# Inverse covariance operators of multivariate nonstationary time series

Jonas Krampe and Suhasini Subba Rao

September 3, 2024

**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 (2021), 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 (2021) 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 $\ell_2$/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($\infty$) 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 Section 4 and 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. 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 $\ell_2$ and $\ell_{2,p}$ to denote the sequence spaces $\ell_2 = \{u = (\ldots, u_{-1}, u_0, u_1, \ldots); u_j \in \mathbb{R}$ and $\sum_{j\in\mathbb{Z}} u_j^2 < \infty\}$ and $\ell_{2,p} = \{v = (\ldots, v_{-1}, v_0, v_1, \ldots); v_j \in \mathbb{R}^p$ and $\sum_{j\in\mathbb{Z}} \|v_j\|_2^2 < \infty\}$. On the spaces $\ell_2$ 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 $s$th element of the $a$th (column) space. Suppose $\{A_{s_1,s_2}\}_{s_1,s_2}$ are $p \times p$-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 $\boldsymbol{A} : \ell_{2,p} \to \ell_{2,p}$ in the sense that if $\boldsymbol{A}x = y$, then $y = (\ldots, y_{-1}, y_0, y_1, \ldots)$ where for all $t \in \mathbb{Z}$, $y_t \in \mathbb{R}^p$ and $y_t = \sum_{\tau\in\mathbb{Z}} A_{t,\tau} x_\tau$. Similarly, suppose $\{B_{s_1,s_2}\}_{s_1,s_2}$ are $p_1 \times p_2$ dimensional matrices, like $\boldsymbol{A}$, we define the infinite dimensional operator $\boldsymbol{B} = (B_{s_1,s_2}; s_1, s_2 \in \mathbb{Z})$ where $\boldsymbol{B} : \ell_{2,p_2} \to \ell_{2,p_1}$. 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 common factors[1] but allows for high dimensional sparse time series. 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) $\ell_2$-norm $\|\cdot\|_2$ of the $p\times p$ matrices. However, no conditions are placed on the $\ell_1$-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 $\ell_2$ sense as $p$ grows, then the conditional interactions are also controlled in the $\ell_2$-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 let $\zeta(j) = v(\log[v(j)])/v(j)$, where $v(\cdot) = \max(1, |\cdot|)$. The proofs in this section can be found in Section 4.1.

## 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

[1]By common 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.

[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 $\lambda_{\sup}$.

$\boldsymbol{C}(-\infty, T) = (C_{t,\tau}; t, \tau \leq T)$. As will be clear later in the paper, the inverse of $\boldsymbol{C}(-\infty, T)$ 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 $t, \tau \leq T$*

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

The above result shows that if the pairwise interaction between the components is bounded with a certain rate in the $\ell_2$-sense then the conditional interactions are also bounded with a certain rate in the $\ell_2$-sense, see Remark 3.1 for a discussion on the role of the dimension $p$.

**Remark 2.1.** *A key ingredient in the proof of Theorem 2.1 is that the inverse of banded matrices decay geometrically; see Demko et al. (1984) and our adapted version for block-banded matrices in Lemma 4.1. Because of this, if the 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 $[C^{(a,b)}]_{t,\tau} = C^{(a,b)}_{t,\tau} = \mathbb{C}\mathrm{ov}[X^{(a)}_t, X^{(b)}_\tau]$. $\widetilde{\boldsymbol{C}}$ 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 (2021)).*

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 $\sum_{s\in\mathbb{Z}}(1+|s|^{\kappa})\|C_s\|_2 < \infty$ is equivalent to a $\kappa$-times differentiability condition on the spectral density $f(\cdot) = (2\pi)^{-1}\sum_{s\in\mathbb{Z}} C_s \exp(-is\cdot)$. To elaborate, since the operator is block Toeplitz it has the representation $\boldsymbol{C} = (\int_0^{2\pi} f(\omega)\exp(i(t-\tau)\omega)d\omega)_{t,\tau}$ and $\boldsymbol{D} = (\int_0^{2\pi} f(\omega)^{-1}\exp(i(t-\tau)\omega)d\omega)_{t,\tau}$. This means the differentiability condition transfers from $f$ to $f^{-1}$ i.e., if $f$ is $\kappa$-times differentiable, then $f^{-1}$ is $\kappa$-times differentiable. Consequently, $\sum_{s\in\mathbb{Z}}(1+|s|^{\kappa})\|C_s\|_2 < \infty$ implies $\sum_{s\in\mathbb{Z}}(1+|s|^{\kappa})\|D_s\|_2 < \infty$. This global condition implies for all $t,\tau \in \mathbb{Z}$ that $\|C_{t-\tau}\|_2 \leq Kv(t-\tau)^{-\kappa}$ and $\|D_{t-\tau}\|_2 \leq Kv(t-\tau)^{-\kappa}$. Interestingly, the individual condition that yields the 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)($\infty$) 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($\infty$) coefficients in the linear projection of $X_T$ onto the space spanned by $\overline{\mathrm{sp}}(X_{T-1},X_{T-2},\ldots)$ i.e.,

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

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

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

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 bound for nonstationary multivariate time series, between the VAR($\infty$) 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 inverse identity 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*

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

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 is represented by a (vector) $AR(\infty)$ model. In the stationary context, building an autoregressive representation of an increasing order relates to the Cholesky decomposition of $\mathbb{V}\mathrm{ar}(X_1, \dots, X_n)^{-1}$ where the $i$th model is given by the $i$th line. The AR$(\infty)$ 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 fundamental 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 $\max_{T>b} \max_{1\le j\le b} |\Phi_{T,T-1,j} - \Phi_{T,b,j}| \le C(\log b)^{\kappa-1} b^{-\kappa+3}$. With Theorem 2.2 we cannot directly compare the coefficients of the two finite AR models (order $T-1$ and order $b$), however it is possible to use this theorem to obtain tighter bounds for their result. To be precise

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

The above leads to the bound (without the log-factors) $\max_{T>b} \max_{1\le j\le b} |\Phi_{T,T-1,j} - \Phi_{T,b,j}| = O(b^{-\kappa+3/2})$ instead of $O(b^{-\kappa+3})$.

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

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

noting that we have dropped $T$ as it is not necessary under stationarity. This implies $\sum_{j=1}^{d}\|\Phi_{d,j} - \Phi_j\|_2 \le \mathcal{K}d^{-\kappa}$. Based on the discussion in Section 2.2, at the individual level this means if $\|C_s\|_2 \le Kv(s)^{-\kappa-\varepsilon}$, then $\sum_{j=1}^{d}\|\Phi_{d,j} - \Phi_j\|_2 \le \mathcal{K}d^{-\kappa+1}$, whereas Theorem 2.2 gives $\sum_{j=1}^{d}\|\Phi_{d,j} - \Phi_j\|_2 \le \mathcal{K}d^{-\kappa+3/2}$. 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 Cramèr 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 non-linear 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($\infty$) representation. Finally (in Section 3.4) we apply these results to show that the smoothness conditions also transfer to the partial covariances. Unless stated otherwise, the proofs in this section can be found in Section 4.2.

