Modified Whittle estimation of multilateral models on a lattice

P.M. Robinson*, J. Vidal Sanz

London School of Economics, Houghton Street, London WC2A 2AE, UK

Abstract

In the estimation of parametric models for stationary spatial or spatio-temporal data on a d-dimensional lattice, for $d \geqslant 2$, the achievement of asymptotic efficiency under Gaussianity, and asymptotic normality more generally, with standard convergence rate, faces two obstacles. One is the "edge effect", which worsens with increasing d. The other is the possible difficulty of computing a continuous-frequency form of Whittle estimate or a time domain Gaussian maximum likelihood estimate, due mainly to the Jacobian term. This is especially a problem in "multilateral" models, which are naturally expressed in terms of lagged values in both directions for one or more of the d dimensions. An extension of the discrete-frequency Whittle estimate from the time series literature deals conveniently with the computational problem, but when subjected to a standard device for avoiding the edge effect has disastrous asymptotic performance, along with finite sample numerical drawbacks, the objective function lacking a minimum-distance interpretation and losing any global convexity properties. We overcome these problems by first optimizing a standard, guaranteed non-negative, discrete-frequency, Whittle function, without edge-effect correction, providing an estimate with a slow convergence rate, then improving this by a sequence of computationally convenient approximate Newton iterations using a modified, almost-unbiased periodogram, the desired asymptotic properties being achieved after finitely many steps. The asymptotic regime allows increase in both directions of all d dimensions, with the central limit theorem established after re-ordering as a triangular array. However our work offers something new for "unilateral" models also. When the data are non-Gaussian, asymptotic variances of all parameter

* Corresponding author. Tel.: +44 20 7955 7516; fax: +44 20 7955 6592.
E-mail address: p.m.robinson@lse.ac.uk (P.M. Robinson).

estimates may be affected, and we propose consistent, non-negative definite estimates of the asymptotic variance matrix.

AMS 1991 subject classification: 62M30; 62F10; 62F12

Keywords: Spatial data; Multilateral modelling; Whittle estimation; Edge effect; Consistent variance estimation

1. Introduction

Consider a stationary process x_t defined on a d-dimensional lattice, t being a multiple index $(t_1, \dots, t_d)$ with $t_j \in \mathbb{Z} = \{0, \pm 1, \dots\}$, $j = 1, \dots, d$, and having a spectral density $f(\lambda)$, $\lambda = (\lambda_1, \dots, \lambda_d)$, $\lambda \in \Pi^d$, $\Pi = (-\pi, \pi]$. This paper is concerned with large sample inference on an unknown m-dimensional column vector θ_0, given a known functional form $f(\lambda; \theta)$, $\theta \in \Theta \subset \mathbb{R}^m$, such that $f(\lambda; \theta_0) \equiv f(\lambda)$, for x_t observed on the rectangular lattice $\mathbb{N} = \{t : -n_{Li} \leqslant t_i \leqslant n_{Ui}, i = 1, \dots, d\}$, $n_{Ui}, n_{Li} \geqslant 0$, $i = 1, \dots, d$. Define $n_i = n_{Li} + n_{Ui} + 1$, $n = \prod_{i=1}^d n_i$, and regard each $n_i = n_i(n)$ as a function of the total number of observations n. Though we only introduce parameter estimates that are based on such a full lattice, our asymptotic construction regards observations as arising singly; the sequence of estimates is defined only with respect to increase in one or the other of the n_i but we can nest the consequent n sequence in $\mathbb{Z}_+ = \{1, 2, \dots\}$. A mild degree of regularity in the n_i, across i, is implied by the following assumption.

A1. For all sufficiently large n, there exist $\xi > 0$, $C_1 > 0$ such that

$$n_i(n) \geqslant C_1 n^{\xi}, \quad i = 1, \dots, d. \tag{1.1}$$

Remark 1. The inequality between arithmetic and geometric means indicates that

$$\sum_{i=1}^{d} n_i^{-1}(n) \geqslant d n^{-1/d}, \tag{1.2}$$

so that $\xi \leqslant 1/d$, the equality here indicating that all n_i increase at the same, $n^{1/d}$, rate. Assumption A1 can hold if, for all i, only one of n_{Ui} and n_{Li} increases unboundedly with n, so that the common random fields prescription $n_{Li} \equiv 0$ is included. It is sometimes artificial to suppose that further sampling is only possible in particular directions.

Remark 2. Domains of observation are often bounded, and "infill" asymptotics (see e.g. [5, 40]) has appeal. This would require modelling x_t continuously across the domain; our goal is to justify useful rules of inference rather than explore issues of interpolation.

The asymptotic properties we aim for in estimating θ_0 are efficiency when x_t is Gaussian, and $n^{\frac{1}{2}}$-consistency and asymptotic normality much more generally.

Definition 1. An estimate $\hat{\theta}$ of θ_0 is said to satisfy Property E if $n^{1/2}(\hat{\theta} - \theta_0)$ converges in distribution to a $\mathcal{N}\left(0, \Phi^{-1}\Psi\Phi^{-1}\right)$ variate, where Φ and Ψ are non-singular matrices given by

$$\Phi = (2\pi)^{-d} \int_{\Pi^d} \partial(\lambda; \theta_0)\partial'(\lambda; \theta_0)\, d\lambda, \quad \partial(\lambda; \theta) = \frac{\partial \log f(\lambda; \theta)}{\partial \theta},$$
$$\Psi = 2\Phi + \kappa \left\{(2\pi)^{-d} \int_{\Pi^d} \partial(\lambda; \theta_0)\, d\lambda\right\} \left\{(2\pi)^{-d} \int_{\Pi^d} \partial(\lambda; \theta_0)\, d\lambda\right\}', \tag{1.3}$$

the prime denoting transposition and κ as defined in the following assumption.

A2. x_t has representation

$$x_t = \mu + \sum_j \beta_{t-j}\varepsilon_j, \quad \sum_j \left|\beta_j\right| < \infty,$$

where the ε_j are independent and identically distributed with zero mean, unit variance and finite fourth cumulant, denoted κ, and $\sum_j$ denotes $\sum_{j \in \mathbb{Z}^d}$.

Remark 3. Alternative conditions to A2 such as mixing conditions and instantaneous nonlinear transformations of a weakly dependent Gaussian process (see e.g. [10,33,34,44]) would be strictly neither stronger nor weaker than A2. The linearity assumption is natural in the context of models discussed in the following section and for the quadratic statistics on which our estimates are based. Note the possibly "multilateral" character of the representation for x_t. Summability of the β_j in A2 is mild by the standard of many weak dependence conditions, and A2 would also be natural in an extension of our work to adaptive estimation, where efficiency improvements are achieved in the presence of unknown, non-Gaussian distribution for ε_t.

Remark 4. Deriving good estimates of θ_0 is also of value in efficient trend estimation, when μ is replaced by a more general regression function.

The following section motivates the above setting and the estimation procedure to be introduced, by discussing a particular class of $f(\lambda; \theta)$. Section 3 reviews the background to the estimation problem, and Sections 4 and 5 present an estimation strategy and asymptotic results. A Monte Carlo study of finite sample performance is reported in Section 6. Section 7 provides consistent estimates of Ψ when x_t is non-Gaussian. Proofs for Sections 4, 5 and 7 are left to Sections 8 and 9.

2. Multilateral and Unilateral ARMA Models

For $z = (z_1, ..., z_d)$ having complex-valued elements, and $\theta \in \Theta$, define

$$a(z; \theta) = \sum_{j_1=-p_{L1}}^{p_{U1}} \cdots \sum_{j_d=-p_{Ld}}^{p_{Ud}} a_j(\theta) \prod_{i=1}^{d} z_i^{j_i}, \tag{2.1}$$

$$b(z;\theta) = \sum_{j_1=-q_{L1}}^{q_{U1}} \cdots \sum_{j_d=-q_{Ld}}^{q_{Ud}} b_j(\theta) \prod_{i=1}^{d} z_i^{j_i}, \tag{2.2}$$

for $j = (j_1, ..., j_d)$, given finite integers $p_{Li} \geqslant 0$, $p_{Ui} \geqslant 0$, $q_{Li} \geqslant 0$, $q_{Ui} \geqslant 0$, and real-valued functions $a_j(\theta)$, $b_j(\theta)$. We call (2.1) and (2.2) multivariate polynomials, even though they can involve negative powers. Denoting by $B = (B_1, ..., B_d)$ the operator such that $\prod_{i=1}^{d} B_i^{j_i} x_t = x_{t-j}$, where $t - j$ is the multiple index $(t_1 - j_1, ..., t_d - j_d)$, suppose x_t has the autoregressive moving average (ARMA) representation

$$\begin{aligned} &\text{ARMA}\,(p_{L1}, p_{U1}; ...; p_{Ld}, p_{Ud} : q_{L1}, q_{U1}; ...; q_{Ld}, q_{Ud}) : a(B;\theta_0)(x_t - \mu) \\ &\quad = b(B;\theta_0)\varepsilon_t, \end{aligned} \tag{2.3}$$

where $\mu = Ex_t$, ε_t satisfies A2 and

$$a(z;\theta) \neq 0,\ b(z;\theta) \neq 0, \quad \text{for } |z_i| = 1,\ i = 1, ..., d,\ \theta \in \Theta. \tag{2.4}$$

Under these conditions, $f(\lambda)$ is finite and positive, and we take

$$f(\lambda;\theta) = (2\pi)^{-d} \left| b\,(e(i\lambda);\theta)\,/a\,(e(i\lambda);\theta) \right|^2, \quad \theta \in \Theta, \tag{2.5}$$

with $e(z) = (e^{z_1}, ..., e^{z_d})$. The summability condition in A2 is satisfied by (2.3), (2.4). Special cases of (2.3) are the autoregressive (AR) model AR $(p_{L1}, p_{U1}; ...; p_{Ld}, p_{Ud})$ when $b(z;\theta_0) \equiv 1$ and the moving average (MA) model MA $(q_{L1}, q_{U1}; ...; q_{Ld}, q_{Ud})$ when $a(z;\theta_0) \equiv 1$. [25] introduced a class of model that overlaps with (2.5). [28,29] developed other classes of models with desirable properties.

Any of the $p_{Li}, p_{Ui}, q_{Li}, q_{Ui}$ in (2.3) can be positive, so these ARMA structures can be "multilateral", and they provide a flexible approach to modelling. It is necessary that θ be identifiable from $f(\lambda;\theta)$, $\lambda \in \Pi^d$, if x_t is Gaussian or, more generally, if information is confined to second moments of x_t. This requires first that θ be identifiable from $a(z;\theta)^{-1}b(z;\theta)$. In the general ARMA case it is necessary that a and b not be over-specified, so they have no common factor, which implies, since A2 fixes $E\varepsilon_t^2 = 1$, a suitable normalization of a or b, such as $b_0(\theta) \equiv 1$. These requirements are innocuous in the AR or MA special cases, but $|a\,(z;\theta)|^2$, $|b\,(z;\theta)|^2$ need not uniquely determine $a\,(z;\theta)$, $b\,(z;\theta)$. A given $a(z;\theta)$, with real-valued coefficients, can be replaced by $\tilde{a}(z;\theta) = \prod_{i=1}^{d} z_i^{j_i} a(z;\theta)$ for any positive or negative integer j_i, but this involves a trivial translation on $\mathbb{Z}^d$, which can be viewed as locating the innovation at $t - j$ rather than t (see [46]), and is thus disregarded. To indicate a more substantive concern, write for $h \geqslant 1$

$$a(z;\theta) = \prod_{j=1}^{h} a_j(z;\theta), \quad \text{all } \theta \in \Theta, \tag{2.6}$$

where the $a_j(z;\theta)$ are non-constant multivariate polynomials, with coefficients that can be complex-valued. When $h > 1$, $a(z;\theta)$ is said to be factorizable, and if $a_j(z;\theta)$ is not factorizable, it is said to be irreducible (see, e.g. [45, pp. 58–62]). Denote by $a_j(z^{-1};\theta)$ the function obtained by replacing z_i by z_i^{-1}, for $i = 1, ..., d$, in $a_j(z;\theta)$. If all $a_j(z;\theta)$ are irreducible, those of the 2^h functions $\prod_{j=1}^{h} a_j\left(z^{\pm 1};\theta\right)$ with real-valued coefficients are indistinguishable.

When $d = 1$, and t denotes time, ambiguity is commonly avoided by focussing on "unilateral" models. Here, an irreducible factorization has $h = p_{L1} + p_{U1}$, and $a(z;\theta)$ is indistinguishable from a $(p_{L1}+p_{U1})$th-degree polynomial in z with all powers non-negative, the usual automatic choice (and given (2.4) there is no loss of generality in specifying all its zeros to be outside the unit circle, the usual "stationarity" condition). On the other hand the requirement that coefficients be real can eliminate possibilities; for example, commencing from $a(z;\theta) = \theta_1 + \theta_2 z + \theta_3 z^2$, with complex-valued zeros, where θ_j is the jth element of θ, there is no equivalent bilateral AR(1, 1) model.

