



University of Essex

Department of Economics

Discussion Paper Series

No. 767 June 2015

Autoregressive Spatial Spectral Estimates

Abhimanyu Gupta

Note : The Discussion Papers in this series are prepared by members of the Department of Economics, University of Essex, for private circulation to interested readers. They often represent preliminary reports on work in progress and should therefore be neither quoted nor referred to in published work without the written consent of the author.

Autoregressive spatial spectral estimates

Abhimanyu Gupta*
University of Essex
Wivenhoe Park
Colchester CO4 3SQ, UK

June 11, 2015

Abstract

Autoregressive spectral density estimation for stationary random fields on a regular spatial lattice has many advantages relative to kernel based methods. It provides a guaranteed positive-definite estimate even when suitable edge-effect correction is employed, is simple to compute using least squares and necessitates no choice of kernel. We truncate a true half-plane infinite autoregressive representation to estimate the spectral density. The truncation length is allowed to diverge in all dimensions in order to avoid the potential bias which would accrue due to truncation at a fixed lag-length. Consistency and strong consistency of the proposed estimator, both uniform in frequencies, are established. Under suitable conditions the asymptotic distribution of the estimate is shown to be zero-mean normal and independent at fixed distinct frequencies, mirroring the behaviour for time series. A small Monte Carlo experiment examines finite sample performance. We illustrate the technique by applying it to Los Angeles house price data and a novel analysis of voter turnout data in a US presidential election. Technically the key to the results is the covariance structure of stationary random fields defined on regularly spaced lattices. We study this in detail and show the covariance matrix to satisfy a generalization of the Toeplitz property familiar from time series analysis.

JEL classifications: C14, C18, C21

Keywords: Spectral density estimation, spectral analysis, random field, covariance matrix, spatial process, strong consistency, lattice data, central limit theorem

* *Email:* a.gupta@essex.ac.uk, *Telephone:* +44-1206-872597

1 Introduction

We are concerned with nonparametric estimation of the spectral density of a zero-mean stationary scalar random field x_t , $t = (t_1, \dots, t_d)'$ with $t_j \in \mathbb{Z}$, $j = 1, \dots, d$, where \mathbb{Z} denotes the set of integers, using an autoregressive technique. Such data may be found in environmental, agricultural, regional and urban economics settings, and are likely to become more prevalent with the rapid advances in remote sensing and GIS software capabilities. The analysis of spatial data has seen a great deal of recent econometric work. In particular, models for lattice data have attracted interest, but primarily in the spatial domain. [Robinson \(2008\)](#) considers tests of spatial correlation for lattice data among a host of other settings while [Hidalgo \(2009\)](#) considers testing for correct parametric covariogram specification. [Roknossadati and Zarepour \(2010\)](#) provide theory for M -estimation (in a class of unilateral models) and [Hidalgo and Seo \(2014\)](#) propose omnibus-type specification tests. [Jenish \(2014\)](#) also considers a nonlinear autoregressive model on a regular lattice as a motivating example in her analysis of a spatial semiparametric model. The natural analogies between lattice and time series data suggests a more central role for frequency domain analysis. The well established study of cycles in time series data via the frequency domain has spatial counterparts. High frequency spatial components may be interpreted as corresponding to phenomena (possibly noise phenomena) that change rapidly over the space, while low frequency components that change less frequently are more structural. For Tokyo land price data [Matsuda and Yajima \(2009\)](#) argue that accurately estimating the spectrum over low frequencies is more desirable than over high frequencies, interpreting the latter as noise and the former as the structural factors of interest. In this context they specify that high frequency noise can include environmental factors, air and noise pollution and sunshine. In this paper we analyse an example with house price data from Los Angeles, and also voting turnout data for US counties in a presidential election. In the second setting high frequency components affecting turnout include weather and demographics which will change very frequently over space. Low frequency components include voting laws, type of election and closeness of the race which can be easily seen to not change very frequently over space. Thus the spectral density can be informative about the power of the low and high frequency components, which will be reflected in the strengths of the peaks.

Recent statistical contributions for lattice data include [Robinson \(2012\)](#), [McElroy and Holan \(2014\)](#) and [Abramovich and Lahav \(2015\)](#), to name a few. Frequency domain techniques are commonly employed and irregular spatial lattices also considered (cf. [Matsuda and Yajima \(2009\)](#), [Bandyopadhyay et al. \(2015\)](#)), but we focus on the regular case. Irregular spacing will disturb the Toeplitz property of the covariance matrix that we exploit for our results, but a more practical reason suggests itself. Many economic data sets can be gridded into cells and the analysis of properties carried out as if the data is observed on a regular lattice of size determined by the number of

grid cells, thereby avoiding many of the problems with irregular data summarized in [Bandyopadhyay et al. \(2015\)](#). [Bronars and Jansen \(1987\)](#) use this method to build a spatial model of unemployment rates in US counties. Our empirical example of presidential election voter turnout data across US counties is in this spirit, but we also grid data in our house price example. [Chen and Nordhaus \(2011\)](#) use regular gridded measures of nighttime lights visible from space as a proxy for economic statistics in countries where such data may not be reliably collected. Statistics Finland collects data by map grid for the whole of Finland, from $250\text{m} \times 250\text{m}$ to $5\text{km} \times 5\text{km}$ cells, while the Geographically based Economic data project seeks to expand gridding globally, cf. http://www.stat.fi/tup/ruututietokanta/index_en.html, <http://gecon.yale.edu/> and [Nordhaus \(2008\)](#). Examples of gridding irregularly spaced data also arise in statistics, cf. [Fuentes \(2007\)](#) and references therein.

Nonparametric spectral estimates for spatial data have typically focused on tapered autocovariance or periodogram based techniques, see e.g. [Yuan and Subba Rao \(1993\)](#), [Politis and Romano \(1996\)](#), [Robinson \(2007\)](#) and [Vidal Sanz \(2009\)](#). Our autoregressive approach allows us to consider nonparametric estimates of the spectral density without the practitioner having to choose a taper or kernel. She simply needs to fit autoregressive models by least squares. For lattice processes autoregressive estimation has another advantage, connected with the edge-effect, which matters when $d = 2$ and worsens with increasing d (cf. [Section 2](#)). [Guyon \(1982\)](#) suggested a version of the covariance estimates which eliminates the bias (asymptotically), but this was criticised by [Dahlhaus and Künsch \(1987\)](#) as it could yield possible negative kernel based spectral density estimates. The latter suggested tapering the covariance estimates, but introduced ambiguity arising from the choice of an appropriate taper. [Robinson and Vidal Sanz \(2006\)](#) propose an alternative, but again there is an element of ambiguity due to the practitioner having to choose a function. On the other hand, autoregressive spectral estimation delivers a guaranteed non-negative estimate even when using edge-effect correction.

For the case of regularly-spaced time series ($d = 1$), [Berk \(1974\)](#) assumes an infinite, one-sided autoregressive representation for x_t , driven by white noise, and provides results on the consistency and asymptotic normality of spectral density estimates with the order of the autoregression allowed to diverge with sample size. We seek to extend this approach to spatial processes. There is some related work in the signal processing literature, see e.g. [Tjøstheim \(1981\)](#), [McClellan \(1982\)](#) and [Wester et al. \(1990\)](#), but under the assumption that the true model is finite, which is a parametric approach that may lead to bias.

The results in this paper overcome two technical hurdles that arise in the transition from $d = 1$ to $d > 1$: the structure of the covariance matrix of stationary spatial process and the number of unique covariances that occur in such a matrix. For the benefit of readers primarily interested in applying the techniques, we treat these hurdles in appendices. Readers interested in the technical details may refer to the appendices.

We also mention here that the asymptotic normality result established by us serves to stress that the difference between the time series and spatial cases is not merely that of extension. The sufficient condition restricting the growth rate of the parameter space when $d = 1$ cannot be regarded as simply a particular case of our theorem for $d > 1$, as we discuss in detail in Section 4.

The paper is structured as follows: Section 2 contains some preliminary results used throughout the paper and theorems on consistency and strong consistency of the truncated AR predictors. Section 3 introduces the spectral density estimate and establishes its uniform consistency and uniform strong consistency. Section 4 records results on the asymptotic distribution of the truncated AR predictors as well as the spectral estimate. Section 5 contains a small Monte Carlo study of finite sample performance, also comparing our estimates with periodogram based ones. Section 6 applies our techniques to two economic data sets. In Appendix A we derive bounds for absolute moments of partial sums of rather general lattice processes, while Appendix B demonstrates that when the spatial process is stationary and has a finite half-plane AR representation the covariance structure satisfies a generalisation of the Toeplitz property familiar from the theory of stationary time series. It also provides an upper bound on the number of unique autocovariances that occur in the covariance matrix of finite, stationary and unilateral processes. Proofs of theorems and lemmas are contained in Appendices C and D respectively.

2 Consistency of truncated AR predictors

Whittle (1954) observed that the estimation of the parameters of multilateral autoregressive processes by least squares leads to inconsistency. This is due to the presence in the likelihood function of a Jacobean term which depends on the parameters. A representation on a ‘half-plane’ permits least squares estimation, however, while in general Whittle likelihood based estimates lack a closed form. He showed, quite generally, that multilateral spatial processes have a (possibly infinite) unilateral representation. Helson and Lowdenslager (1958, 1961) showed that even more generally all stationary, purely non-deterministic spatial processes have a half-plane (i.e. unilateral), infinite, moving-average representation. Whittle (1954) points out that the recovery of the parameters of the original multilateral scheme from the unilateral representation is not as straightforward as with, say, a bilateral $d = 1$ model, indeed even impossible. On the other hand, the unilateral representation is extremely useful if our interest is in prediction or spectral density estimation. As in Tjøstheim (1983) we define the half-plane as all t in the set

$$S_{1+}^{\infty} = \{t_1 > 0; t_1 = 0, t_2 > 0; \dots; t_1 = \dots = t_{d-1} = 0, t_d > 0\} \cap \mathbb{Z}^d. \quad (2.1)$$

The special case with $t_i \geq 0$, $i = 1, \dots, d$, is referred to as a quarter-plane. Write

$z = (z_1, \dots, z_d)'$ with complex-valued elements and $s = (s_1, \dots, s_d)'$ with integer-valued elements, and $z^s = \prod_{j=1}^d z_j^{s_j}$. Define the rational function (see [Rosenblatt \(1985\)](#), p. 228) $B(z) = \sum_{s \in S_{1+}^\infty \cup \mathbf{0}} b_s z^s$, with $\mathbf{0}$ the d -dimensional zero vector. Then we assume

Assumption A. There exist unknown scalars b_s and iid random variables ϵ_t , $t \in \mathbb{Z}^d$ with $E\epsilon_t = 0$ and $E\epsilon_t^2 = \sigma^2$ such that

$$x_t = \sum_{s \in S_{1+}^\infty \cup \mathbf{0}} b_s \epsilon_{t-s}, \quad \sum_{s \in S_{1+}^\infty \cup \mathbf{0}} |b_s| < \infty, \quad b_{\mathbf{0}} \neq 0. \quad (2.2)$$

$B(z)$ is bounded away from zero for $|z_i| = 1$, $i = 1, \dots, d$.

Martingale assumptions can replace the iid imposition on ϵ_t , but we choose to avoid these as they rely on notions of ordering that can be arbitrary. Writing $\Pi = (-\pi, \pi]^d$, denote by $f(\lambda)$ the spectral density of x_t , $\lambda \in \Pi$. If $\int_\Pi \log f(\lambda) d\lambda > -\infty$, then, e.g., [Helson and Lowdenslager \(1958\)](#) and [Korezlioglu and Loubaton \(1986\)](#) prove that Assumption [A](#) will hold, extending the Wold decomposition of time series analysis. Under Assumption [A](#)

$$f(\lambda) = \frac{\sigma^2}{(2\pi)^d} \left| \sum_{s \in S_{1+}^\infty \cup \mathbf{0}} b_s e^{i\lambda' s} \right|^2, \quad \lambda \in \Pi. \quad (2.3)$$

$\left| \sum_{s \in S_{1+}^\infty \cup \mathbf{0}} b_s e^{i\lambda' s} \right|$ being bounded and bounded away from zero guarantees the invertibility of x_t i.e. the existence of d_s , $s \in S_{1+}^\infty$, such that

$$x_t = \sum_{s \in S_{1+}^\infty} d_s x_{t-s} + \epsilon_t, \quad t \in \mathbb{Z}^d, \quad \sum_{s \in S_{1+}^\infty} |d_s| < \infty. \quad (2.4)$$

Writing $D(z) = 1 - \sum_{s \in S_{1+}^\infty} d_s z^s$, Wiener's Lemma ([Rudin \(1973\)](#) p. 266), implies that if $D(z) \neq 0$ for $|z_i| = 1$, $i = 1, \dots, d$, and $B(z)$ is bounded away from zero for $|z_i| = 1$, $i = 1, \dots, d$, then $\sum_{s \in S_{1+}^\infty \cup \mathbf{0}} |b_s| < \infty$, implying a regularity condition on $f(\lambda)$. By Assumption [A](#) and (2.3) there exist real numbers m, M satisfying $0 < m \leq M < \infty$, such that

$$m \leq f(\lambda) \leq M. \quad (2.5)$$

As x_t is stationary, we define the autocovariances $\gamma(k) = Ex_t x_{t+k}$ with $t, k \in \mathbb{Z}^d$.

Lemma 2.1. *Suppose Assumption [A](#) holds. Then $\sum_{k \in \mathbb{Z}^d} |\gamma(k)| < \infty$.*

We denote by C a positive, arbitrarily large but finite generic constant, independent of N .

Assumption B. For some $v \in (1, 2]$, $E|\epsilon_t|^{2v} \leq C$ for all $t \in \mathbb{Z}^d$.

Expressing the moment condition in terms of v delivers conditions restricting the rate of growth of the parameter space relative to sample size that become more stringent as $v \rightarrow 1$. We observe x_t on the rectangular lattice $\mathbb{L} = \{t \in \mathbb{Z}^d : -n_{L_i} \leq t_i \leq n_{U_i}, i = 1, \dots, d\}$, $n_{U_i}, n_{L_i} \geq 0$, $i = 1, \dots, d$. Define $n_i = n_{L_i} + n_{U_i} + 1$, $i = 1, \dots, d$, and $N = \prod_{i=1}^d n_i$.

Consistency in our setting is only possible if sample size increases in all directions, and mild regularity in this increase is implied by

Assumption C. For each n_{L_i} , $i = 2, \dots, d$, and n_{U_i} , $i = 2, \dots, d$, and sufficiently large N , there exists $\chi > 0$ and $c_1 > 0$ such that

$$n_{U_1}(N) \geq c_1 N^\chi, \quad n_i(N) \geq c_1 N^\chi. \quad (2.6)$$

Robinson and Vidal Sanz (2006) point out that $\chi \leq 1/d$ always. We will first obtain a least squares predictor for x_t based on a truncated autoregression of order $p = (p_{L_1}, p_{U_1}; \dots; p_{L_d}, p_{U_d})$, for non-negative integers p_{L_i} , p_{U_i} , $i = 1, \dots, d$. In view of the half-plane representation we can *a priori* set, say, $p_{L_1} = 0$. Now define

$$S[-p_L, p_U] = \{t \in \mathbb{L} : -p_{L_i} \leq t_i \leq p_{U_i}, i = 1, \dots, d\} \cap S_{1+}^\infty, \quad (2.7)$$

which is the truncated set of dependence ‘lags’. Denote $p_i = p_{L_i} + p_{U_i}$, $i = 1, \dots, d$. Let $\mathfrak{h}(p)$ denote the total number of autoregressive parameters to be estimated in the truncated predictor. Then

$$\mathfrak{h}(p) = p_{U_d} + \sum_{j=1}^{d-1} \prod_{i=j+1}^d (p_i + 1) p_{U_j}. \quad (2.8)$$

Our asymptotic theory consists of finding divergent (as $N \rightarrow \infty$) functions $p_{L_i} = p_{L_i}(N)$, $p_{U_i} = p_{U_i}(N)$, $i = 1, \dots, d$ such that we can consistently approximate the infinite representation with truncated predictors. Thus $n_i(N) \geq c_1 N^\chi$ in Assumption C is taken to hold as both n_{L_i} and n_{U_i} diverge with N . We emphasize the dependence of the orders on N , but for notational convenience suppress explicit reference to this. The practitioner may prefer to choose only one truncation length for each dimension. In this case $p_{L_i} = p_{U_i} = p_{U_1} = p^*$, $i = 2, \dots, d$, and (2.8) indicates that $\mathfrak{h}(p^*) = ((2p^* + 1)^d - 1)/2$. A more flexible approach to modelling can be to choose a divergent sequence \bar{p} (dependent on N , and diverging slower than N) and take $p_{L_i} = p_{U_i} = p_{U_1}$ to be the sequence $[(n_i/N) \bar{p}]$, $i = 2, \dots, d$, where $[x]$ denotes the integer part of x .

Write $n_s = \prod_{i=1}^d (n_i - s_i)$ for non-negative integers s_i , $i = 1, \dots, d$ and introduce the covariance estimates

$$\hat{\gamma}(k) = n_p^{-1} \sum_{t(p,n)}'' x_t x_{t+k}, \quad k \in S[-p_L, p_U],$$

where it is assumed that $n_i > p_i \geq 0$ for $i = 1, \dots, d$ and the sum $\sum_{t(p,n)}''$ is defined analogously to Section A with respect to n and p (see (A.3) for range of summation). It contains n_p terms by definition. The estimates $\hat{\gamma}(k)$ incorporate the device for edge-effect correction suggested by Guyon (1982). Consider instead the estimates $\tilde{\gamma}(k) = N^{-1} \sum_{t(|k|,n)}'' x_t x_{t+k}$, where $|k| = (|k_1|, \dots, |k_d|)'$. Then for fixed k , as the $n_i \rightarrow \infty$,

the bias of $\tilde{\gamma}(k)$ for $\gamma(k)$ is of order $\sum_{i=1}^d n_i^{-1}$. The inequality between arithmetic and geometric means indicates that $\sum_{i=1}^d n_i^{-1} \geq dN^{-\frac{1}{d}}$ with equality implying that the n_i all increase at the same, $N^{\frac{1}{d}}$, rate. This inequality implies that the bias of $\tilde{\gamma}(k)$ is of order no less than $N^{-\frac{1}{d}}$. It is clear that this worsens with increasing d , but for $d = 1$ gives the usual ‘parametric’ rate of bias. We assumed that x_t has zero mean, but this may be relaxed to $Ex_t = \alpha$, $t \in \mathbb{Z}^d$. In this case lag k covariance estimates can be $\gamma^*(k) = n_p^{-1} \sum_{t \in \mathbb{Z}^d}'' (x_t - \bar{x})(x_{t+k} - \bar{x})$, where $\bar{x} = N^{-1} \sum_{t \in \mathbb{Z}^d} x_t$, and the latter is readily shown to be $N^{\frac{1}{2}}$ -consistent for α .

Lemma 2.2. *With $n = (n_1, n_2, \dots, n_d)'$, for such positive integers n_i and integers k_i that satisfy $n_i > |k_i|$ for $i = 1, \dots, d$, let*

$$S_{kn} = n_{|k|}^{-1} \sum_{t \in \mathbb{Z}^d}'' u_t, \quad u_t = \sum_{r \in \mathbb{Z}^d} \sum_{s \in \mathbb{Z}^d} \xi_{rs,t}, \quad t \in \mathbb{L}, \quad (2.9)$$

with the $\xi_{rs,t}$ zero mean, independent (over $t \in \mathbb{L}$) random variables. For some $w' \in (1, 2]$, suppose there exist $\eta_{1,r}, \eta_{2,r}$, $r \in \mathbb{Z}^d$, such that

$$E |\xi_{rs,t}|^{w'} \leq |\eta_{1,r} \eta_{2,s}|^{w'}, \quad \sum_{r \in \mathbb{Z}^d} |\eta_{j,r}| < \infty, \quad j = 1, 2, \quad (2.10)$$

for all $r, s \in \mathbb{Z}^d$ and $t \in \mathbb{L}$. Then $E |S_{kn}|^{w'} \leq C n_{|k|}^{1-w'}$.