## 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 $C_{t,\tau}^{(N)} = \mathbb{C}\text{ov}[X_{t,N}, X_{\tau,N}]$ (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*

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

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

*(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 $N$*

$$\|C_{t,\tau}^{(N)} - C_{t-\tau}(t/N)\|_2 \leq \frac{K}{v(t-\tau)^{\kappa-1}} \min\left(\frac{1}{N}, \frac{2}{v(t-\tau)}\right). \tag{9}$$

*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 $|t - \tau| < 2N$ (which is within the classical locally stationary framework), 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 $|t - \tau| \geq 2N$ (as can happen if $t, \tau \in \mathbb{Z}$), 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 the 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($\infty$)(tv-VMA) process)**.** *Consider the tv-VMA($\infty$)*

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

*where $\{\varepsilon_t\}_t$ are uncorrelated random variables with zero mean and variance $I_p$ (see Dahlhaus (1997) and Dahlhaus and Polonik (2006) for the case $p = 1$). In order for the process to be well defined certain summability or decay conditions need to be imposed on the coefficients $\{\Psi_{t,j}\}$. As in Dahlhaus (1997), 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 B.5 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 $\{X_{t,N}\}$. For this, suppose 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 $\|\Psi_{t,j}^{(N)} - \Psi_j(t/N)\| \leq Kv(j)^{-\kappa}/N$.*[3] *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 $\sup_u \|C_r(u)\|_2 \leq$*

[3]Note that without a change in rate, the condition can be weakened to $\|\Psi_{t,j}^{(N)} - \Psi_j(t/N)\| \leq Kv(j)^{-(\kappa-1)}/N$. For illustrative purposes, we use the rate $-\kappa$ here.

*$K/v(r)^{\kappa}$ (this follows from Lemma B.5). 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 $\kappa$ 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.

**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 for any $\varepsilon > 0$ we have $\lambda_{\inf} - \varepsilon \leq \lambda_{\inf}(\boldsymbol{C}(u)) \leq \lambda_{\sup}(\boldsymbol{C}(u)) \leq \lambda_{\sup} + \varepsilon$.*

*Proof.* See Appendix A. □

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

$$\sup_u \|D_{t-\tau}(u)\|_2 \leq \mathcal{K}\zeta(t-\tau)^{-\kappa+1}. \tag{10}$$

Assumption 3.1(i) is difficult to directly verify given a particularly nonstationary model. 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*

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

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

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

*Proof.* See Appendix A. □

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

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

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

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

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 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; define $X_{t,N} = G_{t,N}(\varepsilon_t, \varepsilon_{t-1}, \ldots)$ (or $X_{t,N} = G(X_{t-1,N}, \ldots, X_{t-p,N}; t/N)$), this is the general nonlinear physical dependence model described in Zhou and Wu (2009), Dahlhaus et al. (2019), Karmakar et al. (2021) and Ding and Zhou (2021)) and the auxillary stationary time series $X_t(u) = G(u; \varepsilon_t, \varepsilon_{t-1}, \ldots)$ (or $X_t(u) = G(X_{t-1}(u), \ldots, X_{t-p}(u); u)$). If these two models together satisfy Assumption 3.1(ii,iii) (and certain conditions on the spectral density of $\{X_t(u)\}_t$), then Assumption 3.1(i) also holds. Thus the results in the following sections apply to these models.

### 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 3.3.** *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 $u, v \in \mathbb{R}$*

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

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

$$\left\| \left[ \boldsymbol{D}^{(N)} - \boldsymbol{D}(t/N) \right]_{t,\tau} \right\|_2 \leq \mathcal{K}\zeta(t-\tau)^{\kappa-2} \min(1/N, 2\zeta(t-\tau)), \tag{12}$$

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

An important consequence of Theorem 3.3 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 differentiablity conditions also transfer from $C_r(u)$

to $D_r(u)$. E.g., if one starts with the condition that for all $r$ $\sup_u \|\frac{dC_r(u)}{du}\|_2 \leq K\zeta(r)^{\kappa-1}$, then using the same arguments as those used in the proof of Theorem 3.3 (outlined after the proof of Theorem 3.3) we have

$$\|\frac{dD_r(u)}{du}\|_2 \leq \mathcal{K}\zeta(r)^{\kappa-1}. \tag{13}$$

Smoothness and differentiability conditions on $\boldsymbol{D}^{(N)}$ and $\boldsymbol{D}(u)$ are used in Basu and Subba Rao (2021) (stated in Assumption 4.2) to obtain certain rates of decay on the Fourier transform of $\boldsymbol{D}^{(N)}$. Theorem 3.3 and (13) 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 3.3.

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

$$X_{t,N} = \sum_{j=1}^{d} \Phi_j(t/N)X_{t-j,N} + \Sigma(t/N)^{1/2}\varepsilon_t, \qquad t \in \mathbb{Z}, \tag{14}$$

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 (14) 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) we can use Theorem 2.1 and 3.3, to transfer the information on the rate of decay of the inverse covariance operator to the covariance operator itself.

**Corollary 3.1** (Application to VAR models)**.** *Suppose that the multivariate time series $\{X_{t,N}\}_t$ has the time-varying VAR($d$) representation in (14), where we assume there exists a $\delta > 0$ and $\gamma$ where*

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

*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.* See Appendix A. □

**Remark 3.2.** *As mentioned after Theorem 3.3 smoothness conditions in terms of differentiability transfer between $\boldsymbol{C}(\cdot)$ and $\boldsymbol{D}(\cdot)$. Applied to the above corollary, this implies that smoothness conditions formulated in terms of differentiability of the transition matrices $\Phi_j(\cdot)$ transfer to $\boldsymbol{D}(\cdot)$ and consequently 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 have shown in (4) that under certain conditions all nonstationary time series have an AR$(\infty)$ representation. But there is no guarantee that the AR parameters are smooth. However, 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

$$X_{T,N} = \sum_{j=1}^{\infty} \Phi_{T,j}^{(N)} X_{T-j,N} + \varepsilon_{T,N}, \tag{15}$$

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$(\infty)$ representation

$$X_t(u) = \sum_{j=1}^{\infty} \Phi_j(u) X_{t-j}(u) + \varepsilon_t(u), \tag{16}$$

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 3.4.** *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 (15) with $\Sigma_T^{(N)} = \mathbb{V}\mathrm{ar}[\varepsilon_{T,N}]$. Additionally, let $\{X_t(u)\}_t$ be the locally stationary approximation defined in (16).*

*(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 $j \geq 1$*

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

A potential benefit of Theorem 3.4 is that it could be used to develop a bootstrap procedure for nonstationary time series by transferring the widely used AR-sieve to the locally stationary setup.

**Remark 3.3** (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 $\mathcal{H}^{(N)} = \overline{\mathrm{sp}}(X_{t,N}^{(c)}; t \in \mathbb{Z}, 1 \leq c \leq p)$ 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 $\mathcal{H}^{(N)} - (X^{(c)}; c \in \mathcal{S}) = \overline{\mathrm{sp}}[X_{s,N}^{(c)}; s \in \mathbb{Z}, c \in \mathcal{S}']$ 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

$$X_{t,N}^{(a)|\text{-}\mathcal{S}} \quad := \quad X_{t,N}^{(a)} - P_{\mathcal{H}^{(N)}-(X^{(c)};c\in\mathcal{S})}(X_{t,N}^{(a)}), t \in \mathbb{Z}. \tag{17}$$

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

$$\Delta_{t,\tau,N}^{\text{-}\{a,b\}} = \begin{pmatrix} \rho_{t,\tau,N}^{(a,a)|\text{-}\{a,b\}} & \rho_{t,\tau,N}^{(a,b)|\text{-}\{a,b\}} \\ \rho_{t,\tau,N}^{(b,a)|\text{-}\{a,b\}} & \rho_{t,\tau,N}^{(b,b)|\text{-}\{a,b\}} \end{pmatrix} := \mathbb{C}\mathrm{ov}\left[\begin{pmatrix} X_{t,N}^{(a)|\text{-}\{a,b\}} \\ X_{t,N}^{(b)|\text{-}\{a,b\}} \end{pmatrix}, \begin{pmatrix} X_{\tau,N}^{(a)|\text{-}\{a,b\}} \\ X_{\tau,N}^{(b)|\text{-}\{a,b\}} \end{pmatrix}\right] \tag{18}$$

and self partial covariance

$$\rho_{t,\tau,N}^{(a,a)|\text{-}\{a\}} = \mathbb{C}\mathrm{ov}[X_{t,N}^{(a)|\text{-}\{a\}}, X_{\tau,N}^{(a)|\text{-}\{a\}}]. \tag{19}$$

