

Autoregressive spatial spectral estimates

Abhimanyu Gupta*[†]

Department of Economics
University of Essex

November 20, 2017

Abstract

Nonparametric spectral density estimates find many uses in econometrics. For stationary random fields on a regular spatial lattice, we propose an autoregressive nonparametric spectral density estimate that is guaranteed positive even when suitable edge-effect correction is employed and is simple to compute using least squares. Our estimate is based on truncating a true half-plane infinite autoregressive representation, while also allowing the truncation length to diverge in all dimensions to avoid the potential bias due to truncation at a fixed lag-length. Uniform consistency of the proposed estimate is established, and new criteria for order selection are also suggested and studied in practical settings. 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. Technically the key to the results is the covariance structure of stationary random fields defined on regularly spaced lattices. We 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, HAC estimation, random field, covariance matrix, spatial process, lattice data, central limit theorem.

*I thank co-editor Oliver Linton for comments that significantly improved the paper over a previous version as well as the associate editor and a number of anonymous referees. I am grateful to Peter Robinson for detailed comments that led to substantial improvements, and to Swati Chandna, William Dunsmuir, Graham Elliott, Javier Hidalgo, Gordon Kemp and participants at SETA 2015, Tokyo, for very helpful comments and discussions.

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

1 Introduction

The analysis of spatial data has seen a great deal of recent econometric work. In this paper we are interested in the nonparametric estimation of the spectral density of a spatial process, using an autoregressive technique that has several attractive features. These include estimation based on least squares model fitting, easier handling of the edge effect and the possibility of establishing a central limit theorem at distinct frequencies.

Several approaches to dealing with spatial data are available in the econometric literature. One strand builds on the spatial autoregressive (SAR) model of [Cliff and Ord \(1972\)](#), assuming the existence of a known economic (not necessarily geographic) distance between units, and crucial theoretical contributions here include [Kelejian and Prucha \(1999\)](#) and [Lee \(2004\)](#). There is a very active research programme on SAR models, which are particularly attractive as they entail no knowledge of locations. On the other hand many spatial data sets are observed on \mathbb{R}^d , $d > 1$, implying that the locations are known and the distances between units is geographic, e.g. [Sain and Cressie \(2007\)](#) develop a Markov random field model for multivariate spatial data and apply it to data in environmental economics. For such data notions of dependence such as mixing and near epoch dependence have been developed, see e.g. [Jenish and Prucha \(2009, 2012\)](#).

Assuming then that the locations of the data have a geographic interpretation, we follow a large econometric literature (see e.g. [Conley \(1999\)](#), [Conley and Molinari \(2007\)](#), [Bester et al. \(2011\)](#), [Wang et al. \(2013\)](#) and [Bester et al. \(2016\)](#)) in assuming that the distance between spatial observations is such that the locations can be mapped to the regular integer lattice \mathbb{Z}^d . Such data may also 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, see also [Section 1.2](#) for further discussion. The structural models of local social interactions studied by [Topa \(2001\)](#) and [Conley and Topa \(2007\)](#) are linked with certain processes defined on integer lattices in the interacting particle systems literature. Lattice data arise also in the spatial econometrics literature. [Robinson \(2008, 2011\)](#) considers tests of spatial correlation and asymptotic theory for nonparametric regression with lattice data amongst a host of other settings while [Roknossadati and Zarepour \(2010\)](#) provide theory for M -estimation in a class of unilateral models. [Jenish \(2016\)](#) also considers a nonlinear autoregressive model on a regular lattice as a motivating example in her analysis of a spatial semiparametric model. In their study of non-nested spatial correlation tests [Delgado and Robinson \(2015\)](#) use a lattice setting in their Monte Carlo simulations.

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 $t_j \in \mathbb{Z}$, $j = 1, \dots, d$. The spectral density of a spatial process can be useful in many ways in economic applications, some of which we describe below.

Studentization by nonparametric spectrum estimates in semiparametric estimation and HAC estimation: In semiparametric estimation econometricians are frequently faced with establishing asymptotic normality of a sequence of random variables of the form $N^{-\frac{1}{2}} \sum_{t \in \mathbb{N}} z_t$,

where z_t has spectral density matrix $F(\cdot)$ and $\mathbb{N} \subseteq \mathbb{Z}^d$. Under a number of weak dependence conditions it can be shown that

$$N^{-\frac{1}{2}} \sum_{t \in \mathbb{N}} z_t \xrightarrow{d} N\left(0, (2\pi)^d F(0)\right). \quad (1.1)$$

Thus construction of valid inference rules that make use of (1.1) require a consistent estimate of $F(0)$. The studentization of statistics based on (1.1) by a nonparametric spectrum estimate is common practice in econometrics (its use in statistics dates back to [Jowett \(1955\)](#)) and underlies the particularly important computation of heteroskedasticity and autocorrelation robust (HAC) estimates of the covariance matrix of a spatial process. Consider the spatial regression model

$$y_t = x_t' \beta + \eta_t, \quad (1.2)$$

where η_t is a stationary spatial process, $t \in \mathbb{N}$ and the cardinality of \mathbb{N} is N . Suppose that we have obtained $N^{\frac{1}{2}}$ -consistent estimates of the finite-dimensional parameter β and wish to obtain HAC standard errors. As in the time series case (cf. [Newey and West \(1987\)](#), [Andrews \(1991\)](#)), a HAC estimate is essentially a smoothed estimate of the spectral density matrix of a stationary process at zero frequency.

Various types of spatial HAC estimates have been proposed in the literature. [Kelejian and Prucha \(2007\)](#) propose an estimate based on weights derived from economic distances between units, implying that knowledge of locations is not needed. Their estimator is generalized by [Kim and Sun \(2011\)](#), while [Vogelsang \(2012\)](#) builds a fixed- b asymptotic theory for spatial-dependence robust HAC estimates proposed by [Driscoll and Kraay \(1998\)](#). More general theory that also applies to panel data models with fixed effects is available in [Kim and Sun \(2013\)](#). On the other hand, [Robinson and Thawornkaiwong \(2012\)](#) study a HAC formula based on geographical data, treating the points of observation as discretized locations on a rectangular grid, while [Conley \(1999\)](#) (possibly the first to suggest a spatial HAC estimate) and [Bester et al. \(2016\)](#) assume, like this paper, that locations may be mapped to \mathbb{Z}^d , a setting also employed by [Kim and Sun \(2011, 2013\)](#) to discuss some of their conditions and conduct Monte Carlo simulations. An interesting aspect of spectrum estimation in the context of HAC estimates is that the acronym emphasizes ‘heteroskedasticity’ as much as ‘autocorrelation’, but because the estimate is of the spectrum of a stationary process at zero frequency only the autocorrelation is explicitly allowed for. To justify heteroskedasticity-robustness, familiar limit theorems for non-identically distributed variates are invoked, see e.g. [Robinson \(2005\)](#) for a discussion.

Spectral estimation for errors in efficient frequency domain weighted regression: In some cases what is of interest is not the spectrum of some observed process. For instance, in the context of efficient semiparametric estimation of time series regression models via frequency domain weighted regression, interest centres on spectral estimates of the errors of the model, see e.g. [Hannan \(1970\)](#), chapter 7. It is natural then that for efficient estimation of β in (1.2), when the distribution of η_t is nonparametric, frequency domain weighted regression

be employed, and spectral estimates be based on the residuals.

Specification and goodness of fit testing: Nonparametric spectrum estimates are required by [Hidalgo \(2009\)](#), who considers testing for correct parametric covariogram specification for lattice processes. A similar need arises in [Hidalgo and Seo \(2014\)](#), who propose omnibus-type specification tests. [Hidalgo \(2009\)](#) is equivalent to a test for whether the spectral density function of a spatial process is of prescribed parametric form. The nonparametric estimate of the spectrum via the technique given in this paper can be employed for this as well.

Frequency domain analysis of spatial processes: The natural analogy between lattice and time series data suggests a more central role for frequency domain analysis. 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.

1.1 Kernel versus autoregressive nonparametric spectral estimation

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 \(2009\)](#). The difference between this approach and an autoregressive one is analogous to estimating a nonparametric regression function by kernels or series, with the former providing an estimate that is local in nature and the latter approximating the function globally.

For lattice processes autoregressive estimation can be helpful when handling the edge-effect. When $d = 1$ the loss of data at the end of the series while estimating autocovariances has a negligible asymptotic effect, but this edge-effect, or end-effect, matters when $d = 2$ and worsens with increasing d ; see [Section 2](#) for a more detailed discussion. [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 and allows us to establish what is, to the best of our knowledge, the first central limit theorem for a nonparametric spatial spectral estimate. [Bester et al. \(2016\)](#)'s fixed- b asymptotic theory captures the shape of the sampling region, implying edge-effects are reflected in the reference distribution that generates critical values. Their results may be used to obtain a nonstandard limit for the spectral estimator that depends on edge-effects, in contrast to our

results where the limit is normal.

An advantage of the kernel based estimator over our autoregressive approach is that it can be applied directly to the data given the availability of a distance measure between observations. This avoids the potentially cumbersome, albeit feasible, mapping to a lattice that we assume.

1.2 Regular and irregular lattices

Frequency domain techniques are also employed with data on irregular spatial lattices (see [Matsuda and Yajima \(2009\)](#), [Bandyopadhyay et al. \(2015\)](#)), however irregular spacing will disturb the Toeplitz property of the covariance matrix that we exploit for our results. But another, more practical, reason suggests itself for focussing on regular lattices. 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. \(2015\)](#). Our empirical example of presidential election voter turnout data across US counties, presented in the supplementary material, is in this spirit. [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, see http://www.stat.fi/tup/ruututietokanta/index_en.html, <http://gecon.yale.edu/> and [Nordhaus \(2008\)](#). Gridding of irregularly spaced data is commonplace in statistics, see e.g. [Fuentes \(2007\)](#) and references therein.

1.3 Scope and structure of this paper

For the case of regularly-spaced time series ($d = 1$), [Berk \(1974\)](#) assumes an infinite, one-sided autoregressive representation for x_t , driven by independent innovations, 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 a 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 an online appendix. 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 AR coefficients 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 introduces the setting employed throughout the paper and a theorem on consistency of the truncated AR predictors. Section 3 introduces the spectral density estimate and establishes its uniform 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 as well as applying our method to data generated by a Cliff and Ord (1973)-type spatial autoregressive (SAR) model, which entails nonstationarity. We also suggest some model selection criteria to choose the truncation point, without rigorous proof of their asymptotic properties, and study their finite-sample performance. Proofs of all results presented in the paper are contained in Appendix A while Appendix B presents a set of lemmas with proofs mostly included in the online appendix. The online appendix also contains an empirical example, bounds for absolute moments of partial sums of rather general lattice processes, a generalisation of the Toeplitz property familiar from the theory of stationary time series and an upper bound on the number of unique autocovariances that occur in the covariance matrix of finite, stationary and unilateral processes.

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 Jacobian 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, because one need not attempt to recover any underlying parameters of a multilateral model to study the properties of interest. Instead a straightforward least squares estimate can deliver a spectrum estimate for the process, regardless of its possibly multilateral underlying structure. Thus the half-plane representation places no serious limitation on the dependence structure of the process.

As in Tjøstheim (1983) we define the half-plane as all t in the set

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

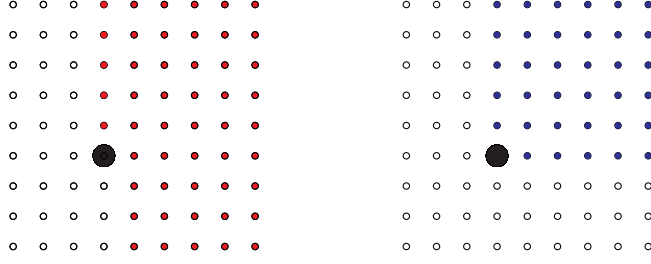


Figure 2.1: Half-plane illustration for $d = 2$. In the left panel solid red dots form S_{1+}^{∞} while in the right panel solid blue dots illustrate the quarter-plane. The large black solid dot marks the origin.

The special case with $t_i \geq 0$, $i = 1 \dots, d$, is referred to as a quarter-plane. When $d = 2$, Figure 2.1 graphically illustrates the half-plane and 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 $z^s = \prod_{j=1}^d z_j^{s_j}$. For real numbers b_s , $s \in S_{1+}^{\infty} \cup \mathbf{0}$, 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 independent random variables ϵ_t , $t \in \mathbb{Z}^d$, with $E\epsilon_t = 0$ and $E\epsilon_t^2 = \sigma^2$ such that

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

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

Martingale assumptions can replace the independence imposition on ϵ_t , but we choose to avoid these as they rely on notions of ordering that can be arbitrary. Identity of distribution is not assumed at this stage. 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 with white noise ϵ_t and $\sum_{s \in S_{1+}^{\infty} \cup \mathbf{0}} |b_s|^2 < \infty$, extending the Wold decomposition of time series analysis. We require $\sum_{s \in S_{1+}^{\infty} \cup \mathbf{0}} |b_s| < \infty$ to rule out potential long-memory, which entails unbounded $f(\lambda)$. Thus Assumption A does rule out stationary spatial processes with long memory and certainly doesn't cover the entire class of stationary processes, while nonstationary processes are not theoretically considered in the paper.

