum_{j=0}^{\infty} \Psi_{\tau,j}^{(N)} \varepsilon_{\tau-j}) = \sum_{j\in\mathbb{Z}} \Psi_{t,j}^{(N)} (\Psi_{\tau,j+\tau-t}^{(N)})^{\top},$$

where we set $\Psi_{t,j}^{(N)} = 0$ for $j < 0$. Using the above decay condition on $\Psi_{t,j}^{(N)}$ and Lemma A.4 we have $\|C_{t,\tau}^{(N)}\|_2 \leq Kv(t-\tau)^{\kappa}$; thus Assumption 3.1(ii) holds. We now introduce the locally stationary approximation to $\{X_{t,N}\}$. Analogous to Dahlhaus (1997) and Dahlhaus and Polonik (2006) (for the case $p = 1$), we assume there exists a Lipschitz continuous matrix function $\Psi_j(\cdot)$ where $\sup_{u\in\mathbb{R}} \|\Psi_j(u)\|_2 \leq Kv(j)^{-\kappa}$, $\sup_{u\in\mathbb{R}} \|\Psi_j(u) - \Psi_j(v)\|_2 \leq K|u-v|v(j)^{-\kappa}$, and $\|\Psi_{t,j}^{(N)} - \Psi_j(t/N)\| \leq Kv(j)^{-\kappa}/N$. Using this, we define the stationary process $\{X_t(u)\}_t$ where $X_t(u) = \sum_{j=0}^{\infty} \Psi_{t,j}(u)\varepsilon_{t-j}$ which has autocovariance $C_r(u) = \sum_{j\in\mathbb{Z}} \Psi_j(u)\Psi_{j+r}(u)^{\top}$ (where we set $\Psi_j(u) = 0$ for $j < 0$). Note $\sup_u \|C_r(u)\|_2 \leq K/v(r)^{\kappa}$

(this follows from Lemma A.4). Furthermore, under these conditions we have

$$
\begin{aligned}
\|C_{t,\tau}^{(N)} - C_{t-\tau}(t/N)\|_2 &\leq \sum_{j\in\mathbb{Z}} \|\Psi_{t,j}^{(N)} - \Psi_j(t/N)\|_2 \|\Psi_{\tau,j+\tau-t}^{(N)}\|_2 \\
&\quad + \sum_{j\in\mathbb{Z}} \|\Psi_j(t/N)\|_2 \big(\|\Psi_{j+\tau-t}(t/N) - \Psi_{j+\tau-t}(\tau/N)\|_2 \\
&\quad + \|\Psi_{j+\tau-t}(\tau/N) - \Psi_{\tau,j+\tau-t}^{(N)}\|_2\big) \\
&\leq \frac{\mathcal{K}}{N} \sum_{j\in\mathbb{Z}} \left(\frac{1}{v(j)^\kappa v(j+t-\tau)^{\kappa-1}} + \frac{|t-\tau|}{v(j)^\kappa v(j+t-\tau)^\kappa} \right) \\
&\leq \frac{\mathcal{K}}{N v(t-\tau)^{\kappa-1}}.
\end{aligned}
$$

Thus Assumption 3.1(iii) holds. We observe that this example illustrates why the rate drops from κ to $\kappa - 1$ in $\|C_{t,\tau}^{(N)} - C_{t-\tau}(t/N)\|_2$; there is an additional "cost" due to the inclusion of the term $|t - \tau|$.

In Example 3.2 (in Section 3.2) we show that Assumption 3.1(i) is also satisfied (for sufficiently large N).

3.2 Properties of the locally stationary covariance

In this subsection we show that positive definiteness of $\boldsymbol{C}^{(N)}$ transfers to $\boldsymbol{C}(\cdot)$ under the stated smoothness condition. Conversely, we show that also the other direction holds i.e., for a sufficiently large N_0 positive definiteness of $\boldsymbol{C}(u)$ implies that $\boldsymbol{C}^{(N)}$ is also positive definite (for $N > N_0$).

Theorem 3.1 (Positive definiteness of $\boldsymbol{C}(u)$)**.** *Suppose Assumption 3.1 holds. Then, for all $u \in \mathbb{R}$ $\{C_r(u)\}_r$ is a positive definite sequence where $\lambda_{\inf} \leq \lambda_{\inf}(\boldsymbol{C}(u)) \leq \lambda_{\sup}(\boldsymbol{C}(u)) \leq \lambda_{\sup}$.*

Proof. In Appendix C.1. □

Under the above theorem, $\{C_r(u)\}_r$ is a positive definite sequence. Consequently by Kolmogorov's extension theorem there exists a stationary multivariate time series $\{X_t(u)\}_{t\in\mathbb{Z}}$ which has $\{C_r(u)\}_{r\in\mathbb{Z}}$ as its autocovariance function. This justifies calling $\{X_{t,N}\}_{t\in\mathbb{Z}}$ a "locally" second order stationary time series. A further implication of Lemma 3.1 is that the inverse of $\boldsymbol{C}(u)$ exists, which we denote by $\boldsymbol{D}(u) = \boldsymbol{C}(u)^{-1} = \{D_{t-\tau}(u); t, \tau \in \mathbb{Z}\}$. Like $\boldsymbol{C}(u)$, $\boldsymbol{D}(u)$ is also block Toeplitz and by Theorem 2.1 the

$p \times p$-dimension matrix $D_{t-\tau}(u)$ has the bound

$$\sup_u \|D_{t-\tau}(u)\|_2 \le \mathcal{K}\zeta(t-\tau)^{-\kappa+1}. \tag{11}$$

For a given nonstationary time series model, Assumption 3.1(i) is difficult to directly verify. However, we now show that given a positive definite sequence $\{C_r(u)\}_r$ which satisfies Assumption 3.1(ii,iii), then Assumption 3.1(i) holds. For the univariate case, a similar result is given in (Ding and Zhou, 2021, Proposition 2.9).

Theorem 3.2. *Suppose $\{X_{t,N}\}_{t\in\mathbb{Z}}$ is a locally stationary time series whose covariance $C^{(N)} = (C^{(N)}_{t,\tau}; t, \tau \in \mathbb{Z})$ satisfies Assumption 3.1(ii,iii). Let $f(\omega; u) = \sum_{r\in\mathbb{Z}} C_r(u)\exp(ir\omega)$ be the local spectral density. If*

$$0 < \gamma_{\inf} \le \inf_u \inf_\omega \lambda_{\min}(f(\omega; u)) \le \sup_u \sup_\omega \lambda_{\max}(f(\omega; u)) \le \gamma_{\sup} < \infty, \tag{12}$$

then there exists a N_0, $\lambda_{\inf}$ and $\lambda_{\sup}$ where for all $N \ge N_0$ we have

$$0 < \lambda_{\inf} \le \lambda_{\inf}(\mathbf{C}^{(N)}) \le \lambda_{\sup}(\mathbf{C}^{(N)}) \le \lambda_{\sup} < \infty.$$

Proof. In Appendix C.1. □

Equipped with the above results, we return to Example 3.1.

Example 3.2 (Example 3.1, continued)**.** *We define the local spectral density as*

$$f(\omega; u) = [\sum_{j=0}^{\infty} \Psi_j(t/N)\exp(-ij\omega)][\sum_{j=0}^{\infty} \Psi_j(t/N)\exp(ij\omega)]^\top.$$

Under the conditions of Example 3.1 we have $\sup_u \sup_\omega \lambda_{\max}(f(\omega; u)) \le \sum_{j\in\mathbb{Z}} Kv(j)^{-\kappa} =: \gamma_{\sup} < \infty$. Furthermore, if we have a non-vanishing filter in the sense

$$\inf_{u\in\mathbb{R}, z\in\mathbb{C}, |z|=1} \lambda_{\min}(\sum_{j=0}^{\infty} \Psi_j(u)z^j) \ge \gamma_{\inf}^{1/2} > 0,$$

then $\inf_u \inf_\omega \lambda_{\min}(f(\omega; u)) \ge \gamma_{\inf}$. Thus the conditions in Theorem 3.2 are satisfied, and for a sufficiently large N_0, there exists $0 < \lambda_{\inf} \le \lambda_{\sup} < \infty$ such that for all $N \ge N_0$ we have

$$0 < \lambda_{\inf} \le \lambda_{\inf}(\mathbf{C}^{(N)}) \text{ and } \lambda_{\sup}(\mathbf{C}^{(N)}) \le \lambda_{\sup} < \infty.$$

In summary, the results in this section tell us the following. If an array of nonstationary time series satisfy Assumption 3.1, then there exists a stationary time series $\{X_t(u)\}$ whose covariance is $\{C_r(u)\}$. Conversely, if we define a nonstationary time series $\{X_{t,N}\}_t$ with covariance $C^{(N)}$ and an accompanying stationary time series $\{X_t(u)\}_t$ whose covariances satisfy (12) and Assumption 3.1(ii,iii), then the positive definite condition in Assumption 3.1(i) holds. One important application of this result is given in Example 3.1. However, the same result holds for more general models, including the models which satisfy the physical dependence conditions considered in Zhou and Wu (2009), Dahlhaus et al. (2019), Karmakar et al. (2021), Zhang and Wu (2021) and Ding and Zhou (2021)). In the following theorem we make this precise.

Theorem 3.3. *Suppose that $\{X_{t,N}\}_t$ is a zero mean multivariate time series of dimension p with the causal representation $X_{t,N} = G_{t,N}(\mathcal{F}_t)$ where $\mathcal{F}_t = (\varepsilon_t, \varepsilon_{t-1}, \ldots)$ and $\{\varepsilon_t\}$ are independent, identically distributed (iid) random vectors of dimension p. Associated with $\{X_{t,N}\}$ we define the multivariate stationary time series $\{X_t(u)\}_t$ where $X_t(u) = G(u, \mathcal{F}_t)$. Using $X_{t,N}$ and $X_t(u)$ we define the error process*

$$e_{t,N} = X_{t,n} - X_t(t/N) = E_{t,N}(\mathcal{F}_t).$$

*and difference process $X_t^{v_1,v_2} = (X_t(v_1) - X_t(v_2))$. Suppose $\{\widetilde{\varepsilon}_t\}_t$ are iid random vectors that are independent of $\{\varepsilon_t\}$ but with the same distribution and define $\mathcal{F}_{t|\{t-j\}} = (\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-j+1}, \widetilde{\varepsilon}_{t-j},$
$\varepsilon_{t-j-1}, \ldots)$. Then we define the coupled processes as $X_{t,N|\{t-j\}} = G_{t,n}(\mathcal{F}_{t|\{t-j\}})$, $X_{t|\{t-j\}}(u) = G(u, \mathcal{F}_{t|\{t-j\}})$, $X_{t|\{t-j\}}^{v_1,v_2} = G(v_1, \mathcal{F}_{t|\{t-j\}}) - G(v_2, \mathcal{F}_{t|\{t-j\}})$ and $e_{t,n|\{t-j\}} = E_{t,N}(\mathcal{F}_{t|\{t-j\}})$. Suppose the following hold:*

(A) Spectral-norm physical dependence

$$\begin{aligned}
&\sup_N \sup_t \| \mathbb{V}\mathrm{ar}(X_{t,N} - X_{t,N|\{t-j\}})\|_2 \leq K\delta_j \\
&\sup_u \| \mathbb{V}\mathrm{ar}(X_{t|\{t-j\}}(u) - X_{t|\{t-j\}}(u))\|_2 \leq K\delta_j \\
&\sup_t \| \mathbb{V}\mathrm{ar}(X_t^{v_1,v_2} - X_{t|\{t-j\}}^{v_1,v_2})\|_2 \leq K|v_1 - v_2|\delta_j \\
&\sup_N \sup_t \| \mathbb{V}\mathrm{ar}(e_{t,N} - e_{t,N|\{t-j\}})\|_2 \leq KN^{-1}\delta_j,
\end{aligned}$$

where $\delta_j = v(j)^{-\kappa}$, $\kappa > 3$ and K is a finite constant.

(B) Let $C_r(u) = \mathbb{C}\mathrm{ov}[X_0(u), X_r(u)]$ and $f(\omega; u) = \sum_{r\in\mathbb{Z}} C_r(u)\exp(ir\omega)$. Then we as-

sume the spectral density matrices satisfy

$$0 < \inf_{u,\omega} \lambda_{\inf} f(\omega; u) \leq \sup_{u,\omega} \lambda_{\sup} f(\omega; u) < \infty.$$

Under the above conditions, Assumption 3.1(ii,iii) is satisfied (with the same κ as that given in the conditions) and for a sufficiently large N_0, Assumption 3.1(i) is satisfied.

Proof. In Appendix C.1. □

We observe that in the theorem above the physical dependence condition (A) is described in terms of a spectral-norm of a variance.

Remark 3.2. *It is worth mentioning that Condition (A) in Theorem 3.3 is equivalent to*

$$\| \mathbb{V}\mathrm{ar}(X_{t,N} - X_{t,N|\{t-j\}})\|_2 = (\max_{\|x\|_2=1} \mathrm{E}[|x^\top (X_{t,N} - X_{t,N|\{t-j\}})|^2])^{1/2}.$$

Using the latter representation a generalisation to a bound on the qth moment: $(\max_{\|x\|_2=1} \mathrm{E}[|x^\top (X_{t,N} - X_{t,N|\{t-j\}})|^q])^{1/q}$ is possible, thus generalising physical dependence in terms of any norm. Recently, **?** *used such a generalisation.*

Example 3.3 (Locally stationary stochastic recurrence equations)**.** *We now show that the nonstationary stochastic recurrence models studied in Subba Rao (2006) and Dahlhaus et al. (2019) satisfy the conditions in Theorem 3.3.*

Let us suppose that $\{X_{t,N}\}$ has the representation

$$X_{t,N} = A(t/N, \varepsilon_t) X_{t-1,N} + b(t/N, \varepsilon_t)$$

where $\{\varepsilon_t\}$ are iid random vectors. The above model includes time-varying random coefficient vector autogressive models, time-varying vector GARCH models and Bilinear models (if the ε_t in $A(t/N, \varepsilon_t)$ were changed to ε_{t-1}) as special cases.

Based on the above model we define the stationary time series model

$$X_t(u) = A(u, \varepsilon_t) X_{t-1}(u) + b(u, \varepsilon_t).$$

Suppose $\sup_u \|\mathrm{E}[A(u, \varepsilon_t) A(u, \varepsilon_t)^\top]\|_2 < \rho < 1$, $\sup_u \|\mathrm{E}[b(u, \varepsilon_t) b(u, \varepsilon_t)^\top]\|_2 < \infty$ *and for all* v_1 *and* v_2 $\|\mathrm{E}[(A(v_1, \varepsilon_t) - A(v_2, \varepsilon_t))(A(v_1, \varepsilon_t) - A(v_2, \varepsilon_t))^\top]\|_2 \leq K|v_1 - v_2|$ *and* $\|\mathrm{E}[(b(v_1, \varepsilon_t) - b(v_2, \varepsilon_t))(b(v_1, \varepsilon_t) - b(v_2, \varepsilon_t))^\top]\|_2 \leq K|v_1 - v_2|$. *Under these conditions it can be shown* *that $X_{t,N}$ and $X_t(u)$ almost surely have the causal solution*

$$\begin{aligned} X_{t,N} &= g_{t,N}(\mathcal{F}_t) = \sum_{s=0}^{\infty} \prod_{i=0}^{s-1} A((t-i)/N, \varepsilon_{t-i}) b((t-s)/N, \varepsilon_{t-s}) \\ X_t(u) &= g(u, \mathcal{F}_t) = \sum_{s=0}^{\infty} \prod_{i=0}^{s-1} A(u, \varepsilon_{t-i}) b(u, \varepsilon_{t-s}). \end{aligned}$$

Further in Lemma C.4 we show that Condition (A) in Theorem 3.3 holds with

$$\delta_j = K(\sum_{s=j}^{\infty} s^{3/2} \rho^{(s-1)/2})^2.$$

3.3 Locally stationary approximations of the inverse covariance

In this section we show that properties on the covariance operator $\boldsymbol{C}^{(N)}$ transfer to the inverse covariance operator $\boldsymbol{D}^{(N)} = (\boldsymbol{C}^{(N)})^{-1}$. Specifically, in the following theorem we show that the relationship between $\boldsymbol{C}^{(N)}$ and $\boldsymbol{C}(u)$ in Assumption 3.1(ii,iii) carry over to $\boldsymbol{D}^{(N)}$ and $\boldsymbol{D}(u) = \boldsymbol{C}(u)^{-1}$ up to a (small) loss in rate. This result is used to show "approximate" smoothness of the time-varying VAR coefficients in representation (4).

Theorem 3.4. *Suppose Assumption 3.1 holds. Then for all $t, \tau \in \mathbb{Z}$, $D_{t-\tau}(u)$ is Lipschitz, in the sense that for all $u, v \in \mathbb{R}$*

$$\|D_{t-\tau}(u) - D_{t-\tau}(v)\|_2 \le \mathcal{K}|u-v|\zeta(\tau - t)^{\kappa-1}. \tag{13}$$

Furthermore, we have for all $t, \tau \in \mathbb{Z}$

$$\left\| \left[\boldsymbol{D}^{(N)} - \boldsymbol{D}(t/N) \right]_{t,\tau} \right\|_2 \le \mathcal{K}\zeta(t-\tau)^{\kappa-2} \min(1/N, 2\zeta(t-\tau)), \tag{14}$$

where $\mathcal{K}$ is a finite constant that is independent of u, v, t, τ.

Proof. In Appendix C.3. □

An important consequence of Theorem 3.4 is that when working with $\boldsymbol{C}$ and $\boldsymbol{D}$ it is enough to put smoothness conditions on one of them as the smoothness transfers to the other. In particular, conditions can be stated in terms of the covariance of the original time series. Furthermore, we note that differentiability conditions also transfer from $C_r(u)$ to $D_r(u)$. E.g., if one starts with the condition that for all r $\sup_u \|\frac{dC_r(u)}{du}\|_2 \le K\zeta(r)^{\kappa-1}$, then using the same arguments as those used in the proof of Theorem 3.4 (outlined after the proof of Theorem 3.4 in Appendix C.3) we have

$$\|\frac{dD_r(u)}{du}\|_2 \leq \mathcal{K}\zeta(r)^{\kappa-1}. \tag{15}$$

Smoothness and differentiability conditions on $\boldsymbol{D}^{(N)}$ and $\boldsymbol{D}(u)$ are used in Basu and Subba Rao (2022) (stated in Assumption 4.2) to obtain certain rates of decay on the Fourier transform of $\boldsymbol{D}^{(N)}$. Theorem 3.4 and (15) show that these conditions can be equivalently stated in terms of smoothness and differentiability conditions on covariance $\boldsymbol{C}^{(N)}$ and $\boldsymbol{C}(u)$. It is worth noting that the loss in the rate of decay for the inverse in Section 2 is also present in Theorem 3.4.

We now state a result that is analogous to Theorem 3.4, but for one-sided matrices. This result will be useful in proving Theorem 3.6 (below) on smoothness properties of time-varying VAR representations.

Theorem 3.5. *Suppose Assumption 3.1 holds and let* $\boldsymbol{C}^{(N)}(-\infty,T) = (C_{t,\tau}^{(N)}; t,\tau \leq T)$ *and* $\boldsymbol{C}(-\infty,T;u) = (C_{t,\tau}(u); t,\tau \leq T)$ *Then for all* $t,\tau \leq T$ *we have*

$$\begin{aligned} & \| \left[\boldsymbol{C}^{(N)}(-\infty,T)^{-1} - \boldsymbol{C}(-\infty,T;T/N)^{-1}\right]_{t,\tau} \|_2 \\ \leq \; & \mathcal{K}\zeta(t-\tau)^{\kappa-2}\min(1/N, 2\zeta(t-\tau)) \end{aligned}$$

Proof. In Appendix C.3 □

We now apply Theorem 3.4 to the popular time-varying VAR model. Let us suppose that $\{X_{t,N}\}$ has the tv-VAR(d) representation

$$X_{t,N} = \sum_{j=1}^{d} \Phi_j(t/N) X_{t-j,N} + \Sigma(t/N)^{1/2}\varepsilon_t, \qquad t \in \mathbb{Z}, \tag{16}$$

where $\{\varepsilon_t\}_t$ are uncorrelated random vectors with variance I_p. In contrast to the tv-VAR representation given in (6), the tv-VAR model is defined with Lipschitz conditions on the matrices $\Phi_j(\cdot)$ and $\Sigma(\cdot)$. The tv-VAR(d) model with smooth AR coefficients as defined in (16) is attractive because its coefficients are straightforward to interpret and has been used in econometrics and in neuroscience (see, for example, Ding et al. (2017); Safikhani and Shojaie (2020); Yan et al. (2021)). Let $\boldsymbol{C}^{(N)}$ denote the covariance corresponding to $\{X_{t,N}\}$. Obtaining a rate of decay for the covariance by directly analyzing $\boldsymbol{C}^{(N)}$ is unwieldy (see Künsch (1995) for the univariate proof). However, we show below that starting with the inverse $\boldsymbol{D}^{(N)} = (\boldsymbol{C}^{(N)})^{-1}$ (which is a banded matrix, since $X_{t,N}$ has a tvVAR(p) representation) we can use Theorem 2.1 and 3.4, to transfer the information on the rate of decay of the inverse covariance operator to the covariance operator itself.

Corollary 3.1 (Application of Theorem 3.4 to tvVAR models)**.** *Suppose that the multivariate time series $\{X_{t,N}\}_t$ has the time-varying VAR(d) representation in (16), where we assume there exists a $\delta > 0$ and γ where*

$$\inf_{u\in\mathbb{R},z\in\mathbb{C},|z|\leq 1+\delta} \lambda_{\min}(I_p - \sum_{j=1}^{d} \Phi_j(u)z^j) \geq \gamma > 0, \tag{17}$$

and the matrices $\Phi_j(\cdot)$ are Lipschitz continuous in the sense that $\|\Phi_j(u) - \Phi_j(v)\|_2 \leq K|u-v|$. We further assume that $\Sigma(\cdot)$ is Lipschitz continuous in the sense that $\|\Sigma(u) - \Sigma(v)\|_2 \leq K|u-v|$ and for all $u \in \mathbb{R}$ $\Sigma(u)$ is positive definite (with eigenvalues that are bounded from above and away from zero uniformly over all u). Let $\boldsymbol{C}^{(N)}$ denote the covariance operator of $\{X_{t,N}\}_t$ and $C_r(u) = \int_0^{2\pi} f(\omega;u)\exp(-ir\omega)d\omega$, where $f(\omega;u) = [I_p - \sum_{j=1}^{d}\Phi_j(u)\exp(-ij\omega)]^{-1}\Sigma(u)([I_p - \sum_{j=1}^{d}\Phi_j(u)\exp(ij\omega)]^{-1})^{\top}$. Then, there exists an N_0 and $0<\rho<1$ such that for all $N > N_0$ we have $\|C_{t,\tau}^{(N)}\|_2 \leq \mathcal{K}\rho^{|t-\tau|}$, $\|C_r(u) - C_r(v)\|_2 \leq \mathcal{K}|u-v|\rho^{|t-\tau|}$, and $\|C_{t,\tau}^{(N)} - C_{t-\tau}(t/N)\|_2 \leq \mathcal{K}\rho^{|t-\tau|}/N$.