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 $\{X_t(u) = (X_t^{(1)}(u),\dots,X_t^{(p)}(u))^\top\}_t$ locally approximates the partial covariance of $\{X_{t,N} = (X_{t,N}^{(1)},\dots,X_{t,N}^{(p)})^\top\}_t$. To do this, analogous to (17), (18) and (19) we define

$$X_t^{(a)|\text{-}\mathcal{S}}(u) \;:=\; X_{t,N}^{(a)}(u) - P_{\mathcal{H}_u-(X_u^{(c)};c\in\mathcal{S})}(X_t^{(a)}(u)) \text{ for } t\in\mathbb{Z}, \tag{20}$$

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

and self partial covariance

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

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 $\boldsymbol{C}^{(e,f)} = (\mathbb{C}\text{ov}[X_{t,N}^{(e)}, X_{\tau,N}^{(f)}]; t,\tau\in\mathbb{Z})$. The presence of $\boldsymbol{C}_{\mathcal{S}',\mathcal{S}'}^{-1}$ in the above expression explains why the results in the previous sections (in particular Theorem 3.3) are necessary for proving the result.

**Theorem 3.5.** *Suppose Assumption 3.1 holds and let* $\Delta_{t,\tau,N}^{\text{-}\{a,b\}}$, $\rho_{t,\tau,N}^{(a,a)|\text{-}\{a\}}$, $\Delta_{t-\tau}^{\text{-}\{a,b\}}(u)$ *and* $\rho_{t-\tau}^{(a,a)|\text{-}\{a\}}(u)$ *be defined as in (18), (19), (21) and (22). Then for all* $a,b\in\{1,\dots,p\}$

$$\begin{aligned}
\|\Delta_{t,\tau,N}^{\text{-}\{a,b\}} - \Delta_{t-\tau}^{\text{-}\{a,b\}}(t/N)\|_2 &\le \mathcal{K}\zeta(t-\tau)^{\kappa-2}\min(1/N,\zeta(t-\tau)) && (23)\\
\|\Delta_{t-\tau}^{\text{-}\{a,b\}}(u) - \Delta_{t-\tau}^{\text{-}\{a,b\}}(v)\|_2 &\le \mathcal{K}|u-v|\zeta(t-\tau)^{\kappa-1} && (24)\\
\|\rho_{t,\tau,N}^{(a,a)|\text{-}\{a\}} - \rho_{t-\tau}^{(a,a)|\text{-}\{a\}}(t/N)\|_2 &\le \mathcal{K}\zeta(t-\tau)^{\kappa-2}\min(1/N,\zeta(t-\tau)) && (25)\\
\textit{and } \|\rho_{t-\tau}^{(a,a)|\text{-}\{a\}}(u) - \rho_{t-\tau}^{(a,a)|\text{-}\{a\}}(v)\|_2 &\le \mathcal{K}|u-v|\zeta(t-\tau)^{\kappa-1}, && (26)
\end{aligned}$$

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

*Proof.* See Appendix A. $\square$

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 3.5 it can be shown that

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

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.

# 4 Proofs of the Main Results

Before proceeding with the proofs, we need to introduce further 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, $A_{s_1,s_2}^{(a,b)} = (e_{s_1} \otimes e_a)^\top \boldsymbol{A}(e_{s_2} \otimes e_b)$ 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 $\ell_2$ and $I_{-k}$ the identity operator after removing the $k$th 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. Furthermore, for an operator in $\ell_{2,p}$ we use $(I_{-k} \otimes I_p)$ to remove the $p$-dimensional row I guess you mean the $k$th block row of dimension $p$? and $(I_{-s} \otimes I_{-a}) =: I_{-(a,s)}$ to remove the $(a, s)$ row.

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

$$\boldsymbol{U}^{-1} = \begin{pmatrix} \widetilde{\boldsymbol{A}} & \widetilde{\boldsymbol{B}} \\ \widetilde{\boldsymbol{C}} & \widetilde{\boldsymbol{D}} \end{pmatrix} = \begin{pmatrix} \widetilde{\boldsymbol{A}} & -\widetilde{\boldsymbol{A}}\boldsymbol{B}\boldsymbol{D}^{-1} \\ -\boldsymbol{D}^{-1}\boldsymbol{C}\widetilde{\boldsymbol{A}} & \boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{C}\widetilde{\boldsymbol{A}}\boldsymbol{B}\boldsymbol{D}^{-1} \end{pmatrix} \tag{27}$$

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

$$\widetilde{\boldsymbol{A}} = (\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1} \text{ and } \widetilde{\boldsymbol{D}} = (\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}. \tag{28}$$

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

$$\boldsymbol{A} - \widetilde{\boldsymbol{A}}^{-1} = \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}. \tag{29}$$

We will make frequent uses of (27) and (29) in the proofs below.

## 4.1 Proofs 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

$$|B_{t,\tau}| \leq \frac{(1+\sqrt{r})^2}{b} \rho^{\lfloor |t-\tau|/M \rfloor + 1}, \tag{30}$$

where $r = b/a$ and $\rho = (\sqrt{r}-1)/(\sqrt{r}+1)$ with $a = \inf_{v \in \ell_2, \|v\|_2=1} \langle v, \boldsymbol{A}v \rangle$ and $b = \sup_{v \in \ell_2, \|v\|_2=1} \langle v, \boldsymbol{A}v \rangle$. 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 generalisation of (30) for block matrices.

**Lemma 4.1.** *Let $\boldsymbol{A}$ be a 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 $k$th $p$-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.* See Appendix B.1. $\square$

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 $|t-\tau| \geq 2$. Define the set $\mathcal{U} = \{\tau \in \mathbb{Z}, \tau \neq t\}$, and denote $C_{t,\mathcal{U}} = (e_t \otimes I_p)^\top \boldsymbol{C} (I_{-t} \otimes I_p)$ and $C_{\mathcal{U},\mathcal{U}} = (I_{-t} \otimes I_p)^\top \boldsymbol{C} (I_{-t} \otimes I_p)$. Without loss of generality we consider a permuated version of $\boldsymbol{C}$, which contains $C_{t,t}$ in the top left hand corner of $\boldsymbol{C}$

$$\boldsymbol{C} = \begin{pmatrix} C_{t,t} & C_{t,\mathcal{U}} \\ C_{\mathcal{U},t} & C_{\mathcal{U},\mathcal{U}} \end{pmatrix}.$$

Using the permuted $\boldsymbol{C}$ makes the proof easier to follow. The same proof applies to the non-permuted $\boldsymbol{C}$ too. Note that in this paper usually all infinite dimensional operators are in bold. Thus $C_{t,\mathcal{U}}, C_{\mathcal{U},t}$ and $C_{\mathcal{U},\mathcal{U}}$ should be in bold. But this makes the notation in the proof quite overbearing, so for this proof we have dropped the bold for these sub-operators. Using the block matrix inversion formula (see Tretter (2008), page 35; for operators) we have

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

From the above we observe that the $D_{t,\tau}$ matrix is

$$D_{t,\tau} = -D_{t,t}^{-1} C_{t,\mathcal{U}} C_{\mathcal{U},\mathcal{U}}^{-1} (I_{-t} \otimes I_p)^\top (e_\tau \otimes I_p),$$

using that $\lambda_{\sup}^{-1} \leq D_{t,t} \leq \lambda_{\inf}^{-1}$ we have $\|D_{t,\tau}\|_2 \leq \lambda_{\sup} \|(C_{t,\mathcal{U}} C_{\mathcal{U},\mathcal{U}}^{-1})(I_{-t}^\top e_\tau \otimes I_p)\|_2$. Thus we focus on bounding the induced $\ell_2$-norm of $A_{t,\tau} = (C_{t,\mathcal{U}} C_{\mathcal{U},\mathcal{U}}^{-1})(I_{-t}^\top e_\tau \otimes I_p)$.