Unilateral structures have been studied when $d \geqslant 2$ also, and these are covered as special cases of our multilateral model. These may have a natural unilateral representation as, for example, when $d = 2$ in the AR(0, 1; 0, 1) model, where

$$a(z;\theta) = 1 - \theta_1 z_1 - \theta_2 z_2, \ \ |\theta_1| + |\theta_2| < 1; \ b(z;\theta) = e^{\theta_3/2}. \tag{2.7}$$

For (2.7) there are simple unilateral AR and (infinite) MA representations on a quadrant, such that x_t (ε_t) is expressed in terms of ε_s (x_s) for $s_j \leqslant t_j$, all j On the other hand [21,26,42,43] discussed conditions under which models that might initially be expressed in multilateral form have infinite AR and MA representations on a quadrant. More general representations have also been referred to as "unilateral". Under conditions easily satisfied by (2.3), (2.4) and A2, x_t has an infinite linear MA representation in orthogonal innovations ζ_s for $s \preceq t$, with square summable coefficients, where $\preceq$ denotes lexicographic order. This extends the Wold representation theorem, and there is a corresponding unilateral infinite AR representation if also $f(\lambda)$ is everywhere positive; see [10,15,26,46].

Such unilateral representations form a framework for extending to $d \geqslant 2$ ARMA order-determination methods and AR nonparametric spectral estimation methods; see [19,42]. They have also been employed in parametric modelling (e.g. [9,18,48]). However, for $d \geqslant 2$ a multilateral finite ARMA given by (2.3) cannot necessarily be represented as a unilateral finite ARMA, as demonstrated in a simple example in [46], where $d = 2$, $m = 1$, $a(z;\theta) = 1 + \theta^2 - \theta\left(z_1 + z_2 + z_2^{-1}\right)$, $b(z;\theta) \equiv 1$; [46] gave a closed form expression for the unilateral infinite AR operator. Whittle [46] did not present the derivation, but it is explained by showing its equivalence to $a(z;\theta)\,(1 - \theta z_2)\left(1 - \theta z_2^{-1}\right)^{-1}$. This indicates a trick that applies somewhat more generally; in particular when $d = m = 2$, $a(z;\theta) = 1 + \theta_2^2 - \theta_1 z_1 - \theta_2(z_2 + z_2^{-1})$ and $b(z;\theta) \equiv 1$,

$$\begin{aligned} &a(z;\theta)(1 - \theta_2 z_2)(1 - \theta_2 z_2^{-1})^{-1} \\ &\quad = 1 - 2\theta_2 z_2 + \theta_2^2 z_2^2 + \theta_1\theta_2 z_1 z_2 - \theta_1(1 - \theta_2^2)(1 - \theta_2 z_2^{-1})^{-1} \end{aligned}$$

is unilateral. (The same multilateral model was also considered by [20], but the unilateral form there appears not to have the same spectral density.) However, it does not work in general, where, even in simple cases such as $d = 2$, $m = 1$, $a(z;\theta) = 1 - \theta(z_1 + z_1^{-1} + z_2 + z_2^{-1})$, $b(z;\theta) \equiv 1$, as [46] also noted, formulae for unilateral representations can be intractable. Spatial dimensions may have no natural direction, so the choice of unilateral direction may in any case be arbitrary.

Following [46], lattice multilateral models driven by white noise, such as (2.3), have been discussed in, for example, [1,2,4,8,10,11,22,30,32,35]. The allowance in (2.3) for the $a_j(\theta)$

and $b_j(\theta)$ to depend on a vector θ of possibly small dimension m relative to the number, $\Pi_{i=1}^d (p_{Li} + p_{Ui} + 1) + \Pi_{i=1}^d (q_{Li} + q_{Ui} + 1) - 1$, of ARMA coefficients can ease the identification problem. Symmetry restrictions (see [1]) can be physically natural and lead to real-valued $a(z;\theta)$ or $b(z;\theta)$. Inequality restrictions are easily enforced in estimation and even when arbitrary are less drastic than choosing the direction of a unilateral model. The structure in [31], in which $h = d$ in (2.6) and $a_j(z;\theta)$ varies with z_j only, can reduce the identification problem to the familiar one when $d = 1$. Isotropic assumptions (e.g. [40]) are another way of introducing parsimony. The multilateral spatial aspect itself is only responsible for finitely many observational equivalents, compared to the uncountable infinity due to overspecified ARMA modelling.

3. Background to Estimation and the Edge-Effect

As an alternative to maximum likelihood, the lattice literature has discussed the original, continuous-frequency, form of estimate proposed by Whittle. Define

$$Q_{C1}(\theta) = (2\pi)^{-d} \int_{\Pi^d} \log f(\lambda;\theta) d\lambda, \quad Q_{C2}(\theta;h) = (2\pi)^{-d} \int_{\Pi^d} \frac{h(\lambda)}{f(\lambda;\theta)} d\lambda,$$
$$Q_C(\theta;h) = Q_{C1}(\theta) + Q_{C2}(\theta;h), \quad \hat{\theta}_C(h) = \arg\min_{\Theta} Q_C(\theta;h),$$

for a generic even function $h(\lambda)$. Introduce the periodogram

$$I(\lambda) = (2\pi)^{-d} \sum_j{}' c_j \cos(j.\lambda),$$

where

$$c_j = n^{-1} \sum_{t(j)} (x_t - \bar{x})(x_{t+j} - \bar{x}), \quad \bar{x} = n^{-1} \sum_{t \in \mathbb{N}} x_t,$$

such that $\sum_j'$ is a sum over $1 - n_i \leqslant j \leqslant n_i - 1$, $i = 1, \ldots, d$, $\sum_{t(j)}$ is a sum over $-n_{Li} \leqslant t_i$, $t_i + j_i \leqslant n_{Ui}$, $i = 1, \ldots, d$, and for d-dimensional quantities such as j that are introduced as a multiple subscript rather than a vector we employ the notation $j.\lambda = \sum_{i=1}^d j_i \lambda_i$.

For $d = 1$, $h(\lambda) = I(\lambda)$ is usual. For a finite AR, $Q_{C2}(\theta; I)$ and its derivatives in θ are analytically evaluated as linear combinations of finitely many c_j, but in MA or ARMA models the calculation is less simple. Even in the AR case $Q_{C1}(\theta)$ can be difficult to calculate. In standard parameterizations of unilateral models $Q_{C1}(\theta)$ is the log variance of the one-step-ahead predictor, and depends only on an element of θ functionally unrelated to the remainder, for example in the AR$(0,1;0,1)$ (2.7) we have $f(\lambda;\theta) = (2\pi)^{-2} e^{\theta_3} / \left|1 - \theta_1 e^{i\lambda_1} - \theta_2 e^{i\lambda_2}\right|^2$, so that $Q_{C1}(\theta) = \theta_3 - 2\log(2\pi)$. However in multilateral models it generally depends on the whole of θ, and does not have a neat closed form; even in quite simple models, [46] found only infinite series representations, whose individual terms can be complicated. Also, [48] showed, with $d = 2$, that the time-domain Gaussian pseudo-likelihood can be conveniently handled (even in the presence of missing data) in case of unilateral finite ARMA models, but for multilateral models it poses similar difficulties to $Q_C(\theta; I)$ (see [1]). Given formulae for autocovariances, algorithms for handling block-Toeplitz matrices, such as those in the SLICOT library, can provide rapid computation of the Gaussian pseudo-likelihood. However, convenient formulae for autocovariances are only available in simple versions of ARMA models, especially because of the possibility of repeated roots in the AR operator.

A statistical drawback of $\hat{\theta}_C(I)$ noted by [9] is the edge effect: for fixed j, as the $n_i \to \infty$ the bias of c_j for $\gamma_j = \operatorname{cov}(x_0, x_j)$ is of order $\sum_{i=1}^{d} n_i^{-1}$, which by (1.2) is of order no less than $n^{-1/d}$. As (1.1) suggests, $\hat{\theta}_C(I)$ is n^{ξ}-consistent: for $d = 2$ it is $n^{\frac{1}{2}}$-consistent only when both n_i increase at the same rate, and even then $n^{\frac{1}{2}}\left(\hat{\theta}_C(I) - \theta_0\right)$ converges in distribution to a variate with non-zero mean, while for $d \geqslant 3$ $\hat{\theta}_C(I)$ is never $n^{\frac{1}{2}}$-consistent; thus for $d \geqslant 2$ $\hat{\theta}_C(I)$ lacks Property E.

The computational drawbacks of $\hat{\theta}_C(I)$ can be avoided by extending the discrete form of Whittle estimate considered by [13] in the time series case $d = 1$. Define

$$Q_{D1}(\theta) = \frac{1}{n}\sum_{j\in\mathbb{N}} \log f(\omega_j; \theta), \quad Q_{D2}(\theta; h) = \frac{1}{n}\sum_{j\in\mathbb{N}} \frac{h(\omega_j)}{f(\omega_j; \theta)},$$

$$Q_D(\theta; h) = Q_{D1}(\theta) + Q_{D2}(\theta; h), \quad \hat{\theta}_D(h) = \arg\min_{\Theta} Q_D(\theta; h),$$

where $\omega_j = (2\pi j_1/n_1, ..., 2\pi j_d/n_d)$. Regarding Q_D as an approximation to Q_C, the quadrature rule employed is not arbitrary, since the ω_j are just sufficiently finely spaced for $\hat{\theta}_D(I)$ to have the same asymptotic properties as $\hat{\theta}_C(I)$; a coarser grid, or one fixed with respect to n, would incur asymptotic bias. Q_D is motivated by models in which $f(\lambda; \theta)$ has a simple closed form. This is not always the case. The authors of Refs. [30,40,46,47] stressed models in which the spectral density of an underlying continuous model, on $\mathbb{R}^d$, has simple form, but application of the usual "folding" formula does not produce a neat closed form for $f(\lambda; \theta)$. However in view of (2.5), Q_D is convenient for multilateral ARMA models, as well as ARMA-signal-plus-ARMA-noise ones, also motivated in [46]. Unlike when $d = 1$, signal-plus-noise processes do not necessarily have finite ARMA representations, because a non-negative multivariate trigonometric polynomial cannot necessarily be factored [23]. Likewise [38] motivated reciprocals of such polynomials without requiring an AR representation. Also, [24] discussed an objective function based on a matrix which would be the variance matrix of the data if x_t, $t \in \mathbb{Z}^d$, form a circulant based on x_t, $t \in \mathbb{N}$. This is equivalent to replacing the $f(\omega_j; \theta)$ in $Q_D(\theta; I)$ by quantities which differ if $f(\lambda; \theta)$ is not a finite trigonometric polynomial (so is not an MA), and are in general of complicated form. In cases when $Q_{C1}(\theta)$ is simple to calculate we might consider the hybrid objective function $Q_{C1}(\theta) + Q_{D2}(\theta; I)$; for example in the unilateral special case (2.7) this produces a closed form estimate of θ_0 (as does $Q_C(\theta; I)$). In general, however, the minimum-distance property of $\hat{\theta}_C(I)$ or $\hat{\theta}_D(I)$ may be lost.

The same edge-effect bias is found in $\hat{\theta}_D(I)$ as in $\hat{\theta}_C(I)$, with respect to which [9] suggested replacing $I(\lambda)$ by the almost-unbiased

$$I_*(\lambda) = (2\pi)^{-d} \sum_j{}' c_j^* \cos(j.\lambda),$$

where

$$c_j^* = \left\{ n / \prod_{i=1}^{d} (n_i - |j_i|) \right\} c_j.$$

With $n_{Li} \equiv 0$ and the n_{Ui} increasing, [9] showed that $\hat{\theta}_C(I_*)$ can satisfy Property E, thereby avoiding edge-effect bias similar to A2; see also [16]. Dahlhaus and Künsch [6] criticized $\hat{\theta}_C(I_*)$ as lacking a minimum-distance interpretation and possibly being harder to locate than the minimizer of an objective function that is guaranteed non-negative, citing numerical experience in support.

Theoretical properties of $\hat{\theta}_D(I_*)$ are disastrous. It suffices to look at the very simple case of a unilateral AR(1) with $d = m = 1$, $x_t = \theta x_{t-1} + \varepsilon_t$, $|\theta_0| < 1$, where

$$Q_{D2}(\theta; I_*) = c_0^*(1 + \theta^2) - 2\theta \left(c_1^* + c_{n-1}^* \right) = Q_{C2}(\theta; I_*) - 2\theta x_1 x_n.$$

Since $x_1 x_n$ does not converge to a non-degenerate random variable (its variance tending to $(1 - \theta_0^2)^{-2}$ in the Gaussian case), $\hat{\theta}_D(I_*)$ is not even consistent (see also the bound given by [25] for trapezoidal approximation of periodogram averages). In $Q_{D2}(\theta; I)$ we have $c_{n-1} = x_1 x_n / n = O_p(n^{-1})$ instead of $c_{n-1}^* = x_1 x_n$, so the "aliasing" of lags causes no asymptotic problem, as demonstrated by [13] in case $d = 1$. These observations may explain the large numerical discrepancy between $\hat{\theta}_C(I_*)$ and $\hat{\theta}_D(I_*)$ found by [30].