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

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

The above is the extension of a time series $\text{AR}(\infty)$ specification to the spatial case. For $d = 2$, (2.4) becomes

$$x_{t_1, t_2} = \sum_{s_1 > 0, s_2 \in \mathbb{Z}} d_{s_1, s_2} x_{t_1 - s_1, t_2 - s_2} + \sum_{s_2 > 0} d_{0, s_2} x_{t_1, t_2 - s_2} + \epsilon_{t_1, t_2}, \quad (t_1, t_2)' \in \mathbb{Z}^2,$$

and, for example, a parametric analogue of a time series $\text{AR}(1)$ model would be

$$x_{t_1, t_2} = d_{1, -1} x_{t_1 - 1, t_2 + 1} + d_{1, 0} x_{t_1 - 1, t_2} + d_{1, 1} x_{t_1 - 1, t_2 - 1} + d_{0, 1} x_{t_1, t_2 - 1} + \epsilon_{t_1, t_2}, \quad (t_1, t_2)' \in \mathbb{Z}^2.$$

By Assumption A and (2.3) there exist real numbers m, M satisfying $0 < m \leq M < \infty$, such that

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

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 truncation point relative to sample size that become more stringent as $v \rightarrow 1$. We observe x_t on the rectangular lattice $\mathbb{L} = \{t \in \mathbb{Z}^d : -n_{L_i} \leq t_i \leq n_{U_i}, i = 1, \dots, d\}$, $n_{U_i}, n_{L_i} \geq 0$, $i = 1, \dots, d$, with $n_{L_1} = 0$ without loss of generality in view of the half plane representation. Define $n_i = n_{L_i} + n_{U_i} + 1$, $i = 1, \dots, d$, and $N = \prod_{i=1}^d n_i$. Note that we observe the data at locations on a regular grid in d dimensions, and consistency in our setting is only possible if sample size increases in all directions. We require a mild degree of regularity in this increase across dimensions, for which we introduce

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} , $i = 1, \dots, d$, where 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’. For $d = 2$, (2.7) becomes the set of all $t = (t_1, t_2)' \in \mathbb{L}$ such that $0 \leq t_1 \leq p_{U_1}$ and $-p_{L_2} \leq t_2 \leq p_{U_2}$. Denote $p_i = p_{L_i} + p_{U_i}$, $i = 1, \dots, d$ and let $\mathfrak{h}(p)$ denote the total number of autoregressive coefficients 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 now 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, possibly in the interests of simplicity but more generally if similar regularity in $f(\lambda)$ is anticipated across dimensions. In this case $p_{L_i} = p_{U_i} = p_{U_1} = p^\dagger$, say, $i = 2, \dots, d$, and (2.8) indicates that $\mathfrak{h}(p^\dagger) = \left((2p^\dagger + 1)^d - 1 \right) / 2$. A more flexible and data-driven 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 .

As x_t is stationary, we define the autocovariances $\gamma(k) = Ex_t x_{t+k}$ with $t, k \in \mathbb{Z}^d$. 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 $\hat{\gamma}(k) = n_p^{-1} \sum''_{t(p,n)} x_t x_{t+k}$, $k \in S[-p_L, p_U]$, where $n_i > p_i \geq 0$ for $i = 1, \dots, d$ and the sum $\sum''_{t(p,n)}$ runs over t satisfying $p_i - n_{L_i} < t_i \leq n_i - n_{L_i}$, indicating that there are n_p summands. 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. Using this correction also removes the edge effect when using a kernel based spectral estimate, however this may lead to negative estimates (see Dahlhaus and Künsch (1987)). A referee has pointed out that such negative values may be replaced by a small positive number and thus this issue may not be very serious in practice.

We assume 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 $\gamma^*(k) = n_p^{-1} \sum''_{t(p,n)} (x_t - \bar{x})(x_{t+k} - \bar{x})$, where $\bar{x} = N^{-1} \sum_{t \in \mathbb{L}} x_t$, and the latter is readily shown to be $N^{\frac{1}{2}}$ -consistent for α . This is equivalent to replacing x_t and x_{t+k} by the residuals from a regression on a constant, and one may instead wish to use residuals from a regression that includes explanatory variables z_t , implying that $Ex_t = z_t' \beta$ for some parameter vector β .

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

$$\hat{d}_{\mathfrak{h}(p)} = \hat{\Psi}_{\mathfrak{h}(p)}^{-1} \hat{\psi}_{\mathfrak{h}(p)}, \quad (2.9)$$

and we denote the elements of $\hat{d}_{\mathfrak{h}(p)}$ by $\hat{d}_{s, \mathfrak{h}(p)}, s \in S[-p_L, p_U]$. Here, we 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 $\gamma(k)$. Note that our estimate (2.9) is the least squares estimate and not the Yule-Walker estimate. For spectral estimation, using the former over the latter has support from the results of Lysne and Tjøstheim (1987), for example, who show that the use of Yule-Walker estimates can lead to loss of peaks and strong bias.

A notable difference from the time series case ($d = 1$) is in the dimension of the AR coefficient space. When $d = 1$, this equals the number of unique covariances in $\Psi_{\mathfrak{h}(p)}$, of which there are $\mathfrak{h}(p)$. On the other hand, in the spatial case when $d \geq 2$, these number at most $\mathfrak{C}(p) \geq \mathfrak{h}(p)$. We show this in supplementary appendix S.3.2, which indicates that

$$\mathfrak{C}(p) = 1 + \sum_{l=1}^{d-1} 2^{d-l-1} \sum_{\substack{\#(l=0) \\ \approx 0_d^l}} \prod_{k=1}^d p_k + 2^{d-1} \prod_{k=1}^d p_k, \quad (2.10)$$

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.

Throughout the sequel we assume that $\mathfrak{h}(p)^{-1} + \mathfrak{C}(p)^{-1} \rightarrow 0$, as $N \rightarrow \infty$ and write $d_{\mathfrak{h}(p)}$ for the $\mathfrak{h}(p) \times 1$ vector with elements d_s , $s \in S[-p_L, p_U]$. 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 supplementary appendix S.3. 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. These are the maximum absolute row-sum and spectral norms, respectively. Our first theorem is a consistency result for the estimated predictor coefficients.

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

$$\frac{\mathfrak{C}(p)}{N^{\frac{v-1}{v}}} \rightarrow 0, \text{ as } N \rightarrow \infty, \quad (2.11)$$

and

$$\sum_{t \in S_{1+}^{\infty} \setminus S[-p_L, p_U]} |d_t| \rightarrow 0 \text{ as } N \rightarrow \infty. \quad (2.12)$$

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

Condition (2.12) 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 case, condition (2.11) applies to the dimension of the AR coefficient space because this dimension is equal to the number of unique covariances in $\Psi_{\mathfrak{h}(p)}$. These unique covariances number $\mathfrak{h}(p)$ when $d = 1$, but at most $\mathfrak{C}(p) \geq \mathfrak{h}(p)$ when $d > 1$. Define the error variance estimate based on least squares residuals 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 2.2. *Under the conditions of Theorem 2.1, $\hat{\sigma}_{\mathfrak{h}(p)}^2 \xrightarrow{p} \sigma^2$, as $N \rightarrow \infty$.*

3 Uniform consistency of AR spectral density estimates

We now introduce spectral density estimates. 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

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

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. We present a theorem for uniform consistency below.

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 $\sup_{\lambda \in \Pi} \left| \hat{f}_{\mathfrak{h}(p)}(\lambda) - f(\lambda) \right| \xrightarrow{p} 0$, 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 AR coefficient space relative to sample size is stronger than the one imposed for Theorems 2.1 and 2.2. For example, if $v = 2$ then (2.11) 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 $\mathfrak{h}(p)^{\frac{1}{2}}$.

4 Asymptotic normality

In this section we prove asymptotic normality of the AR spectral estimate $\hat{f}_{\mathfrak{h}(p)}(\lambda)$. We start by establishing the asymptotic distribution of a linear combination of the autoregression coefficient estimates \hat{d}_s .

Theorem 4.1. *Let $d > 1$, $\alpha(p)$ be as in Lemma B.10, Assumptions A, B and C hold, but strengthened to*

$$v = 2 \text{ and } \epsilon_t \text{ identically distributed for all } t \in \mathbb{L}. \quad (4.1)$$

Choose the sequence p as a function of N such that (2.11) holds and

$$\frac{\mathfrak{h}(p)}{N^{\frac{\chi}{4}}} + N^{\frac{1}{2}} \sum_{s \in S_{1+}^{\infty} \setminus S_{[-p_L, p_U]}} |d_s| \rightarrow 0, \text{ as } N \rightarrow \infty. \quad (4.2)$$

Then $\left(N^{\frac{1}{2}}/\mathfrak{h}(p)^{\frac{1}{2}}\right) \alpha(p)' \left(\hat{d}_{\mathfrak{h}(p)} - d_{\mathfrak{h}(p)}\right) \xrightarrow{d} N(0, \sigma^2 \mu)$, as $N \rightarrow \infty$, where μ is defined in (B.11).

Condition (4.2) 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, (4.2) can never be this sharp as $\chi = 1$ at most when $d = 1$, thus reflecting the fundamental difference between 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 . Thus, the explicit imposition of $d > 1$ in the statement of Theorem 4.1 is an important feature as we cannot simply regard the case $d = 1$ as a special case. It is straightforward to extend the argument to allow for the asymptotic distribution of finitely many linear combinations by replacing $\alpha(p)$ with an $\ell \times \mathfrak{h}(p)$ matrix with full row rank, ℓ fixed, but we consider $\ell = 1$ for simplicity.

We now proceed to the asymptotic normality result for $\hat{f}_{\mathfrak{h}(p)}(\lambda)$, which relies on a lemma establishing asymptotic normality of certain linear combinations of the $\hat{d}_{s, \mathfrak{h}(p)}$. We will write $\boldsymbol{\pi} = (\pi, \dots, \pi)'$. Define $\hat{C}_{\mathfrak{h}(p)}(\lambda) = 1 + \sum_{s \in S[-pL, pU]} \hat{d}_{s, \mathfrak{h}(p)} \cos(s'\lambda)$, $C(\lambda) = 1 + \sum_{s \in S_{1+}^{\infty}} d_s \cos(s'\lambda)$, $\hat{S}_{\mathfrak{h}(p)}(\lambda) = \sum_{s \in S[-pL, pU]} \hat{d}_{s, \mathfrak{h}(p)} \sin(s'\lambda)$, $S(\lambda) = \sum_{s \in S_{1+}^{\infty}} d_s \sin(s'\lambda)$, the $2(q+1) \times 1$ vector $t_{\mathfrak{h}(p)}$ to have elements

$$\begin{aligned} & \left(\hat{C}_{\mathfrak{h}(p)}(\mathbf{0}) - C(\mathbf{0})\right), \left(\hat{C}_{\mathfrak{h}(p)}(\lambda_1) - C(\lambda_1)\right), \dots, \left(\hat{C}_{\mathfrak{h}(p)}(\lambda_q) - C(\lambda_q)\right), \\ & \left(\hat{C}_{\mathfrak{h}(p)}(\boldsymbol{\pi}) - C(\boldsymbol{\pi})\right), \left(\hat{S}_{\mathfrak{h}(p)}(\lambda_1) - S(\lambda_1)\right), \dots, \left(\hat{S}_{\mathfrak{h}(p)}(\lambda_q) - S(\lambda_q)\right), \end{aligned} \quad (4.3)$$

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

$$\begin{aligned} \Gamma &= \left(\sigma^2 / (2\pi)^d\right) \text{diag} [1/f(\mathbf{0}), 1/2f(\lambda_1), \dots, 1/2f(\lambda_q), 1/f(\boldsymbol{\pi}), \\ & 1/2f(\lambda_1), \dots, 1/2f(\lambda_q)]. \end{aligned} \quad (4.4)$$

Lemma 4.1. *Under the conditions of Theorem 4.1, $(N/\mathfrak{h}(p))^{\frac{1}{2}} t_{\mathfrak{h}(p)} \xrightarrow{d} N(0, \Gamma)$, as $N \rightarrow \infty$.*

Lemma 4.1 is analogous to results in the time series literature, cf. Parzen (1969), Berk (1974). Now define the $(q+2) \times 1$ vector $s_{\mathfrak{h}(p)}$ to have elements

$$\hat{f}_{\mathfrak{h}(p)}(\mathbf{0}) - f(\mathbf{0}), \hat{f}_{\mathfrak{h}(p)}(\lambda_1) - f(\lambda_1), \dots, \hat{f}_{\mathfrak{h}(p)}(\lambda_q) - f(\lambda_q), \hat{f}_{\mathfrak{h}(p)}(\boldsymbol{\pi}) - f(\boldsymbol{\pi}), \quad (4.5)$$

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

$$\Omega = 2 \text{diag} (2f^2(\mathbf{0}), f^2(\lambda_1), \dots, f^2(\lambda_q), 2f^2(\boldsymbol{\pi})). \quad (4.6)$$

Theorem 4.2. *Let the conditions of Theorem 4.1 hold with (2.11) replaced by (3.1). Then $(N/\mathfrak{h}(p))^{\frac{1}{2}} s_{\mathfrak{h}(p)} \xrightarrow{d} N(0, \Omega)$, as $N \rightarrow \infty$.*

The asymptotic distribution of the spectral density estimates at distinct frequencies mirrors

$d = 2$					$d = 3$				
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 $\hat{f}_{\hat{\eta}(p)}(\cdot)$.