Lemma 2.3. *If Assumptions A and B hold, $E |\hat{\gamma}(k) - \gamma(k)|^v \leq C n_p^{1-v}$.*

Denote by $\psi_{\mathfrak{h}(p)}$ ($\Psi_{\mathfrak{h}(p)}$) the $\mathfrak{h}(p) \times 1$ vector ($\mathfrak{h}(p) \times \mathfrak{h}(p)$ matrix) with typical element $\gamma(k)$ ($\gamma(k - j)$), $j, k \in S[-p_L, p_U]$, and by $\hat{\psi}_{\mathfrak{h}(p)}$ ($\hat{\Psi}_{\mathfrak{h}(p)}$) the $\mathfrak{h}(p) \times 1$ vector ($\mathfrak{h}(p) \times \mathfrak{h}(p)$ matrix) constructed in exactly the same way but using $\hat{\gamma}(k)$ in place of $\gamma(k)$. Also write $\Delta_{\mathfrak{C}(p)} = \hat{\Psi}_{\mathfrak{h}(p)} - \Psi_{\mathfrak{h}(p)}$ and $\delta_{\mathfrak{h}(p)} = \hat{\psi}_{\mathfrak{h}(p)} - \psi_{\mathfrak{h}(p)}$, with $\mathfrak{C}(p)$ denoting an upper bound for the number of unique covariances in $\Psi_{\mathfrak{h}(p)}$ (see Proposition B.1). For a generic rectangular matrix B , we will denote by $\|B\|_R$ and $\|B\|$ the largest absolute row-sum of B and square root of the largest eigenvalue of $B'B$ respectively.

Lemma 2.4. *If Assumptions A and B hold, $E \|\delta_{\mathfrak{h}(p)}\|^v \leq C \mathfrak{h}(p)^v n_p^{1-v}$.*

Lemma 2.5. *If Assumptions A and B hold, $E \|\Delta_{\mathfrak{C}(p)}\|^v \leq C \mathfrak{C}(p)^v n_p^{1-v}$.*

The $d > 1$ case differs from $d = 1$ in the number of unique covariances in $\Psi_{\mathfrak{h}(p)}$, these numbering $\mathfrak{h}(p)$ in the latter case but at most $\mathfrak{C}(p) \geq \mathfrak{h}(p)$ in the former. More details are in Appendix B.2.

Lemma 2.6. *Let ρ be any eigenvalue of $\Psi_{\mathfrak{h}(p)}$. Then, under Assumption A, $(2\pi)^d m \leq \rho \leq (2\pi)^d M$.*

This lemma is a $d > 1$ generalization of the statement in Grenander and Szegö (1984), p. 64.

Corollary 2.7. *Under the conditions of Lemma 2.6, $\|\Psi_{\mathfrak{h}(p)}^{-1}\| \leq C$.*

Writing $d_{\mathfrak{h}(p)}$ for the $\mathfrak{h}(p) \times 1$ vector with elements d_s , $s \in S[-p_L, p_U]$, we identify

$$d_{\mathfrak{h}(p)} = \Psi_{\mathfrak{h}(p)}^{-1} \psi_{\mathfrak{h}(p)}. \quad (2.11)$$

For n_i and p_i satisfying $n_i > p_i$, $i = 1, \dots, d$, define a least squares predictor of order $\mathfrak{h}(p)$ by $\hat{d}_{\mathfrak{h}(p)} = \arg \min_{a_s, s \in S[-p_L, p_U]} n_p^{-1} \sum_{t(p, n)} \left(x_t - \sum_{s \in S[-p_L, p_U]} a_s x_{t-s} \right)^2$. Then

$$\hat{d}_{\mathfrak{h}(p)} = \hat{\Psi}_{\mathfrak{h}(p)}^{-1} \hat{\psi}_{\mathfrak{h}(p)}, \quad (2.12)$$

and we denote the elements of $\hat{d}_{\mathfrak{h}(p)}$ by $\hat{d}_{s, \mathfrak{h}(p)}$, $s \in S[-p_L, p_U]$. Denote by $d_{s, \mathfrak{h}(p)}$ the scalars

$$\arg \min_{a_s, s \in S[-p_L, p_U]} E \left(x_t - \sum_{s \in S[-p_L, p_U]} a_s x_{t-s} \right)^2, \quad (2.13)$$

and the minimum by $\sigma_{\mathfrak{h}(p)}^2$. By a predictor of order $l < \mathfrak{h}(p)$, we will mean \hat{d}_l with the l lags corresponding to the first l subscripts in the first row of $\Psi_{\mathfrak{h}(p)}$ as ordered in Section B. Throughout the sequel we assume that $\mathfrak{h}(p)^{-1} + \mathfrak{C}(p)^{-1} \rightarrow 0$, as $N \rightarrow \infty$.

Theorem 2.1. *Let Assumptions A, B and C hold, the sequence p be chosen as a function of N such that*

$$\frac{\mathfrak{C}(p)}{N^{\frac{v-1}{v}}} \rightarrow 0, \text{ as } N \rightarrow \infty, \quad (2.14)$$

and

$$\sum_{t \in S_{1+}^\infty \setminus S[-p_L, p_U]} |d_t| \rightarrow 0 \text{ as } N \rightarrow \infty. \quad (2.15)$$

Then

$$\left\| \hat{d}_{\mathfrak{h}(p)} - d_{\mathfrak{h}(p)} \right\| \xrightarrow{p} 0, \text{ as } N \rightarrow \infty.$$

Condition (2.15) says that the dependence from ‘distant’ lags must decline sufficiently fast. The result for $d > 1$ differs from the case $d = 1$ in one important sense. In the latter condition (2.14) applies to the dimension of the parameter space, because this dimension equals the number of unique covariances in $\Psi_{\mathfrak{h}(p)}$. Now this is not the case due to (B.4). By restricting the growth of $\mathfrak{C}(p)$ relative to N further, we can strengthen the mode of convergence to almost-sure convergence.

Theorem 2.2. *Let Assumptions A, B, C and (2.15) hold, the sequence p be chosen as a function of N such that*

$$\mathfrak{C}(p) = \mathcal{O} \left(\frac{N^{\frac{v-1}{v}}}{(\log N)^{\frac{v+1}{v}} (\log \log N)^v} \right) \text{ and } \mathfrak{C}(p) < K 2^m \text{ as } N \rightarrow \infty, \quad (2.16)$$

for some integer m such that $2^m \leq N$ and some $K < 1$. Then

$$\left\| \hat{d}_{\mathfrak{h}(p)} - d_{\mathfrak{h}(p)} \right\| \xrightarrow{a.s.} 0, \text{ as } N \rightarrow \infty.$$

Define the error variance estimate as

$$\hat{\sigma}_{\mathfrak{h}(p)}^2 = n_p^{-1} \sum_{t(p,n)}'' \left(x_t - \sum_{s \in S[-p_L, p_U]} \hat{d}_{s, \mathfrak{h}(p)} x_{t-s} \right)^2.$$

Theorem 2.3. Under the conditions of Theorem 2.1, $\hat{\sigma}_{\mathfrak{h}(p)}^2 \xrightarrow{p} \sigma^2$, while under the conditions of Theorem 2.2, $\hat{\sigma}_{\mathfrak{h}(p)}^2 \xrightarrow{a.s.} \sigma^2$, both as $N \rightarrow \infty$.

3 Uniform consistency of AR spectral density estimates

We now introduce spectral density estimates. First, for $\lambda \in \Pi$, the spectral density of x_t under (2.4) is given by $f(\lambda) = \sigma^2 (2\pi)^{-d} \left| 1 - \sum_{s \in S_{1+}^\infty} d_s e^{is'\lambda} \right|^{-2}$, and we estimate this using

$$\hat{f}_{\mathfrak{h}(p)}(\lambda) = \frac{\hat{\sigma}_{\mathfrak{h}(p)}^2}{(2\pi)^d \left| 1 - \sum_{s \in S[-p_L, p_U]} \hat{d}_{s, \mathfrak{h}(p)} e^{is'\lambda} \right|^2}.$$

Berk (1974) established pointwise consistency of such an estimate when $d = 1$, and Bhansali (1980) proved that the convergence is uniform under the same conditions.

Theorem 3.1. Let Assumptions A, B and C hold, the sequence p be chosen as a function of N such that

$$\frac{\mathfrak{C}(p)\mathfrak{h}(p)^{\frac{1}{2}}}{N^{\frac{v-1}{v}}} \rightarrow 0, \text{ as } N \rightarrow \infty, \quad (3.1)$$

and

$$\mathfrak{h}(p)^{\frac{1}{2}} \sum_{t \in S_{1+}^\infty \setminus S[-p_L, p_U]} |d_t| \rightarrow 0, \text{ as } N \rightarrow \infty. \quad (3.2)$$

Then

$$\sup_{\lambda \in \Pi} \left| \hat{f}_{\mathfrak{h}(p)}(\lambda) - f(\lambda) \right| \xrightarrow{p} 0, \text{ as } N \rightarrow \infty.$$

The conditions imposed for this theorem were stronger than those for results in Section 2 in two ways. First, the condition restricting the rate of growth of the parameter space relative to sample size is stronger than the one imposed for Theorems 2.1 and 2.3. For example, if $v = 2$ then (2.14) required $\mathfrak{C}(p)/N^{1/2} \rightarrow 0$ whereas (3.1) in Theorem 3.1 requires $\mathfrak{C}(p)\mathfrak{h}(p)^{1/2}/N^{1/2} \rightarrow 0$. Note that for $d = 1$ the latter reduces to the condition established by Berk (1974), which is, in fact, a particular case of the condition in Robinson (1979). The second aspect of difference is the requirement in (3.2) that the dependence on ‘distant’ lags decline sufficiently fast to overcome norming by $\mathfrak{h}(p)^{\frac{1}{2}}$.

Theorem 3.2. *Let Assumptions [A](#), [B](#), [C](#), [\(3.2\)](#) hold, and choose the sequence p as a function of N such that*

$$\mathfrak{C}(p)\mathfrak{h}(p)^{\frac{1}{2}} = \mathcal{O}\left(\frac{N^{\frac{v-1}{v}}}{(\log N)^{\frac{v+1}{v}}(\log \log N)^v}\right) \text{ and } \mathfrak{C}(p) < K2^m \text{ as } N \rightarrow \infty, \quad (3.3)$$

for some integer m such that $2^m \leq N$ and some $K < 1$. Then

$$\sup_{\lambda \in \Pi} \left| \hat{f}_{\mathfrak{h}(p)}(\lambda) - f(\lambda) \right| \xrightarrow{a.s.} 0, \text{ as } N \rightarrow \infty.$$

4 Asymptotic normality

In this section we take $v = 2$ in Assumption [B](#), this being standardly imposed for central limit theorems in such settings. For any index t in the sum $\sum_{t \in [p], n}''$ we write $X_t(p)$ for the $\mathfrak{h}(p) \times 1$ vector with typical element x_{t-s} , $s \in S[-p_L, p_U]$. Denote by $\alpha(p)$ an $\mathfrak{h}(p) \times 1$ vector of constants, not all zero.

Lemma 4.1. *Let Assumptions [A](#), [B](#), [C](#) and [\(3.1\)](#) hold, with $v = 2$, and*

$$N^{\frac{1}{2}} \sum_{s \in S_{1+}^{\infty} \setminus S[-p_L, p_U]} |d_s| \longrightarrow 0, \text{ as } N \rightarrow \infty, \quad (4.1)$$

Then, as $N \rightarrow \infty$,

$$N^{\frac{1}{2}} \alpha(p)' \left(\hat{d}_{\mathfrak{h}(p)} - d_{\mathfrak{h}(p)} \right) / \mathfrak{h}(p)^{\frac{1}{2}} - N^{\frac{1}{2}} \sum_{t \in [p], n}'' \alpha(p)' \Psi_{\mathfrak{h}(p)}^{-1} X_t(p) \epsilon_t / n_p \mathfrak{h}(p)^{\frac{1}{2}} \xrightarrow{p} 0. \quad (4.2)$$

Lemma 4.2. *Write $D_{\mathfrak{h}(p)}(z) = 1 - \sum_{s \in S[-p_L, p_U]} d_{s, \mathfrak{h}(p)} z^s$ and let Assumption [A](#) hold. Then $\lim_{\mathfrak{h}(p) \rightarrow \infty} D_{\mathfrak{h}(p)}(z) = D(z)$ for $|z_i| \leq 1$, $i = 1, \dots, d$.*

Lemma 4.3. *Let the conditions of Lemma [4.2](#) hold. Write $\boldsymbol{\pi} = (\pi, \dots, \pi)'$, let $w_1 = \bar{u}_1, \dots, w_q = \bar{u}_q$ be complex numbers for some positive integer q , w_0 and u_0 real numbers, for $t \in S_{1+}^{\infty}$, $\lambda_i \in (0, \pi)^d$ define*

$$\beta_t = w_0 + w_1 e^{it' \lambda_1} + \dots + w_q e^{it' \lambda_q} + w_0 e^{it' \boldsymbol{\pi}} + u_1 e^{-it' \lambda_1} + \dots + u_q e^{-it' \lambda_q}, \quad (4.3)$$

and $\alpha(p)$ be the $\mathfrak{h}(p) \times 1$ vector with typical element β_s , $s \in S[-p_L, p_U]$. Then

$$\lim_{\mathfrak{h}(p) \rightarrow \infty} \mathfrak{h}(p)^{-1} \alpha(p)' \Psi_{\mathfrak{h}(p)}^{-1} \alpha(p) = \mu,$$

where $\mu = w_0^2 / (2\pi)^d f(\mathbf{0}) + 2w_1 u_1 / (2\pi)^d f(\lambda_1) + \dots + 2w_q u_q / (2\pi)^d f(\lambda_q) + u_0^2 / (2\pi)^d f(\boldsymbol{\pi})$.

Lemmas [4.1](#), [4.2](#) and [4.3](#) are lattice extensions of results in [Berk \(1974\)](#). The next theorem establishes the asymptotic distribution of a linear combination (with trigonometric

coefficients) of the autoregression coefficient estimates. The distribution is analogous to that derived when $d = 1$.

Theorem 4.1. *Let Assumptions [A](#), [B](#) and [C](#) hold, with $v = 2$ and $d \geq 2$, and $\alpha(p)$ be as in Lemma [4.3](#). Choose the sequence p as a function of N such that [\(2.14\)](#) holds and*

$$\frac{\mathfrak{h}(p)}{N^{\frac{\chi}{4}}} + N^{\frac{1}{2}} \sum_{s \in S_{1+}^{\infty} \setminus S[-p_L, p_U]} |d_s| \longrightarrow 0, \text{ as } N \rightarrow \infty. \quad (4.4)$$

Then, as $N \rightarrow \infty$,

$$\left(N^{\frac{1}{2}} / \mathfrak{h}(p)^{\frac{1}{2}} \right) \alpha(p)' \left(\hat{d}_{\mathfrak{h}(p)} - d_{\mathfrak{h}(p)} \right) \xrightarrow{d} N(0, \sigma^2 \mu), \text{ as } N \rightarrow \infty.$$

It is straightforward to extend the argument to allow for the asymptotic distribution of finitely many linear combinations by replacing $\alpha(p)$ with an $s \times \mathfrak{h}(p)$ matrix with full row rank, s fixed, but we consider $s = 1$ for simplicity. Condition [\(4.4\)](#) presents an important difference from the case when $d = 1$, where the first term on the LHS of the limit is replaced by the much sharper $\mathfrak{h}(p)/N^{\frac{1}{2}}$. On the other hand, [\(4.4\)](#) can never be this sharp as $\chi = 1$ at most when $d = 1$, thus reflecting the difference in proof techniques for time series and lattice cases noted by [Robinson and Vidal Sanz \(2006\)](#), and imposing a considerable tightening on the rate of growth of $\mathfrak{h}(p)$ that strengthens with increasing d . Define

$$\begin{aligned} \hat{C}_{\mathfrak{h}(p)}(\lambda) &= 1 + \sum_{s \in S[-p_L, p_U]} \hat{d}_{s, \mathfrak{h}(p)} \cos(s' \lambda), & C(\lambda) &= 1 + \sum_{s \in S_{1+}^{\infty}} d_s \cos(s' \lambda), \\ \hat{S}_{\mathfrak{h}(p)}(\lambda) &= \sum_{s \in S[-p_L, p_U]} \hat{d}_{s, \mathfrak{h}(p)} \sin(s' \lambda), & S(\lambda) &= \sum_{s \in S_{1+}^{\infty}} d_s \sin(s' \lambda), \end{aligned}$$

the $2(q+1) \times 1$ vector $t_{\mathfrak{h}(p)}$ to have elements

$$\begin{aligned} &\left(\hat{C}_{\mathfrak{h}(p)}(\mathbf{0}) - C(\mathbf{0}) \right), \left(\hat{C}_{\mathfrak{h}(p)}(\lambda_1) - C(\lambda_1) \right), \dots, \left(\hat{C}_{\mathfrak{h}(p)}(\lambda_q) - C(\lambda_q) \right), \\ &\left(\hat{C}_{\mathfrak{h}(p)}(\boldsymbol{\pi}) - C(\boldsymbol{\pi}) \right), \left(\hat{S}_{\mathfrak{h}(p)}(\lambda_1) - S(\lambda_1) \right), \dots, \left(\hat{S}_{\mathfrak{h}(p)}(\lambda_q) - S(\lambda_q) \right), \end{aligned} \quad (4.5)$$

and the $2(q+1) \times 2(q+1)$ matrix

$$\begin{aligned} \Gamma &= \left(\sigma^2 / (2\pi)^d \right) \text{diag} \left(1/f(\mathbf{0}), 1/2f(\lambda_1), \dots, 1/2f(\lambda_q), 1/f(\boldsymbol{\pi}), \right. \\ &\quad \left. 1/2f(\lambda_1), \dots, 1/2f(\lambda_q) \right). \end{aligned} \quad (4.6)$$

Lemma 4.4. *Under the conditions of Theorem [4.1](#),*

$$(N/\mathfrak{h}(p))^{\frac{1}{2}} t_{\mathfrak{h}(p)} \xrightarrow{d} N(0, \Gamma), \text{ as } N \rightarrow \infty.$$

Lemma [4.4](#) is analogous to results in the time series literature, cf. [Parzen \(1969\)](#), [Berk](#)

$d = 2$					$d = 3$				
τ		0.05	0.075	0.10	τ		0.0075	0.015	0.03
n^*	p	MISE	MISE	MISE	n^*	p	MISE	MISE	MISE
5	1	0.1819	0.3873	0.7297	3	1	0.2878	0.9122	0.8654
7	1	0.1217	0.2923	0.5764	4	1	0.1469	0.2439	0.3832
9	1	0.1132	0.2706	0.5301	5	1	0.1330	0.2329	0.3818
9	2	0.0478	0.0691	0.1166	6	1	0.1391	0.2407	0.3933
11	1	0.1092	0.2717	0.5064	7	1	0.1374	0.2405	0.3835
11	2	0.0287	0.0534	0.1052	8	1	0.1364	0.2387	0.3852
11	3	0.0682	0.0890	0.1056	8	2	0.1381	0.2530	0.5170

Table 5.1: Monte Carlo MISE of $\hat{f}_{\mathfrak{h}(p)}(\bar{\lambda})$, $\bar{\lambda} \in \mathbb{G}$

(1974). Now define the $(q+2) \times 1$ vector $s_{\mathfrak{h}(p)}$ to have elements

$$\hat{f}_{\mathfrak{h}(p)}(\mathbf{0}) - f(\mathbf{0}), \hat{f}_{\mathfrak{h}(p)}(\lambda_1) - f(\lambda_1), \dots, \hat{f}_{\mathfrak{h}(p)}(\lambda_q) - f(\lambda_q), \hat{f}_{\mathfrak{h}(p)}(\boldsymbol{\pi}) - f(\boldsymbol{\pi}), \quad (4.7)$$

and the $(q+2) \times (q+2)$ matrix

$$\Omega = 2 \operatorname{diag} (2f^2(\mathbf{0}), f^2(\lambda_1), \dots, f^2(\lambda_q), 2f^2(\boldsymbol{\pi})). \quad (4.8)$$

Theorem 4.2. *Let the conditions of Theorem 4.1 hold with (2.14) replaced by (3.1). Then*

$$(N/\mathfrak{h}(p))^{\frac{1}{2}} s_{\mathfrak{h}(p)} \xrightarrow{d} N(0, \Omega), \text{ as } N \rightarrow \infty.$$

The asymptotic distribution of the spectral density estimates at various frequencies mirrors that in the time series case (cf. Anderson (1971), ch. 9, Berk (1974)), albeit under the stronger condition (4.4) and different condition (3.1).