In [48] the edge effect was handled in a Gaussian pseudo-likelihood by trimming out observations near the edges, thereby retaining non-negativity of the objective function. Dahlhaus and Künsch [6] proposed an estimate $\hat{\theta}_C(I_T)$, where I_T is the periodogram of tapered x_t, so I_T and $Q_C(\theta; I_T)$ (plus a quantity independent of θ) are always non-negative; for $d \leqslant 3$ and the n_i increasing at the same rate, $\hat{\theta}_C(I_T)$ is $n^{\frac{1}{2}}$-consistent and asymptotically normal, and fully satisfies Property E when a bandwidth number is suitably chosen.

4. Estimates with Property E

We propose an estimate of θ_0 that enjoys some computational advantages of discrete-frequency Whittle and achieves Property E, without tapering (but with a form of bandwidth), in a quite general class of processes that includes the ARMA class and ones in which autocorrelation falls off much more slowly. Arbitrarily large d are covered, with arbitrary relative rates of increase of the n_i subject to A1. We introduce first a truncated version of $I_*(\lambda)$,

$$I_g(\lambda) = (2\pi)^{-d} \sum_{|j_i| \leqslant g(n_i), i=1,\ldots,d} \cdots \sum c_j^* e^{-ij.\lambda},$$

where $g(x)$ satisfies assumption

A3. $g(x)$ is a positive, integer-valued, monotonically increasing function such that

$$g(x) \to \infty \ \text{as} \ x \to \infty,$$

and for all sufficiently large positive x

$$g(x) \leqslant C_2 x, \quad \text{some } C_2 < 1.$$

Remark 5. Due to the latter property of g, the $I_g(\omega_j)$, when averaged over j, are immune to the aliasing problems affecting the $I_*(\omega_j)$. The truncation also has effects that are negligible asymptotically but may be significant in finite samples, where it is a source of bias, but also reduces variance due to the c_j^* for large j. There is sensitivity to choice of g, though an overall sample size n that justifies large sample inference in a given parametric model might entail individual n_i that are not very large, in which case the number of candidate integers $g(n_i)$ may not be great. A simple choice is $g(x) = [x/2]$. The aliasing can alternatively be avoided without truncating but instead evaluating I_* over a finer grid of frequencies, as in [25], but ambiguity is only transferred, the computations are heavier, and no asymptotic efficiency is gained.

Like I_*, I_g is not guaranteed non-negative, so $Q_D(\theta;\, I_g)$ has numerical properties similar to those of $Q_C(\theta;\, I_*)$ criticized in [6] and we do not discuss $\hat{\theta}_D(I_g)$. Theorem 5 of [37] suggests that finitely many Newton iterations, based on $Q_D(\theta;\, I_g)$ and commencing from an n^{ζ}-consistent estimate, for any $\zeta \in (0, \frac{1}{2}]$, will satisfy Property E, building on development by [17] and others of the observation in [27] that a single Newton step can convert an $n^{\frac{1}{2}}$-consistent estimate into an asymptotically efficient one.

Define

$$r(\theta) = \frac{1}{n} \sum_{j \in \mathbb{N}} \partial(\omega_j;\, \theta) \left\{ \frac{I_g(\omega_j)}{f(\omega_j;\, \theta)} - 1 \right\}, \quad R(\theta) = \frac{1}{n} \sum_{j \in \mathbb{N}} \partial(\omega_j;\, \theta) \partial'(\omega_j;\, \theta).$$

We propose two alternative recursions. For $\ell = 1, 2$, given an initial estimate $\hat{\theta}_{[1]}^{(\ell)}$ of θ_0, define

$$\hat{\theta}_{[u+1]}^{(1)} = \hat{\theta}_{[u]}^{(1)} + R\left(\hat{\theta}_{[1]}^{(1)}\right)^{-1} r\left(\hat{\theta}_{[u]}^{(1)}\right), \quad u \geqslant 1, \tag{4.1}$$

$$\hat{\theta}_{[u+1]}^{(2)} = \hat{\theta}_{[u]}^{(2)} + R\left(\hat{\theta}_{[u]}^{(2)}\right)^{-1} r\left(\hat{\theta}_{[u]}^{(2)}\right), \quad u \geqslant 1. \tag{4.2}$$

Thus, $\{\hat{\theta}_{[u]}^{(1)}\}$ entails no updating of the inner product matrix R, though $\hat{\theta}_{[1]}^{(1)} = \hat{\theta}_{[1]}^{(2)}$ implies $\hat{\theta}_{[2]}^{(1)} = \hat{\theta}_{[2]}^{(2)}$. Both sequences approximate solutions to the estimating equations $r(\theta) = 0$, which are first-order conditions for minimizing $Q_D(\theta;\, I_g)$. They are both forms of Gauss–Newton iteration. Newton–Raphson famously numerically converges faster, in a suitable neighbourhood of the target, and [37] showed that this can be matched by faster statistical convergence. Robinson [37] stressed the improvements gained by further iterations on an estimate that already has Property E, in reducing the stochastic order of the difference between the iterated estimate and its target, with possible implications for matching higher-order efficiency. In our case Property E is the goal, the difference between R and the Hessian used in Newton–Raphson is of relatively small order, and Property E would be achieved no faster. Moreover, the Hessian is more complicated to compute than R, and unlike R is not guaranteed non-negative definite, thereby presenting possible convergence problems.

We introduce the following additional assumptions.

A4. For ξ as in A1 and g^{-1} the inverse function of g given in A3, the autocovariance function $\gamma_j = cov(x_0, x_j)$ satisfies

$$\sum_j \left\{ \sum_{i=1}^{d} g^{-1}(|j_i|)^{1/(2\xi)} \right\} |\gamma_j| < \infty.$$

A5. In a neighbourhood of θ_0, $f(\lambda;\theta)$ is positive and thrice boundedly differentiable in θ; $f(\lambda;\theta)$ and its first three derivatives in θ are continuous in λ at $\theta = \theta_0$.

A6. Φ is positive definite.

A7. For $\ell = 1, 2$, $\hat{\theta}^{(\ell)}_{[1]} = \theta_0 + O_p(n^{-\zeta})$, for some $\zeta \in (0, \frac{1}{2})$.

Remark 6. Assumption A4 controls the bias. For ARMA models (2.3), $f(\lambda)$ is analytic so the γ_j decay exponentially; thus A4 holds for any $\xi > 0$ and for $g(x) \sim x^\rho$, any $\rho > 0$, allowing heavy truncation in I_g. Again in an ARMA context, A5 relies on smoothness of the functions $a_j(\theta)$, $b_j(\theta)$, while the standard identifiability condition A6 rules out common roots in $a(z;\theta_0)$ and $b(z;\theta_0)$. We postpone discussion of A7 until Section 5.

Theorem 1. *Under Assumptions A1–A7*:

(i) $\hat{\theta}^{(1)}_{[u]}$ *satisfies Property E for all*

$$u > (2\zeta)^{-1}; \tag{4.3}$$

(ii) $\hat{\theta}^{(2)}_{[u]}$ *satisfies Property E for all*

$$u > \frac{\ell n(\zeta)}{\ell n(\frac{1}{2})}. \tag{4.4}$$

The proof is left to Section 8. It follows from the inequality $x^x > (\frac{1}{2})^{\frac{1}{2}}$ for $0 < x < \frac{1}{2}$ that (4.1) requires at least as many iterations as (4.2), indicating a benefit of updating R in (4.2).

Remark 7. Küveri [25] established Property E for an estimate minimizing a discretized form of $Q_C(\theta; I_*)$ that uses a finer grid than in $Q_D(\theta; I_*)$, assuming x_t is Gaussian and $d = 2$; [25] also considered Newton iteration but from a purely numerical perspective, not discussing the choice of initial value or showing achievement of Property E after finitely many steps.

Remark 8. Our methods and theory can be extended relatively straightforwardly to multivariate x_t, with the proviso that the identifiability problem for multivariate versions of the ARMA models of Section 2 will be more acute; cf. [7] in case $d = 1$.

5. Initial Estimates

The $\hat{\theta}_{[1]}^{(\ell)}$ are likely to be implicitly-defined extremum estimates that do not attempt edge-effect correction. A promising candidate on computational grounds is $\hat{\theta}_D(I)$, which has the desired minimum-distance interpretation, minimizing the objective function $Q_D(\theta; I) - n^{-1}\sum_{j\in\mathbb{N}} \log I(\omega_j) - 1$, which is always non-negative and vanishes only when $I(\omega_j) = f(\omega_j; \theta)$ for all $j \in \mathbb{N}$. In the AR case of (2.3) with $a(z; \theta)$ linear in θ, $Q_D(\theta; I)$ is globally convex for all finite n, so that hill-climbing procedures commencing from any starting value will always converge. To indicate how A7 is satisfied, we introduce the following additional assumptions.

A8. Θ is a compact subset of R^m.
A9. θ_0 is an interior point of Θ.
A10. $f(\lambda; \theta) \neq f(\lambda; \theta_0)$, $\theta \in \Theta - \{\theta_0\}$, for all λ in a subset of Π^d of positive measure.
A11. $\sum_j \left(\sum_{i=1}^{d} |j_i|\right) |\gamma_j| < \infty$.

Theorem 2. *Under Assumptions A1, A2, A5, A6 and A8–A11,*

$$\hat{\theta}_D(I) - \theta_0 = O_p(n^{-\xi}), \text{ as } n \to \infty. \tag{5.1}$$

The rate in (5.1) was anticipated in Section 3, but in view of A7 formal justification seems desirable, especially as we later discuss a modified estimate. Theorem 2 relates to results in [9,13,24], so we only comment briefly on the proof. Consistency, with no rate, may be established much as in [13], using A2, A5 and A8–A10. Using A5, A6, the mean value theorem is then applied to the first-order conditions for a minimum of $Q_D(\theta; I)$, around θ_0, as if a central limit theorem is to be proved, but $(\partial/\partial\theta)Q_D(\theta_0; I)$ is then seen to take the order of its expectation, $n^{-\xi}$ (applying A11 and (8.17) of Section 8). A11 is milder than A4, and could be relaxed at cost of a slower rate than in (5.1) and possible increase in the number of recursions needed to achieve Property E. A8 is nearly costless for ARMA models in view of stationarity and invertibility requirements, and need not apply to scale estimation, as a simple elimination indicates.

When the n_i increase at the same rate, we have $\xi = 1/d$, and Table 1 indicates the minimal values of u, $u(1)$ and $u(2)$, satisfying (4.3) and (4.4) when $\hat{\theta}_{[1]}^{(\ell)} = \hat{\theta}_D(I)$ for $\ell = 1, 2$. For the practically most typical d, $\hat{\theta}_{[u]}^{(1)}$ dominates on computational grounds. If the n_i increase at varying speeds, $\xi < 1/d$ so for $\zeta = \xi$ the $u(\ell)$, and the gap between them, can increase.

Table 1
Minimum values $u(\ell)$, $\ell = 1, 2$, of u satisfying (4.3) and (4.4) when $\zeta = 1/d$.

d :	2	3	4	5	6	7	8	9	10
$u(1)$	2	2	3	3	4	4	5	5	6
$u(2)$	2	2	3	3	3	3	4	4	4

Since $\hat{\theta}_D(I)$ is real-valued and only implicitly-defined, strictly speaking it cannot be obtained by finite computation. In practice one is content with accuracy to a given number of decimal places and such a solution can be reached, using numerical search of $Q_D(\theta; I)$, possibly combined with iteration, but even this can be expensive, especially when m is large. From our statistical perspective we want only to satisfy A7, which does not necessarily require a search that is exhaustive but rather one over a grid that becomes suitably finer as n increases. Robinson [37] showed, for a quite general objective function with an $n^{\frac{1}{2}}$-consistent optimizer, that of order $n^{m\psi}$ search points suffice to achieve an n^{ψ}-consistent estimate, $\psi \leqslant \frac{1}{4}$. To correspondingly approximate $\hat{\theta}_D(I)$, define by G_n a set of points that is regularly-spaced throughout Θ, and such that $\#\{\theta : \theta \in G_n\} \geqslant C_3 n^{m\psi}$, $C_3 > 0$, and denote

$$\hat{\theta}_D^{(s)}(I) = \arg\min_{\theta \in G_n} Q_D(\theta; I).$$

Theorem 3. *Under Assumptions A1, A2, A5, A6 and A8–A11,*

$$\hat{\theta}_D^{(s)}(I) - \theta_0 = O_p(n^{-\psi}), \quad \text{as } n \to \infty, \tag{5.2}$$

for $\psi \leqslant \xi/2$.

We omit the proof because it largely applies Theorem 8 of [37], whose conditions are checkable much as would be done in proving Theorem 2; [37] requires that $\sup_{\Theta} |Q_D(\theta; I) - Q(\theta)| = O_p(n^{-\zeta})$ for $\zeta = \frac{1}{2}$, where $Q(\theta)$ is the probability limit of $Q_D(\theta; I)$, whereas only $\zeta = \xi$ is possible, explaining the weaker result (5.2) that emerges by following his method of proof.

The strategy justified in Theorems 1 and 3 stresses statistical and computational considerations to achieve Property E in a finite, relatively well-defined, number of simple steps. However, a comprehensive search of $Q_D(\theta; I)$, guided by advice from numerical analysis, and iterating (4.1) or (4.2) to achieve numerical convergence, would obviously be desirable.