In Lemma 4.1 we obtain the rate of decay for the entries of the inverse of a block M-banded matrix. We leverage on this result to obtain a rate of decay for $\boldsymbol{C}_{\mathcal{U},\mathcal{U}}^{-1}$. Thus the main outline in the proof below is to (a) replace $\boldsymbol{C}_{\mathcal{U},\mathcal{U}}$ with a (block) banded matrix (b) obtain a bound on the replacement error and (c) balance the rate of decay of the inverse banded matrix approximation with the replacement error (both of which depend on the bandwidth $M$).

To align the dimensions of $C_{\mathcal{U},\mathcal{U}}$ and $\boldsymbol{C}$ we will pad $C_{\mathcal{U},\mathcal{U}}$ with zeros in the sense that $(I_{-t} \otimes I_p) C_{\mathcal{U},\mathcal{U}} (I_{-t} \otimes I_p)^\top$ and $\boldsymbol{C}$ are identical everywhere except at the $t$th $p$-dimensional row/column. Using this notation, we define the following banded matrix. We replace $C_{\mathcal{U},\mathcal{U}}$ with a banded matrix $\boldsymbol{B}_M$ of bandwidth $M$ in the sense that for all $p \times p$ sub-matrices and $s_1, s_2 \in \mathbb{Z}$

$$((I_{-t} \otimes I_p) \boldsymbol{B}_M (I_{-t} \otimes I_p)^\top)_{s_1,s_2} = \mathbb{1}(|s_1 - s_2| \leq M)((I_{-t} \otimes I_p) C_{\mathcal{U},\mathcal{U}} (I_{-t} \otimes I_p)^\top)_{s_1,s_2}, \quad (31)$$

where $\mathbb{1}$ denotes the indicator function. To invert $C_{\mathcal{U},\mathcal{U}}$ we use the identity $C_{\mathcal{U},\mathcal{U}}^{-1} =$

$\boldsymbol{B}_M^{-1}[I + \boldsymbol{B}_M^{-1}(C_{\mathcal{U},\mathcal{U}} - \boldsymbol{B}_M)]^{-1}$. From Lemma B.1, equation (55) we show that $\|C_{\mathcal{U},\mathcal{U}} - \boldsymbol{B}_M\|_2 \leq 2K/(\kappa-1)(M-1)^{-\kappa+1}$. Thus if $M$ is sufficiently large, i.e., $M > 1 + (2K/(\kappa-1)^{1/(\kappa-1)}$, we have $\|\boldsymbol{C}_{\mathcal{U},\mathcal{U}} - \boldsymbol{B}_M\|_2 < 1$, this leads to the Neumann series $C_{\mathcal{U},\mathcal{U}}^{-1} = \boldsymbol{B}_M^{-1}[I + \sum_{s=1}^{\infty}(-1)^s[\boldsymbol{B}_M^{-1}(C_{\mathcal{U},\mathcal{U}} - \boldsymbol{B}_M)]^s]$. Substituting the above into $A_{t,\tau} = (C_{t,\mathcal{U}}C_{\mathcal{U},\mathcal{U}}^{-1})(I_{-t}^{\top}e_\tau \otimes I_p)$ gives

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

By applying the triangular inequality to the above we have $\|A_{t,\tau}\|_2 \leq J_1 + J_2$ where

$$\begin{aligned}
J_1 &= \|C_{t,\mathcal{U}}\boldsymbol{B}_M^{-1}(I_{-t}^{\top}e_\tau \otimes I_p)\|_2 \\
\text{and } J_2 &= \|C_{t,\mathcal{U}}\boldsymbol{B}_M^{-1}\sum_{s=1}^{\infty}(-1)^s[\boldsymbol{B}_M^{-1}(C_{\mathcal{U},\mathcal{U}} - \boldsymbol{B}_M)]^s](I_{-t}^{\top}e_\tau \otimes I_p)\|_2.
\end{aligned}$$

We now bound $J_1$ and $J_2$. By using Lemma B.1, we bound $J_2$ with

$$J_2 \leq \|C_{t,\mathcal{U}}\boldsymbol{B}_M^{-1}\|_2\sum_{s=1}^{\infty}(K^2M^{-\kappa+1})^s \leq 2K/(\kappa-1)(M-1)^{-\kappa+1}.$$

We next bound $J_1$. First, we expand the matrix multiplication of $C_{t,\mathcal{U}}\boldsymbol{B}_M^{-1}$, then use the sub-multiplicativety of $\|\cdot\|_2$. This gives

$$J_1 \leq \sum_{s\in\mathbb{Z},s\neq t}\|C_{t,s}\|_2 \cdot \|((I_{-t}\otimes I_p)\boldsymbol{B}_M^{-1}(I_{-t}\otimes I_p))_{s,\tau}\|_2.$$

Under Assumption 2.1 we have $\|C_{t,s}\|_2 \leq Kv(t-s)^{-\kappa}$, whereas Lemma 4.1 gives

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

where $r_M = \lambda_{\sup,M}/\lambda_{\inf,M}$, $\lambda_{\sup,M} \leq \lambda_{\sup} + 2K/(\kappa-1)(M-1)^{-\kappa+1}$, $\lambda_{\inf,M} \geq \lambda_{\inf} - 2K/(\kappa-1)(M-1)^{-\kappa+1}$, and $\rho_M = (\sqrt{r_M}-1)/(\sqrt{r_M}+1)$. If $M > (2K/(\kappa-1)\max(2/\lambda_{\inf}, 1/\lambda_{\sup}))^{(1/(\kappa-1)} + 1$, 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 $r = \lambda_{\text{sup}}/\lambda_{\text{inf}}$. With this, we have

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

Hence

$$J_1 \leq \frac{2K(1+2\sqrt{r})^2}{\lambda_{\text{sup}}}\sum_{s\in Z, s\neq t}|s-t|^{-\kappa}\rho^{\lfloor |s-\tau|/M\rfloor+1} \leq \frac{2K(1+2\sqrt{r})^2}{\lambda_{\text{sup}}}\sum_{s\in Z}\rho^{|s|/M}\frac{1}{v(s-t+\tau)^{\kappa}}.$$

Thus when $M > K_c := (2K/(\kappa-1)\max(2/\lambda_{\text{inf}}, 1/\lambda_{\text{sup}}))^{(1/(\kappa-1)}+1$, the above bounds for $J_1$ and $J_2$ hold and we have $\|D_{t,\tau}\|_2 \leq \lambda_{\text{sup}}(J_1+J_2)$.