5 Monte Carlo simulations

We examined finite-sample behaviour in a set of Monte Carlo simulations. As in Robinson and Vidal Sanz (2006) and Robinson (2007) we generated x_t using

$$x_t = \sigma \epsilon_t + \sigma \tau \sum_{s_1=-1}^1 \cdots \sum_{\substack{s_d=-1 \\ s \neq \mathbf{0}}}^1 \epsilon_{t-s}, \quad (5.1)$$

for $d = 2, 3$, similar to one considered in Haining (1978). Then

$$f(\lambda) = \frac{\sigma^2}{2\pi^d} \{1 + \tau \nu_d(\lambda)\}, \quad (5.2)$$

$d = 2$							
τ		0.05		0.075		0.10	
n^*	p	Bias	Std. Dev.	Bias	Std. Dev.	Bias	Std. Dev.
5	1	0.0238	0.0665	0.0681	0.3063	0.1496	0.4834
7	1	0.0121	0.0312	0.0367	0.0661	0.0876	0.1523
9	1	0.0112	0.0228	0.0321	0.0513	0.0739	0.1024
9	2	0.0090	0.0830	0.0014	0.0551	-0.0028	0.0495
11	1	0.0105	0.0200	0.0317	0.0434	0.0678	0.0838
11	2	0.0023	0.0268	0.0002	0.0317	-0.0063	0.0383
11	3	0.0212	0.1978	0.0263	0.1191	0.0319	0.1165

$d = 3$							
τ		0.0075		0.015		0.03	
n^*	p	Bias	Std. Dev.	Bias	Std. Dev.	Bias	Std. Dev.
3	1	0.0064	0.1164	0.0023	0.0161	0.0728	0.8927
4	1	-0.0001	0.0021	0.0003	0.0032	0.0041	0.0094
5	1	-0.0002	0.0013	0.0000	0.0019	0.0029	0.0058
6	1	-0.0004	0.0010	-0.0002	0.0014	0.0028	0.0041
7	1	-0.0003	0.0008	-0.0002	0.0010	0.0024	0.0037
8	1	-0.0003	0.0007	-0.0002	0.0008	0.0022	0.0033
8	2	0.0003	0.0020	0.0027	0.0049	0.0207	0.0314

Table 5.2: Monte Carlo bias and standard deviation of $\hat{f}_{\mathbf{b}(p)}(\mathbf{0})$

with $\nu_d(\lambda) = \prod_{i=1}^d (1 + 2 \cos \lambda_j) - 1$. Robinson and Vidal Sanz (2006) show that a sufficient condition for invertibility of (5.1) is

$$|\tau| < \left(3^d - 1\right)^{-1}. \quad (5.3)$$

We took $\mathbb{L} = \{t : -n^* \leq t_i \leq n^*, i = 1, \dots, d\}$, i.e. $n_{L_i} = n_{U_i} = n^*$ for all $i = 1, \dots, d$, implying $N = (2n^* + 1)^d$, and generated NID(0,1) ϵ_t (so $\sigma^2 = 1$) on \mathbb{L} in each of the 500 replications. We experimented with more values of τ and n^* than Robinson (2007), using the following specifications:

$$\begin{aligned} d &= 2 : \tau = 0.05, 0.075, 0.10; \quad n^* = 5, 7, 9, 11 \\ d &= 3 : \tau = 0.0075, 0.015, 0.03; \quad n^* = 3, 4, 5, 6, 7, 8. \end{aligned}$$

We maintained $p_{L_i} = p_{U_i} = p_{U_1} = p$, $i = 2, \dots, d$ and for $d = 2$ took $p = 1$ for $n^* = 5, 7$; $p = 1, 2$ for $n^* = 9$ and $p = 1, 2, 3$ for $n^* = 11$, while for $d = 3$ we took $p = 1$ for $n^* = 3, 4, 5, 6, 7$; $p = 1, 2$ for $n^* = 8$. The choices of τ satisfy (5.3).

Π was discretized with gaps of 0.10 in each dimension and we call this grid \mathbb{G} . In Table 5.1 we report Monte Carlo mean integrated squared error (MISE), which we define

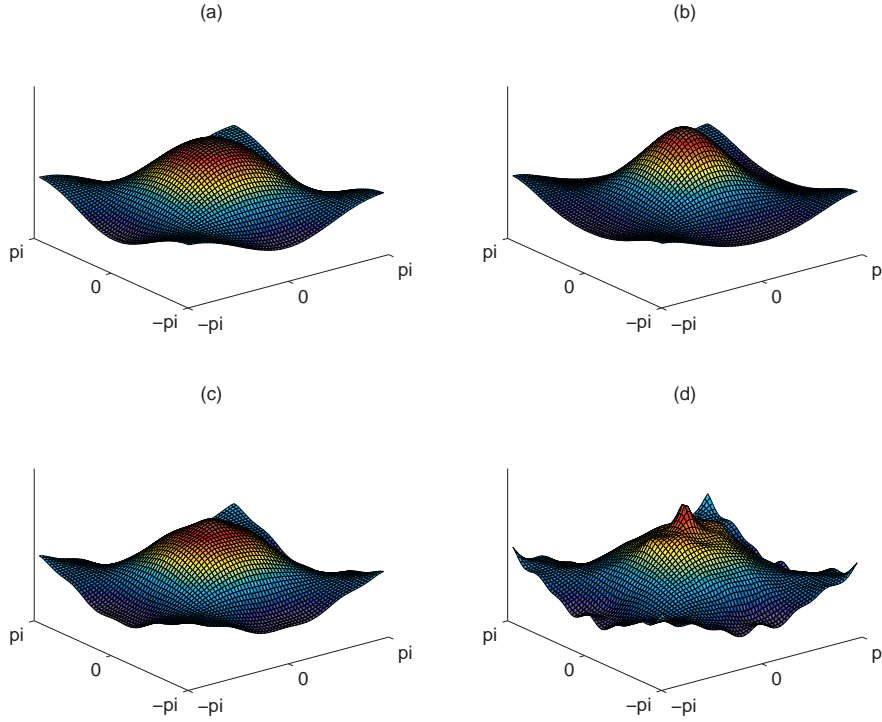


Figure 5.1: Spectral estimates for $d = 2$, $n^* = 11$, $\tau = 0.05$. (a) True spectrum (b) AR estimate with $p = 1$ (c) AR estimate with $p = 2$ (d) AR estimate with $p = 3$

as the Euclidean norm of $\hat{f}_{\hat{h}(p)}(\bar{\lambda}) - f(\bar{\lambda})$ evaluated at frequencies in \mathbb{G} , i.e. $\text{MISE} = \left\{ \sum_{\bar{\lambda} \in \mathbb{G}} \left(\hat{f}_{\hat{h}(p)}(\bar{\lambda}) - f(\bar{\lambda}) \right)^2 \right\}^{\frac{1}{2}}$.

Regardless of the value of d , MISE is smaller for smaller values of τ . As n^* increases MISE decreases for each value of τ , but not monotonically when $d = 3$. In the following discussion any triple is to be read as (n^*, d, p) . The MISE for $(9, 2, 1)$ dominates that for $(9, 2, 2)$ for any value of τ , and likewise the MISE for $(11, 2, 1)$ compared to $(11, 2, 2)$. However there is a cost in allowing increase of p and that is reflected in the MISE for $(11, 2, 3)$ dominating that for $(11, 2, 2)$. Similar patterns are seen for other values of n^* but the results for bigger p than those shown are not worth reporting for either value of d . The case $(8, 3, 1)$ exhibits very little change from $(7, 3, 1)$, while $(8, 3, 2)$ performs worse than $(8, 3, 1)$ for all values of τ .

Table 5.2 reports Monte Carlo bias and standard deviation (SD) at $\lambda = \mathbf{0}$. The biases decrease monotonically for all values of τ when $d = 2$, while for $d = 3$ the decrease is not monotonic always, although the values seem quite acceptable. The biases are much smaller for $d = 3$, almost vanishing for larger n^* and smaller τ . These features match those of the untapered and tapered periodogram based estimates in Robinson (2007), but there other features that differ from that paper. All biases there are negative, but we find that they are mostly positive. For $d = 2$ our biases sometimes dominate (in absolute value) those in Robinson (2007) but can become better than untapered estimates e.g. for $n^* = 9$. We experimented with more values of n^* (not reported) for both $d = 2$ and

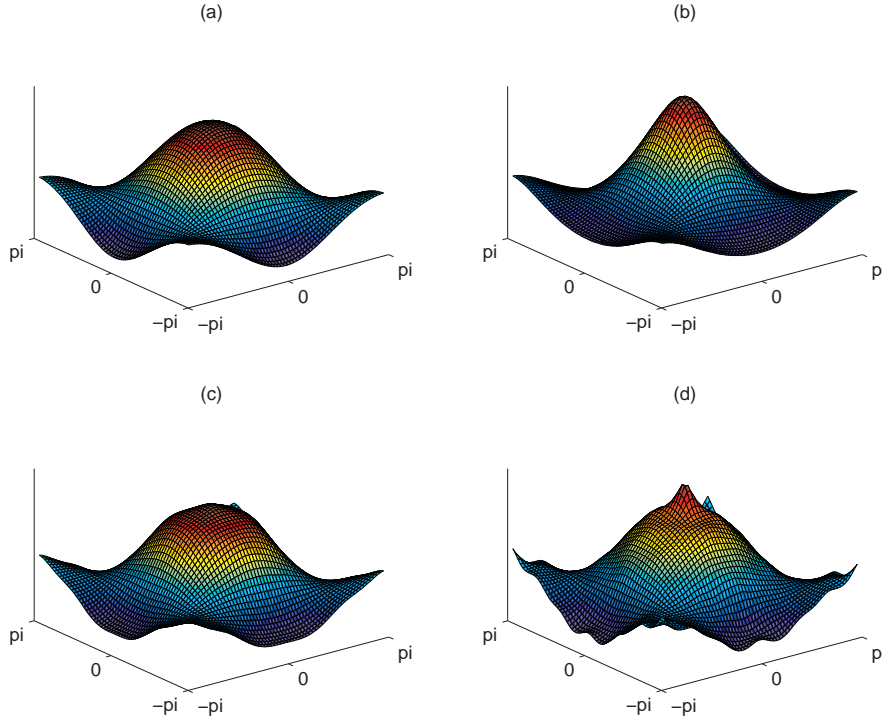


Figure 5.2: Spectral estimates for $d = 2$, $n^* = 11$, $\tau = 0.075$. (a) True spectrum (b) AR estimate with $p = 1$ (c) AR estimate with $p = 2$ (d) AR estimate with $p = 3$

$d = 3$ and find that the biases are acceptable for small values of p . Unlike Robinson (2007) we find that the SD also reduces monotonically with n^* and for $d = 3$ becomes zero up to two decimal places when $n^* \geq 4$ for all τ , with just one exception for $(8, 3, 2)$. For $d = 2$ such behaviour is not observed, but SD does decline as n^* increases.

The behaviour of estimates relative to true spectra for $d = 2$ is illustrated graphically over \mathbb{G} with $n^* = 11$ for $\tau = 0.05, 0.075, 0.10$ in Figures 5.1, 5.2 and 5.3 respectively. In each figure the top-left surface, labelled (a), plots the true spectral density. The figures labelled (b), (c), (d) show plots of the autoregressive spectral density estimate computed using $p = 1, 2, 3$, respectively. All spectra are plotted on a \log_{10} scale. Figure 5.1 shows that the estimated spectrum when $\tau = 0.05$ has too sharp a peak for $p = 1$, but this flattens to one resembling the true peak for $p = 2$. As seen in Tables 5.1 and 5.2, estimates worsen for $p = 3$, illustrated by the choppy and very sharp-peaked surface in Figure 5.1(d). For $\tau = 0.075$, Figure 5.2 exhibits similar features, with $p = 2$ giving the best estimate. Finally, for $\tau = 0.10$ we see again from Figure 5.3 that $p = 2$ does best but compared with Figures 5.1(c) and 5.2(c) the contours of the true spectrum are not as well estimated, as observed numerically in Tables 5.1 and 5.2. For $p = 3$ the estimated surface exhibits poor properties by flattening, as opposed to the sharp peaks exhibited by Figures 5.1(d) and 5.2(d).

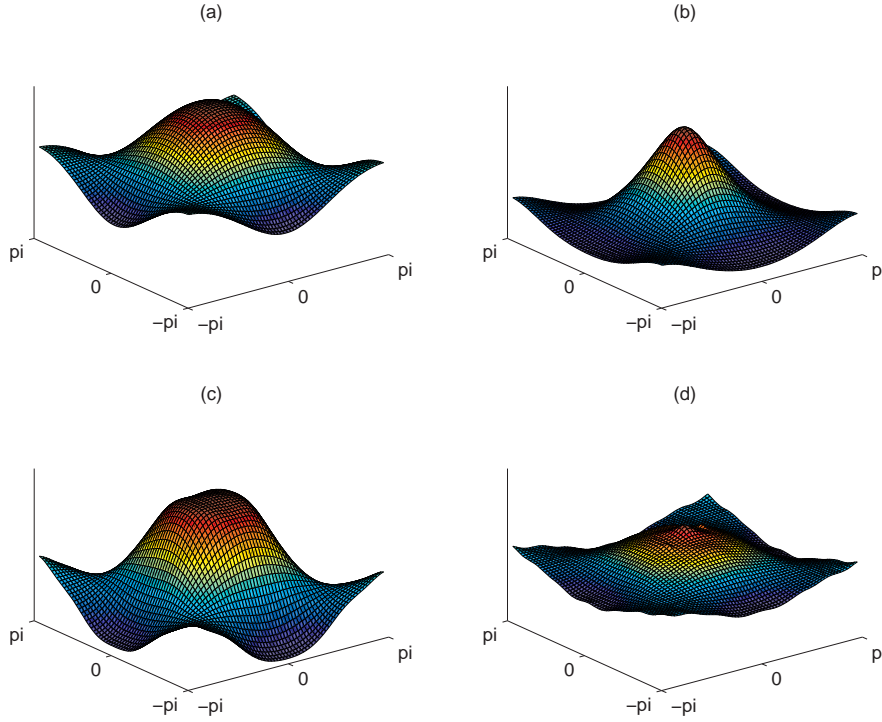


Figure 5.3: Spectral estimates for $d = 2$, $n^* = 11$, $\tau = 0.10$. (a) True spectrum (b) AR estimate with $p = 1$ (c) AR estimate with $p = 2$ (d) AR estimate with $p = 3$

6 Empirical examples

Data for both our examples is available at www.spatial-statistics.com.

6.1 Los Angeles house price data

We use median house price data for census blocks in California from the 1990 census from [Pace and Barry \(1997\)](#). We confine our analysis to the city of Los Angeles. The data is gridded as follows: an 8×20 grid of square cells, each with about a 4.8 km edge is superimposed on Los Angeles, from 33.752°N to 34.152°N and 117.439°W to 118.439°W . The grid covers a total of 5450 observations. The average of the median house values for each cell is calculated and the 160 such observations form our sample. There are 8 empty cells, to which we assign the value zero in the spirit of the time series missing data literature. The gridding is shown in [Figure 6.1](#). Smaller grid cells would lead to more empty cells, and [Bronars and Jansen \(1987\)](#) note that while choice of cell size is somewhat arbitrary it is analogous to selecting quarterly, monthly or weekly data in time series analysis. House price data is not a zero mean process, so we subtract the sample mean using the whole sample from each cell as remarked in [Section 2](#).

Our lag order choices are $p_{U_1} = 1$, $p_{L_2} = p_{U_2} = 5$, which enable clear identification of peaks. Some spatial generalisations of various time series information criteria for selecting lag orders have been discussed in the literature for the quarter-plane case, c.f. [Tjøstheim \(1981\)](#) for a generalisation of the Final Prediction Error (FPE) criterion of

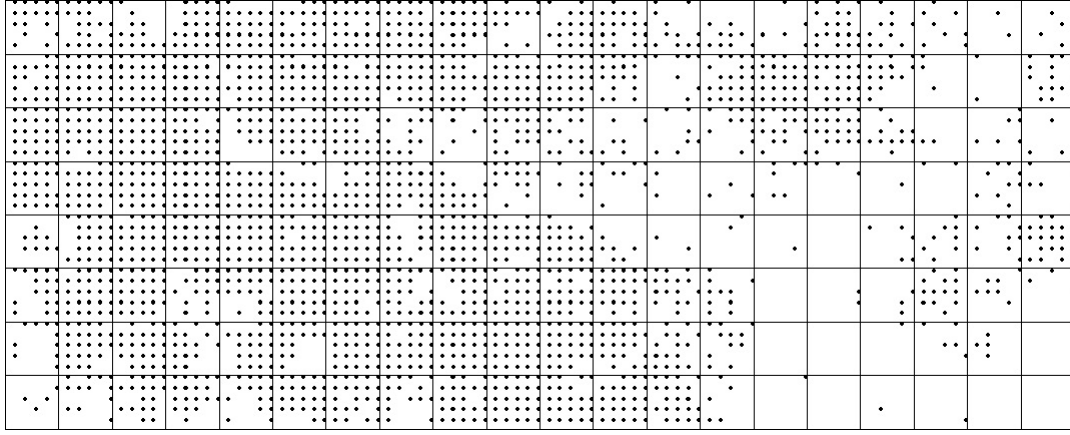


Figure 6.1: Gridded Los Angeles median house price data

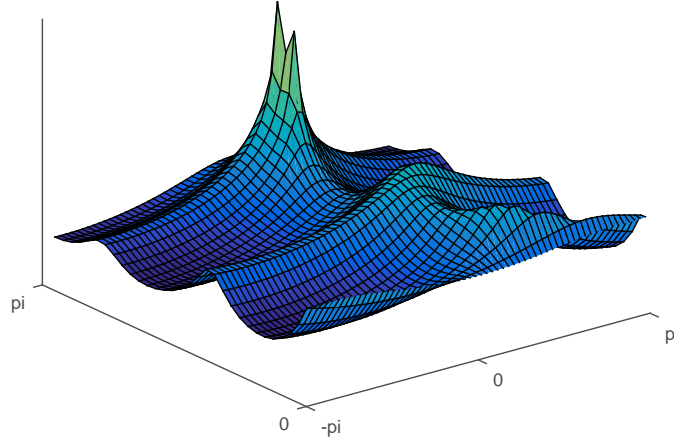


Figure 6.2: Spectral density estimate for Los Angeles median house price data

[Akaike \(1970\)](#) and the Bayesian Information Criterion (BIC) of [Schwarz \(1978\)](#). However half-plane extensions seem to be unavailable in the literature and ad hoc extensions we tried do not work very well. To be precise using $FPE = \hat{\sigma}_{\mathfrak{h}(p)}^2(N + \mathfrak{h}(p))/(N - \mathfrak{h}(p))$ leads to overfitting and makes it hard to identify peaks clearly, in this example suggesting $p_{U_1} = 1$, $p_{L_2} = p_{U_2} = 7$. This is clearly an area for future research.