τ		0.05	0.075	0.10
n^*	m^*	MISE	MISE	MISE
5	1	0.2610	0.3896	0.4956
7	1	0.2323	0.3528	0.4682
7	2	0.2464	0.3750	0.5262
9	1	0.2121	0.3405	0.5205
9	2	0.2305	0.3757	0.5110
9	3	0.2296	0.3716	0.4955
11	1	0.2257	0.3495	0.4734
11	2	0.2221	0.3719	0.4861
11	3	0.2288	0.3712	0.4957
11	4	0.2373	0.3788	0.5271

Table 5.2: Monte Carlo MISE of $\hat{f}^T(\cdot)$.

that in the time series case (cf. [Anderson \(1971\)](#), ch. 9, [Berk \(1974\)](#)), albeit under the stronger condition (4.2) and different condition (3.1).

5 Monte Carlo simulations

We examined finite-sample behaviour in two sets of Monte Carlo simulations, one with stationary and regularly-spaced data and the second with nonstationary and possibly irregularly-spaced data.

5.1 Stationary and regularly-spaced data

As in [Robinson and Vidal Sanz \(2006\)](#) and [Robinson \(2007\)](#) we generated x_t using

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

$d = 2$		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

$d = 3$		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 5.3: Monte Carlo bias and standard deviation of $\hat{f}_{b(p)}(\mathbf{0})$.

for $d = 2, 3$, similar to a model considered in [Haining \(1978\)](#). Then

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

with $\nu_d(\lambda) = \prod_{j=1}^d (1 + 2 \cos \lambda_j) - 1$. [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 $\mathbb{L} = \{t : -n^* \leq t_i \leq n^*, i = 1, \dots, d\}$, 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:

$$d = 2 : \tau = 0.05, 0.075, 0.10; \quad n^* = 5, 7, 9, 11;$$

$$d = 3 : \tau = 0.0075, 0.015, 0.03; \quad n^* = 3, 4, 5, 6, 7, 8.$$

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

τ		0.05		0.075		0.10	
n^*	m^*	Bias	Std. Dev.	Bias	Std. Dev.	Bias	Std. Dev.
5	1	-0.0141	0.0019	-0.0226	0.0029	-0.0237	0.0064
7	1	-0.0095	0.0023	-0.0138	0.0037	-0.0236	0.0057
7	2	-0.0125	0.0013	-0.0214	0.0022	-0.0307	0.0031
9	1	-0.0090	0.0028	-0.0205	0.0029	-0.0248	0.0052
9	2	-0.0103	0.0016	-0.0186	0.0019	-0.0316	0.0029
9	3	-0.0121	0.0010	-0.0199	0.0018	-0.0278	0.0025
11	1	-0.0102	0.0024	-0.0159	0.0031	-0.0258	0.0053
11	2	-0.0112	0.0014	-0.0187	0.0021	-0.0271	0.0030
11	3	-0.0110	0.0010	-0.0201	0.0016	-0.0283	0.0029
11	4	-0.0124	0.0008	-0.0206	0.0012	-0.0307	0.0021

Table 5.4: Monte Carlo bias and standard deviation of $\hat{f}^T(\mathbf{0})$.

τ	0.05			0.075			0.10		
p	$\widehat{\text{FPE}}$	$\widetilde{\text{FPE}}$	$\overline{\text{FPE}}$	$\widehat{\text{FPE}}$	$\widetilde{\text{FPE}}$	$\overline{\text{FPE}}$	$\widehat{\text{FPE}}$	$\widetilde{\text{FPE}}$	$\overline{\text{FPE}}$
1	0.9751	0.9900	0.9825	0.9660	0.9808	0.9734	0.9386	0.9529	0.9458
2	0.9702	1.0115	0.9906	0.9561	0.9967	0.9762	0.9220	0.9612	0.9414
3	0.9618	1.0456	1.0028	0.9488	1.0315	0.9892	0.9115	0.9909	0.9503
4	0.9489	1.0933	1.0183	0.9343	1.0765	1.0026	0.8989	1.0357	0.9646

Table 5.5: Order selection using FPE, $d = 2$ and $n^* = 11$.

$n^* = 3, 4, 5, 6, 7$; $p = 1, 2$, for $n^* = 8$. The choices of τ satisfy (5.3).

We also compare our estimates with those obtained using a tapered average periodogram estimate, the aim of tapering being to mitigate the edge effect, cf. Dahlhaus and Künsch (1987). In particular, let $h_N(t) = 2^{-d} \prod_{i=1}^d h(t_i/n_i)$, where $h(\cdot)$ is a function taking values in $[0, 1]$ and define the tapered periodogram of x_t by

$$I_x^T(\lambda) = \left(\sum_{t \in \mathbb{L}} h_N^2(t) \right)^{-1} \left| \sum_{t \in \mathbb{L}} h_N(t) x_t e^{-it'\lambda} \right|^2, \quad \lambda \in \Pi. \quad (5.4)$$

Like Hidalgo and Seo (2014), we opt for the cosine bell taper, viz. $h(z) = 1 - \cos(2\pi z)$. The estimator we use is one commonly used in time series analysis, see e.g. Brillinger (1975), p.132, subsequently also studied in the spatial context by e.g. Robinson (2007), and (with some abuse of notation in the sum index) is given by

$$\hat{f}^T(\lambda) = (2\pi)^{-d} \left[\left(2 \prod_{i=1}^d m_i \right) + 1 \right]^{-1} \sum_{j=-m}^m I_x^T(\lambda + \lambda_j^{\mathcal{F}}), \quad (5.5)$$

where $m = (m_1, \dots, m_d)'$, the m_i are non-negative integer sequences satisfying $m_i + m_i/n_i \rightarrow$

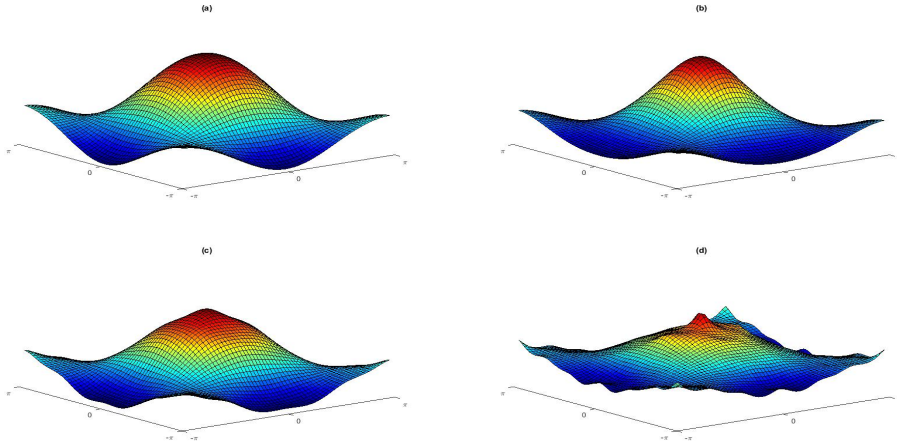


Figure 5.1: Spectral estimates for $d = 2$, $n^* = 11$, $\tau = 0.05$. (a) True spectrum (b) AR estimate with $p = 1$ (c) AR estimate with $p = 2$ (d) AR estimate with $p = 3$

0, $i = 1, \dots, d$, and $\lambda_k^{\mathcal{F}} = (\lambda_{k_1}^{\mathcal{F}}, \dots, \lambda_{k_d}^{\mathcal{F}})'$ is the k -th Fourier frequency, i.e. $\lambda_{k_i}^{\mathcal{F}} = 2\pi k_i/n_i$, with $k_1 = 0, \dots, n_1$, and $k_i = 0, \pm 1, \dots, \pm n_i$, $i = 2, \dots, d$. We will report results for $d = 2$ and take $m_1 = m_2 = m^* = 1, 2, 3, 4$, with higher values of m^* for bigger n^* .

Π is discretized with gaps of 0.10 in each dimension and we call this grid \mathbb{G} . In Tables 5.1 and 5.2 we report Monte Carlo mean integrated squared error (MISE) for $\hat{f}_{\mathfrak{h}(p)}(\lambda)$ and $\hat{f}^T(\lambda)$ respectively. This is defined as the Euclidean norm of $\hat{f}_{\mathfrak{h}(p)}(\bar{\lambda}) - f(\bar{\lambda})$ evaluated at frequencies in \mathbb{G} , i.e. $\text{MISE} = \left\{ \sum_{\bar{\lambda} \in \mathbb{G}} \left(\hat{f}_{\mathfrak{h}(p)}(\bar{\lambda}) - f(\bar{\lambda}) \right)^2 \right\}^{\frac{1}{2}}$, with an analogous definition for $\hat{f}^T(\lambda)$.

We first analyze Table 5.1. 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 . Moving to Table 5.2, we observe that MISE of $\hat{f}^T(\cdot)$ is generally much larger than for $\hat{f}_{\mathfrak{h}(p)}(\cdot)$. In fact, the MISE of the latter is smaller for any values of n^* , p and m^* for both $\tau = 0.05, 0.075$. Only when $\tau = 0.10$ does it become greater when $n^* \leq 7$, but here too with larger sample size, i.e. $n^* \geq 9$, it becomes much smaller than that of the periodogram based estimate as p increases.

Tables 5.3 and 5.4 report Monte Carlo bias and standard deviation (SD) for $\hat{f}_{\mathfrak{h}(p)}(\mathbf{0})$ and $\hat{f}^T(\mathbf{0})$, respectively. 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 . Like

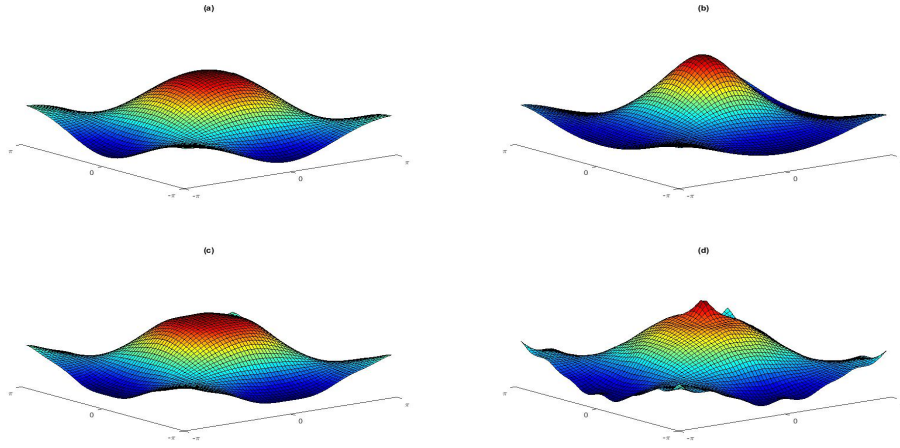


Figure 5.2: Spectral estimates for $d = 2$, $n^* = 11$, $\tau = 0.075$. (a) True spectrum (b) AR estimate with $p = 1$ (c) AR estimate with $p = 2$ (d) AR estimate with $p = 3$

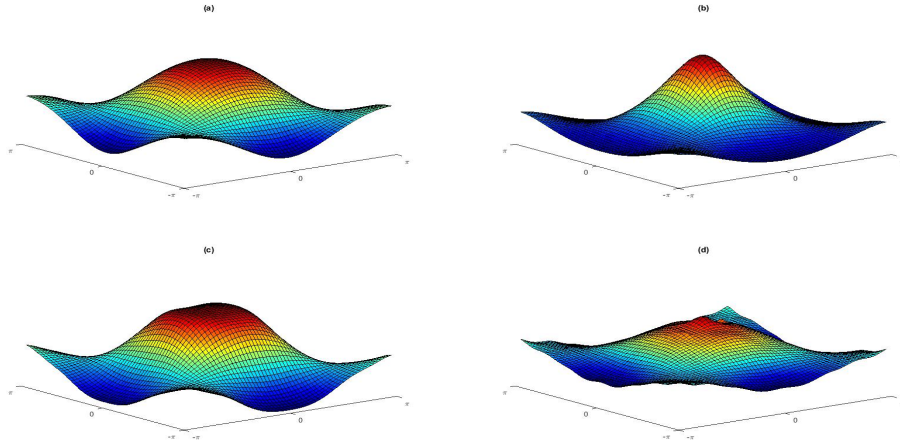


Figure 5.3: Spectral estimates for $d = 2$, $n^* = 11$, $\tau = 0.10$. (a) True spectrum (b) AR estimate with $p = 1$ (c) AR estimate with $p = 2$ (d) AR estimate with $p = 3$