The final part in the proof is to balance the two bounds $J_1$ and $J_2$. For this, 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 $J_1$ and $J_2$ it can be shown that

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

On the other hand, if $|t-\tau|$ is small, i.e., $M_{t-\tau}<K_c$ then it can be shown that

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

The above gives the following global bound for all $t,\tau\in\mathbb{Z}$

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

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

To prove (2), we only need to slightly modified the arguments. We use $\mathcal{U} = \{\tau \leq$

$T, \tau \neq t\}$ and obtain

$$\boldsymbol{C}(-\infty; T) = \begin{pmatrix} C_{t,t} & C_{t,\mathcal{U}} \\ C_{\mathcal{U},t} & C_{\mathcal{U},\mathcal{U}} \end{pmatrix}.$$

This leads to

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

Then, we can follow the same strategy. Note that the sums involved are now from $-\infty$ to $T$ but they are bounded from above by the sums given in (1). $\square$

*Proof of Theorem 2.2*. As mentioned in Section 2.3 in order to connect the coefficients $\{\Phi_{T,j}\}$ and $\Phi_{T,d,j}$ we first need to connect the inverses of $\boldsymbol{C}(T-d,T)$ and $\boldsymbol{C}(-\infty,T)$. For this let

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

where $\boldsymbol{C}(-\infty, T-d, T) = (C_{t,\tau}; t \leq T-d, T-d+1 \leq \tau \leq T)$ and $\boldsymbol{C}(T-d,T) = (C_{t,\tau}; T-d+1 \leq t, \tau \leq T)$. In order to compare the AR coefficients we use the block decomposition of $\boldsymbol{C}(-\infty, T)^{-1}$

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

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 prove the result we start by showing that for all $1 \leq t, \tau \leq d$ we have

$$\|[\boldsymbol{C}(T-d,T)^{-1} - \widetilde{\boldsymbol{D}}(T-d,T)]_{T-t,T-\tau}\|_2 \quad \leq \quad \mathcal{K}\zeta(d-t)^{\kappa-3/2}\zeta(d-\tau)^{\kappa-3/2}. \quad (32)$$

Using Schur's complement we have

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

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

$$\begin{aligned} & \|[\boldsymbol{C}(T-d,T)^{-1} - \widetilde{\boldsymbol{D}}(T-d,T)]_{T-t,T-\tau}\|_2 \\ \leq \; & \lambda_{\sup} \|(e_{T-t} \otimes I_p)^\top \widetilde{\boldsymbol{D}}(-\infty, T-d, T)\|_2 \|(e_{T-\tau} \otimes I_p)^\top \widetilde{\boldsymbol{D}}(-\infty, T-d, T)\|_2 . \end{aligned} \tag{33}$$

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 B.3 and using Theorem 2.1 we have

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

Substituting the above into (33) we have

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

This proves (32).

We now compare the bottom rows of $\boldsymbol{C}(T-d,T)^{-1}$ and $\widetilde{\boldsymbol{D}}(T-d,T)$ and (32) to prove (7). That is setting $t = 0$ and $\tau = j$ gives

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

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).

## 4.2 Proofs of results in Section 3

*Proof of Theorem 3.3.* We begin with the proof of (11). Note that $\boldsymbol{C}^{-1} = \boldsymbol{D}$. Using the classical matrix inverse expansion we have

$$\begin{aligned}\boldsymbol{D}(u) - \boldsymbol{D}(v) &= \boldsymbol{C}(u)^{-1} - \boldsymbol{C}(v)^{-1} = \boldsymbol{C}(u)^{-1}[\boldsymbol{C}(v) - \boldsymbol{C}(u)]\boldsymbol{C}(v)^{-1} \\ &= \boldsymbol{D}(u)[\boldsymbol{C}(v) - \boldsymbol{C}(u)]\boldsymbol{D}(v). \end{aligned} \tag{34}$$

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 B.5 and $\mathcal{K}$ is finite constant, independent of $u, v, t, \tau$. This proves (11).

To prove (12), we note that using the classical inverse matrix expansion (analogous to (34)) 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

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

where the last bound follows from Lemma B.5. This proves (12). $\square$

*Proof of* (13). By using (34) 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 3.3, the condition $\sup_u \|\frac{dC_r(u)}{du}\|_2 \leq K\zeta(r)^{\kappa-1}$ and Lemma B.5. $\square$

To prove Theorem 3.4, below, we require the following corollary of Theorem 3.3.

**Corollary 4.1.** *Suppose Assumption 3.1 holds and let* $\boldsymbol{C}^{(N)}(-\infty,T)=(C^{(N)}_{t,\tau};t,\tau\leq T)$ *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.* The result uses that $\|[\boldsymbol{C}^{(N)}(-\infty,T)^{-1}]_{s_1,s_2}\leq\mathcal{K}\zeta(s_1-s_2)^{\kappa-1}$ 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 3.3. $\square$

*Proof of Theorem 3.4.* To prove the result we start with the following identities

$$\begin{aligned}\Phi^{(N)}_{T,j} &= -([\boldsymbol{C}^{(N)}(-\infty,T)^{-1}]_{T,T})^{-1}[\boldsymbol{C}(-\infty,T)^{-1}]_{T,T-j} \\ \text{and }\Phi_j(u) &= -([\boldsymbol{C}(-\infty,T;u)^{-1}]_{T,T})^{-1}[\boldsymbol{C}(-\infty,T;u)^{-1}]_{T,T-j}\end{aligned}\qquad(35)$$

where $\boldsymbol{C}^{(N)}(-\infty,T)=(C^{(N)}_{t,\tau};t,\tau\leq T)$ and $\boldsymbol{C}(-\infty,T;u)=(C_{t,\tau}(u);t,\tau\leq T)$. These identities together with Corollary 4.1 will be used to prove the result.

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

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

Thus by using Corollary 4.1 (with $t=T$ and $\tau=T$) we have

$$\begin{aligned}\|\Sigma^{(N)}_t-\Sigma(t/N)\|_2 \leq\;& \|([\boldsymbol{C}(-\infty,T;T/N)^{-1}]_{T,T})^{-1}\|_2\cdot\|[\boldsymbol{C}(-\infty,T;T/N)^{-1}-\boldsymbol{C}^{(N)}(-\infty,T)^{-1}]_{T,T}\|_2\\ &\times\|([\boldsymbol{C}^{(N)}(-\infty,T)^{-1}]_{T,T})^{-1}\|_2\leq\mathcal{K}N^{-1}.\end{aligned}\qquad(36)$$

This proves the first part of (i)

To prove the second part of (ii), we use (35) to give the decomposition $\Phi^{(N)}_{t,j}-\Phi_j(t/N)=J_1+J_2$, 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

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

where we have used the bounds in Theorem 2.1 and (36) in the above. Using a similar argument (and Corollary 4.1 (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 3.3, and we omit the details. $\square$

### Acknowledgements

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). SSR research was partially supported by the National Science Foundation (grant DMS-1812128).