Proof. In Appendix C.3. □

Remark 3.3 (Differentiability of the tv-VAR covariance)**.** *As mentioned after Theorem 3.4 in equation (15), smoothness conditions in terms of differentiability transfer between $\boldsymbol{C}(\cdot)$ and $\boldsymbol{D}(\cdot)$. For the tv-VAR model this implies that smoothness conditions formulated in terms of differentiability of the transition matrices $\Phi_j(\cdot)$ transfer to $\boldsymbol{D}(\cdot)$, then from equation (15) they transfer to $\boldsymbol{C}(\cdot)$. Ding et al. (2017), Lemma 3.1 also prove that differentiability of $\Phi_1(\cdot)$ implies differentiability of the covariance for tv-VAR(1) models. They show this result by directly connecting the covariance to $\Phi_1(\cdot)$ through the tv-VAR(1) model. However, their proof requires the additional condition that $\|\Phi_1\|_1 = \max_{\|x\|_1=1}\|\Phi_1 x\|_1 < 1$, which places quite strict conditions on the VAR parameters.*

We apply Corollary 3.1 to the time-varying ARCH process.

Example 3.4 (The time-varying ARCH(p) process)**.** *The time-varying ARCH model is* *defined as follows. Let*

$$X_{t,N} = \sigma_{t,N} Z_t \qquad \sigma_{t,N}^2 = a_0(t/N) + \sum_{j=1}^{p} a_j(t/N) X_{t-j}^2,$$

where $\{Z_t\}$ *are iid random variables with mean zero and variance one and the coefficients* $a_j(\cdot)$ *are Lipschitz continuous and such that* $\inf_u a_0(u) > 0$*, for* $1 \leq j \leq p$ $a_j(\cdot) \geq 0$ *and are such that* $\sup_u (\mathrm{E}[Z_0^4])^{1/2} \sum_{j=1}^{p} a_j(u) < 1$*. Under these conditions it can be shown that* $\sup_{t,N} \mathrm{E}[X_{t,N}^4] < \infty$ *and that* $X_{t,N}^2$ *has the tvAR(p) representation*

$$X_{t,N}^2 = a_0(t/N) + \sum_{j=1}^{p} a_j(t/N) X_{t-j,N}^2 + \varepsilon_{t,N}$$

where $\varepsilon_{t,N} = \sigma_{t,N}^2 (Z_t^2 - 1)$*. The condition that* $\sup_u \sum_{j=1}^{p} a_j(u) < 1$ *(which is implied by the condition* $\sup_u (\mathrm{E}[Z_0^4])^{1/2} \sum_{j=1}^{p} a_j(u) < 1$*) implies that (17) holds. Thus by applying Corollary 3.1, Assumption 3.1 holds for the time-varying ARCH process.*

We have shown in (4) that under certain conditions all nonstationary time series have an AR(∞) representation. But there is no guarantee that the AR parameters are smooth. Below, we show below that under the locally stationary conditions in Assumption 3.1 a smooth *approximation* is possible.

We recall from (4) that $\{X_{T,N}\}_t$ has the representation

$$X_{T,N} = \sum_{j=1}^{\infty} \Phi_{T,j}^{(N)} X_{T-j,N} + \varepsilon_{T,N}, \tag{18}$$

where $\{\varepsilon_{T,N}\}_{t \in \mathbb{Z}}$ are uncorrelated random vectors with $\Sigma_{T,N} = \mathbb{V}\mathrm{ar}[\varepsilon_{T,N}]$. We have shown in Section 3.2 that under Assumption 3.1 there exists a stationary time series $\{X_t(u)\}_t$ with autocovariance $\{C_r(u)\}_r$. Using the arguments leading to (4), it can be shown that $\{X_t(u)\}_t$ has the VAR(∞) representation

$$X_t(u) = \sum_{j=1}^{\infty} \Phi_j(u) X_{t-j}(u) + \varepsilon_t(u), \tag{19}$$

where $\varepsilon_t(u)$ are uncorrelated random vectors with variance $\Sigma(u) = \mathbb{V}\mathrm{ar}[\varepsilon_t(u)]$. In the following theorem we show that $\{\Phi_{T,j}^{(N)}\}$ can be approximated by the stationary VAR coefficients $\{\Phi_j(u)\}$.

Theorem 3.6. *Suppose the array of time series $\{X_{T,N}\}_t$ satisfy Assumption 3.1 and let $\{\Phi_{t,j}^{(N)}\}_j$ be defined as in (18) with $\Sigma_T^{(N)} = \mathbb{V}\mathrm{ar}[\varepsilon_{T,N}]$. Additionally, let $\{X_t(u)\}_t$ be the locally stationary approximation defined in (19).*

(i) Then for all $T \in \mathbb{Z}$ and $j \geq 1$ we have

$$\begin{aligned} \|\Sigma_T^{(N)} - \Sigma(T/N)\|_2 &\leq \frac{\mathcal{K}}{N} \\ \text{and } \|\Phi_{T,j}^{(N)} - \Phi_j(T/N)\|_2 &\leq \mathcal{K}\zeta(j)^{\kappa-2}\min(2\zeta(j), 1/N) \end{aligned}$$

(ii) For all $u_1, u_2 \in \mathbb{R}$ and $j \geq 1$

$$\begin{aligned} \|\Sigma(u_1) - \Sigma(u_2)\|_2 &\leq \mathcal{K}|u_1 - u_2| \\ \text{and } \|\Phi_j(u_1) - \Phi_j(u_2)\|_2 &\leq \mathcal{K}\zeta(j)^{\kappa-1}|u_1 - u_2|. \end{aligned}$$

Proof. In Appendix C.3 □

Remark 3.4 (Approximation and estimation by finite order tv-VAR)**.** *From above theorem, if a process is locally stationary then it can be approximated by an time-varying VAR(∞) time series with slowly varying parameters. Consequently by using Meyer and Kreiss (2015) this implies that the locally stationary time series can be approximated with a finite order time-varying VAR(d) with slowly varying parameters. More precisely, let $\{\Phi_{d,j}(u)\}_{j=1}^d$ denote the finite order (stationary) VAR(d) parameters associated with the vector autocovariance $\{C_r(u)\}_r$. Then by Meyer and Kreiss (2015) we have* $\sup_u \sum_{j=1}^d \|\Phi_{d,j}(u) - \Phi_j(u)\|_2 \leq \mathcal{K}d^{-\kappa}$. *This result together with Theorems 2.2, 3.6 and the triangle inequality gives*

$$\sum_{j=1}^{d} \|\Phi_{T,d,j} - \Phi_{d,j}(T/N)\|_2 \leq \mathcal{K}\left(N^{-1} + \zeta(d)^{\kappa-3/2}\right).$$

A potential application of Theorem 3.6 is that it could be used (i) in forecasting and (ii) to develop a bootstrap procedure for nonstationary time series by transferring the widely used stationary AR-sieve to the locally stationary setup. Both procedures would require estimators of the the finite order time-varying VAR parameters $\{\Phi_{d,j}(T/N)\}$. One could estimate this using local kernel methods or the sieve estimation method described in Ding and Zhou (2020).

Remark 3.5 (Innovations and Kolomogorov's formula)**.** *An immediate implication of the* *above result is that the time varying innovation variance* $\Sigma_t^{(N)}$ *can be approximated by Kolomogorov's formula*

$$det[\Sigma_t^{(N)}] = \int_{-\pi}^{\pi} \log det[f(t/N;\omega)]d\omega + O(1/N)$$

where $f(u;\omega) = \sum_{r\in\mathbb{Z}} C_r(u)\exp(ir\omega)$*. A similar result was obtained in Liu et al. (2021), Proposition 1 for a specific class of locally stationary time series.*

3.4 The partial covariance of a locally stationary time series

The partial covariance is commonly used in the analysis of time series as a measure of linear dependence between two time series after accounting for all the other components in the time series. For stationary time series, the analysis is typically conducted through the partial spectral coherence which is the standardized Fourier transform of the partial covariance, and is, conveniently, a function of the spectral density matrix function (cf. Priestley (1981); Brillinger (2001); Dahlhaus (2000a); Krampe and Paparoditis (2022)). For nonstationary time series the time-varying partial spectral coherence can be defined as a function of the localized inverse spectral density, as was done in Park et al. (2014). However, as far as we are aware, there are no results that connect this definition (of the time-varying partial spectral coherence) to the actual partial covariance of the underlying nonstationary time series.

We use the results on inverse covariances (developed in Section 3.3) to show that the partial covariance of a locally stationary time series (as defined in Assumption 3.1) can be approximated by a smooth function, which, in turn, is the partial covariance of the locally stationary approximation $\{X_t(u)\}_t$. We show below that this result can be used to justify using the time-varying partial spectral coherence as an approximation of the Fourier transform of the localized partial covariance.

We start by defining the partial covariance for nonstationary time series. For this, let $\mathcal{H}^{(N)} = \overline{\text{sp}}(X_{t,N}^{(c)}; t \in \mathbb{Z}, 1 \leq c \leq p)$ denote the space spanned by the entire multivariate time series. Furthermore, let $\mathcal{S} \subseteq \{1, \ldots, p\} =: V$ be a set of indices referring to components of the time series and $\mathcal{H}^{(N)} - (X^{(c)}; c \in \mathcal{S}) = \overline{\text{sp}}[X_{s,N}^{(c)}; s \in \mathbb{Z}, c \in \mathcal{S}']$ be the space spanned by the entire time series of the components in $\mathcal{S}'$ only, where $\mathcal{S}'$ denotes the complement of $\mathcal{S}$. Let $P_{\mathcal{M}}(Y)$ denote the orthogonal projection of $Y \in \mathcal{H}^{(N)}$ onto the subspace $\mathcal{M}$. For any $\mathcal{S} \subseteq V$, we define the residual of $X_{t,N}^{(a)}$ after projecting on

$\mathcal{H}^{(N)} - (X^{(c)}; c \in \mathcal{S})$ as

$$X_{t,N}^{(a)|\text{-}\mathcal{S}} \quad := \quad X_{t,N}^{(a)} - P_{\mathcal{H}^{(N)}-(X^{(c)};c\in\mathcal{S})}(X_{t,N}^{(a)}), t \in \mathbb{Z}. \tag{20}$$

In the definitions below we focus on the two sets $\mathcal{S} = \{a, b\}$ and $\mathcal{S} = \{a\}, a, b \in V, a \neq b$. Using the above, we define the partial covariance

$$\Delta_{t,\tau,N}^{\text{-}\{a,b\}} = \begin{pmatrix} \rho_{t,\tau,N}^{(a,a)|\text{-}\{a,b\}} & \rho_{t,\tau,N}^{(a,b)|\text{-}\{a,b\}} \\ \rho_{t,\tau,N}^{(b,a)|\text{-}\{a,b\}} & \rho_{t,\tau,N}^{(b,b)|\text{-}\{a,b\}} \end{pmatrix} := \mathbb{C}\text{ov}\left[\begin{pmatrix} X_{t,N}^{(a)|\text{-}\{a,b\}} \\ X_{t,N}^{(b)|\text{-}\{a,b\}} \end{pmatrix}, \begin{pmatrix} X_{\tau,N}^{(a)|\text{-}\{a,b\}} \\ X_{\tau,N}^{(b)|\text{-}\{a,b\}} \end{pmatrix}\right] \tag{21}$$

and self partial covariance

$$\rho_{t,\tau,N}^{(a,a)|\text{-}\{a\}} = \mathbb{C}\text{ov}[X_{t,N}^{(a)|\text{-}\{a\}}, X_{\tau,N}^{(a)|\text{-}\{a\}}]. \tag{22}$$

As will become clear in the proof of the following theorem $\Delta_{t,\tau,N}^{\text{-}\{a,b\}}$ and $\rho_{t,\tau,N}^{(a,a)|\text{-}\{a\}}$ can be expressed in terms of the matrix operator $\boldsymbol{C}^{(N)}$ and its inverse. Under Assumption 3.1 and by Theorem 3.1 there exists a stationary time series $\{X_t(u)\}_t$ which has covariance $\boldsymbol{C}(u)$, that locally approximates $\boldsymbol{C}^{(N)}$. Using $\boldsymbol{C}(u)$ we will define the partial covariances corresponding to the stationary time series $\{X_t(u)\}_t$. In the theorem below we show that the partial covariances of $\{X_t(u) = (X_t^{(1)}(u), \ldots, X_t^{(p)}(u))^\top\}_t$ locally approximates the partial covariance of $\{X_{t,N} = (X_{t,N}^{(1)}, \ldots, X_{t,N}^{(p)})^\top\}_t$. To do this, analogous to (20), (21) and (22) we define

$$X_t^{(a)|\text{-}\mathcal{S}}(u) \quad := \quad X_{t,N}^{(a)}(u) - P_{\mathcal{H}_u-(X_u^{(c)};c\in\mathcal{S})}(X_t^{(a)}(u)) \text{ for } t \in \mathbb{Z}, \tag{23}$$

$$\Delta_{t-\tau}^{\text{-}\{a,b\}}(u) = \begin{pmatrix} \rho_{u,t-\tau}^{(a,a)|\text{-}\{a,b\}} & \rho_{u,t-\tau}^{(a,b)|\text{-}\{a,b\}} \\ \rho_{u,t-\tau}^{(b,a)|\text{-}\{a,b\}} & \rho_{u,t-\tau}^{(b,b)|\text{-}\{a,b\}} \end{pmatrix} := \mathbb{C}\text{ov}\left[\begin{pmatrix} X_t^{(a)|\text{-}\{a,b\}}(u) \\ X_t^{(b)|\text{-}\{a,b\}}(u) \end{pmatrix}, \begin{pmatrix} X_\tau^{(a)|\text{-}\{a,b\}}(u) \\ X_\tau^{(b)|\text{-}\{a,b\}}(u) \end{pmatrix}\right] \tag{24}$$

and self partial covariance

$$\rho_{t-\tau}^{(a,a)|\text{-}\{a\}}(u) = \mathbb{C}\text{ov}[X_t^{(a)|\text{-}\{a\}}(u), X_\tau^{(a)|\text{-}\{a\}}(u)]. \tag{25}$$

We note that a key ingredient in the proof of the theorem below is that the partial covariance can be expressed as

$$\mathbb{V}\text{ar}\left[X_{t,N}^{(e)|\text{-}\{a,b\}}; t \in \mathbb{Z}, e \in \{a, b\}\right] = \boldsymbol{C}_{\mathcal{S},\mathcal{S}} - \boldsymbol{C}_{\mathcal{S},\mathcal{S}'}\boldsymbol{C}_{\mathcal{S}',\mathcal{S}'}^{-1}\boldsymbol{C}_{\mathcal{S},\mathcal{S}'}^\top,$$

where $\mathcal{S} = \{a, b\}$, $\boldsymbol{C}_{\mathcal{S},\mathcal{S}} = (\boldsymbol{C}^{(e,f)}; e, f \in \mathcal{S})$ (similarly for $\boldsymbol{C}_{\mathcal{S},\mathcal{S}'}$ and $\boldsymbol{C}_{\mathcal{S}',\mathcal{S}'}$) and $\boldsymbol{C}^{(e,f)} = (\mathbb{C}\text{ov}[X^{(e)}_{t,N}, X^{(f)}_{\tau,N}]; t, \tau \in \mathbb{Z})$. The presence of $\boldsymbol{C}^{-1}_{\mathcal{S}',\mathcal{S}'}$ in the above expression explains why the results in the previous sections (in particular Theorem 3.4) are necessary for proving the result.

Theorem 3.7. *Suppose Assumption 3.1 holds and let further* $\Delta^{-\{a,b\}}_{t,\tau,N}$, $\rho^{(a,a)|-\{a\}}_{t,\tau,N}$, $\Delta^{-\{a,b\}}_{t-\tau}(u)$, *and* $\rho^{(a,a)|-\{a\}}_{t-\tau}(u)$ *be defined as in (21), (22), (24) and (25). Then for all* $a, b \in \{1, \ldots, p\}$

$$\begin{aligned}
\|\Delta^{-\{a,b\}}_{t,\tau,N} - \Delta^{-\{a,b\}}_{t-\tau}(t/N)\|_2 &\leq \mathcal{K}\zeta(t-\tau)^{\kappa-2}\min(1/N, \zeta(t-\tau)) &(26)\\
\|\Delta^{-\{a,b\}}_{t-\tau}(u) - \Delta^{-\{a,b\}}_{t-\tau}(v)\|_2 &\leq \mathcal{K}|u-v|\zeta(t-\tau)^{\kappa-1} &(27)\\
\|\rho^{(a,a)|-\{a\}}_{t,\tau,N} - \rho^{(a,a)|-\{a\}}_{t-\tau}(t/N)\|_2 &\leq \mathcal{K}\zeta(t-\tau)^{\kappa-2}\min(1/N, \zeta(t-\tau)) &(28)\\
\textit{and } \|\rho^{(a,a)|-\{a\}}_{t-\tau}(u) - \rho^{(a,a)|-\{a\}}_{t-\tau}(v)\|_2 &\leq \mathcal{K}|u-v|\zeta(t-\tau)^{\kappa-1}, &(29)
\end{aligned}$$

where $\mathcal{K}$ *is a positive generic constant.*

Proof. In Appendix C.3. □

The above result provides the tools to prove the following. Let $\{X_{t,N}\}_t$ be an array of nonstationary time series that satisfy Assumption 3.1 and $\{C_r(u)\}_r$ the corresponding stationary approximation covariance. Let $f(\omega; u) = \sum_{r\in\mathbb{Z}} C_r(u)e^{ir\omega}$ and $\Gamma(\omega; u) = f(\omega; u)^{-1}$. Using the stationary partial spectral coherence (see Priestley (1981), Section 9.3 and Dahlhaus (2000a)), the localized (complex) partial spectral coherence is defined as

$$g_{a,b}(\omega; u) = -\frac{\Gamma^{(a,b)}(\omega; t/N)}{(\Gamma^{(a,a)}(\omega; t/N)\Gamma^{(b,b)}(\omega; t/N))^{1/2}},$$

where $\Gamma^{(a,b)}(\omega)$ denotes the (a, b) entry of the matrix $\Gamma(\omega; u)$. Under Assumption 3.1 (for $\kappa > 3$) and by using Theorem 3.7 it can be shown that

$$\frac{\sum_{r\in\mathbb{Z}} \rho^{(a,b)|-\{a,b\}}_{t,t+r,N} \exp(ir\omega)}{\sqrt{\sum_{r\in\mathbb{Z}} \rho^{(a,a)|-\{a,b\}}_{t,t+r,N} \exp(ir\omega) \sum_{r\in\mathbb{Z}} \rho^{(b,b)|-\{a,b\}}_{t,t+r,N} \exp(ir\omega)}} = g_{a,b}(\omega; t/N) + O(N^{-1}).$$

In other words, the estimated local partial spectral coherence (based on an estimator of the local spectral density function) is an estimator of the Fourier transform of the partial covariances of the nonstationary time series localised about time point t. This justifies using local spectral density estimation approaches for estimating the partial covariance.

Acknowledgements

The authors would like to thank two anonymous referees for their comments and suggestions, which greatly improved all aspects of the paper. JK was supported by the Research Center (SFB) 884 "Political Economy of Reforms"(Project B6), funded by the German Research Foundation (DFG), and acknowledges the partial support of DFG (travel grant 493207657) and National Institute of Health (grants R01GM135926 and R21NS120227). SSR research was partially supported by the National Science Foundation (grant DMS-1812128 and DMS-2210726).

References

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

Basu, S. and Subba Rao, S. (2022). Graphical models for nonstationary time series. *To appear Annals of Statistics*.

Böttcher, A. and Grudsky, S. M. (2000). *Toeplitz matrices, asymptotic linear algebra, and functional analysis*. Birkhäuser Verlag, Basel.

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

Cheng, R. and Pourahmadi, M. (1993). Baxter's inequality and convergence of finite predictors of multivariate stochastic processess. *Probability Theory and Related Fields*, 95(1):115–124.

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

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

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

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

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

Dahlhaus, R. and Subba Rao, S. (2006). Statistical inference for time-varying ARCH processes. *Ann. Statist.*, 34(3):1075–1114.

Demko, S., Moss, W. F., and Smith, P. W. (1984). Decay rates for inverses of band matrices. *Mathematics of computation*, 43(168):491–499.

Ding, X., Qiu, Z., and Chen, X. (2017). Sparse transition matrix estimation for high-dimensional and locally stationary vector autoregressive models. *Electronic Journal of Statistics*, 11(2):3871–3902.

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

Ding, X. and Zhou, Z. (2021). Auto-regressive approximations to non-stationary time series, with inference and applications. *https://arxiv.org/abs/2112.00693*.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and statistics*, 82(4):540–554.

Hannan, E. and Deistler, M. (1988). *The Statistical Theory of Linear Systems*, volume 70. SIAM.

Karmakar, S., Richter, S., and Wu, W. B. (2021). Simultaneous inference for time-varying models. *Journal of Econometrics*.

Krampe, J., Kreiss, J.-P., and Paparoditis, E. (2021). Bootstrap based inference for sparse high-dimensional time series models. *Bernoulli*, 27(3):1441–1466.

Krampe, J. and McMurry, T. L. (2021). Estimating wold matrices and vector moving average processes. *Journal of Time Series Analysis*, 42(2):201–221.

Krampe, J. and Paparoditis, E. (2021). Sparsity concepts and estimation procedures for high-dimensional vector autoregressive models. *Journal of Time Series Analysis*, 42(5-6):554–579.

Krampe, J. and Paparoditis, E. (2022). Spectral analysis of high-dimensional time series. *Preprint*.

Künsch, H. R. (1995). A note on causal solutions for locally stationary AR-processes. *Technical Report*.

Liu, Y., Taniguchi, M., and Ombao, H. (2021). Statistical inference for local granger causality. *https://arxiv.org/abs/2103.00209*.

Lütkepohl, H. (1996). *Handbook of Matrices*. Wiley.

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

Meyer, M. and Kreiss, J.-P. (2015). On the vector autoregressive sieve bootstrap. *Journal of Time Series Analysis*, 36(3):377–397.

Park, T., Eckley, A. I., and Ombao, H. (2014). Estimating time-evolving partial coherence between signals via multivariate locally stationary wavelet processes. *IEEE Transactions in Signal Processing*, 62:5420–5250.

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

Safikhani, A. and Shojaie, A. (2020). Joint structural break detection and parameter estimation in high-dimensional nonstationary VAR models. *Journal of the American Statistical Association*, 0(0):1–14.

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

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

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

Truquet, L. (2019). Local stationarity and time-inhomogeneous markov chains. *Ann. Statist.*, 47:2023–2050.

Vogt, M. (2012). Nonparametric regression for locally stationary time series. *Ann. Statist.*, 40:2601–2633.

Wiener, N. and Masani, P. (1958). The prediction theory of multivariate stochastic processes, ii. *Acta Mathematica*, 99(1):93–137.

Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences*, 102:14150–14154.

Yan, Y., Gao, J., and Peng, B. (2021). On time-varying VAR models: Estimation, testing and impulse response analysis. *arXiv preprint arXiv:2111.00450*.

Zhang, D. and Wu, W. B. (2021). Second-order asymptotics for high dimensional locally stationary processes. *Annals of Statistics*, 49:233–254.