The estimated spectrum is plotted on a \log_{10} scale in [Figure 6.2](#). Due to symmetry we only plot the results over $(-\pi, \pi] \times [0, \pi]$. There are very clear and strong peaks at low frequency, illustrating power in structural low frequency components. As discussed in the introduction these correspond to components that change infrequently over the space. High frequency peaks are not very strong, but clear ripples are visible indicating moderate power in noise components that change frequently over space such as pollution, sunshine, crime rate and proximity to busy roads.

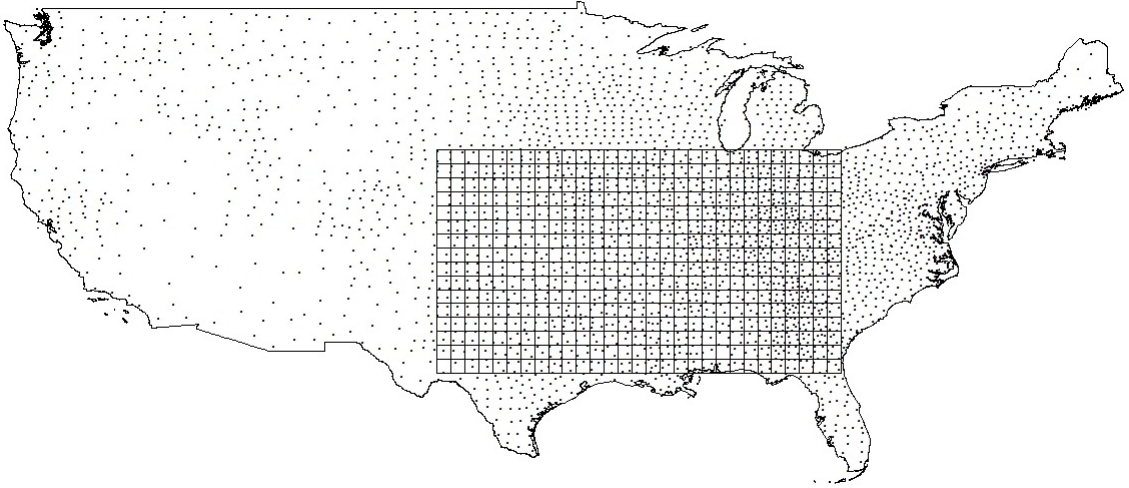


Figure 6.3: Gridded county level US voter turnout data

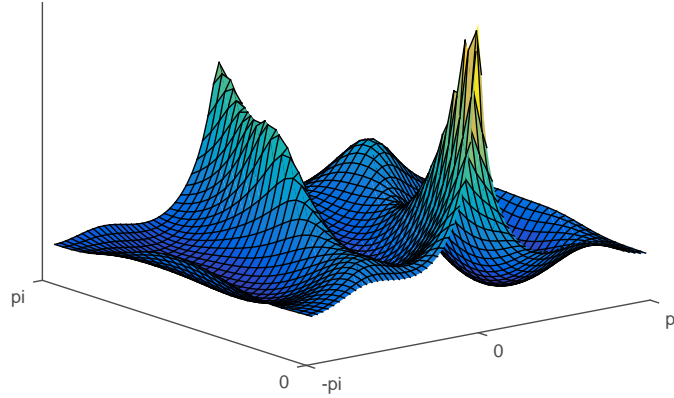


Figure 6.4: Spectral density estimate for county level US presidential election data

6.2 US presidential election voter turnout data

In this example we study county level voter turnout (defined as votes cast divided by total population) data from the 1980 US presidential election, used in [Pace and LeSage \(2003\)](#). Following a strategy similar to [Bronars and Jansen \(1987\)](#) we grid the data over a 16×29 grid of square cells, each with about a 69.3 km edge, from 30.20°N to 41.72°N and 81.52°W to 102.4°W . As [Figure 6.3](#) illustrates, the choice of coordinates gives the largest possible sample size while accounting for the irregular border and coastline of the US, as well as the relative sparsity of observations west of our imposed North-South border that runs from Nebraska to Texas. The grid covers a total of 1539 counties, and the voter turnout is taken as recorded at the centroid of each county. The average of the voter turnout for the centroids that lie in each cell is calculated, and the sample mean subtracted from each cell, yielding 464 observations. There are no empty cells and since a centroid can only appear in one cell there is no overlap. We take $p_{U_1} = p_{L_2} = p_{U_2} = 3$, noting again that the FPE suggested $p_{U_1} = p_{L_2} = p_{U_2} = 5$ leads to an uninformative spectrum.

The estimated spectrum is plotted in Figure 6.4, again over half of the frequency plane and on a \log_{10} scale. There is a very strong peak at low frequency, indicating the power in low frequency structural components. The 1980 election was a historic one, with Ronald Reagan defeating Jimmy Carter in a landslide victory. Thus the closeness of the race could be interpreted to not change very frequently over space, contributing to a strong low frequency component. However there are high frequency ripples and one strong high frequency peak, though not as strong as the low frequency one. Our analysis suggests that turnout rates in this election were influenced quite strongly by factors that change with high frequency over space, such as weather and demographic composition of the electorate. The latter includes age distribution, racial distribution, gender distribution and socio-economic distribution in the various counties. This seems reasonable in an election in which an incumbent Democratic candidate was heavily defeated by a Republican challenger.

Acknowledgements

I am grateful to Professor Peter M. Robinson for detailed comments that led to a much improved version. Helpful comments from Gordon Kemp are also acknowledged.

A Bounds for moments of partial sums of lattice processes

In this appendix we establish bounds for w -th absolute moments of partial sums of a class of lattice processes, with $w \in (1, 2]$. The class of processes under consideration is one that arises in many applications, so the result may be of independent interest due to its generality. Consider a scalar lattice process $\{\zeta_t : t \in \mathbb{L}\}$ defined by $\zeta_t = \sum_{s^1 \in \mathbb{Z}^d} \dots \sum_{s^q \in \mathbb{Z}^d} \xi_{st}$, $t \in \mathbb{L}$, where $\mathbf{s} = (s^1, \dots, s^q)$. This definition covers situations where certain statistics of spatial processes may be expressible in terms of products of sums of random variables. Assume that this process satisfies the following conditions:

Assumption D. ξ_{st} are mean-zero and independent over $t \in \mathbb{L}$.

Assumption E. For some $w \in (1, 2]$ positive constants $\{\eta_{ks} : s \in \mathbb{Z}^d, 1 \leq k \leq q\}$, $\{a_t : t \in \mathbb{L}\}$ exist such that

$$E |\xi_{st}|^w < \eta_{\mathbf{s}}^w a_t^w, \quad (\text{A.1})$$

where $\eta_{\mathbf{s}} = \prod_{k=1}^q \eta_{ks^k}$ and

$$\sum_{s \in \mathbb{Z}^d} \eta_{ks} < \infty, \quad 1 \leq k \leq q. \quad (\text{A.2})$$

The result in this section is similar to Lemma 1 of [Robinson \(1978\)](#) for $d = 1$, but he allowed for martingale ξ_{st} , but as discussed earlier we avoid this.

Before we can introduce our result, we need to establish some more notation and illustrate it with examples. Write $L = (L_1, \dots, L_d)'$, $0 < L_i \leq n_{L_i} + n_{U_i}$ for $i = 1, \dots, d$, and define $S_L = \sum'_{t(L)} \zeta_t$, where $\sum'_{t(L)}$ runs over t satisfying $-n_{L_i} < t_i \leq L_i - n_{L_i}$. There are $\prod_{i=1}^d L_i$ summands in this sum. For any multiple index $t \in \mathbb{Z}^d$, write $|t| = (|t_1|, \dots, |t_d|)'$. Also write $M = (M_1, \dots, M_d)'$, M_i possibly negative, with $|M_i| < L_i$, and define $S_{ML} = \sum''_{t(|M|, L)} \zeta_t$, where $\sum''_{t(|M|, L)}$ runs over t satisfying

$$\begin{aligned} -n_{L_i} < t_i \leq L_i - |M_i| - n_{L_i}; & \text{ if } M_i < 0, \\ M_i - n_{L_i} < t_i \leq L_i - n_{L_i}; & \text{ if } M_i \geq 0, \end{aligned} \quad (\text{A.3})$$

indicating that there are $\prod_{i=1}^d (L_i - |M_i|)$ summands in this sum. If $M_i \geq 0$ for each $i = 1, \dots, d$ then, unlike in time series, $S_{ML} \neq S_L - S_M$. In the d -dimensional lattice case we may write $S_{ML} = S_L - S_{ML}^*$ with $S_{ML}^* = \sum^*_{t(M, L)} \zeta_t$, $\sum^*_{t(M, L)}$ running over t satisfying $-n_{L_i} < t_i \leq L_i$ with at least one $i = 1, \dots, d$ for which $t_i \leq M_i - n_{L_i}$. There are $\prod_{i=1}^d L_i - \prod_{i=1}^d (L_i - M_i)$ summands in this sum. For $d = 2$, S_L consists of the sum of observations at those points in the intersection of points to the north-east of $(-n_{L_1} + 1, -n_{L_2} + 1)$ and to the south-west of (L_1, L_2) . S_M is visualised similarly. S_{ML} consists of the sum of observations at those points in the intersection of points to the north-east of $(-n_{L_1} + M_1 + 1, -n_{L_2} + M_2 + 1)$ and to the south-west of (L_1, L_2) . Figure [A.1](#) illustrates these definitions for $d = 2$; $n_{L_1} = n_{L_2} = 0$; $n_{U_1} = n_{U_2} = 6$; $(L_1, L_2) = (4, 4)$ and $(M_1, M_2) = (2, 2)$. Observations summed in S_L

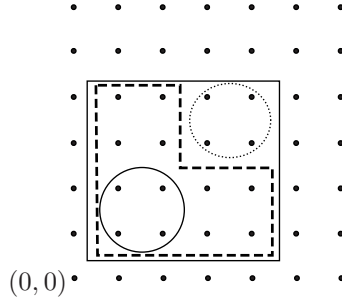


Figure A.1: Illustration of S_L, S_M, S_{ML}^* and S_{ML} , $d = 2$, $n_{L_1} = n_{L_2} = 0$; $n_{U_1} = n_{U_2} = 6$; $(L_1, L_2) = (4, 4)$ and $(M_1, M_2) = (2, 2)$.

are those recorded at points within the solid-bordered boxed area. For S_M , S_{ML}^* and S_{ML} the points of observation are in the solid-bordered circular area, dashed polygonal area and dotted circular area respectively. An alternative way of writing $\sum''_{t(|M|, L)}$ is $\sum_{t, t-M \in \mathbb{L}_L}$ where $\mathbb{L}_L = \{t \in \mathbb{Z}^d : -n_{L_i} \leq t_i \leq L_i - n_{L_i}, i = 1, \dots, d\}$. Now define $b_{wL} = 0$ if $L = (L_1, \dots, L_d)$, $L_i \geq 0$ for $i = 1, \dots, d$ with at least one $L_i = 0$, and $b_{wL} = \sum'_{t(L)} a_t^w$ if $L = (L_1, \dots, L_d)$, $L_i > 0$ for $i = 1, \dots, d$. Similarly define $b_{wML} = 0$ if $L - |M| = (L_1 - |M_1|, \dots, L_d - |M_d|)$, $L_i - |M_i| \geq 0$ for $i = 1, \dots, d$ with at least one $L_i - |M_i| = 0$, and $b_{wML} = \sum''_{t(|M|, L)} a_t^w$ if $L - |M| = (L_1 - |M_1|, \dots, L_d - |M_d|)$, $L_i - |M_i| > 0$ for $i = 1, \dots, d$. We are now in a position to prove the main result of this section.

Lemma A.1. *Let Assumptions **D** and **E** hold. Then $E|S_{ML}|^w < C b_{wML}$.*

Note that we did not impose stationarity of ζ_t , nor did we use any half-plane representation for ζ_t . In view of this Lemma A.1 is quite general.

B Properties of covariance matrices of autoregressive lattice processes

B.1 A spatial generalisation of the Toeplitz property

In this appendix we generalize the Toeplitz property of covariance matrices for stationary time series with finite autoregressive representations to stationary spatial processes with finite half-plane or quarter-plane representations. It is necessary to introduce an ordering of the elements of \mathbb{Z}^d in order to write the objects of interest in matrical and vectorial form. Such an ordering can be carried out in many ways and as long as a consistent ordering is followed it should not matter which particular one is used. However certain orderings may be more beneficial in obtaining a clearer picture of the structure of the covariance matrix. We consider the cases $d = 2$ and $d = 3$, and then discuss the situation for general d . We also illustrate the relevant quarter-plane situations first and then build on this treatment to explain the differences in the half-plane case, the latter being more

complicated due to negative entries in the indices. The definitions are recursive in nature.

$d = 2$

This case is discussed quite extensively in the signal-processing literature for instance in Tjøstheim (1981) and Wester et al. (1990).

Quarter-plane representations

Here $p_{L_2} = 0$. For each $l = 0, \dots, p_{U_1}$, define $\check{\psi}_l^{(1)}(p)$ to be the $(p_{U_2} + 1) \times 1$ vector with typical i -th element $\gamma(l, i)$, $i = 0, \dots, p_{U_2}$, and $\check{\psi}^{(2)}(p) = \left(\check{\psi}_0^{(1)}(p), \check{\psi}_1^{(1)}(p), \dots, \check{\psi}_{p_{U_1}}^{(1)}(p) \right)'$, the latter a nested vector of dimension $(p_{U_2} + 1) \times (p_{U_1} + 1)$. Finally denote by $\psi_{\mathfrak{h}(p)}$ the $(p_{U_1} + 1)(p_{U_2} + 1) - 1 \times 1$ vector got by removing the first element of $\check{\psi}_{\mathfrak{h}(p)}$, which has dimension $\mathfrak{h}(p) \times 1$. For each $l = 0, \dots, p_{U_1}$, define $\check{\Psi}_l^{(1)}(p)$ to be the $(p_{U_2} + 1) \times (p_{U_2} + 1)$ Toeplitz matrix with typical (i, j) -th element $\gamma(l, i - j)$, $i, j = 0, \dots, p_{U_2}$, $\check{\Psi}_{\mathfrak{h}(p)}$ to be the block-Toeplitz matrix of (block) dimension $(p_{U_1} + 1)$ and (i, j) -th block $\check{\Psi}_{i-j}^{(1)}(p)$, $i, j = 0, \dots, p_{U_1}$, so

$$\check{\Psi}_{\mathfrak{h}(p)} = \begin{pmatrix} \check{\Psi}_0^{(1)}(p) & \check{\Psi}_{-1}^{(1)}(p) & \dots & \check{\Psi}_{-p_{U_1}}^{(1)}(p) \\ \check{\Psi}_1^{(1)}(p) & \check{\Psi}_0^{(1)}(p) & \dots & \check{\Psi}_{-p_{U_1}+1}^{(1)}(p) \\ \vdots & \vdots & \ddots & \vdots \\ \check{\Psi}_{p_{U_1}}^{(1)}(p) & \check{\Psi}_{p_{U_1}-1}^{(1)}(p) & \dots & \check{\Psi}_0^{(1)}(p) \end{pmatrix}.$$

Denote by $\Psi_{\mathfrak{h}(p)}$ the $(p_{U_1} + 1)(p_{U_2} + 1) - 1 \times (p_{U_1} + 1)(p_{U_2} + 1) - 1$ matrix formed by deleting the first row and first column of $\check{\Psi}_{\mathfrak{h}(p)}$. Then the dimension of $\Psi_{\mathfrak{h}(p)}$ is $\mathfrak{h}(p) \times \mathfrak{h}(p)$.

Half-plane representations

Here we have $p_{L_2} > 0$. For each $l = 0, \dots, p_{U_1}$, define $\check{\psi}_l^{(1)}(p)$ as the $(p_2 + 1) \times 1$ vector with typical i -th element $\gamma(l, i)$, $i = -p_{L_2}, \dots, p_{U_2}$, and $\check{\psi}_{\mathfrak{h}(p)}$ as the $(p_2 + 1) \times (p_{U_1} + 1) \times 1$ nested vector with i -th block $\check{\psi}_i^{(1)}(p)$, $i = 0, \dots, p_{U_1}$. $\check{\psi}_{\mathfrak{h}(p)}$ has dimension $(p_{U_1} + 1)(p_2 + 1) \times 1$ with $(p_{U_1} + 1)(p_2 + 1) = \mathfrak{h}(p) + p_{L_2} + 1$. Therefore, unlike in the quarter-plane situation, we will now denote by $\psi_{\mathfrak{h}(p)}$ the $\mathfrak{h}(p) \times 1$ vector formed by deleting the first $p_{L_2} + 1$ elements of $\check{\psi}_{\mathfrak{h}(p)}$. For each $l = 0, \dots, p_{U_1}$, define $\check{\Psi}_l^{(1)}(p)$ to be the $(p_2 + 1) \times (p_2 + 1)$ Toeplitz matrix with typical (i, j) -th element $\gamma(l, i - j)$, $i, j = 0, \dots, p_2$. Now, define $\check{\Psi}_{\mathfrak{h}(p)}$ to be the block-Toeplitz matrix of (block) dimension $(p_{U_1} + 1) \times (p_{U_1} + 1)$ and (i, j) -th block $\check{\Psi}_{i-j}^{(1)}(p)$, $i, j = 0, \dots, p_{U_1}$. So we have

$$\check{\Psi}_{\mathfrak{h}(p)} = \begin{pmatrix} \check{\Psi}_0^{(1)}(p) & \check{\Psi}_{-1}^{(1)}(p) & \dots & \check{\Psi}_{-p_{U_1}}^{(1)}(p) \\ \check{\Psi}_1^{(1)}(p) & \check{\Psi}_0^{(1)}(p) & \dots & \check{\Psi}_{-p_{U_1}+1}^{(1)}(p) \\ \vdots & \vdots & \ddots & \vdots \\ \check{\Psi}_{p_{U_1}}^{(1)}(p) & \check{\Psi}_{p_{U_1}-1}^{(1)}(p) & \dots & \check{\Psi}_0^{(1)}(p) \end{pmatrix}.$$

$\check{\Psi}_{\mathfrak{h}(p)}$ has dimension $(p_{U_1} + 1)(p_2 + 1) \times (p_{U_1} + 1)(p_2 + 1)$ with $(p_{U_1} + 1)(p_2 + 1) = \mathfrak{h}(p) + p_{L_2} + 1$. Again, unlike in the quarter-plane case, we will denote by $\Psi_{\mathfrak{h}(p)}$ the $\mathfrak{h}(p) \times \mathfrak{h}(p)$ matrix formed by deleting the first $p_{L_2} + 1$ rows and columns of $\check{\Psi}_{\mathfrak{h}(p)}$.

