Econometric Theory, **39**, 2023, 950–988.
doi:10.1017/S0266466622000226

NONPARAMETRIC PREDICTION WITH SPATIAL DATA

ABHIMANYU GUPTA
University of Essex

JAVIER HIDALGO
London School of Economics

We describe a (nonparametric) prediction algorithm for spatial data, based on a canonical factorization of the spectral density function. We provide theoretical results showing that the predictor has desirable asymptotic properties. Finite sample performance is assessed in a Monte Carlo study that also compares our algorithm to a rival nonparametric method based on the infinite *AR* representation of the dynamics of the data. Finally, we apply our methodology to predict house prices in Los Angeles.

1. INTRODUCTION

Random models for spatial or spatio-temporal data play an important role in many disciplines of economics, such as environmental, urban, development or agricultural economics as well as economic geography, among others. When data is collected over time such models are termed 'noncausal' and have drawn interest in economics, see for instance Breidt, Davis, and Trindade (2001) among others for some early examples. Other studies may be found in the special volume by Baltagi, Kelejian, and Prucha (2007) or Cressie (1993). Classic treatments include the work by Mercer and Hall (1911) on wheat crop yield data (see also Gao, Lu, and Tjøstheim (2006)) or Batchelor and Reed (1918) which was employed as an example and analysed in the celebrated paper by Whittle (1954). Other illustrations are given in Cressie and Huang (1999), see also Fernandez-Casal, Gonzalez-Manteiga, and Febrero-Bande (2003). With a view towards applications in environmental and agricultural economics, Mitchell, Genton, and Gumpertz (2005) employed a model of the type studied in this paper to analyse the effect of carbon dioxide on crops, whereas Genton and Koul (2008) examine the yield of barley in UK. The latter manuscripts shed light on how these models can be useful when there is evidence of spatial movement, such as that of pollutants, due to winds or ocean currents.

Research of the first author was supported by the Economic and Social Research Council (ESRC) grant ES/R006032/1. Research of the second author was supported by STICERD, LSE. Address correspondence to Abhimanyu Gupta, Department of Economics, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK; e-mail: a.gupta@essex.co.uk.

Doubtless one of the main aims when analysing data is to provide predicted values of realizations of the process. More specifically, assume that we have a realization $\mathcal{X}_n = \left\{x_{t_i}\right\}_{i=1}^n$ at locations $t_1, \ldots, t_n$ of a process $\{x_t\}_{t \in \mathcal{D}}$, where $\mathcal{D}$ is a subset of $\mathbb{R}^d$. We wish then to predict the value of x_t at some unobserved location t_0, say x_{t_0}. For instance in a time series context, we wish to predict the value x_{n+1} at the unobserved location (future time) $n+1$ given a stretch of data $x_1, .., x_n$. It is often the case that the predictor of x_{t_0} is based on a weighted average of the data $\mathcal{X}_n$, that is

$$\widehat{x}_{t_0} = \sum_{i=1}^{n} \beta_i x_{t_i}, \tag{1.1}$$

where the weights $\beta_1, \ldots, \beta_n$ are chosen to minimize the $\mathcal{L}_2$-risk function

$$E\left(x_{t_0} - \sum_{i=1}^{n} b_i x_{t_i}\right)^2$$

with respect to $b_1, \ldots, b_n$. With spatial data, the solution in (1.1) is referred as the Kriging predictor, see Stein (1999), which is also the best linear predictor for x_{t_0}. Notice that under Gaussianity or our Condition $C1$ below, the best linear predictor is also the best predictor. It is important to bear in mind that with spatial data prediction is also associated with both interpolation as well as extrapolation.

The optimal weights $\{\beta_i\}_{i=1}^n$ in (1.1) depend on the covariogram (or variogram) structure of $\left\{x_{t_1}, \ldots, x_{t_n}; x_{t_0}\right\} =: \left\{\mathcal{X}_n; x_{t_0}\right\}$, see among others Stein (1999) or Cressie (1993). That is, denoting the covariogram by $Cov\left(x_{t_i}, x_{t_j}\right) =: C\left(t_i, t_j\right)$ and assuming stationarity, so that $C\left(t_i, t_j\right) =: C\left(t_i - t_j\right)$, we have that the best linear predictor (1.1) becomes

$$\widehat{x}_{t_0} = \gamma'(t_0)\, \boldsymbol{C}^{-1} \mathcal{X}_n, \tag{1.2}$$

where we have assumed without loss of generality that $Ex_t = 0$ and

$$\boldsymbol{C} = \left\{C\left(t_i - t_j\right)\right\}_{i,j=1}^{n}; \quad \gamma'(t_0) = Cov\left(\mathcal{X}_n; x_{t_0}\right) = E\left\{x_{t_0}\left(x_{t_1}, \ldots, x_{t_n}\right)\right\}.$$

When the data is regularly observed, that is on a lattice, the unknown covariogram function $C(h)$ is replaced by its sample analogue

$$\widehat{C}(h) = \frac{1}{|n(h)|} \sum_{n(h)} x_{t_i} x_{t_j},$$

where $n(h) = \left\{\left(t_i, t_j\right) : t_i - t_j = h\right\}$ and $|n(h)|$ denotes the cardinality of the set $n(h)$. When the data is not regularly spaced some modifications of $\widehat{C}(h)$ have been suggested, see Cressie (1993, p.70) for details. One problem with the above estimator $\widehat{C}(h)$ is that it can only be employed for lags h which are found in the data, and hence the Kriging predictor (1.2) cannot be computed if $t_i - t_0 \neq h$ for any h such that $n(h)$ is not an empty set. To avoid this problem a typical solution is to

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

assume some specific parametric function $C(h) =: C(h;\theta)$, so that one computes (1.2) with $C\left(h;\widehat{\theta}\right)$ replacing $C(h)$ therein, where $\widehat{\theta}$ is some estimator of θ.

In this paper, we shall consider the situation when the spatial data is collected on a lattice. This may occur as a consequence of some planned experiment or due to a systematic sampling scheme, or when we can regard the (possibly non-gridded) observations as the result of aggregation over a set of covering regions rather than values at a particular site, see, e.g., Conley (1999), Conley and Molinari (2007), Bester, Conley, and Hansen (2011), Wang, Iglesias, and Wooldridge (2013), Nychka et al. (2015), Bester et al. (2016). As a result of this ability to map locations to a regular grid, lattice data are frequently studied in the econometrics literature, see, e.g., Roknossadati and Zarepour (2010), Robinson (2011) and Jenish (2016). Nonsystematic patterns may occur, although these might arise as a consequence of missing observations, see Jenish and Prucha (2012) for a study that covers irregular spatial data.

However contrary to the solution given in (1.2), our aim is to provide an estimator of (1.1) without assuming any particular parameterization of the dynamic or covariogram structure of the data a priori, for instance without assuming any particular functional form for the covariogram $C(h)$. The latter might be of interest as we avoid the risk that misspecification might induce on the predictor. In this sense, this paper may be seen as a spatial analog of contributions in a standard time series context such as Bhansali (1974) and Hidalgo and Yajima (2002).

The remainder of the paper is organized as follows. In the next section, we describe the multilateral and unilateral representation of the data and their links with a Wold-type decomposition. We also describe the canonical factorization of the spectral density function, which plays an important role in our prediction methodology described in Section 3, wherein we examine its statistical properties. Section 4 describes a small Monte-Carlo experiment to gain some information regarding the finite sample properties of the algorithm, and compares our frequency domain predictor to a potential 'space-domain' competitor. Because land value and real-estate prices comprise classical applications of spatial methods, see, e.g., Iversen Jr. (2001), Banerjee et al. (2004), Majumdar et al. (2006), we apply the procedures to prediction of house prices in Los Angeles in Section 5. Finally, Section 6 gives a summary of the paper, whereas the proofs are confined to the mathematical appendix.

2. MULTILATERAL AND UNILATERAL REPRESENTATIONS

Before we describe how to predict the value of the process $\{x_t\}_{t\in\mathbb{Z}^d}$ at unobserved locations, for $d \geq 1$, it is worth discussing what do we understand by multilateral and unilateral representations of the process and, more importantly, the link with the Wold-type decomposition. Recall that in the prediction theory of stationary time series, i.e., when $d = 1$, the Wold decomposition plays a key role. For that purpose, and using the notation that for any $a \in \mathbb{Z}^d$, $a = (a[1],\ldots,a[d])$, so that $t-j$ stands for $(t[1]-j[1],\ldots.,t[d]-j[d])$, we shall assume that the (lattice) process

$\{x_t\}_{t\in\mathbb{Z}^d}$ admits a representation given by

$$x_t = \sum_{j\in\mathbb{Z}^d} \psi_j \varepsilon_{t-j}, \quad \sum_{j\in\mathbb{Z}^d} \left\{ \sum_{\ell=1}^{d} j^2[\ell] \right\} |\psi_j| < \infty, \tag{2.1}$$

where the ε_t are independent and identically distributed random variables with zero mean, unit variance and finite fourth moments. The model in (2.1) denotes the dynamics of x_t and it is known as the multilateral representation of $\{x_t\}_{t\in\mathbb{Z}^d}$. It is worth pointing out that a consequence of the latter representation is that the sequence $\{\varepsilon_t\}_{t\in\mathbb{Z}^d}$ loses its interpretation as being the "prediction" error of the model, and thus they can no longer be regarded as innovations, as was first noticed by Whittle (1954). When $d = 1$, this multilateral representation gives rise to so-called noncausal models or, in Whittle's terminology, linear *transect* models. These models can be regarded as forward looking and have gained some consideration in economics, see for instance Lanne and Saikkonen (2011), Davis, Klüppelberg, and Steinkohl (2013), Lanne and Saikkonen (2013) or Cavaliere, Nielsen, and Rahbek (2020).

It is worth remarking that, contrary to $d = 1$, it is not sufficient for the coefficients ψ_j in (2.1) to be $O\left(|j|^{-3-\eta}\right)$ for any $\eta > 0$ as our next example illustrates. Indeed for instance take $d = 2$, an often encountered setting with real data, and suppose that $\psi_j = (j[1]+j[2])^{-4} = O\left(\|j\|^{-4}\right)$. However it is known that the sequence $\left\{\sum_{\ell=1}^{2} j^2[\ell]\right\}|\psi_j|$ is not summable. That is, see for instance Limaye and Zeltser (2009),

$$c_N^{-1} = \left(\sum_{j[1],j[2]=1}^{N} \left\{ \sum_{\ell=1}^{2} j^2[\ell] \right\} |\psi_j| \right)^{-1} \underset{N\to\infty}{\to} 0. \tag{2.2}$$

One classical parameterization of (2.1) is the *ARMA* field model

$$\begin{aligned} &P(L)\,x_t = Q(L)\,\varepsilon_t, \\ &\quad P(z) = \sum_{j\in\mathbb{Z}_1^d} \alpha_j z^j; \quad \alpha_0 = 1;\ Q(z) = \sum_{j\in\mathbb{Z}_2^d} \beta_j z^j; \quad \beta_0 = 1, \end{aligned}$$

where $\mathbb{Z}_1^d$ and $\mathbb{Z}_2^d$ are finite subsets of $\mathbb{Z}^d$ and henceforth $z^j = \prod_{\ell=1}^{d} z[\ell]^{j[\ell]}$ with the convention that $0^0 = 1$. As an example, we have the $ARMA\,(-k_1,k_2;-\ell_1,\ell_2)$ field

$$\sum_{j=-k_1}^{k_2} \alpha_j x_{t-j} = \sum_{j=-\ell_1}^{\ell_2} \beta_j \varepsilon_{t-j}, \quad \alpha_0 = \beta_0 = 1. \tag{2.3}$$

As mentioned above, the Wold decomposition, and hence the concept of past and future, plays a key role in the theory of prediction when $d = 1$. However, contrary to the situation when $d = 1$, an intrinsic problem with spatial or lattice data is that we cannot assign a unique meaning to the concept of "past" and/or "future".

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

One immediate consequence is then that different definitions of what might be considered as past (or future) will yield different Wold-type decompositions. More specifically, denote a "half-plane" of $\mathbb{Z}^2$ according to the lexicographical (dictionary) ordering "$\prec$" defined as

$$j \prec k \Leftrightarrow (j[1] < k[1]) \text{ or } (j[1] = k[1] \vee j[2] < k[2]), \tag{2.4}$$

where herewith we shall consider the case when $d = 2$, often encountered with real data. The half-plane defined by "$\prec$" is illustrated in Figure 1. Following earlier work by Helson and Lowdenslager (1958, 1961), there exists then a Wold-type representation of the (spatial) process $\{x_t\}_{t\in\mathbb{Z}^2}$ given by

$$x_t = \vartheta_t + \sum_{0\prec j} \zeta_j \vartheta_{t-j}, \quad \sum_{0\prec j} |\zeta_j| < \infty, \tag{2.5}$$

where $\{\vartheta_t\}_{t\in\mathbb{Z}^2}$ is a zero mean white noise sequence with finite second moments σ_ϑ^2. It is worth recalling that ϑ_t once again has the interpretation of being the "one-step" prediction error. Often (2.5) is called a unilateral representation of $\{x_t\}_{t\in\mathbb{Z}^2}$ as opposed to the multilateral representation in (2.1). See also Whittle (1954) for some earlier work on multilateral versus unilateral representations. As an example, (2.3) becomes a unilateral or causal model when $\ell_1 = k_1 = 0$. (2.5) might be regarded as a particular way to model the dependence of x_t induced by the lexicographic ordering in (2.4). Of course, the choice of the "half-plane" of $\mathbb{Z}^2$ according to the associated chosen lexicographic ordering is not the only possible one. That is, a different choice of "half-plane" of $\mathbb{Z}^2$, induced by the lexicographic ordering, will yield a "similar" but different representation of x_t to that given in (2.5). As will become clear in the next section, the choice of a specific lexicographic ordering, or its associated half-plane, will depend very much on practical purposes. For instance, the choice of (2.4) will depend on the location where we wish to predict x_t. Last but not least it is worth, and important, mentioning that the sequences $\{\varepsilon_t\}_{t\in\mathbb{Z}^2}$ and $\{\vartheta_t\}_{t\in\mathbb{Z}^2}$ are not the same. Recall that a similar phenomenon occurs when $d = 1$ and the practitioner allows for noncausal/bilateral representations of the sequence x_t. When this is the case, the "bilateral or noncausal" representation has errors which are independent and identically distributed, whereas for its "unilateral or causal"representation, the corresponding errors are only a white noise sequence.

It is clear from the introduction that to provide accurate and valid (linear) predictions (or interpolations), a key component is to obtain the covariogram function of the sequence $\{x_t\}_{t\in\mathbb{Z}^2}$, that is $C(h) = Cov(x_t, x_{t+h})$, which is related to the spectral density function $f(\lambda)$ via the expression

$$C(h) = \int_{\Pi^2} f(\lambda) e^{-ih\cdot\lambda} d\lambda, \quad h \in \mathbb{Z}^2,$$

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

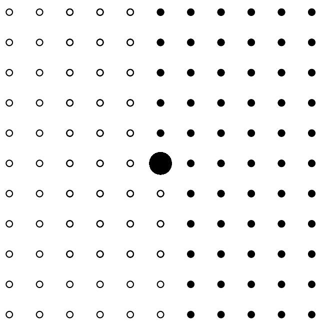


FIGURE 1. Half-plane illustration for $d = 2$. Circles form the half plane "$\prec 0$" while solid dots form the half plane "$0 \prec$". The large black solid dot marks the origin.

where $\Pi = (-\pi, \pi]$. Henceforth the notation "$h \cdot \lambda$" means the inner product of the vectors h and λ. It is worth observing that we can factorize $f(\lambda)$ as

$$f(\lambda) = \frac{\sigma_\varepsilon^2}{(2\pi)^2} |\Psi(\lambda)|^2 =: \frac{\sigma_\vartheta^2}{(2\pi)^2} |\Upsilon(\lambda)|^2, \quad \lambda \in \Pi^2,$$

where $\sigma_\varepsilon^2 = E\varepsilon_t^2$ and $\sigma_\vartheta^2 = E\vartheta_t^2$, and

$$\Psi(\lambda) = \sum_{j \in \mathbb{Z}^2} \psi_j e^{-ij\cdot\lambda}; \qquad \Upsilon(\lambda) = 1 + \sum_{0 \prec j} \zeta_j e^{-ij\cdot\lambda}.$$

The latter displayed expressions indicate that either $\Psi(\lambda)$ or $\Upsilon(\lambda)$ summarize the covariogram structure of $\{x_t\}_{t \in \mathbb{Z}^2}$.