Robinson (2007), we find that all biases for $\hat{f}^T(\mathbf{0})$ are negative, unlike the mostly positive values observed for $\hat{f}_{\hat{h}(p)}(\mathbf{0})$. For $d = 2$ the biases of $\hat{f}_{\hat{h}(p)}(\mathbf{0})$ sometimes dominate (in absolute value) those of $\hat{f}^T(\mathbf{0})$, but can become better e.g. for $n^* = 11$. For both $\hat{f}_{\hat{h}(p)}(\mathbf{0})$ and $\hat{f}^T(\mathbf{0})$ we find that the smallest values of SD also reduces monotonically with n^* . For $d = 3$, SD of $\hat{f}_{\hat{h}(p)}(\mathbf{0})$ 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 $\hat{f}_{\hat{h}(p)}(\cdot)$ 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

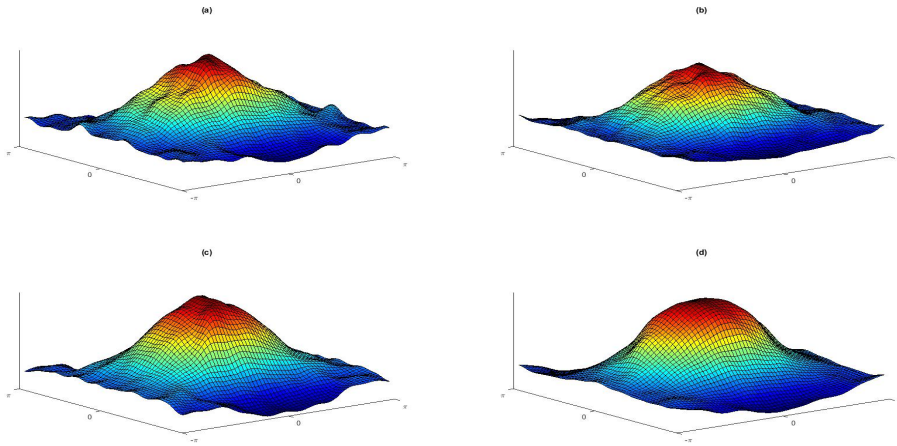


Figure 5.4: Average periodogram spectral estimates for $d = 2$, $n^* = 11$, $\tau = 0.075$. (a) $m^* = 1$ (b) $m^* = 2$ (c) $m^* = 3$ (d) $m^* = 4$.

the true peak for $p = 2$. As seen in Tables 5.1 and 5.3, 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 (visually) 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 5.3. 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). We illustrate using a similar figure for $\hat{f}^T(\cdot)$ in Figure 5.4, but to conserve space only report results for $\tau = 0.075$, so that the true spectrum is displayed in Figure 5.2(a). We see from panels (a) and (b) that when $m^* \leq 2$ a number of ridges appear in the plot and it doesn't quite resemble the true spectrum. However the resemblance becomes much stronger in panels (c) and (d), which display the estimated spectrum for $m^* = 3$ and $m^* = 4$ respectively, and looks the closest in panel (d). An interesting point to note is the greater levels of local detail (e.g. dimples and ridges) in the plots of Figure 5.4 as opposed to Figures 5.1, 5.2 and 5.3. This is in keeping with the 'local' character of the kernel based estimate $\hat{f}^T(\cdot)$ in comparison with the 'global' approximation that a sieve estimate like $\hat{f}_{\mathfrak{h}(p)}(\cdot)$ entails.

5.2 Order selection

We now discuss order selection for the AR spectral estimate. We begin by briefly discussing the time series case, where we write $\mathfrak{h}(p) = pU_1$ as p for simplicity. Shibata (1981) deduced that for a Gaussian, stationary linear process x_t , $t \in \mathbb{Z}$, the asymptotic lower bound for the integrated relative squared error (IRSE) of the AR spectral estimate, defined as $\int_{\Pi} \left\{ \left(\hat{f}_p(\lambda) - f(\lambda) \right) / f(\lambda) \right\}^2 d\lambda$, is achieved by the p^* minimizing $S_N(p) = (N + 2p)\hat{\sigma}_p^2$. $\hat{f}_{p^*}(\cdot)$ is then termed an optimal or asymptotically efficient AR spectral estimate. The result requires $\sum_{s=1}^{\infty} |d_s| < \infty$ and also that x_t does not degenerate to a finite order AR process.

If the Gaussianity assumption is dropped, [Karagrigoriou \(1997\)](#) establishes similar results under the conditions $\sup_t E\epsilon_t^{16} < \infty$ and

$$\sum_{s=1}^{\infty} s |d_s| < \infty. \quad (5.6)$$

The condition (5.6) is required to satisfy a form of mixing condition on the cumulants introduced by [Brillinger \(1969\)](#). For $d > 1$ and some generic vector norm $\|\cdot\|_{gen}$ a sufficient condition would be

$$\sum_{s \in S_{1+}^{\infty}} \|s\|_{gen} |d_s|, \quad (5.7)$$

indeed conditions of this type have been shown to be related to other notions of mixing for lattice processes, see e.g. Corollary (1.7.2), pg. 32, in [Guyon \(1995\)](#) and [Robinson \(2007\)](#). Conditions like (5.6) and (5.7) impose greater smoothness on the spectral density as compared to absolute summability of the d_s . As both [Shibata \(1981\)](#) and [Karagrigoriou \(1997\)](#) show, the optimality property of p^* minimizing $S_N(p)$ is shared by Akaike Information Criterion (AIC)-like criteria, such as the AIC itself and Final Prediction Error (FPE). On the other hand, these criteria are inconsistent. Furthermore, [Karagrigoriou \(1997\)](#) also shows that consistent criteria such as BIC (Bayesian Information Criterion) and the ϕ criterion of [Hannan and Quinn \(1979\)](#) are asymptotically inefficient in the sense of [Shibata \(1981\)](#), i.e. they do not attain a lower bound for the IRSE. A referee has pointed out that consistency of order selection criteria is proved in settings where the true autoregression is finite, whereas we consider an approximation to a true infinite model (as does [Shibata \(1981\)](#)). Thus we opt for potentially optimal, as opposed to potentially consistent, criteria in the following paragraph. We say ‘potentially’ because we do not establish asymptotic properties of the criteria in this paper. Further support for this choice is provided by the results of [Shibata \(1986\)](#), which show that a consistent model selection criterion can lead to the loss of the $N^{\frac{1}{2}}$ -consistency property of parameter estimates.

In view of the above discussion, we propose versions of the FPE criterion. For spatial processes, the FPE has been extended at least in the quarter-plane case, see e.g. [Tjøstheim \(1981\)](#). Following this approach, for the half-plane setting we can take

$$\widehat{\text{FPE}}(p) = \hat{\sigma}_{\mathfrak{h}(p)}^2 \left(\frac{N + \mathfrak{h}(p)}{N - \mathfrak{h}(p)} \right), \quad (5.8)$$

However, the preceding sections stress that unlike in the case when $d = 1$ or indeed the quarter-plane case, we have $\mathfrak{C}(p) \geq \mathfrak{h}(p)$ when using a half-plane representation. Thus while we fit an AR model with $\mathfrak{h}(p)$ coefficients in fact the estimation is based on up to $\mathfrak{C}(p)$ autocovariances. The formula in (5.8) penalizes only with respect to $\mathfrak{h}(p)$, but realizing that in fact the correct penalty term is $\mathfrak{C}(p)$ suggests a degrees of freedom correction of the form

$$\widetilde{\text{FPE}}(p) = \hat{\sigma}_{\mathfrak{h}(p)}^2 \left(\frac{N + \mathfrak{C}(p)}{N - \mathfrak{C}(p)} \right). \quad (5.9)$$

Finally because $\mathfrak{C}(p)$ is only an upper bound, the degrees of freedom penalty imposed in (5.9) may be too strong so, taking $\mathfrak{a}(p) = (\mathfrak{h}(p) + \mathfrak{C}(p))/2$, we define

$$\overline{\text{FPE}}(p) = \hat{\sigma}_{\mathfrak{h}(p)}^2 \left(\frac{N + \mathfrak{a}(p)}{N - \mathfrak{a}(p)} \right). \quad (5.10)$$

Note that the paper does not present rigorous justification of these criteria, rather treating them as reasonable empirical rules.

The order selection criteria (5.8), (5.9) and (5.10) are displayed in Table 5.5, for $d = 2$ and $n^* = 11$. For all values of τ , $\widehat{\text{FPE}}(p)$ tends to overfit, which corresponds to our discussion above about this criterion imposing an insufficiently large penalty for larger truncation points. On the other hand, both $\widetilde{\text{FPE}}(p)$ and $\overline{\text{FPE}}(p)$ suggest a very parsimonious fit with $p = 1$ in all but one case, and from the analysis of Figures 5.1, 5.2 and 5.3 this certainly seems a better automatic data-driven choice of p than $p \geq 4$ suggested by $\widehat{\text{FPE}}(p)$. Furthermore, $\overline{\text{FPE}}(p)$ suggests $p = 2$ when $\tau = 0.10$, which seems to be a very reasonable choice based on the discussions above. Further illustration of the order selection criteria with real data is provided in supplementary appendix S.1.

5.3 Nonstationary and possibly irregularly-spaced data

In this subsection we illustrate the method further by also applying it to data simulated from a nonstationary process with irregular spacing. For $n^* = 9, 15$, let W be an $N \times N$ circulant matrix with zero diagonals and first row given by $(0, 1/2, 0, \dots, 0, 1/2)$. In each of 500 replications, generate the $N \times 1$ vector of data x by the Cliff and Ord (1973)-type SAR model as follows

$$x = (I_N - 0.05W)^{-1} \epsilon, \quad (5.11)$$

where ϵ is generated by independent standard normal draws. Clearly, the elements x_t , $t = 1, \dots, N$, of x are nonstationary stochastic processes. We compare spectrum estimates from two settings: the first one assumes the availability of data at all lattice points, i.e. a regular lattice, and the second method assumes a certain proportion of lattice points have no data available, i.e. an irregular lattice. These missing points of observation are randomly assigned in each Monte Carlo trial and are approximately 8% of the data when $n^* = 9$ and 7% of the data when $n^* = 15$. Following the approach of practitioners, missing data are replaced with zeros. We wish to compare the performance of these estimates, which we carry out via analysis of Figures 5.5 and 5.6 for the regular and irregular lattices respectively. Figure 5.5 plots the estimated spectrum for $n^* = 9$ when $p = 1, 2$ in panels (a) and (b) respectively, and for $n^* = 15$ when $p = 1, 2, 3, 4$ in panels (c), (d), (e) and (f) respectively. Figure 5.6 displays exactly the same plots for the irregular lattice. In both figures we see clearly the pitfalls of overfitting: too large a value of p leads to a spectrum that is very uninformative, evident in panels (b), (e) and (f) of both figures. On the other hand, comparing the Figures 5.5(a) and 5.6(a) we see that irregular spacing leads to a smaller estimated peak at frequency $\mathbf{0}$. In the same vein, when $n^* = 15$ a comparison of Figures 5.5(c) and 5.6(c) suggests that a

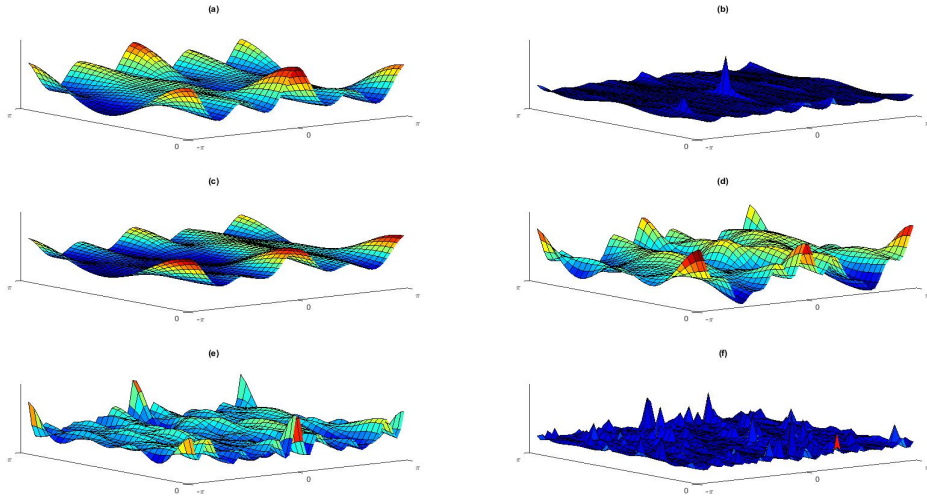


Figure 5.5: Spectral estimates for $d = 2$, $n^* = 9, 15$, $\tau = 0.05$, nonstationary data generated by (5.11). Regular lattice. (a) $n^* = 9, p = 1$, (b) $n^* = 9, p = 2$, (c) $n^* = 15, p = 1$, (d) $n^* = 15, p = 2$, (e) $n^* = 15, p = 3$, (f) $n^* = 15, p = 4$.

similar problem does not arise in this case, perhaps due to loss of data being mitigated by a larger sample. However, this conclusion may be misleading. Indeed, when $p = 2$ we can compare Figures 5.5(d) and 5.6(d) to notice that the peak at frequency $\mathbf{0}$ is indeed flatter in the latter, as is a ripple that runs through the spectrum at across the zero frequency axis in one dimension.