6. Monte Carlo Study of Finite-Sample Performance

A Monte Carlo study was carried out to study the finite-sample performance of our estimates. We first consider the simple symmetric multilateral model

$$x_t = \sigma_0 \varepsilon_t + \rho_0 \sigma_0 \sum_{\substack{j_1=-1 \\ j \neq (0,\ldots,0)}}^{1} \cdots \sum_{j_d=-1}^{1} \varepsilon_{t-j}. \tag{6.1}$$

This is an MA $(1, 1; \ldots; 1, 1)$ representation defined as in Section 2 with $a(z; \theta) \equiv 1$, $b_j(\theta) = \sigma$ for $j = (0, \ldots, 0)$, $b_j(\theta) = \sigma\rho$ for $j = (\pm 1, \ldots, \pm 1)$, and $b_j(\theta) \equiv 0$ otherwise, taking $\theta = (\rho, \theta)'$. [11] discussed a similar model. We deduce that

$$f(\lambda; \theta) = \frac{\sigma^2}{(2\pi)^d} \left\{1 + \rho v_d(\lambda_1, \ldots, \lambda_d)\right\}^2,$$

where

$$v_d(\lambda_1, ..., \lambda_d) = \prod_{j=1}^{d}(1 + 2\cos\lambda_j) - 1.$$

An "invertibility" condition satisfying (2.4) is

$$|\rho_0| < \left(3^d - 1\right)^{-1}. \tag{6.2}$$

For given n^*, we generated NID(0, 1) ε_t for $t_\ell = 0, \pm1, ..., \pm(n^* + 1)$, $\ell = 1, ..., d$, and then x_t $t \in \mathbb{N} = \{t : t_\ell = 0, \pm1, ..., \pm n^*, \ell = 1, ..., d\}$, using (6.1). Thus we study only the regular case $n_{Li} = n_{Ui} = n^*$, $i = 1, ..., d$, with $n = (2n^* + 1)^d$.

The experiment was carried out for $d = 2$ and 3, with the following specifications:

$$\begin{aligned} d &= 2 : \rho_0 = 0.05, 0.1;\ \ \sigma_0 = 1;\ (n, g) = (121, 2), (121, 5), (361, 4), (361, 9), \\ d &= 3 : \rho_0 = 0.015, 0.03;\ \sigma_0 = 1;\ (n, g) = (125, 1), (125, 2), (343, 1), (343, 3), \end{aligned}$$

where $g = g(n_i) = g(2n^* + 1)$. The g's were determined by the rules $g = [n^*/2]$ and $g = [n^*]$, noting that $n^* = 5, 9$ for $d = 2$ and $n^* = 2, 3$ for $d = 3$. The n^* were chosen so as to make n relatively stable across d. Note that (6.2) is satisfied.

The initial estimate $\hat{\theta}_{[1]} = \hat{\theta}^{(1)}_{[1]} = \hat{\theta}^{(2)}_{[1]}$ was computed according to the scheme justified in Theorem 3. Notice that our parameterization allows σ to be eliminated, leaving an objective function

$$M(\rho) = \log\hat{\sigma}^2(\rho) + \frac{2}{n}\sum_{j\in\mathbb{N}}\log\left\{1 + \rho v_d(\omega_j)\right\},$$

where

$$\hat{\sigma}^2(\rho) = \frac{(2\pi)^d}{n}\sum_{j\in\mathbb{N}}\frac{I(\omega_j)}{\left\{1 + \rho v_d(\omega_j)\right\}^2}.$$

We took $\hat{\theta}_{[1]} = \left(\hat{\rho}_{[1]}, \hat{\sigma}^2(\hat{\rho}_{[1]})\right)'$, where $\hat{\rho}_{[1]}$ minimizes $M(\rho)$ over a set $G_n^{(d)}$, such that

$$\begin{aligned} G_n^{(2)} &= \left\{r : r = \frac{j}{16n^{\frac{1}{4}}},\ j = 0, \pm1, ...;\ \ |r| < 1/8\right\}, \\ G_n^{(3)} &= \left\{r : r = \frac{j}{52n^{1/6}},\ j = 0, \pm1, ...;\ \ |r| < 1/26\right\}, \end{aligned}$$

indicating equally-spaced points over the set (6.2). Thus $G_n^{(2)}$ contains about $4n^{\frac{1}{4}}$ points, and $G_n^{(3)}$ about $4n^{1/6}$. Notice that G_n of Theorem 3 contains of order $n^{1/d}$ points on the basis of $m = 2$ and $\xi = 1/d$, since it was assumed there that an m-dimensional search is carried out. Due to the elimination of σ we can get the $n^{1/(2d)}$ -consistency of $\hat{\theta}_D^{(s)}(I)$ in the statement of Theorem 3 by searching over $G_n^{(d)}$.

Table 2
Monte Carlo bias (standard deviation) with $d = 2$, $\rho = 0.05$

n, g	121, 2	121, 5	361, 4	361, 9
$\hat{\rho}_{[1]}$	-.0081 (.0275)	-.0081 (.0275)	-.0046 (.0147)	-.0046 (.0147)
$\hat{\rho}^{(1)}_{[3]}$	-.0065 (.0291)	-.0046 (.0280)	-.0032 (.0145)	-.0028 (.0145)
$\hat{\rho}^{(2)}_{[3]}$	-.0064 (.0290)	-.0046 (.0279)	-.0032 (.0145)	-.0027 (.0145)

Table 3
Monte Carlo bias (standard deviation) with $d = 2$, $\rho = 0.10$

n, g	121, 2	121, 5	361, 4	361, 9
$\hat{\rho}_{[1]}$	-.0184 (.0265)	-.0184 (.0277)	-.0097 (.0148)	-.0047 (.0148)
$\hat{\rho}^{(1)}_{[3]}$	-.0083 (.0331)	-.0088 (.0277)	-.0064 (.0144)	-.0058 (.0145)
$\hat{\rho}^{(2)}_{[3]}$	-.0087 (.0324)	-.0089 (.0276)	-.0064 (.0144)	-.0058 (.0145)

Table 4
Monte Carlo bias (standard deviation) with $d = 3$, $\rho = 0.015$

n, g	125, 1	125, 2	343, 1	343, 3
$\hat{\rho}_{[1]}$	-.0053 (.0125)	-.0053 (.0125)	-.0044 (.0091)	-.0044 (.0091)
$\hat{\rho}^{(1)}_{[4]}$	-.0038 (.0168)	.0023 (.0197)	-.0015 (.0113)	.0000 (.0113)
$\hat{\rho}^{(2)}_{[3]}$	-.0040 (.0165)	-.0020 (.0197)	-.0015 (.011)	-.0002 (.0110)

Both sequences of iterations (4.1) and (4.2) were pursued. Property E is first achieved by $\hat{\rho}^{(1)}_{[3]}$ and $\hat{\rho}^{(2)}_{[3]}$ for $d = 2$, and by $\hat{\rho}^{(1)}_{[4]}$ and $\hat{\rho}^{(2)}_{[3]}$ for $d = 3$. We report Monte Carlo bias and standard deviation, on the basis of 100 replications, for $d = 2$ with $\rho = 0.05$ in Table 2, $d = 2$ with $\rho = 0.01$ in Table 3, $d = 3$ with $\rho = 0.015$ in Table 4, and $d = 3$ with $\rho = 0.03$ in Table 5. A constant feature is that the outcomes of iterations (4.1) and (4.2) were almost identical, which is in line with the theory since both employ the minimum number of iterations necessary to achieve Property E. Biases are predominantly negative. The bias-reductions achieved in Table 2 are not great though the bias of $\hat{\rho}_{[1]}$ is about 16% of ρ when $n = 121$, and nearly 10% when $n = 361$, and the percentage reductions are about 20% and 30% respectively. These are greater in Table 3, more than halving the bias in case of the smaller sample size. As feared, the iterations produce overall a worsening in standard deviation (though there is a slight improvement for $d = 2$ and $n = 361$). For $d = 2$ and $n = 121$ the smaller g does worst, for $d = 3$ and $n = 125$ it does best; though we expect to reduce variability by omitting long lags from the periodogram, it could be increased by also omitting short ones. As expected, biases were mostly smaller for the larger g. Notice the enormous percentage bias reductions achieved by (4.1) and (4.2) when $d = 3$ and $n = 343$.

Table 5
Monte Carlo bias (standard deviation) with $d = 3$, $\rho = 0.03$

n, g	125, 1	125, 2	343, 1	343, 3
$\hat{\rho}_{[1]}$	-.0115 (.0121)	-.0015 (.0121)	-.0089 (.0091)	-.0089 (.0091)
$\hat{\rho}^{(1)}_{[4]}$	-.0038 (.0224)	.0051 (.0314)	-.0001 (.0151)	.0006 (.0132)
$\hat{\rho}^{(2)}_{[3]}$	-.0048 (.0202)	.0017 (.0214)	.0006 (.0179)	-.0000 (.0123)

Table 6
Monte Carlo bias (standard deviation) with $d = 4$, $\rho = 0.015$

n, g	625, 1	625, 2	2401, 1	2401, 3
$\hat{\rho}_{[1]}$	-.0067 (.0094)	-.0067 (.0094)	-.0050 (.0050)	-.0050 (.0050)
$\hat{\rho}^{(1)}_{[5]}$	.0022 (.0104)	.0044 (.0129)	.0005 (.0066)	.0006 (.0060)
$\hat{\rho}^{(2)}_{[4]}$	.0024 (.0108)	.0042 (.0123)	.0005 (.0066)	.0006 (.0060)

Table 7
Monte Carlo bias (standard deviation) with $d = 4$, $\rho = 0.03$

n, g	625, 1	625, 2	2401, 1	2401, 3
$\hat{\rho}_{[1]}$	-.0150 (.0090)	-.0150 (.0090)	-.0123 (.0048)	-.0123 (.0048)
$\hat{\rho}^{(1)}_{[5]}$	-.0024 (.0125)	.0020 (.0155)	.0010 (.0072)	.0004 (.0072)
$\hat{\rho}^{(2)}_{[4]}$	-.0031 (.0128)	.0028 (.0167)	.0011 (.0075)	.0005 (.0071)

The spatio-temporal model with $d = 4$,

$$x_t = \sigma_0 \varepsilon_t + \rho_0 \sigma_0 \underset{(j_1, j_2, j_3) \neq (0,0,0)}{\sum_{j_1=-1}^{1} \sum_{j_2=-1}^{1} \sum_{j_3=-1}^{1}} \varepsilon_{t_1 - j_1, t_2 - j_2, t_3 - j_3, t_4 - 1},$$

was also simulated. This is unilateral with respect to the fourth, "time" dimension, and

$$f(\lambda; \theta) = \frac{\sigma^2}{(2\pi)^4} \left\{ 1 + \rho^2 v_3(\lambda_1, \lambda_2, \lambda_3) + 2\rho v_3(\lambda_1, \lambda_2, \lambda_3) \cos \lambda_4 \right\}.$$

We took $\sigma_0^2 = 1$ $\rho_0 = 0.015, 0.03$ and $(n, g) = (625, 1), (625, 2), (2401, 1), (2401, 3)$, the n resulting from $n^* = 2$ and 3. Tables 6 and 7 mostly reveal little difference between the outcomes of (4.1) and (4.2). Both recursions definitely worsen standard deviation, but there are substantial absolute bias reductions, which seem especially welcome as $\hat{\rho}_{[1]}$ exhibits biases between -$\rho/3$ and -$\rho/2$; the recursions also mostly reverse the sign of the bias.

7. Variance Matrix Estimation

When x_t is Gaussian, estimates satisfying Property E are asymptotically efficient, and have limiting variance matrix $2\Phi^{-1}$, so Theorem 1 can be applied in approximate inference on θ_0 by consistently estimating Φ by $\hat{\Phi} = R(\hat{\theta})$, where $\hat{\theta}$ is any consistent estimate of θ_0. More generally, if we can partition θ in the ratio $m_a : m_b$ as $\theta = (\theta_a', \theta_b')'$, and correspondingly $\partial(\lambda; \theta) = \left(\partial_a(\lambda; \theta)', \partial_b(\lambda; \theta)'\right)'$, such that $\int_{\Pi^d} \partial_a(\lambda; \theta_0) d\lambda = 0$ and $\partial_b(\lambda; \theta_0)$ is constant, then the leading $m_a \times m_a$ sub-matrix of $\Phi^{-1}\Psi\Phi^{-1}$ is twice the inverse of the leading $m_a \times m_a$ sub-matrix of Φ (which is block-diagonal), irrespective of whether or not $\kappa = 0$. Such circumstances occur in standard unilateral parameterizations of ARMA models, where $m_b = 1$ and $(2\pi)^{-d} \int_{\Pi^d} \log f(\lambda; \theta) d\lambda = \log \theta_b$, say, but not in non-standard parameterizations, such as signal-plus-noise and multilateral models, as the discussion of $Q_{C1}(\theta)$ in Section 3 suggests. Here, asymptotic inference requires consistently estimating Ψ, for which several approaches have been suggested in case $d = 1$.