When $d = 1$ and the sequence $\{x_t\}_{t \in \mathbb{Z}}$ is purely nondeterministic we know, see Whittle (1961, *p*.26) or Brillinger (1981, Theorem 3.8.4), that the spectral density $f(\lambda)$ admits a representation

$$f(\lambda) =: \exp(\alpha_0) |A(\lambda)|^{-2} = \exp\left\{\alpha_0 + 2\sum_{k=1}^{\infty} \alpha_k \cos(k\lambda)\right\},$$

where by definition $A(\lambda) =: \exp\left\{-\sum_{k=1}^{\infty} \alpha_k e^{ik\cdot\lambda}\right\}$. The latter expression is referred to as the canonical factorization of the spectral density function and is also known as Bloomfield's model. One important consequence of the canonical factorization is that the sequence $\{x_t\}_{t \in \mathbb{Z}}$ can be written as

$$x_t + \sum_{j=1}^{\infty} a_j x_{t-j} = \vartheta_t,$$

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

where ϑ_t is a zero mean white noise sequence with finite second moments and a_j are the Fourier coefficients of $A(\lambda)$, that is

$$a_j = \frac{1}{2\pi}\int_{-\pi}^{\pi} A(\lambda)\, e^{ij\cdot\lambda} d\lambda; \quad 0 < j,$$

with $2\pi \exp(\alpha_0) = \sigma_\vartheta^2$, i.e., the one-step prediction error. However, more importantly, denoting

$$B(\lambda) = A^{-1}(\lambda) = \exp\left\{\sum_{k=1}^{\infty} \alpha_k e^{ik\cdot\lambda}\right\},$$

we have that its Fourier coefficients equal the coefficients ζ_j in (2.5).

Whittle (1954, Section 6), Section 6, signalled that a similar argument can be used when $d > 1$. However a formal and theoretical justification for a canonical factorization of $f(\lambda)$ when $d > 1$ was discussed in Korezlioglu and Loubaton (1986), see also Solo (1986). More specifically, they show that the spectral density function of $\{x_t\}_{t\in\mathbb{Z}^2}$ might be characterized using the representation

$$f(\lambda) =: \exp(\alpha_0)\, |A(\lambda)|^{-2} = \exp\left\{\alpha_0 + 2\sum_{0\prec k} \alpha_k \cos(k\cdot\lambda)\right\}, \tag{2.6}$$

where

$$A(\lambda) =: \exp\left\{-\sum_{0\prec k} \alpha_k e^{ik\cdot\lambda}\right\}, \tag{2.7}$$

which is sometimes known as the *Cepstrum* model by Solo (1986), who notes that if $0 < f(\lambda) < M$ then the representation of the spectral density in (2.6) or in (2.7) exists, see also McElroy and Holan (2014). Note that the coefficients α_k in (2.7) are the Fourier coefficients of $\log(f(\lambda))$, that is

$$\alpha_k = \frac{1}{2\pi^2}\int_{\widetilde{\Pi}^2} \log(f(\lambda)) \cos(k\cdot\lambda)\, d\lambda, \qquad 0 \prec k \text{ and } k = 0, \tag{2.8}$$

where $\widetilde{\Pi}^2 = [0,\pi] \times \Pi$, that is $\lambda \in \widetilde{\Pi}^2$ if $\lambda[1] \in [0,\pi]$ and $\lambda[2] \in \Pi$.

As it is the case when $d = 1$, there is a relationship between the representation in (2.5) and (2.6) or (2.7), i.e., between the coefficients ζ_j and α_k. So, it will be convenient to discuss the relationship between the representations of the sequence $\{x_t\}_{t\in\mathbb{Z}^2}$ in the "frequency" and "space" domains. The link among these coefficients turns out to play a crucial role in our prediction algorithm. For that purpose, consider the lexicographic ordering given in (2.4). Then, denoting the Fourier coefficients of $A(\lambda)$ by

$$a_j = \frac{1}{(2\pi)^2}\int_{\Pi^2} A(\lambda)\, e^{ij\cdot\lambda} d\lambda; \quad 0 \prec j, \tag{2.9}$$

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

and $a_0 = 1$, the sequence $\{x_t\}_{t\in\mathbb{Z}^2}$ has a unilateral representation given by

$$x_t + \sum_{0\prec j} a_j x_{t-j} = \vartheta_t, \tag{2.10}$$

where $\{\vartheta_t\}_{t\in\mathbb{Z}^2}$ is the sequence given in (2.5). But also we have that the coefficients ζ_j in (2.5) are the Fourier coefficients of $B(\lambda) =: A^{-1}(\lambda) = \exp\left\{\sum_{0\prec k} \alpha_k e^{-ik\cdot\lambda}\right\}$. That is,

$$\zeta_j = \frac{1}{(2\pi)^2}\int_{\Pi^2} B(\lambda)\, e^{ij\cdot\lambda} d\lambda, \quad 0\prec j; \quad \zeta_0 = 1,$$

see Section 1.2 of Korezlioglu and Loubaton (1986, Section 1.2). The latter might be considered as an extension of the canonical factorization given in Brillinger (1981) to the case $d > 1$. However, one key aspect is that there is a direct link between α_k and the coefficients of the Wold-type decomposition of its autoregressive representation, that is a_j or ζ_j, and α_k. This observation will be important for our prediction methodology in the next section.

3. PREDICTION ALGORITHM

The purpose of the section is to present and examine a prediction algorithm, extending the methodology in Bhansali (1974) or Hidalgo and Yajima (2002), to the case when $d = 2$. Similar to the aforementioned work, a key component of the methodology will be based on the canonical factorization of the spectral density in (2.6). Due to the rather unusual notation in this paper, we have decided to collate it at this stage for convenience. Given two vectors a and b, $a \geq (\leq)\, b$ means that $a[\ell] \geq (\leq)\, b[\ell]$ for all $\ell = 1,2$. Denote

$$\Pi_n^2 = \left\{\lambda_{k[\ell]} = \frac{2\pi k[\ell]}{n[\ell]},\ \ k[\ell] = 0, \pm 1, \ldots, \pm \tilde{n}[\ell] =: \frac{n[\ell]}{2},\ \ \ell = 1,2\right\},$$

where $\lambda_k = \left(\lambda_{k[1]}, \lambda_{k[2]}\right)$ are the Fourier frequencies and $\widetilde{\Pi}_n^2 = \left\{\lambda_k \in \Pi_n^2 : \lambda_{k[1]} > 0\right\}$. Finally, we denote

$$\int_{\lambda\preceq\pi}^{+} = \int_{\lambda[1]=0}^{\pi}\int_{\lambda[2]=-\pi}^{\pi}; \quad \int_{\lambda\preceq\pi}^{-} = \int_{\lambda[1]=-\pi}^{0}\int_{\lambda[2]=-\pi}^{\pi}; \tag{3.1}$$

$$\int_{a\leq\lambda\leq b} = \int_{\lambda[1]=a[1]}^{b[1]}\int_{\lambda[2]=a[2]}^{b[2]}.$$

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

Similarly, we denote

$$\sum_{j\preceq J}{}^{+}c_j=\sum_{j[2]=1}^{J[2]}c_{0,j[2]}+\sum_{j[1]=1}^{J[1]}\sum_{j[2]=1-J[2]}^{J[2]}c_{j[1],j[2]};$$
$$\sum_{j\preceq J}{}^{-}c_j=\sum_{j[2]=1-J[2]}^{0}c_{0,j[2]}+\sum_{j[1]=1-J[1]}^{0}\sum_{j[2]=1-J[2]}^{J[2]}c_{j[1],j[2]}; \tag{3.2}$$
$$\sum_{a\leq t\leq b}=\sum_{t[1]=a[1]}^{b[1]}\sum_{t[2]=a[2]}^{b[2]},$$

where we are using the convention that for any $k\in\mathbb{Z}^2$, we write d_k as

$d_k=d_{k[1],k[2]}$.

Observe that $\sum_{j\preceq J}^{+}+\sum_{j\preceq J}^{-}=\sum_{-J<j\preceq J}$, and likewise $\int_{\lambda\preceq\pi}^{+}+\int_{\lambda\preceq\pi}^{-}=\int_{\lambda\in\Pi^2}$.

Before we describe our prediction algorithm, we shall introduce our set of regularity conditions.

Condition C1 (**a**) $\{\vartheta_t\}_{t\in\mathbb{Z}^2}$ in (2.5) is a zero mean white noise sequence of random variables with variance σ_ϑ^2 and finite 4*th* moments, with $\kappa_{4,\vartheta}$ denoting the fourth cumulant of ϑ_t.

(**b**) The unilateral *Moving Average* representation of $\{x_t\}_{t\in\mathbb{Z}^2}$ in (2.5) can be written (or it has a representation) as a unilateral *Autoregressive* model

$$x_t+\sum_{0\prec j}a_jx_{t-j}=\vartheta_t. \tag{3.3}$$

(**c**) The coefficients in ζ_j in (2.5) satisfy

$$\sum_{0\prec j}\left\{\sum_{\ell=1}^{2}j^4[\ell]\right\}\left|\zeta_j\right|<\infty.$$

Condition C2 $n=(n[1],n[2])$ satisfies that $n[1]\asymp n[2]$ where "$a\asymp b$" means that $K^{-1}\leq a/b\leq K$ for some finite positive constant K.

We now comment on Conditions $C1$ and $C2$. First, Condition $C2$ can be generalized to allow for different rates of convergence to zero of $n^{-1}[\ell]$, $\ell=1,2$. However, for notational simplicity, we prefer to keep it as it stands. Condition $C1$ could have been written in terms of the multilateral representation in (2.1). However since the prediction employs the representation in (3.3) or (2.5), we have opted to write $C1$ as it stands. Part (a) of Condition $C1$ seems to be a minimal condition for our results below to hold true. Sufficient regularity conditions required for the validity of the expansion in (3.3) is $\Upsilon(z)$ be nonzero for any $z[\ell]$, $\ell=1,2$. The latter condition guarantees that $f(\lambda)>0$ for all $\lambda\in\widetilde{\Pi}^2$. Part (**c**) entails that the spectral density $f(\lambda)$ is 4 times continuously differentiable. This is needed if one wants to achieve a similar rate of approximation of sums by their integrals when $d=1$ and the function is twice continuously differentiable. Indeed whereas when $d=1$, we have that

$$\frac{1}{n}\sum_{i=1}^{n} g\left(\frac{i}{n}\right) - \int_0^1 g(x) = \frac{1}{n}\left(g(0)-g(1)\right) + O\left(n^{-2}\right),$$

with two continuous derivatives for $g(x)$, to have a "similar" result when $d=2$ one needs $g(x)$ to be 4 times continuously differentiable. See Lemma 6 in the appendix for some extra insight.

We now discuss the methodology to predict the value of x_t at an unobserved location without imposing any specific parametric model for $f(\lambda)$. In addition, as a by-product, we provide a simple estimator of the coefficients ζ_j or a_j. First, $A(\lambda)$ and expression (2.8) suggest that to compute an estimator of the coefficients α_j and/or a_j, it suffices to obtain an estimator of $f(\lambda)$. To that end, for a generic sequence $\{v_t\}_{t=1}^{n}$, we shall define the *discrete Fourier transform*, *DFT*, as

$$\begin{aligned} w_v(\lambda) &= \frac{1}{(n[1]n[2])^{1/2}} \sum_{1\le t\le n} v_t e^{-it\cdot\lambda} \\ &=: \frac{1}{\mathbf{n}^{1/2}} \sum_{1\le t\le n} v_t e^{-it\cdot\lambda}, \end{aligned}$$

using, in what follows, the notation that for any $g=(g[1],g[2])$,

$$\mathbf{g} = g[1]g[2]. \tag{3.4}$$

Also we define the *periodogram* as

$$I_v(\lambda) = \frac{1}{(2\pi)^2}\left|w_v(\lambda)\right|^2; \qquad \lambda \in \widetilde{\Pi}^2.$$

In real applications, in order to make use of the *fast Fourier transform*, the *periodogram* will be evaluated at the Fourier frequencies λ_k.

However as noted by Guyon (1982), due to non-negligible end effects (the edge effect), the bias of the periodogram does not converge to zero fast enough when $d>1$. We therefore proceed as in Dahlhaus and Künsch (1987), and employ the *tapered periodogram* defined as

$$I_v^T(\lambda_j) = \frac{1}{(2\pi)^2}\left|w_v^T(\lambda_j)\right|^2; \quad w_v^T(\lambda_j) = \frac{1}{\left(\sum_{1\le t\le n} h_t^2\right)^{1/2}} \sum_{1\le t\le n} h_t v_t e^{it\cdot\lambda_j}, \tag{3.5}$$

where $w_v^T(\lambda_j)$ denotes the taper discrete Fourier transform, *DFT*. One common taper is the *cosine-bell (or Hanning)* function, which is defined as

$$h_t = \frac{1}{4} h_{t[1]} h_{t[2]}; \quad h_{t[\ell]} = \left(1-\cos\left(\frac{2\pi t[\ell]}{n[\ell]}\right)\right), \tag{3.6}$$

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

see Brillinger (1981). It is worth observing the *cosine-bell* taper *DFT* is related to $w_v(\lambda)$ by the equality

$$w_v^T(\lambda_j) = \frac{1}{6}\prod_{\ell=1}^{2}\left[-w_v\left(\lambda_{j[\ell]-1}\right) + 2w_v\left(\lambda_{j[\ell]}\right) - w_v\left(\lambda_{j[\ell]+1}\right)\right]. \tag{3.7}$$

In this paper we shall explicitly consider the *cosine-bell*, although the same results follow employing other taper functions such as *Parzen or Kolmogorov* tapers (Brillinger, 1981). This is formalized in the next condition.

Condition C3 $\{h_t\}_{t=1}^n$ is the cosine-bell taper function in (3.6).

Using notation in (3.4), we shall estimate $f(\lambda)$ by the average tapered periodogram

$$\widehat{f}(\lambda) = \frac{1}{4\mathbf{m}}\sum_{-m<\ell\leq m} I_x^T(\lambda+\lambda_\ell), \tag{3.8}$$

where $m[\ell]/n[\ell] + m[\ell]^{-1} = o(1)$, for $\ell = 1,2$. Next, we denote $\widetilde{\lambda}_k = \left(\widetilde{\lambda}_{k[1]}, \widetilde{\lambda}_{k[2]}\right)'$, for $k[1] = 0,1,\ldots,M[1] =: \tilde{n}[1]/m[1]$ and $k[2] = 0, \pm 1,\ldots, \pm M[2] =: \tilde{n}[2]/m[2]$, where

$$\widetilde{\lambda}_{k[\ell]} = \frac{\pi k[\ell]}{M[\ell]};\quad \ell = 1,2.$$

Bearing in mind (3.2), denoting $\mathcal{M} = \{j : (0 \prec j) \wedge (-M < j \leq M)\}$ and abbreviating $\phi\left(\widetilde{\lambda}_k\right)$ by ϕ_k for a generic function $\phi(\lambda)$, we estimate the coefficients a_j, as

$$\widehat{a}_j = \frac{1}{4\mathbf{M}}\sum_{-M<k\leq M}\widehat{A}_k e^{ij\cdot\widetilde{\lambda}_k}, \quad j\in\mathcal{M}; \tag{3.9}$$

$$\widehat{A}_k = \overline{\widehat{A}}_{-k} = \exp\left\{-\sum_{j\preceq M}{}^{+}\widehat{\alpha}_j e^{-ij\cdot\widetilde{\lambda}_k}\right\}, \qquad k\in\mathcal{M}\cup\{0\};$$

$$\widehat{\alpha}_j = \frac{1}{2\mathbf{M}}\sum_{k\preceq M}{}^{+}\cos\left(j\cdot\widetilde{\lambda}_k\right)\log\widehat{f}_k, \qquad j\in\mathcal{M}\cup\{0\}. \tag{3.10}$$

It is also worth defining the quantities (3.9) and (3.10) when $\widehat{f}(\lambda)$ is replaced by $f(\lambda)$, that is

$$\widetilde{f}(\lambda) = \frac{1}{4\mathbf{m}}\sum_{-m<k\leq m} f(\lambda+\lambda_k).$$

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

We have

$$
\begin{aligned}
\widetilde{a}_{j,n} &= \frac{1}{4\mathbf{M}} \sum_{-M<k\le M} \widetilde{A}_{k,n} e^{ij\cdot\widetilde{\lambda}_k} && j \in \mathcal{M} \\
\widetilde{A}_{k,n} &= \overline{\widetilde{A}}_{-k,n} = \exp\left\{ -\sum_{j\le M}{}^{+} \widetilde{\alpha}_{j,n} e^{-ij\cdot\widetilde{\lambda}_k} \right\} && k \in \mathcal{M}\cup\{0\} \\
\widetilde{\alpha}_{j,n} &= \frac{1}{2\mathbf{M}} \sum_{k\le M}{}^{+} \cos\left(j\cdot\widetilde{\lambda}_k\right) \log \widetilde{f}_k && j \in \mathcal{M}\cup\{0\},
\end{aligned}
\tag{3.11}
$$

and we also denote

$$
\begin{aligned}
a_{j,n} &= \frac{1}{4\mathbf{M}} \sum_{-M<k\le M} A_{k,n} e^{ij\cdot\widetilde{\lambda}_k} && j \in \mathcal{M} \\
A_{k,n} &= \overline{A}_{-k,n} = \exp\left\{ -\sum_{j\le M}{}^{+} \alpha_{j,n} e^{-ij\cdot\widetilde{\lambda}_k} \right\} && k \in \mathcal{M}\cup\{0\} \\
\alpha_{j,n} &= \frac{1}{2\mathbf{M}} \sum_{k\le M}{}^{+} \cos\left(j\cdot\widetilde{\lambda}_k\right) \log f_k && j \in \mathcal{M}\cup\{0\}.
\end{aligned}
\tag{3.12}
$$