A Proofs

This section contains proof of all results in the paper. We will write $\Delta_{\mathfrak{C}(p)} = \hat{\Psi}_{\mathfrak{h}(p)} - \Psi_{\mathfrak{h}(p)}$ and $\delta_{\mathfrak{h}(p)} = \hat{\psi}_{\mathfrak{h}(p)} - \psi_{\mathfrak{h}(p)}$ throughout the appendices.

Proof of Theorem 2.1. We have

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

so that the norm of the LHS above is bounded by

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

Now $\left\| \hat{\Psi}_{\mathfrak{h}(p)}^{-1} \right\| \leq \left\| \hat{\Psi}_{\mathfrak{h}(p)}^{-1} - \Psi_{\mathfrak{h}(p)}^{-1} \right\| + \left\| \Psi_{\mathfrak{h}(p)}^{-1} \right\| \leq \left(\left\| \hat{\Psi}_{\mathfrak{h}(p)}^{-1} \right\| \left\| \Delta_{\mathfrak{C}(p)} \right\| + 1 \right) \left\| \Psi_{\mathfrak{h}(p)}^{-1} \right\|$, 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 B.5 it follows that $\left\| \Delta_{\mathfrak{C}(p)} \right\| \xrightarrow{p} 0$ if $\mathfrak{C}(p)^v n_p^{1-v} \rightarrow 0$, i.e., $\mathfrak{C}(p)^v N^{1-v} \left(\prod_{i=1}^d (1 - n_i^{-1} p_i) \right)^{1-v} \rightarrow 0$, which is true by (2.11). Thus $\text{plim}_{N \rightarrow \infty} \left\| \hat{\Psi}_{\mathfrak{h}(p)}^{-1} \right\| \leq \lim_{N \rightarrow \infty} \left\| \Psi_{\mathfrak{h}(p)}^{-1} \right\| < \infty$, from Corollary B.7. Now we deal with the factor in parentheses in (A.1). By Lemma B.4, Markov's

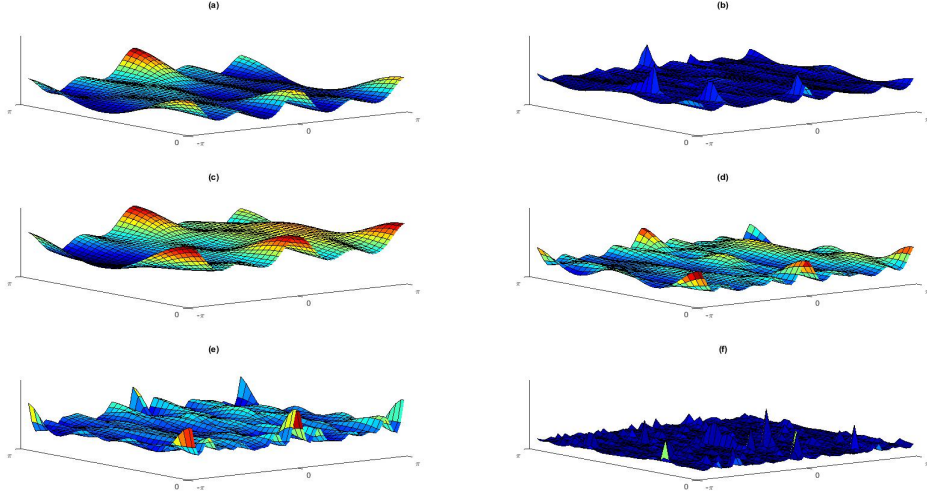


Figure 5.6: Spectral estimates for $d = 2$, $n^* = 9, 15$, $\tau = 0.05$, nonstationary data generated by (5.11). Irregular lattice. (a) $n^* = 9, p = 1$, (b) $n^* = 9, p = 2$, (c) $n^* = 15, p = 1$, (d) $n^* = 15, p = 2$, (e) $n^* = 15, p = 3$, (f) $n^* = 15, p = 4$.

inequality and (2.11), $\|\delta_{\mathfrak{h}(p)}\| \xrightarrow{p} 0$. For the second term, we have $\|\Delta_{\mathfrak{C}(p)}\| \xrightarrow{p} 0$ and also $\|d_{\mathfrak{h}(p)}\| = \left(\sum_{s \in S[-pL, pU]} d_s^2\right)^{\frac{1}{2}} \leq \sum_{s \in S[-pL, pU]} |d_s| \leq \sum_{s \in S_{1+}^{\infty}} |d_s| < \infty$. Thus the second term converges to zero in probability. Finally, for the third term note that Assumption A 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 $\gamma(k) = E x_t x_{t-k} = \sum_{t \in S_{1+}^{\infty}} d_t \gamma(t-k)$, $k \in S_{1+}^{\infty}$, so $\|\Psi_{\mathfrak{h}(p)} d_{\mathfrak{h}(p)} - \psi_{\mathfrak{h}(p)}\|^2$ is

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

using Lemma B.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 (2.12), completing the proof. Note that we have also shown that

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

by Markov's inequality. \square

Proof of Theorem 2.2. Note that $\hat{\gamma}(\mathbf{0}) = n_p^{-1} \sum_{t(p,n)}'' x_t^2$. Using the definition of least squares and some algebra, we may write $\hat{\sigma}_{\mathfrak{h}(p)}^2 - \sigma^2$ as $n_p^{-1} \sum_{t(p,n)}'' \left(x_t - \sum_{s \in S[-p_L, p_U]} \hat{d}_{s, \mathfrak{h}(p)} x_{t-s} \right)^2 - \sigma^2$, which equals

$$\begin{aligned} \hat{\gamma}(\mathbf{0}) - \hat{d}'_{\mathfrak{h}(p)} \hat{\psi}_{\mathfrak{h}(p)} - \sigma^2 &= \hat{\gamma}(\mathbf{0}) - \left(\hat{d}_{\mathfrak{h}(p)} - d_{\mathfrak{h}(p)} \right)' \hat{\psi}_{\mathfrak{h}(p)} - d'_{\mathfrak{h}(p)} \hat{\psi}_{\mathfrak{h}(p)} - \gamma(\mathbf{0}) + \sum_{t \in S_{1+}^{\infty}} d_t \gamma(t) \\ &= \hat{\gamma}(\mathbf{0}) - \gamma(\mathbf{0}) - \left(\hat{d}_{\mathfrak{h}(p)} - d_{\mathfrak{h}(p)} \right)' \psi_{\mathfrak{h}(p)} - d'_{\mathfrak{h}(p)} \delta_{\mathfrak{h}(p)} - \left(\hat{d}_{\mathfrak{h}(p)} - d_{\mathfrak{h}(p)} \right)' \delta_{\mathfrak{h}(p)} - d'_{\mathfrak{h}(p)} \psi_{\mathfrak{h}(p)} \\ &+ \sum_{t \in S_{1+}^{\infty}} d_t \gamma(t). \end{aligned}$$

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

$$\begin{aligned} \hat{\sigma}_{\mathfrak{h}(p)}^2 - \sigma^2 &= (\hat{\gamma}(\mathbf{0}) - \gamma(\mathbf{0})) - \left(\hat{d}_{\mathfrak{h}(p)} - d_{\mathfrak{h}(p)} \right)' \psi_{\mathfrak{h}(p)} - d'_{\mathfrak{h}(p)} \delta_{\mathfrak{h}(p)} \\ &- \left(\hat{d}_{\mathfrak{h}(p)} - d_{\mathfrak{h}(p)} \right)' \delta_{\mathfrak{h}(p)} + \sum_{t \in S_{1+}^{\infty} \setminus S[-p_L, p_U]} d_t \gamma(t). \end{aligned} \quad (\text{A.3})$$

The first term on the RHS converges to 0 in probability by Lemma B.3 and Markov's inequality, the second by Theorem 2.1 and Lemma B.1, the third term by Lemma B.4, (2.11) and Assumption A and the fourth term by Theorem 2.1, Lemma B.4 and (2.11). For the fifth term, convergence to zero follows by (2.12) and Lemma B.1. Note that we have also proved

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

because $\mathfrak{h}(p) \leq \mathfrak{C}(p)$ and $\lim_{N \rightarrow \infty} N/n_p = 1$. \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

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

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

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

On the other hand,

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

and

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

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

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

by the Cauchy Schwarz inequality. By (3.1), (3.2) and (A.2), we have $\mathfrak{h}(p)^{\frac{1}{2}} \left\| \hat{d}_{\mathfrak{h}(p)} - d_{\mathfrak{h}(p)} \right\| = \mathcal{O}_p \left(\mathfrak{C}(p) \mathfrak{h}(p)^{\frac{1}{2}} / N^{\frac{v-1}{v}} \right)$, implying that (A.9) is negligible. We have then shown that

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

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

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

with probability approaching 1 as $N \rightarrow \infty$. The identity $a^2 - b^2 = (a-b)^2 + 2b(a-b)$ implies $\left| \left| \hat{D}_{\mathfrak{h}(p)}(e^{i\lambda}) \right|^2 - \left| D(e^{i\lambda}) \right|^2 \right|$ is bounded by

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

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

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

Because (A.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 (A.6), (A.11), (A.13) and Theorem 2.2. \square

Proof of Theorem 4.1. By Lemma B.8 and (2.2), we need to establish the asymptotic distribution of

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

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

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

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

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

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 independent 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 $N^{-\frac{1}{2}} \sum_{j=1}^N v_N(j)$. For μ defined in (B.11), we now show that

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

The uncorrelatedness of $\delta_N(j)$ implies $E v_N^2(j) = (N^2 \sigma^2 / n_p^2 \mathfrak{h}(p)) \alpha(p)' \Psi_{\mathfrak{h}(p)}^{-1} \Psi_{\mathfrak{h}(p), M} \Psi_{\mathfrak{h}(p)}^{-1} \alpha(p)$ for any j , where $\Psi_{\mathfrak{h}(p), M}$ denotes the symmetric matrix with elements $\sigma^2 \sum_{r \in S[-M, M]} b_r b_{r+s-t}$, with $s, t \in S[-p_L, p_U]$. Elementary inequalities together with (A.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

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

Now $N^{-1} \sum_{j=1}^N E v_N^2(j)$ is bounded by $(N^2 \sigma^2 / n_p^2 \mathfrak{h}(p)) \|\alpha(p)\|^2 \|\Psi_{\mathfrak{h}(p)}^{-1}\|^2 \|\Psi_{\mathfrak{h}(p), M} - \Psi_{\mathfrak{h}(p)}\| + \sigma^2 (N^2 \sigma^2 / n_p^2 \mathfrak{h}(p)) \alpha(p)' \Psi_{\mathfrak{h}(p)}^{-1} \alpha(p)$. The first term is easily seen to be $o(\eta)$ as $N \rightarrow \infty$, by (A.18), while the second converges to $\sigma^2 \mu$ as $N \rightarrow \infty$ by Lemma B.10 and because $N/n_p \rightarrow 1$ as $N \rightarrow \infty$. Thus (A.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

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

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

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

By (A.17), $E(N^{-1} u_N^2(j)) = \sigma^{-2} \mu^{-1} (\sigma^2 \mu + o(\eta) + o(1)) = \mathcal{O}(1)$ uniformly in j , implying that $N^{-1} u_N^2(j)$ is a uniformly integrable array under (4.1), whence (A.19) follows on noticing that its LHS 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, (A.20) is proved if we show

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

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

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

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

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

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

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 (A.23) because $\ell_{j, N, i}$ depend on N . They show that (A.23) is $\mathcal{O}(N^{-\frac{\kappa}{2}})$, whence (A.21) is $\mathcal{O}(\mathfrak{h}(p)N^{-\frac{\kappa}{4}}) = o(1)$, unlike when $d = 1$, when (A.23) is $\mathcal{O}(N^{-1})$ and (A.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 Lemma 4.1. 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 $\lambda \in \Pi$, we can replace $C(\lambda)$ and $S(\lambda)$ in (4.3) 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 B.8 and Theorem 4.1 immediately provide the joint asymptotic normality of (4.3), 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 B.10, 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 $(N/\mathfrak{h}(p))^{\frac{1}{2}} (\hat{S}_{\mathfrak{h}(p)}(\lambda_j) - S(\lambda_j))$. It is easy to show using this method that the asymptotic variance of the sum of any pair of terms (4.3) is the sum of the asymptotic variances, implying that the asymptotic covariance matrix is diagonal. \square

Proof of Theorem 4.2. By (3.1) and (4.2), $(N/\mathfrak{h}(p))^{\frac{1}{2}} (\hat{\sigma}_{\mathfrak{h}(p)}^2 - \sigma^2) = o_p(1)$, with the five terms on the RHS of (A.3) shown to be negligible as in Berk (1974), noting that the bound achieved in (A.4) can be sharpened to the extent required under the conditions of the the-

orem. 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 4.1 and the delta method, so we omit the details. \square

B Lemmas

All proofs not provided below are included in the online appendix.