For unilateral models, [14] proposed a consistent estimate of Ψ, involving time-domain filtering, that is advantageously guaranteed to be non-negative definite (n.n.d.), but seems difficult to extend to multilateral spatial models. The frequency-domain proposal in [41], for estimating $\int_{\Pi^2} \rho(\lambda, \chi) f_4(\lambda, \chi, -\chi)\, d\lambda\, d\chi$, where f_4 is the fourth cumulant spectral density of x_t, and ρ is a continuous function on Π^2, does seem to be extendable, indeed it does not assume linearity of x_t so it affords some robustness. However, it is somewhat complicated, it requires choice of a kernel function and bandwidth, and the resulting estimate of Ψ does not seem to be necessarily n.n.d. Chiu [3] proposed that $n^{-2} \sum_{j\in\mathbb{N}} \sum_{k\in\mathbb{N}} \rho(\omega_j)\rho(\omega_k) I(\omega_j) I(\omega_k)$, with ρ now a continuous function on Π, consistently estimates something with an additive component $(2\pi)^{-1} \int_{\Pi^2} \rho(\lambda)\rho(\chi) f_4(\lambda, -\lambda, \chi)$ $d\lambda\, d\chi$, the others being functionals of f and easily estimable. However, this estimate is actually uninformative about f_4; it equals $\left\{n^{-1} \sum_{j\in\mathbb{N}} \rho(\omega_j) I(\omega_j)\right\}^2 \to_p \{(2\pi)^{-1} \int_{\Pi} \rho(\lambda)$ $f(\lambda) d\lambda\}^2$.

A simple estimate of Ψ that is clearly n.n.d. is

$$\frac{1}{n} \sum_{j\in\mathbb{N}} \partial(\omega_j; \hat{\theta}) \partial'\left(\omega_j; \hat{\theta}\right) \left\{\frac{I(\omega_j)}{f(\omega_j; \hat{\theta})} - 1\right\}^2. \tag{7.1}$$

Consistency is anticipated due to the approximate independence, across the ω_j, of the factor in braces in (7.1). Eq. (7.1) can advantageously still be consistent when Ψ lacks the simple structure in (1.3) which is due to the linearity in A2; for example under α-mixing, which would require a moment condition of order greater than 4. We study in more detail an estimate which exploits linearity, seems new even in case $d = 1$, and applies also to long range dependent processes.

Since Φ is consistently estimated by $\hat{\Phi}$, and Ξ by $\hat{\Xi} = n^{-1} \sum_{j\in\mathbb{N}} \partial(\omega_j; \hat{\theta})$, it suffices to estimate κ. Given $\hat{\varepsilon}_t$, $t \in \mathbb{N}$, introduce

$$\hat{\mu}_2 = n^{-1} \sum_{t\in\mathbb{N}} \hat{\varepsilon}_t^2, \quad \hat{\mu}_4 = n^{-1} \sum_{t\in\mathbb{N}} \hat{\varepsilon}_t^4. \tag{7.2}$$

An obvious estimate of κ is $\tilde{\kappa} = \hat{\mu}_4 - 3$, but $2\hat{\Phi} + \tilde{\kappa}\hat{\Xi}\hat{\Xi}'$ is not necessarily n.n.d. However, since $2\left(\hat{\Phi} - \hat{\Xi}\hat{\Xi}'\right)$ and $\left(\hat{\mu}_4 - \hat{\mu}_2^2\right)\hat{\Xi}\hat{\Xi}'$ are both n.n.d., so is their sum $2\hat{\Phi} + \left(\hat{\mu}_4 - \hat{\mu}_2^2 - 2\right)$ $\hat{\Xi}\hat{\Xi}'$, which is also consistent for Ψ if $\hat{\mu}_2$ and $\hat{\mu}_4$ are consistent for $E\varepsilon_t^2$ and $E\varepsilon_t^4$ (explaining the introduction of $\hat{\mu}_2$ despite $E\varepsilon_t^2 = 1$ being given). It remains to obtain $\hat{\varepsilon}_t$ that achieve this property.

For finite AR models, this is straightforward. Define

$$\hat{\varepsilon}_t^{(1)} = a\left(B; \hat{\theta}\right)(x_t - \bar{x}), \quad t \in \mathbb{N},$$

with a given by (2.1) and x_s replaced by $\bar{x}$ when $s \notin \mathbb{N}$. Other models, in particular multilateral MA and ARMA ones, may be difficult to invert, and require proxies for x_s for all $s \notin \mathbb{N}$. For such models we develop an approach of [36] (intended for unilateral models with $d = 1$) which assumes knowledge of $\alpha(z; \theta)$ of z and θ such that $f(\lambda; \theta) = (2\pi)^{-d}\left|\alpha(e(i\lambda); \theta)\right|^{-2}$; for example in the ARMA model (2.3), $\alpha(z; \theta) = a(z; \theta)/b(z; \theta)$. Define $w(\lambda) = \{(2\pi)^d n\}^{-\frac{1}{2}} \sum_{t\in\mathbb{N}} x_t e^{it.\lambda}$ and

$$\hat{\varepsilon}_t^{(2)} = (2\pi)^{d/2} n^{-\frac{1}{2}} \sum_{j\in\mathbb{N}} \alpha\left(e(i\omega_j); \hat{\theta}\right) w(\omega_j) e^{-it\cdot\omega_j}, \quad t \in \mathbb{N}. \tag{7.3}$$

When expressed in the time domain, (7.3) effectively treats x_t on $\mathbb{Z}^d$ as a circulant, with observations on $\mathbb{N}$ repeated periodically; we show that, as with $\hat{\theta}_D(I)$, the consequent error is asymptotically negligible, and (7.3) is computationally advantageous when α is a simple function, as in ARMA models, and in making double use of the fast Fourier transform. Robinson [36] studied convergence of $\hat{\varepsilon}_t^{(1)}, \hat{\varepsilon}_t^{(2)}$ and their use in kernel probability density estimation (in the unilateral $d = 1$ case) but did not employ them in estimating moments.

We introduce the following assumptions.

A12. For all $\lambda \in \Pi^d$, $\alpha(e(i\lambda); \theta)$ is boundedly differentiable in a neighbourhood of θ_0, it is non-zero and has absolutely convergent Fourier series at $\theta = \theta_0$, and x_t has representation

$$\alpha(B; \theta_0)(x_t - \mu) = \varepsilon_t, \quad t \in \mathbb{Z}^d,$$

where the ε_t are independent and identically distributed with zero mean, unit variance and finite fourth moment $\dot{\mu}_4$.

A13. $\hat{\theta} = \theta_0 + O_p(n^{-\zeta})$ for $\zeta > \frac{1}{4}$.

Remark 9. A12 implies knowledge of a factorization of $f(\lambda; \theta)$, but it entails no strengthening of the fourth moment condition in A2, and holds for stationary and invertible ARMA processes with coefficients that are smooth in θ, as well as for many processes with long-range dependence; there, the summability of β_j assumed in A2 will not hold, but square summability does, as under A12, while in long-range-dependent models AR weights are typically absolutely convergent. It would be possible to still cover ARMA processes by strengthening A12 but relaxing A13 to only consistency of $\hat{\theta}$. However, in the context of estimating $\Phi^{-1}\Psi\Phi^{-1}$, we already have an $n^{\frac{1}{2}}$-consistent estimate of θ_0, though the $\hat{\theta}_{[1]}^{(\ell)}$ in A7 also satisfy A13 if $\zeta = 1/d$ for $d \leqslant 3$.

The following theorem is proved in Section 9.

Theorem 4. *Let Assumptions* A12 *and* A13 *hold. Then with $\alpha(z;\theta) = a(z;\theta)$ for $i = 1$, as $n \to \infty$*

$$\hat{\mu}_2^{(i)} \to_p E\varepsilon_0^2, \quad \hat{\mu}_4^{(i)} \to_p E\varepsilon_0^4, \quad i = 1, 2. \tag{7.4}$$

If, further, Assumptions A1, A2, A5 *and* A6 *hold*,

$$2\Phi^{-1} + \left(\hat{\mu}_4^{(i)} - \hat{\mu}_2^{(i)2} - 2\right)\left(\hat{\Phi}^{-1}\hat{\Xi}\right)\left(\hat{\Phi}^{-1}\hat{\Xi}\right)', \quad i = 1, 2, \tag{7.5}$$

are non-negative definite and as $n \to \infty$ converge in probability to $\Phi^{-1}\Psi\Phi^{-1}$.

8. Proof of Theorem 1

Introduce the artificial estimate

$$\hat{\theta} = \theta_0 + R(\theta_0)^{-1} r(\theta_0).$$

It suffices to show that $\hat{\theta}$ has Property E and

$$\hat{\theta}_{[u]}^{(\ell)} - \hat{\theta} = o_p(n^{-\frac{1}{2}}), \quad \ell = 1, 2, \tag{8.1}$$

when u satisfies (4.3) for $\ell = 1$ and (4.4) for $\ell = 2$.

The first statement follows on showing

$$n^{\frac{1}{2}} r(\theta_0) \to_d \mathcal{N}(0, \Psi) \tag{8.2}$$

and

$$R(\theta_0) \to_p \Phi. \tag{8.3}$$

With respect to the second write, with $\tilde{\theta}_{[u]}^{(1)} = \hat{\theta}_{[1]}^{(1)}$, $\tilde{\theta}_{[u]}^{(2)} = \hat{\theta}_{[u]}^{(2)}$,

$$\begin{aligned}\hat{\theta}_{[u+1]}^{(\ell)} - \hat{\theta} &= \hat{\theta}_{[u]}^{(\ell)} - \theta_0 + R\left(\tilde{\theta}_{[u]}^{(\ell)}\right)^{-1} r\left(\hat{\theta}_{[u]}^{(\ell)}\right) - R(\theta_0)^{-1} r(\theta_0) \\ &= \left\{R\left(\tilde{\theta}_{[u]}^{(\ell)}\right)^{-1} - R(\theta_0)^{-1}\right\} r(\theta_0) + \left\{I_m + R\left(\tilde{\theta}_{[u]}^{(\ell)}\right)^{-1} \tilde{S}_{[u]}^{(\ell)}\right\}\left(\hat{\theta}_{[u]}^{(\ell)} - \theta_0\right),\end{aligned}$$

where I_m is the m-rowed identity matrix and $\tilde{S}^{(\ell)}_{[u]}$ is the matrix obtained by evaluating each row of $S(\theta) = \left(\partial/\partial\theta'\right) r(\theta)$ at a point on the line segment between $\hat{\theta}^{(\ell)}_{[u]}$ and θ_0. On showing

$$\left\| R\left(\tilde{\theta}^{(\ell)}_{[u]}\right)^{-1} - R(\theta_0)^{-1} \right\| = O_p\left(\left\|\tilde{\theta}^{(\ell)}_{[u]} - \theta_0\right\|\right), \tag{8.4}$$

$$\left\| I_m + R\left(\tilde{\theta}^{(\ell)}_{[u]}\right) \tilde{S}^{(\ell)}_{[u]} \right\| = O_p\left(\left\|\tilde{\theta}^{(\ell)}_{[u]} - \theta_0\right\| + n^{-\frac{1}{2}}\right), \tag{8.5}$$

where $\|A\| = \left\{tr(AA')\right\}^{\frac{1}{2}}$ for any matrix A, we deduce

$$\hat{\theta}^{(\ell)}_{[u+1]} - \hat{\theta} = O_p\left(\left(n^{-\frac{1}{2}} + \left\|\hat{\theta}^{(\ell)}_{[u]} - \theta_0\right\|\right)\left\|\tilde{\theta}^{(\ell)}_{[u]} - \theta_0\right\|\right).$$

As in [37] we have the solutions

$$\hat{\theta}^{(1)}_{[u+1]} - \hat{\theta} = O_p\left(\left\|\hat{\theta}^{(1)}_{[1]} - \theta_0\right\|^{u+1}\right) + o_p(n^{-\frac{1}{2}}) = O_p\left(n^{-(u+1)\xi}\right) + o_p(n^{-\frac{1}{2}}),$$

$$\hat{\theta}^{(2)}_{[u+1]} - \hat{\theta} = O_p\left(\left\|\hat{\theta}^{(2)}_{[1]} - \theta_0\right\|^{2^u}\right) + o_p(n^{-\frac{1}{2}}) = O_p\left(n^{-2^u\xi}\right) + o_p(n^{-\frac{1}{2}}),$$

whence (8.1) holds under (4.3) and (4.4) respectively.

The proof of (8.4) involves standard application of the mean value theorem, given A5, A6 and (8.3), which follows immediately from continuity of $\partial(\lambda;\theta_0)$. The proof of (8.5) uses similar arguments, the fact that

$$I_m + R(\theta)^{-1}S(\theta) = I_m - R(\theta)^{-1}n^{-1}\sum_{j\in\mathbb{N}} \partial(\omega_j;\theta)\partial'(\omega_j;\theta)\frac{I_g(\omega_j)}{f(\omega_j;\theta)}$$
$$+R(\theta)^{-1}n^{-1}\sum_{j\in\mathbb{N}}\frac{\partial^2\log f(\omega_j;\theta)}{\partial\theta\partial\theta'}\left\{\frac{I_g(\omega_j)}{f(\omega_j;\theta)} - 1\right\},$$

and arguments employed in the proof of (8.2), which we now consider.