# A Proofs of remaining results in Sections 3.3 and 3.4

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

$$G_u(\omega) = \sum_{r\in\mathbb{Z}} C_r(u)\exp(ir\omega) \tag{37}$$

is bounded above and below by the $\lambda_{\text{sup}}$ and $\lambda_{\text{inf}}$ respectively (for all $\omega$). 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 $\lambda_{\text{sup}}$ and $\lambda_{\text{inf}}$ (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 $T_{u,N}$

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

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 $\boldsymbol{C}(u)$

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

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

$$G_{u,M}(\omega) = \frac{1}{M} \sum_{t,\tau=T_{u,N}-M/2+1}^{T_{u,N}+M/2} C_{t-\tau}(u) \exp(i(t-\tau)\omega) = (x_\omega \otimes I_p)^* \boldsymbol{C}_M(u) (x_\omega \otimes I_p), \quad (38)$$

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

$$G_{u,M}^{(N)}(\omega) = \frac{1}{M} \sum_{t,\tau=T_{u,N}-M/2+1}^{T_{u,N}+M/2} C_{t,\tau}^{(N)} \exp(i(t-\tau)\omega) = (x_\omega \otimes I_p)^* \boldsymbol{C}_{u,M}^{(N)} (x_\omega \otimes I_p). \quad (39)$$

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

$$\lambda_{\inf} \leq \|G_{u,M}^{(N)}(\omega)\|_2 \leq \lambda_{\sup} \text{ for all } N,\ M \text{ and } \omega. \quad (40)$$

By using Lemma B.2, equation (59) we have

$$\sup_\omega \|G_{u,M}(\omega) - G_{u,M}^{(N)}(\omega)\|_2 \leq \mathcal{K}\frac{M}{N}, \quad (41)$$

where $\mathcal{K}$ is a generic constant that depends only on $K$ and $\kappa$. 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 B.2, equation (60) we have $\sup_\omega \|G_u(\omega) - G_{u,M}(\omega)\|_2 \leq \mathcal{K}/M$. By using this and (41) 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 (40) 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 $\varepsilon > 0$ $\lambda_{\inf} - \varepsilon \leq \|G_u(\omega)\|_2 \leq \lambda_{\sup} + \varepsilon$, as required. $\square$

*Proof of Theorem 3.2.* To prove 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 the bound for the difference between the two matrices is small). 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 $N$). And the key observation is that the banded matrix embeds an infinite number of $(M+1)\times(M+1)$-dimensional block matrices, where each block matrix can be approximated by a stationary matrix. It can be shown that lower and upper bounds for the eigenvalues of the locally stationary approximation block matrix are given by $\gamma_{\inf}$ and $\gamma_{\sup}$. This approximation gives a bound for the eigenvalues of each block matrix. 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" 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 $M = 2m, m \in \mathbb{N}$. Using Lemma B.1 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) banded matrix

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

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) we have with Lemma B.4

$$\begin{aligned}
&\|\boldsymbol{C}^{(N)}(s-m,s+m)-\boldsymbol{C}(s-m,s+m;s/N)\|_2 \\
\leq\ & \sup_{t\in(s-m,s+m)}\sum_{\tau=s-m}^{s+m}\|C_{t,\tau}-C_{t-\tau}(s/N)\|_2 \leq \mathcal{K}\frac{m}{N},
\end{aligned} \tag{42}$$

where $\mathcal{K}$ is a generic constant that holds for all $N$ and $s$. The condition $0<\gamma_{\inf}\leq \inf_u\inf_\omega \lambda_{\min}(f(\omega;u)) \leq \sup_u\sup_\omega \lambda_{\max}(f(\omega;u)) \leq \gamma_{\sup}<\infty$ implies (see, among others, (Basu and Michailidis, 2015, Proposition 2.3)) that for all $u\in\mathbb{R}$ $\lambda_{\min}[\boldsymbol{C}(s-m,s+m;u)] \geq \inf_\omega \lambda_{\min}[f(\omega;u)] \geq \gamma_{\inf}$ and $\lambda_{\max}[\boldsymbol{C}(s-m,s+m;u)] \leq \inf_\omega \lambda_{\max}[f(\omega;u)] \leq \gamma_{\sup}$. Therefore by using (42) and the above we have

$$\left(\gamma_{\inf}-\mathcal{K}\frac{m}{N}\right)\|x_{s-m,s+m}\|_2 \leq x_{s-m,s+m}^\top \boldsymbol{C}^{(N)}(s-m,s+m)x_{s-m,s+m} \quad\leq\quad \left(\gamma_{\sup}+\mathcal{K}\frac{m}{N}\right)\|x_{s-m,s+m}\|_2. \tag{43}$$

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

$$x^\top \boldsymbol{C}_M^{(N)}x = \sum_{\ell\in\mathbb{Z}}\sum_{r=-M}^{M} x_\ell^\top C_{\ell,\ell+r}x_{\ell+r} \tag{44}$$

with the overlapping block matrix inner-product

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

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

$$X_M^\top \boldsymbol{O}_M X_M = \sum_{\ell\in\mathbb{Z}}\sum_{r=-M}^{M}\left(\frac{M+1-|r|}{M+1}\right)x_\ell^\top C_{\ell,\ell+r}x_{\ell+r}. \tag{45}$$

Using (44) and (45) 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

$$\begin{aligned}\left\|x^{\top}\boldsymbol{C}_M^{(N)}x - X_M^{\top}\boldsymbol{O}_M X_M\right\|_2 &\leq \frac{1}{M+1}\sum_{\ell\in\mathbb{Z}}\sum_{r=-m}^{m}\frac{|r|}{v(r)^{\kappa}}\|x_{\ell}\|_2\|x_{\ell+r}\|_2 \\ &\leq \frac{2}{M+1}\left(\sum_{r=1}^{\infty}\frac{1}{v(r)^{\kappa-1}}\right)\sum_{\ell\in\mathbb{Z}}\|x_{\ell}\|_2^2 = \frac{2}{M+1}\left(\sum_{r=1}^{\infty}\frac{1}{v(r)^{\kappa-1}}\right), \qquad (46)\end{aligned}$$

where the last line follows because $\|x\|_2 = \sum_{\ell\in\mathbb{Z}}\|x_{\ell}\|_2^2 = 1$. Finally we use (43) 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 (46), $\|\boldsymbol{C}^{(N)} - \boldsymbol{C}_M^{(N)}\|_2 \leq \mathcal{K}M^{-\kappa+1}$ 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. □

*Proof of Corollary 3.1.* We show the result follows from Theorem 3.2, Lemma 4.1, and Theorem 3.3 using the inverse matrix $\boldsymbol{D}^{(N)} = (\boldsymbol{C}^{(n)})^{-1}$ which has simple properties. 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

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

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 $\gamma_1$ and $\gamma_2$ that $0 < \gamma_1 \leq \inf_u \inf_\omega \lambda_{\min}(f(\omega;u)^{-1}) \leq \sup_u \sup_\omega \lambda_{\max}(f(\omega;u)^{-1}) \leq \gamma_2 < \infty$ and thus the eigenvalues of $\boldsymbol{D}(u)$ are uniformly bounded away from $\gamma_1$ and $\gamma_2$. Let $\boldsymbol{C}(u) = \boldsymbol{D}(u)^{-1} = (C_{t-\tau}; t,\tau \in \mathbb{Z})$. Then by using Lemma 4.1 we have

$$\sup_u \|C_r(u)\|_2 \leq \mathcal{K}\rho^{|r|} \tag{47}$$

for some $0 < \rho < 1$. Further, by using (11) (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

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

The Lipschitz conditions on $\Phi_j(\cdot)$ together with (47) and (48) 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 4.1. Further, the conditions in Theorem 3.3 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. $\square$

We now prove Theorem 3.5. 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 $[\boldsymbol{C}^{(e,f)}]_{t,\tau} = \mathbb{C}\mathrm{ov}[X_{t,N}^{(e)}, X_{t,N}^{(f)}]$. 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 $[\boldsymbol{C}^{(e,f)}(u)]_{t,\tau} = \mathbb{C}\mathrm{ov}[X_t^{(e)}(u), X_t^{(f)}(u)]$. This representation is instrumental in proving the result below.