Lemma B.1. *Suppose Assumption A holds. Then $\sum_{k \in \mathbb{Z}^d} |\gamma(k)| < \infty$.*

Proof. Standard. \square

Lemma B.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*

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

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

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

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 B.3. *If Assumptions A and B hold, $E |\hat{\gamma}(k) - \gamma(k)|^v \leq C n_p^{1-v}$.*

Lemma B.4. *If Assumptions A and B hold, $E \|\delta_{\mathfrak{h}(p)}\|^v \leq C \mathfrak{h}(p)^v n_p^{1-v}$.*

Lemma B.5. *If Assumptions A and B hold, $E \|\Delta_{\mathfrak{e}(p)}\|^v \leq C \mathfrak{e}(p)^v n_p^{1-v}$.*

Proof. Write $\check{\Delta}_{\mathfrak{e}(p)} = \hat{\Psi}_{\mathfrak{h}(p)} - \check{\Psi}_{\mathfrak{h}(p)}$, 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

$$\|\Delta_{\mathfrak{e}(p)}\| \leq \|\Delta_{\mathfrak{e}(p)}\|_R \leq \|\check{\Delta}_{\mathfrak{e}(p)}\|_R. \quad (\text{B.3})$$

We will now bound the absolute row-sums of $\check{\Delta}_{\mathfrak{e}(p)}$ uniformly over all rows. Consider a typical row of $\check{\Delta}_{\mathfrak{e}(p)}$. This has $\hat{\gamma}(l_1 - \bar{l}_1, l_2 - \bar{l}_2, \dots, l_d - j_d) - \gamma(l_1 - \bar{l}_1, l_2 - \bar{l}_2, \dots, l_d - j_d)$; $j_d = 0, \dots, p_d$, 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

$$\sum_{d-1} \sum_{j_d=0}^{p_d} \left| \hat{\gamma}(l_1 - \bar{l}_1, l_2 - \bar{l}_2, \dots, l_d - j_d) - \gamma(l_1 - \bar{l}_1, l_2 - \bar{l}_2, \dots, l_d - j_d) \right| \quad (\text{B.4})$$

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), (B.4) is bounded by

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

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 ST.1. This bound is clearly uniform over all possible rows. So using Hölder's inequality and Lemma B.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 (B.3). □

Lemma B.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 B.7. *Under the conditions of Lemma B.6, $\|\Psi_{\mathfrak{h}(p)}^{-1}\| \leq C$.*

For any index t in the sum $\sum_{t(|p|,n)}''$ 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 B.8. *Let Assumptions A, B, C and (3.1) hold, with $v = 2$, and*

$$N^{\frac{1}{2}} \sum_{s \in S_{1+}^{\infty} \setminus S[-p_L, p_U]} |d_s| \longrightarrow 0, \text{ as } N \rightarrow \infty, \quad (\text{B.5})$$

Then, as $N \rightarrow \infty$,

$$N^{\frac{1}{2}} \alpha(p)' \left(\hat{d}_{\mathfrak{h}(p)} - d_{\mathfrak{h}(p)} \right) / \mathfrak{h}(p)^{\frac{1}{2}} - N^{\frac{1}{2}} \sum_{t(p,n)}'' \alpha(p)' \Psi_{\mathfrak{h}(p)}^{-1} X_t(p) \epsilon_t / n_p \mathfrak{h}(p)^{\frac{1}{2}} \xrightarrow{p} 0. \quad (\text{B.6})$$

Proof. 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 (B.6) equals

$$\begin{aligned} & N^{\frac{1}{2}}\alpha(p)'\hat{\Psi}_{\mathfrak{h}(p)}^{-1}\Delta\mathfrak{C}(p)\Psi_{\mathfrak{h}(p)}^{-1}\sum_{t(p,n)}''X_t(p)\bar{\epsilon}_{t,\mathfrak{h}(p)}/n_p\mathfrak{h}(p)^{\frac{1}{2}} \\ & + N^{\frac{1}{2}}\alpha(p)'\Psi_{\mathfrak{h}(p)}^{-1}\sum_{t(p,n)}''X_t(p)\sum_{s\in S_{1+}^{\infty}\setminus S[-p_L,p_U]}d_sx_{t-s}/n_p\mathfrak{h}(p)^{\frac{1}{2}}. \end{aligned} \quad (\text{B.7})$$

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 B.7. Thus the square of the second term in (B.7) 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), so $E\left\|\sum_{t(p,n)}''X_t(p)\bar{\epsilon}_{t,\mathfrak{h}(p)}\right\|^2 = \mathcal{O}\left(\mathfrak{h}(p)N^{\frac{1}{2}}n_p^{\frac{1}{2}}\sum_{s\in S_{1+}^{\infty}\setminus S[-p_L,p_U]}d_s^2\right)$, and the first term in (B.7) is in turn of order $\mathcal{O}_p\left(\mathfrak{h}(p)^{\frac{1}{2}}\|\Delta\mathfrak{C}(p)\|N^{\frac{1}{4}}n_p^{\frac{1}{4}}\sum_{s\in S_{1+}^{\infty}\setminus S[-p_L,p_U]}|d_s|\right) = \mathcal{O}_p\left(\mathfrak{h}(p)^{\frac{1}{2}}\mathfrak{C}(p)/n_p^{\frac{1}{2}}\right)\alpha_p(1)$, by Lemmas B.5, B.7, (B.5). This is negligible by (3.1). \square

Lemma B.9. Write $D_{\mathfrak{h}(p)}(z) = 1 - \sum_{s\in S[-p_L,p_U]}d_{s,\mathfrak{h}(p)}z^s$, $D(z) = 1 - \sum_{s\in S_{1+}^{\infty}}d_s z^s$ and let Assumption A hold. Then $\lim_{\mathfrak{h}(p)\rightarrow\infty}D_{\mathfrak{h}(p)}(z) = D(z)$ for $|z_i| \leq 1$, $i = 1, \dots, d$.

Proof. We can take $\lambda = \mathbf{0}$ in Theorem 2.2 of Baxter (1962)¹, as in Berk (1974), and obtain

$$\sum_{r\in S[-p_L,p_U]\cup\mathbf{0}}\left|d_{r,\mathfrak{h}(p)}/\sigma_{\mathfrak{h}(p)}^2 - d_r/\sigma^2\right| \leq C \sum_{r\in S_{1+}^{\infty}\setminus S[-p_L,p_U]}|d_r|/\sigma^2, \quad (\text{B.8})$$

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

$$\sigma_{\mathfrak{h}(p)}^2 - \sigma^2 = \gamma(\mathbf{0}) - d'_{\mathfrak{h}(p)}\psi_{\mathfrak{h}(p)} - \sigma^2 = \sum_{r\in S_{1+}^{\infty}\setminus S[-p_L,p_U]}d_r\gamma(r) \rightarrow 0, \quad (\text{B.9})$$

as $\mathfrak{h}(p) \rightarrow \infty$, by (2.4) and Lemma B.1. Combining (B.8) and (B.9) yields the result. \square

Lemma B.10. Let the conditions of Lemma B.9 hold. 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

$$\beta_t = w_0 + w_1e^{it'\lambda_1} + \dots + w_qe^{it'\lambda_q} + w_0e^{it'\pi} + u_1e^{-it'\lambda_1} + \dots + u_qe^{-it'\lambda_q}, \quad (\text{B.10})$$

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

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

¹Meyer et al. (2016) have proved a version of this inequality under (5.7), which is a regularity condition that implies greater smoothness of $f(\lambda)$.

where

$$\mu = w_0^2/(2\pi)^d f(\mathbf{0}) + 2w_1 u_1/(2\pi)^d f(\lambda_1) + \cdots + 2w_q u_q/(2\pi)^d f(\lambda_q) + u_0^2/(2\pi)^d f(\boldsymbol{\pi}). \quad (\text{B.11})$$

References

- Anderson, T. W., (1971). *The Statistical Analysis of Time Series*. John Wiley & Sons.
- Andrews, D. W. K., (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
- Bandyopadhyay, S., Lahiri, S. N., Nordman, D. J., (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.
- Bester, C. A., Conley, T. G., Hansen, C. B., (2011). Inference with dependent data using cluster covariance estimators. *Journal of Econometrics* 165, 137–151.
- Bester, C. A., Conley, T. G., Hansen, C. B., Vogelsang, T. J., (2016). Fixed-b asymptotics for spatially dependent robust nonparametric covariance matrix estimators. *Econometric Theory* 32, 154–186.
- Bhansali, R. J., (1980). Autoregressive and window estimates of the inverse correlation function. *Biometrika* 67, 551–566.
- Brillinger, D. R., (1969). Asymptotic properties of spectral estimates of second order. *Biometrika* 56, 375–390.
- Brillinger, D. R., (1975). *Time Series: Data Analysis and Theory*. Holt, Rinehart and Winston.
- Chen, X., Nordhaus, W. D., (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.
- Cliff, A. D., Ord, J. K., (1973). *Spatial Autocorrelation*. London: Pion.
- Cliff, A. D., Ord, J. K., (1972). Testing for spatial autocorrelation among regression residuals. *Geographical Analysis* 4, 267–284.
- Conley, T. G., (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics* 92, 1–45.
- Conley, T. G., Molinari, F., (2007). Spatial correlation robust inference with imperfect distance information. *Journal of Econometrics* 140, 76–96.
- Conley, T. G., Topa, G., (2007). Estimating dynamic local interactions models. *Journal of Econometrics* 140, 282–303.
- Dahlhaus, R., Künsch, H., (1987). Edge effects and efficient parameter estimation for stationary random fields. *Biometrika* 74, 877–882.

- Delgado, M., Robinson, P. M., (2015). Non-nested testing of spatial correlation. *Journal of Econometrics* 187, 385–401.
- Driscoll, J. C., Kraay, A. C., (1998). Consistent covariance matrix estimation with spatially-dependent panel data. *Review of Economics and Statistics* 80, 549–560.
- Fuentes, M., (2007). Approximate likelihood for large irregularly spaced spatial data. *Journal of the American Statistical Association* 102, 321–331.
- Grenander, U., Szegö, G., (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.
- Guyon, X., (1995). *Random Fields on a Network: Modeling, Statistics and Applications*. Springer-Verlag, New York.
- 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.
- Hannan, E. J., Quinn B. G., (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B* 41, 190–195.
- Helson, H., Lowdenslager, D., (1958). Prediction theory and Fourier series in several variables. *Acta Mathematica* 99, 165–202.
- Helson, H., Lowdenslager, D., (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., Seo, M. H., (2014). Specification tests for lattice processes. *Econometric Theory* 31, 294–336.
- Jenish, N., (2016). Spatial semiparametric model with endogenous regressors. *Econometric Theory* 32, 714–739.
- Jenish, N., Prucha, I. R., (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics* 150, 86–98.
- Jenish, N., Prucha, I. R., (2012). On spatial processes and asymptotic inference under near-epoch dependence. *Journal of Econometrics* 170, 178 – 190.
- Jowett, G. H., (1955). The comparison of means of sets of observations from sections of independent stochastic series. *Journal of the Royal Statistical Society. Series B* 17, 208–227.
- Karagrigoriou, A., (1997). Asymptotic efficiency of the order selection of a nongaussian AR process. *Statistica Sinica* 7, 407–423.
- Kelejian, H. H., Prucha, I. R., (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40, 509–533.
- Kelejian, H. H., Prucha, I. R., (2007). HAC estimation in a spatial framework. *Journal of Econometrics* 140, 131–154.

- Kim, M. S., Sun, Y., (2011). Spatial heteroskedasticity and autocorrelation consistent estimation of covariance matrix. *Journal of Econometrics* 160, 349–371.
- Kim, M. S., Sun, Y., (2013). Heteroskedasticity and spatiotemporal dependence robust inference for linear panel models with fixed effects. *Journal of Econometrics* 177, 85–108.
- Korezlioglu, H., Loubaton, P., (1986). Spectral factorization of wide sense stationary processes on \mathbb{Z}^2 . *Journal of Multivariate Analysis* 19, 24–47.
- Lee, L. F., (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72, 1899–1925.
- Lysne, D., Tjøstheim, D., (1987). Loss of spectral peaks in autoregressive spectral estimation. *Biometrika* 74, 200–206.
- Matsuda, Y., Yajima, Y., (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.
- Meyer, M., Jentsch, C., Kreiss J.-P., (2017). Baxter’s inequality and sieve bootstrap for random fields. *Bernoulli* 23, 2988–3020.
- Newey, W. K., West, K. D., (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation-consistent covariance matrix. *Econometrica* 55, 703–708.
- Nordhaus, W. D., (2008). New metrics for environmental economics: Gridded economic data. *The Integrated Assessment Journal* 8, 73–84.
- 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., Romano J. P., (1996). On flat-top kernel spectral density estimators for homogeneous random fields. *Journal of Statistical Planning and Inference* 51, 41 – 53.
- Robinson, P. M., (1979). Distributed lag approximation to linear time-invariant systems. *The Annals of Statistics* 7, 507–515.
- Robinson, P. M., (2005). Robust covariance matrix estimation: HAC estimates with long memory/antipersistence correction. *Econometric Theory* 21, 171–180.
- 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., (2011). Asymptotic theory for nonparametric regression with spatial data. *Journal of Econometrics* 165, 5–19.
- Robinson, P. M., Thawornkaiwong, S., (2012). Statistical inference on regression with spatial dependence. *Journal of Econometrics* 167, 521–542.
- Robinson, P. M., Vidal Sanz, J., (2006). Modified Whittle estimation of multilateral models on a lattice. *Journal of Multivariate Analysis* 97, 1090–1120.
- Roknossadati, S. M., Zarepour, M., (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.