Write $\tau(\lambda) = \partial(\lambda;\theta_0)/f(\lambda)$ and then $r(\theta_0) = r_1 + r_2$, where

$$r_1 = n^{-1}\sum_{j\in\mathbb{N}}\tau(\omega_j)\left\{I_g(\omega_j) - EI_g(\omega_j)\right\},$$
$$r_2 = n^{-1}\sum_{j\in\mathbb{N}}\tau(\omega_j)\left\{EI_g(\omega_j) - f(\omega_j)\right\}.$$

For brevity of proof assume $\mu = 0$ and replace $x_t - \bar{x}$ by x_t; it is straightforward to show that this has negligible effect, $\bar{x}$ being $n^{\frac{1}{2}}$-consistent for μ under A2. Now

$$EI_g(\lambda) - f(\lambda) = (2\pi)^{-d}\sum_{j:|j_i|>g(n_i),\ \text{some } i}\cdots\sum\gamma_j\cos(j.\lambda).$$

This is bounded by

$$K\sum_{i=1}^{d}\sum_{|j_i|>g(n_i)}\sum_{|j_k|<\infty,k\neq i}\cdots\sum\left|\gamma_j\right| \leqslant K\sum_{i=1}^{d}n_i^{-1/(2\xi)}\sum_{|j_i|>g(n_i)}g^{-1}(|j_i|)^{1/(2\xi)}$$
$$\times\sum_{|j_k|<\infty,k\neq i}\cdots\sum\left|\gamma_j\right| = o(n^{-\frac{1}{2}})$$

under A1 and A4, K being a generic, positive constant. Thence $r_2 = o(n^{-\frac{1}{2}})$ and it suffices to establish (8.2) with $r(\theta_0)$ replaced by r_1.

Introduce the Cesaro sum of the multiple Fourier series of $\tau(\lambda)$,

$$\tau_L(\lambda) = \sum_{\ell\in A_L}\prod_{i=1}^{d}\left(1-\frac{|\ell_i|}{L}\right)\tau_\ell e^{-i\ell.\lambda},$$

for $\ell = (\ell_1, ..., \ell_d)$, $A_L = \{\ell : |\ell_i| \leqslant L, i = 1, ..., d\}$ and

$$\tau_\ell = (2\pi)^{-d}\int_{\Pi^d}\tau(\lambda)e^{i\ell.\lambda}d\lambda.$$

Fix $\eta_1 > 0$. By continuity of $\tau(\lambda)$ we can choose L such that

$$\sup_{\lambda}|\tau(\lambda)-\tau_L(\lambda)| < \eta_1. \tag{8.6}$$

Writing

$$r_{1L} = n^{-1}\sum_{j\in\mathbb{N}}\tau_L(\omega_j)\left\{I_g(\omega_j)-EI_g(\omega_j)\right\},$$

$r_1 - r_{1L}$ has mean zero and variance

$$n^{-2}\sum_{j\in\mathbb{N}}\sum_{k\in\mathbb{N}}\tilde{\tau}_L(\omega_j)\tilde{\tau}_L(\omega_k)\text{cov}\left\{I_g(\omega_j), I_g(\omega_k)\right\}$$

$$= \{(2\pi)^d n\}^{-2}\sum_{j\in\mathbb{N}}\sum_{k\in\mathbb{N}}\tilde{\tau}_L(\omega_j)\tilde{\tau}_L(\omega_k)\left\{\sum_u{}''\sum_v{}''\text{cov}(c_u^*, c_v^*)e^{i(v.\omega_k-u.\omega_j)}\right\}, \tag{8.7}$$

where $\tilde{\tau}_L(\lambda) = \tau(\lambda)-\tau_L(\lambda)$ and $\sum_u'' = \sum\cdots\sum_{|u_i|\leqslant g(n_i)}$, $i = 1, ..., d$. The proof that (8.7) $= o(n^{-1})$ is somewhat different from that (in the time series literature) when I_g is replaced by I in r_{1L}. With $n(u) = \Pi_{i=1}^{d}(n_i - |u_i|)$, the term in braces in (8.7) is

$$\sum_u{}''\sum_v{}''[n(u)n(v)]^{-1}\sum_{s(u)}\sum_{t(v)}\{\gamma_{t-s-u}\gamma_{t+v-s}+\gamma_{t-s}\gamma_{t-s+v-u}$$
$$+\,cum\,(x_s, x_{s+u}, x_t, x_{t+v})\}\,e^{i(v.\omega_k-u.\omega_j)}$$

$$= \sum_u{}'' \sum_v{}'' [n(u)n(v)]^{-1} \sum_{s(u)} \sum_{t(v)} \Bigg[\int_{\Pi^d} \int_{\Pi^d} f(\lambda) f(\chi)$$

$$\times \left\{ e^{i(t-s-u).\lambda - i(t+v-s).\chi} + e^{i(t-s).\lambda - i(t-s+v-u).\chi} \right\} \, d\lambda \, d\chi$$

$$+\kappa \sum_\ell \beta_{s-\ell} \beta_{s+u-\ell} \beta_{t-\ell} \beta_{t+v-\ell} \Bigg] e^{i(v.\omega_k - u.\omega_j)}. \tag{8.8}$$

The contribution to (8.7) from the first term in braces in (8.8) is

$$\{(2\pi)^d n\}^{-2} \int_{\Pi^d} \int_{\Pi^d} \sum_{j \in \mathbb{N}} \sum_{k \in \mathbb{N}} \tilde{\tau}_L(\omega_j) \tilde{\tau}_L(\omega_k) \sum_u{}'' \sum_v{}'' \{n(u)n(v)\}^{-1}$$

$$\times e^{-iu.(\lambda+\omega_j) - iv.(\chi-\omega_k)} \sum_{s(u)} \sum_{t(v)} e^{i(t-s).(\lambda-\chi)} f(\lambda) f(\chi) \, d\lambda \, d\chi$$

$$= \{(2\pi)^d n\}^{-2} \int_{\Pi^d} \int_{\Pi^d} \left\{ \sum_{j \in \mathbb{N}} \tilde{\tau}_L(\omega_j) \sum_u{}'' n(u)^{-1} e^{-iu.(\lambda+\omega_j)} \sum_{s(u)} e^{is.(\chi-\lambda)} \right\}$$

$$\times \left\{ \sum_{k \in \mathbb{N}} \tilde{\tau}_L(\omega_k) \sum_v{}'' n(v)^{-1} e^{iv.(\omega_k-\chi)} \sum_{t(v)} e^{it.(\lambda-\chi)} \right\} f(\lambda) f(\chi) \, d\lambda \, d\chi.$$

By the Schwarz inequality and A5 this is bounded by a constant times

$$\{(2\pi)^d n\}^{-2} \left\{ \int_{\Pi^d} \int_{\Pi^d} \left\| \sum_{j \in \mathbb{N}} \tilde{\tau}_L(\omega_j) \sum_u{}'' n(u)^{-1} e^{-iu.(\lambda+\omega_j)} \sum_{s(u)} e^{is.(\chi-\lambda)} \right\|^2 d\lambda \, d\chi \right.$$

$$\left. \times \int_{\Pi^d} \int_{\Pi^d} \left\| \sum_{k \in \mathbb{N}} \tilde{\tau}_L(\omega_k) \sum_v{}'' n(v)^{-1} e^{iv.(\omega_k-\chi)} \sum_{t(v)} e^{it.(\lambda-\chi)} \right\|^2 d\lambda \, d\chi \right\}^{\frac{1}{2}}$$

$$= n^{-2} \sum_u{}'' n(u)^{-1} \left\| \sum_{j \in \mathbb{N}} \tilde{\tau}_L(\omega_j) e^{-iu.\omega_j} \right\|^2$$

since $\sum_{s(u)} 1 = n(u)$. For $|u_i| \leqslant g(n_i)$, $i = 1, ..., d$, A3 implies that $n(u)^{-1} \leqslant K n^{-1}$, so the last displayed expression is bounded by a constant times

$$n^{-3} \sum_u{}'' \left\| \sum_{j \in \mathbb{N}} \tilde{\tau}_L(\omega_j) e^{-iu.\omega_j} \right\|^2 \leqslant n^{-3} \sum_u{}''' \left\| \sum_{j \in \mathbb{N}} \tilde{\tau}_L(\omega_j) e^{-iu.\omega_j} \right\|^2, \tag{8.9}$$

where $\sum_u'''$ is the sum $\sum \cdots \sum_{1-n_i \leqslant u_i \leqslant n_i, i=1,\ldots,d}$. Because

$$\sum_{u_\ell=1-n_\ell}^{0} e^{2\pi i(k_\ell - j_\ell)/n_\ell} = \sum_{u_\ell=1}^{n_\ell} e^{2\pi i(k_\ell - j_\ell) n_\ell} = n_\ell 1(j_\ell = k_\ell) \tag{8.10}$$

for $1 \leqslant j_\ell, k_\ell \leqslant n_\ell$, it follows that the bound in (8.9) is

$$2^d n^{-2} \sum_{j \in \mathbb{N}} \left\| \tilde{\tau}_L(\omega_j) \right\|^2 \leqslant 2^d \eta^2 n^{-1}.$$

The contribution to (8.7) from the second term in braces in (8.8) is readily found to be of the same order. The contribution to (8.7) from the fourth cumulant term in (8.8) is bounded by

$$\begin{aligned}
& K n^{-2} \sum_u{}'' \sum_v{}'' \{n(u)n(v)\}^{-1} \left\| \sum_{j \in \mathbb{N}} \tilde{\tau}_L(\omega_j) e^{-iu.\omega_j} \right\| \left\| \sum_{k \in \mathbb{N}} \tilde{\tau}_L(\omega_k) e^{iv.\omega_k} \right\| \\
& \quad \times \sum_{s(u)} \sum_{t(v)} \sum_{\ell} \left| \beta_{s-\ell} \beta_{s+u-\ell} \beta_{t-\ell} \beta_{t+v-\ell} \right| \\
& \quad \leqslant K n^{-4} \sum_u{}'' \sum_v{}'' \left\{ \left\| \sum_{j \in \mathbb{N}} \tilde{\tau}_L(\omega_j) e^{-iu.\omega_j} \right\|^2 + \left\| \sum_{k \in \mathbb{N}} \tilde{\tau}_L(\omega_k) e^{iv.\omega_k} \right\|^2 \right\} \\
& \quad \times \sum_{s(u)} \sum_{t(v)} \sum_{\ell} \left| \beta_{s-\ell} \beta_{s+u-\ell} \beta_{t-\ell} \beta_{t+v-\ell} \right| \\
& \quad \leqslant K n^{-4} \sum_u{}'' \left\| \sum_{j \in \mathbb{N}} \tilde{\tau}_L(\omega_j) e^{-iu.\omega_j} \right\|^2 \sum_{s(u)} \sum_{\ell} \left| \beta_{s-\ell} \right| \sum_t \left| \beta_{t-\ell} \right| \sum_v \left| \beta_{t+v-\ell} \right| \\
& \quad \leqslant K n^{-3} \sum_u{}'' \left\| \sum_{j \in \mathbb{N}} \tilde{\tau}_L(\omega_j) e^{-iu.\omega_j} \right\|^2 \leqslant K \eta^2 n^{-1}
\end{aligned}$$

as before.

We now wish to show that for fixed L

$$n^{\frac{1}{2}} r_{1L} \to_d \mathcal{N}(0, \Psi_L), \tag{8.11}$$

where

$$\begin{aligned}
\Psi_L = {} & \frac{2}{(2\pi)^d} \int_{\Pi^d} \tau_L(\lambda) \tau_L'(\lambda) f(\lambda)^2 d\lambda \\
& + \kappa \left\{ \int_{\Pi^d} \tau_L(\lambda) f(\lambda)\, d\lambda \right\} \left\{ \int_{\Pi^d} \tau_L'(\lambda) f(\lambda)\, d\lambda \right\}.
\end{aligned}$$

Using (8.10),

$$r_{1L} = (2\pi)^{-d} \sum_{\ell \in A_L} \prod_{i=1}^{d} \left(1 - \frac{|\ell_i|}{L} \right) \tau_\ell \left(c_\ell^* - \gamma_\ell \right)$$

for n sufficiently large, because then $L + g(n_i) < n_i$ for all i and there is no contribution from aliased terms. In view of A2,

$$c_\ell^* - \gamma_\ell = n(\ell)^{-1} \sum_j \sum_k \beta_j \beta_k \sum_{t(\ell)} \left\{ \varepsilon_{t-j}\varepsilon_{t+\ell-k} - 1(j = k - \ell) \right\}. \tag{8.12}$$

Fix $\eta_2 > 0$. We may choose M such that

$$\sum_{j \notin A_M} \left| \beta_j \right| < \eta_2.$$

The difference between (8.12) and

$$\begin{aligned} q_{\ell,M} &= n(\ell)^{-1} \sum_{j, j+\ell \in A_M} \beta_j \beta_{j+\ell} \sum_{t(\ell)} \left(\varepsilon_{t-j}^2 - 1 \right) \\ &+ n(\ell)^{-1} \sum_{\substack{j \in A_M \\ k \neq j+\ell}} \sum_{k \in A_M} \beta_j \beta_k \sum_{t(\ell)} \varepsilon_{t-j}\varepsilon_{t+\ell-k} \end{aligned} \tag{8.13}$$

has mean zero and variance that is readily shown to be $O\left(\eta_2 n^{-1}\right) = o(n^{-1})$ as $\eta_2 \to 0$. In view of the Cramer–Wold device we seek to establish asymptotic normality of

$$n^{\frac{1}{2}} \sum_{\ell \in A_L} a_\ell q_{\ell,M} \tag{8.14}$$

for arbitrary a_ℓ, not all zero. In other words, we establish asymptotic normality of a linear combination of finitely many terms of the forms

$$n^{\frac{1}{2}} n(\ell)^{-1} \sum_{t(\ell)} \left\{ \varepsilon_{t-j}\varepsilon_{t+\ell-k} - 1 \right\}, \quad j \neq k - \ell,$$

and

$$n^{\frac{1}{2}} n(\ell)^{-1} \sum_{t(\ell)} \left(\varepsilon_{t-j}^2 - 1 \right),$$

since L and M are fixed.