Zhou, Z. and Wu, W. B. (2009). Local linear quantile estimation for nonstationary time series. *Ann. Statist.*, 37(5B):2696–2729.

A Supplementary material

A.1 Summary of results in supplementary material

In order to navigate the appendix we summarize below the contents and main results in the appendix.

- Appendix A gives all the background lemmas.

  In Appendix A.2 we state all the block operator identities that are required in this paper. In Appendix A.3 we state and derive several matrix norm inequalities. This includes a Cauchy-Schwarz type bound for the spectral norm of cross covariance matrices (Lemma A.3).

- In Appendix B we prove the results in Section 2. A fundamental result required in the proof is Lemma B.1 which gives a bound on the entries for the inverse of block banded matrices.

- In Appendix C we prove the results for Section 3.

  The proofs for Theorems 3.1 and 3.2 are given in Appendix C.1. In Appendix C.2 we consider models which satisfy the physical dependence conditions first proposed in Wu (2005). The proof of Theorems 3.4, 3.5 and 3.7 follow a similar set of arguments and are given in Appendix C.3.

A.2 Notation and background

Before proceeding with the proofs, we need to introduce some notation. We define below unit vectors of appropriate dimension to select sub-matrices or elements from the operator $\boldsymbol{A} : \ell_{2,p} \to \ell_{2,p}$. That is, $A_{s_1,s_2} = (e_{s_1} \otimes I_p)^\top \boldsymbol{A}(e_{s_2} \otimes I_p)$, where $\otimes$ is the Kronecker product and I_p denotes the identity operator in $\mathbb{R}^p$. Furthermore, $A^{(a,b)}_{s_1,s_2} = (e_{s_1} \otimes e_a)^\top \boldsymbol{A}(e_{s_2} \otimes e_b)$ and we introduce the short notation for this unit vector as $e_{(a,s)} = (e_s \otimes e_a)$.

In the proofs below we will often consider sub-matrices, where one column or row has been removed. To set-up the matrix notation for this, let I denote the identity operator in ℓ_2 and I_{-k} the identity operator after removing the kth row, i.e., for $u \in \ell_2$, $I_{-k}u = (\ldots, u_{-1}, u_0, u_1, \ldots, u_{k-1}, u_{k+1}, \ldots)$. The same notation is used for operators in $\mathbb{R}^p$ and similar spaces. This results in the following operations applied to an operator $\boldsymbol{A} : \ell_{2,p} \to \ell_{2,p}$:

- $(e_{s_1} \otimes e_a)^\top \boldsymbol{A}(e_{s_2} \otimes e_b) = A^{(a,b)}_{s_1,s_2}$.
- $(I_{-k} \otimes I_p)\boldsymbol{A}$ removes from the infinite dimensional matrix p rows (of infinite length) so that $A^{(a,b)}_{k,i}$ is removed for all $i \in \mathbb{Z}, a, b \in \{1, \ldots, p\}$.
- $\boldsymbol{A}(I_{-k} \otimes I_p)^\top$ removes from the infinite dimensional matrix p columns (of infinite length) so that $A^{(a,b)}_{i,k}$ is removed for all $i \in \mathbb{Z}, a, b \in \{1, \ldots, p\}$.
- $(I_{-k} \otimes I_p)\boldsymbol{A}(I_{-k} \otimes I_p)^\top =: \widetilde{\boldsymbol{A}}$ is an infinite dimensional matrix where $A^{(a,b)}_{i,k}$ and $A^{(a,b)}_{k,i}$ are removed for all $i \in \mathbb{Z}, a, b \in \{1, \ldots, p\}$.
- $(I_{-k} \otimes I_p)^\top \widetilde{\boldsymbol{A}}(I_{-k} \otimes I_p) = \boldsymbol{B}$ is an infinite dimensional matrix where p zero columns and rows (of infinite length) are added so that $B^{(a,b)}_{i,k} = 0, B^{(a,b)}_{k,i} = 0$ for all $i \in \mathbb{Z}, a, b \in \{1, \ldots, p\}$. Additionally, for all $s_1 \neq k$ and $s_2 \neq k$ we have $((I_{-k} \otimes I_p)^\top \widetilde{\boldsymbol{A}}(I_{-k} \otimes I_p))_{s_1,s_2} = (\boldsymbol{A})_{s_1,s_2}$.
- $(I_{-k} \otimes I_{-a})\boldsymbol{A}$ removes from the infinite dimensional matrix $p-1$ rows (of infinite length) so that $A^{(c,b)}_{k,i}$ is removed for all $i \in \mathbb{Z}, a, b \in \{1, \ldots, p\}, c \neq a$.
  Similarly, for the other operations used above.
  We denote $(I_{-k} \otimes I_{-a}) =: I_{-(a,k)}$.
- We have that $(I_{-k} \otimes I_p)^\top (I_{-k} \otimes I_p)$ is the identity operator on the reduced space and $(I_{-k} \otimes I_p)(I_{-k} \otimes I_p)^\top + (e_k \otimes I_p)(e_k \otimes I_p)^\top = \boldsymbol{I} = (I \otimes I_p)$, where $\boldsymbol{I}$ is the identify on the full space. Furthermore, $(e_k \otimes I_p)^\top (I_{-k} \otimes I_p) = 0$.

- For $x \in \ell_q, q \in [1,\infty]$ we define $\|x\|_q = (\sum_{l\in Z} x_l^q)^{1/q}$ and $\|x\|_\infty = \max_{l\in\mathbb{Z}} |x_l|$. For an operator $\boldsymbol{B} : \ell_2 \to \ell_2$, we also define the ℓ_q-induced norms, that is for $q \in [1,\infty]$ we set $\|\boldsymbol{B}\|_q =: \sup_{\|x\|_q=1, x\in\ell_2} \|\boldsymbol{B}x\|_q$, where $\|\boldsymbol{B}\|_\infty = \sup_{s_1\in\mathbb{Z}} \sum_{s_2\in\mathbb{Z}} |B_{s_1,s_2}|$.

An important tool in the proofs is the inversion and manipulation of infinite dimensional (block) matrices. Under certain conditions on both the matrices and the spaces we can treat these in much the same way as finite dimensional matrices. An identity that we will make frequent use of is the analogous version of the block inversion identity but for infinite dimensional operators. Suppose that $\boldsymbol{U} : (S_1, S_2) \to (S_1, S_2)$ where S_1 and S_2 are two Hilbert spaces and

$$\boldsymbol{U} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix}.$$

If the eigenvalues of $\boldsymbol{U}$ are bounded away from zero and from infinite, then using equation (1.7.4) in Tretter (2008), page 43 (setting $\lambda = 0$) for the inversion of block operator matrices we have

$$\boldsymbol{U}^{-1} = \begin{pmatrix} \widetilde{\boldsymbol{A}} & \widetilde{\boldsymbol{B}} \\ \widetilde{\boldsymbol{C}} & \widetilde{\boldsymbol{D}} \end{pmatrix} = \begin{pmatrix} \widetilde{\boldsymbol{A}} & -\widetilde{\boldsymbol{A}}\boldsymbol{B}\boldsymbol{D}^{-1} \\ -\boldsymbol{D}^{-1}\boldsymbol{C}\widetilde{\boldsymbol{A}} & \boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{C}\widetilde{\boldsymbol{A}}\boldsymbol{B}\boldsymbol{D}^{-1} \end{pmatrix} \quad (30)$$

where from Definition 1.6.1 in Tretter (2008), page 35 we have

$$\widetilde{\boldsymbol{A}} = (\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1} \text{ and } \widetilde{\boldsymbol{D}} = (\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}. \quad (31)$$

An immediately consequence of the above is that the difference in the block diagonal entries is

$$\boldsymbol{A} - \widetilde{\boldsymbol{A}}^{-1} = \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C} = \boldsymbol{B}(\widetilde{\boldsymbol{D}} - \widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{A}}^{-1}\widetilde{\boldsymbol{B}})\boldsymbol{C}. \quad (32)$$

We will make frequent use of (30) and (32) in the proofs.

A.3 Some background results

Lemma A.1. *Suppose that $\{A_\ell\}_{\ell=1}^\infty$ is a sequence of $p \times p$ dimensional matrices and $\sum_{\ell=1}^\infty \|A_\ell\|_2^2 < \infty$. Define the sequence space $\ell_{2,p}^+ = \{w = (v_1, v_2, \ldots) : v_j \in \mathbb{R}^p\}$ and the* *linear operator* $\boldsymbol{A} = (A_\ell; \ell \geq 0)$*, where* $\boldsymbol{A} : \ell_{2,p}^{+} \to \ell_{2,p}^{+}$*. Then*

$$\|\boldsymbol{A}\|_2 \leq (\sum_{\ell=1}^{\infty} \|A_\ell\|_2^2)^{1/2}$$

Proof. Let $x = (x_1, x_2, \ldots)$ where $x_l \in \mathbb{R}^p$. By definition of the $\|\cdot\|_2$ operator norm we have

$$\begin{aligned}
\|\boldsymbol{A}\|_2 = & \sup_{\|x\|_2=1, x\in \ell_{2,p,1}} x^\top \boldsymbol{A}^\top \boldsymbol{A} x = \sup_{\|x\|_2=1, x\in \ell_{2,p,1}} (\sum_{l_1,l_2=1}^{\infty} x_{l_1}^\top A_{l_1}^\top A_{l_2} x_{l_2})^{1/2} \\
& \leq \sup_{\|x\|_2=1, x\in \ell_{2,p,1}} \sum_{l=1}^{\infty} \|x_l\|_2 \|A_l\|_2 \\
& \leq \sup_{\|x\|_2=1, x\in \ell_{2,p,T}} (\sum_{l=1}^{\infty} \|x_l\|_2^2)^{1/2} (\sum_{l=1}^{\infty} \|A_l\|_2^2)^{1/2} \text{ (by the Cauchy-Schwarz inequality)} \\
& = (\sum_{l=1}^{\infty} \|A_l\|_2^2)^{1/2},
\end{aligned}$$

thus proving the result. □

We use the following result in the proof of Lemma B.2 and Theorem 3.1.

Lemma A.2. *Let* $\boldsymbol{B}$ *be a symmetric linear operator from* $\ell_{2,p}$ *to* $\ell_{2,p}$ *with* $\|\boldsymbol{B}\|_2 < \infty$*. Then,*

$$\|\boldsymbol{B}\|_2 \leq \max_{s_1} \sum_{s_2\in\mathbb{Z}} \|B_{s_1,s_2}\|_2$$

Proof. To prove the result we define the following operator based on $\boldsymbol{B}$. Let $\widetilde{\boldsymbol{B}} = (\|B_{s_1,s_2}\|_2)_{s_1,s_2}$ be an operator from ℓ_2 to ℓ_2. Since $\boldsymbol{B}$ is symmetric, we have

$$\begin{aligned}
\|\boldsymbol{B}\|_2 = & \sup_{\|x\|_2=1} x^\top \boldsymbol{B} x = \sup_{\|x\|_2=1} \sum_{s_1,s_2\in\mathbb{Z}} x_{s_1}^\top B_{s_1,s_2} x_{s_2} \leq \sup_{\|x\|_2=1} \sum_{s_1,s_2\in\mathbb{Z}} \|x_{s_1}\|_2 \|B_{s_1,s_2}\|_2 \|x_{s_2}\|_2 \\
& = \|\widetilde{\boldsymbol{B}}\|_2 \leq \|\widetilde{\boldsymbol{B}}\|_\infty = \max_{s_1} \sum_{s_2\in\mathbb{Z}} \|B_{s_1,s_2}\|_2.
\end{aligned}$$

This proves the result. □

The following lemma is a generalisation of the Cauchy-Schwarz inequality to the spectral norm of matrices.

Lemma A.3. *Let X and Y be finite dimensional random vectors (not necessarily of the same dimension). Then, we have*

$$\| \mathbb{C}\mathrm{ov}(Y, X)\|_2^2 \leq \| \mathbb{V}\mathrm{ar}(X)\|_2 \| \mathbb{V}\mathrm{ar}(Y)\|_2.$$

A generalisation of the above result is to the case that A and B denote two conformable random matrices. Then

$$\|\mathrm{E}(AB)\|_2^2 \leq \|\mathrm{E}(AA^\top)\|_2 \|\mathrm{E}(BB^\top)\|_2$$

Proof. To prove the result we start by first assuming that $\mathbb{V}\mathrm{ar}(X)$ is strictly positive definite and later relax this condition to the case that $\mathbb{V}\mathrm{ar}(X)$ is non-negative definite. Let

$$\mathbb{V}\mathrm{ar}((X^\top, Y^\top)^\top)) = \begin{pmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{pmatrix}.$$

Since Σ is a positive semi-definite matrix, $\Sigma_{2,2} - \Sigma_{2,1}\Sigma_{1,1}^{-1}\Sigma_{1,2}$ is a positive semi-definite matrix. Hence, we have (see, for example, (Lütkepohl, 1996, p.76))

$$\|\Sigma_{2,2}\|_2 \geq \|\Sigma_{2,1}\Sigma_{1,1}^{-1}\Sigma_{1,2}\|_2 = \|\Sigma_{2,1}\Sigma_{1,1}^{-1/2}\|_2^2.$$

Thus,

$$\| \mathbb{C}\mathrm{ov}(Y, X)\|_2^2 = \|\Sigma_{2,1}\Sigma_{1,1}^{-1/2}\Sigma_{1,1}^{1/2}\|_2^2 \leq \|\Sigma_{2,1}\Sigma_{1,1}^{-1/2}\|_2^2 \|\Sigma_{1,1}^{1/2}\|_2^2 \leq \|\Sigma_{2,2}\|_2 \|\Sigma_{1,1}\|_2.$$

We now generalise the proof to the case that $\Sigma_{11} = \mathbb{V}\mathrm{ar}(X)$ is non-negative definite. For this note that we have the eigenvalue decomposition $\mathbb{V}\mathrm{ar}(X) = B\Lambda B^*$. In the case that $\mathbb{V}\mathrm{ar}(X)$ is only positive semi-definite but not positive definite, we have for some $r < p$ that $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0)$, where $\lambda_j > 0$ are the ordered positive eigenvalues. Let $R = \{1, \ldots, r\}$. We then define $Z = I_{p,R}^\top B^* X$. Then note that $\mathbb{V}\mathrm{ar}(Z) = I_{p,R}^\top \Lambda I_{p,R}$, i.e, positive definite, $\| \mathbb{V}\mathrm{ar}(Z)\|_2 = \| \mathbb{V}\mathrm{ar}(X)\|_2$, and $|B\|_2 = 1$. Furthermore, we have $X = BI_{p,R}Z$. This implies with the previous result for positive definite variances

$$\begin{aligned} \| \mathbb{C}\mathrm{ov}(Y, X)\|_2 &= \| \mathbb{C}\mathrm{ov}(Y, Z) I_{p,R}^\top B^*\|_2 \leq \| \mathbb{V}\mathrm{ar}(Y)\|_2 \| \mathbb{V}\mathrm{ar}(Z)\|_2 \|I_{p,R}^\top B\|_2 \\ &= \| \mathbb{V}\mathrm{ar}(X)\|_2 \| \mathbb{V}\mathrm{ar}(Y)\|_2. \end{aligned}$$

For the generalisation to matrices, suppose that A and B are random matrices, where

$$\mathrm{E}\begin{pmatrix} B \\ A \end{pmatrix}\begin{pmatrix} B^\top & A^\top \end{pmatrix} = \mathrm{E}\begin{pmatrix} BB^\top & BA^\top \\ AB^\top & AA^\top \end{pmatrix}.$$

Let $\Sigma_{1,1} = \mathrm{E}BB^\top$, $\Sigma_{1,2} = \mathrm{E}BA^\top$, $\Sigma_{2,1} = \mathrm{E}AB^\top$, and $\Sigma_{2,2} = \mathrm{E}AA^\top$. Then, we can follow the previous arguments. □

Remark A.1 (Generalisation of Lemma A.3 to Infinite dimensional operators)**.** *Suppose that the eigenvalues of the symmetric positive semi-definite operator* $\boldsymbol{\Sigma}$ *are bounded, and*

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^* & \boldsymbol{D} \end{pmatrix}.$$

By using the same arguments as those in Lemma A.3 we have

$$\|\boldsymbol{B}\|_2^2 \leq \|\boldsymbol{A}\|_2 \|\boldsymbol{D}\|_2.$$

An application of the Lemma A.3 is in obtaining a bound for the spectral norm of the variance of infinite sums. Suppose the random matrix Y has the representation

$$Y = \sum_{j=0}^{\infty} U_j,$$

where $\{U_j\}$ are random matrices. Then

$$\|\mathrm{E}[YY^\top]\|_2 \leq \sum_{j_1,j_2=0}^{\infty} \|\mathrm{E}[U_{j_1}U_{j_2}^\top]\|_2.$$

By applying Lemma A.3 to bound $\|\mathrm{E}[U_{j_1}U_{j_2}^\top]\|_2$ we have

$$\|\mathrm{E}[YY^\top]\|_2 \leq \left(\sum_{j=0}^{\infty} \|\mathrm{E}[U_jU_j^\top]\|_2^{1/2}\right)^2. \tag{33}$$

The above bound will be used to prove the results in Example 3.3.

The following lemma is used in the proofs of Theorems 3.4, 3.6 and 3.7.

Lemma A.4. *Let* $v(j) = \max(1, |j|)$ *and* $\zeta(j) = v(\log[v(j)])/v(j)$. *For all* $y \in \mathbb{R}$ *and*

$p \geq 2$ we have

$$\sum_{j\in\mathbb{Z}} v(j)^{-p} v(j+y)^{-p} \leq (\pi^2+3) v(y-1)^{-p} \tag{34}$$

and

$$\sum_{j\in\mathbb{Z}} \zeta(j)^{p} \zeta(j+y)^{p} \leq 20 \zeta(y-1)^{p} \tag{35}$$

Further, suppose that $p, q, r \geq 2$ then

$$\sum_{j\in\mathbb{Z}} v(j)^{-q} v(j+y)^{-p} \leq (\pi^2+3) v(y-1)^{-\min(p,q)}, \tag{36}$$

$$\sum_{j\in\mathbb{Z}} \zeta(j)^{p} \zeta(j+y)^{q} \leq 20 \zeta(y-1)^{\min(p,q)}, \tag{37}$$

$$\sum_{s_1,s_2\in\mathbb{Z}} v(s_1+t)^{-p} v(s_1+s_2)^{-q} v(s_2+\tau)^{-r} \leq (\pi^2+3)^2 v(t-\tau-2)^{-\min(p,q,r)}, \tag{38}$$

and

$$\sum_{s_1,s_2\in\mathbb{Z}} \zeta(s_1+t)^{p} \zeta(s_1+s_2)^{q} \zeta(s_2+\tau)^{-r} \leq 400 \zeta(t-\tau-2)^{\min(p,q,r)} \tag{39}$$

Proof. First note that $\sum_{k=1}^{\infty} k^{-2} = \pi^2/6$. The strategy is to split the sum in several parts and for each part we pull one of the factors out of say, of $v(j)^{-p}v(j+y)^{-p}$, leverage on the pulled factor and show that the remaining sum is finite.

We first prove (34). Without loss of generality, let $y > 0$. We have

$$\sum_{j\in\mathbb{Z}} v(j)^{-p} v(j+y)^{-p} = I_1 + I_2 + I_3,$$

where

$$
\begin{aligned}
I_1 &= \sum_{j=0}^{\infty} v(j)^{-p} v(j+y)^{-p} \le (\pi^2/6+1) v(y)^{-p}, \\
I_2 &= 2v(y-1)^{-p} + \sum_{j=-y+2}^{-y/2} v(j)^{-p} v(j+y)^{-p} + \sum_{j=-y/2+1}^{-2} v(j)^{-p} v(j+y)^{-p} \\
&\le 2v(y-1)^{-p} + 2v(y/2)^{-p} 2^{-p+2} \le 2v(y-1)^{-p} + v(y)^{-p} 8(\pi^2/6-1) \\
&\le 2v(y-1)^{-p} + v(y)^{-p}(2/3\pi^2) \\
I_3 &= \sum_{j=-\infty}^{-y} v(j)^{-p} v(j+y)^{-p} \le (\pi^2/6+1) v(y)^{-p}.
\end{aligned}
$$

The bounds for I_1, I_2 and I_3 prove (34).

To proof (35), note first that

$$
\sum_{k=1}^{\infty} \zeta(k)^2 = 1 + \sum_{k=2}^{\infty} \zeta(k)^2 \le 1 + \int_1^{\infty} (\log(x)/x)^2 dx = 1 + 2.
$$

We will also use that $\zeta(\cdot)$ is monotonic decreasing after $\zeta(3)$, and $\zeta(1) = 1$, $\zeta(2) = \zeta(4) < \zeta(3)$. We start by follow the arguments as in the proof of (34) by splitting the sum into three parts we have $\sum_{j\in\mathbb{Z}} \zeta(j)^p \zeta(j+y)^p = I_1 + I_2 + I_3$ (where I_1, I_2 and I_3 are the same as those in the proof of (34) but with $\zeta(\cdot)$ replacing $v(\cdot)^{-1}$). Without loss of generality we prove the result for $y \ge 3$. For $y \ge 3$ and using the monotonicity property of $\zeta(\cdot)$ we have

$$
I_1 = \sum_{j=0}^{\infty} \zeta(j)^p \zeta(j+y)^p \le 3\zeta(y)^p
$$

and by the same argument

$$
I_3 = \sum_{j=-\infty}^{-y} \zeta(j)^p \zeta(j+y)^p \le 3\zeta(y)^p.
$$

Next we bound I_2. For this we use that

$$
\zeta(y/2)^p \sum_{j=2}^{\infty} \zeta(j)^p \le \zeta(y) 2^p \sum_{j=2}^{\infty} \zeta(j)^p \le \zeta(y) (\sum_{j=1}^{\infty} \zeta(j)^p + \sum_{j=1}^{\infty} \zeta(j+1/2)^p) \le 6\zeta(y).
$$

This gives

$$I_2 = \sum_{j=-y}^{-1} \zeta(j)^p \zeta(j+y)^p = 2\zeta(y-1)^p + \sum_{j=-y+2}^{-y/2} \zeta(j)^p \zeta(j+y)^p + \sum_{j=-y/2+1}^{-2} \zeta(j)^p \zeta(j+y)^p$$
$$\leq 2\zeta(y-1)^p + 12\zeta(y)$$
$$\leq 14\zeta(y-1)^p.$$

Thus we have bounds for the terms I_1, I_2 and I_3 in $\sum_{j\in\mathbb{Z}} \zeta(j)^p \zeta(j+y)^p$, which proves (35).

The proof of (36) uses that $v(j)^{-p} > v(j)^{-q}$, then the result immediately follows from (34).