We shall now begin describing how we can predict a value x_t at the location $s = (s[1], s[2])$ such that $1 \le s[1] \le n[1]$ and $1 \le s[2] \le n[2]$. For instance, we wish to predict the unobserved value x_s

$$
\begin{array}{cccccccc}
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & (s[1],s[2])? & \bullet \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet
\end{array}
\quad \text{or} \quad
\begin{array}{cccccc}
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & \bullet & (s[1],s[2])? \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet
\end{array}
$$

Now, the location of s suggests that a convenient unilateral representation of x_t appears to be

$$
x_t = -\sum_{k[2]=1}^{\infty} a_{0,k[2]}\, x_{t[1],t[2]-k[2]} - \sum_{k[1]=1}^{\infty} \sum_{k[2]=-\infty}^{\infty} a_k\, x_{t-k} + \vartheta_t, \tag{3.13}
$$

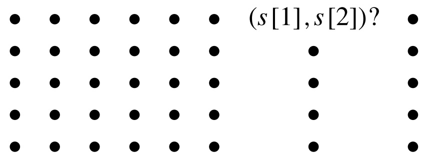
which comes from the lexicographic ordering in (2.4). Since we need to estimate the coefficients a_k, the prediction will then become

$$
\widehat{x}_{s[1],s[2]} = -\sum_{k[2]=1}^{M[2]} \widehat{a}_{0,k[2]}\, x_{s[1],s[2]-k[2]} - \sum_{k[1]=1}^{M[1]} \sum_{k[2]=1-M[2]}^{M[2]} \widehat{a}_k\, x_{s-k}, \tag{3.14}
$$

where $\widehat{a}_k\, x_{s-k} =: \widehat{a}_{k[1],k[2]}\, x_{s[1]-k[1],s[2]-k[2]}$. However, it may be very plausible that the value of M is such that we may not observe the process at some of the locations employed to compute (3.14). That is, consider the situation where we want to

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

predict x_s



Because x_{s-k} is not observed at locations for which, say, $k[1]=1$ and $k[2]<0$, it implies that to compute (3.14), we then need first to obtain a predictor for these values of x_{s-k}. It is obvious that we may predict those unobserved locations as described in expression (3.18). However, it becomes apparent that we can avoid this extra computational burden. Indeed, this is so as the relative location $(s[1],s[2])$ suggests that the practitioner might employ the Wold-type representation

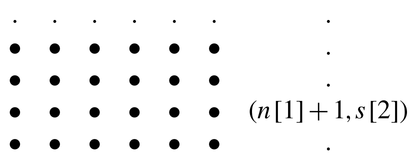
$$x_t = -\sum_{k[1]=1}^{\infty} a_{k[1],0}\, x_{t[1]-k[1],t[2]} - \sum_{k[2]=1}^{\infty}\sum_{k[1]=-\infty}^{\infty} a_k\, x_{t-k} + \vartheta_t \tag{3.15}$$

which can be regarded as induced by the lexicographic ordering

$$j \prec k \Leftrightarrow (j[2]<k[2]) \text{ or } (j[2]=k[2] \vee j[1]<k[1]). \tag{3.16}$$

Note that the lexicographic ordering (3.16) is as that in (2.4) but swapping $j[2]$ for $j[1]$. From here, we proceed as with (3.14) but with the "coordinates" [2] and [1] changing their roles.

Finally, consider the case where the location we wish to predict x_s is $(n[1]+1,s[2])$. That is,



Now, the location of $s=:(n[1]+1,s[2])$ suggests that the more convenient representation of x_s appears to be that in (3.13) which comes from the lexicographic ordering in (2.4), and hence our prediction is given in (3.14). That is, since we need to estimate the coefficients a_k, the prediction will then become

$$\widehat{x}_{n[1]+1,s[2]} = -\sum_{k[2]=1}^{M[2]} \widehat{a}_{0,k[2]}\, x_{n[1]+1,s[2]-k[2]} - \sum_{k[1]=1}^{M[1]}\sum_{k[2]=1-M[2]}^{M[2]} \widehat{a}_k\, x_{s-k}. \tag{3.17}$$

However to compute the prediction we also need to replace the unobserved x_s by its prediction. As with "standard" time series when we wish to predict beyond 1 period ahead, this is done by recursion, that is we make use of formula (3.14) starting say from the value $x_{n[1]+1,s[2]-M[2]}$. Once we have "predicted" the value for this observation, we then predict $x_{n[1]+1,s[2]-M[2]+1}$ and so on. For instance, for any

$r[1] = 0, \ldots, r$ and $r[2] = 0, \ldots, r = \min\{n[2]/8; M[2]\}$,

$$\widehat{x}_{t[1]-r[1],t[2]-r[2]} = -\sum_{k[2]=1}^{M[2]} \widehat{a}_{0,k[2]}\, \widehat{x}_{t[1]-r[1],t[2]-r[2]-k[2]} - \sum_{k[1]=1}^{M[1]} \sum_{k[2]=1-M[2]}^{M[2]} \widehat{a}_{k[1],k[2]}\, \widehat{x}_{t[1]-r[1]-k[1],t[2]-r[2]-k[2]}, \tag{3.18}$$

where we take the convention that $\widehat{x}_s = x_s$ if the location were observed and $=: 0$ when $s[2] < -r$ or $\{s[2] < 0 \wedge s[1] < n[1] - r\}$. Finally, if we were interested to predict x_t at the unobserved location $(t[1], n[2]+1)$, then it suggests to employ the lexicographic ordering in (3.16) and hence the representation given in (3.15), and then we would proceed as above but again with the "coordinates" [2] and [1] changing their roles.

Before we examine the statistical properties of $\widehat{x}_t$ in (3.14) or (3.17), we shall look at those of $\widehat{\alpha}_j$ or $\widehat{A}_j$. For that purpose, denote

$$\delta_j := 1 \text{ if } j = 0 \text{ and } := 0 \text{ otherwise} \tag{3.19}$$

$$\phi_k = \begin{cases} i\frac{1-\cos(k\pi)}{k\pi}, & \text{if } k \in \mathbb{N}^+ \\ 1, & \text{if } k = 0. \end{cases}$$

Also, denote $\{\boldsymbol{\xi}_j\}_j$ the Fourier coefficients of $g(\lambda)$ given by

$$g(\lambda) = \frac{1}{6}\left(f_{11}(\lambda) + f_{22}(\lambda)\right); \quad \lambda \in \Pi^2 \tag{3.20}$$

$$f_{\ell_1\ell_2}(\lambda) = \frac{\partial^2}{\partial\lambda_{[\ell_1]}\partial\lambda_{[\ell_2]}} f(\lambda);\ \ell_1, \ell_2 = 1, 2.$$

Notice that Condition $C1$ implies that $g(\lambda)$ is twice continuous differentiable, so that $\{\boldsymbol{\xi}_j\}_j$ is summable.

We introduce one extra condition relating the rate of increase of $m[\ell]$ with respect of $n[\ell]$.

Condition C4 $n[\ell], m[\ell] \to \infty$, for $\ell = 1, 2$, such that

$$\frac{n^3[\ell]}{m^4[\ell]} + \frac{m[\ell]}{n[\ell]} \to 0 \quad \ell = 1, 2.$$

THEOREM 1. *Under $C1 - C4$, for any finite integer J, we have that*

(a) $\mathbf{n}^{1/2}\left(\widehat{\alpha}_j - \widetilde{\alpha}_{j,n}\right)_{j=1}^{J} \xrightarrow{d} \mathcal{N}(0, \Omega_\alpha)$,

(b) $\widetilde{\alpha}_{j,n} - \alpha_{j,n} = O\left(\mathbf{M}^{-1}\xi_j + \mathbf{M}^2\right), \quad j = 1, \ldots, J$,

where Ω_a is a diagonal matrix whose (j,j)-th element is $1 + \left(1 + \kappa_{4,\vartheta}\right)\delta_j$.

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

Remark. Because $\sigma_\vartheta^2 = 2\pi\exp(\alpha_0)$, we have that $\widehat{\sigma}_\vartheta^2 =: 2\pi\exp(\widehat{\alpha}_0)$ is a consistent estimator of σ_ϑ^2. Indeed, by standard delta methods, the proof follows using Theorem 1 and that Lemma 6 implies that $\alpha_{0,n} - \alpha_0 = O\left(\mathbf{M}^{-1/2}\right)$.

THEOREM 2. *Under $C1-C4$, for any finite integer J, we have that*

(a) $\mathbf{m}^{1/2}\left(\widehat{A}_j - \widetilde{A}_{j,n}\right)_{j=1}^{J} \xrightarrow{d} \mathcal{N}^c(0,\Omega_A)$,

(b) $\widetilde{A}_{j,n} - A_{j,n} = \dfrac{1}{\mathbf{M}} g_j A_{j,n} + o\left(\mathbf{m}^{-1/2}\right), \quad j = 1,\ldots,J,$

where $g_j = g\left(\widetilde{\lambda}_j\right)$ is given in (3.20) and $\mathcal{N}^c(0,\Omega_A)$ denotes a complex normal random variable with the (j_1,j_2)-th element of Ω_A given by

$$\Omega_{A,j_1j_2} = 2\left(\delta_{j_1[1]-j_2[1]} + 2^{-1}\phi_{j_1[1]}\phi_{j_2[1]} - i\phi_{j_1[1]-j_2[1]}\right)\delta_{j_1[2]\pm j_2[2]} A_{j_1}\overline{A}_{j_2}\,.$$

We shall now denote $a_\upsilon = 0$ if $\upsilon \prec 0$.

THEOREM 3. *Under $C1-C4$, for any finite integer J, we have that*

(a) $\mathbf{n}^{1/2}\left(\widehat{a}_j - \widetilde{a}_{j,n}\right)_{j=1}^{J} \xrightarrow{d} \mathcal{N}(0,\Omega_a)$,

(b) $\mathbf{n}^{1/2}\left(\widetilde{a}_{j,n} - a_{j,n}\right) \to 0,$

where a typical element (j_1,j_2) of Ω_a, with $j_1 \preceq j_2$, is $\sum_{0\preceq k} a_k a_{k+j_2-j_1}$.

Once we have obtained the asymptotic properties of the estimators of a_j, for $0 \prec j$ and $j \le M$, we are in a position to examine the asymptotic properties of the predictor $\widehat{x}_s$ in (3.14) or (3.17). To that end, denote by $\left\{x_t^*\right\}_{t\in\mathbb{Z}^2}$ a new independent replicate sequence with the same statistical properties of the original sequence $\{x_t\}_{t\in\mathbb{Z}^2}$ not used in the estimation of the spectral density function. Then let $\widehat{x}_s^*$ be as in (3.14) but with $\widehat{x}_t$ replaced by x_t^* there, that is

$$\widehat{x}_{s[1],s[2]} = -\sum_{k[2]=1}^{M[2]} \widehat{a}_{0,k[2]}\, x^*_{s[1],s[2]-k[2]} - \sum_{k[1]=1}^{M[1]} \sum_{k[2]=1-M[2]}^{M[2]} \widehat{a}_k x^*_{s-k},$$

or (3.17) but with $\widehat{x}_t$ being replaced by x_t^* there, that is

$$\widehat{x}^*_{t[1]-r[1],t[2]-r[2]} = -\sum_{k[2]=1}^{M[2]} \widehat{a}_{0,k[2]}\, \widehat{x}^*_{t[1]-r[1],t[2]-r[2]-k[2]} - \sum_{k[1]=1}^{M[1]} \sum_{k[2]=1-M[2]}^{M[2]} \widehat{a}_k x^*_{(s-r)-k}.$$

THEOREM 4. *Under $C1-C4$, we have that*

(a) $AE\left(\widehat{x}^*_{s[1],s[2]} - x^*_{s[1],s[2]}\right)^2 = \sigma_\vartheta^2$,

(b) $AE\left(\widehat{x}^*_{n[1]+1,t[2]} - x^*_{n[1]+1,t[2]}\right)^2 = \left(1 + \sum_{k[2]=1}^{\infty} \zeta^2_{0,k[2]}\right)\sigma_\vartheta^2$,

*where AE denotes the "*Asymptotic Expectation*".*

4. MONTE CARLO EXPERIMENT

We examine the finite-sample behavior of our algorithm in a set of Monte Carlo simulations. As in Robinson and Vidal Sanz (2006) and Robinson (2007) we used the model

$$x_t = \epsilon_t + \tau \sum_{\substack{s_1=-1 \\ s\neq 0}}^{1} \sum_{s_2=-1}^{1} \epsilon_{t-s}, \tag{4.1}$$

similar to one considered in Haining (1978). Then

$$f(\lambda) = (2\pi)^{-2}\{1+\tau\nu(\lambda)\}, \tag{4.2}$$

with $\nu(\lambda) = \prod_{j=1}^{2}\left(1+2\cos\lambda_j\right)-1$. Robinson and Vidal Sanz (2006) show that a sufficient condition for invertibility of (4.1) is

$$|\tau| < 1/8. \tag{4.3}$$

We first generated a 40×41 lattice using (4.1), with $\tau = 0.05, 0.075, 0.10$ and the ϵ_t drawn independently from three different distributions for each τ: $U(-5,5)$, $N(0,1)$ and $\chi_9^2 - 9$. The aim of this section is to examine the performance of both prediction algorithms in predicting the $20,20$-th element of this lattice. We did this by assuming a situation in which the practitioner has available data sets of various sizes, generated from (4.1). To permit a clear like-for-like comparison of improvement in performance as sample size increases, we construct the prediction coefficients using the samples generated in each replication and then use these to construct predictions for the 20,20-th element of the 40×41 lattice.

We took $n[1] = n^*+1$ and $n[2] = 2n^*+1$, for some positive integer n^*, implying $\mathbf{n} = (2n^*+1)(n^*+1)$, and generated iid ϵ_t from each of the three distributions mentioned in the previous paragraph. In each of the 1000 replications we experimented with $\tau = 0.05, 0.075, 0.10$ and $n^* = 5, 10, 20$ and 40. The choices of τ satisfy (4.3).

Given the different sample sizes in each dimension, we can experiment with more values of $m[1], m[2]$ and p_1, p_2 as n^* increases. We make the following choices:

$$m[1] = m[2] = 1; p^* = p_1 = p_2 = 1,2, \text{ when } n^* = 5,$$

$$m[1] = 1,2; m[2] = 1,2; p^* = p_1 = p_2 = 1,2,3, \text{ when } n^* = 10,$$

$$m[1] = 1,2,3; m[2] = 1,2,3,4,5; p^* = p_1 = p_2 = 1,2,3, \text{ when } n^* = 20,$$

$$m[1] = m[2] = 1,2,3,4,5; p^* = p_1 = p_2 = 1,2,3,4,5 \text{ when } n^* = 40.$$

The flexible exponential approach requires a nonparametric estimate of $f(\lambda)$. Two such estimates are available to use: the first one based on the tapered periodogram described in (3.8), which we denote $\hat{f}(\lambda)$, and the second based on the autoregressive approach in Gupta (2018). The latter also provides a rival

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

prediction methodology based on a nonparametric algorithm using AR model fitting, extending well established results for $d = 1$, see Bhansali (1978) and Lewis and Reinsel (1985). The idea is first to obtain a least squares predictor based on a truncated autoregression of order $p = \left(p_{L_1}, p_{U_1}; p_{L_2}, p_{U_2}\right)$, for non-negative integers p_{L_ℓ}, p_{U_ℓ}, $\ell = 1,2$, with the truncation allowed to diverge as $N \to \infty$. That is, we approximate the infinite unilateral representation in (2.10) by one of increasing order.

In view of the half-plane representation we can *a priori* set, say, $p_{L_2} = 0$ when considering $\preccurlyeq$. If we could observe the AR prediction coefficients a_k, say, a prediction of x_s based on $\preccurlyeq$ could be constructed as

$$\check{x}_s = \sum_{k \in S[-p_L, p_U]} a_k \check{x}_{s-k}, \tag{4.4}$$

where $S[-p_L, p_U]$ is the intersection of the set $\left\{t \in \mathbb{L} : -p_{L_\ell} \le t_\ell \le p_{U_\ell},\ \ell = 1,2\right\}$ with the prediction half-plane. This is the spatial version of one-step prediction and again we follow the convention that $\check{x}_s = x_s$ if x_s is observed. However (4.4) is not feasible and needs to be replaced by an approximate version, as described below.