We map $\mathbb{Z}^d$ into $\mathbb{Z}_+$ in order to employ a standard martingale central limit theorem for triangular arrays. There is considerable literature on asymptotic theory for random fields, including work based on multilateral models, e.g. [22], on the basis of unidirectional increase, i.e. with only the n_{Ui} increasing. For $k \geqslant 1$, denote by $C_k^{(d)}$ the lattice points on the surface of the d -dimensional cube with vertices $(\pm k, ..., \pm k)$; there are $m_k^{(d)} = (2k+1)^d - (2k-1)^d$ such points. Consider an arbitrary ordering of the points $j \in C_k^{(b)}$, namely $j_{(1)}^{(k)}, ..., j_{(m_k^{(d)})}^{(k)}$. Introduce a function $\phi : \mathbb{Z}^d \to \mathbb{Z}_+$ such that

$$\begin{aligned} \phi(0, ..., 0) &= 1 \\ \phi\left(j_{(1)}^{(1)}\right) &= 2, ..., \phi\left(j_{(3^d-1)}^{(1)}\right) = 3^d, \\ &\vdots \qquad\qquad \vdots \\ \phi\left(j_{(1)}^{(k)}\right) &= (2k-1)^d + 1, ..., \phi\left(j_{((2k+1)^d-(2k-1)^d)}^{(k)}\right) = (2k+1)^d, \end{aligned}$$

and so on. For example, in case $d = 2$ we might have the "spiral" ordering

$$j_{(1)}^{(k)} = (-k, k), \quad j_{(2)}^{(k)} = (-k, 1-k), ..., j_{(3^d-1)} = (1-k, -k).$$

When $n_{Li} = n_{Ui} = n^*$ for all i, so $\mathbb{N} = A_{2n^*+1}$, the $(2n^*+1)^d$ observations have thus accumulated first at $\{0, ..., 0\}$, followed by $C_1^{(d)}, ..., C_{n^*}^{(d)}$, in that order.

For more general circumstances, define

$$\psi_n(j) = \phi(j) - \#\{k : k \notin \mathbb{N};\ \phi(k) < \phi(j)\}, \quad j \in \mathbb{N};$$

thus, having ordered on $A_{\max}(n_{Li}, n_{Ui}, i = 1, ..., d)$ we drop points outside $\mathbb{N}$ and then close up the gaps, re-labelling and preserving the order. Introduce the triangular array $\delta_n(s)$, $1 \leqslant s \leqslant n$, of iid variates with zero mean, variance 1 and fourth cumulant κ, such that

$$\delta_n\left(\psi_n(j)\right) = \varepsilon_j, \quad j \in \mathbb{N}.$$

Considering now the contribution to (8.14) from the "squared" terms ε_{t-j}^2 in $q_{\ell,M}$,

$$\sum_{t(\ell)} \left(\varepsilon_{t-j}^2 - 1\right) \tag{8.15}$$

differs from

$$\sum_{t\in\mathbb{N}} \left(\varepsilon_t^2 - 1\right) \tag{8.16}$$

by

$$O\left(\sum_{i=1}^{d} \prod_{j=1, j\neq i}^{d} n_j\right) = O\left(n \sum_{i=1}^{d} n_i^{-1}\right) = O\left(n^{1-\xi}\right) \tag{8.17}$$

terms, uniformly in $j \in A_M$, $\ell \in A_L$. Thus, because the $\varepsilon_t^2 - 1$ are iid with zero mean and finite variance, the difference between (8.15) and (8.16) is $O_p\left(n^{(1-\xi)/2}\right)$. As for product terms, note that in

$$\sum_{t(\ell)} \varepsilon_{t-j}\varepsilon_{t+\ell-k} \tag{8.18}$$

we have for each summand either $\phi(t-j) > \phi(t+\ell-k)$ or $\phi(t-j) < \phi(t+\ell-k)$. Overall there are $n - O\left(n^{1-\xi}\right)$ summands, and, possibly after finite translation across

$\mathbb{Z}^d$, each can be written in the form $\delta_n(s)\delta_n(s - r_{sn}(j,k,\ell))$ for suitable s and positive integer $r_{sn}(j,k,\ell)$. Thus because these summands are uncorrelated across s, (8.18) differs by $O_p\left(n^{(1-\xi)/2}\right)$ from

$$\sum_{s=1}^{n} \delta_n(s)\delta_n\left(s - r_{sn}(j,k,\ell)\right).$$

It follows from this discussion that (8.14) differs by $o_p(1)$ from $n^{-\frac{1}{2}}\sum_{s=1}^{n} u_n(s)$, where

$$u_n(s) = \left\{\delta_n^2(s) - 1\right\} \sum_{\ell \in A_L} a_\ell\{n/n(\ell)\} \sum_{j, j+\ell \in A_M} \beta_j \beta_{j+\ell}$$
$$+\delta_n(s) \sum_{\ell \in A_L} a_\ell\{n/n(\ell)\} \sum_{\substack{j \in A_M, k \in A_M \\ k \neq j+\ell}}\sum \beta_j \beta_k \delta_n\left(s - r_{sn}(j,k,\ell)\right).$$

The $u_n(s)$ thus comprise a martingale difference array. Denote by $F_{s,n}$ the σ-field of events generated by $\delta_n(t)$, $t \leqslant s$. It follows from [12, Chapter 2], [39] that if

$$\lim_{n\to\infty} n^{-1} \sum_{s=1}^{n} E u_n^2(s) \tag{8.19}$$

is positive and finite and

$$n^{-1}\sum_{s=1}^{n} E\left\{u_n^2(s) 1\left(|u_n(s)| \geqslant \eta_3 n^{\frac{1}{2}}\right)\right\} \to 0, \quad \text{all } \eta_3 > 0, \tag{8.20}$$

$$n^{-1}\sum_{s=1}^{n}\left[E\left\{u_n^2(s)\middle| F_{s-1,n}\right\} - E u_n^2(s)\right] \to_p 0, \tag{8.21}$$

then

$$n^{-\frac{1}{2}}\sum_{s=1}^{n} u_n(s) \to_d \mathcal{N}(0, \sigma^2),$$

where σ^2 is given by (8.19).

To prove (8.20) write $u_n(s) = u_{1n}(s) + u_{2n}(s)$, where $u_{1n}(s)$ consists of the terms in $\{\delta_n^2(s) - 1\}$. It suffices to show that

$$n^{-1}\sum_{s=1}^{n} E\left\{u_{in}^2(s) 1\left(|u_{in}(s)| > \eta_s n^{\frac{1}{2}}\right)\right\} \to 0, \text{ all } \eta_3 > 0, \quad i = 1, 2.$$

For $i = 1$ this follows from identity of distribution and finite fourth moment of the $\delta_n(s)$, boundedness of $n/n(\ell)$ and summability of the β_j. For $i = 2$ it follows from the same facts after applying Cauchy and elementary inequalities.

Next consider (8.21), which is equivalent to

$$
\begin{aligned}
& n^{-1}\sum_{s=1}^{n}\left[\left\{\sum_{\ell\in A_L} a_\ell \frac{n}{n(\ell)} \sum_{\substack{j\in A_M \\ k\neq j+\ell}}\sum_{k\in A_M} \beta_j\beta_k\delta_n\left(s-r_{sn}(j,k,\ell)\right)\right\}^2\right. \\
& \qquad \left. -E\left\{\sum_{\ell\in A_L} a_\ell \frac{n}{n(\ell)} \sum_{\substack{j\in A_M \\ k\neq j+\ell}}\sum_{k\in A_M} \beta_j\beta_k\delta_n\left(s-r_{sn}(j,k,\ell)\right)\right\}^2\right] \\
& \qquad +2E\varepsilon_0^3 n^{-1}\sum_{s=1}^{n}\left\{\sum_{\ell\in A_L} a_\ell \frac{n}{n(\ell)} \sum_{j,j+\ell\in A_M} \beta_j\beta_{j+\ell}\right\} \qquad (8.22) \\
& \times\left\{\sum_{\ell\in A_L} a_\ell \frac{n}{n(\ell)} \sum_{\substack{j\in A_M \\ k\neq j+\ell}}\sum_{k\in A_M} \beta_j\beta_k\delta_n\left(s-r_{sn}(j,k,\ell)\right)\right\} \to_p 0
\end{aligned}
$$

because the squared terms in $\delta_n^2(s)-1$ contribute nothing due to independence. For any fixed $j_{(i)}, k_{(i)} \in A_M$ and $\ell_{(i)} \in A_L$, $i=1,2$, consider

$$
n^{-1}\sum_{s=1}^{n}\left\{\delta_n(s-r_{sn1})\delta_n(s-r_{sn2}) - E\delta_n(s-r_{sn1})\delta_n(s-r_{sn2})\right\}, \qquad (8.23)
$$

where $r_{sni} = r_{sn}\left(j_{(i)}, k_{(i)}, \ell_{(i)}\right)$. Now (8.23) has mean zero and variance

$$
\begin{aligned}
n^{-2}\sum_{s=1}^{n}\sum_{t=1}^{n}&[E\delta_n(s-r_{sn1})\delta_n(t-r_{tn1})E\delta_n(s-r_{sn2})\delta_n(t-r_{tn2}) \\
&+E\delta_n(s-r_{sn1})\delta_n(t-r_{tn2})E\delta_n(s-r_{sn2})\delta_n(t-r_{tn1}) \\
&+cum\left\{\delta_n(s-r_{sn1}), \delta_n(t-r_{tn1}), \delta_n(s-r_{sn2}), \delta_n(t-r_{tn2})\right\}]. \qquad (8.24)
\end{aligned}
$$

All summands are finite. Summands for $s=t$ contribute $O(n^{-1})$. For $s\neq t$, there is a difference from the case $d=1$ in that the r_{sni} depend on n, but because $C_k^{(d)}$ has $O(k^{d-1})$ lattice points as $k\to\infty$, and the surface of $\mathbb{N}$ has $O\left(\sum_{i=1}^{d}\prod_{j=1,j\neq i}^{d} n_j\right)$ lattice points, and because of (8.17), it follows that $r_{sni}=O(n^{1-\xi})$ uniformly as $n\to\infty$. Thus, splitting the sum into two parts, one containing terms for which $|s-t|\leqslant n^{1-\xi/2}$ and one terms for which $|s-t|>n^{1-\xi/2}$ the first component contributes $O(n^{-\xi/2})$ to (8.24), and the second, zero. Since only finitely many terms of form (8.23) are involved, and because clearly $n^{-1}\sum_{i=1}^{n}\delta_n\left(s-r_{sn}(j,k,\ell)\right)=O_p(n^{-\frac{1}{2}})$, (8.21) is established.

We can evaluate (8.19) as

$$\sum_{\ell\in A_L}\sum_{m\in A_L} a_\ell a_m \Bigg\{ \sum_{i\in A_M}\sum_{j\in A_M}\sum_{k,k-i+j-\ell+m\in A_M} \beta_i\beta_j\beta_k\beta_{k-i+j-\ell+m}$$
$$+\sum_{i\in A_M}\sum_{j\in A_M}\sum_{k,k+i-j-\ell+m\in A_M} \beta_i\beta_j\beta_k\beta_{k+i-j+\ell+m}$$
$$+\kappa\left(\sum_{j,j+\ell\in A_M}\beta_j\beta_{j+\ell}\right)\left(\sum_{j,j+m\in A_M}\beta_j\beta_{j+m}\right)\Bigg\}.$$

Since this differs by $O(\eta_2)$ from

$$\sum_{\ell\in A_L}\sum_{m\in A_L} a_\ell a_m \left\{\sum_i\sum_j\sum_k \beta_i\beta_j\beta_k(\beta_{k-i+j-\ell+m}+\beta_{k+i-j+\ell+m})+\kappa\gamma_\ell\gamma_m\right\}$$
$$=\sum_{\ell\in A_L}\sum_{m\in A_L} a_\ell a_m \left[(2\pi)^{-d}\int_{\Pi^d} f(\lambda)^2\exp\{i(\ell-m)\lambda+i(\ell+m)\lambda\}\,d\lambda+\kappa\gamma_\ell\gamma_m\right]$$

we deduce (8.11) via Bernstein's lemma. From (8.6), $\Psi_L\to\Psi$ as $L\to\infty$, so we then likewise deduce (8.2). □

9. Proof of Theorem 4

Given (7.4), we have already justified the claims about (7.5), and for (7.4) we only prove the second statement with $i=2$, because the other proofs are easier. We have

$$\hat{\mu}_4^{(2)}-\mu_4=n^{-1}\sum_{t\in\mathbb{N}}\left(\hat{\varepsilon}_t^{(2)4}-\varepsilon_t^4\right)+n^{-1}\sum_{t\in\mathbb{N}}\left(\varepsilon_t^4-\mu_4\right).$$

The second term on the right is $o_p(1)$ by the law of large numbers, while by the identity $x^4-y^4=(x-y)(x^3+x^2y+xy^2+y^3)$ and H ölder's inequality the first term is $o_p(1)$ if

$$n^{-1}\sum_{t\in\mathbb{N}}\left(\hat{\varepsilon}_t^{(2)}-\varepsilon_t\right)^4\to_p 0. \tag{9.1}$$

Write

$$\hat{\varepsilon}_t^{(2)}-\varepsilon_t=E_t+F_t,$$

where

$$E_t=(2\pi)^{d/2}n^{-\frac{1}{2}}\sum_{j\in\mathbb{N}}\left\{\alpha\left(e(i\omega_j);\hat{\theta}\right)-\alpha\left(e(i\omega_j);\theta_0\right)\right\}w(\omega_j)e^{-it.\omega_j},$$
$$F_t=(2\pi)^{d/2}n^{-\frac{1}{2}}\sum_{j\in\mathbb{N}}\alpha\left(e(i\omega_j);\theta_0\right)w(\omega_j)e^{-it.\omega_j}-\varepsilon_t.$$

Again, for brevity we assume $\mu=0$ and replace $x_t-\bar{x}$ by x_t.