To prove (38), let us suppose wlog that $p \leq q \leq r$, then by using (36) we have

$$\begin{aligned}
\sum_{s_1,s_2\in\mathbb{Z}} v(s_1+t)^{-r} v(s_1+s_2)^{-p} v(s_2+\tau)^{-q} &= \sum_{s_1\in\mathbb{Z}} v(s_1+t)^{-r} \sum_{s_2\in\mathbb{Z}} v(s_1+s_2)^{-p} v(s_2+\tau)^{-q} \\
&\leq (\pi^2+3) \sum_{s_1\in\mathbb{Z}} v(s_1+t)^{-r} v(s_1-\tau-1)^{-p} \\
&\leq (\pi^2+3)^2 \sum_{s_1\in\mathbb{Z}} v(t-\tau-2)^{-p}
\end{aligned}$$

where the last two lines follow from (36). This proves the result. (37) and (39) follow analogously. □

B Proof of results in Section 2

The proof of Theorem 2.1 is based on decomposing $\boldsymbol{C}^{-1}$ in terms of the inverse of a banded block matrix and its remainder, and balancing these two terms. An important result on the inverse of banded matrices is given in Demko et al. (1984), Theorem 2.4. Specifically, they consider positive definite infinite dimensional matrices of the form $\boldsymbol{A} : \ell_2 \to \ell_2$ where $\boldsymbol{A} = (A_{t,\tau}; t, \tau \in \mathbb{Z})$ ($A_{t,\tau} \in \mathbb{R}$). They show that if $\boldsymbol{A}$ has bandwidth M (in the sense $A_{t,\tau} = 0$ if $|t-\tau| > M$) and $\boldsymbol{A}^{-1} = \boldsymbol{B} = (B_{t,\tau}; t, \tau \in \mathbb{Z})$, then

$$|B_{t,\tau}| \leq \frac{(1+\sqrt{r})^2}{b} \rho^{\lfloor |t-\tau|/M \rfloor + 1}, \tag{40}$$

where $\rho = (\sqrt{r}-1)/(\sqrt{r}+1)$, $r = b/a$, $b = \sup_{v\in\ell_2,\|v\|_2=1} \langle v, \boldsymbol{A}v\rangle$, and $a = \inf_{v\in\ell_2,\|v\|_2=1} \langle v, \boldsymbol{A}v\rangle$. An interesting application of this results is given in Ding and Zhou (2021), who use it to obtain a rate of decay for the parameters in an autoregressive approximation. As our results are in the multivariate (possibly high dimensional) setting we require a bound on the block entries of a banded matrix (and not just the individual entries). Thus in the following lemma we obtain a generalisation of (40) for block matrices.

Lemma B.1. *Let $\boldsymbol{A}$ be a positive definite linear operator on $\ell_{2,p}$ where $\boldsymbol{A} = (A_{t,\tau}; t, \tau \in \mathbb{Z})$ and $A_{t,\tau}$ is a $p \times p$ dimensional matrix. We suppose that $\boldsymbol{A}$ is block-banded with bandwidth M and block-size p in the sense that for all s_1, s_2 with $|s_1 - s_2| > M$, $A_{s_1,s_2} = 0$. Let $b = \sup_{v\in\ell_{2,p},\|v\|_2=1}\langle v, \boldsymbol{A}v\rangle$, and $a = \inf_{v\in\ell_{2,p},\|v\|_2=1}\langle v, \boldsymbol{A}v\rangle$. Furthermore, $r = b/a$, $\rho = (\sqrt{r}-1)/(\sqrt{r}+1)$. Let $\boldsymbol{B} = \boldsymbol{A}^{-1} = (B_{t,\tau}; t, \tau \in \mathbb{Z})$ (where $B_{t,\tau}$ is a $p \times p$ dimensional matrix). Then, the following bound holds for all $p \times p$ sub-matrices and $t \neq \tau$*

$$\|B_{t,\tau}\|_2 \leq \frac{(1+\sqrt{r})^2}{b}\rho^{\lfloor |t-\tau|/M \rfloor + 1}$$

where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x.

Let $\widetilde{\boldsymbol{A}} = (I_{-k} \otimes I_p)^\top \boldsymbol{A}(I_{-k} \otimes I_p)$ be a sub-matrix without the kth p-dimensional row and column, where $k \in \mathbb{Z}$. Then, for $\widetilde{\boldsymbol{B}} = \widetilde{\boldsymbol{A}}^{-1}$ with $\widetilde{B}_{t,\tau} = (((I_{-k} \otimes I_p)\widetilde{\boldsymbol{A}}^{-1}(I_{-k} \otimes I_p)^\top)_{t,\tau}; t, \tau \in \mathbb{Z})$ the following bound holds for all $p \times p$ sub-matrices and $t \neq \tau$

$$\|\widetilde{B}_{t,\tau}\|_2 \leq \frac{(1+\sqrt{r})^2}{b}\rho^{\lfloor |t-\tau|/M \rfloor + 1}.$$

Proof. The proof is based on the proof of Proposition 2.2 in Demko et al. (1984), with a modification to allow for block matrices. We use the notation from Proposition 2.2 in Demko et al. (1984). More precisely, let Π_n denote the space of polynomials up to order n. A key ingredient in the proof is the following classical result from spectral theory. Suppose A is a positive definite operator, then

$$\|\boldsymbol{A}^{-1} - p(\boldsymbol{A})\|_2 \leq \max_{x\in[a,b]} |1/x - p(x)|,$$

where p is a real polynomial and recall $b = \sup_{v\in\ell_{2,p},\|v\|_2=1}\langle v, \boldsymbol{A}v\rangle$ and $a = \inf_{v\in\ell_{2,p},\|v\|_2=1}\langle v, \boldsymbol{A}v\rangle$. Set $r = b/a$, $\rho = (\sqrt{r}-1)/(\sqrt{r}+1)$. For any complex valued function f on K, define the norm $\|f\|_K = \sup\{|f(z)| : z \in K\}$ (thus $\|1/x - p(x)\|_{[a,b]} = \max_{x\in[a,b]} |1/x - p(x)|$). Proposition 2.1, Demko et al. (1984) show that

$$\inf\{\|1/x - p(x)\|_{[a,b]} : p \in \Pi_n\} = \frac{(1+\sqrt{r})^2}{b}\rho^{n+1}. \tag{41}$$

Using this result we define the polynomial

$$p_n^* = \arg_{p\in\Pi_n} \inf\{\|1/x - p(x)\|_{[a,b]} : p \in \Pi_n\}. \tag{42}$$

We note for any polynomial p_n of order n and M block-banded matrix $\boldsymbol{A}$ with block size p, if $|t-\tau| \geq nM$ then $p_n(\boldsymbol{A})_{t,\tau} \equiv 0$ where $p_n(\boldsymbol{A}))_{t,\tau}$ denotes the (t,τ) $p\times p$ dimension block matrix in $p_n(\boldsymbol{A})$.

For a given t and τ, set $n = \lfloor |t-\tau|/M \rfloor$. Let p_n^* be defined as in (42). Then by definition of n we have $p_n^*(\boldsymbol{A})_{t,\tau} = 0$. Since $B_{t,\tau} = (\boldsymbol{A}^{-1})_{t,\tau}$ this gives

$$\begin{aligned}\|B_{t,\tau}\|_2 &= \|(\boldsymbol{A}^{-1} - p_n^*(\boldsymbol{A}))_{t,\tau}\|_2 \leq \|\boldsymbol{A}^{-1} - p_n^*(\boldsymbol{A})\|_2 = \|1/x - p_n^*(x)\|_{[a,b]} \\ &= \tfrac{(1+\sqrt{r})^2}{b}\rho^{\lfloor |t-\tau|/M\rfloor + 1},\end{aligned}$$

where the last part follows from (41). This completes the proof of the first assertion.

For the second assertion, our strategy is to extend $\widetilde{\boldsymbol{A}}$ such that is an operator from $\ell_{2,p}$ to $\ell_{2,p}$ and possesses the same banded-scheme as $\boldsymbol{A}$. Then, we apply the results we derived in the first assertion to this extended $\widetilde{\boldsymbol{A}}$, hence we obtain an inverse with the desired properties. Lastly, we show that when shrinking the inverse of the extended $\widetilde{\boldsymbol{A}}$ to the space of $\widetilde{\boldsymbol{A}}$, we obtain an inverse of $\widetilde{\boldsymbol{A}}$. This idea can be formalised as follows. The extended $\widetilde{\boldsymbol{A}}$ is obtained by $(I_{-k}\otimes I_p)\widetilde{\boldsymbol{A}}(I_{-k}\otimes I_p)^\top + c(e_k \otimes I_p)(e_k\otimes I_p)^\top =: \boldsymbol{E}$. We have that $\boldsymbol{E}$ is a block-banded matrix. Additionally, if we set $c = \|(e_k\otimes I_p)^\top \boldsymbol{A}(e_k\otimes I_p)\|_2$ the largest and smallest eigenvalues of E can be bounded by those of A, that is its largest and smallest eigenvalues are bounded above from b and below from a, respectively. Hence, the previous assertion applies to $\boldsymbol{E}$. We now show $(I_{-k}\otimes I_p)^\top \boldsymbol{D}^{-1}(I_{-k}\otimes I_p) = (\widetilde{\boldsymbol{A}})^{-1}$ which gives the assertion. For this, we show $(I_{-k}\otimes I_p)^\top \boldsymbol{D}^{-1}(I_{-k}\otimes I_p)\widetilde{\boldsymbol{A}} = (I_{-k}\otimes I_p)^\top(I_{-k}\otimes I_p)$ and use the uniqueness of the inverse operator. The calculation is

$$\begin{aligned}(I_{-k}\otimes I_p)^\top \boldsymbol{D}^{-1}(I_{-k}\otimes I_p)\widetilde{\boldsymbol{A}} =& (I_{-k}\otimes I_p)^\top((I_{-k}\otimes I_p)\widetilde{\boldsymbol{A}}(I_{-k}\otimes I_p)^\top + c(e_k\otimes I_p)(e_k\otimes I_p)^\top)^{-1} \\ &\times (I_{-k}\otimes I_p)(\widetilde{\boldsymbol{A}}(I_{-k}\otimes I_p)^\top + c(e_k\otimes I_p)(e_k\otimes I_p)^\top \\ &- c(e_k\otimes I_p)(e_k\otimes I_p)^\top)(I_{-k}\otimes I_p) \\ =& (I_{-k}\otimes I_p)^\top(I_{-k}\otimes I_p) + 0.\end{aligned}$$

Thus, $(I_{-k}\otimes I_p)^\top \boldsymbol{D}^{-1}(I_{-k}\otimes I_p)$ is an inverse of $\widetilde{\boldsymbol{A}}$ and the second assertion follows. □

We now apply the above result to a specific banded matrix (required in the proof of

Theorem 2.1). Define the integer set $t^c = \{\tau \in \mathbb{Z}, \tau \neq t\}$, and $\boldsymbol{C}_{t^c,t^c} = (I_{-t} \otimes I_p)^\top \boldsymbol{C}(I_{-t} \otimes I_p)$ (this is operator $\boldsymbol{C}$ but with the tth block row and column removed). We define $\boldsymbol{B}_M$ as the Mth banded version of $\boldsymbol{C}_{t^c,t^c}$ as follows. For all $p \times p$ sub-matrices and $s_1, s_2 \in \mathbb{Z}$ let

$$((I_{-t} \otimes I_p)\boldsymbol{B}_M(I_{-t} \otimes I_p)^\top)_{s_1,s_2} = \mathbb{1}(|s_1 - s_2| \leq M)((I_{-t} \otimes I_p)\boldsymbol{C}_{t^c,t^c}(I_{-t} \otimes I_p)^\top)_{s_1,s_2}, \quad (43)$$

where $\mathbb{1}$ denotes the indicator function.

The following lemma is used in the proof of Theorems 2.1 and 3.2.

Lemma B.2. *[Properties of $\boldsymbol{B}_M$] Suppose Assumption 2.1 is satisfied and let $\boldsymbol{B}_M$ be a (symmetric) banded matrix defined as in (43). Define the space of vectors*

$$\ell_{2,p}^{-t} = \{v = (\ldots, v_{t-1}, v_{t+1}, v_{t+2}, \ldots); v_j \in \mathbb{R}^p, \sum_{j \neq t} \|v_j\|_2^2 < \infty\}$$

and the eigenvalues

$$a_M = \inf_{v \in \ell_{2,p}^{-t}, \|v\|_2 = 1} \langle v, \boldsymbol{B}_M v \rangle \text{ and } b_M = \sup_{v \in \ell_{2,p}^{-t}, \|v\|_2 = 1} \langle v, \boldsymbol{B}_M v \rangle.$$

Then

$$\|\boldsymbol{C}_{t^c,t^c} - \boldsymbol{B}_M\|_2 \leq 2\frac{K}{(\kappa - 1)}(M-1)^{-\kappa+1}, \quad (44)$$

$$a_M \geq \lambda_{\inf} - 2\frac{K}{(\kappa-1)}(M-1)^{-\kappa+1}, b_M \leq \lambda_{\sup} + 2\frac{K}{(\kappa-1)}(M-1)^{-\kappa+1} \quad (45)$$

and if M is such that $\lambda_{\inf} - 2\frac{K}{(\kappa-1)}(M-1)^{-\kappa+1} > 0$, then

$$\|\boldsymbol{B}_M^{-1}\|_2 \leq \left(\lambda_{\inf} - 2\frac{K}{(\kappa-1)}(M-1)^{-\kappa+1}\right)^{-1}. \quad (46)$$

The same rates to the banded matrices associated with $\boldsymbol{C}$ or $\boldsymbol{C}(-\infty; T]$.

Proof. We first prove (44). For this, we first expand $\boldsymbol{C}_{t^c,t^c} - \boldsymbol{B}_M$ with zero such that it is an operator from $\ell_{2,p}$ to $\ell_{2,p}$ again. Then, we apply Lemma A.2 to obtain

$$\|\boldsymbol{C}_{t^c,t^c} - \boldsymbol{B}_M\|_2 = \leq \sup_{s_1} \sum_{s_2} \|((I_{-k} \otimes I_p)(\boldsymbol{C}_{t^c,t^c} - \boldsymbol{B}_M)(I_{-k} \otimes I_p))_{s_1,s_2}\|_2$$
$$\leq \sum_{|s|>M} \frac{K}{|s|^{-\kappa}} \leq 2K \sum_{s>M} \int_{s-1}^{s} x^{-\kappa} dx = 2\frac{K}{(\kappa-1)}(M-1)^{-\kappa+1}.$$

where the first bound on last line above follows from Assumption 2.1.

To prove (45) we use that $\boldsymbol{B}_M = \boldsymbol{C}_{t^c,t^c} + (\boldsymbol{C}_{t^c,t^c} - \boldsymbol{B}_M)$ and the eigenvalues of $\boldsymbol{C}_{t^c,t^c}$ are in $[\lambda_{\inf}, \lambda_{\sup}]$. Thus, with (44) we have

$$\lambda_{\inf}(\boldsymbol{B}_M) \geq \lambda_{\inf} - 2\frac{K}{(\kappa-1)}(M-1)^{-\kappa+1} \text{ and } \lambda_{\sup}(\boldsymbol{B}_M) \leq \lambda_{\sup} + 2\frac{K}{(\kappa-1)}(M-1)^{-\kappa+1}. \quad (47)$$

The proof of (46) immediately follows from (45). □

Using the above lemma we now prove Theorem 2.1.

Proof of Theorem 2.1. For (1) we focus here on the case $t \neq \tau$ and $|t-\tau| \geq 2$.

To motivate the proof, we first describe a more direct but naive approach which does not give sufficiently sharp bounds. One strategy is to directly approximate $\boldsymbol{D}$ with the inverse of a (block) banded matrix, say $\boldsymbol{A}_M$, and then use the Neuman series to bound its error. I.e. use an expansion of the form

$$\boldsymbol{D} = (\boldsymbol{A}_M + (\boldsymbol{D} - \boldsymbol{A}_M))^{-1} = \boldsymbol{A}_M^{-1} + \sum_{j=1}^{\infty} [\boldsymbol{A}_M^{-1}(\boldsymbol{D} - \boldsymbol{A}_M)]^j,$$

which holds when M is large enough such that $\|\boldsymbol{A}_M^{-1}(\boldsymbol{D} - \boldsymbol{A}_M)\|_2 < 1$. The ℓ_2 bound of the above is

$$\begin{aligned}
\|\boldsymbol{D}_{t,\tau}\|_2 &\leq \|(\boldsymbol{A}_M^{-1})_{t,\tau}\|_2 + \sum_{j=1}^{\infty} \|\boldsymbol{A}_M^{-1}(\boldsymbol{D} - \boldsymbol{A}_M)\|_2^j \\
&\leq \|\boldsymbol{A}_M^{-1}\|_2 + \|\boldsymbol{A}_M^{-1}(\boldsymbol{D} - \boldsymbol{A}_M)\|_2 (1 - \|\boldsymbol{A}_M^{-1}(\boldsymbol{D} - \boldsymbol{A}_M)\|_2)^{-1} \\
&= I_1 + I_2.
\end{aligned}$$

By using Lemma B.1 we can show that

$$I_1 \leq \frac{(1+\sqrt{r_M})^2}{\lambda_{\sup,M}} \rho_M^{\lfloor |s-\tau|/M \rfloor + 1},$$

where $r_M = \lambda_{\sup,M}/\lambda_{\inf,M}$ $\rho_M = (\sqrt{r_M} - 1)/(\sqrt{r_M} + 1)$ and $\lambda_{\inf,M}$ and $\lambda_{\sup,M}$ are the eigenvalues of $\boldsymbol{A}_M$. It can also be shown that

$$I_2 \leq \frac{C}{(\kappa - 1)}(M-1)^{-\kappa+1}.$$

This leads to the rate

$$\|\boldsymbol{D}_{t,\tau}\|_2 \leq \frac{(1+\sqrt{r_M})^2}{\lambda_{\sup,M}}\rho_M^{\lfloor |s-\tau|/M \rfloor + 1} + \frac{C}{(\kappa-1)}(M-1)^{-\kappa+1}.$$

However, this bound does not adequately utilize the rate of decay of the entries of $\boldsymbol{C}$. Instead we take an indirect approach, where we rewrite $\boldsymbol{D}$ as the inverse of a block matrix, where the relevant entries are the inverse of one submatrix of $\boldsymbol{C}$ *multipled* with another submatrix of $\boldsymbol{C}$. The latter term allows us to leverage on the rate of decay of the entries of $\boldsymbol{C}$. We describe this approach below.

Define the integer set $t^c = \{\tau \in \mathbb{Z}, \tau \neq t\}$, and denote $\boldsymbol{C}_{t,t^c} = (e_t \otimes I_p)^\top \boldsymbol{C}(I_{-t} \otimes I_p)$ and $\boldsymbol{C}_{t^c,t^c} = (I_{-t} \otimes I_p)^\top \boldsymbol{C}(I_{-t} \otimes I_p)$. Without loss of generality we consider a permuted version of $\boldsymbol{C}$, which contains $C_{t,t}$ in the top left hand corner of $\boldsymbol{C}$, where

$$\boldsymbol{C} = \begin{pmatrix} C_{t,t} & \boldsymbol{C}_{t,t^c} \\ \boldsymbol{C}_{t^c,t} & \boldsymbol{C}_{t^c,t^c} \end{pmatrix}.$$

Setting $\boldsymbol{U} = \boldsymbol{C}$, $\boldsymbol{A} = C_{t,t}$, $\boldsymbol{B} = \boldsymbol{C}_{t,t^c}$, $\boldsymbol{C} = \boldsymbol{C}_{t,t^c}^\top$, $\boldsymbol{D} = \boldsymbol{C}_{t^c,t^c}$ and applying the block matrix operator inversion formula in (30) we have

$$\boldsymbol{D} = \boldsymbol{C}^{-1} = \begin{pmatrix} D_{t,t} & -D_{t,t}^{-1}\boldsymbol{C}_{t,t^c}\boldsymbol{C}_{t^c,t^c}^{-1} \\ -\boldsymbol{C}_{t^c,t^c}^{-1}\boldsymbol{C}_{t^c,t}D_{t,t}^{-1} & (\boldsymbol{C}_{t^c,t^c} - \boldsymbol{C}_{t^c,t}C_{t,t}^{-1}\boldsymbol{C}_{t,t^c})^{-1} \end{pmatrix}.$$

Using the above $D_{t,\tau}$ can be written as

$$D_{t,\tau} = -D_{t,t}^{-1}\boldsymbol{C}_{t,t^c}\boldsymbol{C}_{t^c,t^c}^{-1}(I_{-t} \otimes I_p)^\top (e_\tau \otimes I_p),$$

using that $\lambda_{\sup}^{-1} \leq D_{t,t} \leq \lambda_{\inf}^{-1}$ we have $\|D_{t,\tau}\|_2 \leq \lambda_{\sup}\|(\boldsymbol{C}_{t,t^c}\boldsymbol{C}_{t^c,t^c}^{-1})(I_{-t}^\top e_\tau \otimes I_p)\|_2$. Thus for the remainder of the proof, we focus on bounding the induced ℓ_2-norm of

$$A_{t,\tau} = (\boldsymbol{C}_{t,t^c}\boldsymbol{C}_{t^c,t^c}^{-1})(I_{-t}^\top e_\tau \otimes I_p).$$

An outline in the proof is to (a) replace $\boldsymbol{C}_{t^c,t^c}^{-1}$ with the inverse of a (block) banded matrix (b) use the Neuman series to obtain a bound on the replacement error and (c) finally balance the rate of decay of the inverse banded matrix approximation of $(\boldsymbol{C}_{t,t^c}\boldsymbol{C}_{t^c,t^c}^{-1})(I_{-t}^{\top}e_{\tau}\otimes I_p)$ with the spectral norm of the approximation error (both of which depend on the bandwidth M).