Writing $p_\ell = p_{L_\ell} + p_{U_\ell}$, we assume throughout that $n[\ell] > p_\ell$ for $\ell = 1,2$, and denote $n_p = \prod_{\ell=1}^{2} (n[\ell] - p_\ell)$, $\mathfrak{h}(p) = p_{U_2} + (p_1 + 1) p_{U_2}$, i.e., the cardinality of $S[-p_L, p_U]$. Suppose that the data are observed on $\left\{(t_1, t_2) : n_{L_1} \le t_1 \le n_{U_1}, -n_{L_2} \le t_2 \le n_{U_2}\right\}$. Define a least squares predictor of order $\mathfrak{h}(p)$ by

$$\check{d}_p = arg\min_{a_k, k \in S[-p_L, p_U]} n_p^{-1} \sum_{j(p,n)}^{\prime\prime} \left(x_j - \sum_{k \in S[-p_L, p_U]} a_k x_{k-j} \right)^2, \tag{4.5}$$

where $\sum_{j(p,n)}^{\prime\prime}$ runs over $\left\{(j_1, j_2) : p_1 - n_{L_1} < j_1 \le n_{U_1} + 1, p_2 - n_{L_2} < j_2 \le n_{U_2} + 1\right\}$. We denote the elements of $\check{d}_p$ by $\check{d}_p(k)$, $k \in S[-p_L, p_U]$, and the minimum value by $\check{\sigma}_p^2$. A feasible half-plane prediction based on a fitted autoregression of order p is given by

$$\check{x}_{p,s} = \sum_{k \in S[-p_L, p_U]} \check{d}_p(k) \check{x}_{s-k}. \tag{4.6}$$

The autoregressive nonparametric spectrum estimate is defined as

$$\check{f}(\lambda) = \frac{\check{\sigma}_p^2}{(2\pi)^2 \left| 1 - \sum_{k \in S[-p_L, p_U]} \check{d}_p(k) e^{ik'\lambda} \right|^2}. \tag{4.7}$$

A predictor of x_s based on (3.14) using $\hat{f}(\lambda)$ (respectively $\check{f}(\lambda)$) is denoted $\hat{x}_s$ (respectively $\tilde{x}_s$), while a predictor based on (4.6) is denoted $\check{x}_s$ as mentioned above.

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

TABLE 1. Monte Carlo RMSE of prediction with $n^* = 5$, model (4.1).

$\epsilon_t \sim U(-5,5)$										
τ		0.05	0.075	0.10	0.05	0.075	0.10	0.05	0.075	0.10
$(m[1],m[2])$	p^*		$\hat{x}_{20,20}$			$\tilde{x}_{20,20}$			$\check{x}_{20,20}$	
(1,1)	1	0.5273	0.5252	0.5269	0.5143	0.5123	0.5144	0.4279	0.4575	0.4813
(1,1)	2				0.5141	0.5123	0.5144	0.4806	0.5051	0.5261
$\epsilon_t \sim N(0,1)$										
τ		0.05	0.075	0.10	0.05	0.075	0.10	0.05	0.075	0.10
$(m[1],m[2])$	p^*		$\hat{x}_{20,20}$			$\tilde{x}_{20,20}$			$\check{x}_{20,20}$	
(1,1)	1	1.2916	1.1360	1.1072	1.2713	1.1120	1.0810	1.0487	1.0226	0.9874
(1,1)	2				1.2712	1.1120	1.0811	1.0829	1.0589	1.0313
$\epsilon_t \sim \chi^2_9 - 9$										
τ		0.05	0.075	0.10	0.05	0.075	0.10	0.05	0.075	0.10
$(m[1],m[2])$	p^*		$\hat{x}_{20,20}$			$\tilde{x}_{20,20}$			$\check{x}_{20,20}$	
(1,1)	1	2.0651	1.9869	1.8656	2.0199	1.9435	1.8238	2.2475	2.1666	2.0835
(1,1)	2				2.0197	1.9435	1.8239	2.5720	2.5355	2.4353

Let $\vec{x}_{r,s}$ be a generic predictor of x_s in replication r, $r = 1,\dots,1000$. We report a statistic called the root mean squared error (RMSE) of prediction, defined as

$$RMSE(\vec{x}_s) = \left\{ \frac{1}{1000} \sum_{r=1}^{1000} \left(\vec{x}_{r,s} - x_s \right)^2 \right\}^{\frac{1}{2}}. \tag{4.8}$$

The results are reported in Tables 1-4. We observe an improvement in prediction performance as n^* increases, and also as the bandwidths ($(m[1],m[2])$ and p^*) increase as function of n^*. This is as expected in the theory. Nevertheless, even for rather small sample sizes the RMSE is acceptable. For example, for $\epsilon_t \sim U(-5,5)$ and $\epsilon_t \sim N(0,1)$ with $n^* = 5$, we can obtain predictions with RMSE that are not radically different from the $n^* = 10$ case, even though this change in n^* entails a sample that is nearly four times larger (231 against 66). In comparison the RMSE with the smaller sample size can be quite close to those obtained with more data in some cases, cf. $\check{x}_{20,20}$ for any error distribution.

For the smallest sample size $\check{x}_{20,20}$ can outperform $\hat{x}_{20,20}$ and $\tilde{x}_{20,20}$, but with increasing n^* the latter two clearly begin to dominate. An inspection of Tables 1-4 reveals that the use of the flexible exponential algorithm proposed in this paper together with either the tapered periodogram or the AR spectral estimator of Gupta (2018) outperforms autoregressive prediction in moderate to large sample sizes. There is little to choose from between the two best performing algorithms, and a practitioner might choose to use either one. However the AR prediction is clearly dominated by our algorithm.

Table 2. Monte Carlo RMSE of prediction with $n^* = 10$, model (4.1).

$\epsilon_t \sim U(-5,5)$										
τ		0.05	0.075	0.10	0.05	0.075	0.10	0.05	0.075	0.10
$(m[1],m[2])$	p^*		$\hat{x}_{20,20}$			$\tilde{x}_{20,20}$			$\check{x}_{20,20}$	
(1,1)	1	0.4212	0.4200	0.4158	0.3999	0.3989	0.3946	0.4250	0.4554	0.4808
(2,2)	2	0.5288	0.5308	0.5336	0.5161	0.5184	0.5217	0.4325	0.4626	0.4896
(1,2)	3	0.3859	0.3806	0.3776	0.3849	0.3792	0.3757	0.4390	0.4630	0.4894
$\epsilon_t \sim N(0,1)$										
τ		0.05	0.075	0.10	0.05	0.075	0.10	0.05	0.075	0.10
$(m[1],m[2])$	p^*		$\hat{x}_{20,20}$			$\tilde{x}_{20,20}$			$\check{x}_{20,20}$	
(1,1)	1	1.2711	1.3066	1.2479	1.2487	1.2862	1.2283	1.0517	1.0152	0.9788
(2,2)	2	1.1526	1.1248	1.0953	1.1318	1.1016	1.0700	1.0575	1.0252	0.9912
(1,2)	3	1.2689	1.2123	1.1755	1.2433	1.1868	1.1501	1.0700	1.0376	1.0006
$\epsilon_t \sim \chi^2_9 - 9$										
τ		0.05	0.075	0.10	0.05	0.075	0.10	0.05	0.075	0.10
$(m[1],m[2])$	p^*		$\hat{x}_{20,20}$			$\tilde{x}_{20,20}$			$\check{x}_{20,20}$	
(1,1)	1	0.9720	0.8856	0.8149	0.9805	0.8932	0.8217	2.0183	1.9234	1.8353
(2,2)	2	2.0694	1.9571	1.8487	2.0256	1.9150	1.8080	2.0018	1.8313	1.6798
(1,2)	3	1.3581	1.2594	1.1650	1.5024	1.4011	1.3045	2.1316	1.9801	1.8485

5. AN APPLICATION TO HOUSE PRICE PREDICTION IN LOS ANGELES

In this section we show how the techniques established in the paper can be used to predict house prices. This can be of interest in real estate and urban economics, as well as for property developers. Indeed, spatial methods are frequently used in these fields, as studied for instance by Iversen Jr. (2001), Banerjee et al. (2004) and Majumdar et al. (2006). We use median house price data for census blocks in California from the 1990 census from Pace and Barry (1997), available at www.spatial-statistics.com. We confine our analysis to the city of Los Angeles. The data is gridded as follows: a 14×23 grid of square cells is superimposed on Los Angeles, from 33.75°N to 34.17°N and 117.75°W to 118.44° W. The grid covers a total of 5259 observations. The average of the median house values for each cell is calculated and the 322 such observations form our sample. The gridding is shown in Figure 2, in which the 8 empty cells are filled and marked with a cross. We wish to predict the house price for these cells. House price data is not a zero mean process, so we subtract the sample mean using the whole sample from each cell.

We proceed in the following way: to obtain the coefficients $\hat{a}_\ell$, and $\check{d}(\ell)$ in (3.14) and (4.6) we use the 14×19 sublattice formed of the first 19 columns of

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

TABLE 3. Monte Carlo RMSE of prediction with $n^* = 20$, model (4.1).

$\epsilon_t \sim U(-1,1)$										
τ		0.05	0.075	0.10	0.05	0.075	0.10	0.05	0.075	0.10
$(m[1],m[2])$	p^*		$\hat{x}_{20,20}$			$\tilde{x}_{20,20}$			$\check{x}_{20,20}$	
(1,1)	1	0.2946	0.2865	0.2806	0.2989	0.2905	0.2843	0.4245	0.4549	0.4806
(1,2)	2	0.3917	0.3861	0.3817	0.3942	0.3881	0.3832	0.4263	0.4587	0.4867
(1,3)	2	0.4526	0.4427	0.4335	0.4448	0.4350	0.4258			
(2,3)	2	0.4273	0.4170	0.4084	0.4136	0.4036	0.3953			
(2,4)	2	0.3792	0.3763	0.3732	0.3783	0.3749	0.3713			
(3,4)	4	0.4274	0.4238	0.4216	0.4246	0.4203	0.4174	0.4326	0.4617	0.4889
(3,5)	3	0.4270	0.4238	0.4216	0.4243	0.4203	0.4173	0.4271	0.4593	0.4871
$\epsilon_t \sim N(0,1)$										
τ		0.05	0.075	0.10	0.05	0.075	0.10	0.05	0.075	0.10
$(m[1],m[2])$	p^*		$\hat{x}_{20,20}$			$\tilde{x}_{20,20}$			$\check{x}_{20,20}$	
(1,1)	1	1.1985	1.2252	1.1586	1.2006	1.2274	1.1612	1.0515	1.0146	0.9783
(1,2)	2	1.2035	0.9851	0.9384	1.1964	0.9775	0.9311	1.0532	1.0145	0.9800
(1,3)	2	1.0145	0.8611	0.8164	1.0234	0.8709	0.8262			
(2,3)	2	0.9794	0.9397	0.8998	0.9850	0.9457	0.9061			
(2,4)	2	1.2033	1.1699	1.1390	1.1771	1.1439	1.1132			
(3,4)	4	1.1636	1.1374	1.1112	1.1456	1.1187	1.0919	1.0572	1.0161	0.9786
(3,5)	3	1.1634	1.1371	1.1110	1.1454	1.1185	1.0917	1.0594	1.0170	0.9795
$\epsilon_t \sim \chi^2_9 - 9$										
τ		0.05	0.075	0.10	0.05	0.075	0.10	0.05	0.075	0.10
$(m[1],m[2])$	p^*		$\hat{x}_{20,20}$			$\tilde{x}_{20,20}$			$\check{x}_{20,20}$	
(1,1)	1	1.0012	0.5964	0.5140	1.0421	0.6479	0.5613	1.9713	1.8688	1.7806
(1,2)	2	0.6229	0.5203	0.4430	0.7138	0.6053	0.5215	1.9134	1.7618	1.5989
(1,3)	2	0.9546	0.8202	0.6878	1.0768	0.9393	0.8037			
(2,3)	2	0.9984	0.8924	0.7884	1.1088	0.9986	0.8906			
(2,4)	2	1.3377	1.2419	1.1488	1.4835	1.3859	1.2906			
(3,4)	4	1.7939	1.6942	1.5952	1.7852	1.6818	1.5790	1.9971	1.8303	1.6848
(3,5)	3	1.7913	1.6904	1.5911	1.7825	1.6780	1.5749	1.9484	1.7935	1.6429

cells. This sublattice contains no missing observations. Once the coefficients are obtained we construct predictions using the remaining 4×19 sublattice, in a step-by-step manner. The shaded-and-crossed cell (8,20) is predicted first, followed by (8,21) and (8,22). We then predict (4,21), followed by (7,22), (9,23), (6,23) and (1,23).

The predicted values are tabulated for various values of $(m[1],m[2])$ and (p_1,p_2) in Tables 5 and 6. The predicted values are quite stable across the choices

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

TABLE 4. Monte Carlo RMSE of prediction with $n^* = 40$, model (4.1).

$\epsilon_t \sim U(-5,5)$										
τ		0.05	0.075	0.10	0.05	0.075	0.10	0.05	0.075	0.10
$(m[1],m[2])$	p^*		$\hat{x}_{20,20}$			$\tilde{x}_{20,20}$			$\check{x}_{20,20}$	
(1,1)	1	0.2034	0.2049	0.2066	0.2138	0.2151	0.2165	0.4228	0.4541	0.4800
(2,2)	2	0.2880	0.2821	0.2763	0.2924	0.2862	0.2801	0.4234	0.4561	0.4844
(3,3)	3	0.3229	0.3133	0.3039	0.3100	0.3008	0.2917	0.4234	0.4560	0.4843
(4,4)	4	0.3987	0.3946	0.3907	0.3768	0.3728	0.3691	0.4243	0.4567	0.4848
(5,5)	5	0.3766	0.3712	0.3656	0.3734	0.3678	0.3619	0.4253	0.4575	0.4852
$\epsilon_t \sim N(0,1)$										
τ		0.05	0.075	0.10	0.05	0.075	0.10	0.05	0.075	0.10
$(m[1],m[2])$	p^*		$\hat{x}_{20,20}$			$\tilde{x}_{20,20}$			$\check{x}_{20,20}$	
(1,1)	1	1.0378	1.0174	0.9847	1.0430	1.0226	0.9897	1.0516	1.0145	0.9782
(2,2)	2	1.0132	0.9739	0.9349	1.0156	0.9765	0.9379	1.0512	1.0155	0.9809
(3,3)	3	1.0971	1.0582	1.0201	1.0716	1.0331	0.9954	1.0508	1.0134	0.9755
(4,4)	4	1.1024	1.0610	1.0191	1.0782	1.0377	0.9967	1.0511	1.0136	0.9758
(5,5)	5	1.0987	1.0490	0.9989	1.0538	1.0054	0.9566	1.0517	1.0141	0.9760
$\epsilon_t \sim \chi_9^2 - 9$										
τ		0.05	0.075	0.10	0.05	0.075	0.10	0.05	0.075	0.10
$(m[1],m[2])$	p^*		$\hat{x}_{20,20}$			$\tilde{x}_{20,20}$			$\check{x}_{20,20}$	
(1,1)	1	1.0228	0.9107	0.7998	1.0813	0.9668	0.8535	1.9533	1.8618	1.7730
(2,2)	2	0.5742	0.4864	0.4025	0.6323	0.5412	0.4533	1.8933	1.7385	1.5718
(3,3)	3	0.7560	0.6947	0.6335	0.8305	0.7655	0.7007	1.9050	1.7582	1.6037
(4,4)	4	0.8226	0.7490	0.6752	0.8315	0.7571	0.6825	1.9156	1.7701	1.6184
(5,5)	5	0.8275	0.7496	0.6720	0.9196	0.8378	0.7563	1.9327	1.7883	1.6380

$(m[1],m[2]) = (2,2),(2,3)$ using either the periodogram or AR spectral estimate. They most closely match those obtained when $(p_1,p_2) = (2,2),(2,3)$ in (4.6). In the latter case we compare in Table 7 the order selection criteria proposed by Gupta (2018), which include the usual FPE and BIC (denoted with a $\widehat{\ }$) as well as the corrected version that account for the spatial case (denoted with $\widetilde{\ }$ and $\bar{\ }$). The FPE tends to favor longer lag lengths no matter which version is used, as do $\widehat{BIC}$ and $\overline{BIC}$. However the latter as well as $\widetilde{FPE}$ are not monotonically decreasing in lag length, unlike $\widehat{BIC}$, $\widehat{FPE}$ and $\overline{FPE}$. Thus the latter three are likely to overfit and seem undesirable. If we impose a selection rule that picks the desirable lag order as the first instance when the selection criteria shows an increase with lag length, then we get $(p_1,p_2) = (2,2)$ using $\widetilde{FPE}$ and $\overline{BIC}$. $\widetilde{BIC}$ indicates a choice of $(p_1,p_2) = (1,2)$, on the other hand. All considered, it seems that $(p_1,p_2) = (2,2)$ is a reasonable choice.

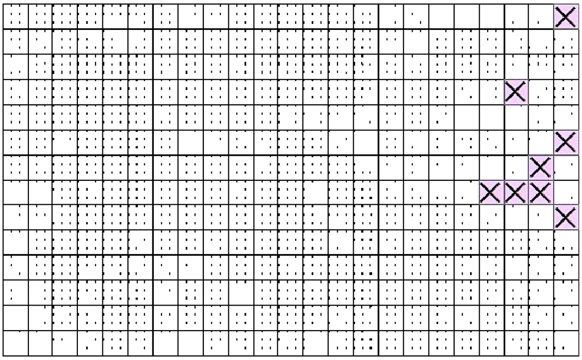


FIGURE 2. Gridded Los Angeles median house price data.