*Proof of Theorem 3.5.* We first prove (23) and (24). We start by obtaining an expression for

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

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

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

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 $2 \times 2$, $2 \times (p-2)$ 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

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

and

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

Then, we have

$$\begin{aligned}\Delta_{t,\tau,N}^{\text{-}\{a,b\}} &= [\boldsymbol{A}^{(1,2)} - \boldsymbol{B}^{(1,2)}(\boldsymbol{E}^{(1,2)})^{-1}(\boldsymbol{B}^{(1,2)})^\top]_{t,\tau} \\ \text{and } \Delta_{t-\tau}^{\text{-}\{a,b\}}(u) &= [\boldsymbol{A}^{(1,2)}(u) - \boldsymbol{B}^{(1,2)}(u)(\boldsymbol{E}^{(1,2)}(u))^{-1}(\boldsymbol{B}^{(1,2)}(u))^\top]_{t,\tau}. \qquad (48)\end{aligned}$$

We use the above representations to prove (23). Using (48) we have

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

where

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

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 3.3). 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

$$
\begin{aligned}
J_3 &= \| \sum_{s_1,s_2\in\mathbb{Z}} [\boldsymbol{B}^{(1,2)}]_{t,s_1}[(\boldsymbol{E}^{(1,2)})^{-1} - (\boldsymbol{E}^{(1,2)}(t/N))^{-1}]_{s_1,s_2}[\boldsymbol{B}^{(1,2)}(t/N))^{\top}]_{s_2,\tau}\|_2 \\
&\leq \sum_{s_1,s_2\in\mathbb{Z}} \|[\boldsymbol{B}^{(1,2)}]_{t,s_1}\|_2 \cdot \|[(\boldsymbol{E}^{(1,2)})^{-1} - (\boldsymbol{E}^{(1,2)}(t/N))^{-1}]_{s_1,s_2}\|_2 \cdot \|(\boldsymbol{B}^{(1,2)}(t/N))^{\top}]_{s_2,\tau}\|_2
\end{aligned} \tag{49}
$$

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

$$
\|[\boldsymbol{B}^{(1,2)}]_{t,s_1}\|_2 \leq Kv(t-s_1)^{-\kappa} \text{ and } \|(\boldsymbol{B}^{(1,2)}(t/N))^{\top}]_{s_2,\tau}\|_2 \leq Kv(s_2-\tau)^{-\kappa}. \tag{50}
$$

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 3.3 to $(\boldsymbol{E}^{(1,2)})^{-1}$ and this gives

$$
\|((\boldsymbol{E}^{(N),(1,2)})^{-1} - (\boldsymbol{E}^{(1,2)}(s_1/N))^{-1})_{s_1,s_2}\|_2 \leq \mathcal{K}\zeta(s_1-s_2)^{\kappa-2}\min(1/N, 2\zeta(s_1-s_2)) \tag{51}
$$

and

$$
\|((\boldsymbol{E}^{(1,2)}(s_1/N))^{-1} - (\boldsymbol{E}^{(1,2)}(t/N))^{-1})_{s_1,s_2}\|_2 \leq \mathcal{K}|s_1-t|\zeta(s_1-s_2)^{\kappa-1}/N. \tag{52}
$$

Substituting (50), (51) and (52) into (49) 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 B.5.

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 (50), 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

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

thus proving (23). The proof of (24) follows a similar technique.

Finally, the proofs for (25) and (26) are the same as the proofs for (23) and (24), thus we omit the details. $\square$

# B Technical lemmas

## B.1 Proofs and lemmas for the proof of Theorem 2.1

*Proof of Lemma 4.1.* The proof is based on the proof of Proposition 2.2 in Demko et al. (1984), with a small modification to allow block matrices. We use the notation from Proposition 2.2 in Demko et al. (1984). More precisely, let $\Pi_n$ denote the space of polynomials up to order $n$. A key ingredient in the proof is the following result from spectral theory

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

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 $\|1/x - p(x)\|_{[a,b]} = \max_{x\in[a,b]}|1/x - p(x)|$). In

Proposition 2.1, Demko et al. (1984) show that

$$\inf\{\|1/x - p(x)\|_{[a,b]} : p \in \Pi_n\} = \frac{(1+\sqrt{r})^2}{b}\rho^{n+1}. \tag{53}$$

Using this, we define the polynomial

$$p_n^* = \arg_{p\in\Pi_n} \inf\{\|1/x - p(x)\|_{[a,b]} : p \in \Pi_n\}. \tag{54}$$

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 $(t,\tau)$ $p\times p$ dimension block matrix in $p_n(\boldsymbol{A})$.

For a given $t$ and $\tau$, set $n = \lfloor |t-\tau|/M \rfloor$. Let $p_n^*$ be defined as in (54). 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

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

where the last part follows from (53). This completes the proof of the first assertion.

For the second assertion we slightly modify $\widetilde{\boldsymbol{A}}$, invert it and link the modified matrix to the inverse of $\widetilde{\boldsymbol{A}}$. Since $\widetilde{\boldsymbol{A}}$ is missing a row and column, the idea is to extend this matrix to its original dimension. For this, note first that $\boldsymbol{D} := (I_{-k}\otimes I_p)\widetilde{\boldsymbol{A}}(I_{-k}\otimes I_p)^\top + c(e_k \otimes I_p)(e_k \otimes I_p)^\top$ is a block-banded matrix and with $c = \|(e_k\otimes I_p)^\top \boldsymbol{A}(e_k\otimes I_p)\|_2$ its largest and smallest eigenvalues are bounded by $b$ and $a$. Hence, the previous assertion applies to $\boldsymbol{D}$. Let $\boldsymbol{I}$ be the identity operator on $\ell_{2,p}$. Then, we have $(I_{-k}\otimes I_p)(I_{-k}\otimes I_p)^\top$ is $\boldsymbol{I}$ without the $k$th $p$-dimensional row/column. This implies $\|(I_{-k}\otimes I_p)(I_{-k}\otimes I_p)^\top \boldsymbol{D}(I_{-k}\otimes I_p)(I_{-k}\otimes I_p)^\top)_{t,\tau}\|_2 \leq \|(\boldsymbol{D})_{t,\tau}\|_2$. Additionally, we have the following $(I_{-k}\otimes I_p)^\top(I_{-k}\otimes I_p)$ is the identity operator on the reduced space, $(I_{-k}\otimes I_p)(I_{-k}\otimes I_p)^\top + (e_k\otimes I_p)(e_k\otimes I_p)^\top = \boldsymbol{I}$, and $(e_k\otimes I_p)^\top(I_{-k}\otimes I_p) = 0$. 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.

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

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. $\square$

We now apply the above result to $\boldsymbol{B}_M$ defined in (31).