Let $\boldsymbol{B}_M$ denote the Mth banded matrix version of $\boldsymbol{C}_{t^c,t^c}$; the precise definition is given in (43). By Lemma B.2, equation (44) we have the bound

$$\|\boldsymbol{C}_{t^c,t^c}-\boldsymbol{B}_M\|_2 \leq 2\frac{K}{(\kappa-1)}(M-1)^{-\kappa+1}. \tag{48}$$

Using $\boldsymbol{B}_M$ we write $\boldsymbol{C}_{t^c,t^c}^{-1}$ as a Neumann series

$$\boldsymbol{C}_{t^c,t^c}^{-1} = \boldsymbol{B}_M^{-1}[I+\boldsymbol{B}_M^{-1}(\boldsymbol{C}_{t^c,t^c}-\boldsymbol{B}_M)]^{-1} = \boldsymbol{B}_M^{-1}[I+\sum_{s=1}^{\infty}(-1)^s[\boldsymbol{B}_M^{-1}(\boldsymbol{C}_{t^c,t^c}-\boldsymbol{B}_M)]^s],$$

noting that the above expanion holds, when $M > 1+(2K/(\kappa-1))^{1/(\kappa-1)}$ thus $\|\boldsymbol{C}_{t^c,t^c}-\boldsymbol{B}_M\|_2 < 1$. Substituting the above into $A_{t,\tau} = (\boldsymbol{C}_{t,t^c}\boldsymbol{C}_{t^c,t^c}^{-1})(I_{-t}^{\top}e_{\tau}\otimes I_p)$ gives for all $t,\tau\in\mathbb{Z}$

$$\begin{aligned} A_{t,\tau} &= \boldsymbol{C}_{t,t^c}\boldsymbol{B}_M^{-1}[I+\sum_{s=1}^{\infty}(-1)^s[\boldsymbol{B}_M^{-1}(\boldsymbol{C}_{t^c,t^c}-\boldsymbol{B}_M)]^s](I_{-t}^{\top}e_{\tau}\otimes I_p) \\ &= \boldsymbol{C}_{t,t^c}\boldsymbol{B}_M^{-1}(I_{-t}^{\top}e_{\tau}\otimes I_p)+\boldsymbol{C}_{t,t^c}\boldsymbol{B}_M^{-1}\sum_{s=1}^{\infty}(-1)^s[\boldsymbol{B}_M^{-1}(\boldsymbol{C}_{t^c,t^c}-\boldsymbol{B}_M)]^s](I_{-t}^{\top}e_{\tau}\otimes I_p). \end{aligned}$$

By applying the triangular inequality to the above we have $\|A_{t,\tau}\|_2 \leq J_{1,t,\tau}+J_{2,t,\tau}$ where

$$\begin{aligned} J_{1,t,\tau} &= \|\boldsymbol{C}_{t,t^c}\boldsymbol{B}_M^{-1}(I_{-t}^{\top}e_{\tau}\otimes I_p)\|_2 \\ \text{and } J_{2,t,\tau} &= \|\boldsymbol{C}_{t,t^c}\boldsymbol{B}_M^{-1}\sum_{s=1}^{\infty}(-1)^s[\boldsymbol{B}_M^{-1}(\boldsymbol{C}_{t^c,t^c}-\boldsymbol{B}_M)]^s](I_{-t}^{\top}e_{\tau}\otimes I_p)\|_2. \end{aligned}$$

We now bound $J_{1,t,\tau}$ and $J_{2,t,\tau}$. By using the sub-multiplicativity of $\|\cdot\|_2$ we bound $J_{2,t,\tau}$ with

$$J_{2,t,\tau} \leq \|\boldsymbol{C}_{t,t^c}\boldsymbol{B}_M^{-1}\|_2\sum_{s=1}^{\infty}(\|[\boldsymbol{B}_M^{-1}\|_2\|(\boldsymbol{C}_{t^c,t^c}-\boldsymbol{B}_M)]\|_2)^s. \tag{49}$$

By using Lemma B.2, if M is such that $\lambda_{\inf}-2\frac{K}{(\kappa-1)}(M-1)^{-\kappa+1}>0$, then

$$\|[\boldsymbol{B}_M^{-1}\|_2\|(\boldsymbol{C}_{t^c,t^c}-\boldsymbol{B}_M)]\|_2 \leq \left(\lambda_{\inf}-2\frac{K}{(\kappa-1)}(M-1)^{-\kappa+1}\right)^{-1}2\frac{K}{(\kappa-1)}(M-1)^{-\kappa+1}.$$

Thus for

$$M > 1 + (2K/[\min(1, \lambda_{\inf}/2)(\kappa - 1)])^{1/(\kappa-1)} := M_2,$$

we have $\|[\boldsymbol{B}_M^{-1}\|_2\|(\boldsymbol{C}_{t^c,t^c} - \boldsymbol{B}_M)]\|_2 < 1$. Hence, we obtain the geometric sum

$$\begin{aligned} J_{2,t,\tau} &\leq \|\boldsymbol{C}_{t,t^c}\boldsymbol{B}_M^{-1}\|_2 \sum_{s=1}^{\infty} \|[\boldsymbol{B}_M^{-1}\|_2\|(\boldsymbol{C}_{t^c,t^c} - \boldsymbol{B}_M)]\|_2^s \\ &\leq \frac{2K/(\kappa-1)(M-1)^{-\kappa+1}}{1 - \|[\boldsymbol{B}_M^{-1}\|_2\|(\boldsymbol{C}_{t^c,t^c} - \boldsymbol{B}_M)]\|_2} = 2\mathcal{K}/(\kappa-1)(M-1)^{-\kappa+1} =: \widetilde{J}_{2,t,\tau}, \end{aligned}$$

where the last line of the above follows from Lemma B.2. In summary, for $M > M_2$ we have

$$J_{2,t,\tau} \;\; \leq \;\; 2\mathcal{K}/(\kappa-1)(M-1)^{-\kappa+1} = \widetilde{J}_{2,t,\tau}. \tag{50}$$

Next we bound $J_{1,t,\tau}$. We start by expanding $\boldsymbol{C}_{t,t^c}\boldsymbol{B}_M^{-1}$, then use the sub-multiplicativity of $\|\cdot\|_2$ to give

$$J_{1,t,\tau} \;\; \leq \sum_{s\in\mathbb{Z}, s\neq t} \|C_{t,s}\|_2 \cdot \|((I_{-t}\otimes I_p)\boldsymbol{B}_M^{-1}(I_{-t}\otimes I_p)^\top)_{s,\tau}\|_2. \tag{51}$$

We bound the terms inside of the sum $\sum_{s\in\mathbb{Z},s\neq t} \|C_{t,s}\|_2 \cdot \|((I_{-t}\otimes I_p)\boldsymbol{B}_M^{-1}(I_{-t}\otimes I_p)^\top)_{s,\tau}\|_2$. Under Assumption 2.1 we have $\|C_{t,s}\|_2 \leq Kv(t-s)^{-\kappa}$. To bound the second term, we use Lemma B.1

$$\|((I_{-t}\otimes I_p)\boldsymbol{B}_M^{-1}(I_{-t}\otimes I_p)^\top)_{s,\tau}\|_2 \leq \frac{(1+\sqrt{r_M})^2}{\lambda_{\sup,M}}\rho_M^{\lfloor |s-\tau|/M\rfloor+1}, \tag{52}$$

where $r_M = \lambda_{\sup,M}/\lambda_{\inf,M}$ $\rho_M = (\sqrt{r_M}-1)/(\sqrt{r_M}+1)$ and $\lambda_{\sup,M}$ and $\lambda_{\inf,M}$ are such that

$$\lambda_{\sup,M} \;\; \leq \;\; \lambda_{\sup} + 2\frac{K}{\kappa-1}(M-1)^{-\kappa+1} \text{ and } \lambda_{\inf,M} \geq \lambda_{\inf} - 2\frac{K}{\kappa-1}(M-1)^{-\kappa+1}.$$

This gives a bound for $r_M, \lambda_{\sup,M}$ and ρ_M in terms of $r, \lambda_{\sup}, \rho$ and M. To remove the dependency of M in these we choose M such that

$$M > \left(\frac{2K}{\kappa-1}\max(2\lambda_{\inf}^{-1}, \lambda_{\sup}^{-1})\right)^{1/(\kappa-1)} + 1 := M_1$$

For $M > M_1$ we have $\lambda_{\inf,M} \geq \lambda_{\inf}/2$ and $\lambda_{\sup,M} \leq 2\lambda_{\sup}$. This means, $r_M \leq 4r$ and

$\rho_M \leq (2\sqrt{r}-1)/(2\sqrt{r}+1) =: \rho$, where $r = \lambda_{\sup}/\lambda_{\inf}$. Substituting this into (52) gives

$$\|((I_{-t}\otimes I_p)\boldsymbol{B}_M^{-1}(I_{-t}\otimes I_p)^\top)_{s,\tau}\|_2 \leq \frac{2(1+2\sqrt{r})^2}{\lambda_{\sup}}\rho^{\lfloor |s-\tau|/M\rfloor+1}. \tag{53}$$

Substituting (53) and $\|C_{t,s}\|_2 \leq Kv(t-s)^{-\kappa}$ into (51) we have

$$\begin{aligned} J_{1,t,\tau} &\leq \tfrac{2K(1+2\sqrt{r})^2}{\lambda_{\sup}}\textstyle\sum_{s\in Z, s\neq t}|s-t|^{-\kappa}\rho^{\lfloor |s-\tau|/M\rfloor+1} \\ &\leq \tfrac{2K(1+2\sqrt{r})^2}{\lambda_{\sup}}\textstyle\sum_{s\in Z}\rho^{|s|/M}\tfrac{1}{v(s-t+\tau)^\kappa} = \widetilde{J}_{1,t,\tau}. \end{aligned} \tag{54}$$

Thus when $M > K_c$ where

$$K_c := \max(M_1, M_2) = \left(\frac{2K}{\kappa-1}\max(2\lambda_{\inf}^{-1}, \lambda_{\sup}^{-1}, 1)\right)^{1/(\kappa-1)} + 1$$

the bounds $J_{1,t,\tau}$ and $J_{2,t,\tau}$ in (54) and (50) hold and we have

$$\|D_{t,\tau}\|_2 \leq \lambda_{\sup}(\widetilde{J}_{1,t,\tau} + \widetilde{J}_{2,t,\tau}),$$

where $\widetilde{J}_{1,t,\tau}$ and $\widetilde{J}_{2,t,\tau}$ are defined in (54) and (50) respectively.

The final part in the proof is to balance the two bounds $\widetilde{J}_{1,t,\tau}$ and $\widetilde{J}_{2,t,\tau}$. For each $t,\tau \in \mathbb{Z}$ we set $M = M_{t-\tau} := -\frac{|t-\tau|\log(\rho)}{2(\kappa-1)\log(|t-\tau|)}$ (note $0 < \rho < 1$). When $|t-\tau|$ is sufficiently large i.e., $M_{t-\tau} \geq K_c$ by substituting $M_{t-\tau}$ into the bounds for $J_{1,t,\tau}$ and $J_{2,t,\tau}$ it can be shown that

$$\|D_{t,\tau}\|_2 \;\leq\; 2K(1+2\sqrt{r})^2(2^\kappa + 2S_\kappa)|t-\tau|^{-\kappa+1} + \frac{2K}{\kappa-1}\left(\frac{|\log(\rho)|}{2(\kappa-1)}\frac{|t-\tau|}{\log|t-\tau|} - 1\right)^{-\kappa+1} \tag{55}$$

Note that the above expression, though unwieldy gives the desired decay $\zeta(t-\tau)^{\kappa-1}$. However, if $|t-\tau|$ is small, i.e., $M_{t-\tau} \leq K_c$ the bound (54) does not hold and we use an alternative bound for $\|D_{t,\tau}\|_2$. It is easily seen that $\|D_{t,\tau}\|_2 \leq \|\boldsymbol{D}\|_2 \leq \lambda_{\inf}^{-1}$. We rewrite the above in a similar form as (54) (but with different contants)

$$\|D_{t,\tau}\|_2 \leq \lambda_{\inf}^{-1} \leq (\lambda_{\inf}\min(\lambda_{\inf}/2, \lambda_{\sup}))^{-1}\frac{2K}{\kappa-1}\left(\frac{|\log(\rho)|}{2(\kappa-1)}\frac{|t-\tau|}{\log|t-\tau|} - 1\right)^{-\kappa+1}. \tag{56}$$

Combining (54) and (55) gives the following global bound for all $t, \tau \in \mathbb{Z}$

$$\begin{aligned}\|D_{t,\tau}\|_2 &\leq 2K(1+2\sqrt{r})^2(2^\kappa + 2S_\kappa)v(t-\tau)^{-\kappa+1} \\ &\quad + \max(1, (\min(\lambda_{\inf}/2, \lambda_{\sup})\lambda_{\inf})^{-1})\frac{2K}{(\kappa-1)}\left(\frac{|\log(\rho)|}{2(\kappa-1)}\frac{v(t-\tau)}{v(\log v(t-\tau))} - 1\right)^{-\kappa+1} \\ &\leq \mathcal{K}\zeta(t-\tau)^{\kappa-1}.\end{aligned}$$

Note that in the proof we have carefully tracked all the constants, to demonstrate that the constants only depend on $\lambda_{\inf}, \lambda_{\sup}, K$ and κ. To reduce notation, in the remainder of the paper we use a generic constant $\mathcal{K}$.

To prove (2), we only need a slight modification of the above arguments. We define the integer set $t_T^c = \{\tau \leq T, \tau \neq t\}$ and obtain

$$\boldsymbol{C}(-\infty; T) = \begin{pmatrix} C_{t,t} & \boldsymbol{C}_{t,t_T^c} \\ \boldsymbol{C}_{t_T^c,t} & \boldsymbol{C}_{t_T^c,t_T^c} \end{pmatrix}.$$

This leads to

$$[\boldsymbol{C}(-\infty; T)^{-1}]_{t,\tau} = -[\boldsymbol{C}(-\infty; T)^{-1}]_{t,t}^{-1}\boldsymbol{C}_{t,t_T^c}\boldsymbol{C}_{t_T^c,t_T^c}^{-1}(I_{-t} \otimes I_p)^\top(e_\tau \otimes I_p).$$

Then, we follow the same strategy as above. Note that the sums occurring now are going from from $-\infty$ to T instead before in deriving (1) in which they are from $-\infty$ to ∞. $\square$

Proof of Theorem 2.2. We first recall that the coefficients $\{\Phi_{T,j}\}$ and $\Phi_{T,d,j}$ are embedded in the last rows of $\boldsymbol{C}(-\infty, T)^{-1}$ and $\boldsymbol{C}(T-d, T)^{-1}$ respectively. Therefore we first need to connect the inverses of $\boldsymbol{C}(-\infty, T)$ and $\boldsymbol{C}(T-d, T)$. For this, we write $\boldsymbol{C}(-\infty, T)$ in terms of the following block matrix

$$\boldsymbol{C}(-\infty, T) = \begin{pmatrix} \boldsymbol{C}(-\infty, T-d) & \boldsymbol{C}(-\infty, T-d, T) \\ \boldsymbol{C}(-\infty, T-d, T)^\top & \boldsymbol{C}(T-d, T) \end{pmatrix},$$

where $\boldsymbol{C}(-\infty, T-d, T) = (C_{t,\tau}; t \leq T-d, T-d+1 \leq \tau \leq T)$ and $\boldsymbol{C}(T-d, T) = (C_{t,\tau}; T-d+1 \leq t, \tau \leq T)$. Next we represent $\boldsymbol{C}(-\infty, T)^{-1}$ as a block operator (analogous to $\boldsymbol{C}(-\infty, T)$)

$$\boldsymbol{C}(-\infty, T)^{-1} = \begin{pmatrix} \widetilde{\boldsymbol{D}}(-\infty, T-d) & \widetilde{\boldsymbol{D}}(-\infty, T-d, T) \\ \widetilde{\boldsymbol{D}}(-\infty, T-d, T)^\top & \widetilde{\boldsymbol{D}}(T-d, T) \end{pmatrix}.$$

Note we have used the notation $\widetilde{\boldsymbol{D}}$ to show that they are not the inverse of the corresponding submatrix of $\boldsymbol{C}$. To evaluate $\boldsymbol{C}(T-d,T)^{-1}-\widetilde{\boldsymbol{D}}(T-d,T)$ we apply the second identity in (32) where we set $\boldsymbol{U}=\boldsymbol{C}(-\infty,T)^{-1}$, $\boldsymbol{A}=\widetilde{\boldsymbol{D}}(-\infty,T-d)$, $\boldsymbol{B}=\widetilde{\boldsymbol{D}}(-\infty,T-d,T)$, $\boldsymbol{C}=\widetilde{\boldsymbol{D}}(-\infty,T-d,T)^{\top}$, $\boldsymbol{D}=\widetilde{\boldsymbol{D}}(T-d,T)$ and $\widetilde{\boldsymbol{D}}=\boldsymbol{C}(T-d,T)$. This gives

$$\boldsymbol{C}(T-d,T)^{-1}-\widetilde{\boldsymbol{D}}(T-d,T)=-\widetilde{\boldsymbol{D}}(-\infty,T-d,T)\widetilde{\boldsymbol{D}}(-\infty,T-d)^{-1}\widetilde{\boldsymbol{D}}(-\infty,T-d,T)^{\top}.$$

Thus block-wise for all $1\leq t,\tau\leq d$ we have

$$\begin{aligned}&[\boldsymbol{C}(T-d,T)^{-1}-\widetilde{\boldsymbol{D}}(T-d,T)]_{T-t,T-\tau}\\ =\ &-[(e_{T-t}\otimes I_p)^{\top}\widetilde{\boldsymbol{D}}(-\infty,T-d,T)]\widetilde{\boldsymbol{D}}(-\infty,T)^{-1}[(e_{T-\tau}\otimes I_p)^{\top}\widetilde{\boldsymbol{D}}(-\infty,T-d,T)]^{\top}.\end{aligned}$$

Using the above we obtain the bound

$$\begin{aligned}&\|[\boldsymbol{C}(T-d,T)^{-1}-\widetilde{\boldsymbol{D}}(T-d,T)]_{T-t,T-\tau}\|_2\\ \leq\ &\lambda_{\sup}\|(e_{T-t}\otimes I_p)^{\top}\widetilde{\boldsymbol{D}}(-\infty,T-d,T)\|_2\|(e_{T-\tau}\otimes I_p)^{\top}\widetilde{\boldsymbol{D}}(-\infty,T-d,T)\|_2. \quad (57)\end{aligned}$$

Next we obtain a bound for the matrix rows $(e_{T-t}\otimes I_p)^{\top}\widetilde{\boldsymbol{D}}(-\infty,T-d,T)=(\widetilde{\boldsymbol{D}}(-\infty,T-d,T)_{T-t,\ell};\ell<T)$. By applying Lemma A.1 and using Theorem 2.1 we have

$$\begin{aligned}\|(e_{T-t}\otimes I_p)^{\top}\widetilde{\boldsymbol{D}}(-\infty,T-d,T)\|_2 &\leq (\sum_{\ell=-\infty}^{T-d-1}\|\widetilde{\boldsymbol{D}}(-\infty,T-d,T)_{T-t,\ell}\|_2^2)^{1/2}\\ &\leq \mathcal{K}(\sum_{\ell=-\infty}^{T-d-1}\zeta(T-t-\ell)^{2(\kappa-1)})^{1/2}\leq \mathcal{K}\zeta(d-t)^{\kappa-3/2}.\end{aligned}$$

Substituting the above into (56) for all $1\leq t,\tau\leq d$ we have

$$\|[\boldsymbol{C}(T-d,T)^{-1}-\widetilde{\boldsymbol{D}}(T-d,T)]_{T-t,T-\tau}\|_2 \ \leq\ \mathcal{K}\zeta(d-t)^{\kappa-3/2}\zeta(d-\tau)^{\kappa-3/2}. \quad (58)$$

We now return to the VAR coefficients. Using the block inverse operator identity in (30) it can be shown that $1\leq j\leq d$

$$\Phi_{T,d,j}-\Phi_{T,j}=-[\boldsymbol{C}(T-d,T)^{-1}]_{T,T}^{-1}[\boldsymbol{C}(T-d,T)^{-1}]_{T,T-j}+[\widetilde{\boldsymbol{D}}(T-d,T)]_{T,T}^{-1}[\widetilde{\boldsymbol{D}}(T-d,T)]_{T,T-j},$$

(the bottom rows of $\boldsymbol{C}(T-d,T)^{-1}$ and $\widetilde{\boldsymbol{D}}(T-d,T)$ respectively). Using the above and

(57) we will prove (7). Setting $t=0$ and $\tau=j$ in (57) gives

$$\begin{aligned}\|\Phi_{T,d,j}-\Phi_{T,j}\|_2 &\leq \lambda_{\sup}\|[\boldsymbol{C}(T-d,T)^{-1}-\widetilde{\boldsymbol{D}}(T-d,T)]_{T,T-j}\|_2\\ &\quad+\lambda_{\sup}\|[[\boldsymbol{C}(T-d,T)^{-1}]_{T,T}^{-1}-\widetilde{\boldsymbol{D}}(T-d,T)]_{T,T}^{-1}\|_2\\ &\leq \mathcal{K}\zeta(d)^{\kappa-3/2}\zeta(d-j)^{\kappa-3/2}.\end{aligned}$$

This proves (7). Using (7) we immediately obtain (8). □

Note that projection methods can also be used to prove the above result (and the same bound obtained). In this case the proof would be similar to that given in the proof of Theorem 3.2 in Meyer et al. (2017) (in the context of spatially stationary processes).

C Proofs of results in Section 3

C.1 Proofs of results in Section 3.2

The following lemma is used in the proof of Theorem 3.1.

Lemma C.1. *Suppose Assumption 3.1 holds and let $G_{u,M}(\omega)$, $G_{u,M}^{(N)}(\omega)$ and $G_u(\omega)$ be defined as in (61), (62) and (60) respectively. Then*

$$\sup_{\omega}\|G_{u,M}(\omega)-G_{u,M}^{(N)}(\omega)\|_2\leq\mathcal{K}\frac{M}{N} \tag{59}$$

and

$$\sup_{\omega}\|G_u(\omega)-G_{u,M}(\omega)\|_2\leq\mathcal{K}\left(\frac{1}{M}+\frac{1}{M^{\kappa-1}}\right) \tag{60}$$

where $\mathcal{K}$ is a constant that only depends on K and κ.

Proof. Under Assumption 3.1(iii) we have

$$\begin{aligned}\|G_{u,M}^{(N)}(\omega)-G_{u,M}(\omega)\|_2 &\leq \frac{1}{M}\sum_{t,\tau=T_{u,N}-M/2+1}^{T_{u,N}+M/2}\|C_{t,\tau}^{(N)}-C_{t-\tau}(u)\|_2\\ &\leq \frac{1}{M}\sum_{t,\tau=T_{u,N}-M/2+1}^{T_{u,N}+M/2}\left(\frac{1}{Nv(t-\tau)^{\kappa-1}}+\frac{|(T_{u,N}-t)|}{Nv(t-\tau)^{\kappa}}\right)\leq\mathcal{K}\frac{M}{N},\end{aligned}$$

this proves (58). To prove (59) we use that

$$G_u(\omega) \;=\; G_{u,M}(\omega) + \frac{1}{M}\sum_{|r|\leq M/2}|r|C_r(u)\exp(ir\omega) + \sum_{|r|>M/2}C_r(u)\exp(ir\omega).$$

Under Assumption 3.1(iii) we have $\|C_r(u)\|_2 \leq K/v(r)^{\kappa}$ (where $\kappa > 2$), thus

$$\|G_u(\omega) - G_{u,M}(\omega)\|_2 \leq \frac{1}{M}\sum_{|r|\leq M/2}|r|\|C_r(u)\|_2 + \sum_{|r|>M/2}\|C_r(u)\|_2 \leq \mathcal{K}\left(\frac{1}{M}+\frac{1}{M^{\kappa-1}}\right).$$

Thus proving the result. □

We are now equiped to prove Theorem 3.1.

Proof of Theorem 3.1. Our aim is to show that the $\|\cdot\|_2$-norm of the matrix function

$$G_u(\omega) = \sum_{r\in\mathbb{Z}} C_r(u)\exp(ir\omega) \tag{61}$$

is bounded above and below by the λ_{sup} and λ_{inf} respectively (for all ω). Since $\boldsymbol{C}(u)$ is a (block) Toeplitz matrix then by Toeplitz theorem (see Toeplitz (1911) and Böttcher and Grudsky (2000), Theorem 1.1) this would immediately prove that the eigenvalues of $\boldsymbol{C}(u)$ are bounded above and below by λ_{sup} and λ_{inf} (thus proving the result).