By direct calculation, using (8.10) again,

$$F_t = \sum_{s \notin \mathbb{N}} \alpha_{t-s} x_s + \sum_{s \in \mathbb{N}} x_s \sum_{k \neq 0} \alpha_{t-s+k(n)},$$

where $\alpha_j = (2\pi)^{-d} \int_{\Pi^d} \alpha\left(e(i\lambda); \theta_0\right) e^{-ij.\lambda} d\lambda$ and $k(n) = (k_1 n_1, ..., k_d n_d)$. It follows from A12 that x_t has a linear representation as in A2 but with the β_j possibly being only square-summable. Nevertheless,

$$Ex_t^4 = 3\left(\sum_j \beta_j^2\right)^2 + \sum_j \beta_j^4 E\varepsilon_{t-j}^4 \leqslant K\left(\sum_j \beta_j^2\right)^2 < \infty.$$

Thus

$$E\left(\sum_{s \notin \mathbb{N}} \alpha_{t-s} x_s\right)^4 \leqslant K\left(\sum_{s \notin \mathbb{N}} |\alpha_{t-s}|\right)^4 \leqslant K \sum_{s \notin \mathbb{N}} |\alpha_{t-s}|.$$

It follows that

$$\begin{aligned} n^{-1} \sum_{t \in \mathbb{N}} E\left(\sum_{s \notin \mathbb{N}} \alpha_{t-s} x_s\right)^4 &\leqslant Kn^{-1} \sum_{t \in \mathbb{N}} \sum_{s \notin \mathbb{N}} |\alpha_{t-s}| \\ &\leqslant Kn^{-1} \sum_j \left|\alpha_j\right| \prod_{\ell=1}^d \left\{|j_\ell| \, 1\left(|j_\ell| \leqslant n_\ell\right) + n_\ell 1\left(|j_\ell| \geqslant n_\ell\right)\right\}, \end{aligned}$$

which tends to zero as $n \to \infty$ by summability of the α_j and the Toeplitz lemma. Beginning in the same way,

$$E\left(\sum_{s \in \mathbb{N}} x_s \sum_{k \neq 0} \alpha_{t+s+k(n)}\right)^4 \leqslant K\left(\sum_{s \in \mathbb{N}} \sum_{k \neq 0} \left|\alpha_{t-s+k(n)}\right|\right)^4.$$

For any of the finitely many k such that $|k_\ell| \leqslant 1$ for all ℓ, and $k_\ell \neq 0$ for some ℓ,

$$\begin{aligned} n^{-1} \sum_{t \in \mathbb{N}} \left(\sum_{s \in \mathbb{N}} \left|\alpha_{t-s+k(n)}\right|\right)^4 &\leqslant Kn^{-1} \sum_{t \in \mathbb{N}} \sum_{s \in \mathbb{N}} \left|\alpha_{t-s+k(n)}\right| \\ &\leqslant Kn^{-1} \sum_{j \in \mathbb{N}_2} \left|\alpha_j\right| \prod_{\ell=1}^d |j_\ell|, \end{aligned}$$

where $\mathbb{N}_2 = \{j : |j_\ell| \leqslant 2n_\ell, \ell = 1, ..., d\}$. This is $o(1)$ as before. Denoting by $\mathbb{K}$ the remaining $k \in \mathbb{Z}^d$, by elementary inequalities the proof that $n^{-1} \sum_{t \in \mathbb{N}} EF_t^4 \to 0$ is completed by the calculation

$$n^{-1}\sum_{t\in\mathbb{N}}\left(\sum_{s\in\mathbb{N}}\sum_{k\in\mathbb{K}}\left|\alpha_{t-s+k(n)}\right|\right)^4 \leqslant K\sum_{\ell=1}^{d}\sum_{j:|j_\ell|\geqslant n_\ell}\left|\alpha_j\right| \to 0,$$

by summability of α_j.

Finally,

$$n^{-1}\sum_{t\in\mathbb{N}}E_t^4 \leqslant n^{-1}\left(\sum_{t\in\mathbb{N}}E_t^2\right)^2 \tag{9.2}$$

and from (8.10)

$$\begin{aligned}\sum_{t\in\mathbb{N}}E_t^2 &= (2\pi)^d\sum_{j\in\mathbb{N}}\left|\alpha\left((e(i\omega_j);\hat{\theta}\right)-\alpha\left(e(i\omega_j);\theta_0\right)\right|^2 I(\omega_j)\\ &\leqslant K\left\|\hat{\theta}-\theta_0\right\|^2\sum_{j\in\mathbb{N}}I(\omega_j)\leqslant K\left\|\hat{\theta}-\theta_0\right\|^2\sum_{t\in\mathbb{N}}x_t^2\end{aligned}$$

with probability approaching 1 as $n \to \infty$, in view of A12 and A13. Then (9.2) $= O_p(n^{1-4\zeta}) = o_p(1)$ for $\zeta > \frac{1}{4}$. This completes the proof of (9.1). □

Acknowledgements

This research was supported by a Leverhulme Trust Personal Research Professorship and ESRC Grant R000233962. Thanks are due to Fabrizio Iacone for carrying out the numerical work in Section 6 and two referees for helpful comments.

References

[1] M.M. Ali, Analysis of stationary spatial-temporal processes: estimation and prediction, Biometrika 66 (1979) 513–518.

[2] J. Besag, Spatial interaction and the statistical analysis of lattice systems (with Discussion), J.R. Statist. Soc. Ser. B 36 (1974) 192–236.

[3] S.-T. Chiu, Weighted least squares estimators on the frequency domain for the parameters of a time series, Ann. Statist. 16 (1988) 1315–1326.

[4] A.D. Cliff, J.K. Ord, Spatial Processes: Models & Applications, Pion, London, 1981.

[5] N. Cressie, Statistics for Spatial Data, second ed., Wiley, New York, 1993.

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

[7] M. Deistler, W.T.M. Dunsmuir, E.J. Hannan, Vector linear time series models: corrections and extensions, Adv. Appl. Probab. 10 (1978) 360–372.

[8] A.C. Gleeson, C.A. McGilchrist, Bilateral processes on a rectangular lattice, Australian J. Statist. 22 (1980) 197–206.

[9] X. Guyon, Parameter estimation for a stationary process on a d-dimensional lattice, Biometrika 69 (1982) 95–106.

[10] X. Guyon, Random Fields on a Network, Springer, New York, 1995.

[11] R. Haining, The moving average model for spatial interaction, Trans Inst. Brit. Geographers 3 (1978) 202–225.
[12] P. Hall, C.C. Heyde, Martingale Limit Theory and its Applications, Academic Press, New York, 1980.
[13] E.J. Hannan, The asymptotic theory of linear time series models, J. Appl. Probab. 10 (1973) 130–145.
[14] E.J. Hannan, W.T.M. Dunsmuir, M. Deistler, Estimation of vector ARMAX models, J. Multivariate Anal. 10 (1980) 275–295.
[15] H. Helson, D. Lowdenslager, Prediction theory and Fourier series in several variables, Acta Math. 99 (1958) 165–202.
[16] C.C. Heyde, R. Gay, Smoothed periodogram asymptotics and estimation for processes and fields with possible long range dependence, Stoch. Proc. Appl. 45 (1993) 169–182.
[17] Y. Hosoya, M. Taniguchi, A central limit theorem for stationary processes and the parametric estimation of linear processes, Ann. Statist. 10 (1982) 132–153.
[18] D. Huang, Central limit theorem for two-parameter martingale differences with application to stationary random fields, Sci. China 35 (1992) 413–425.
[19] D. Huang, V.V. Anh, Estimation of spatial ARMA models, Australian J. Statist. 34 (1992) 513–530.
[20] A.K. Jain, Advances in mathematical models for image processing, Proc. IEEE 69 (1981) 502–528.
[21] J. Jiang, Parameter estimation of spatial AR model, Chin. Ann. Math. 12B (1991) 432–444.
[22] J. Jiang, Uniform convergence rate of sample ACV and ACR for linear spatial series under more general martingale condition, Adv. Math. 20 (1991) 39–50.
[23] R.L. Kashyap, Characterization and estimation of two-dimensional ARMA models, IEEE Trans. Information Theory IT-30 (1984) 736–745.
[24] J.T. Kent, K.V. Mardia, Spectral and circulant approximations to the likelihood for stationary Gaussian random fields, J. Statist. Planning Inference 50 (1996) 379–394.
[25] H.T. Kiiveri, Fitting spatial correlation models: approximating a likelihood approximation, Australian J. Statist. 34 (1992) 497–512.
[26] H. Korezlioglu, P. Loubaton, Spectral factorization of wide sense stationary processes on Z^2, J. Multivariate Anal. 19 (1988) 24–47.
[27] L. LeCam, On the asymptotic theory of estimation and testing hypotheses, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, vol 1, University of California Press, Berkeley (1956) 129–156.
[28] C. Ma, Stationary spatio-temporal covariance models, J. Multivariate Anal. 86 (2003) 97–107.
[29] C. Ma, Spatial autoregression and related spatio-temporal models, J. Multivariate Anal. 88 (2004) 152–162.
[30] K.V. Mardia, R.J. Marshall, Maximum likelihood estimation of models for residual covariance in spatial regression, Biometrika 71 (1984) 135–146.
[31] R.J. Martin, A subclass of lattice processes applied to a problem in planar sampling, Biometrika 66 (1979) 209–217.
[32] P.A.P. Moran, Necessary conditions for Markovian processes on a lattice, J. Appl. Probab. 10 (1973) 605–612.
[33] B.S. Nahapetian, The central limit theorem for random fields with mixing conditions, Adv. Probab. 6 (1980) 531–548.
[34] V. de Oliveira, B. Kedem, D.A. Short, Bayesian prediction of transformed Gaussian random fields, J. Amer. Statist. Assoc. 92 (1997) 1422–1433.
[35] B. Ranneby, Stochastic models of variation in time and space, In: Statistics in Theory and Practice: Essays in Honour of Bertil Matérn, Swedish University of Agricultural Sciences, Umea, 1982, pp. 227–245.
[36] P.M. Robinson, Time series residuals with application to probability density estimation, J. Time Series Anal. 8 (1987) 329–344.
[37] P.M. Robinson, The stochastic difference between econometric statistics, Econometrica 56 (1988) 531–548.
[38] Yu A. Rosanov, On Gaussian fields with given conditional distributions, Probab. Theory Appl. 12 (1967) 381–391.
[39] D.J. Scott, Central limit theorems for martingales using a Skorokhod representation approach, Adv. Appl. Probab. 5 (1973) 119–137.
[40] M.L. Stein, Interpolation of Spatial Data: Some Theory for Kriging, Springer, NewYork, 1999.
[41] M. Taniguchi, On estimation of the integrals of the fourth order cumulant spectral density, Biometrika 69 (1982) 117–122.

[42] D. Tjostheim, Statistical spatial series modelling II: some further results on unilateral lattice processes, Adv. Appl. Probab. 15 (1983) 562–584.
[43] D. Tjostheim, Statistical spatial series modelling, Adv. Appl. Probab. 10 (1978) 130–154 Correction Adv. Appl. Probab. 16 (1984) 120.
[44] L.T. Tran, S. Yakowitz, Nearest neighbor estimators for random fields, J. Multivariate Anal. 44 (1993) 23–46.
[45] B.K. van der Waerden, Modern Algebra, vol. 1, Ungar, New York, 1953.
[46] P. Whittle, On stationary processes in the plane, Biometrika 41 (1954) 434–449.
[47] P. Whittle, Topographic correlation power-law, covariance functions, and diffusion, Biometrika 49 (1963) 305–314.
[48] Q. Yao, P. Brockwell, Gaussian maximum likelihood estimation for ARMA models II: spatial processes, Preprint, 2002.