$d = 3$

Quarter-plane representations

In this case $p_{L_2} = p_{L_3} = 0$. We build the definitions analogously to the $d = 2$ case. For $l = 0, \dots, p_{U_1}$ and $m = 0, \dots, p_{U_2}$, define $\check{\psi}_{l,m}^{(1)}(p)$ to be the $(p_{U_3} + 1) \times 1$ vector with typical i -th element $\gamma(l, m, i)$, $i = 0, \dots, p_{U_3}$ and $\check{\psi}_m^{(2)}(p)$ as the $(p_{U_3} + 1) \times (p_{U_1} + 1) \times 1$ nested vector with i -th block $\check{\psi}_{i,m}^{(1)}(p)$, $i = 0, \dots, p_{U_1}$, and finally $\check{\psi}_{\mathfrak{h}(p)}$ as the twice nested $\prod_{i=1}^3 (p_{U_i} + 1) \times 1$ block vector with i -th block $\check{\psi}_i^{(2)}(p)$, $i = 0, \dots, p_{U_2}$. Then denote by $\psi_{\mathfrak{h}(p)}$ the $\prod_{i=1}^3 (p_{U_i} + 1) - 1$ -dimensional vector formed by deleting the first element of $\check{\psi}_{\mathfrak{h}(p)}$, which is $\mathfrak{h}(p) \times 1$. We now define the matrices. For $l = 0, \dots, p_{U_1}$ and $m = 0, \dots, p_{U_2}$, define $\check{\Psi}_{l,m}^{(1)}(p)$ to be the $(p_{U_3} + 1) \times (p_{U_3} + 1)$ Toeplitz matrix with typical (i, j) -th element $\gamma(l, m, i - j)$, $i, j = 0, \dots, p_{U_3}$ and $\check{\Psi}_m^{(2)}(p)$ to be the block-Toeplitz with Toeplitz blocks matrix of (block) dimension $(p_{U_1} + 1)$ and (i, j) -th block given by $\check{\psi}_{i-j,m}^{(1)}(p)$, $i, j = 0, \dots, p_{U_1}$, and then write $\check{\Psi}_{\mathfrak{h}(p)}$ for the (thrice) block-Toeplitz matrix of (block) dimension $(p_{U_2} + 1) \times (p_{U_2} + 1)$ and (i, j) -th block given by $\check{\Psi}_{i-j}^{(2)}(p)$, $i, j = 0, \dots, p_{U_2}$. Now denote by $\Psi_{\mathfrak{h}(p)}$ the $\prod_{i=1}^3 (p_{U_i} + 1) - 1$ -dimensional matrix formed by deleting the first row and first column of $\check{\Psi}_{\mathfrak{h}(p)}$. Then the dimension of $\Psi_{\mathfrak{h}(p)}$ is $\mathfrak{h}(p) \times \mathfrak{h}(p)$.

Half-plane representations

Now $p_{L_2} > 0$ or/and $p_{L_3} > 0$. For $l = 0, \dots, p_{U_1}$ and $m = -p_{L_2}, \dots, p_{U_2}$, define $\check{\psi}_{l,m}^{(1)}(p)$ to be the $(p_3 + 1) \times 1$ vector with typical i -th element $\gamma(l, m, i)$, $i = -p_{L_3}, \dots, p_{U_3}$, $\check{\psi}_m^{(2)}(p)$ to be the $(p_3 + 1) \times (p_{U_1} + 1) \times 1$ nested vector with i -th block $\check{\psi}_{i,m}^{(1)}(p)$, $i = 0, \dots, p_{U_1}$ and $\check{\psi}_{\mathfrak{h}(p)}$ to be the $\prod_{i=1}^3 (p_i + 1) \times 1$ nested vector with i -th block $\check{\psi}_i^{(2)}(p)$, $i = -p_{L_2}, \dots, p_{U_2}$. $\check{\psi}_{\mathfrak{h}(p)}$ has dimension $\prod_{i=1}^3 (p_i + 1)$ and also $\prod_{i=1}^3 (p_i + 1) = \mathfrak{h}(p) + p_{L_3} + p_{L_2}(p_3 + 1) + 1$. Therefore, unlike in the quarter-plane situation, we will now denote by $\psi_{\mathfrak{h}(p)}$ the $\mathfrak{h}(p) \times 1$ vector formed by the following procedure:

1. Delete each of the $\check{\psi}_{0,m}^{(1)}(p)$, $m = -p_{L_2}, \dots, -1$.
2. Delete the first $p_{L_3} + 1$ elements from $\check{\psi}_0^{(2)}(p)$.

The total elements then deleted are $p_{L_2}(p_3 + 1) + p_{L_3} + 1$ in number, and the dimension of $\psi_{\mathfrak{h}(p)}$ follows. For the matrices, we again proceed similarly. For $l = 0, \dots, p_{U_1}$ and $m = -p_{L_2}, \dots, p_{U_2}$, define $\check{\Psi}_{l,m}^{(1)}(p)$ to be the $(p_3 + 1) \times (p_3 + 1)$ Toeplitz matrix with

typical (i, j) -th element $\gamma(l, m, i - j)$, $i, j = -p_{L_3}, \dots, p_{U_3}$, $\check{\Psi}_m^{(2)}(p)$ to be the block-Toeplitz with Toeplitz blocks matrix of (block) dimension $(p_{U_1} + 1)$ and (i, j) -th block $\check{\Psi}_{i-j, m}^{(1)}(p)$, $i, j = 0, \dots, p_{U_1}$, and $\check{\Psi}_{\mathfrak{h}(p)}$ to be the (thrice) block-Toeplitz matrix of (block) dimension $(p_2 + 1) \times (p_2 + 1)$ and (i, j) -th block $\check{\Psi}_{i-j}^{(2)}(p)$, $i, j = -p_{L_2}, \dots, p_{U_2}$. Now denote by $\Psi_{\mathfrak{h}(p)}$ the $\prod_{i=1}^3 (p_{U_i} + 1) - 1$ -dimensional matrix formed by deleting those rows and columns of $\check{\Psi}^{(3)}(p)$ corresponding to the elements of $\check{\Psi}_{\mathfrak{h}(p)}$ deleted earlier. For instance, if the i -th element of $\check{\psi}_{\mathfrak{h}(p)}$ was deleted then we delete the i -th row and i -th column of $\check{\Psi}_{\mathfrak{h}(p)}$. We repeat this for each deleted element of $\check{\psi}_{\mathfrak{h}(p)}$. Then the dimension of $\check{\Psi}_{\mathfrak{h}(p)}$ is $\mathfrak{h}(p) \times \mathfrak{h}(p)$.

General d

Quarter-plane representations

In this case we have $p_{L_2} = p_{L_3} = \dots = p_{L_d} = 0$. For $l_i = 0, \dots, p_{U_i}$, $i = 1, \dots, d - 1$, define $\check{\psi}_{l_1, \dots, l_{d-1}}^{(1)}(p)$ to be the $(p_{U_d} + 1) \times 1$ vector with typical i -th element $\gamma(l_1, \dots, l_{d-1}, i)$, $i = 0, \dots, p_{U_d}$, for $l_i = 0, \dots, p_{U_i}$, $i = 1, \dots, d - 1$ define $\check{\psi}_{l_2, \dots, l_{d-1}}^{(2)}(p)$ to be the nested vector of (nested) dimension $(p_{U_1} + 1)$ and i -th block $\check{\psi}_{i, l_2, \dots, l_{d-1}}^{(1)}(p)$, $i = 0, \dots, p_{U_1}$, and proceeding in this manner, for $l_{d-1} = 0, \dots, p_{U_{d-1}}$ define $\check{\psi}_{l_{d-1}}^{(d-1)}(p)$ to be the nested vector of (nested) dimension $(p_{U_{d-2}} + 1) \times 1$ and i -th block $\check{\psi}_{i, l_{d-1}}^{(d-2)}(p)$, $i = 0, \dots, p_{U_{d-2}}$. Finally, define $\check{\psi}_{\mathfrak{h}(p)}$ to be the nested vector of (nested) dimension $(p_{U_d} + 1)$ and i -th block $\check{\psi}_i^{(d-1)}(p)$, $i = 0, \dots, p_{U_{d-1}}$. Now denote by $\psi_{\mathfrak{h}(p)}$ the $\prod_{i=1}^d (p_{U_i} + 1) - 1$ -dimensional vector formed by deleting the first element of $\check{\psi}_{\mathfrak{h}(p)}$. Then the dimension of $\psi_{\mathfrak{h}(p)}$ is $\mathfrak{h}(p) \times 1$. For the matrices, for $l_i = 0, \dots, p_{U_i}$, $i = 1, \dots, d - 1$, we define $\check{\Psi}_{l_1, \dots, l_{d-1}}^{(1)}(p)$ to be the $(p_{U_d} + 1)$ -dimensional Toeplitz matrix with typical (i, j) -th element $\gamma(l_1, \dots, l_{d-1}, i - j)$, $i, j = 0, \dots, p_{U_d}$, for $l_i = 0, \dots, p_{U_i}$, $i = 2, \dots, d - 1$ define $\check{\Psi}_{l_2, \dots, l_{d-1}}^{(2)}(p)$ to be the block Toeplitz with Toeplitz blocks matrix of (nested) dimension $(p_{U_1} + 1)$ and (i, j) -th block $\check{\Psi}_{i-j, l_2, \dots, l_{d-1}}^{(1)}(p)$, $i, j = 0, \dots, p_{U_1}$, and, proceeding recursively, for $l_{d-1} = 0, \dots, p_{U_{d-1}}$ we define $\check{\Psi}_{l_{d-1}}^{(d-1)}(p)$ to be the nested block-Toeplitz matrix of (block) dimension $(p_{U_{d-2}} + 1) \times (p_{U_{d-2}} + 1)$ and (i, j) -th block $\check{\Psi}_{i-j, l_{d-1}}^{(d-2)}(p)$, $i, j = 0, \dots, p_{U_{d-2}}$. The next step consists of defining $\check{\Psi}_{\mathfrak{h}(p)}$ to be the block-Toeplitz matrix of (block) dimension $(p_{U_{d-1}} + 1) \times (p_{U_{d-1}} + 1)$ and (i, j) -th block $\check{\Psi}_{i-j}^{(d-1)}(p)$, $i, j = 0, \dots, p_{U_{d-1}}$. Now denote by $\Psi_{\mathfrak{h}(p)}$ the $\prod_{i=1}^d (p_{U_i} + 1) - 1$ -dimensional square matrix formed by deleting the first row and first column of $\check{\Psi}_{\mathfrak{h}(p)}$. Clearly the dimension of $\Psi_{\mathfrak{h}(p)}$ is $\mathfrak{h}(p) \times \mathfrak{h}(p)$.

Half-plane representations

Now $p_{L_i} > 0$ for some $i = 1, \dots, d$. For $l_i = -p_{L_i}, \dots, p_{U_i}$, $i = 1, \dots, d - 1$; $p_{L_1} = 0$, define $\check{\psi}_{l_1, \dots, l_{d-1}}^{(1)}(p)$ to be the $(p_d + 1) \times 1$ vector with typical element $\gamma(l_1, \dots, l_{d-1}, i)$, $i = -p_{L_d}, \dots, p_{U_d}$. Next, for $l_i = -p_{L_i}, \dots, p_{U_i}$, $i = 2, \dots, d - 1$ define $\check{\psi}_{l_2, \dots, l_{d-1}}^{(2)}(p)$

to be the nested vector of (nested) dimension $(p_{U_1} + 1)$ and i -th block $\check{\psi}_{i,l_2,\dots,l_{d-1}}^{(1)}(p)$, $i = 0, \dots, p_{U_1}$. Proceeding in this manner, for $l_{d-1} = -p_{L_{d-1}}, \dots, p_{U_{d-1}}$ we define $\check{\psi}_{l_{d-1}}^{(d-1)}(p)$ to be the nested vector of (nested) dimension $(p_{d-2} + 1) \times 1$ and i -th block $\check{\psi}_{i,l_{d-1}}^{(d-2)}(p)$, $i = -p_{L_{d-2}}, \dots, p_{U_{d-2}}$. Finally, define $\check{\psi}_{\mathfrak{h}(p)}$ to be the nested vector of (nested) dimension $(p_d + 1)$ and i -th block $\check{\psi}_i^{(d-1)}(p)$, $i = -p_{L_{d-1}}, \dots, p_{U_{d-1}}$. Now $\check{\psi}_{\mathfrak{h}(p)}$ is $\prod_{i=1}^d (p_i + 1) \times 1$ where we note that $p_{L_1} = 0$, so $\prod_{i=1}^d (p_i + 1) = \mathfrak{h}(p) + p_{L_d} + p_{L_{d-1}}(p_d + 1) + \dots + p_{L_2}(p_3 + 1) \dots (p_d + 1) + 1$. Define $\psi_{\mathfrak{h}(p)}$ as the $\mathfrak{h}(p) \times 1$ vector formed using the following procedure:

- (1) Delete each of $\check{\psi}_{0,l_2,\dots,l_{d-1}}^{(1)}(p)$, $l_2 = -p_{L_2}, \dots, -1$ and $l_i = -p_{L_i}, \dots, p_{U_i}$, $i = 3, \dots, d-1$.
- (2) Delete each of $\check{\psi}_{0,l_3,\dots,l_{d-1}}^{(2)}(p)$, $l_3 = -p_{L_3}, \dots, -1$ and $l_i = -p_{L_i}, \dots, p_{U_i}$, $i = 4, \dots, d-1$.
- \vdots
- $(d-2)$ Delete each of the $\check{\psi}_{0,l_{d-1}}^{(d-2)}(p)$, $l_{d-1} = -p_{L_{d-1}}, \dots, -1$.
- $(d-1)$ Delete the first $p_{L_d} + 1$ elements of $\check{\psi}_0^{(d-1)}(p)$.

Thus $p_{L_2}(p_3 + 1) \dots (p_d + 1) + \dots + p_{L_{d-1}}(p_d + 1) + p_{L_d} + 1$ elements are deleted, and the dimension of $\psi_{\mathfrak{h}(p)}$ is $\mathfrak{h}(p) \times 1$. By construction $\psi_{\mathfrak{h}(p)}$ has elements $\gamma(s)$, $s \in S[-p_L, p_U]$. We now define the matrices. For $l_1 = 0, \dots, p_{U_1}$ and $l_i = -p_{L_i}, \dots, p_{U_i}$, $i = 2, \dots, d-1$, define $\check{\Psi}_{l_1,\dots,l_{d-1}}^{(1)}(p)$ to be the $(p_d + 1)$ -dimensional Toeplitz matrix with typical (i, j) -th element $\gamma(l_1, \dots, l_{d-1}, i - j)$, $i, j = -p_{L_d}, \dots, p_{U_d}$. Next, for $l_i = -p_{L_i}, \dots, p_{U_i}$, $i = 2, \dots, d-1$ define $\check{\Psi}_{l_2,\dots,l_{d-1}}^{(2)}(p)$ to be the block Toeplitz with Toeplitz blocks matrix of (nested) dimension $(p_{U_1} + 1)$ and (i, j) -th block $\check{\Psi}_{i-j,l_2,\dots,l_{d-1}}^{(1)}(p)$, $i, j = 0, \dots, p_{U_1}$. Proceeding in this manner, for $l_{d-1} = -p_{L_{d-1}}, \dots, p_{U_{d-1}}$ we define $\check{\Psi}_{l_{d-1}}^{(d-1)}(p)$ to be the nested block-Toeplitz matrix of (block) dimension $(p_{d-2} + 1) \times (p_{d-2} + 1)$ and (i, j) -th block $\check{\Psi}_{i-j,l_{d-1}}^{(d-2)}(p)$, $i, j = -p_{L_{d-2}}, \dots, p_{U_{d-2}}$. Finally, define $\check{\Psi}_{\mathfrak{h}(p)}$ to be the block-Toeplitz matrix of (block) dimension $(p_{d-1} + 1) \times (p_{d-1} + 1)$ and (i, j) -th block $\check{\Psi}_{i-j}^{(d-1)}(p)$, $i, j = -p_{L_{d-1}}, \dots, p_{U_{d-1}}$. So in this (most general case) case we obtain the general form of the covariance matrix as

$$\check{\Psi}_{\mathfrak{h}(p)} = \begin{pmatrix} \check{\Psi}_0^{(d-1)}(p) & \check{\Psi}_{-1}^{(d-1)}(p) & \dots & \check{\Psi}_{-p_{d-1}}^{(d-1)}(p) \\ \check{\Psi}_1^{(d-1)}(p) & \check{\Psi}_0^{(d-1)}(p) & \dots & \check{\Psi}_{-p_{d-1}+1}^{(d-1)}(p) \\ \vdots & \vdots & \ddots & \vdots \\ \check{\Psi}_{p_{d-1}}^{(d-1)}(p) & \check{\Psi}_{p_{d-1}-1}^{(d-1)}(p) & \dots & \check{\Psi}_0^{(d-1)}(p) \end{pmatrix}.$$

Now denote by $\Psi_{\mathfrak{h}(p)}$ the matrix formed by deleting those rows and columns of $\check{\Psi}_{\mathfrak{h}(p)}$ corresponding to the elements deleted from $\check{\psi}_{\mathfrak{h}(p)}$ above. Then the dimension of $\Psi_{\mathfrak{h}(p)}$ is $\mathfrak{h}(p) \times \mathfrak{h}(p)$.

We can straightforwardly extend a representation for $\Psi_{\mathfrak{h}(p)}^{-1}$ given for $d = 1$ by [Akaike \(1969\)](#) and [Kromer \(1970\)](#). Label the indices of the elements of the first row of $\Psi_{\mathfrak{h}(p)}$

from left to right as $j_0, \dots, j_{\mathfrak{h}(p)-1}$, $j_0 \equiv \mathbf{0}$. Define $\Sigma_{\mathfrak{h}(p)} = \text{diag}(\sigma_0^2, \dots, \sigma_{\mathfrak{h}(p)-1}^2)$, with σ_l^2 , $l = 0, \dots, \mathfrak{h}(p) - 1$, defined as in (2.13). The lag indices in the predictor for a generic l are defined by the first l indices in the first row of $\Psi_{\mathfrak{h}(p)}$. Defining

$$L_{\mathfrak{h}(p)} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ d_{j_1,1} & 1 & 0 & \dots & 0 \\ & & \dots & & \\ d_{j_{\mathfrak{h}(p)-1}, \mathfrak{h}(p)-1} & \dots & & d_{j_{\mathfrak{h}(p)-1}, \mathfrak{h}(p)-1} & 1 \end{pmatrix}, \quad (\text{B.1})$$

we have

$$\Psi_{\mathfrak{h}(p)}^{-1} = L'_{\mathfrak{h}(p)} \Sigma_{\mathfrak{h}(p)}^{-1} L_{\mathfrak{h}(p)}. \quad (\text{B.2})$$

B.2 Counting covariances in stationary and unilateral lattice autoregressive models

Autoregressive models on d -dimensional lattices can generate covariance matrices of the form $\Psi_{\mathfrak{h}(p)}$ which differ from those in the time series case in the number of unique covariances amongst their elements. Consider a stationary time series x_t with an $\text{AR}(k)$ (here $\mathfrak{h}(k) = k$) representation $x_t = \sum_{j=1}^k a_j x_{t-j} + \epsilon_t$ for which Ψ_k is a Toeplitz matrix with k unique autocovariances, which is also the dimension of the matrix. On the other hand, consider a 2-dimensional lattice process x_t with an $\text{AR}(0, 1; 1, 1)$ representation. In this case

$$\Psi_{\mathfrak{h}(0,1;1,1)} = \begin{pmatrix} \gamma(0,0) & \gamma(-1,0) & \gamma(-1,2) & \gamma(-1,1) \\ & \gamma(0,0) & \gamma(0,2) & \gamma(0,1) \\ & & \gamma(0,0) & \gamma(0,-1) \\ & & & \gamma(0,0) \end{pmatrix},$$

which is a 4×4 matrix with 6 unique covariances. While the above may suggest that the number of unique covariances in such matrices is $\prod_{i=1}^d (p_i + 1)$, this is in fact incorrect as the following example shows. A 2-dimensional lattice process x_t with an $\text{AR}(0, 2; 1, 1)$ representation has $\Psi_{\mathfrak{h}(0,2;1,1)}$ given by

$$\begin{pmatrix} \gamma(0,0) & \gamma(-1,0) & \gamma(-2,0) & \gamma(-1,2) & \gamma(-2,2) & \gamma(-1,1) & \gamma(-2,1) \\ & \gamma(0,0) & \gamma(-1,0) & \gamma(0,2) & \gamma(-1,2) & \gamma(0,1) & \gamma(-1,1) \\ & & \gamma(0,0) & \gamma(1,2) & \gamma(0,2) & \gamma(1,1) & \gamma(0,1) \\ & & & \gamma(0,0) & \gamma(-1,0) & \gamma(0,-1) & \gamma(-1,-1) \\ & & & & \gamma(0,0) & \gamma(1,-1) & \gamma(-1,0) \\ & & & & & \gamma(0,0) & \gamma(-1,-1) \\ & & & & & & \gamma(0,0) \end{pmatrix},$$

which is a 7×7 matrix with 11 unique covariances, and the latter obviously does not equal $(p_1 + 1) \times (p_2 + 1) = 12$. We will provide an upper bound for the number of unique

covariances in $\check{\Psi}_{\mathfrak{h}(p)}$ for general d .