For a given $u \in \mathbb{R}$ and $N \in \mathbb{N}$ we define the integer $T_{u,N}$ as $T_{u,N} = \lfloor uN \rfloor$ (where $\lfloor x \rfloor$ denotes the largest integer smaller than x). Let $M \in 2\mathbb{N}$ and define an $M\times M$-dimensional submatrix of $\boldsymbol{C}^{(M)}$ that is centred about $T_{u,N}$

$$\boldsymbol{C}_{u,M}^{(N)} \;:=\; (C^{(N)}_{T_{u,N}+s_1,T_{u,N}+s_2})_{s_1,s_2=-M/2+1,\ldots,M/2} =: (I^{(M)}_{T_{u,N}}\otimes I_p)^{\top}\boldsymbol{C}^{(N)}(I^{(M)}_{T_{u,N}}\otimes I_p).$$

We show below that if M is sufficiently small, then $\boldsymbol{C}_{u,M}^{(N)}$ is an approximation of the $M \times M$-dimensional submatrix of $\boldsymbol{C}(u)$

$$\boldsymbol{C}_M(u) \;:=\; (C_{s_1-s_2}(u))_{s_1,s_2=-M/2+1,\ldots,M/2} =: (I^{(M)}_{(u)}\otimes I_p)^{\top}\boldsymbol{C}(u)(I^{(M)}_{(u)}\otimes I_p).$$

We start by obtaining a finite approximation of $G_u(\omega)$ in terms of $\boldsymbol{C}_M(u)$. Let

$$G_{u,M}(\omega) = \frac{1}{M}\sum_{t,\tau=T_{u,N}-M/2+1}^{T_{u,N}+M/2} C_{t-\tau}(u)\exp(i(t-\tau)\omega) = (x_\omega\otimes I_p)^{*}\boldsymbol{C}_M(u)(x_\omega\otimes I_p), \tag{62}$$

where $x_\omega = 1/\sqrt{M}(\exp(-it\omega))_{t=T_{u,N}-M/2+1,\ldots,T_{u,N}+M/2}$. Using $\boldsymbol{C}_{u,M}^{(N)}$ for each $M \in 2\mathbb{N}$ and $\omega \in [0, 2\pi]$ we define the quantity

$$G_{u,M}^{(N)}(\omega) = \frac{1}{M} \sum_{t,\tau=T_{u,N}-M/2+1}^{T_{u,N}+M/2} C_{t,\tau}^{(N)} \exp(i(t-\tau)\omega) = (x_\omega \otimes I_p)^* \boldsymbol{C}_{u,M}^{(N)} (x_\omega \otimes I_p). \tag{63}$$

Since $\boldsymbol{C}_{u,N}^{(M)}$ is a finite dimensional submatrix of $\boldsymbol{C}^{(N)}$, for $N > N_0$, the eigenvalues of $\boldsymbol{C}_{u,M}^{(N)}$ are bounded above and below by $\lambda_{\inf}$ and $\lambda_{\sup}$ respectively. Then, since $\|x_\omega\|_2 = 1$ we have

$$\lambda_{\inf} \leq \|G_{u,M}^{(N)}(\omega)\|_2 \leq \lambda_{\sup} \text{ for all } N,\, M \text{ and } \omega. \tag{64}$$

By using Lemma C.1, equation (58) we have

$$\sup_\omega \|G_{u,M}(\omega) - G_{u,M}^{(N)}(\omega)\|_2 \leq \mathcal{K}\frac{M}{N}, \tag{65}$$

where $\mathcal{K}$ is a generic constant that depends only on K and κ. The above immediately implies $\lambda_{\inf} - \mathcal{K}M/N \leq \|G_{u,M}(\omega)\|_2 \leq \lambda_{\sup} + \mathcal{K}M/N$. Finally we return to $G_u(\omega)$. Using Lemma C.1, equation (59) we have $\sup_\omega \|G_u(\omega) - G_{u,M}(\omega)\|_2 \leq \mathcal{K}/M$. By using this and (64) we have

$$\|G_u(\omega)\|_2 = \|G_{u,M}^{(N)}(\omega)\|_2 + O\left(\frac{M}{N} + \frac{1}{M}\right).$$

Finally, we set $M = 2\lfloor N^{1/2} \rfloor$ and substitute it into the above, this together with (63) gives

$$\lambda_{\inf} - \frac{\mathcal{K}}{N^{1/2}} \leq \|G_u(\omega)\|_2 \leq \lambda_{\sup} + \frac{\mathcal{K}}{N^{1/2}}.$$

As this holds for all $N > N_0$ we have that for any $\varepsilon > 0$ $\lambda_{\inf} - \varepsilon \leq \|G_u(\omega)\|_2 \leq \lambda_{\sup} + \varepsilon$. Thus leading to the required result. □

Proof of Theorem 3.2. We start by giving a short overview of the proof. To show that Assumption 3.1(i) holds (a uniform bound on the eigenvalues of $\boldsymbol{C}^{(N)}$) for a sufficiently large N, we first replace the infinite dimensional matrix $\boldsymbol{C}^{(N)}$ with an infinite dimensional banded matrix $\boldsymbol{C}_M^{(N)}$ (where we obtain a bound for $\|\boldsymbol{C}^{(N)} - \boldsymbol{C}_M^{(N)}\|_2$). The central part of the proof is to obtain a bound for the eigenvalues of $\boldsymbol{C}_M^{(N)}$ (that is uniform over a sufficiently large N). The key observation is that the banded matrix embeds an infinite number of overlapping $(M+1)\times(M+1)$-dimensional block matrices, where each block matrix can be approximated by an $(M+1)\times(M+1)$-dimensional block matrix whose entries consist of a stationary autocovariance. We will show that a lower and upper bound for the eigenvalues of each stationary approximation block matrix is $\gamma_{\inf}$ and $\gamma_{\sup}$ respectively. This yields a bound for the eigenvalues of each $(M+1)\times(M+1)$-dimensional block matrices in $\boldsymbol{C}_M^{(N)}$. Finally, motivated by the proof of Proposition 2.9 in Ding and Zhou (2021), we show that the eigenvalues of the banded matrix $\boldsymbol{C}_M^{(N)}$ can be bounded by the eigenvalues of "overlapping" $(M+1)\times(M+1)$-dimensional block matrices. This will prove the result.

We start by defining the infinite dimensional (block) banded matrix $\boldsymbol{C}_M^{(N)}$ where for all $t,\tau\in\mathbb{Z}$ the entries are defined by $[\boldsymbol{C}_M^{(N)}]_{t,\tau}=\mathbb{1}(|t-\tau|\leq M)C_{t,\tau}$. Without loss of generality we assume that $M=2m$ where $m\in\mathbb{N}$. Using Lemma B.2 we have $\|\boldsymbol{C}^{(N)}-\boldsymbol{C}_M^{(N)}\|_2\leq\mathcal{K}M^{-\kappa+1}$. Our aim is to obtain bounds for $x^\top\boldsymbol{C}_M^{(N)}x$ where $x=(\ldots,x_{-1},x_0,x_1,\ldots)^\top\in\ell_{2,p},x_l\in\mathbb{R}^p$ and $\|x\|_2=1$. To do this we define the $(M+1)p$-dimensional shifting subsequence $x_{s-m,s+m}=(x_{s-m},\ldots,x_{s+m})^\top$ and the $(M+1)p\times(M+1)p$ dimensional (block) matrix

$$\boldsymbol{C}^{(N)}(s-m,s+m)=(C_{t,\tau}^{(N)};s-m\leq t,\tau\leq s+m).$$

It can be shown that for $|t-\tau|\leq M+1$ the entries of $\boldsymbol{C}_M^{(N)}$ can be written in terms of $\boldsymbol{C}^{(N)}(s-m,s+m)$

$$[C_M^{(N)}]_{t,t+r}=\begin{cases}\frac{1}{M-|r|}\sum_{s=0}^{2m-r}C^{(N)}(t-s,t+2m-s)_{(s+1,s+1+r-2m)} & r\geq 0\\ \frac{1}{M-|r|}\sum_{s=0}^{2m-|r|}C^{(N)}(t-2m+s,t+s)_{(s+1+r-2m,s+1)} & r<0\end{cases}$$

For each $u\in\mathbb{Z}$ we define the stationary approximation matrix $\boldsymbol{C}(s-m,s+m;u)$

$$\boldsymbol{C}(s-m,s+m;u)=(C_{t-\tau}(u);s-m\leq t,\tau\leq s+m).$$

Under Assumption 3.1(iii) and using Lemma A.2 we have

$$\begin{aligned}&\|\boldsymbol{C}^{(N)}(s-m,s+m)-\boldsymbol{C}(s-m,s+m;s/N)\|_2\\ \leq\ &\sup_{t\in(s-m,s+m)}\sum_{\tau=s-m}^{s+m}\|C_{t,\tau}-C_{t-\tau}(s/N)\|_2\leq\mathcal{K}\frac{m}{N},\end{aligned}\tag{66}$$

where $\mathcal{K}$ is a generic constant that holds for all N and s. The condition

$$0 < \gamma_{\inf} \leq \inf_u \inf_\omega \lambda_{\inf}(f(\omega;u)) \leq \sup_u \sup_\omega \lambda_{\sup}(f(\omega;u)) \leq \gamma_{\sup} < \infty$$

implies (see, among others, (Basu and Michailidis, 2015, Proposition 2.3)) that for all $u \in \mathbb{R}$

$$\begin{aligned}\gamma_{\inf} &\leq \inf_\omega \lambda_{\inf}[f(\omega;u)] \leq \lambda_{\inf}[\boldsymbol{C}(s-m,s+m;u)] \\ &\leq \lambda_{\sup}[\boldsymbol{C}(s-m,s+m;u)] \leq \sup_\omega \lambda_{\sup}[f(\omega;u)] \leq \gamma_{\sup}.\end{aligned}$$

Therefore by using (65) and the above we have

$$\begin{aligned}\left(\gamma_{\inf} - \mathcal{K}\frac{m}{N}\right)\|x_{s-m,s+m}\|_2 &\leq x_{s-m,s+m}^\top \boldsymbol{C}^{(N)}(s-m,s+m)x_{s-m,s+m} \\ &\leq \left(\gamma_{\sup} + \mathcal{K}\frac{m}{N}\right)\|x_{s-m,s+m}\|_2.\end{aligned} \tag{67}$$

This gives a bound for each block. Next we obtain a bound between

$$x^\top \boldsymbol{C}_M^{(N)} x = \sum_{\ell\in\mathbb{Z}} \sum_{r=-M}^{M} x_\ell^\top C_{\ell,\ell+r} x_{\ell+r} \tag{68}$$

with the overlapping block matrix inner-product

$$X_M^\top \boldsymbol{O}_M X_M := \frac{1}{M+1}\sum_{s\in\mathbb{Z}} x_{s-m,s+m}^\top \boldsymbol{C}^{(N)}(s-m,s+m)x_{s-m,s+m}.$$

Note we have not formally defined X_M or O_M but have simply set it to equal the above. Basic algebra gives

$$X_M^\top \boldsymbol{O}_M X_M = \sum_{\ell\in\mathbb{Z}} \sum_{r=-M}^{M} \left(\frac{M+1-|r|}{M+1}\right) x_\ell^\top C_{\ell,\ell+r} x_{\ell+r}. \tag{69}$$

Using (67) and (68) we have

$$x^\top \boldsymbol{C}_M^{(N)} x - X_M^\top \boldsymbol{O}_M X_M = \frac{1}{M+1}\sum_{\ell\in\mathbb{Z}} \sum_{r=-M}^{M} |r| x_\ell^\top C_{\ell,\ell+r} x_{\ell+r}.$$

Hence under Assumption 3.1(ii) we have

$$\left\|x^\top \boldsymbol{C}_M^{(N)} x - X_M^\top \boldsymbol{O}_M X_M\right\|_2 \leq \frac{1}{M+1}\sum_{\ell\in\mathbb{Z}}\sum_{r=-m}^{m}\frac{|r|}{v(r)^\kappa}\|x_\ell\|_2\|x_{\ell+r}\|_2$$
$$\leq \frac{2}{M+1}\left(\sum_{r=1}^{\infty}\frac{1}{v(r)^{\kappa-1}}\right)\sum_{\ell\in\mathbb{Z}}\|x_\ell\|_2^2 = \frac{2}{M+1}\left(\sum_{r=1}^{\infty}\frac{1}{v(r)^{\kappa-1}}\right), \quad (70)$$

where the last line follows because $\|x\|_2 = \sum_{\ell\in\mathbb{Z}}\|x_\ell\|_2^2 = 1$. Finally, we obtain an upper and lower bound for $X_M^\top \boldsymbol{O}_M X_M$. We use (66) to give

$$\frac{(\gamma_{\inf} - \mathcal{K}m/N)}{M}\sum_{s\in\mathbb{Z}}\|x_{s-m,s+m}\|_2^2 \leq X_M^\top \boldsymbol{O}_M X_M \leq \frac{(\gamma_{\sup} + \mathcal{K}m/N)}{M}\sum_{s\in\mathbb{Z}}\|x_{s-m,s+m}\|_2^2.$$

Using that $\sum_{s\in\mathbb{Z}}\|x_{s-m,s+m}\|_2^2 = (M+1)\|x\|_2^2 = (M+1)$ we have

$$\gamma_{\inf} - \mathcal{K}m/N \leq X_M^\top \boldsymbol{O}_M X_M \leq \gamma_{\sup} + \mathcal{K}m/N.$$

Hence by using (69), $\|\boldsymbol{C}^{(N)} - \boldsymbol{C}_M^{(N)}\|_2 \leq \mathcal{K}M^{-\kappa+1}$ and setting $m = \lfloor N^{1/\kappa}\rfloor$ we have

$$\gamma_{\inf} - \mathcal{K}N^{-1+1/\kappa} \leq x^\top \boldsymbol{C}^{(N)} x \leq \gamma_{\sup} + \mathcal{K}N^{-1+1/\kappa},$$

where $\mathcal{K}$ is generic constant that does not depend on N or M. Thus for a sufficiently large N we have the result. □

C.2 Proofs for spectral-norm physical dependence systems

In order to prove Theorem 3.3 we require the following lemma.

Lemma C.2. *Suppose $\{V_t\}_t$ and $\{U_t\}_t$ are zero mean multivariate time series of dimension p that have the causal representation $V_t = v_t(\mathcal{F}_t)$ and $U_t = u_t(\mathcal{F}_t)$ where $\mathcal{F}_t = (\varepsilon_t, \varepsilon_{t-1}, \ldots)$ and $\{\varepsilon_t\}$ are iid random vectors. Let $\{\widetilde{\varepsilon}_t\}_t$ are iid random vectors with the same distribution as $\{\varepsilon_t\}_t$ but independent of them and set $\mathcal{F}_{t|\{t-j\}} = (\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-j+1}, \widetilde{\varepsilon}_{t-j}, \varepsilon_{t-j-1}, \ldots)$. Let $U_{t|\{t-j\}} = u_t(\mathcal{F}_{t|\{t-j\}})$ and $V_{t|\{t-j\}} = v_t(\mathcal{F}_{t|\{t-j\}})$. We assume that $\{V_t\}$ and $\{U_t\}$ satisfy the spectral-norm physical dependence conditions*

$$\sup_t \|\mathbb{V}\mathrm{ar}((U_t - U_{t|\{t-j\}})\|_2 \leq A\delta_j \text{ and } \sup_t[\mathbb{V}\mathrm{ar}(V_t - V_{t|\{t-j\}})\|_2 \leq B\delta_j$$

where $\delta_j = v(j)^{-\kappa}$ *and* $\kappa > 1$. *Then*

$$\|\,\mathbb{C}\mathrm{ov}(U_t, V_\tau)\|_2 \leq AB\delta_{|t-\tau|}\sum_{j=0}^{\infty}\delta_j.$$

Proof. To prove the result we write both V_t and U_t as the sum of martingale differences and represent $\mathbb{C}\mathrm{ov}(U_t, V_\tau)$ as the covariance of the martingale difference. This representation together with the physical dependence condition will prove the result. The details are below.

With a small abuse of notation we define the sigma algebra $\mathcal{F}_t = \sigma(\varepsilon_t, \varepsilon_{t-1}, \ldots)$. Since $U_t, V_t \in \mathcal{F}_t$, almost surely we can represent U_t and V_t as the infinite sum

$$U_t = \sum_{j=0}^{\infty} D_U(t,j) \text{ and } V_t = \sum_{j=0}^{\infty} D_V(t,j),$$

where $D_U(t,j) = \mathrm{E}(U_t|\mathcal{F}_{t-j}) - \mathrm{E}(U_t|\mathcal{F}_{t-j-1})$ and $D_V(t,j) = \mathrm{E}(V_t|\mathcal{F}_{t-j}) - \mathrm{E}(V_t|\mathcal{F}_{t-j-1})$. Without loss of generality we assume that $t - j_1 < \tau - j_2$, then by iterated expectations

$$\begin{aligned} \mathrm{E}\left[D_U(t,j_1)D_V(\tau,j_2)^\top\right] &= \mathrm{E}\left[D_U(t,j_1)\mathrm{E}[D_V(\tau,j_2)^\top|\mathcal{F}_{t-j_1}]\right] \\ &= \mathrm{E}\left[D_U(t,j_1)\mathrm{E}[(\mathrm{E}(V_t|\mathcal{F}_{t-j}) - \mathrm{E}(V_t|\mathcal{F}_{t-j-1}))^\top|\mathcal{F}_{t-j_1}]\right] = 0 \end{aligned} \tag{71}$$

where the above holds because for any $t - j_1 < \tau - i$, $\mathrm{E}[\mathrm{E}(V_\tau|\mathcal{F}_{\tau-i})|\mathcal{F}_{t-j_1}] = \mathrm{E}[V_\tau|\mathcal{F}_{t-j_1}]$. By a similar argument, $\mathrm{E}\left[D_U(t,j_1)D_V(\tau,j_2)^\top\right] = 0$ for $t - j_1 > \tau - j_2$.

We use (70) to write $\mathbb{C}\mathrm{ov}[V_t, U_\tau]$ as the product of martingale differences. Using (70) and assuming $t < \tau$ we have

$$\mathbb{C}\mathrm{ov}[V_t, U_\tau] = \sum_{j=\tau-t}^{\infty} \mathrm{E}[D_V(t,j)D_U(\tau,\tau-t+j)^\top].$$

Applying Lemma A.3 to the above gives

$$\|\,\mathbb{C}\mathrm{ov}[V_t, U_\tau]\|_2 \leq \sum_{j=\tau-t}^{\infty} \|\,\mathbb{V}\mathrm{ar}[D_V(t,j)]\|_2^{1/2}\|\,\mathbb{V}\mathrm{ar}[D_U(\tau,\tau-t+j)]\|_2^{1/2}. \tag{72}$$

Finally to bound the above expectations we use the physical dependence condition and the observation

$$D_V(t,j) = \mathrm{E}[V_t - V_{t|\{t-j\}}|\mathcal{F}_{t-j}].$$

Thus by using the law of total variance we have

$$\| \mathbb{V}\text{ar}[D_V(t,j)]\|_2^{1/2} \leq \| \mathbb{V}\text{ar}[(V_t - V_{t|\{t-j\}})]\|_2^{1/2} \leq B\delta_j.$$

By the same argument we have $\| \mathbb{V}\text{ar}[D_U(\tau,\tau-t+j)^2]\|_2^{1/2} \leq A\delta_{\tau-t+j}$. Substituting these bounds into (71) and using that $\delta_j = v(j)^{-\kappa}$ we have

$$|\mathbb{C}\text{ov}[V_t, U_\tau]| \leq AB\delta_{\tau-t}\sum_{j=0}^{\infty}\delta_j.$$

A similar bound holds for the case $t > \tau$. Thus proving the result. □

Proof of Theorem 3.3. We first show that condition (A) implies that Assumption 3.1(ii,iii) hold.

By using Lemma C.2 (with $V_t = W_t = X_{t,N}$) and the spectral-norm physical dependence condition on $\{X_{t,N}\}$ it immediately follows that Assumption 3.1(ii) holds.

We now show that $C_r(u) = \mathbb{C}\text{ov}[X_0(u), X_r(u)]$ satisfies Assumption 3.1(iii) parts (a),(b) and (c). Assumption 3.1(iii) part (a) holds by definition of $C_r(u)$. Assumption 3.1(iii) part (b) follows from Lemma C.2 and the spectral-norm physical dependence condition on $\{X_t(u)\}_t$. To show that Assumption 3.1(iii) part (c) holds we treat the case $|u-v| \leq 1$ and $|u-v| > 1$ separately. For $|u-v| > 1$ by using (b) we have

$$\begin{aligned} \| \mathbb{C}\text{ov}(X_t(u), X_\tau(u)) - \mathbb{C}\text{ov}(X_t(v), X_\tau(v))\|_2 &\leq \| \mathbb{C}\text{ov}(X_t(u), X_\tau(u))\|_2 + \| \mathbb{C}\text{ov}(X_t(v), X_\tau(v))\|_2 \\ &\leq \frac{2K}{v(t-\tau)^\kappa} \leq \frac{2K|u-v|}{v(t-\tau)^\kappa}, \end{aligned}$$

thus (c) holds for $|u-v| > 1$. For the case $|u-v| \leq 1$, we first use condition i) which states that $\widetilde{X}_t^{u,v} = X_t(u) - X_t(v)$ satisfies the physical dependence condition $\| \mathbb{V}\text{ar}(\widetilde{X}_t^{u,v} - \widetilde{X}_{t|\{t-j\}}^{u,v})\|_2 \leq |u-v|K\delta_j$. Thus, by using the expansion $X_\tau(v) = X_\tau(u) + \widetilde{X}_\tau^{u,v}$ and applying Lemma C.2 we have

$$\begin{aligned} \|C_{t-\tau}(u) - C_{t-\tau}(v)\|_2 &\leq \left[\| \mathbb{C}\text{ov}[X_t(v), \widetilde{X}_\tau^{u,v}]\|_2 + \| \mathbb{C}\text{ov}[\widetilde{X}_t^{u,v}, X_\tau(v)]\|_2 + \| \mathbb{C}\text{ov}[\widetilde{X}_\tau^{u,v}, \widetilde{X}_t^{u,v}\|_2\right] \\ &\leq \frac{K|u-v|}{v(t-\tau)^\kappa} + \frac{K|u-v|}{v(t-\tau)^\kappa} + \frac{K|u-v|^2}{v(t-\tau)^\kappa} \leq \frac{3K|u-v|}{v(t-\tau)^\kappa}, \end{aligned}$$

where the last inequality follows due to the condition $|u-v| \leq 1$. This proves that Assumption 3.1(ii)(c) holds.