TABLE 5. Los Angeles house price predictions, in '00,000 US Dollars.

$(m[1],m[2])$		(8,20)	(8,21)	(8,22)	(4,21)	(7,22)	(9,23)	(6,23)	(1,23)
(1,1)	$\widehat{}$	0.5645	0.5540	0.5384	0.6976	0.6671	0.1962	0.9337	0.4518
	$\sim$	0.5244	0.5529	0.5523	0.6643	0.6256	0.1969	0.8770	0.4348
(1,2)	$\widehat{}$	0.8054	0.5312	0.8791	1.2740	0.8797	1.0049	1.2029	0.6338
	$\sim$	0.7043	0.4290	0.7684	1.1141	0.7693	0.8788	1.0520	0.5543
(2,1)	$\widehat{}$	1.0691	0.8761	0.3926	1.4462	1.9503	0.7717	0.8519	1.0789
	$\sim$	0.9828	0.8133	0.4722	1.3612	1.7429	0.6726	0.8448	1.0044
(2,2)	$\widehat{}$	2.0717	1.7439	1.4680	3.0421	2.4033	2.3360	2.3242	2.0935
	$\sim$	1.8934	1.4566	1.1207	2.7803	2.1965	2.1349	2.1242	1.9134
(2,3)	$\widehat{}$	1.9970	1.6205	1.3149	2.9325	2.3167	2.2518	2.2404	2.0181
	$\sim$	1.7926	1.3057	0.9511	2.6323	2.0796	2.0213	2.0111	1.8115
(1,3)	$\widehat{}$	0.8150	0.5414	0.8899	1.2892	0.8902	1.0169	1.2173	0.6414
	$\sim$	0.7218	0.4460	0.7874	1.1419	0.7885	0.9007	1.0782	0.5681

6. CONCLUSION

In this paper we have dealt with the problem of prediction when the data $\{x_t\}_{t\in\mathbb{Z}^2}$ is collected on a lattice. To do so, we considered unilateral representations of $\{x_t\}_{t\in\mathbb{Z}^2}$ and in particular the canonical factorization of the spectral density function, the latter being possible as observed by Whittle (1954). Our approach does not need any parameterization of the model (i.e., the covariogram structure of the data), so we avoid the consequences that a wrong parameterization can have in the predictor. We have also compared our methodology to one based on the space domain by using a finite approximation of the unilateral autoregressive model in (3.3).

However, it might be interesting to examine how our proposed methodology compares with one based on the conditional autoregressive (*CAR*) representation

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

TABLE 6. Los Angeles house price predictions using (4.6), in '00,000 US Dollars.

(p_1,p_2)	(8,20)	(8,21)	(8,22)	(4,21)	(7,22)	(9,23)	(6,23)	(1,23)
(1,1)	1.6525	1.2876	1.0325	2.4728	1.6663	2.1204	1.6825	1.7266
(1,2)	1.5373	1.2516	0.8473	2.3741	1.7944	1.8452	1.7626	1.5659
(2,1)	1.8138	1.6433	1.1237	2.4881	1.8545	2.3404	1.7547	2.3233
(2,2)	1.8703	1.9858	1.4089	2.6079	2.4100	2.4992	2.0862	2.2252
(3,2)	1.7925	1.9298	1.4400	2.4489	2.6472	2.5962	2.4021	1.9518
(4,3)	2.4465	2.2325	1.9075	2.0319	3.1207	3.5439	2.8234	2.0841

TABLE 7. Los Angeles house price predictions using (4.6), BIC and FPE.

(p_1,p_2)	$\widehat{\text{BIC}}$	$\widetilde{\text{BIC}}$	$\overline{\text{BIC}}$	$\widehat{\text{FPE}}$	$\widetilde{\text{FPE}}$	$\overline{\text{FPE}}$
(1,1)	0.5979	0.6001	0.5990	0.5472	0.5639	0.5555
(1,2)	0.5973	0.5979	0.5976	0.5144	0.5183	0.5164
(2,1)	0.5953	0.6042	0.5997	0.4766	0.5377	0.5062
(2,2)	0.5946	0.6008	0.5977	0.4351	0.4728	0.4535
(3,2)	0.5943	0.6105	0.6024	0.4022	0.5018	0.4489
(4,3)	0.5824	0.6081	0.5953	0.2129	0.3058	0.2543

of Besag (1974). That is, let x_t be given by

$$\begin{aligned} x_t &= \mu + E[x_t \mid x_r; r \neq t] + u_t \\ &= \mu + \sum_{r\neq t} \zeta_{t-r} x_r + u_t, \qquad \zeta_{t-r} = \zeta_{r-t}. \end{aligned} \tag{6.1}$$

Note that our definition in (6.1) implies that x_t is, among other characteristics, homogeneous. The representation of x_t given in (6.1) suggests to predict a value x_t at a location $s = (s[1], s[2])$, $1 \le s[1] \le n[1]$ and $1 \le s[2] \le n[2]$, by

$$\widehat{x}_s = \widehat{\mu} + \sum_{-M<r-s<M;\, r\neq s} \widehat{\zeta}_{r-s} x_r, \tag{6.2}$$

where $\widehat{\mu}$ and $\widehat{\zeta}_{r-s}$ are respectively the least squares estimator of μ and ζ_{r-s}, and with the convention that $x_r = 0$ if it were not observed. This is in the same spirit as we did with our predictor in (3.14). On the other hand, if we were interested in predicting a value x_t at a location $s = (n[1]+1, s[2])$, we might then use

$$\widehat{x}_{n[1]+1,s[2]} = \widehat{\mu} + \sum_{-M<r-s<M;\, r\neq s} \widehat{\zeta}_{r-s} x_r. \tag{6.3}$$

However to compute the prediction we would also need to replace the unobserved x_r by its prediction as in (3.17). The latter might be done in an iterative fashion similar to what we did in (3.18).

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

MATHEMATICAL APPENDIX

A. Proofs of Theorems

To simplify the notation, we shall write $\sum_{j\preceq J}$ instead of $\sum_{j\preceq J}^{+}$ given in (3.2). That is,

$$\sum_{k\preceq M} d_k = \sum_{k[2]=1}^{M[2]} d_{0,k[2]} + \sum_{k[1]=1}^{M[1]} \sum_{k[2]=1-M[2]}^{M[2]} d_{k[1],k[2]}. \tag{A.1}$$

Also recalling our notation in (3.4), that is for any $g=(g[1],g[2])$,

$\mathbf{g} = g[1]g[2]$,

we may write, for instance, that

$$\frac{1}{M[1]M[2]}\sum_{k\preceq M} d_k =: \frac{1}{\mathbf{M}}\sum_{k\preceq M} d_k$$

$$\frac{1}{M[1]M[2]}\sum_{-M<k\preceq M} d_k =: \frac{\mathbf{1}}{\mathbf{M}}\sum_{-M<k\preceq M} d_k.$$

Finally, when needed for the sake of notational simplicity, we might take $M[1]=M[2]$ and $n[1]=n[2]$, in which case $\mathbf{n}=n^2[1]$ and $\mathbf{m}=m^2[1]$ say.

A.1. Proof of Theorem 1

We shall examine part (**a**), since part (**b**) follows by Lemma 2 and standard arguments. By the Cramér-Wold device, it suffices to show that for a finite set of constants φ_j, $j=1,\ldots,J$,

$$\mathbf{n}^{1/2}\sum_{j=1}^{J}\varphi_j\left(\widehat{\alpha}_j-\widetilde{\alpha}_{j,n}\right) \xrightarrow{d} \mathcal{N}\left(0, \sum_{j=1}^{J}\varphi_j^2\left(1+\left(1+\kappa_{4,\vartheta}\right)\delta_j\right)\right). \tag{A.2}$$

First, by definition of $\widehat{\alpha}_j$ and $\widetilde{\alpha}_{j,n}$, we have that

$$\widehat{\alpha}_j-\widetilde{\alpha}_{j,n} = \frac{1}{2\mathbf{M}}\sum_{k\preceq M}\log\left(\frac{\widehat{f}_k}{\widetilde{f}_k}\right)\cos\left(j\cdot\widetilde{\lambda}_k\right). \tag{A.3}$$

Because standard inequalities and then Lemma 3 yield that

$$\sup_{k\preceq M}\left|\frac{\widehat{f}_k-\widetilde{f}_k}{\widetilde{f}_k}\right|^2 \le \sum_{k\preceq M}\left|\frac{\widehat{f}_k-\widetilde{f}_k}{\widetilde{f}_k}\right|^2 = O_p\left(\frac{\mathbf{M}}{\mathbf{m}}\right) = o_p(1), \tag{A.4}$$

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

we obtain that the left side of (A.3) is, by Lemma 3,

$$\frac{1}{2\mathbf{M}}\sum_{k\preceq M}\frac{\widehat{f}_k-\widetilde{f}_k}{\widetilde{f}_k}\cos\left(j\cdot\widetilde{\lambda}_k\right)+O_p\left(\mathbf{m}^{-1}\right)$$

$$=\frac{1}{2\mathbf{M}}\sum_{k\preceq M}\frac{\widehat{f}_k-\widetilde{f}_k}{f_k}\cos\left(j\cdot\widetilde{\lambda}_k\right)+\frac{1}{2\mathbf{M}}\sum_{k\preceq M}\left(\frac{\widehat{f}_k-\widetilde{f}_k}{f_k}\right)\left(\frac{f_k-\widetilde{f}_k}{\widetilde{f}_k}\right)\cos\left(j\cdot\widetilde{\lambda}_k\right)$$

$$+o_p\left(\mathbf{n}^{-1/2}\right), \tag{A.5}$$

after using Taylor series expansion of $\log(z)$ around $z=1$ and Condition $C4$. Now, the absolute value of the second term on the right of the last displayed expression is bounded by

$$\frac{1}{2\mathbf{M}}\sum_{k\preceq M}\left|\frac{\widehat{f}_k-\widetilde{f}_k}{f_k}\right|\left|\frac{f_k-\widetilde{f}_k}{\widetilde{f}_k}\right|=O\left(\frac{1}{\mathbf{M}\mathbf{m}^{1/2}}\right)=o\left(\mathbf{n}^{-1/2}\right) \tag{A.6}$$

by Lemmas 2 and 3. So, we conclude that

$$\mathbf{n}^{1/2}\left(\widehat{\alpha}_j-\widetilde{\alpha}_{j,n}\right)=\frac{\mathbf{n}^{1/2}}{2\mathbf{M}}\sum_{k\preceq M}\frac{\widehat{f}_k-\widetilde{f}_k}{f_k}\cos\left(j\cdot\widetilde{\lambda}_k\right)+o_p(1)$$

$$=\frac{1}{2\mathbf{n}^{1/2}}\sum_{k\preceq n}\frac{I_x^T(\lambda_k)-f(\lambda_k)}{f(\lambda_k)}h_{k,n}(j)+o_p(1),$$

where $h_{k,n}(j)$ is a step function defined as

$$h_{k,n}(j)=f_p^{-1}f(\lambda_k)\cos\left(j\cdot\widetilde{\lambda}_k\right)$$

when $2p[\ell]-1<\frac{k[\ell]}{m[\ell]}<2p[\ell]+1$ and $1\le p[1]<M[1]$, $1-M[2]<p[2]\le M[2]$. Now, using Lemma 1, we have that for all j,

$$\sum_{k\preceq n}\left(\frac{I_x^T(\lambda_k)}{f(\lambda_k)}-\frac{(2\pi)^2 I_\vartheta^T(\lambda_k)}{\sigma_\vartheta^2}\right)h_{k,n}(j)=o_p\left(\mathbf{n}^{1/2}\right).$$

So, we conclude that the left side of (A.2) is

$$\mathbf{n}^{1/2}\sum_{j=1}^{J}\varphi_j\left(\widehat{\alpha}_j-\widetilde{\alpha}_{j,n}\right)=\sum_{j=1}^{J}\varphi_j\frac{1}{\mathbf{n}^{1/2}}\sum_{k\preceq n}\left(\frac{(2\pi)^2 I_\vartheta^T(\lambda_k)}{\sigma_\vartheta^2}-1\right)h_{k,n}(j)+o_p(1)$$

$$=\sum_{j=1}^{J}\varphi_j\frac{1}{\mathbf{n}^{1/2}}\sum_{k\preceq n}\left(\frac{(2\pi)^2 I_\vartheta^T(\lambda_k)}{\sigma_\vartheta^2}-1\right)\cos\left(j\cdot\widetilde{\lambda}_k\right)\left(1+o_p(1)\right)$$

after we observe that Condition $C1$ implies that

$$h_{k,n}(j)=\cos\left(j\cdot\widetilde{\lambda}_k\right)\left(1+\frac{1}{\mathbf{M}^{1/2}}\left(\frac{\partial f(\lambda_k)}{\partial\lambda[1]}+\frac{\partial f(\lambda_k)}{\partial\lambda[2]}\right)+O\left(\frac{1}{\mathbf{M}}\right)\right).$$

Recall that we might assume that $M[1] = M[2]$ and our notation in (3.4). From here the conclusion is standard proceeding as in the proof of Theorems 1 and 2 of Hidalgo (2009), see also Robinson and Vidal-Sanz (2006), and so it is omitted. ■

A.2. Proof of Theorem 2

Define $\widehat{d}_j = \log \widehat{A}_j$, $\widetilde{d}_{j,n} = \log \widetilde{A}_{j,n}$ and $d_{j,n} = \log A_{j,n}$. We begin with part (**b**). First by definition,

$$\widetilde{d}_{j,n} - d_{j,n} =: \sum_{k \preceq M} \left(\widetilde{\alpha}_{k,n} - \alpha_{k,n}\right) e^{-ik\cdot\widetilde{\lambda}_j}, \tag{A.7}$$

which by Taylor expansion of $\log\left(\widetilde{f}_r / f_r\right)$, $(B.1)$ in Lemma 2 and Condition $C4$, it is

$$\frac{1}{2\mathbf{M}} \sum_{k \preceq M} \sum_{r \preceq M} \left\{ \left(\frac{\widetilde{f}_r - f_r}{f_r} \right) + \frac{1}{2} \left(\frac{\widetilde{f}_r - f_r}{f_r} \right)^2 \right\} \cos\left(k \cdot \widetilde{\lambda}_r\right) e^{-ik\cdot\widetilde{\lambda}_j} + o\left(\frac{1}{\mathbf{m}^{1/2}} \right)$$

$$= \frac{1}{2\mathbf{M}} \sum_{k \preceq M} \left\{ \sum_{r \preceq M} \left\{ \left(\frac{\widetilde{f}_r - f_r}{f_r} \right) + \frac{1}{72\mathbf{M}^2} g_r^2 \right\} \cos\left(k \cdot \widetilde{\lambda}_r\right) e^{-ik\cdot\widetilde{\lambda}_j} \right\} + o\left(\frac{1}{\mathbf{m}^{1/2}} \right).$$

Now using the inequality

$$\left| \sum_{k \preceq M} e^{-ik\cdot\widetilde{\lambda}_p} \right| \leq K \widetilde{\lambda}_{p[1]}^{-1} \widetilde{\lambda}_{p[2]}^{-1}, \tag{A.8}$$

we have that

$$\left| \frac{1}{\mathbf{M}^3} \sum_{k \preceq M} \left\{ \sum_{r \preceq M} g_r^2 \cos\left(k \cdot \widetilde{\lambda}_r\right) e^{-ik\cdot\widetilde{\lambda}_j} \right\} \right| \leq \frac{K}{\mathbf{M}^3} \sum_{r \preceq M} g_r^2 \left| \sum_{k \preceq M} \cos\left(k \cdot \widetilde{\lambda}_r\right) e^{-ik\cdot\widetilde{\lambda}_j} \right|$$
$$= O\left(\frac{\log^2 \mathbf{M}}{\mathbf{M}^2} \right).$$

Thus using that $2\cos x = e^{ix} + e^{-ix}$, the right side of (A.7) becomes

$$\frac{1}{2\mathbf{M}} \sum_{k \preceq M} \left\{ \sum_{r \preceq M} \left(\frac{\widetilde{f}_r - f_r}{f_r} \right) \cos\left(k \cdot \widetilde{\lambda}_r\right) e^{-ik\cdot\widetilde{\lambda}_j} \right\} + O\left(\frac{\log^2 \mathbf{M}}{\mathbf{M}^2} \right)$$

$$= \frac{1}{2\mathbf{M}} \sum_{r \preceq M} \left\{ \left(\frac{\widetilde{f}_r - f_r}{f_r} \right) \sum_{k \preceq M} \left(\frac{e^{ik\cdot\widetilde{\lambda}_{r-j}} + e^{ik\cdot\widetilde{\lambda}_{-r-j}}}{2} \right) \right\} + O\left(\frac{\log^2 \mathbf{M}}{\mathbf{M}^2} \right)$$

$$= \frac{1}{12\mathbf{M}} \left\{ \sum_{k \preceq M} \frac{1}{\mathbf{M}} \sum_{r \preceq M} g_r \cos\left(k \cdot \widetilde{\lambda}_r\right) \right\} e^{-ik\cdot\widetilde{\lambda}_j} + O\left(\frac{1}{\mathbf{M}^2} \sum_{r \preceq M} \frac{1}{r \pm j} \right)$$

$$= \frac{1}{6\mathbf{M}} g_j + o\left(\frac{1}{\mathbf{m}^{1/2}} \right),$$

where in the second equality we have used (B.1) and (A.8) and then Condition $C4$ and for the third equality that

$$\frac{1}{\mathbf{M}}\sum_{r\preceq M} g_r \cos\left(k\cdot\widetilde{\lambda}_r\right) = \int g(\lambda)\cos(k\cdot\lambda)\,d\lambda + O(\mathbf{M}^{-1})$$
$$=: \xi_k + \varphi_k O(\mathbf{M}^{-1/2}) + O(\mathbf{M}^{-1})$$

using Lemma 6 and then that $\sum_{k\preceq M}\xi_k e^{-ik\cdot\lambda} = g(\lambda)+O(\mathbf{M}^{-1})$ since $g(\lambda)$, given in (3.20), is twice continuously differentiable so that $|\xi_k| = O(|k|^{-3})$ and $|\varphi_k| = O(|k|^{-2})$. From here we conclude the proof of part (**b**) by standard algebra.