Proposition B.1. *Suppose that $\{x_t : t \in \mathbb{L}\}$ is a stationary random field with the representation (2.4). Then the number of unique covariances in $\check{\Psi}_{\mathfrak{h}(p)}$ does not exceed*

$$\mathfrak{C}(p) = 1 + \sum_{l=1}^{d-1} 2^{d-l-1} \sum_{\substack{\#(l=0) \\ \approx 0_d^l}} \prod_{k=1}^d p_k + 2^{d-1} \prod_{k=1}^d p_k, \quad (\text{B.3})$$

where $\sum_{\#(l=0)}$ sums over all the possible ways in which $(p_1, p_2, \dots, p_d)'$ can have l entries equal to 0 and the product $\prod_{k=1, \approx 0_d^l}^d$ multiplies over k such that the l zero entries of $(p_1, p_2, \dots, p_d)'$ are excluded.

The proof follows by counting combinations across dimensions and is omitted. Also, it is clear from the formulae (2.8) and (B.3) that

$$\mathfrak{h}(p) \leq \mathfrak{C}(p), \quad (\text{B.4})$$

for all d . We now illustrate the formula with examples. For $d = 1$ with $p_1 = k$ (an AR(k) specification) $\check{\Psi}_k$ is Toeplitz with first row $(\gamma(0), \dots, \gamma(k))$, and the formula (B.3) delivers a bound that holds with equality. For $d = 2$ the formula indicates a maximum of $1 + 2^0(p_1 + p_2) + 2^1 p_1 p_2 = 1 + p_1 + p_2 + 2p_1 p_2$ unique covariances, delivering bounds of 8 and 13 for the AR(0, 1; 1, 1) and AR(0, 2; 1, 1) models respectively, while for $d = 3$ there are at most $1 + 2^0(p_1 + p_2 + p_3) + 2^1(p_1 p_2 + p_1 p_3 + p_2 p_3) + 2^2 p_1 p_2 p_3$ unique covariances. If equal truncation lengths are chosen in each dimension, so that $p_{U_i} = p_{L_i} = p$ for each $i = 1, \dots, d$, we have $p_1 = p$ and $p_i = 2p$ for $i = 2, \dots, d$. Then the formulae become $1 + 3p + 4p^2$ and $1 + 5p + 20p^2 + 16p^3$ respectively.

C Proofs of theorems

Proof of Theorem 2.1. We have

$$\begin{aligned} \hat{d}_{\mathfrak{h}(p)} - d_{\mathfrak{h}(p)} &= \hat{\Psi}_{\mathfrak{h}(p)}^{-1} \left(\hat{\psi}_{\mathfrak{h}(p)} - \hat{\Psi}_{\mathfrak{h}(p)} d_{\mathfrak{h}(p)} \right) \\ &= \hat{\Psi}_{\mathfrak{h}(p)}^{-1} \left(\delta_{\mathfrak{h}(p)} - \Delta_{\mathfrak{C}(p)} d_{\mathfrak{h}(p)} + \psi_{\mathfrak{h}(p)} - \Psi_{\mathfrak{h}(p)} d_{\mathfrak{h}(p)} \right) \end{aligned}$$

so that the norm of the LHS above is bounded by

$$\left\| \hat{\Psi}_{\mathfrak{h}(p)}^{-1} \right\| \left(\left\| \delta_{\mathfrak{h}(p)} \right\| + \left\| \Delta_{\mathfrak{C}(p)} \right\| \left\| d_{\mathfrak{h}(p)} \right\| + \left\| \Psi_{\mathfrak{h}(p)} d_{\mathfrak{h}(p)} - \psi_{\mathfrak{h}(p)} \right\| \right). \quad (\text{C.1})$$

Now $\left\| \hat{\Psi}_{\mathfrak{h}(p)}^{-1} \right\| \leq \left\| \hat{\Psi}_{\mathfrak{h}(p)}^{-1} - \Psi^{-1} \right\| + \left\| \Psi_{\mathfrak{h}(p)}^{-1} \right\| \leq \left(\left\| \hat{\Psi}_{\mathfrak{h}(p)}^{-1} \right\| \left\| \Delta_{\mathfrak{C}(p)} \right\| + 1 \right) \left\| \Psi_{\mathfrak{h}(p)}^{-1} \right\|$, so $\left\| \hat{\Psi}_{\mathfrak{h}(p)}^{-1} \right\| \left(1 - \left\| \Psi_{\mathfrak{h}(p)}^{-1} \right\| \left\| \Delta_{\mathfrak{C}(p)} \right\| \right) \leq \left\| \Psi_{\mathfrak{h}(p)}^{-1} \right\|$. Using Markov's inequality and Lemma 2.5 it

follows that $\|\Delta_{\mathfrak{C}(p)}\| \xrightarrow{p} 0$ if

$$\mathfrak{C}(p)^v n_p^{1-v} \rightarrow 0, \text{ i.e., } \mathfrak{C}(p)^v N^{1-v} \left(\prod_{i=1}^d \left(1 - \frac{p_i}{n_i} \right) \right)^{1-v} \rightarrow 0,$$

which is true by (2.14). Thus $\lim_{N \rightarrow \infty} \|\hat{\Psi}_{\mathfrak{h}(p)}^{-1}\| \leq \lim_{N \rightarrow \infty} \|\Psi_{\mathfrak{h}(p)}^{-1}\| < \infty$, from Corollary 2.7. Now we deal with the factor in parentheses in (C.1). By Lemma 2.4, Markov's inequality and (2.14), $\|\delta_{\mathfrak{h}(p)}\| \xrightarrow{p} 0$. For the second term, we have $\|\Delta_{\mathfrak{C}(p)}\| \xrightarrow{p} 0$ and also $\|d_{\mathfrak{h}(p)}\| = \left(\sum_{s \in S[-p_L, p_U]} d_s^2 \right)^{\frac{1}{2}} \leq \sum_{s \in S[-p_L, p_U]} |d_s| \leq \sum_{s \in S_{1+}^\infty} |d_s| < \infty$. Thus the second term converges to zero in probability. Finally, for the third term note that (2.2) implies that $E\epsilon_t x_{t-k} = \sum_{s \in S_{1+}^\infty \cup \mathbf{0}} b_s E\epsilon_t \epsilon_{t-k-s} = 0$, $k \in S_{1+}^\infty$, $t \in \mathbb{L}$, because $k + s = 0$ is not possible due to our definition of half-plane (2.1). This indicates that $\gamma(k) = Ex_t x_{t-k} = \sum_{t \in S_{1+}^\infty} d_t \gamma(t-k)$, $k \in S_{1+}^\infty$, so $\|\Psi_{\mathfrak{h}(p)} d_{\mathfrak{h}(p)} - \psi_{\mathfrak{h}(p)}\|^2$ is

$$\begin{aligned} & \sum_{s \in S[-p_L, p_U]} \left(\sum_{t \in S[-p_L, p_U]} d_s \gamma(t-s) - \gamma(s) \right)^2 \\ &= \sum_{s \in S[-p_L, p_U]} \left(\sum_{t \in S[-p_L, p_U]} d_s \gamma(t-s) - \sum_{t \in S_{1+}^\infty} d_t \gamma(t-s) \right)^2 \\ &= \sum_{s \in S[-p_L, p_U]} \left(\sum_{t \in S_{1+}^\infty \setminus S[-p_L, p_U]} d_t \gamma(t-s) \right)^2 \\ &\leq \sum_{s \in S[-p_L, p_U]} \left(\sum_{t \in S_{1+}^\infty \setminus S[-p_L, p_U]} d_t^2 \right) \left(\sum_{t \in S_{1+}^\infty \setminus S[-p_L, p_U]} \gamma(t-s)^2 \right) \\ &= \left(\sum_{s \in S[-p_L, p_U]} \sum_{t \in S_{1+}^\infty \setminus S[-p_L, p_U]} \gamma(t-s)^2 \right) \left(\sum_{t \in S_{1+}^\infty \setminus S[-p_L, p_U]} d_t^2 \right) \\ &\leq C \sum_{s \in \mathbb{Z}^d} \gamma(s)^2 \sum_{t \in S_{1+}^\infty \setminus S[-p_L, p_U]} d_t^2 = C \sum_{t \in S_{1+}^\infty \setminus S[-p_L, p_U]} d_t^2, \end{aligned}$$

using Lemma 2.1. Thus $\|\Psi_{\mathfrak{h}(p)} d_{\mathfrak{h}(p)} - \psi_{\mathfrak{h}(p)}\| \leq C \sum_{t \in S_{1+}^\infty \setminus S[-p_L, p_U]} |d_t|$, which converges to zero as $N \rightarrow \infty$ due to (2.15), completing the proof. Note that we have also shown that

$$\|\hat{d}_{\mathfrak{h}(p)} - d_{\mathfrak{h}(p)}\| = \mathcal{O}_p \left(\frac{\mathfrak{C}(p)}{N^{\frac{v-1}{v}}} \right), \quad (\text{C.2})$$

by Markov's inequality. \square

Proof of Theorem 2.2. We first prove that $\|\Delta_{\mathfrak{C}(p)}\| \xrightarrow{a.s.} 0$ and $\|\delta_{\mathfrak{h}(p)}\| \xrightarrow{a.s.} 0$, as $N \rightarrow \infty$.

$E \|\Delta_{\mathfrak{C}(p)}\|^v$ is bounded (as in the proof of Lemma 2.5) by a constant times

$$\mathfrak{C}(p)^v N^{1-v} \leq C \left\{ (\log N)^{v+1} (\log \log N)^v \right\}^{-1} \leq C m^{-v}, \quad (\text{C.3})$$

by (2.16), so that $\|\Delta_{\mathfrak{C}(p)}\|$ converges completely to zero, and therefore almost surely. An identical proof holds for $\|\delta_{\mathfrak{h}(p)}\|$, whence the proof follows that of Theorem 2.1. \square

Proof of Theorem 2.3. Note that $\hat{\gamma}(\mathbf{0}) = n_p^{-1} \sum_{t(p,n)}'' x_t^2$. Using standard algebraic manipulation and the definition of least squares we may write $\hat{\sigma}_{\mathfrak{h}(p)}^2 - \sigma^2$ as

$$\begin{aligned} & n_p^{-1} \sum_{t(p,n)}'' \left(x_t - \sum_{s \in S[-p_L, p_U]} \hat{d}_{s, \mathfrak{h}(p)} x_{t-s} \right)^2 - \sigma^2 \\ &= \hat{\gamma}(\mathbf{0}) - \hat{d}'_{\mathfrak{h}(p)} \hat{\psi}_{\mathfrak{h}(p)} - \sigma^2 \\ &= \hat{\gamma}(\mathbf{0}) - \left(\hat{d}_{\mathfrak{h}(p)} - d_{\mathfrak{h}(p)} \right)' \hat{\psi}_{\mathfrak{h}(p)} - d'_{\mathfrak{h}(p)} \hat{\psi}_{\mathfrak{h}(p)} - \hat{\gamma}(\mathbf{0}) + \sum_{t \in S_{1+}^\infty} d_t \gamma(t) \\ &= \hat{\gamma}(\mathbf{0}) - \gamma(\mathbf{0}) - \left(\hat{d}_{\mathfrak{h}(p)} - d_{\mathfrak{h}(p)} \right)' \psi_{\mathfrak{h}(p)} - d'_{\mathfrak{h}(p)} \Delta_{\mathfrak{C}(p)} \\ &\quad - \left(\hat{d}_{\mathfrak{h}(p)} - d_{\mathfrak{h}(p)} \right)' \Delta_{\mathfrak{C}(p)} - d'_{\mathfrak{h}(p)} \psi_{\mathfrak{h}(p)} + \sum_{t \in S_{1+}^\infty} d_t \gamma(t). \end{aligned}$$

Since $d'_{\mathfrak{h}(p)} \psi_{\mathfrak{h}(p)} = \sum_{s \in S[-p_L, p_U]} d_s \gamma(s)$, we can write

$$\begin{aligned} \hat{\sigma}_{\mathfrak{h}(p)}^2 - \sigma^2 &= (\hat{\gamma}(\mathbf{0}) - \gamma(\mathbf{0})) - \left(\hat{d}_{\mathfrak{h}(p)} - d_{\mathfrak{h}(p)} \right)' \psi_{\mathfrak{h}(p)} - d'_{\mathfrak{h}(p)} \Delta_{\mathfrak{C}(p)} \\ &\quad - \left(\hat{d}_{\mathfrak{h}(p)} - d_{\mathfrak{h}(p)} \right)' \Delta_{\mathfrak{C}(p)} + \sum_{t \in S_{1+}^\infty \setminus S[-p_L, p_U]} d_t \gamma(t). \end{aligned}$$

The first term on the RHS converges to 0 in probability by Lemma 2.3 and Markov's inequality, the second by Theorem 2.1 and Lemma 2.1, the third term by Lemma 2.4, (2.14) and Assumption A and the fourth term by Theorem 2.1, Lemma 2.4 and (2.14). For the fifth term, convergence to zero follows by (2.15) and Lemma 2.1. Note that we have also proved

$$\hat{\sigma}_{\mathfrak{h}(p)}^2 - \sigma^2 = \mathcal{O}_p \left(\frac{\mathfrak{C}(p)}{N^{\frac{v-1}{v}}} + \sum_{t \in S_{1+}^\infty \setminus S[-p_L, p_U]} d_t^2 \right), \quad (\text{C.4})$$

because $\mathfrak{h}(p) \leq \mathfrak{C}(p)$ and $\lim_{N \rightarrow \infty} N/n_p = 1$. The almost sure convergence proof is similar and omitted. \square

Proof of Theorem 3.1. We recall $D(e^{i\lambda}) = 1 - \sum_{s \in S_{1+}^\infty} d_s e^{is'\lambda}$ and define $\hat{D}_{\mathfrak{h}(p)}(e^{i\lambda}) =$

$1 - \sum_{s \in S[-p_L, p_U]} \hat{d}_{s, \mathfrak{h}(p)} e^{is'\lambda}$. Then

$$\hat{f}_{\mathfrak{h}(p)}(\lambda) - f(\lambda) = \frac{\sigma^2 \left(\left| \hat{D}_{\mathfrak{h}(p)}(e^{i\lambda}) \right|^2 - |D(e^{i\lambda})|^2 \right) - |D(e^{i\lambda})|^2 (\hat{\sigma}_{\mathfrak{h}(p)}^2 - \sigma^2)}{(2\pi)^d |D(e^{i\lambda})|^2 \left| \hat{D}_{\mathfrak{h}(p)}(e^{i\lambda}) \right|^2}. \quad (\text{C.5})$$

Because $D(e^{i\lambda}) = \sigma^2 \{(2\pi)^d f(\lambda)\}^{-1}$, by (2.5) we have

$$c \leq D(e^{i\lambda}) \leq C, \text{ uniformly in } \lambda \in \Pi. \quad (\text{C.6})$$

On the other hand $\hat{D}_{\mathfrak{h}(p)}(e^{i\lambda}) = \hat{\sigma}^2(p) \{(2\pi)^d \hat{f}(\lambda)\}^{-1}$, so that

$$\sup_{\lambda \in \Pi} \left| \hat{D}_{\mathfrak{h}(p)}(e^{i\lambda}) \right| \leq \sup_{\lambda \in \Pi} \left| \hat{D}_{\mathfrak{h}(p)}(e^{i\lambda}) - D(e^{i\lambda}) \right| + \sup_{\lambda \in \Pi} |D(e^{i\lambda})| \quad (\text{C.7})$$

and

$$\inf_{\lambda \in \Pi} \left| \hat{D}_{\mathfrak{h}(p)}(e^{i\lambda}) \right| \geq \inf_{\lambda \in \Pi} |D(e^{i\lambda})| - \sup_{\lambda \in \Pi} \left| \hat{D}_{\mathfrak{h}(p)}(e^{i\lambda}) - D(e^{i\lambda}) \right|. \quad (\text{C.8})$$

$\left| \hat{D}_{\mathfrak{h}(p)}(e^{i\lambda}) - D(e^{i\lambda}) \right|$ is bounded by

$$\begin{aligned} & \sum_{s \in S[-p_L, p_U]} \left| \hat{d}_{s, \mathfrak{h}(p)} - d_s \right| \left| e^{is'\lambda} \right| + \sum_{s \in S_{1+}^\infty \setminus S[-p_L, p_U]} |d_s| \left| e^{is'\lambda} \right| \\ & \leq \mathfrak{h}(p)^{\frac{1}{2}} \left\| \hat{d}_{\mathfrak{h}(p)} - d_{\mathfrak{h}(p)} \right\| + \sum_{s \in S_{1+}^\infty \setminus S[-p_L, p_U]} |d_s|, \end{aligned} \quad (\text{C.9})$$

by the Cauchy Schwarz inequality. By (3.1) and (3.2), we conclude from (C.2) that

$$\mathfrak{h}(p)^{\frac{1}{2}} \left\| \hat{d}_{\mathfrak{h}(p)} - d_{\mathfrak{h}(p)} \right\| = \mathcal{O}_p \left(\frac{\mathfrak{C}(p) \mathfrak{h}(p)^{\frac{1}{2}}}{N^{\frac{v-1}{v}}} \right),$$

implying that (C.9) is negligible. We have then shown that

$$\sup_{\lambda \in \Pi} \left| \hat{D}_{\mathfrak{h}(p)}(e^{i\lambda}) - D(e^{i\lambda}) \right| \xrightarrow{p} 0. \quad (\text{C.10})$$

Using (C.6), (C.7) and (C.8) together with (C.10) implies that

$$c \leq \hat{D}_{\mathfrak{h}(p)}(e^{i\lambda}) \leq C, \text{ uniformly in } \lambda \in \Pi, \quad (\text{C.11})$$

with probability approaching 1 as $N \rightarrow \infty$. The identity $a^2 - b^2 = (a - b)^2 + 2b(a - b)$

implies $\left| \left| \hat{D}_{\mathfrak{h}(p)}(e^{i\lambda}) \right|^2 - |D(e^{i\lambda})|^2 \right|$ is bounded by

$$\left(\hat{D}_{\mathfrak{h}(p)}(e^{i\lambda}) - D(e^{i\lambda}) \right)^2 + 2 \left| D(e^{i\lambda}) \right| \left| \hat{D}_{\mathfrak{h}(p)}(e^{i\lambda}) - D(e^{i\lambda}) \right|, \quad (\text{C.12})$$

where the RHS converges to 0 in probability uniformly in λ by (C.10) and (C.11) so that

$$\sup_{\lambda \in \Pi} \left| \left| \hat{D}_{\mathfrak{h}(p)}(e^{i\lambda}) \right|^2 - |D(e^{i\lambda})|^2 \right| \xrightarrow{p} 0. \quad (\text{C.13})$$