Finally to prove that (10) holds, we use a similar technique as above. We focus on the case $|t-\tau| > N$ and $|t-\tau| \leq N$ separately. For $|t-\tau| > N$ and by using the above for the bounds for $C_{t,\tau}^{(N)}$ and $C_{t-\tau}(t/N)$ it can be shown that $\|C_{t,\tau}^{(N)} - C_{t-\tau}(t/N)\|_2 \leq 2Kv(t-\tau)^{-\kappa}$. On the other hand for $|t-\tau| \leq N$ we use that

$$\begin{aligned} X_{t,N} &= X_t(t/N) + e_{t,N} \\ \text{and } X_{\tau,N} &= X_\tau(\tau/N) + e_{\tau,N} = X_\tau(t/N) + \{X_\tau(\tau/N) - X_\tau(t/N)\} + e_{\tau,N}. \end{aligned}$$

Substituting the above into $\mathbb{C}\text{ov}(X_{t,N}, X_{\tau,N})$ gives

$$\mathbb{C}\text{ov}(X_{t,N}, X_{\tau,N}) = \mathbb{C}\text{ov}\left[X_t(t/N) + e_{t,N}, X_\tau(t/N) + \{X_\tau(\tau/N) - X_\tau(t/N)\} + e_{\tau,N}\right].$$

Expanding out the above covariance and using that $\mathbb{C}\text{ov}[X_t(t/N), X_\tau(t/N)] = C_{t-\tau}(t/N)$ we have

$$\begin{aligned} & \|C_{t,\tau}^{(N)} - C_{t-\tau}(t/N)\|_2 \\ \leq\ & \|\mathbb{C}\text{ov}[X_t(t/N), (X_\tau(\tau/N) - X_\tau(t/N))]\|_2 + \|\mathbb{C}\text{ov}[X_t(t/N), e_{\tau,N}]\|_2 \\ & + \|\mathbb{C}\text{ov}[e_{t,N}, X_\tau(t/N)]\|_2 + \|\mathbb{C}\text{ov}[e_{t,N}, (X_\tau(t/N) - X_\tau(\tau/N))]\|_2 + \|\mathbb{C}\text{ov}[e_{t,N}, e_{\tau,N}]\|_2. \end{aligned}$$

Under the spectral-norm physical dependence conditions (and by using Lemma C.2) we have

$$\begin{aligned} & \|C_{t,\tau}^{(N)} - C_{t-\tau}(t/N)\|_2 \\ \leq\ & K\left(|t-\tau|N^{-1}v(t-\tau)^{-\kappa} + 2\cdot N^{-1}v(t-\tau)^{-\kappa} + \frac{|t-\tau|}{N^2}v(t-\tau)^{-\kappa} + N^{-2}v(t-\tau)^{-\kappa}\right) \\ \leq\ & K\left(2\cdot N^{-1}v(t-\tau)^{-\kappa+1} + 3\cdot N^{-1}v(t-\tau)^{-\kappa}\right) \\ \leq\ & 5K\cdot N^{-1}v(t-\tau)^{-\kappa+1}, \end{aligned}$$

where the last line is due to $|t-\tau| \leq N$. The bounds for the two cases, $|t-\tau| \leq N$ and $|t-\tau| > N$ show that Assumption 3.1(iii) equation (10) holds.

Finally, under condition (B) and by applying Theorem 3.2 we have that Assumption 3.1(i) holds. □

In order to study the properties of the stochastic recurrence equation defined in Example 3.3 we state a general result for the time series $\{Y_t\}$ where

$$Y_t = G_t(\mathcal{F}_t) = \sum_{s=0}^{\infty} B_{t-s+1,t}(\mathcal{F}_{t-s+1,t}) b_{t-s}(\varepsilon_{t-s}) \tag{73}$$

with $\mathcal{F}_t = (\varepsilon_t, \varepsilon_{t-1}, \ldots)$, $\mathcal{F}_{t-s,t} = (\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-s})$ and $\{\varepsilon_t\}$ are iid random variables. Using Y_t we define the following coupled process

$$\begin{aligned} Y_{t|\{t-j\}} &= G_t(\mathcal{F}_{t|\{t-j\}}) = \sum_{s=0}^{t-j+2} B_{t-s+1,t}(\mathcal{F}_{t-s+1,t}) b_{t-s}(\varepsilon_{t-s}) \\ &\quad + B_{t-j+1,t}(\mathcal{F}_{t-s,t}) b_{t-j}(\widetilde{\varepsilon}_{t-j}) + \sum_{s=0}^{t-j+2} B_{t-s+1,t}(\mathcal{F}_{t-s+1,t}) b_{t-s}(\varepsilon_{t-s}), \end{aligned} \tag{74}$$

where $\mathcal{F}_{t|\{t-j\}} = (\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-j+1}, \widetilde{\varepsilon}_{t-j}, \varepsilon_{t-j-1}, \ldots)$ and if $s < j$ $\mathcal{F}_{t-s,t|\{t-j\}} = \mathcal{F}_{t-s,t}$ else $\mathcal{F}_{t-s,t|\{t-j\}} = (\varepsilon_t, \varepsilon_{t-1}, \ldots, \widetilde{\varepsilon}_{t-j}, \varepsilon_{t-j-1} \ldots, \varepsilon_{t-s})$ and $\widetilde{\varepsilon}_{t-j}$ is independent of ε_{t-j}.

To prove the following result we use that if B_1 and B_2 are conformable independent random matrices then

$$\|\mathrm{E}(B_1 B_2 B_2^\top B_1^\top)\|_2 \leq \|\mathrm{E}[B_1 B_1^\top]\|_2 \|\mathrm{E}[B_2 B_2^\top]\|_2. \tag{75}$$

Taking this further, if $B_1, \ldots, B_K$ are independent conformable random matrices then

$$\|\mathrm{E}(B_1 B_2 \ldots B_K B_K^\top \ldots B_2^\top B_1^\top)\|_2 \leq \prod_{i=1}^{K} \|\mathrm{E}[B_i B_i^\top]\|_2. \tag{76}$$

To simplify notation in the proofs below for the random vector or matrix X we let $V(X) = \mathrm{E}[XX^\top]$.

Lemma C.3. *Let Y_t and $Y_{t|\{t-j\}}$ be defined as in (72) and (73) respectively. Then we have*

$$\|V(Y_t - Y_{t|\{t-j\}})\|_2 \leq 4 \Big(\sum_{s=j}^{\infty} \|V[b_{t-s}(\varepsilon_{t-s})]\|_2^{1/2} \|V[B_{t-s+1,t}(\mathcal{F}_{t-s+1,t})]\|_2^{1/2}\Big)^2$$

Proof. Using (72) and (73) we have

$$Y_t - Y_{t|\{t-j\}} = B_{t-j+1,t}(\mathcal{F}_{t-j+1,t})[b_{t-j}(\varepsilon_{t-j}) - b_{t-j}(\widetilde{\varepsilon}_{t-j})] + \sum_{s=j+1}^{\infty} [B_{t-s+1,t}(\mathcal{F}_{t-s+1,t}) - B_{t-s+1,t}(\mathcal{F}_{t-s+1,t|\{t-j\}})]b_{t-s}(\varepsilon_{t-s}). \quad (77)$$

Applying (33) to the above gives

$$\|V(Y_t - Y_{t|\{t-j\}})\|_2^{1/2} \leq \|V(B_{t-j+1,t}(\mathcal{F}_{t-j+1,t})[b_{t-j}(\varepsilon_{t-j}) - b_{t-j}(\widetilde{\varepsilon}_{t-j})])\|_2^{1/2} + \sum_{s=j+1}^{\infty} \|V([B_{t-s+1,t}(\mathcal{F}_{t-s+1,t}) - B_{t-s+1,t}(\mathcal{F}_{t-s+1,t|\{t-j\}})]b_{t-s}(\varepsilon_{t-s}))\|_2^{1/2}.$$

By applying (74) to each of the terms above we have

$$\|V(Y_t - Y_{t|\{t-j\}})\|_2^{1/2} \leq \|V(b_{t-j}(\varepsilon_{t-j}) - b_{t-j}(\widetilde{\varepsilon}_{t-j}))\|_2^{1/2}\|V[B_{t-j+1,t}(\mathcal{F}_{t-j+1,t})]\|_2^{1/2} + \sum_{s=j+1}^{\infty} \|V[B_{t-s+1,t}(\mathcal{F}_{t-s+1,t}) - B_{t-s+1,t}(\mathcal{F}_{t-s+1,t|\{t-j\}})]\|_2^{1/2}\|V(b_{t-s}(\varepsilon_{t-s}))\|_2^{1/2}. \quad (78)$$

We now bound the terms inside the above sum. By using Lemma A.3 we have

$$\begin{aligned}
&\|V(b_{t-j}(\varepsilon_{t-j}) - b_{t-j}(\widetilde{\varepsilon}_{t-j}))\|_2 \leq \|V(b_{t-j}(\varepsilon_{t-j}))\|_2 + \|V(b_{t-j}(\widetilde{\varepsilon}_{t-j}))\|_2 \\
&+2\|\mathrm{E}[b_{t-j}(\varepsilon_{t-j})b_{t-j}(\widetilde{\varepsilon}_{t-j})^\top]\|_2 \\
\leq\ & \|V(b_{t-j}(\varepsilon_{t-j}))\|_2 + \|V(b_{t-j}(\widetilde{\varepsilon}_{t-j}))\|_2 + 2\|V(b_{t-j}(\varepsilon_{t-j}))\|_2^{1/2}\|V(b_{t-j}(\widetilde{\varepsilon}_{t-j}))]\|_2^{1/2} \\
=\ & 4\|V(b_{t-j}(\varepsilon_{t-j}))\|_2,
\end{aligned}$$

where the last line follows from the fact that $V(b_{t-j}(\varepsilon_{t-j})) = V(b_{t-j}(\widetilde{\varepsilon}_{t-j}))$. By a similar set of arguments we have

$$\|V[B_{t-s+1,t}(\mathcal{F}_{t-s+1,t}) - B_{t-s+1,t}(\mathcal{F}_{t-s+1,t|\{t-j\}})]\|_2 \leq 4\|V[B_{t-s+1,t}(\mathcal{F}_{t-s+1,t})]\|_2.$$

Substituting these bounds into (77) gives the result. □

We now apply the above result to the nonstationary model

$$X_{t,N} = A(t/N, \varepsilon_t)X_{t-1,N} + b(t/N, \varepsilon_t)$$

and its stationary approximation

$$X_t(u) = A(u, \varepsilon_t)X_{t-1}(u) + b(u, \varepsilon_t).$$

In the case $\sup_u \|\mathrm{E}[A(u, \varepsilon_t)A(u, \varepsilon_t)^\top]\|_2 < 1$, then both $X_{t,N}$ and $X_t(u)$ admit the causal solutions

$$\begin{aligned} X_{t,N} &= g_{t,N}(\mathcal{F}_t) = \sum_{s=0}^{\infty}\prod_{i=0}^{s-1} A((t-i)/N, \varepsilon_{t-i})b((t-s)/N, \varepsilon_{t-s}) \\ X_t(u) &= g(u, \mathcal{F}_t) = \sum_{s=0}^{\infty}\prod_{i=0}^{s-1} A(u, \varepsilon_{t-i})b(u, \varepsilon_{t-s}). \end{aligned}$$

We will use Lemma C.3 to prove the assertion in Example 3.3. The above expansions for $X_{t,N}$ and $X_t(u)$ allow us to apply Lemma C.3 to obtain the physical dependence bound. In the same spirit we require analogous expansions for $e_{t,N} = X_{t,N} - X_t(t/N)$ and $\widetilde{X}_t^{v_1,v_2} = X_t(v_1) - X_t(v_1)$

$$\begin{aligned} e_{t,N} &= \sum_{s=0}^{\infty}\prod_{i=0}^{s-1}\left[A\left(\frac{(t-i)}{N}, \varepsilon_{t-i}\right) b\left(\frac{(t-s)}{N}, \varepsilon_{t-s}\right) - \prod_{i=0}^{s-1} A\left(\frac{t}{N}, \varepsilon_{t-i}\right) b\left(\frac{t}{N}, \varepsilon_{t-s}\right)\right] \\ &= \sum_{s=0}^{\infty}\sum_{k=0}^{s-1}\prod_{i=0}^{k-1} A\left(\frac{t-i}{n}, \varepsilon_{t-i}\right)\left[A\left(\frac{t-k}{n}, \varepsilon_{t-k}\right) - A\left(\frac{t}{n}, \varepsilon_{t-k}\right)\right] \\ &\quad \times \prod_{i=k+1}^{s-1} A\left(\frac{t}{n}, \varepsilon_{t-i}\right) b(\frac{t-s}{n}, \varepsilon_{t-s}) \\ &\quad + \sum_{s=0}^{\infty}\prod_{i=0}^{s-1} A(u, \varepsilon_{t-i})\left[b\left(\frac{t-s}{N}, \varepsilon_{t-s}\right) - b\left(\frac{t}{n}, \varepsilon_{t-s}\right)\right]. \end{aligned}$$

and

$$\begin{aligned} \widetilde{X}_t^{v_1,v_2} &= \sum_{s=0}^{\infty}\sum_{k=0}^{s-1}[\prod_{i=0}^{k-1} A(v_1, \varepsilon_{t-i})][A(v_1, \varepsilon_{t-k}) - A(v_2, \varepsilon_{t-k})][\prod_{i=k+1}^{s-1} A(v_2, \varepsilon_{t-i})]b(\frac{t-s}{n}, \varepsilon_{t-s}) + \\ &\quad \sum_{s=0}^{\infty}\prod_{i=0}^{s-1} A(v_2, \varepsilon_{t-i})[b(v_1, \varepsilon_{t-s}) - b(v_2, \varepsilon_{t-s})]. \end{aligned}$$

Using the above expansion we can prove Example 3.3, which is given in the following lemma.

Lemma C.4. *Suppose* $\sup_u \|\mathrm{E}[A(u, \varepsilon_t)A(u, \varepsilon_t)^\top]\|_2 < \rho < 1$, $\sup_u \|\mathrm{E}[b(u, \varepsilon_t)b(u, \varepsilon_t)^\top]\|_2 <$

∞ *and for all* v_1 *and* v_2 $\|\mathrm{E}[(A(v_1,\varepsilon_t)-A(v_2,\varepsilon_t))(A(v_1,\varepsilon_t)-A(v_2,\varepsilon_t))^\top]\|_2 \leq K|v_1-v_2|$ *and* $\|\mathrm{E}[(b(v_1,\varepsilon_t)-b(v_2,\varepsilon_t))(b(v_1,\varepsilon_t)-b(v_2,\varepsilon_t))^\top]\|_2 \leq K|v_1-v_2|$. *Then*

$$\begin{aligned}
\sup_N \sup_t \|\mathbb{V}\mathrm{ar}(X_{t,N}-X_{t,N|\{t-j\}})\|_2 &\leq K\rho^j \\
\sup_u \|\mathbb{V}\mathrm{ar}(X_{t|\{t-j\}}(u)-X_{t|\{t-j\}}(u))\|_2 &\leq K\rho^j \\
\|\mathbb{V}\mathrm{ar}(X_t^{v_1,v_2}-X_{t|\{t-j\}}^{v_1,v_2}])\|_2 &\leq K|v_1-v_2|(\sum_{s=j}^{\infty} s\rho^{(s-1)/2})^2 \\
\textit{and } \sup_N \sup_t \|\mathbb{V}\mathrm{ar}(e_{t,N}-e_{t,N|\{t-j\}})\|_2 &\leq KN^{-1}(\sum_{s=j}^{\infty} s^{3/2}\rho^{(s-1)/2})^2.
\end{aligned}$$

Proof. To prove the result we obtain bounds for $\|V[b_{t-s}(\varepsilon_{t-s})]\|_2$ and $\|V[B_{t-s+1,t}(\mathcal{F}_{t-s+1,t})]\|_2$ and apply Lemma C.3. By using that $\{\varepsilon_t\}_t$ are iid random vectors and applying (75) we have

$$\left\| V\left(\prod_{i=0}^{s-1} A(u,\varepsilon_{t-i})\right)\right\|_2 \leq \|V(A(u,\varepsilon_0))\|_2^s \leq \rho^s,$$

$$\left\| V\left(\prod_{i=0}^{s-1} A((t-i)/n,\varepsilon_{t-i})\right)\right\|_2 \leq \prod_{i=0}^{s-1}\|V(A((t-i)/n,\varepsilon_0))\|_2 \leq \rho^s.$$

Further, by using (33) we have

$$\begin{aligned}
&\left\| V\left(\sum_{k=0}^{s-1}[\prod_{i=0}^{k-1} A(v_1,\varepsilon_{t-i})][A(v_1,\varepsilon_{t-k})-A(v_2,\varepsilon_{t-k})][\prod_{i=k+1}^{s-1} A(v_2,\varepsilon_{t-i})]\right)\right\|_2 \\
\leq & \left(\sum_{k=0}^{s-1}\left\| V\left([\prod_{i=0}^{k-1} A(v_1,\varepsilon_{t-i})][A(v_1,\varepsilon_{t-k})-A(v_2,\varepsilon_{t-k})][\prod_{i=k+1}^{s-1} A(v_2,\varepsilon_{t-i})]\right)\right\|_2^{1/2}\right)^2 \\
\leq & \left(\sum_{k=0}^{s-1}\|V(A(v_1,\varepsilon_{t-i})\|_2^{k/2}\|V[A(v_1,\varepsilon_{t-k})-A(v_2,\varepsilon_{t-k})\|_2^{1/2}\|V(A(v_2,\varepsilon_{t-i})])\|_2^{(s-k-1)/2}\right)^2 \\
\leq & Ks^2\rho^{s-1}|v_1-v_2|
\end{aligned}$$

and

$$\left\|V\left(\sum_{k=0}^{s-1}[\prod_{i=0}^{k-1}A(\frac{t-i}{n},\varepsilon_{t-i})][A(\frac{t-k}{n},\varepsilon_{t-k})-A(\frac{t}{n},\varepsilon_{t-k})][\prod_{i=k+1}^{s-1}A(\frac{t}{n},\varepsilon_{t-i})]\right)\right\|_2 \leq \frac{Ks^3\rho^{s-1}}{n}.$$

We use the above bounds obtain the result.

To bound $\|\mathbb{V}\mathrm{ar}(X_{t,N}-X_{t,N|\{t-j\}})\|_2$ we set

$$B_{t-s+1,t}(\mathcal{F}_{t-s+1,t}) = \prod_{i=0}^{s-1}A((t-i)/n,\varepsilon_{t-i}) \text{ and } b_{t-s}(\varepsilon_{t-s}) = b((t-s)/n,\varepsilon_{t-s}).$$

Now by applying Lemma C.3 we have

$$\sup_N\sup_t \|\mathbb{V}\mathrm{ar}[X_{t,N}-X_{t,N|\{t-j\}})\|_2 \quad \leq \quad (K\sum_{s=j}\rho^{s/2})^{1/2} \leq K\rho^j.$$

By using similar arguments we obtain the remaining bounds. □

C.3 Proof of results in Sections 3.3 and 3.4

Proof of Theorem 3.4. We begin with the proof of (13). Note that $\boldsymbol{C}^{-1}=\boldsymbol{D}$. Using the classical matrix inverse expansion we have

$$\begin{aligned}\boldsymbol{D}(u)-\boldsymbol{D}(v) &= \boldsymbol{C}(u)^{-1}-\boldsymbol{C}(v)^{-1} = \boldsymbol{C}(u)^{-1}[\boldsymbol{C}(v)-\boldsymbol{C}(u)]\boldsymbol{C}(v)^{-1}\\ &= \boldsymbol{D}(u)[\boldsymbol{C}(v)-\boldsymbol{C}(u)]\boldsymbol{D}(v). \end{aligned} \tag{79}$$

Thus by the Lipschitz continuity of $\boldsymbol{C}$ (see Assumption 3.1(iii)) and Theorem 2.1, we have

$$\begin{aligned}&\|D_{t-\tau}(u)-D_{t-\tau}(v)\|_2 = \sum_{s_1,s_2\in\mathbb{Z}}(\boldsymbol{D}(v))_{t,s_1}(\boldsymbol{C}(u)-\boldsymbol{C}(v))_{s_1,s_2}(\boldsymbol{D}(u))_{s_2,\tau}\\ &\leq K\mathcal{K}^2\sum_{s_1,s_2\in\mathbb{Z}}\zeta(t-s_1)^{\kappa-1}\frac{|u-v|}{v(s_2-s_1)^{\kappa}}\zeta(s_2-\tau)^{\kappa-1}\\ &= K\mathcal{K}^2\sum_{s_1,s_2\in\mathbb{Z}}\zeta(s_1)^{\kappa-1}\frac{|u-v|}{v(s_2+\tau-t-s_1)^{\kappa}}\zeta(s_2)^{\kappa-1}\\ &\leq 49K\mathcal{K}^2|u-v|\zeta(\tau-t)^{\kappa-1},\end{aligned}$$

where the last inequality follows from Lemma A.4 and $\mathcal{K}$ is finite constant, independent of u,v,t,τ. This proves (13).

To prove (14), we note that using the classical inverse matrix expansion (analogous to (77)) we have

$$\boldsymbol{D}^{(N)} - \boldsymbol{D}(t/N) \;=\; \boldsymbol{D}^{(N)}\left(\boldsymbol{C}(t/N) - \boldsymbol{C}^{(N)}\right)\boldsymbol{D}(t/N).$$

Theorem 2.1 gives bounds for the entries in $\boldsymbol{D}(t/N)$ and $\boldsymbol{D}^{(N)}$. On the other hand, Assumption 3.1 gives the bound

$$\begin{aligned}\|\left(\boldsymbol{C}(t/N) - \boldsymbol{C}^{(N)}\right)_{s_1,s_2}\|_2 &\leq \|\left(\boldsymbol{C}(t/N) - \boldsymbol{C}(s_1/N)\right)_{s_1,s_2}\|_2 + \|\left(\boldsymbol{C}(s_1/N) - \boldsymbol{C}^{(N)}\right)_{s_1,s_2}\|_2 \\ &\leq K\left(\min\left(\frac{|t-s_1|}{Nv(s_1-s_2)^{\kappa}}, \frac{2}{v(s_1-s_2)^{\kappa}}\right) + \frac{1}{Nv(s_1-s_2)^{\kappa-1}}\right).\end{aligned}$$

Substituting these bounds into $[\boldsymbol{D}^{(N)}\left(\boldsymbol{C}(t/N) - \boldsymbol{C}^{(N)}\right)\boldsymbol{D}(t/N)]_{t,\tau}$ gives

$$\begin{aligned}&\|(\boldsymbol{D}^{(N)} - \boldsymbol{D}(t/N))_{t,\tau}\|_2 \\ \leq\;& K\mathcal{K}^2\sum_{s_1,s_2\in\mathbb{Z}}\zeta(t-s_1)^{\kappa-1}\left(v(s_1-s_2)^{-\kappa}\min\left(\frac{|t-s_1|}{N},2\right) + \frac{1}{Nv(s_1-s_2)^{\kappa-1}}\right)\zeta(\tau-s_2)^{\kappa-1} \\ \leq\;& K\mathcal{K}^2\min\left(\sum_{s_1,s_2\in\mathbb{Z}}\zeta(t-s_1)^{\kappa-2}\times\frac{1}{Nv(s_1-s_2)^{\kappa}}\times\zeta(\tau-s_2)^{\kappa-1},\right. \\ &\left. 2\sum_{s_1,s_2\in\mathbb{Z}}\zeta(t-s_1)^{\kappa-1}\times\frac{1}{v(s_1-s_2)^{\kappa}}\times\zeta(\tau-s_2)^{\kappa-1}\right) \\ &+K\mathcal{K}^2\sum_{s_1,s_2\in\mathbb{Z}}\zeta(t-s_1)^{\kappa-1}\times\frac{1}{Nv(s_1-s_2)^{\kappa-1}}\times\zeta(\tau-s_2)^{\kappa-1} \\ \leq\;& 98K\mathcal{K}^2\zeta(t-\tau)^{\kappa-2}\min(1/N, 2\zeta(t-\tau)),\end{aligned}$$

where the last bound follows from Lemma A.4. This proves (14). □

Proof of equation (15). By using (77) we have

$$D_r(u) - D_r(v) = \sum_{s_1,s_2\in\mathbb{Z}} D_{s_1}(u)[C_{s_1}(u) - C_{s_2}(v)]D_{s_2-r}(v).$$

Let $h \in \mathbb{R}\backslash\{0\}$, and substitute $v = u+h$ and $u = u$ into the above to give

$$[D_r(u) - D_r(u+h)]/h = \sum_{s_1,s_2\in\mathbb{Z}} D_{s_1}(u)\frac{[C_{s_1}(u) - C_{s_2}(u+h)]}{h}D_{s_2-r}(u+h).$$

Taking the limit $h \to 0$ (and using dominated convergence to exchange limit and sum) gives the entry-wise matrix derivative

$$\frac{dD_r(u)}{du} = -\sum_{s_1,s_2} D_{s_1}(u)\frac{dC_{s_1-s_2}(u)}{du}D_{s_2-r}(u)$$

and the bound

$$\left\|\frac{dD_r(u)}{du}\right\|_2 \leq \sum_{s_1,s_2}\|D_{s_1}(u)\|_2\|\frac{dC_{s_1-s_2}(u)}{du}\|_2\|D_{s_2-r}(u)]\|_2 \leq \mathcal{K}\zeta(r)^{\kappa-1},$$

where the last inequality follows from Theorem 3.4, the condition $\sup_u \|\frac{dC_r(u)}{du}\|_2 \leq K\zeta(r)^{\kappa-1}$ and Lemma A.4. □

Proof of Corollary 3.1. To prove the result we start with the inverse matrix $\boldsymbol{D}^{(N)} = (\boldsymbol{C}^{(n)})^{-1}$ which we show below has simple easily derivable properties. We then apply Theorem 3.2, Lemma B.1, and Theorem 3.4 to obtain analogous properties on its inverse $\boldsymbol{C}^{(N)} = (\boldsymbol{D}^{(n)})^{-1}$.