Next, we show part (**a**). By Cramér-Wold device, it suffices to examine that for any set of finite constants $\varphi_{q_1},\ldots,\varphi_{q_2}$, the behavior of

$$\mathbf{m}^{1/2}\sum_{j=q_1}^{q_2}\varphi_j\left(\widehat{A}_j - A_{j,n}\right).$$

First, by definitions of $\widehat{A}_{j,n}$ and $\widetilde{A}_{j,n}$, we have that

$$\mathbf{m}^{1/2}\left(\widehat{d}_j - \widetilde{d}_{j,n}\right) = -\mathbf{m}^{1/2}\sum_{k\preceq M}\left(\widehat{\alpha}_k - \widetilde{\alpha}_{k,n}\right)e^{-ik\cdot\widetilde{\lambda}_j} \tag{A.9}$$
$$= -\mathbf{m}^{1/2}\sum_{k\preceq M}\frac{1}{2\mathbf{M}}\sum_{s\preceq M}\left(\frac{\widehat{f}_s - \widetilde{f}_s}{\widetilde{f}_s}\right)\cos\left(k\cdot\widetilde{\lambda}_s\right)e^{-ik\cdot\widetilde{\lambda}_j} + o_p(1)$$
$$= \sum_{k\preceq M}\frac{\mathbf{m}^{1/2}}{2\mathbf{n}}\sum_{s\preceq n}\rho_s h_{s,n}(k)\,e^{-ik\cdot\widetilde{\lambda}_j} + o_p(1)$$

proceeding as in the proof of Theorem 1, where $\rho_s = (2\pi)^2\sigma_{\vartheta}^{-2}I_{\vartheta}^{T}(\lambda_s)$ and $h_{s,n}(k)$ were defined there. So, because $n[\ell] = 2M[\ell]m[\ell]$ for $\ell = 1,2$, denoting $\psi_{s,n}(j) = (2\mathbf{M})^{-1/2}\sum_{k\preceq M}h_{s,n}(k)\,e^{-ik\cdot\widetilde{\lambda}_j}$, we conclude that

$$\mathbf{m}^{1/2}\sum_{j=1}^{J}\varphi_j\left(\widehat{d}_j - \widetilde{d}_{j,n}\right) = \frac{1}{2\mathbf{n}^{1/2}}\sum_{s\preceq n}\rho_s\sum_{j=1}^{J}\varphi_j\psi_{s,n}(j) + o_p(1) \tag{A.10}$$
$$\xrightarrow{d}\mathcal{N}\left(0,\varphi' V\varphi\right)$$

proceeding as in the proof of Theorem 1, where $\varphi' = (\varphi_1,\ldots,\varphi_J)$ and

$$V = \lim_{n\to\infty}\sum_{j_1,j_2=1}^{J}\varphi_{j_1}\varphi_{j_2}\frac{1}{\mathbf{n}}\sum_{\ell\preceq n}\psi_{\ell,n}(j_1)\,\overline{\psi_{\ell,n}}(j_2)$$
$$= \lim_{n\to\infty}\sum_{j_1,j_2=1}^{J}\varphi_{j_1}\varphi_{j_2}\frac{1}{2\mathbf{M}^2}\sum_{\ell\preceq M}\sum_{k_1,k_2\preceq M}e^{-ik_1\widetilde{\lambda}_{j_1}+ik_2\widetilde{\lambda}_{j_2}}\cos\left(k_1\cdot\widetilde{\lambda}_\ell\right)\cos\left(k_2\cdot\widetilde{\lambda}_\ell\right)$$
$$= 2^{-1}\sum_{j_1,j_2=1}^{J}\varphi_{j_1}\varphi_{j_2}\left(\delta_{j_1[1]-j_2[2]} + 2^{-1}\phi_{j_1[1]}\phi_{j_2[1]} - i\phi_{j_1[1]-j_2[1]}\right)$$

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

by Lemma 4. From here the conclusion of the theorem follows by standard delta arguments. ■

A.3. Proof of Theorem 3

We begin with part (**a**). To that end, it suffices to show that

$$\mathbf{n}^{1/2}\sum_{\upsilon=p}^{q}\varphi_{\upsilon}\left(\widehat{a}_{\upsilon}-\widetilde{a}_{\upsilon,n}\right)\overset{d}{\to}\mathcal{N}\left(0,\sum_{\upsilon_1,\upsilon_2=p}^{q}\varphi_{\upsilon_1}\varphi_{\upsilon_2}\Omega_{a,\upsilon_1,\upsilon_2}\right). \quad \textbf{(A.11)}$$

By definition of $\widehat{a}_{\upsilon}-\widetilde{a}_{\upsilon,n}$ and Taylor expansion of $\widehat{A}_j-\widetilde{A}_{j,n}$, a typical component on the left of (A.11) is

$$\frac{\mathbf{n}^{1/2}}{4\mathbf{M}}\sum_{-M<j\le M}\left(\widehat{d}_j-\widetilde{d}_{j,n}\right)\widetilde{A}_{j,n}e^{i\upsilon\cdot\widetilde{\lambda}_j}+\frac{\mathbf{n}^{1/2}}{4\mathbf{M}}\sum_{-M<j\le M}\left|\widehat{d}_j-\widetilde{d}_{j,n}\right|^2\left|\widetilde{A}_{j,n}\right|\left(1+o_p(1)\right). \quad \textbf{(A.12)}$$

Now (A.10) implies that

$$\mathbf{m}\left|\widehat{d}_j-\widetilde{d}_{j,n}\right|^2=K\frac{1}{\mathbf{n}}\left|\sum_{k\preceq n}\rho_k\psi_{k,n}(j)\right|^2+o_p(1)=O_p(1),$$

by Theorem A of Serfling (1980), p. 14 because Condition $C1$ implies that $x^4(t)$ is uniformly integrable and Theorem 2 and the continuous mapping theorem implies that $\left|\mathbf{n}^{-1/2}\sum_{k\preceq n}\rho_k\psi_{k,n}(j)\right|^2\to_d\chi^2$. So, using (A.7) we conclude that (A.12) is

$$-\frac{\mathbf{n}^{1/2}}{4\mathbf{M}}\sum_{-M<j\le M}\left(\sum_{k\preceq M}\left(\widehat{\alpha}_k-\widetilde{\alpha}_{k,n}\right)e^{-ik\cdot\widetilde{\lambda}_j}\right)\widetilde{A}_{j,n}e^{i\upsilon\cdot\widetilde{\lambda}_j}+O_p\left(\frac{\mathbf{n}^{1/2}}{\mathbf{m}}\right)$$

$$=-\frac{\mathbf{n}^{1/2}}{4\mathbf{M}^2}\sum_{-M<j\le M}\left(\sum_{k,r\preceq M}\left(\frac{\widehat{f}_r-\widetilde{f}_r}{\widetilde{f}_r}\right)\cos\left(k\cdot\widetilde{\lambda}_r\right)e^{-ik\cdot\widetilde{\lambda}_j}\right)\widetilde{A}_{j,n}e^{i\upsilon\cdot\widetilde{\lambda}_j}+O_p\left(\frac{\mathbf{n}^{1/2}}{\mathbf{m}}\right).$$

$$=-\frac{\mathbf{n}^{1/2}}{4\mathbf{M}}\sum_{r\preceq M}\left(\frac{\widehat{f}_r-\widetilde{f}_r}{\widetilde{f}_r}\right)\left(\sum_{k\preceq M}\cos\left(k\cdot\widetilde{\lambda}_r\right)a_{\upsilon-k}\right)+O_p\left(\mathbf{M}^{-1/2}\right) \quad \textbf{(A.13)}$$

$$=:\mathbf{n}^{1/2}\varrho_{n,\upsilon-k}+O_p\left(\mathbf{M}^{-1/2}\right),$$

where in the first equality we use (A.9) and in the second equality that Lemma 6 implies that

$$\frac{\mathbf{n}^{1/2}}{\mathbf{M}}\sum_{r\preceq M}\left(\frac{\widehat{f}_r-\widetilde{f}_r}{\widetilde{f}_r}\right)\sum_{k\preceq M}\cos\left(k\cdot\widetilde{\lambda}_r\right)\left\{\frac{1}{4\mathbf{M}}\sum_{-M<j\le M}\widetilde{A}_{j,n}e^{-i(k-\upsilon)\cdot\widetilde{\lambda}_j}-a_{k-\upsilon}\right\}$$

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

$$= \frac{\mathbf{n}^{1/2}}{\mathbf{M}} \sum_{r \preceq M} \left(\frac{\widehat{f}_r - \widetilde{f}_r}{\widetilde{f}_r} \right) \sum_{k \preceq M} \cos\left(k \cdot \widetilde{\lambda}_r\right) \left\{ \frac{1}{\mathbf{M}} h_{k-\upsilon} + O\left(\frac{\mathbf{1}}{\mathbf{M}^2}\right) \right\}$$

$$= O_p\left(\mathbf{M}^{-1/2}\right)$$

because $\left(\widehat{f}_r - \widetilde{f}_r\right)/\widetilde{f}_r = O_p\left(\mathbf{m}^{-1/2}\right)$ and $\{h_k\}_k$ is a summable sequence.

So, we conclude that the left side of (A.11) is

$$-\sum_{\upsilon=p}^{q} \varphi_\upsilon \mathbf{n}^{1/2} \varrho_{n,\upsilon-k} + o_p(1) \xrightarrow{d} \mathcal{N}(0,V),$$

where $V = \lim_{M\to\infty} \mathbf{M}^{-1} \sum_{r \preceq M} \left(\sum_{\upsilon=p}^{q} \varphi_\upsilon \sum_{k \preceq M} \cos(k \cdot \lambda_r) a_{k-\upsilon} \right)^2$.

We now conclude because using that $2\cos(x) = e^{ix} + e^{-ix}$, $2\cos(x)\cos(y) = \cos(x+y) + \cos(x-y)$, a typical component of V is $\varphi_{\upsilon_1}\varphi_{\upsilon_2}$ times

$$\lim_{M\to\infty} \sum_{k_1,k_2 \preceq M} a_{k_1-\upsilon_1} a_{k_2-\upsilon_2} \frac{1}{\mathbf{M}} \sum_{r \preceq M} \cos\left(k_1 \cdot \widetilde{\lambda}_r\right) \cos\left(k_2 \cdot \widetilde{\lambda}_r\right)$$

$$= \lim_{M\to\infty} \sum_{k_1,k_2 \preceq M} a_{k_1-\upsilon_1} a_{k_2-\upsilon_2} \frac{1}{2\mathbf{M}} \sum_{r \preceq M} \left(\cos\left((k_1+k_2)\cdot \widetilde{\lambda}_r\right) + \cos\left((k_1-k_2)\cdot \widetilde{\lambda}_r\right) \right)$$

$$= \lim_{M\to\infty} \sum_{k_1,k_2 \preceq M} a_{k_1-\upsilon_1} a_{k_2-\upsilon_2} \frac{1}{2M[1]} \sum_{r[1]=1}^{M[1]} \left\{ \cos\left((k_1[1]+k_2[1]) \widetilde{\lambda}_{r[1]}\right) \delta_{k_2[2]+k_1[2]} \right.$$
$$\left. + \cos\left((k_1[1]-k_2[1]) \widetilde{\lambda}_{r[1]}\right) \delta_{k_2[2]-k_1[2]} \right\}$$

$$= \sum_{0 \preceq k} a_k a_{k+\upsilon_2-\upsilon_1} - \lim_{M\to\infty} \frac{1}{M[1]} \sum_{k_1 \neq k_2 \preceq M;\, k_1 \pm k_2[1] = 1,3,\ldots,[M/2]} a_{k_1-\upsilon_1} a_{k_2-\upsilon_2},$$

where we have taken $\upsilon_1 \preceq \upsilon_2$. From here the conclusion is standard since a_υ is summable.

Part (**b**) follows by similar arguments to those in (A.13) and Lemma 2, so it is omitted. ■

A.4. Proof of Theorem 4

We begin with part (**a**). For that purpose, denote

$$\dot{x}_s^* = \sum_{k[2]=M[2]+1}^{\infty} a_{0,k[2]}\, x^*_{s[1],s[2]-k[2]} + \left(\sum_{k[1]=1}^{M[1]} \sum_{k[2]=1-M[2]}^{M[2]} \right)^{\perp} a_k x^*_{s-k}$$

$$\ddot{x}_s^* = \sum_{k[2]=1}^{M[2]} a_{0,k[2]}\, x^*_{s[1],s[2]-k[2]} + \sum_{k[1]=1}^{M[1]} \sum_{k[2]=1-M[2]}^{M[2]} a_k x^*_{s-k}$$

$$x^*_{n,s} = -\sum_{k[2]=1}^{M[2]} a_{n,(0,k[2])}\, x^*_{s[1],s[2]-k[2]} - \sum_{k[1]=1}^{M[1]} \sum_{k[2]=1-M[2]}^{M[2]} a_{n,k} x^*_{s-k},$$

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

where $\left(\sum_{k[1]=1}^{M[1]}\sum_{k[2]=1-M[2]}^{M[2]}\right)^{\perp} = \sum_{k[1]=1}^{\infty}\sum_{k[2]=-\infty}^{\infty} - \sum_{k[1]=1}^{M[1]}\sum_{k[2]=1-M[2]}^{M[2]}$. Then,

$$x_s^* - \widehat{x}_s^* = \vartheta_s - \dot{x}_s^* + \left(x_{n,s}^* - \widehat{x}_s^*\right) - \left(x_{n,s}^* + \ddot{x}_s^*\right). \tag{A.14}$$

The second moment of $\dot{x}_s^*$ is clearly $o(1)$ since $\sum_{k[2]=-\infty}^{\infty}\left|a_{k[1],k[2]}\right| < K$ for any $k[1]$ and $\sum_{k[1]=1}^{\infty}\left|a_{k[1],k[2]}\right| < K$ for any $k[2]$ and that $M \to \infty$. Next, the second moment of the last term on the right of (A.14) is bounded by

$$2E\left(\sum_{k[2]=1}^{M[2]}\left(a_{n,(0,k[2])} - a_{0,k[2]}\right) x_{s[1],s[2]-k[2]}^*\right)^2 + 2E\left(\sum_{k[1]=1}^{M[1]}\sum_{k[2]=1-M[2]}^{M[2]}\left(a_{n,k} - a_k\right) x_{s-k}^*\right)^2 = o(1),$$

by Lemma 7 and that the covariance of x_s^* is summable. Thus, it remains to examine the behavior of $x_{n,s}^* - \widehat{x}_s^*$ on the right of (A.14), which is

$$\sum_{k[2]=1}^{M[2]}\left(\widehat{a}_{0,k[2]} - \widetilde{a}_{n,(0,k[2])}\right) x_{s[1],s[2]-k[2]}^* + \sum_{k[1]=1}^{M[1]}\sum_{k[2]=1-M[2]}^{M[2]}\left(\widehat{a}_k - \widetilde{a}_{n,k}\right) x_{s-k}^* + \sum_{k[2]=1}^{M[2]}\left(\widetilde{a}_{n,(0,k[2])} - a_{n,(0,k[2])}\right) x_{s[1],s[2]-k[2]}^* + \sum_{k[1]=1}^{M[1]}\sum_{k[2]=1-M[2]}^{M[2]}\left(\widetilde{a}_{n,k} - a_{n,k}\right) x_{s-k}^*. \tag{A.15}$$

Now Theorem 3 part (**b**) and summability of the covariance of x_s^* yields that the second moment of the second term of (A.15) is $o(1)$. So, to complete the proof of part (**a**), we need to look at the first term, which is

$$\begin{aligned}
&\sum_{k[2]=1}^{M[2]}\left(\widehat{a}_{0,k[2]} - \widetilde{a}_{n,(0,k[2])} - \varrho_{n,(0,k[2])}\right) x_{s[1],s[2]-k[2]}^* \\
&+ \sum_{k[1]=1}^{M[1]}\sum_{k[2]=1-M[2]}^{M[2]}\left(\widehat{a}_k - \widetilde{a}_{n,k} - \varrho_{n,k}\right) x_{s-k}^* \\
&+ \sum_{k[2]=1}^{M[2]}\varrho_{n,(0,k[2])}\, x_{s[1],s[2]-k[2]}^* + \sum_{k[1]=1}^{M[1]}\sum_{k[2]=1-M[2]}^{M[2]}\varrho_{n,k} x_{s-k}^* \\
=& \sum_{k[2]=1}^{M[2]}\varrho_{n,(0,k[2])}\, x_{s[1],s[2]-k[2]}^* + \sum_{k[1]=1}^{M[1]}\sum_{k[2]=1-M[2]}^{M[2]}\varrho_{n,k} x_{s-k}^* + O_p\left(\mathbf{M}^{1/2}\mathbf{n}^{-1/2}\right)
\end{aligned}$$

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

because using expression (A.13), $\left|\widehat{a}_k - a_{k,n} - \varrho_k\right| = O_p\left(\mathbf{n}^{-1/2}\mathbf{M}^{-1/2}\right)$. Finally,

$$E\left|\sum_{k[1]=1}^{M[1]}\sum_{k[2]=1-M[2]}^{M[2]} \varrho_{n,k} x^*_{s-k}\right|$$

$$\leq \sum_{k[1]=1}^{M[1]}\sum_{k[2]=1-M[2]}^{M[2]} \left(E\varrho^2_{n,k}\right)^{1/2}\left(Ex^{*2}_{s-k}\right)^{1/2} = O\left(\mathbf{M}\mathbf{n}^{-1/2}\right) = o(1)$$

by triangle and Cauchy-Schwarz inequalities and then Condition $C4$. Similarly we have that $E\left|\sum_{k[2]=1}^{M[2]} \varrho_{n,0,k[2]}\; x^*_{s[1],s[2]-k[2]}\right| = o(1)$, which concludes the proof of part (**a**).