Because (C.5) implies that

$$\left| \hat{f}_{\mathfrak{h}(p)}(\lambda) - f(\lambda) \right| \leq \frac{\sigma^2 \left| \left| \hat{D}_{\mathfrak{h}(p)}(e^{i\lambda}) \right|^2 - |D(e^{i\lambda})|^2 \right| + |D(e^{i\lambda})|^2 \left| \hat{\sigma}_{\mathfrak{h}(p)}^2 - \sigma^2 \right|}{(2\pi)^d |D(e^{i\lambda})|^2 \left| \hat{D}_{\mathfrak{h}(p)}(e^{i\lambda}) \right|^2},$$

the theorem now follows by (C.6), (C.11), (C.13) and Theorem 2.3. \square

Proof of Theorem 4.1. By Lemma 4.1 and (2.2), we need to establish the asymptotic distribution of

$$\left(N^{\frac{1}{2}} / n_p \mathfrak{h}(p)^{\frac{1}{2}} \right) \sum_{r \in S_{1+}^{\infty} \cup \mathbf{0}} b_r \sum_{s \in S[-p_L, p_U]} \alpha(p)' \Psi_{\mathfrak{h}(p)}^{(s)} \sum_{t(p,n)}'' \epsilon_{t-r-s} \epsilon_t, \quad (\text{C.14})$$

with $\Psi_{\mathfrak{h}(p)}^{(s)}$ denoting a typical column of $\Psi_{\mathfrak{h}(p)}$. Fixing $\eta > 0$, in view of (2.2) we can choose a positive integer M such that

$$\sum_{r \notin S[-M, M]} b_r < \eta / \mathfrak{h}(p)^{\frac{1}{2}}, \quad (\text{C.15})$$

where $S[-M, M] = \{t_i : |t_i| \leq M, i = 1, \dots, d\}$. Note that $r \notin S[-M, M]$ if and only if $r \in S_{1+}^{\infty} \setminus S[-M, M]$. The difference between (C.14) and

$$g_{\mathfrak{h}(p), M} = \left(N^{\frac{1}{2}} / n_p \mathfrak{h}(p)^{\frac{1}{2}} \right) \sum_{r \in S[-M, M]} b_r \sum_{s \in S[-p_L, p_U]} \alpha(p)' \Psi_{\mathfrak{h}(p)}^{(s)} \sum_{t(p,n)}'' \epsilon_{t-r-s} \epsilon_t \quad (\text{C.16})$$

is readily shown to have mean zero and variance that is $\mathcal{O}(\eta^2 N n_p^{-1}) = o(1)$, as $\eta \rightarrow 0$, because $N/n_p = \mathcal{O}(1)$. Thus we establish asymptotic normality of $g_{\mathfrak{h}(p), M}$. A martingale central limit theorem of Scott (1973) can be applied by mapping \mathbb{Z}^d into \mathbb{Z}_+ , as in Robinson and Vidal Sanz (2006). They denote by $C_k^{(d)}$ the lattice points of on the surface of the d -dimensional cube with vertices $(\pm k, \dots, \pm k)$, and arbitrarily order them as $t_{(1)}^{(k)}, \dots, t_{m_k^{(d)}}^{(k)}$, with $m_k^{(d)} = (2k+1)^d - (2k-1)^d$. Introduce the function $\phi : \mathbb{Z}^d \rightarrow \mathbb{Z}_+$,

defined as

$$\begin{aligned}
\phi(\mathbf{0}) &= 1 \\
\phi\left(t_{(1)}^{(1)}\right) &= 2, \dots, \phi\left(t_{3^d-1}^{(1)}\right) = 3^d \\
&\vdots \\
\phi\left(t_{(1)}^{(k)}\right) &= (2k-1)^d + 1, \dots, \phi\left(t_{m_k^{(d)}}^{(k)}\right) = (2k+1)^d,
\end{aligned}$$

and $\theta_N(t) = \phi(t) - \#\{s : s \notin \mathbb{L}; \phi(s) < \phi(t)\}$, $t \in \mathbb{L}$. Having thus ordered on the integer vertices of a hypercube containing \mathbb{L} , we drop points outside \mathbb{L} and re-label after closing gaps and preserving order. Now define the triangular array $\delta_N(j)$, $j = 1, \dots, N$, of iid random variables with zero mean, variance σ^2 and finite fourth moment by $\delta_N(\theta_N(t)) = \epsilon_t$, $t \in \mathbb{L}$. For each summand in $\sum_{t(p,n)} \epsilon_{t-r-s} \epsilon_t$ either $\phi(t-r-s) < \phi(t)$ or $\phi(t-r-s) > \phi(t)$, and there are a total of $N - \mathcal{O}(N^{1-\chi})$ summands, each of which can be written as $\delta_N(j) \delta_N(j - \ell_{j,N}(s, r))$ for suitable j and $\ell_{j,N}(s, r) \in \mathbb{Z}_+$ (possibly after finite translation across \mathbb{Z}^d). Define

$$v_N(j) = \left(N/n_p \mathfrak{h}(p)^{\frac{1}{2}}\right) \sum_{r \in S[-M, M]} b_r \sum_{s \in S[-p_L, p_U]} \alpha(p)' \Psi_{\mathfrak{h}(p)}^{(s)} \delta_N(j) \delta_N(j - \ell_{j,N}(s, r));$$

thus by uncorrelatedness of $v_N(j)$ over j , $g_{\mathfrak{h}(p), M}$ differs by $\mathcal{O}_p\left(\mathfrak{h}(p) N^{-\frac{\chi}{2}}\right) = o_p(1)$ from $N^{-\frac{1}{2}} \sum_{j=1}^N v_N(j)$. We now show that

$$\lim_{N \rightarrow \infty} N^{-1} \sum_{j=1}^N E v_N^2(j) = \sigma^2 \mu + o(\eta). \quad (\text{C.17})$$

The uncorrelatedness and identity of distribution of $\delta_N(j)$ implies that

$$E v_N^2(j) = (N^2 \sigma^2 / n_p^2 \mathfrak{h}(p)) \alpha(p)' \Psi_{\mathfrak{h}(p)}^{-1} \Psi_{\mathfrak{h}(p), M} \Psi_{\mathfrak{h}(p)}^{-1} \alpha(p), \quad \text{any } j,$$

where, with $s, t \in S[-p_L, p_U]$, $\Psi_{\mathfrak{h}(p), M}$ denotes the symmetric matrix with elements $\sigma^2 \sum_{r \in S[-M, M]} b_r b_{r+s-t}$. Elementary inequalities together with (C.15) imply that the latter differ from a typical element of $\Psi_{\mathfrak{h}(p)}$ by $\sigma^2 \sum_{r \notin S[-M, M]} b_r b_{r+s-t} = \mathcal{O}(\eta^2 / \mathfrak{h}(p))$, whence

$$\|\Psi_{\mathfrak{h}(p), M} - \Psi_{\mathfrak{h}(p)}\| = \mathcal{O}(\eta^2) = o(\eta), \quad \text{as } N \rightarrow \infty. \quad (\text{C.18})$$

Now $N^{-1} \sum_{j=1}^N E v_N^2(j)$ is bounded by

$$(N^2 \sigma^2 / n_p^2 \mathfrak{h}(p)) \|\alpha(p)\|^2 \left\| \Psi_{\mathfrak{h}(p)}^{-1} \right\|^2 \|\Psi_{\mathfrak{h}(p), M} - \Psi_{\mathfrak{h}(p)}\| + \sigma^2 \mu + o(1),$$

because $N/n_p \rightarrow 1$ as $N \rightarrow \infty$. The first term on the RHS above is easily seen to be $o(\eta)$ as $N \rightarrow \infty$, by (C.18). Thus (C.17) is established. The $v_N(j)$ form a martingale

difference array. Denote by $\mathcal{F}_{k,N}$ the σ -field of events generated by $\delta_N(j)$, $j \leq k$. Writing $u_N(j) = v_N(j)/\sigma\mu^{\frac{1}{2}}$, Theorem 2 of Scott (1973) implies that if

$$N^{-1} \sum_{j=1}^N E \left\{ u_N^2(j) \mathbf{1} \left(|u_N(j)| \geq \varrho N^{\frac{1}{2}} \right) \right\} \rightarrow 0, \quad \text{all } \varrho > 0, \quad (\text{C.19})$$

$$N^{-1} \sum_{j=1}^N [E \{ u_N^2(j) | \mathcal{F}_{j-1,N} \} - E u_N^2(j)] \xrightarrow{p} 0, \quad (\text{C.20})$$

then $N^{-\frac{1}{2}} \sum_{j=1}^N v_N(j) \xrightarrow{d} N(0, \sigma^2 \mu)$.

By (C.17), $E(u_N^2(j)) = \sigma^{-2} \mu^{-1} (\sigma^2 \mu + o(\eta) + o(1)) = O(1)$ uniformly in j , implying that $N^{-1} u_N^2(j)$ is a uniformly integrable array, whence (C.19) follows on noticing that the LHS of this is bounded above by $\max_{j=1, \dots, N} E \{ u_N^2(j) \mathbf{1} (u_N^2(j) \geq \varrho^2 N) \}$. Next, (C.20) is proved if we show

$$\begin{aligned} & N^{-1} \sum_{j=1}^N \left[\left\{ \mathfrak{h}(p)^{-\frac{1}{2}} \sum_{r \in S[-M, M]} b_r \sum_{s \in S[-p_L, p_U]} \alpha(p)' \Psi_{\mathfrak{h}(p)}^{(s)} \delta_N(j - \ell_{j,N}(s, r)) \right\}^2 \right. \\ & \left. - E \left\{ \mathfrak{h}(p)^{-\frac{1}{2}} \sum_{r \in S[-M, M]} b_r \sum_{s \in S[-p_L, p_U]} \alpha(p)' \Psi_{\mathfrak{h}(p)}^{(s)} \delta_N(j - \ell_{j,N}(s, r)) \right\}^2 \right] \xrightarrow{p} 0. \end{aligned} \quad (\text{C.21})$$

Fix $s_{(i)} \in S[-M, M]$ and $r_{(i)} \in S[-p_L, p_U]$, $i = 1, 2$, define

$\ell_{j,N,i} = \ell_{j,N}(s_{(i)}, r_{(i)})$ and consider

$$N^{-1} \sum_{j=1}^N \{ \delta_N(j - \ell_{j,N,1}) \delta_N(j - \ell_{j,N,2}) - E \delta_N(j - \ell_{j,N,1}) \delta_N(j - \ell_{j,N,2}) \}. \quad (\text{C.22})$$

Clearly (C.22) has mean zero, while its variance is

$$\begin{aligned} & N^{-2} \sum_{j=1}^N \sum_{k=1}^N [E \delta_N(j - \ell_{j,N,1}) \delta_N(k - \ell_{k,N,1}) E \delta_N(j - \ell_{j,N,2}) \delta_N(k - \ell_{k,N,2}) \\ & + E \delta_N(j - \ell_{j,N,1}) \delta_N(k - \ell_{k,N,2}) E \delta_N(j - \ell_{j,N,2}) \delta_N(k - \ell_{k,N,1}) \\ & + \text{cum} \{ \delta_N(j - \ell_{j,N,1}), \delta_N(k - \ell_{k,N,1}), \delta_N(j - \ell_{j,N,2}), \delta_N(k - \ell_{k,N,2}) \}], \end{aligned} \quad (\text{C.23})$$

where $\text{cum} \{x, y, z, w\}$ denotes the joint cumulant of x, y, z, w . Robinson and Vidal Sanz (2006) noted that, for $d > 1$, the $s \neq t$ terms have a non-zero contribution to (C.23) because $\ell_{j,N,i}$ depend on N . They show that (C.23) is $\mathcal{O}(N^{-\frac{\kappa}{2}})$, whence (C.21) is $\mathcal{O}(\mathfrak{h}(p) N^{-\frac{\kappa}{4}}) = o(1)$, unlike when $d = 1$, when (C.23) is $\mathcal{O}(N^{-1})$ and (C.21) is $\mathcal{O}(\mathfrak{h}(p) N^{-\frac{1}{2}})$ (cf Berk (1974)). The theorem now follows by Bernstein's Lemma (see

e.g. Hannan (1970) pg. 242). \square

Proof of Theorem 4.2. By (C.4), (3.1) and (4.4), $(N/\mathfrak{h}(p))^{\frac{1}{2}} \left(\hat{\sigma}_{\mathfrak{h}(p)}^2 - \sigma^2 \right) = o_p(1)$. Because $\hat{f}_{\mathfrak{h}(p)}(\lambda) = \hat{\sigma}_{\mathfrak{h}(p)}^2 (2\pi)^{-d} \left(\hat{C}_{\mathfrak{h}(p)}(\lambda)^2 + \hat{S}_{\mathfrak{h}(p)}(\lambda)^2 \right)^{-1}$ the proof is standard by Lemma 4.4 and the delta method, so we omit the details. \square

D Proofs of lemmas

Proof of Lemma 2.1. Standard. \square

Proof of Lemma 2.2. The result follows from Lemma A.1 taking $N = n$, $M = k$, $q = 2$ and $a_t = 1$ for all $t \in \mathbb{L}$. \square

Proof of Lemma 2.3. For $\hat{\gamma}(k) - \gamma(k)$ to be of the form of S_{pn} in Lemma 2.2, define $\xi_{rs,t} = b_r b_{r-k} (\epsilon_{t-r}^2 - \sigma^2)$, $s = r - k$; $= b_r b_s \epsilon_{t-r} \epsilon_{t-k-s}$, $s \neq r - k$. Then the $\xi_{rs,t}$ are clearly zero-mean. They are independent because the ϵ_t are. Therefore, they satisfy Assumption D. By the c_r -inequality, Cauchy-Schwarz inequality and Assumption B,

$$\begin{aligned} E |\xi_{rs,t}|^v &\leq 2 |b_r b_{r-k}|^v \left(E |\epsilon_{t-r}|^{2v} + \sigma^{2v} \right) \leq C |b_r b_{r-k}|^v, s = r - k, \\ E |\xi_{rs,t}|^v &\leq |b_r b_s|^v \left(E |\epsilon_{t-r}|^{2v} E |\epsilon_{t-s}|^{2v} \right)^{\frac{1}{2}} \leq C |b_r b_s|^v, s \neq r - k, \end{aligned}$$

verifying that (2.10) holds since the b_r are absolutely summable. The result follows immediately from Lemma 2.2. \square

Proof of Lemma 2.4.

$$\begin{aligned} E \|\delta_{\mathfrak{h}(p)}\|^v &\leq E \left(\sum_{s \in S[-p_L, p_U]} |\hat{\gamma}(s) - \gamma(s)| \right)^v \\ &\leq \mathfrak{h}(p)^{v-1} \sum_{s \in S[-p_L, p_U]} E |\hat{\gamma}(s) - \gamma(s)|^v \\ &\leq C \mathfrak{h}(p)^{v-1} \sum_{s \in S[-p_L, p_U]} n_p^{1-v} = C \mathfrak{h}(p)^v n_p^{1-v}, \end{aligned}$$

using Hölder's inequality and Lemma 2.3. \square

Proof of Lemma 2.5. Write $\check{\Delta}_{\mathfrak{C}(p)} = \hat{\Psi}_{\mathfrak{h}(p)} - \check{\Psi}_{\mathfrak{h}(p)}$, where $\hat{\Psi}_{\mathfrak{h}(p)}$ is constructed in the obvious way using estimated covariances. Using the inequality $\|B\| \leq \|B\|_R$ for symmetric matrices B , we have

$$\|\Delta_{\mathfrak{C}(p)}\| \leq \|\Delta_{\mathfrak{C}(p)}\|_R \leq \|\check{\Delta}_{\mathfrak{C}(p)}\|_R. \quad (\text{D.1})$$

We will now bound the absolute row-sums of $\check{\Delta}_{\mathfrak{C}(p)}$ uniformly over all rows. Consider a typical row of $\check{\Delta}_{\mathfrak{C}(p)}$. This consists of

$$\hat{\gamma}(l_1 - \bar{l}_1, l_2 - \bar{l}_2, \dots, l_d - j_d) - \gamma(l_1 - \bar{l}_1, l_2 - \bar{l}_2, \dots, l_d - j_d); \quad j_d = 0, \dots, p_d,$$

for some l_1, \dots, l_d , $l_i = 0, \dots, p_i$ and all $\bar{l}_1, \dots, \bar{l}_{d-1}$, $\bar{l}_i = 0, \dots, p_i$. It follows that a typical absolute row sum is

$$\sum_{d-1} \sum_{j_d=0}^{p_d} |\hat{\gamma}(l_1 - \bar{l}_1, l_2 - \bar{l}_2, \dots, l_d - j_d) - \gamma(l_1 - \bar{l}_1, l_2 - \bar{l}_2, \dots, l_d - j_d)| \quad (\text{D.2})$$

with \sum_{d-1} running over $\bar{l}_1, \dots, \bar{l}_{d-1}$, $\bar{l}_i = 0, \dots, p_i$. Since the summands are absolute values of the elements of a row of a Toeplitz matrix (by construction), (D.2) is bounded by

$$2 \sum_{d-1} \sum_{k_d=-p_d}^{p_d} |\hat{\gamma}(l_1 - \bar{l}_1, l_2 - \bar{l}_2, \dots, k_d) - \gamma(l_1 - \bar{l}_1, l_2 - \bar{l}_2, \dots, k_d)|$$

which in turn is bounded by

$$2 \sum_{\text{unique covariances}} |\hat{\gamma}(k) - \gamma(k)|,$$

there being $\mathfrak{C}(p)$ terms in the sum by Proposition B.1. This bound is clearly uniform over all possible rows. So using Hölder's inequality and Lemma 2.3

$$\begin{aligned} E \|\check{\Delta}_{\mathfrak{C}(p)}\|_R^v &\leq 4^v E \left(\sum_{\text{unique covariances}} |\hat{\gamma}(k) - \gamma(k)| \right)^v \\ &\leq 8 \mathfrak{C}(p)^{1-v} \sum_{\text{unique covariances}} E |\hat{\gamma}(k) - \gamma(k)|^v \\ &\leq C \mathfrak{C}(p)^{1-v} \sum_{\text{unique covariances}} n_p^{1-v} = C \mathfrak{C}(p)^v n_p^{1-v}. \end{aligned}$$

Then the result follows from the above and (D.1). □

Proof of Lemma 2.6. Eigenvalues of $\Psi_{\mathfrak{h}(p)}$ are determined by the generalized Toeplitz form $\sum_{j,k \in S[-p_L, p_U]} \xi_j \gamma(j-k) \xi_k$, for real numbers ξ_s , $s \in S[-p_L, p_U]$, $\sum_{s \in S[-p_L, p_U]} \xi_s^2 = 1$, summing over $j, k \in S[-p_L, p_U]$ by construction of $\Psi_{\mathfrak{h}(p)}$. This equals

$$\sum_{j,k \in S[-p_L, p_U]} \int_{\Pi} e^{i(j-k)'\lambda} f(\lambda) d\lambda \quad \xi_j \xi_k = \int_{\Pi} \left| \sum_{j \in S[-p_L, p_U]} e^{ij'\lambda} \xi_j \right|^2 f(\lambda) d\lambda$$

$$\begin{aligned}
&\in \left| \sum_{j \in S[-p_L, p_U]} \xi_j \right|^2 \left[m \int_{\Pi} d\lambda, M \int_{\Pi} d\lambda \right] = \sum_{j \in S[-p_L, p_U]} \xi_j^2 \left[m \int_{\Pi} d\lambda, M \int_{\Pi} d\lambda \right] \\
&= \left[(2\pi)^d m, (2\pi)^d M \right],
\end{aligned}$$

using $\gamma(j-k) = \int_{\Pi} e^{i(j-k)'\lambda} f(\lambda) d\lambda$ and (2.5). \square