- Sain, S. R., Cressie, N., (2007). A spatial model for multivariate lattice data. *Journal of Econometrics* 140, 226 – 259.
- 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.
- Shibata, R., (1981). An optimal autoregressive spectral estimate. *Annals of Statistics* 9, 300–306.
- Shibata, R., (1986). Consistency of model selection and parameter estimation. *Journal of Applied Probability* 23, 127–141.
- 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.
- Topa, G., (2001). Social interactions, local spillovers and unemployment. *Review of Economic Studies* 68, 261–295.
- Vidal Sanz, J., (2009). Automatic spectral density estimation for random fields on a lattice via bootstrap. *TEST* 18, 96–114.
- Vogelsang, T. J., (2012). Heteroskedasticity, autocorrelation, and spatial correlation robust inference in linear panel models with fixed-effects. *Journal of Econometrics* 166, 303–319.
- Wang, H., Iglesias, E. M., Wooldridge J. M., (2013). Partial maximum likelihood estimation of spatial probit models. *Journal of Econometrics* 172, 77–89.
- Wester, R., Tummala, M., Therrien, C. W., (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., Subba Rao, T., (1993). Higher order spectral estimation for random fields. *Multi-dimensional Systems and Signal Processing* 4, 7–22.

Supplementary appendix to ‘Autoregressive spatial spectral estimates’

Abhimanyu Gupta

November 20, 2017

This supplementary appendix contains an empirical example in Section S.1, bounds for absolute moments of partial sums of rather general lattice processes in Section S.2, a generalisation of the Toeplitz property familiar from the theory of stationary time series and an upper bound on the number of unique autocovariances that occur in the covariance matrix of finite, stationary and unilateral processes in Section S.3, and proofs of most of the lemmas in Appendix B of the paper. It also strengthens the mode of convergence of estimates to almost sure convergence in Section S.5

S.1 Empirical example

The data used is available at www.spatial-statistics.com. 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 S.1.1 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. Smaller grid cells would lead to 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. Voter turnout is not a zero mean process, so we subtract the sample mean using the whole sample from each cell as remarked in Section 2.

We now apply the order selection methods given in (5.8), (5.9) and (5.10). Table S.1.1 tabulates all the versions of FPE for various truncation order choices $p^* = p_{U_1} = p_{L_2} = p_{U_2}$. The advantage of our degree of freedom correction in $\widetilde{\text{FPE}}(p)$ is immediately clear. The ‘naive’ extension given by $\widehat{\text{FPE}}(p)$ continues to fall and suggest $p^* = 5$, which leads to an extremely choppy and uninformative spectrum, as discussed in the next paragraph. On the other hand, $\widetilde{\text{FPE}}(p)$ gives a choice of $p^* = 3$, the estimated spectrum from which is more informative. The average degrees of freedom corrected version $\overline{\text{FPE}}(p)$ suffers from the same overfitting problem as $\widehat{\text{FPE}}(p)$.

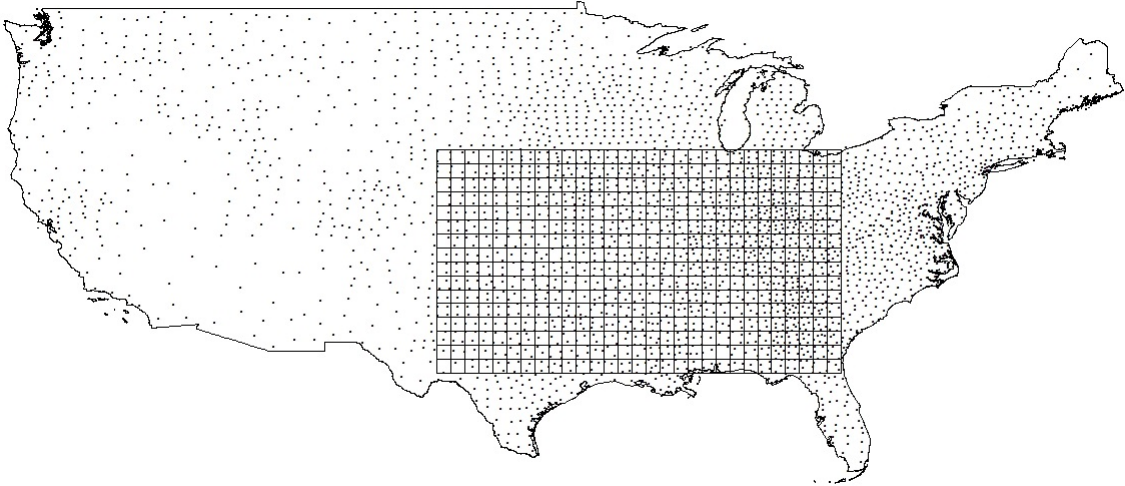


Figure S.1.1: Gridded county level US voter turnout data

p^*	$\widehat{\text{FPE}}(p)$	$\widetilde{\text{FPE}}(p)$	$\overline{\text{FPE}}(p)$
1	0.0032	0.0033	0.0033
2	0.0026	0.0027	0.0026
3	0.0023	0.0025	0.0024
4	0.0023	0.0026	0.0024
5	0.0020	0.0026	0.0023

Table S.1.1: Order selection via FPE for county level US presidential election data

These findings are illustrated in Figure S.1.2, which plots the estimated spectrum for various choices of p on a \log_{10} scale. Due to symmetry we only plot the results over $(-\pi, \pi] \times [0, \pi]$. There is a very strong peak at low frequency, indicating the power in low frequency structural components. For $p = 1$, panel (a) shows that the estimated spectrum tends to not pick up the features of the spectrum and seems to suffer from underfitting. Looking at panel (b), we see that matters improve when $p = 2$ and more features of the spectrum, in particular two peaks, can be discerned. For $p = 3$ (panel (c)) the peaks appear to at their sharpest and best defined, while the signs of overfitting that only just start to appear in panel (d) ($p = 4$) progressively get more deleterious in panels (e) and (f), i.e. when $p = 5$ and 6.

The periodogram spectral estimate with $m_1 = m_2 = m^* = 2, 4, 6, 8$ is plotted in Figure S.1.3. The estimate has features that broadly match those seen in the AR estimate in Figure S.1.2, viz. it also exhibit two strong peaks. However the peaks are sharper and well-defined in Figure S.1.2, whereas local features and ripples seem to better represented in Figure S.1.3. Given the local and global character of the tapered periodogram and AR estimate respectively, it is not surprising to see these features in the spectrum estimate. Furthermore, larger values of m^* lead to better defined peaks.

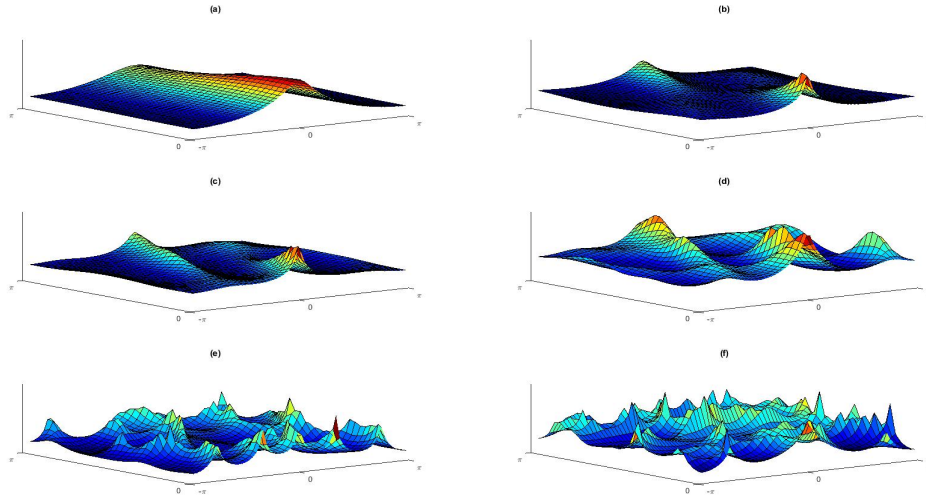


Figure S.1.2: Autotregressive spectral density estimate for county level US presidential election data, (a) $p = 1$, (b) $p = 2$, (c) $p = 3$, (d) $p = 4$, (e) $p = 5$, (f) $p = 6$.

S.2 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 A. ξ_{st} are mean-zero and independent over $t \in \mathbb{L}$.

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

$$E |\xi_{st}|^w < \eta_{\mathbf{s}}^w a_t^w, \quad (\text{S.2.1})$$

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

$$\sum_{s \in \mathbb{Z}^d} \eta_{ks} < \infty, \quad 1 \leq k \leq q. \quad (\text{S.2.2})$$

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 $|M_i| < L_i$,

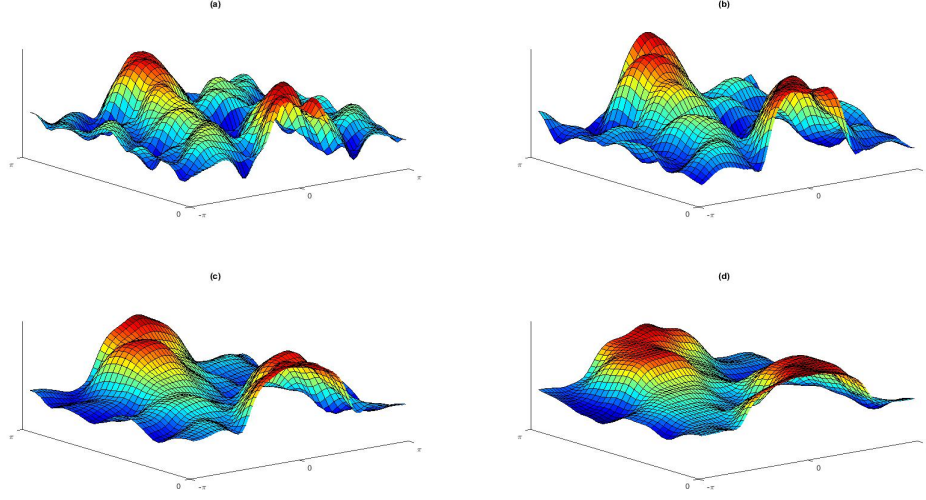


Figure S.1.3: Average tapered periodogram spectral density estimate for county level US presidential election data, (a) $m^* = 2$, (b) $m^* = 4$, (c) $m^* = 6$, (d) $m^* = 8$.

and define $S_{ML} = \sum''_{t(|M|,L)} \zeta_t$, where $\sum''_{t(|M|,L)}$ runs over t satisfying

$$\begin{aligned} -n_{L_i} < t_i \leq L_i - |M_i| - n_{L_i}; & \text{ if } M_i < 0, \\ M_i - n_{L_i} < t_i \leq L_i - n_{L_i}; & \text{ if } M_i \geq 0, \end{aligned} \quad (\text{S.2.3})$$

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 S.2.1 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 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

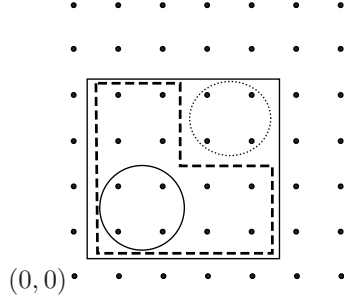


Figure S.2.1: Illustration of S_L, S_M, S_{ML}^* and S_{ML} , $d = 2$, $n_{L_1} = n_{L_2} = 0$; $n_{U_1} = n_{U_2} = 6$; $(L_1, L_2) = (4, 4)$ and $(M_1, M_2) = (2, 2)$.

$i = 1, \dots, d$. We are now in a position to prove the main result of this section.

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

Proof. $S_{MN} = \sum_{t(|M|, N)}'' \sum_{s^1 \in \mathbb{Z}^d} \dots \sum_{s^q \in \mathbb{Z}^d} \xi_{st}$, which is rewritten as

$$S_{MN} = \sum_{s^1 \in \mathbb{Z}^d} \dots \sum_{s^q \in \mathbb{Z}^d} \eta_{1s^1}^{1-1/w} \eta_{1s^1}^{1/w} \sum_{t(|M|, N)}'' (\xi_{st} / \eta_{1s^1}),$$

whence from Hölder's inequality

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

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

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

After q applications of Hölder's inequality and using (S.2.2) we obtain

$$|S_{MN}|^w \leq C \sum_{s^1 \in \mathbb{Z}^d} \dots \sum_{s^q \in \mathbb{Z}^d} \eta_{\mathbf{s}}^{1-w} \left| \sum_{t(|M|, N)}'' \xi_{st} \right|^w. \quad (\text{S.2.4})$$

Also, from von Bahr and Esseen (1965) and (S.2.1)

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

Taking expectations of (S.2.4) and applying the above and (S.2.2) we conclude

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

establishing the lemma. \square

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 SL.1 is quite general. It is similar to Lemma 1 of Robinson (1978) for $d = 1$.