**Lemma B.1.** *[Properties of $\boldsymbol{B}_M$] Suppose Assumption 2.1 is satisfied and let $\boldsymbol{B}_M$ be a banded matrix defined as in (31). 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*

$$\|C_{\mathcal{S},\mathcal{S}} - \boldsymbol{B}_M\|_2 \leq 2K/(\kappa - 1)(M-1)^{-\kappa+1}, \tag{55}$$

$$a_M \geq \lambda_{\min} - 2K/(\kappa-1)(M-1)^{-\kappa+1} \text{ and } b_M \leq \lambda_{\max} + 2K/(\kappa-1)(M-1)^{-\kappa+1} \tag{56}$$

$$\|\boldsymbol{B}_M^{-1}\|_2 \leq (\lambda_{\min} - 2K/(\kappa-1)(M-1)^{-\kappa+1})^{-1} \tag{57}$$

*The same rates apply also if $\boldsymbol{C}$ itself or $\boldsymbol{C}(-\infty; T]$ are approximated by (a corresponding) banded matrix $\boldsymbol{B}_M$.*

*Proof.* We first prove (55). For this, we first expand $C_{\mathcal{S},\mathcal{S}} - \boldsymbol{B}_M$ with zero such that it is an operator from $\ell_{2,p}$ to $\ell_{2,p}$ again. Then, we use Lemma B.4 and obtain

$$\begin{aligned}\|C_{\mathcal{S},\mathcal{S}} - \boldsymbol{B}_M\|_2 =& \|(I_{-k} \otimes I_p)(C_{\mathcal{S},\mathcal{S}} - \boldsymbol{B}_M)(I_{-k} \otimes I_p)^\top\|_2 \leq \sup_{s_1} \sum_{s_2} \|(C_{\mathcal{S},\mathcal{S}} - \boldsymbol{B}_M)_{s_1,s_2}\|_2 \\ \leq& \sum_{|s|>M} \frac{K}{|s|^{-\kappa}} \leq 2K \sum_{s>M} \int_{s-1}^{s} x^{-\kappa} dx = 2K/(\kappa-1)(M-1)^{-\kappa+1}.\end{aligned}$$

To prove (56) we use that $\boldsymbol{B}_M = C_{\mathcal{S},\mathcal{S}} + (C_{\mathcal{S},\mathcal{S}} - \boldsymbol{B}_M)$ and the eigenvalues of $C_{\mathcal{S},\mathcal{S}}$ are in $[\lambda_{\min}, \lambda_{\max}]$. Thus, with (55) we have

$$\lambda_{\inf}(\boldsymbol{B}_M) \geq \lambda_{\min} - 2K/(\kappa-1)(M-1)^{-\kappa+1} \text{ and } \lambda_{\sup}(\boldsymbol{B}_M) \leq \lambda_{\max} + 2K/(\kappa-1)(M-1)^{-\kappa+1} \tag{58}$$

The proof of (57) immediately follows from (56). $\square$

## B.2 Technical lemmas for the proof of results in Section 3

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

**Lemma B.2.** *Suppose Assumption 3.1 holds and let $G_{u,M}(\omega)$, $G_{u,M}^{(N)}(\omega)$ and $G_u(\omega)$ be defined as in (38), (39) and (37) 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 $\kappa$.*

*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 (59). To prove (60) 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. □

**Lemma B.3.** *Suppose that $\{A_\ell\}_{\ell=1}^{\infty}$ is a sequence of $p \times p$ dimensional matrices where $\sum_{\ell=1}^{\infty} \|A_\ell\|_2^2 < \infty$. Define the sequence space $\ell_{2,p,1} = \{w = (v_1, v_2, \ldots) : v_j \in \mathbb{R}^p\}$ and the linear operator $\boldsymbol{A} = (A_\ell; \ell \geq 0)$, where $\boldsymbol{A} : \ell_{2,p,1} \to \mathbb{R}$. 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.1 and Theorem 3.1.

**Lemma B.4.** *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 $B$. Let $\widetilde{\boldsymbol{B}} = (\|B_{s_1,s_2}\|_2)_{s_1,s_2}$ be an operator from $\ell_2$ to $\ell_2$. Since $B$ 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 used in the proofs of Theorems 3.3, 3.4 and 3.5.

**Lemma B.5.** *Let $v(j) = \max(1, |j|)$ and $\zeta(j) = v(\log[v(j)])/v(j)$ We have for some $y \in \mathbb{R}, p \in \mathbb{Z}$ and $p \geq 2$*

$$\sum_{j\in\mathbb{Z}} v(j)^{-p} v(j+y)^{-p} \leq (\pi^2 + 3) v(y-1)^{-p} \tag{61}$$

*and*

$$\sum_{j\in\mathbb{Z}} \zeta(j)^p \zeta(j+y)^p \le 7\zeta(y-3)^p \tag{62}$$

*Further, suppose that $p, q, r \ge 2$ then*

$$\sum_{j\in\mathbb{Z}} v(j)^{-q} v(j+y)^{-p} \le (\pi^2+3) v(y-1)^{-\min(p,q)}, \tag{63}$$

$$\sum_{j\in\mathbb{Z}} \zeta(j)^p \zeta(j+y)^q \le 7\zeta(y-3)^{\min(p,q)}, \tag{64}$$

$$\sum_{s_1,s_2\in\mathbb{Z}} v(s_1+t)^{-p} v(s_1+s_2)^{-q} v(s_2+\tau)^{-r} \le (\pi^2+3)^2 v(t-\tau-2)^{-\min(p,q,r)}, \tag{65}$$

*and*

$$\sum_{s_1,s_2\in\mathbb{Z}} \zeta(s_1+t)^p \zeta(s_1+s_2)^q \zeta(s_2+\tau)^{-r} \le 49\zeta(t-\tau-6)^{\min(p,q,r)} \tag{66}$$

*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 (61). Without loss of generality, let $y > 0$ and $y \in \mathbb{N}$. We have

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

where

$$I_1 = \sum_{j=0}^{\infty} v(j)^{-p} v(j+y)^{-p} \le (\pi^2/6+1) v(y)^{-p},$$

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

$$
I_3 = \sum_{j=-\infty}^{-y} v(j)^{-p} v(j+y)^{-p} \leq (\pi^2/6 + 1)v(y)^{-p}.
$$

Putting the bounds for $I_1 - I_3$ immediately proves (61). Furthermore, we have $v(|y|-1) \geq v(|y|/2)$. This follows immediately for $|y| \geq 2$. For $|y| < 2$ note that $v(|y|-1) = 1 = v(|y|/2)$.

To proof (62), note first that $\sum_{k=1}^{\infty} \zeta(k)^2 = 1 + \sum_{k=2}^{\infty} \zeta(k)^2 \leq 1 + \int_1^{\infty} (\log(x)/x)^2 dx = 1 + 2$. Second note that $\zeta(\cdot)$ is monotonic decreasing after $\zeta(3)$, and we have $1 = \zeta(1)$, $\zeta(2) = \zeta(4) < \zeta(3)$. With this, we can follow the arguments as above and split the sum up into three parts $I_1 + I_2 + I_3$. Wlog let $y \geq 9$. We have

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

Furthermore, since $\zeta(4) \leq 0.5$ and for $y \geq 9, p \geq 2$ it holds $\zeta(y-3)^p + \zeta(y/2)^{2p}(y/2) \leq \zeta(y)^p$, we have

$$
\begin{aligned}
I_2 &= \sum_{j=-y}^{-1} \zeta(j)^p \zeta(j+y)^p = 2\zeta(y-1)^p + \zeta(y-2)^p + 2\zeta(y-3)^p \\
&+ \sum_{j=-y+2}^{-y/2} \zeta(j)^p \zeta(j+y)^p + \sum_{j=-y/2+1}^{-4} \zeta(j)^p \zeta(j+y)^p \\
&\leq 2\zeta(y-1)^p + \zeta(y-2)^p + \zeta(y-3)^p + \zeta(y/2)^{2p}(y/2) + \zeta(y/2)^p \zeta(4)^{p-2} \sum_{j=4}^{\infty} \zeta(j)^2 \\
&\leq 7\zeta(y-3)^p
\end{aligned}
$$

and

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

The proof of (63) uses that $v(j)^{-p} > v(j)^{-q}$, then the result immediately follows from (61).

To prove of (65), let us suppose wlog that $p \leq q \leq r$, then by using (63) 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 (63). This proves the result. (64) and (66) follow analogously. □

# References


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

Basu, S. and Subba Rao, S. (2021). Graphical models for nonstationary time series. *arXiv preprint arXiv:2109.08709*.

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. (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*.

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.

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

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