Proof of Corollary 2.7. If $\|\Psi_{\mathfrak{h}(p)}^{-1}\|$ exists, it is the reciprocal of the smallest eigenvalue, say μ , of $\Psi_{\mathfrak{h}(p)}$. Using Lemma 2.6 we get $\|\Psi_{\mathfrak{h}(p)}^{-1}\| = \mu^{-1} \leq (2\pi)^{-d} m^{-1} \leq C$. \square

Proof of Lemma 4.1. Define $\bar{\epsilon}_{t, \mathfrak{h}(p)} = x_t - \sum_{s \in S[-p_L, p_U]} d_s x_{t-s}$. Then

$$\bar{\epsilon}_{t, \mathfrak{h}(p)} - \epsilon_t = \sum_{s \in S_{1+}^{\infty} \setminus S[-p_L, p_U]} d_s x_{t-s},$$

so that the LHS of (4.2) equals

$$\begin{aligned}
&N^{\frac{1}{2}} \alpha(p)' \hat{\Psi}_{\mathfrak{h}(p)}^{-1} \Delta_{\mathfrak{C}(p)} \Psi_{\mathfrak{h}(p)}^{-1} \sum_{t(p, n)} "X_t(p) \bar{\epsilon}_{t, \mathfrak{h}(p)} / n_p \mathfrak{h}(p)^{\frac{1}{2}} \\
&+ N^{\frac{1}{2}} \alpha(p)' \Psi_{\mathfrak{h}(p)}^{-1} \sum_{t(p, n)} "X_t(p) \sum_{s \in S_{1+}^{\infty} \setminus S[-p_L, p_U]} d_s x_{t-s} / n_p \mathfrak{h}(p)^{\frac{1}{2}}. \quad (\text{D.3})
\end{aligned}$$

Now, $\alpha(p)' \Psi_{\mathfrak{h}(p)}^{-1} X_t(p) / \mathfrak{h}(p)^{\frac{1}{2}}$ is a linear process in lags of ϵ_t , with mean 0 and variance $\mathfrak{h}(p)^{-1} \alpha(p)' \Psi_{\mathfrak{h}(p)}^{-1} \alpha(p) = \mathcal{O}(1)$, by Lemma 2.7. Thus the square of the second term in (D.3) has expectation bounded by a constant times $N^{\frac{1}{2}} n_p^{\frac{1}{2}} \sum_{s \in S_{1+}^{\infty} \setminus S[-p_L, p_U]} d_s^2 \rightarrow 0$, by Lemma 2 of Berk (1974), which also implies that

$$E \left\| \sum_{t(p, n)} "X_t(p) \bar{\epsilon}_{t, \mathfrak{h}(p)} \right\|^2 = \mathcal{O} \left(\mathfrak{h}(p) N^{\frac{1}{2}} n_p^{\frac{1}{2}} \sum_{s \in S_{1+}^{\infty} \setminus S[-p_L, p_U]} d_s^2 \right),$$

so the first term in (D.3) is

$$\mathcal{O}_p \left(\mathfrak{h}(p)^{\frac{1}{2}} \|\Delta_{\mathfrak{C}(p)}\| N^{\frac{1}{4}} n_p^{\frac{1}{4}} \sum_{s \in S_{1+}^{\infty} \setminus S[-p_L, p_U]} |d_s| \right) = \mathcal{O}_p \left(\mathfrak{h}(p)^{\frac{1}{2}} \mathfrak{C}(p) / n_p^{\frac{1}{2}} \right) o_p(1),$$

by Lemmas 2.5, 2.7, (4.1), and is negligible by (3.1). \square

Proof of Lemma 4.2. We can take $\lambda = \mathbf{0}$ in Theorem 2.2 of Baxter (1962), as in Berk (1974), and obtain

$$\sum_{r \in S[-p_L, p_U] \cup \mathbf{0}} \left| d_{r, \mathfrak{h}(p)} / \sigma_{\mathfrak{h}(p)}^2 - d_r / \sigma^2 \right| \leq C \sum_{r \in S_{1+}^{\infty} \setminus S[-p_L, p_U]} |d_r| / \sigma^2, \quad (\text{D.4})$$

with $d_0 = d_{0, \mathfrak{h}(p)} = 1$. Also,

$$\sigma_{\mathfrak{h}(p)}^2 - \sigma^2 = \gamma(\mathbf{0}) - d'_{\mathfrak{h}(p)} \psi_{\mathfrak{h}(p)} - \sigma^2 = \sum_{r \in S_{1+}^\infty \setminus S[-p_L, p_U]} d_r \gamma(r) \rightarrow 0, \quad (\text{D.5})$$

as $\mathfrak{h}(p) \rightarrow \infty$, by (2.4) and Lemma 2.1. Combining (D.4) and (D.5) yields the result. \square

Proof of Lemma 4.3. The proof is a straightforward extension of Theorem 3 of Berk (1974). Label the indices in the first row of $\Psi_{\mathfrak{h}(p)}$ (these are identical to those in the first row of $\hat{\Psi}_{\mathfrak{h}(p)}$) from, left to right, as $j_0, j_1, \dots, j_{\mathfrak{h}(p)-1}$, with $j_0 \equiv \mathbf{0}$. Take

$$\nu(p) = \left(1, e^{ij'_1 \lambda}, \dots, e^{ij'_{\mathfrak{h}(p)-1} \lambda}\right)', \quad \eta(p) = \left(1, e^{ij'_1 \mu}, \dots, e^{ij'_{\mathfrak{h}(p)-1} \mu}\right)'; \lambda, \mu \in \Pi.$$

In view of (B.2) it is sufficient to evaluate $\lim_{\mathfrak{h}(p) \rightarrow \infty} \mathfrak{h}(p)^{-1} \nu(p)' \Psi_{\mathfrak{h}(p)} \eta(p)$, which equals

$$\lim_{\mathfrak{h}(p) \rightarrow \infty} \mathfrak{h}(p)^{-1} \sum_{l=0}^{\mathfrak{h}(p)-1} D_l(e^{-i\lambda}) D_l(e^{-i\mu}) e^{ij'_l(\lambda+\mu)} / \sigma_l^2, \quad (\text{D.6})$$

where $e^z = (e^{z_1}, \dots, e^{z_d})'$ for any $s \in \mathbb{C}^d$. If $\lambda_i = -\mu_i$ or $\lambda_i = \mu_i = \pi$, $i = 1, \dots, d$, the RHS of (D.6) equals $\lim_{l \rightarrow \infty} |D_l(e^{i\lambda})|^2 / \sigma_l^2 = |D(e^{i\lambda})|^2 / \sigma^2 = 1 / (2\pi)^d f(\lambda)$, by Lemma 4.2.

If $e^{ij'_l(\lambda+\mu)} \neq 1$ for all j_l write $D_l(e^{-i\lambda}) D_l(e^{-i\mu}) / \sigma_l^2 = U_l$, $e^{ij'_l(\lambda+\mu)} = V_l$, where $V_l = \sum_{r=1}^l V_r$. Then the RHS of (D.6) equals

$$\begin{aligned} \lim_{\mathfrak{h}(p) \rightarrow \infty} \mathfrak{h}(p)^{-1} \left(\sum_{l=1}^{\mathfrak{h}(p)-2} (U_l - U_{l+1}) V_l + U_{\mathfrak{h}(p)-1} V_{\mathfrak{h}(p)-1} \right) \\ = \lim_{\mathfrak{h}(p) \rightarrow \infty} \mathfrak{h}(p)^{-1} \sum_{l=1}^{\mathfrak{h}(p)-2} (U_l - U_{l+1}) V_l, \end{aligned} \quad (\text{D.7})$$

because $\lim_{\mathfrak{h}(p) \rightarrow \infty} U_{\mathfrak{h}(p)-1} = D(e^{-i\lambda}) D(e^{-i\mu}) / \sigma^2 < C$, by Lemma 4.2, and $V_{\mathfrak{h}(p)-1} = (1 - e^{i\mathfrak{h}(p)(\lambda+\mu)}) / (1 - e^{i(\lambda+\mu)}) = \mathcal{O}(1)$. Then, by Lemma 4.2 it follows that the RHS of (D.7) equals 0. \square

Proof of Lemma 4.4. Since $(N/\mathfrak{h}(p))^{\frac{1}{2}} \sum_{s \in S_{1+}^\infty \setminus S[-p_L, p_U]} d_s e^{is'\lambda} \rightarrow 0$ as $N \rightarrow \infty$, any $\lambda \in \Pi^d$, we can replace $C(\lambda)$ and $S(\lambda)$ in (4.5) by $C_{\mathfrak{h}(p)}(\lambda) = 1 + \sum_{s \in S[-p_L, p_U]} d_s \cos(s'\lambda)$ and $S_{\mathfrak{h}(p)}(\lambda) = \sum_{s \in S[-p_L, p_U]} d_s \sin(s'\lambda)$ respectively. Lemma 4.1 and Theorem 4.1 immediately provide the joint asymptotic normality of (4.5), by the Cramér-Wold device. The asymptotic variance of $(N/\mathfrak{h}(p))^{\frac{1}{2}} (\hat{C}_{\mathfrak{h}(p)}(\mathbf{0}) - C(\mathbf{0}))$ is obtained by taking $w_0 = 1$ and others zero in Lemma 4.3, while for $(N/\mathfrak{h}(p))^{\frac{1}{2}} (\hat{C}_{\mathfrak{h}(p)}(\boldsymbol{\pi}) - C(\boldsymbol{\pi}))$ we take $u_0 = 1$ with others zero. For $j = 1, \dots, q$, take $w_j = u_j = 1/2$ and others zero for $(N/\mathfrak{h}(p))^{\frac{1}{2}} (\hat{C}_{\mathfrak{h}(p)}(\lambda_j) - C(\lambda_j))$, and $w_j = -i/2$, $u_j = i/2$ and others zero for

$(N/\mathfrak{h}(p))^{\frac{1}{2}} \left(\hat{S}_{\mathfrak{h}(p)}(\lambda_j) - S(\lambda_j) \right)$. It is easy to show using this method that the asymptotic variance of the sum of any pair of terms (4.5) is the sum of the asymptotic variances, implying that the asymptotic covariance matrix is diagonal. \square

Proof of Lemma A.1. $S_{MN} = \sum_{t(|M|,N)}'' \sum_{s^1 \in \mathbb{Z}^d} \cdots \sum_{s^q \in \mathbb{Z}^d} \xi_{st}$, which is rewritten as $S_{MN} = \sum_{s^1 \in \mathbb{Z}^d} \cdots \sum_{s^q \in \mathbb{Z}^d} \eta_{1s^1}^{1-1/w} \eta_{1s^1}^{1/w} \sum_{t(|M|,N)}'' (\xi_{st}/\eta_{1s^1})$ whence from Hölder's inequality

$$|S_{MN}|^w \leq \left(\sum_{s \in \mathbb{Z}^d} \eta_{1s} \right)^{w-1} \sum_{s^1 \in \mathbb{Z}^d} \eta_{1s^1}^{1-w} \left| \sum_{s^2 \in \mathbb{Z}^d} \cdots \sum_{s^q \in \mathbb{Z}^d} \sum_{t(|M|,N)}'' \xi_{st} \right|^w.$$

Similarly $\left| \sum_{s^2 \in \mathbb{Z}^d} \cdots \sum_{s^q \in \mathbb{Z}^d} \sum_{t(|M|,N)}'' \xi_{st} \right|^w$ is bounded by

$$\left(\sum_{s \in \mathbb{Z}^d} \eta_{2s} \right)^{w-1} \sum_{s^2 \in \mathbb{Z}^d} \eta_{2s^2}^{1-w} \left| \sum_{s^3 \in \mathbb{Z}^d} \cdots \sum_{s^q \in \mathbb{Z}^d} \sum_{t(|M|,N)}'' \xi_{st} \right|^w.$$

After q applications of Hölder's inequality and using (A.2) we obtain

$$|S_{MN}|^w \leq C \sum_{s^1 \in \mathbb{Z}^d} \cdots \sum_{s^q \in \mathbb{Z}^d} \eta_{\mathbf{s}}^{1-w} \left| \sum_{t(|M|,N)}'' \xi_{st} \right|^w. \quad (\text{D.8})$$

Also, from von Bahr and Esseen (1965) and (A.1)

$$E \left| \sum_{t(|M|,N)}'' \xi_{st} \right|^w \leq C \sum_{t(|M|,N)}'' |\xi_{st}|^w \leq C \eta_{\mathbf{s}}^w \sum_{t(|M|,N)}'' a_t^w.$$

Taking expectations of (D.8) and applying the above and (A.2) we conclude

$$E |S_{MN}|^w \leq C \sum_{s^1 \in \mathbb{Z}^d} \cdots \sum_{s^q \in \mathbb{Z}^d} \eta_{\mathbf{s}} \sum_{t(|M|,N)}'' a_t^w \leq C \sum_{t(|M|,N)}'' a_t^w = C b_{wMN},$$

establishing the lemma. \square

References

- Abramovich, F. and T. Lahav (2015). Sparse additive regression on a regular lattice. *Journal of the Royal Statistical Society Series B* 77, 443–459.
- Akaike, H. (1969). Power spectrum estimation through autoregressive model fitting. *Annals of the Institute of Statistical Mathematics* 21, 407–419.

- Akaike, H. (1970). Statistical predictor identification. *Annals of the Institute of Statistical Mathematics* 22, 203–217.
- Anderson, T. W. (1971). *The Statistical Analysis of Time Series*. John Wiley & Sons.
- Bandyopadhyay, S., S. N. Lahiri, and D. J. Nordman (2015). A frequency domain empirical likelihood method for irregularly spaced spatial data. *The Annals of Statistics* 43, 519–545.
- Baxter, G. (1962). An asymptotic result for the finite predictor. *Mathematica Scandinavica* 10, 137–144.
- Berk, K. N. (1974). Consistent autoregressive spectral estimates. *The Annals of Statistics* 2, 489–502.
- Bhansali, R. J. (1980). Autoregressive and window estimates of the inverse correlation function. *Biometrika* 67, 551–566.
- Bronars, G. and W. Jansen (1987). The geographic distribution of unemployment rates in the U.S. *Journal of Econometrics* 36, 251–279.
- Chen, X. and W. D. Nordhaus (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences of the United States of America* 108, 8589–8594.
- Dahlhaus, R. and H. Künsch (1987). Edge effects and efficient parameter estimation for stationary random fields. *Biometrika* 74, 877–882.
- Fuentes, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *Journal of the American Statistical Association* 102, 321–331.
- Grenander, U. and G. Szegö (1984). *Toeplitz Forms and their Applications*. Chelsea, New York.
- Guyon, X. (1982). Parameter estimation for a stationary process on a d -dimensional lattice. *Biometrika* 69, 95–105.
- Haining, R. P. (1978). The moving average model for spatial interaction. *Transactions of the Institute of British Geographers* 3, 202–225.
- Hannan, E. J. (1970). *Multiple Time Series*. John Wiley & Sons.
- Helson, H. and D. Lowdenslager (1958). Prediction theory and Fourier series in several variables. *Acta Mathematica* 99, 165–202.
- Helson, H. and D. Lowdenslager (1961). Prediction theory and Fourier series in several variables. II. *Acta Mathematica* 106, 175–213.

- Hidalgo, J. (2009). Goodness of fit for lattice processes. *Journal of Econometrics* 151, 113–128.
- Hidalgo, J. and M. H. Seo (2014). Specification tests for lattice processes. *Econometric Theory* 31, 294–336.
- Jenish, N. (2014). Spatial semiparametric model with endogenous regressors. *Econometric Theory FirstView*, 1–26.
- Korezlioglu, H. and P. Loubaton (1986). Spectral factorization of wide sense stationary processes on \mathbb{Z}^2 . *Journal of Multivariate Analysis* 19, 24–47.
- Kromer, R. E. (1970). *Asymptotic Properties of the Autoregressive Spectral Estimator*. Ph. D. thesis, Stanford University.
- Matsuda, Y. and Y. Yajima (2009). Fourier analysis of irregularly spaced data on R^d . *Journal of the Royal Statistical Society Series B* 71, 191–217.
- McClellan, J. H. (1982). Multidimensional spectral estimation. *Proceedings of the IEEE* 70, 1029–1039.
- McElroy, T. S. and S. H. Holan (2014). Asymptotic theory of cepstral random fields. *The Annals of Statistics* 42, 64–86.
- Nordhaus, W. D. (2008). New metrics for environmental economics: Gridded economic data. *The Integrated Assessment Journal* 8, 73–84.
- Pace, R. K. and R. Barry (1997). Sparse spatial autoregressions. *Statistics & Probability Letters* 33, 291 – 297.
- Pace, R. K. and J. P. LeSage (2003). Likelihood dominance spatial inference. *Geographical Analysis* 35, 133–147.
- Parzen, E. (1969). Multiple time series modeling. In P. R. Krishnaiah (Ed.), *Multivariate Analysis II*, pp. 389–410. Academic Press, New York.
- Politis, D. N. and J. P. Romano (1996). On flat-top kernel spectral density estimators for homogeneous random fields. *Journal of Statistical Planning and Inference* 51, 41 – 53.
- Robinson, P. M. (1978). On consistency in time series analysis. *The Annals of Statistics* 6, 215–223.
- Robinson, P. M. (1979). Distributed lag approximation to linear time-invariant systems. *The Annals of Statistics* 7, 507–515.
- Robinson, P. M. (2007). Nonparametric spectrum estimation for spatial data. *Journal of Statistical Planning and Inference* 137, 1024–1034.

- Robinson, P. M. (2008). Correlation testing in time series, spatial and cross-sectional data. *Journal of Econometrics* 147, 5–16.
- Robinson, P. M. (2012). Inference on power law spatial trends. *Bernoulli* 18, 644–677.
- Robinson, P. M. and J. Vidal Sanz (2006). Modified Whittle estimation of multilateral models on a lattice. *Journal of Multivariate Analysis* 97, 1090–1120.
- Roknossadati, S. M. and M. Zarepour (2010). M -estimation for a spatial unilateral autoregressive model with infinite variance innovations. *Econometric Theory* 26, 1663–1682.
- Rosenblatt, M. (1985). *Stationary Sequences and Random Fields*. Birkhäuser.
- Rudin, W. (1973). *Functional Analysis*. McGraw-Hill, New York.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* 6, 461–464.
- Scott, D. J. (1973). Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach. *Advances in Applied Probability* 5, 119–137.
- Tjøstheim, D. (1981). Autoregressive modeling and spectral analysis of array data in the plane. *IEEE Transactions on Geoscience and Remote Sensing* GE-19, 15–24.
- Tjøstheim, D. (1983). Statistical spatial series modelling II: Some further results on unilateral lattice processes. *Advances in Applied Probability* 15, 562–584.
- Vidal Sanz, J. (2009). Automatic spectral density estimation for random fields on a lattice via bootstrap. *Test* 18, 96–114.
- von Bahr, B. and C. G. Esseen (1965). Inequalities for the r th absolute moment of a sum of random variables, $1 \leq r \leq 2$. *The Annals of Mathematical Statistics* 36, 299–303.
- Wester, R., M. Tummala, and C. W. Therrien (1990). Multidimensional autoregressive spectral estimation using iterative methods. In *1990 Conference Record Twenty-Fourth Asilomar Conference on Signals, Systems and Computers*, Volume 1, pp. 442–446.
- Whittle, P. (1954). On stationary processes in the plane. *Biometrika* 41, 434–449.
- Yuan, J. and T. Subba Rao (1993). Higher order spectral estimation for random fields. *Multidimensional Systems and Signal Processing* 4, 7–22.