S.3 Properties of covariance matrices of autoregressive lattice processes

S.3.1 A spatial generalisation of the Toeplitz property

In this appendix we generalise 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 matricial 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 $\check{\psi}_{\mathfrak{h}(p)}$ 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 $\Psi_{i-j}^{(1)}(p)$,

$i, j = 0, \dots, p_{U_1}$, so

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

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

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

$\check{\Psi}_{\mathfrak{h}(p)}$ 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 $\check{\Psi}_{\mathfrak{h}(p)}$ 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 $\Psi_{\mathfrak{h}(p)}$ 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 $\psi_{i, l_2, \dots, l_{d-1}}^{(1)}(p)$, $i = 0, \dots, p_{U_1}$, 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 $l_i = -p_{L_i}, \dots, p_{U_i}$, $i = 4, \dots, d - 1$.
- ⋮
- ($d - 2$) 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 $j_0, \dots, j_{\mathfrak{h}(p)-1}$, $j_0 \equiv \mathbf{0}$. Denote by $d_{s, \mathfrak{h}(p)}$ the scalars

$$\arg \min_{a_s, s \in S[-p_L, p_U]} E \left(x_t - \sum_{s \in S[-p_L, p_U]} a_s x_{t-s} \right)^2, \quad (\text{S.3.1})$$

the minimum by $\sigma_{\mathfrak{h}(p)}^2$, and write $\Sigma_{\mathfrak{h}(p)} = \text{diag}(\sigma_0^2, \dots, \sigma_{\mathfrak{h}(p)-1}^2)$. 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

$$L_{\mathfrak{h}(p)} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ d_{j_1, 1} & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ d_{j_{\mathfrak{h}(p)-1}, \mathfrak{h}(p)-1} & \dots & \dots & d_{j_{\mathfrak{h}(p)-1}, \mathfrak{h}(p)-1} & 1 \end{pmatrix}, \quad (\text{S.3.2})$$

we have

$$\Psi_{\mathfrak{h}(p)}^{-1} = L'_{\mathfrak{h}(p)} \Sigma_{\mathfrak{h}(p)}^{-1} L_{\mathfrak{h}(p)}. \quad (\text{S.3.3})$$

S.3.2 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 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 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 AR(0, 2; 1, 1) representation has $\Psi_{\mathfrak{h}(0,2;2,1)}$ 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 ST.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*

$$\mathfrak{c}(p) = 1 + \sum_{l=1}^{d-1} 2^{d-l-1} \sum_{\substack{\#(l=0) \\ \approx 0_d^l}} \prod_{k=1}^d p_k + 2^{d-1} \prod_{k=1}^d p_k, \quad (\text{S.3.4})$$

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 $\prod_{k=1, \neq 0_d}^d$ multiplies over k such that the l zero entries of $(p_1, p_2, \dots, p_d)'$ are excluded.

Proof. Autocovariances can involve several types of indices. Consider first the number of autocovariances with all nonzero entries in the indices. In each dimension there are p_k possible indices, and in all except the first dimension (in which there are no negative indices due to the half-plane representation), there may be negative indices. Thus there are at most $2^{d-1} \prod_{k=1}^d p_k$ such autocovariances, which is the third term in (S.3.4).

The second term in (S.3.4) arises from counting the maximum number of indices that can have a certain number of zero elements. For example, the case of the first index entry being zero and the rest nonzero will give rise to at most $2^{d-2} \prod_{k=2}^d p_k$ autocovariances with such indices, while case of the first two index entries being zero and the rest nonzero will give rise to at most $2^{d-3} \prod_{k=3}^d p_k$ and so on. The second term in (S.3.4) then accounts for the fact that a single zero index entry may occur in d possible places, two zero index entries may occur in $\binom{d}{2}$ ways, three zero index entries in $\binom{d}{3}$ ways, etc. This completes the explanation of the second term in (S.3.4). Finally, the case with all index elements equal to zero (i.e. the variance) accounts for the last item in (S.3.4). \square

It is clear from the formulae (2.8) and (S.3.4) that

$$\mathfrak{h}(p) \leq \mathfrak{C}(p), \quad (\text{S.3.5})$$

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 (S.3.4) 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.

S.4 Proofs of lemmas in Appendix B

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

Proof of Lemma B.3. For $\hat{\gamma}(k) - \gamma(k)$ to be of the form of S_{pn} in Lemma B.2, define $\xi_{r,s,t} = b_r b_{r-k} (\epsilon_{t-r}^2 - \sigma^2)$, $s = r - k$; $= b_r b_s \epsilon_{t-r} \epsilon_{t-k-s}$, $s \neq r - k$. Then the $\xi_{r,s,t}$ are clearly zero-mean. They are independent because the ϵ_t are. Therefore, they satisfy Assumption A. By the c_r -inequality, Cauchy-Schwarz inequality and Assumption B, we

obtain $E |\xi_{rs,t}|^v \leq 2 |b_r b_{r-k}|^v \left(E |\epsilon_{t-r}|^{2v} + \sigma^{2v} \right) \leq C |b_r b_{r-k}|^v$, $s = r - k$ and $E |\xi_{rs,t}|^v \leq |b_r b_s|^v \left(E |\epsilon_{t-r}|^{2v} E |\epsilon_{t-s}|^{2v} \right)^{\frac{1}{2}} \leq C |b_r b_s|^v$, $s \neq r - k$, verifying that (B.2) holds since the b_r are absolutely summable. The result follows immediately from Lemma B.2. \square

Proof of Lemma B.4.

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

using Hölder's inequality and Lemma B.3. \square

Proof of Lemma B.6. Eigenvalues of $\Psi_{\mathfrak{h}(p)}$ are determined by the generalized Toeplitz form $\sum_{j,k \in S[-pL, pU]} \xi_j \gamma(j-k) \xi_k$, for real numbers ξ_s , $s \in S[-pL, pU]$, $\sum_{s \in S[-pL, pU]} \xi_s^2 = 1$, summing over $j, k \in S[-pL, pU]$ by construction of $\Psi_{\mathfrak{h}(p)}$. This equals

$$\begin{aligned} &\sum_{j,k \in S[-pL, pU]} \int_{\Pi} e^{i(j-k)'\lambda} f(\lambda) d\lambda \xi_j \xi_k = \int_{\Pi} \left| \sum_{j \in S[-pL, pU]} e^{ij'\lambda} \xi_j \right|^2 f(\lambda) d\lambda \\ &\in \left(\int_{\Pi} \left| \sum_{j \in S[-pL, pU]} e^{ij'\lambda} \xi_j \right|^2 d\lambda \right) \times [m, M] = \left(\int_{\Pi} \sum_{j \in S[-pL, pU]} \xi_j^2 d\lambda \right) \times [m, M] \\ &= \left[(2\pi)^d m, (2\pi)^d M \right], \end{aligned}$$

using $\gamma(j-k) = \int_{\Pi} e^{i(j-k)'\lambda} f(\lambda) d\lambda$, (2.5) and the fact that $\int_{\Pi} e^{i(j-k)'\lambda} d\lambda = 0$ if $j \neq k$. \square

Proof of Corollary B.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 B.6 we get $\|\Psi_{\mathfrak{h}(p)}^{-1}\| = \mu^{-1} \leq (2\pi)^{-d} m^{-1} \leq C$. \square

Proof of Lemma B.10. 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

$$\nu(p) = \left(1, e^{ij_1'\lambda}, \dots, e^{ij_{\mathfrak{h}(p)-1}'\lambda} \right)', \quad \eta(p) = \left(1, e^{ij_1'\mu}, \dots, e^{ij_{\mathfrak{h}(p)-1}'\mu} \right)'; \lambda, \mu \in \Pi.$$

For $z \in \mathbb{C}^d$ such that $|z_i| \leq 1$, $i = 1, \dots, d$, define $D_l(z)$ analogously to $D_{\mathfrak{h}(p)}(z)$, but using only the $\hat{d}_{s, \mathfrak{h}(p)}$ corresponding to the l -th 'lag-length' in our ordering, $l = 1, \dots, \mathfrak{h}(p)$. In view of (S.3.3) 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

$$\lim_{\mathfrak{h}(p) \rightarrow \infty} \mathfrak{h}(p)^{-1} \sum_{l=0}^{\mathfrak{h}(p)-1} D_l(e^{-i\lambda}) D_l(e^{-i\mu}) e^{ij_l'(\lambda+\mu)} / \sigma_l^2, \quad (\text{S.4.1})$$

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 (S.4.1) equals $\lim_{l \rightarrow \infty} |D_l(e^{i\lambda})|^2 / \sigma_l^2 = |D(e^{i\lambda})|^2 / \sigma^2 = \left((2\pi)^d f(\lambda) \right)^{-1}$, by Lemma B.9.

If $e^{ij_l(\lambda+\mu)} \neq 1$ for all j_l write $D_l(e^{-i\lambda}) D_l(e^{-i\mu}) / \sigma_l^2 = U_l$, $e^{ij_l(\lambda+\mu)} = V_l$ and $\mathcal{V}_r = \sum_{l=0}^r V_l$. Then the RHS of (S.4.1) equals

$$\begin{aligned} \lim_{\mathfrak{h}(p) \rightarrow \infty} \mathfrak{h}(p)^{-1} \left(\sum_{l=0}^{\mathfrak{h}(p)-2} (U_l - U_{l+1}) \mathcal{V}_l + U_{\mathfrak{h}(p)-1} \mathcal{V}_{\mathfrak{h}(p)-1} \right) \\ = \lim_{\mathfrak{h}(p) \rightarrow \infty} \mathfrak{h}(p)^{-1} \sum_{l=0}^{\mathfrak{h}(p)-2} (U_l - U_{l+1}) \mathcal{V}_l, \end{aligned} \quad (\text{S.4.2})$$

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 B.9, and $\mathcal{V}_{\mathfrak{h}(p)-1} = (1 - e^{i\mathfrak{h}(p)(\lambda+\mu)}) / (1 - e^{i(\lambda+\mu)}) = \mathcal{O}(1)$. Then, by Lemma B.9 it follows that the RHS of (S.4.2) equals 0. \square

Lemmas B.8, B.9 and B.10 are lattice extensions of results in Berk (1974).

S.5 Almost sure convergence of estimates

By restricting the growth of $\mathfrak{C}(p)$ relative to N further, we can strengthen the mode of convergence to almost-sure convergence.

Theorem ST.1. *Let Assumptions A, B, C and (2.12) hold, the sequence p be chosen as a function of N such that*

$$\mathfrak{C}(p) = \mathcal{O} \left(\frac{N^{\frac{v-1}{v}}}{(\log N)^{\frac{v+1}{v}} (\log \log N)^v} \right) \text{ and } \mathfrak{C}(p) < K2^m \text{ as } N \rightarrow \infty, \quad (\text{S.5.1})$$

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

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

Theorem ST.2. *Under the conditions of Theorem ST.1, $\hat{\sigma}_{\mathfrak{h}(p)}^2 \xrightarrow{\text{a.s.}} \sigma^2$, as $N \rightarrow \infty$.*

Theorem ST.3. *Let Assumptions A, B, C, (3.2) hold, and choose the sequence p as a function of N such that*

$$\mathfrak{C}(p) \mathfrak{h}(p)^{\frac{1}{2}} = \mathcal{O} \left(\frac{N^{\frac{v-1}{v}}}{(\log N)^{\frac{v+1}{v}} (\log \log N)^v} \right) \text{ and } \mathfrak{C}(p) < K2^m \text{ as } N \rightarrow \infty, \quad (\text{S.5.2})$$

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

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

Proof of Theorem ST.1. 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 B.5) by a constant times

$$\mathfrak{c}(p)^v N^{1-v} \leq C \left\{ (\log N)^{v+1} (\log \log N)^v \right\}^{-1} \leq C m^{-v}, \quad (\text{S.5.3})$$

by (S.5.1), 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 ST.2. Similar to the proof of Theorem 2.2. \square

Proof of Theorem ST.3. Similar to the proof of Theorem 3.1. \square

References

- Akaike, H. (1969). Power spectrum estimation through autoregressive model fitting. *Annals of the Institute of Statistical Mathematics* 21, 407–419.
- Berk, K. N. (1974). Consistent autoregressive spectral estimates. *The Annals of Statistics* 2, 489–502.
- Bronars, G. and W. Jansen (1987). The geographic distribution of unemployment rates in the U.S. *Journal of Econometrics* 36, 251–279.
- Kromer, R. E. (1970). *Asymptotic Properties of the Autoregressive Spectral Estimator*. Ph. D. thesis, Stanford University.
- Pace, R. K. and J. P. LeSage (2003). Likelihood dominance spatial inference. *Geographical Analysis* 35, 133–147.
- Robinson, P. M. (1978). On consistency in time series analysis. *The Annals of Statistics* 6, 215–223.
- 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.
- 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.