We now show part (**b**), that is when $s =: (n[1]+1, s[2])$. For that purpose, it is convenient to recall our representation in (2.5). The reason is because the prediction error $x^*_{n[1]+1,s[2]} - \widehat{x}_{n[1]+1,s[2]}$ can be written as

$$x^*_{n[1]+1,s[2]} - \widehat{x}_{n[1]+1,s[2]} = \vartheta^*_s + \sum_{0\prec k}\zeta_k\vartheta^*_{s-k} - \sum_{k[2]=1}^{M[2]}\widehat{\zeta}_{0,k[2]}\;\vartheta^*_{n[1]+1,s[2]-k[2]}$$
$$- \sum_{k[1]=1}^{M[1]}\sum_{k[2]=1-M[2]}^{M[2]}\widehat{\zeta}_k\;\vartheta^*_{s-k},$$

where $\widehat{\zeta}_{k[1],k[2]}$ is similar to $\widehat{a}_{k[1],k[2]}$ but where

$$\widehat{\zeta}_k = \frac{1}{4\mathbf{M}}\sum_{-M<\ell\leq M}\widehat{A}_\ell e^{ik\cdot\widetilde{\lambda}_\ell}, \quad k\in\mathcal{M};$$

$$\widehat{A}_\ell = \overline{\widehat{A}}_{-\ell} = \exp\left\{\sum_{j\leq M}{}^{+}\widehat{\alpha}_j e^{-ij\cdot\widetilde{\lambda}_\ell}\right\}, \qquad \ell\in\mathcal{M}\cup\{0\}$$

and $\widehat{\alpha}_j$ as defined in (3.10). Now, it is obvious that we have the same type of (statistical) results for $\widehat{\zeta}_k$ as those obtained for $\widehat{a}_k$, and hence proceeding as in part (**a**), we conclude that

$$x^*_{n[1]+1,s[2]} - \widehat{x}_{n[1]+1,s[2]} = \vartheta^*_{n[1]+1,s[2]} + \sum_{k[2]=1}^{\infty}\zeta_{0,k[2]}\;\vartheta^*_{n[1]+1,s[2]-k[2]} + o_p(1)$$

and that

$$AE\left(x^*_{n[1]+1,s[2]} - \widehat{x}_{n[1]+1,s[2]}\right) = \left(1 + \sum_{k[2]=1}^{\infty}\zeta^2_{0,k[2]}\right)\sigma^2_\vartheta.$$

This concludes the proof of the theorem.

B. Technical Lemmas

To simplify the notation, we abbreviate $\sum_{-m<j\leq m}$ by $\sum_j$ in what follows and recall our notation in (3.4).

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

LEMMA 1. *Under Conditions $C1-C4$ we have that*

$$\widehat{f}_k = \frac{1}{4\mathbf{m}} \sum_j f\left(\lambda_j + \widetilde{\lambda}_k\right) \frac{I_\vartheta^T\left(\lambda_j + \widetilde{\lambda}_k\right)}{\sigma_\vartheta^2} + \epsilon_{n,k},$$

where $\left\{\epsilon_{k,n}\right\}_k$ is a triangular array sequence of r.v.'s such that $E\sup_k \left|\epsilon_{k,n}\right|^2 = o\left(\mathbf{m}^{-1}\right)$.

Proof. The proof follows easily from Lemma 4 of Hidalgo (2009), and so it is omitted. □

LEMMA 2. *Assuming $C1-C4$, $(\mathbf{a})\,\widetilde{\alpha}_{k,n} - \alpha_{k,n} = \mathbf{M}^{-1}\xi_k + O\left(\mathbf{M}^{-2}\right)$ and $(\mathbf{b})\,\alpha_{k,n} - \alpha_k = O(\mathbf{M}^{-1})$.*

Proof. We begin with part (**a**). By definition of $\widetilde{f}_j$ and then Taylor series expansion of $\log(\cdot)$, we have that

$$\widetilde{\alpha}_{k,n} - \alpha_{k,n} = \frac{1}{2\mathbf{M}} \sum_{j \preceq M} \left\{ \left(\frac{\widetilde{f}_j - f_j}{f_j}\right) + \frac{1}{2}\left(\frac{\widetilde{f}_j - f_j}{f_j}\right)^2 (1 + o(1)) \right\} \cos\left(k \cdot \widetilde{\lambda}_j\right).$$

So, it suffices to examine the behavior of $f_j^{-1}\left(\widetilde{f}_j - f_j\right)$. By definition and (3.20),

$$\begin{aligned}
\frac{\widetilde{f}_j - f_j}{f_j} &= \frac{f_j^{-1}}{4\mathbf{m}} \sum_k \left\{f\left(\lambda_{k+mj}\right) - f\left(\lambda_{mj}\right)\right\} \\
&= \frac{f_j^{-1}}{4\mathbf{m}} \sum_k \left\{ \frac{k^2[1]}{n^2[1]} f_{11}\left(\lambda_{mj}\right) + \frac{k^2[2]}{n^2[2]} f_{22}\left(\lambda_{mj}\right) \right\} + O\left(\frac{1}{\mathbf{M}^2}\right) \\
&= \frac{1}{6\mathbf{M}} g_j + O\left(\frac{1}{\mathbf{M}^2}\right),
\end{aligned} \tag{B.1}$$

because $f(\lambda)$ is a four times differentiable function and $\sum_k k^{c_1}[1]k^{c_2}[2] = 0$ if $c_1 + c_2$ is an odd integer. From here the conclusion follows by standard arguments, because $g(\lambda)$ is a continuous differentiable function, so that the Riemman sums converge to their integral counterpart.

Part (**b**) follows using Lemma 6. □

LEMMA 3. *Assuming, $C1-C4$, for all $k = 1, 2, \ldots$*

$$E\left(\widetilde{f}_k^{-1}\left(\widehat{f}_k - \widetilde{f}_k\right)\right)^2 = O\left(\mathbf{m}^{-1}\right).$$

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

Proof. Because $\widetilde{f}_k = (4\mathbf{m})^{-1}\sum_j f(\lambda_{j+mk}) > 0$, the left side of the last displayed equality is, up to multiplicative constants, bounded by

$$E\left(\frac{1}{\mathbf{m}}\sum_j f(\lambda_{j+mk})\left(\frac{I_x^T(\lambda_{j+mk})}{f(\lambda_{j+mk})} - \frac{I_\vartheta^T(\lambda_{j+mk})}{\sigma_\vartheta^2}\right)\right)^2$$

$$+E\left(\frac{1}{\mathbf{m}}\sum_j f(\lambda_{j+mk})\left(\frac{I_\vartheta^T(\lambda_{j+mk})}{\sigma_\vartheta^2} - 1\right)\right)^2.$$

The first term of the last displayed expression is $o(\mathbf{m}^{-1})$ by Lemma 1, whereas the second term follows by standard arguments, as ϑ_t is an *iid* sequence of r.v.'s with finite fourth moments. □

LEMMA 4.

$$\frac{1}{\mathbf{M}^2}\sum_{p\preceq M}\left(\sum_{k_1\preceq M}\cos\left(k_1\cdot\widetilde{\lambda}_p\right)e^{-ik_1\cdot\widetilde{\lambda}_{j_1}}\right)\left(\sum_{k_2\preceq M}\cos\left(-k_2\cdot\widetilde{\lambda}_p\right)e^{ik_2\cdot\widetilde{\lambda}_{j_2}}\right) \tag{B.2}$$

$$=2\left(\delta_{j_1[1]-j_2[1]}+2^{-1}\phi_{j_1[1]}\phi_{j_2[1]}-i\phi_{j_1[1]-j_2[1]}\right)\delta_{j_1[2]\pm j_2[2]}+O(\mathbf{M}^{-1}).$$

Proof. First,

$$\sum_{k\preceq M}e^{-ik\cdot\widetilde{\lambda}_p} =: \sum_{k[2]=1}^{M}e^{-ik[2]\frac{\pi p[2]}{M}}+\sum_{k[1]=1}^{M}e^{-ik[1]\frac{\pi p[1]}{M}}\sum_{k[2]=1-M}^{M}e^{-ik[2]\frac{\pi p[2]}{M}}$$

$$=\sum_{k[2]=1}^{M}e^{-ik[2]\frac{\pi p[2]}{M}}+2M\sum_{k[1]=1}^{M}e^{-ik[1]\frac{\pi p[1]}{M}}\delta_{p[2]}$$

$$=:\mathcal{D}(p[2])+2M\mathcal{D}(p[1])\delta_{p[2]}=\Xi(p[1],p[2]),$$

where for notational simplicity, we assume that $M[1]=M[2]=:M$.

Next, because $2\cos(x)=\exp(ix)+\exp(-ix)$, we have then that (B.2) is

$$\frac{1}{4\mathbf{M}^2}\sum_{p\preceq M}\left(\sum_{k_1\preceq M}\left(e^{-ik_1\cdot\widetilde{\lambda}_{p+j_1}}+e^{-ik_1\cdot\widetilde{\lambda}_{j_1-p}}\right)\right)\left(\sum_{k_2\preceq M}\left(e^{ik_2\cdot\widetilde{\lambda}_{p+j_2}}+e^{-ik_2\cdot\widetilde{\lambda}_{p-j_2}}\right)\right)$$

$$=\frac{1}{4\mathbf{M}^2}\sum_{p\preceq M}\{(\Xi(p[1]+j_1[1],p[2]+j_1[2])+\Xi(j_1[1]-p[1],j_1[2]-p[2]))$$

$$(\Xi(-p[1]-j_2[1],-p[2]-j_2[2])+\Xi(p[1]-j_2[1],p[2]-j_2[2]))\}. \tag{B.3}$$

Let's examine a typical term on the right of (B.3), say

$$\frac{1}{4\mathbf{M}^2}\sum_{p\preceq M}\Xi(p[1]+j_1[1],p[2]+j_1[2])\ \Xi(-p[1]-j_2[1],-p[2]-j_2[2]).$$

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

By definition, the last displayed expression is

$$\frac{1}{4\mathbf{M}^2}\sum_{p\preceq M}\{\mathcal{D}(p[2]+j_1[2])\mathcal{D}(-p[2]-j_2[2])\}$$

$$+\frac{1}{2M^3}\sum_{p\preceq M}\left\{\mathcal{D}(p[2]+j_1[2])\mathcal{D}(-p[1]-j_2[1])\delta_{p[2]+j_2[2]}\right\}$$

$$+\frac{1}{2M^3}\sum_{p\preceq M}\left\{\mathcal{D}(-p[2]-j_2[2])\mathcal{D}(p[1]+j_1[1])\delta_{p[2]+j_1[2]}\right\}$$

$$+\frac{1}{M^2}\sum_{p\preceq M}\left\{\mathcal{D}(p[1]+j_1[1])\delta_{p[2]+j_1[2]}\mathcal{D}(-p[1]-j_2[1])\delta_{p[2]+j_2[2]}\right\}.$$

Because of (A.8), it easy to see that the first term is $O(M^{-1})$, whereas the second term is

$$\frac{1}{2M^3}\sum_{p[1]=1}^{M}\mathcal{D}(j_1[2]-j_2[2])\mathcal{D}(p[1]+j_2[1])$$

$$\le K\frac{1}{M^2}\sum_{p[1]=1}^{M}\mathcal{D}(p[1]+j_2[1])\le K\frac{1}{M}\sum_{p[1]=1}^{M}\frac{1}{(p[1]+j_2[1])_+}$$

$$=O\left(\frac{\log M}{M}\right),$$

so is the third term by symmetry. Finally the fourth term is different from zero if $j_1[2]=j_2[2]$, in which case it becomes

$$\frac{1}{M^2}\sum_{p[1]=1}^{M}\mathcal{D}(p[1]+j_1[1])\mathcal{D}(-p[1]-j_2[1]).$$

Then, proceeding similarly with the other three terms in (B.3), we can conclude, except negligible terms, that it is

$$\frac{1}{M^2}\sum_{p[1]=1}^{M}\{\mathcal{D}(p[1]+j_1[1])\mathcal{D}(-p[1]-j_2[1])$$

$$+\mathcal{D}(j_1[1]-p[1])\mathcal{D}(p[1]-j_2[1])\}\delta_{j_1[2]-j_2[2]}$$

$$\frac{1}{M^2}\sum_{p[1]=1}^{M}\{\mathcal{D}(j_1[1]-p[1])\mathcal{D}(-p[1]-j_2[1])$$

$$+\mathcal{D}(p[1]+j_1[1])\mathcal{D}(p[1]-j_2[1])\}\delta_{j_1[2]+j_2[2]}$$

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

$$
= \frac{2}{M^2} \sum_{k_1, k_2 \preceq M} \left(e^{-i\left(j_1[1]\widetilde{\lambda}_{k_1[1]} - j_2[1]\widetilde{\lambda}_{k_2[1]}\right)} \sum_{p[1]=1}^{M} \cos\left(\left(k_1[1] - k_2[1]\right)\widetilde{\lambda}_{p[1]}\right)\right) \delta_{j_1[2]-j_2[2]}
$$

$$
+ \frac{2}{M^2} \sum_{k_1, k_2 \preceq M} \left(e^{-i\left(j_1[1]\widetilde{\lambda}_{k_1[1]} - j_2[1]\widetilde{\lambda}_{k_2[1]}\right)} \sum_{p[1]=1}^{M} \cos\left(\left(k_1[1] + k_2[1]\right)\widetilde{\lambda}_{p[1]}\right)\right) \delta_{j_1[2]+j_2[2]}.
$$

From here we conclude by Lemma 4 of Hidalgo and Yajima (2002). □

LEMMA 5. *Under Condition C1, we have that*

$$
\sum_{\{M[1] \le k[1]\} \vee \{M[2] \le k[2]\}} \alpha_k e^{-ik\cdot\widetilde{\lambda}_j} = O\left(\mathbf{M}^{-2}\right).
$$

Proof. The proof is standard because four times continuous differentiability of $f(\lambda)$ implies that $\alpha_{k[\ell]} = O\left(k[\ell]^{-5}\right)$ for $\ell = 1, 2$. □

The next lemma is regarding the approximation of integrals by sums. Taking for simplicity that $\check{n} =: n[1] = n[2]$ and recalling our notation, we have then that $j/n =: \left(j[1]/\check{n}, j[2]/\check{n}\right)$. Also, use the standard notation, $|k| = k_1 + k_2$, $k! = k_1!k_1!$, $y^k = y_1^{k_1} y_2^{k_2}$ and for a function $\Upsilon(x)$

$$
\partial^k \Upsilon(x) = \frac{\partial^{|k|}\Upsilon(x)}{\partial x^{|k|}}.
$$

LEMMA 6. *Assume that $\Upsilon(\cdot)$ is a function q times continuously differentiable in $[0,1]^2$. Then,*

$$
\frac{1}{\check{n}^2} \sum_{j[1]=1}^{\check{n}} \sum_{j[2]=1}^{\check{n}} \Upsilon\left(\frac{j}{n}\right) - \int_0^1 \int_0^1 \Upsilon(x)\,dx = \sum_{|k| \le q-1} h_{n,k} \digamma_k + O\left(\frac{1}{\check{n}^q}\right), \tag{B.4}
$$

where $h_{n,k}$ is a sequence such that $h_{n,k} = O\left(\check{n}^{-k}\right)$ and $\digamma_1, \digamma_2, \ldots, \digamma_{q-1}$ are finite constants.