Define the matrix

$$\widetilde{\Phi}_j(t/N) = \begin{cases} I_p & j=0 \\ -\Phi_j(t/N) & 1\leq j\leq p \\ 0 & \text{otherwise} \end{cases}.$$

Using $\{\Phi_j(u)\}_j$ we define the stationary time $X_t(u) = \sum_{j=1}^{d}\Phi_j(u)X_{t-j}(u) + \Sigma(u)^{1/2}\varepsilon_t$. This has the inverse (stationary) covariance $\boldsymbol{D}(u) = (D_{t-\tau}(u); t,\tau\in\mathbb{Z})$ where

$$D_{t-\tau}(u) = \sum_{\ell=0}^{d}\widetilde{\Phi}_\ell(u)^\top\Sigma(u)^{-1}\widetilde{\Phi}_{(t-\tau)+\ell}(u). \tag{80}$$

The corresponding inverse spectral density is $f(\omega;u)^{-1} = \sum_{r\in\mathbb{Z}} D_r(u)\exp(ir\omega)$. Under the stated conditions on the roots associated with $\{\Phi_j(u)\}_r$ we have that for some γ_1 and γ_2 that $0 < \gamma_1 \leq \inf_u\inf_\omega \lambda_{\inf}(f(\omega;u)^{-1}) \leq \sup_u\sup_\omega\lambda_{\sup}(f(\omega;u)^{-1}) \leq \gamma_2 < \infty$ and thus the eigenvalues of $\boldsymbol{D}(u)$ are uniformly bounded away from γ_1 and γ_2. Let $\boldsymbol{C}(u) = \boldsymbol{D}(u)^{-1} = (C_{t-\tau}; t,\tau\in\mathbb{Z})$. Then by using Lemma B.1 we have

$$\sup_u \|C_r(u)\|_2 \leq \mathcal{K}\rho^{|r|} \tag{81}$$

for some $0<\rho<1$. Further, by using (13) (applied to exponential decay rather than polynomial decay) we have $\|C_r(u) - C_r(v)\|_2 \leq \mathcal{K}\rho^{|r|}|u-v|$.

Using the Cholesky decomposition it can be shown that the inverse covariance is $\boldsymbol{D}^{(N)} = (D_{t,\tau}; t, \tau \in \mathbb{Z})$ where

$$D_{t,\tau}^{(N)} = \sum_{\ell=0}^{d} \widetilde{\Phi}_{\ell}\left(\frac{t+\ell}{N}\right)^{\top} \Sigma\left(\frac{t+\ell}{N}\right)^{-1} \widetilde{\Phi}_{(t-\tau)+\ell}\left(\frac{t+\ell}{N}\right). \tag{82}$$

The Lipschitz conditions on $\Phi_j(\cdot)$ together with (78) and (79) imply that $D_{t,\tau}^{(N)}$ is approximated by $D_{t-\tau}(t/N)$. I.e.

$$|D_{t,\tau}^{(N)} - D_{t-\tau}(t/N)\|_2 \leq \begin{cases} \frac{\mathcal{K}}{N} & |t-\tau| \leq d \\ 0 & |t-\tau| > d \end{cases}.$$

Now by using the above and Theorem 3.2 for large enough N the conditions in Assumption 3.1 hold (in terms of the inverse covariance). Therefore for sufficiently large N, the rate $\|C_{t,\tau}^{(N)}\|_2 \leq \mathcal{K}\rho^{|t-\tau|}$ follows from Lemma B.1. Further, the conditions in Theorem 3.4 hold and we have

$$\|C_{t,\tau}^{(N)} - C_{t-\tau}(t/N)\| \leq \mathcal{K}\frac{\rho^{|r|}}{N},$$

which gives $\|C_{t,\tau}^{(N)} - C_{t-\tau}(t/N)\|_2 \leq \mathcal{K}\frac{\rho^{|t-\tau|}}{N}$. Thus we have proved the result. □

Proof of Theorem 3.5. The result uses the bounds $\|[\boldsymbol{C}^{(N)}(-\infty, T)^{-1}]_{s_1,s_2} \leq \mathcal{K}\zeta(s_1-s_2)^{\kappa-1}$ and $\|[\boldsymbol{C}(-\infty, T; u)^{-1}]_{s_1,s_2} \leq \mathcal{K}\zeta(s_1-s_2)^{\kappa-1}$. The assertion follows by the same steps as in the proof of Theorem 3.4. □

Proof of Theorem 3.6. To prove the result we start with the following identities

$$\begin{aligned} \Phi_{T,j}^{(N)} &= -([\boldsymbol{C}^{(N)}(-\infty, T)^{-1}]_{T,T})^{-1}[\boldsymbol{C}(-\infty, T)^{-1}]_{T,T-j} \qquad (83) \\ \text{and } \Phi_j(u) &= -([\boldsymbol{C}(-\infty, T; u)^{-1}]_{T,T})^{-1}[\boldsymbol{C}(-\infty, T; u)^{-1}]_{T,T-j} \end{aligned}$$

where $\boldsymbol{C}^{(N)}(-\infty, T) = (C_{t,\tau}^{(N)}; t, \tau \leq T)$ and $\boldsymbol{C}(-\infty, T; u) = (C_{t,\tau}(u); t, \tau \leq T)$. These identities together with Theorem 3.5 will be used to prove the result.

We first obtain a bound for $\|\Sigma_T^{(N)} - \Sigma(T/N)\|_2$. We note that

$$\begin{aligned}
& \Sigma_T^{(N)} - \Sigma(T/N) = ([\boldsymbol{C}^{(N)}(-\infty,T)^{-1}]_{T,T})^{-1} - ([\boldsymbol{C}(-\infty,T;T/N)^{-1}]_{T,T})^{-1} \\
= \; & ([\boldsymbol{C}(-\infty,T;T/N)^{-1}]_{T,T})^{-1}([\boldsymbol{C}(-\infty,T;T/N)^{-1}]_{T,T} \\
& -[\boldsymbol{C}^{(N)}(-\infty,T)^{-1}]_{T,T})([\boldsymbol{C}^{(N)}(-\infty,T)^{-1}]_{T,T})^{-1}.
\end{aligned}$$

Thus by using Theorem 3.5 (with $t = T$ and $\tau = T$) we have

$$\begin{aligned}
\|\Sigma_t^{(N)} - \Sigma(t/N)\|_2 \;\leq\; & \|([\boldsymbol{C}(-\infty,T;T/N)^{-1}]_{T,T})^{-1}\|_2\|([\boldsymbol{C}^{(N)}(-\infty,T)^{-1}]_{T,T})^{-1}\|_2 \\
& \times\|[\boldsymbol{C}(-\infty,T;T/N)^{-1} - \boldsymbol{C}^{(N)}(-\infty,T)^{-1}]_{T,T}\|_2 \\
\leq\; & \mathcal{K}N^{-1}. \qquad (84)
\end{aligned}$$

This proves the first part of (i)

To prove the second part of (ii), we use (79) to give the decomposition $\Phi_{t,j}^{(N)} - \Phi_j(t/N) = J_1 + J_2$, where

$$\begin{aligned}
J_1 &= -\left[([\boldsymbol{C}^{(N)}(-\infty,T)^{-1}]_{T,T})^{-1} - [\boldsymbol{C}(-\infty,t;t/N)^{-1}]_{T,T})^{-1}\right][\boldsymbol{C}^{(N)}(-\infty,T)^{-1}]_{T,T-j}, \\
J_2 &= -([\boldsymbol{C}(-\infty,T;T/N)^{-1}]_{T,T})^{-1}\left[[\boldsymbol{C}^{(N)}(-\infty,T)^{-1} - \boldsymbol{C}(-\infty,T;T/N)^{-1}]_{T,T-j}\right].
\end{aligned}$$

First we bound J_1 this gives

$$\begin{aligned}
\|J_1\|_2 \;\leq\; & \left\|([\boldsymbol{C}^{(N)}(-\infty,T)^{-1}]_{T,T})^{-1} - ([\boldsymbol{C}(-\infty,T;T/N)^{-1}]_{T,T})^{-1}\right\|_2 \\
& \times\|[\boldsymbol{C}^{(N)}(-\infty,T)^{-1}]_{T,T-j}\|_2 \\
\leq\; & \mathcal{K}\frac{1}{N}\zeta(0)^{\kappa-1} \times \zeta(j)^{\kappa-1}.
\end{aligned}$$

where we have used the bounds in Theorem 2.1 and (80) in the above. Using a similar argument (and Theorem 3.5 (with $t = T$ and $\tau = T - j$)) we have

$$\begin{aligned}
\|J_2\|_2 \;\leq\; & \|([\boldsymbol{C}(-\infty,T;T/N)^{-1}]_{T,T})^{-1}\|_2\left\|[\boldsymbol{C}^{(N)}(-\infty,T)^{-1} - \boldsymbol{C}(-\infty,T;T/N)^{-1}]_{T,T-j}\right\|_2 \\
\leq\; & \mathcal{K}\zeta(j)^{\kappa-2}\min(2\zeta(j), 1/N).
\end{aligned}$$

Altogether this gives $\|\Phi_{T,j}^{(N)} - \Phi_j(T/N)\|_2 \leq \mathcal{K}\zeta(j)^{\kappa-2}\min(2\zeta(j), 1/N)$. Thus we have proved the second part of (i). The proof for (ii) follows a similar method as given in the proof of Theorem 3.4, and we omit the details. □

We now prove Theorem 3.7. To prove this result we will use the alternative representation of the covariance operator $\boldsymbol{C}^{(N)}$ defined in Remark 2.2. With this in mind, we define the sub-operators $\boldsymbol{C}^{(e,f)} : \ell_2 \to \ell_2$ which are infinite dimensional matrices where $[\boldsymbol{C}^{(e,f)}]_{t,\tau} = \mathbb{C}\text{ov}[X_{t,N}^{(e)}, X_{t,N}^{(f)}]$. Note that to reduce cumbersome notation, we have dropped the N from the definition $\boldsymbol{C}^{(e,f)}$. We also define the corresponding "stationary" matrix operators $\boldsymbol{C}^{(e,f)}(u) : \ell_2 \to \ell_2$, where $[\boldsymbol{C}^{(e,f)}(u)]_{t,\tau} = \mathbb{C}\text{ov}[X_t^{(e)}(u), X_t^{(f)}(u)]$. This representation is instrumental in proving the result below.

Proof of Theorem 3.7. We first prove (26) and (27). We start by obtaining an expression for

$$
\begin{aligned}
\mathbb{V}\text{ar}\left[X_t^{(c)|\text{-}\{a,b\}}; t \in \mathbb{Z}, c \in \{1,2\}\right] &= (\Delta_{t,\tau,N}^{\text{-}\{a,b\}}; t, \tau \in \mathbb{Z}) \\
\text{and } \mathbb{V}\text{ar}\left[X_t(u)^{(c)|\text{-}\{a,b\}}; t \in \mathbb{Z}, c \in \{1,2\}\right] &= (\Delta_{t-\tau}^{\text{-}\{a,b\}}(u); t, \tau \in \mathbb{Z}).
\end{aligned}
$$

To simplify notation, and without loss of generality, we focus on the case $a = 1, b = 2$. We will represent the above in terms of block matrices of $\boldsymbol{C}^{(N)}$ and $\boldsymbol{C}(u)$. We define $\boldsymbol{A}^{(1,2)} : \ell_{2,2} \to \ell_{2,2}$, $\boldsymbol{B}^{(1,2)} : \ell_{2,p-2} \to \ell_{2,2}$ and $\boldsymbol{E}^{(1,2)} : \ell_{2,p-2} \to \ell_{2,p-2}$ where

$$
\begin{aligned}
\boldsymbol{A}^{(1,2)} &= \begin{pmatrix} \boldsymbol{C}^{(1,1)} & \boldsymbol{C}^{(1,2)} \\ \boldsymbol{C}^{(2,1)} & \boldsymbol{C}^{(2,2)} \end{pmatrix}, \boldsymbol{B}^{(1,2)} = \begin{pmatrix} \boldsymbol{C}^{(1,3)} & \dots & \boldsymbol{C}^{(1,p)} \\ \boldsymbol{C}^{(2,3)} & \dots & \boldsymbol{C}^{(2,p)} \end{pmatrix} \\
\text{and } \boldsymbol{E}^{(1,2)} &= (\boldsymbol{C}^{(e,f)}; e, f \in \{3, \dots, p\}).
\end{aligned}
$$

Analogously, we define $\boldsymbol{A}^{(1,2)}(u), \boldsymbol{B}^{(1,2)}(u), \boldsymbol{E}^{(1,2)}(u)$. It is clear the operators $\boldsymbol{A}^{(1,2)}$, $\boldsymbol{B}^{(1,2)}$ and $\boldsymbol{E}^{(1,2)}$ are comprised of an infinite number of 2×2, $2 \times (p-2)$ and $(p-2) \times (p-2)$ matrices respectively. To denote these sub-matrices we use the following notation. Suppose $\boldsymbol{H} : \ell_{2,p_1} \to \ell_{2,p_2}$ for some p_1, p_2 then $[\boldsymbol{H}]_{t,\tau} := (I_{p_1} \otimes e_t)^\top B^{(1,2)}(I_{p_2} \otimes e_\tau)$ refers to their $p_1 \times p_2$-dimensional submatrices.

It is well known that the conditional covariance of $X_{t,N}^{(c)}$ and $X_t^{(c)}(u)$ can be represented as the Schur complement

$$
\mathbb{V}\text{ar}\left[X_{t,N}^{(c)|\text{-}\{1,2\}}; t \in \mathbb{Z}, c \in \{1,2\}\right] = \boldsymbol{A}^{(1,2)} - \boldsymbol{B}^{(1,2)}(\boldsymbol{E}^{(1,2)})^{-1}(\boldsymbol{B}^{(1,2)})^\top
$$

and

$$
\mathbb{V}\text{ar}\left[X_t(u)^{(c)|\text{-}\{1,2\}}; t \in \mathbb{Z}, c \in \{1,2\}\right] = \boldsymbol{A}^{(1,2)}(u) - \boldsymbol{B}^{(1,2)}(u)(\boldsymbol{E}^{(1,2)}(u))^{-1}(\boldsymbol{B}^{(1,2)}(u))^\top.
$$

Then, we have

$$\Delta^{\text{-}\{a,b\}}_{t,\tau,N} = [\boldsymbol{A}^{(1,2)} - \boldsymbol{B}^{(1,2)}(\boldsymbol{E}^{(1,2)})^{-1}(\boldsymbol{B}^{(1,2)})^{\top}]_{t,\tau}$$
$$\text{and } \Delta^{\text{-}\{a,b\}}_{t-\tau}(u) = [\boldsymbol{A}^{(1,2)}(u) - \boldsymbol{B}^{(1,2)}(u)(\boldsymbol{E}^{(1,2)}(u))^{-1}(\boldsymbol{B}^{(1,2)}(u))^{\top}]_{t,\tau}. \quad (85)$$

We use the above representations to prove (26). Using (81) we have

$$\|\Delta^{\text{-}\{a,b\}}_{t,\tau,N} - \Delta^{\text{-}\{a,b\}}_{t-\tau}(t/N)\|_2 \leq J_1 + J_2 + J_3 + J_4$$

where

$$J_1 = \|(\boldsymbol{A}^{(1,2)} - \boldsymbol{A}^{(1,2)}(t/N))_{t,\tau}\|_2$$
$$J_2 = \|(\boldsymbol{B}^{(1,2)}(\boldsymbol{E}^{(1,2)})^{-1}(\boldsymbol{B}^{(1,2)} - \boldsymbol{B}^{(1,2)}(t/N))^{\top})_{t,\tau}\|_2$$
$$J_3 = \|(\boldsymbol{B}^{(1,2)}((\boldsymbol{E}^{(1,2)})^{-1} - (\boldsymbol{E}^{(1,2)}(t/N))^{-1})(\boldsymbol{B}^{(1,2)}(t/N))^{\top})_{t,\tau}\|_2$$
$$J_4 = \|((\boldsymbol{B}^{(1,2)} - \boldsymbol{B}^{(1,2)}(t/N))(\boldsymbol{E}^{(1,2)}(t/N))^{-1}(\boldsymbol{B}^{(1,2)}(t/N))^{\top})_{t,\tau}\|_2.$$

Under Assumption 3.1 and by using Theorem 2.1 we bound the terms above (the proof is in the spirit of the proof of Theorem 3.4). Assumption 3.1 directly implies

$$J_1 = \|(\boldsymbol{A}^{(1,2)} - \boldsymbol{A}^{(1,2)}(t/N))_{t,\tau}\|_2 \leq K\frac{1}{Nv(t-\tau)^{\kappa-1}}.$$

The bounds for J_2, J_3 and J_4 are more involved, however all three follow a similar strategy. We focus on obtaining a bound for J_3. Using standard matrix multiplication it can be seen that

$$J_3 = \|\sum_{s_1,s_2\in\mathbb{Z}}[\boldsymbol{B}^{(1,2)}]_{t,s_1}[(\boldsymbol{E}^{(1,2)})^{-1} - (\boldsymbol{E}^{(1,2)}(t/N))^{-1}]_{s_1,s_2}[\boldsymbol{B}^{(1,2)}(t/N))^{\top}]_{s_2,\tau}\|_2$$
$$\leq \sum_{s_1,s_2\in\mathbb{Z}}\|[\boldsymbol{B}^{(1,2)}]_{t,s_1}\|_2 \cdot \|[(\boldsymbol{E}^{(1,2)})^{-1} - (\boldsymbol{E}^{(1,2)}(t/N))^{-1}]_{s_1,s_2}\|_2 \cdot \|(\boldsymbol{B}^{(1,2)}(t/N))^{\top}]_{s_2,\tau}\|_2 \quad (86)$$

To bound $\|[\boldsymbol{B}^{(1,2)}]_{t,s_1}\|_2$ and $\|(\boldsymbol{B}^{(1,2)}(t/N))^{\top}]_{s_2,\tau}\|_2$ we simply use Assumption 3.1, which immediately gives

$$\|[\boldsymbol{B}^{(1,2)}]_{t,s_1}\|_2 \leq Kv(t-s_1)^{-\kappa} \text{ and } \|(\boldsymbol{B}^{(1,2)}(t/N))^{\top}]_{s_2,\tau}\|_2 \leq Kv(s_2-\tau)^{-\kappa}. \quad (87)$$

The bound for $\|[(\boldsymbol{E}^{(1,2)})^{-1} - (\boldsymbol{E}^{(1,2)}(t/N))^{-1}]_{s_1,s_2}\|_2$ needs a little more work. We first note that the covariance operator $\boldsymbol{E}^{(1,2)}$ is a suboperator of $\boldsymbol{C}^{(N)}$, thus it satisfies Assumption 3.1 where $\boldsymbol{E}^{(1,2)}(u)$ is its locally stationary approximation. Therefore we can apply the results of Theorem 3.4 to $(\boldsymbol{E}^{(1,2)})^{-1}$ and this gives

$$\|((\boldsymbol{E}^{(N),(1,2)})^{-1} - (\boldsymbol{E}^{(1,2)}(s_1/N))^{-1})_{s_1,s_2}\|_2 \leq \mathcal{K}\zeta(s_1-s_2)^{\kappa-2}\min(1/N, 2\zeta(s_1-s_2)) \quad (88)$$

and

$$\|((\boldsymbol{E}^{(1,2)}(s_1/N))^{-1} - (\boldsymbol{E}^{(1,2)}(t/N))^{-1})_{s_1,s_2}\|_2 \leq \mathcal{K}|s_1-t|\zeta(s_1-s_2)^{\kappa-1}/N. \quad (89)$$

Substituting (83), (84) and (85) into (82) we have

$$\begin{aligned} J_3 &\leq \mathcal{K}K^2 \sum_{s_1,s_2\in\mathbb{Z}} \frac{1}{v(t-s_1)^{\kappa}} \times \big(\zeta(s_1-s_2)^{\kappa-2}\min(1/N, 2\zeta(s_1-s_2)) \\ &\quad + |s_1-t|\zeta(s_1-s_2)^{\kappa-1}/N\big) \frac{1}{v(s_2-\tau)^{\kappa}} \\ &\leq 2\times(49)K^2\mathcal{K}\zeta(t-\tau)^{\kappa-2}\min(1/N,\zeta(t-\tau)) =: \mathcal{K}\zeta(t-\tau)^{\kappa-2}\min(1/N,\zeta(t-\tau)), \end{aligned}$$

where the last line follows from Lemma A.4.

To bound J_2, we use Theorem 2.1 to give, $\|[(\boldsymbol{E}^{(1,2)})^{-1}]_{s_1,s_2}\| \leq \mathcal{K}\zeta(s_1-s_2)^{(\kappa-1)}$. This together with (83), using the bounds stated in Assumption 3.1(iii) and following the same proof as above we can show that

$$J_2 \leq \mathcal{K}\zeta(t-\tau)^{\kappa-1}/N \text{ and } J_4 \leq \mathcal{K}\zeta(t-\tau)^{\kappa-1}/N.$$

Altogether the bounds for J_1, J_2, J_3 and J_4 prove

$$\|\Delta_{t,\tau,N}^{-\{a,b\}} - \Delta_{t-\tau}^{-\{a,b\}}(t/N)\|_2 \leq \mathcal{K}\zeta(t-\tau)^{\kappa-2}\min(1/N,\zeta(t-\tau))$$

thus proving (26). The proof of (27) follows a similar technique.

Finally, the proofs for (28) and (29) are the same as the proofs for (26) and (27), thus we omit the details. □