Proof. The left side of (B.4) is

$$
\sum_{j[1]=1}^{\check{n}} \sum_{j[2]=1}^{\check{n}} \int_{\frac{j[1]-1}{\check{n}}}^{\frac{j[1]}{\check{n}}} \int_{\frac{j[2]-1}{\check{n}}}^{\frac{j[2]}{\check{n}}} \left(\Upsilon\left(\frac{j}{n}\right) - \Upsilon(x)\right) dx \tag{B.5}
$$

$$
= \sum_{j[1]=1}^{\check{n}} \sum_{j[2]=1}^{\check{n}} \int_{\frac{j[1]-1}{\check{n}}}^{\frac{j[1]}{\check{n}}} \int_{\frac{j[2]-1}{\check{n}}}^{\frac{j[2]}{\check{n}}}
$$

$$
\times \left\{ \sum_{|k| \le q-1} \frac{\partial^k \Upsilon\left(\frac{j}{n}\right)}{k!} \left(x - \frac{j}{n}\right)^k + \sum_{|k|=q} \frac{\partial^k \Upsilon(x(j))}{k!} \left(x - \frac{j}{n}\right)^k \right\} dx,
$$

by Taylor's expansion and where $x(j)$ denotes a point between $(j-1)/n$ and j/n. Now, the right side of (B.5) is

$$\sum_{|k|\le q-1}\frac{1}{\check{n}^{k+2}}\sum_{j[1]=1}^{\check{n}}\sum_{j[2]=1}^{\check{n}}\frac{\partial^k\Upsilon\left(\frac{j}{n}\right)}{k!}+O\left(\frac{1}{\check{n}^q}\right). \tag{B.6}$$

Denoting $\partial^k\Upsilon(x)=\eta_k(x)$ and $\digamma_k=\int_0^1\int_0^1\eta_k(x)\,dx$, $k=1,\ldots,3$, and proceeding as with (B.5), we conclude that (B.6), and hence the left side of (B.4), is

$$\sum_{|k|\le q-1}\frac{K}{\check{n}^k}\digamma_k+O\left(\frac{1}{\check{n}^4}\right)$$

after observing that $\eta_k(x)$, $k=1,\ldots,q-1$, are respectively $q-k$ continuous differentiable functions. □

LEMMA 7. *Under Condition C1, for all p,*

$$a_{p,n}-a_p=\frac{\varrho_p}{\mathbf{M}}+O\left(\mathbf{M}^{-2}\log\mathbf{M}\right),$$

where $\left\{|\varrho_p|\right\}_{p\ge1}$ *is a summable sequence.*

Proof. By definition, $a_{p,n}-a_p$ is

$$\frac{1}{4\mathbf{M}}\sum_{-M<j\le M}\left(A_{j,n}-A_j^*\right)e^{ip\cdot\widetilde{\lambda}_j}+\frac{1}{4\mathbf{M}}\sum_{-M<j\le M}\left(A_j^*-A_j\right)e^{ip\cdot\widetilde{\lambda}_j}$$
$$+\left(\frac{1}{4\mathbf{M}}\sum_{-M<j\le M}A_je^{ip\cdot\widetilde{\lambda}_j}-\frac{1}{4\pi^2}\int_{\Pi^2}A(\lambda)\,e^{ip\cdot\lambda}d\lambda\right), \tag{B.7}$$

where $A_j^*=\exp\left(-\sum_{k\preceq M}\alpha_ke^{-ik\cdot\widetilde{\lambda}_j}\right)$. Because $A(\lambda)$ is four times continuous differentiable, the third term of (B.7) is

$$\mathbf{M}^{-1}\digamma_{2,p}+\mathbf{M}^{-3/2}\digamma_{3,p}+O\left(\mathbf{M}^{-2}\right) \tag{B.8}$$

by Lemma 6. This is the case after we notice that $\digamma_{\ell,p}$ there is given by

$$\digamma_{\ell,p}=:\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\partial^\ell A(\lambda)\,e^{ip\cdot\lambda}d\lambda=O\left(p^\ell a_p\right)=O\left(p^{\ell-5}\right)$$

which is clearly summable since $\ell\le3$, and because $A(-\pi,\lambda[2])=A(\pi,\lambda[2])$ and $A(\lambda[2],-\pi)=A(\lambda[2],\pi)$ for every $\lambda[1],\lambda[2]\in\Pi$ implies that

$$\digamma_{1,p}=:\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left(\frac{\partial A(\lambda)\,e^{ip\cdot\lambda}}{\partial\lambda[1]}+\frac{\partial A(\lambda)\,e^{ip\cdot\lambda}}{\partial\lambda[2]}\right)d\lambda=0.$$

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

Next, the second term of (B.7) is bounded in absolute value by

$$\left| \frac{1}{4\mathbf{M}} \sum_{-M<j\leq M} \left(\exp\left\{ \sum_{p}{}^{\dagger} \alpha_k e^{-ik\cdot\widetilde{\lambda}_j} \right\} - 1 \right) A_j e^{ip\cdot\widetilde{\lambda}_j} \right|$$

$$\leq \frac{K}{4\mathbf{M}} \sum_{-M<j\leq M} \left|A_j\right| \left| \sum_{k}{}^{\dagger} \alpha_k e^{-ik\cdot\widetilde{\lambda}_j} \right| (1+O(1)) = O\left(\mathbf{M}^{-2}\right), \tag{B.9}$$

by Lemma 5 and that $\sum_{-M<j\leq M} \left|A_j\right| = O(\mathbf{M})$, where $\sum_p^{\dagger}$ denotes the summation in

$\mathcal{S}(p) =: \{p : (0 \prec p) \wedge \{(M[1] < p[1]) \vee (M[2] < p[2])\}\}.$

Finally, by definition of $A_{j,n}$ and A_j^*, the first term of $(B.7)$ is

$$\frac{1}{2\mathbf{M}} \sum_{-M<j\leq M} \left(\exp\left\{ \sum_{k\preceq M} \left(\alpha_{k,n} - \alpha_k\right) e^{-ik\cdot\widetilde{\lambda}_j} \right\} - 1 \right) A_j^* e^{ip\cdot\widetilde{\lambda}_j}, \tag{B.10}$$

where using the inequality in (A.8), we have that $\sum_{k\preceq M} \left(\alpha_{k,n} - \alpha_k\right) e^{-ik\cdot\widetilde{\lambda}_j}$ is

$$\sum_{-k\preceq M} \left(\frac{1}{4\mathbf{M}} \sum_{-M<\ell\leq M} \log(f_\ell) \cos\left(k\cdot\widetilde{\lambda}_\ell\right) - \frac{1}{4\pi^2} \int_{\Pi^2} \log(f(\lambda)) \cos(k\cdot\lambda)\, d\lambda \right) e^{-ik\cdot\widetilde{\lambda}_j}$$

$$= \sum_{k\preceq M} \left(\frac{1}{\mathbf{M}} F_{2,k} + \frac{1}{\mathbf{M}^{3/2}} F_{3,k} \right) e^{-ik\cdot\widetilde{\lambda}_j} + O\left(\mathbf{M}^{-1} j^{-1}\right), \tag{B.11}$$

by Lemma 6 and where now $F_{\ell,k} = \int_{\Pi^2} \log(f(\lambda)) \cos(k\cdot\lambda)\, d\lambda$, $\ell = 2,3$. Next, because $\log(f(\lambda))$ is four times continuously differentiable, it implies that $F_{\ell,p} =: O\left(p^{\ell-5}\right)$ and thus by standard arguments, the right side of (B.11) is

$$\mathbf{M}^{-1}\psi_2\left(\widetilde{\lambda}_j\right) + \mathbf{M}^{-3/2}\psi_3\left(\widetilde{\lambda}_j\right) + O\left(\mathbf{M}^{-1} j^{-1}\right)$$

$$\psi_\ell(\lambda) = \sum_{k\preceq\infty} F_{\ell,k} e^{-ik\cdot\lambda}.$$

From here and using Taylor expansion of $\exp(x)$, we obtain that $(B.10)$ is

$$\frac{1}{2\mathbf{M}} \sum_{-M<j\leq M} \left\{ \mathbf{M}^{-1}\psi_2\left(\widetilde{\lambda}_j\right) + \mathbf{M}^{-3/2}\psi_3\left(\widetilde{\lambda}_j\right) \right\} A_j^* e^{ip\cdot\widetilde{\lambda}_j} + O\left(\mathbf{M}^{-2}\log\mathbf{M}\right)$$

$$= \frac{\nu_{p,2}}{\mathbf{M}} + \frac{\nu_{p,3}}{\mathbf{M}^{3/2}} + O\left(\mathbf{M}^{-2}\log\mathbf{M}\right) \tag{B.12}$$

where $\left\{\nu_{p,\ell}\right\}_{p\geq 1}$, $\ell = 2,3$, are the Fourier coefficients of $\psi_\ell(\lambda) A^*(\lambda)$, which are summable because $\psi_\ell(\lambda) A^*(\lambda)$ is $4-\ell$ times differentiable function. The conclusion of the lemma now follows by gathering terms (B.8), (B.9) and (B.12). □

https://doi.org/10.1017/S0266466622000226 Published online by Cambridge University Press

REFERENCES

Baltagi, B. H., H. H. Kelejian, & I. R. Prucha (2007) Analysis of spatially dependent data. *Journal of Econometrics* 140, 1–4.

Banerjee, S., A.E. Gelfand, J.R. Knight, & C.F. Sirmans (2004) Spatial modeling of house prices using normalized distance-weighted sums of stationary processes. *Journal of Business & Economic Statistics* 22, 206–213.

Batchelor, L. & H. Reed (1918) Relation of the variability of yields of fruit trees to the accuracy of field trials. *Journal of Agricultural Research* XII, 245–283.

Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B* 36, 192–236.

Bester, C.A., T.G. Conley, & C.B. Hansen (2011) Inference with dependent data using cluster covariance estimators. *Journal of Econometrics* 165, 137–151.

Bester, C.A., T.G. Conley, C.B. Hansen, & T.J. Vogelsang (2016) Fixed-b asymptotics for spatially dependent robust nonparametric covariance matrix estimators. *Econometric Theory* 32, 154–186.

Bhansali, R.J. (1974) Asymptotic properties of the Wiener–Kolmogorov predictor. I. *Journal of the Royal Statistical Society: Series B* 36, 61–73.

Bhansali, R.J. (1978) Linear prediction by autoregressive model fitting in the time domain. *Annals of Statistics* 60, 224–231.

Breidt, F.J., R.A. Davis, & A.A. Trindade (2001) Least absolute deviation estimation for all-pass time series models. *Annals of Statistics* 29, 919–946.

Brillinger, D.R. (1981) *Time Series: Data Analysis and Theory*. Holden Day.

Cavaliere, G., H.B. Nielsen, & A. Rahbek (2020) Bootstrapping noncausal autoregressions: With applications to explosive bubble modeling. *Journal of Business & Economic Statistics* 38, 55–67.

Conley, T.G. (1999) GMM estimation with cross sectional dependence. *Journal of Econometrics* 92, 1–45.

Conley, T.G. & F. Molinari (2007) Spatial correlation robust inference with imperfect distance information. *Journal of Econometrics* 140, 76–96.

Cressie, N. & H. Huang (1999) Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association* 94, 1330–1340.

Cressie, N.A. (1993) *Statistics for Spatial Data*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons.

Dahlhaus, R. & H. Künsch (1987) Edge effects and efficient parameter estimation for stationary random fields. *Biometrika* 74, 877–882.

Davis, R.A., C. Klüppelberg, & C. Steinkohl (2013) Statistical inference for max-stable processes in space and time. *Journal of the Royal Statistical Society: Series B* 75, 791–819.

Fernandez-Casal, R., W. Gonzalez-Manteiga, & M. Febrero-Bande (2003) Flexible spatio-temporal stationary variogram models. *Statistics and Computing* 13, 127–136.

Gao, J., Z. Lu, & D. Tjøstheim (2006) Estimation in semiparametric spatial regression. *The Annals of Statistics* 34, 1395–1435.

Genton, M.G. & H.L. Koul (2008) Minimum distance inference in unilateral autoregressive lattice processes. *Statistica Sinica* 18, 617–631.

Gupta, A. (2018) Autoregressive spatial spectral estimates. *Journal of Econometrics* 203, 80–95.

Guyon, X. (1982) Parameter estimation for a stationary process on a d-dimensional lattice. *Biometrika* 69, 95–105.

Haining, R.P. (1978) The moving average model for spatial interaction. *Transactions of the Institute of British Geographers* 3, 202–225.

Helson, H. & D. Lowdenslager (1958) Prediction theory and Fourier series in several variables. *Acta Mathematica* 99, 165–202.

Helson, H. & D. Lowdenslager (1961) Prediction theory and Fourier series in several variables. II. *Acta Mathematica* 106, 175–213.

Hidalgo, J. (2009) Goodness of fit for lattice processes. *Journal of Econometrics* 151, 113–128.

Hidalgo, J. & Y. Yajima (2002) Prediction and signal extraction of strongly dependent processes in the frequency domain. *Econometric Theory* 18, 584–624.

IversenJr. , E. (2001) Spatially disaggregated real estate indices. *Journal of Business & Economic Statistics* 19, 341–357.

Jenish, N. (2016) Spatial semiparametric model with endogenous regressors. *Econometric Theory* 32, 714–739.

Jenish, N. & I.R. Prucha (2012) On spatial processes and asymptotic inference under near-epoch dependence. *Journal of Econometrics* 170, 178–190.

Korezlioglu, H. and P. Loubaton (1986). *Spectral factorization of wide sense stationary processes on* $\mathbb{Z}^2$. *Journal of Multivariate Analysis* 19, 24–47.

Lanne, M. & P. Saikkonen (2011) Noncausal autoregressions for economic time series. *Journal of Time Series Econometrics* 3, 1–30.

Lanne, M. & P. Saikkonen (2013) Noncausal vector autoregression. *Econometric Theory* 29, 447–481.

Lewis, R. & G.C. Reinsel (1985) Prediction of multivariate time series by autoregressive model fitting. *Journal of Multivariate Analysis* 16, 393–411.

Limaye, B.V. & M. Zeltser (2009) On the Pringsheim convergence of double series. *Proceedings of the Estonian Academy of Sciences* 58, 108.

Majumdar, A., H.J. Munneke, A.E. Gelfand, S. Banerjee, & C.F. Sirmans (2006) Gradients in spatial response surfaces with application to urban land values. *Journal of Business & Economic Statistics* 24, 77–90.

McElroy, T.S. & S.H. Holan (2014) Asymptotic theory of cepstral random fields. *The Annals of Statistics* 42, 64–86.

Mercer, W.B. & A.D. Hall (1911) The experimental errors of field trials. *Journal of Agricultural Science* IV, 107–132.

Mitchell, M.W., M.G. Genton, & M.L. Gumpertz (2005) Testing for separability of space-time covariances. *Environmetrics* 16, 819–831.

Nychka, D., S. Bandyopadhyay, D. Hammerling, F. Lindgren, & S. Sain (2015) A multiresolution Gaussian process model for the analysis of large spatial datasets. *Journal of Computational and Graphical Statistics* 24, 579–599.

Pace, R.K. & R. Barry (1997) Sparse spatial autoregressions. *Statistics & Probability Letters* 33, 291–297.

Robinson, P.M. (2007) Nonparametric spectrum estimation for spatial data. *Journal of Statistical Planning and Inference* 137, 1024–1034.

Robinson, P.M. (2011) Asymptotic theory for nonparametric regression with spatial data. *Journal of Econometrics* 165, 5–19.

Robinson, P.M. & J. Vidal Sanz (2006) Modified Whittle estimation of multilateral models on a lattice. *Journal of Multivariate Analysis* 97, 1090–1120.

Roknossadati, S.M. & M. Zarepour (2010) M-estimation for a spatial unilateral autoregressive model with infinite variance innovations. *Econometric Theory* 26, 1663–1682.

Serfling, R. (1980) *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons.

Solo, V. (1986) Modeling of two-dimensional random fields by parametric Cepstrum. *IEEE, Transactions on Information Theory* 42, 743–750.

Stein, M. (1999) *Interpolation of Spatial Data: Some Theory for Kriging*. Spinger-Verlag.

Wang, H., E.M. Iglesias, & J.M. Wooldridge (2013) Partial maximum likelihood estimation of spatial probit models. *Journal of Econometrics* 172, 77–89.

Whittle, P. (1954) On stationary processes in the plane. *Biometrika* 41, 434–449.