# MODIFIED WHITTLE ESTIMATION OF MULTILATERAL SPATIAL MODELS

*P. M Robinson*
*J. Vidal Sanz*

THE INSTITUTE FOR FISCAL STUDIES
DEPARTMENT OF ECONOMICS, UCL
cemmap working paper CWP18/03

# Modified Whittle Estimation of Multilateral Spatial Models*

P.M. Robinson and J. Vidal Sanz

London School of Economics

November 21, 2003

**Abstract**

We consider the estimation of parametric models for stationary spatial or spatio-temporal data on a $d$-dimensional lattice, for $d \geq 2$. 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 difficulty of computing a continuous-frequency form of Whittle estimate or a time domain Gaussian maximum likelihood estimate, especially in case of multilateral models, due mainly to the Jacobian term. An extension of the discrete-frequency Whittle estimate from the time series literature deals conveniently with the latter 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

*Research supported by ESRC Grant R000239936. Thanks are due to Fabrizio Iacone for carrying out the numerical work reported in Section 3.

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. A Monte Carlo study of finite sample behaviour is included. The asymptotic regime allows increase in both directions, unlike the usual random fields formulation, with the central limit theorem established after re-ordering as a triangular array. When the data are non-Gaussian, the asymptotic variances of all parameter estimates are likely to be affected, and we provide a consistent, non-negative definite, estimate of the asymptotic variance matrix.

*AMS 2000 subject classifications.* Primary 62M30; secondary 62F10, 62F12.

*Key words and phrases.* Spatial data, multilateral models, Whittle estimation, edge effect, consistent variance estimation.

*Abbreviated title.* Estimation of spatial models.

## 1. INTRODUCTION

Consider a stationary process $x_t$ defined on a $d$-dimensional lattice, $t$ being a multiple index $(t_1, ..., t_d)$ with $t_j \in \mathbb{Z} = \{0, \pm 1, ...\}$, $j = 1, ..., d$, and having a spectral density $f(\lambda)$, $\lambda = (\lambda_1, ..., \lambda_d)$, $\lambda \in \Pi^d$, $\Pi = (-\pi, \pi]$. This paper is concerned with large sample inference on an unknown $m$-dimensional column vector $\theta_0$, given a known functional form $f(\lambda; \theta)$ such that $f(\lambda; \theta_0) \equiv f(\lambda)$.

Such parametric modelling is often approached in terms of linear filtering of a white noise process. For $\theta \in \Theta$, where $\Theta \subset \mathbb{R}^m$ is the set of admissible parameter values, and for $z = (z_1, ..., z_d)$ having complex-valued elements, define

$$a(z;\theta) = \sum_{j_1=-p_{L1}}^{p_{U1}} \cdots \sum_{j_d=-p_{Ld}}^{p_{Ud}} a_j(\theta) \prod_{i=1}^{d} z_i^{j_i}, \quad (1.1)$$

$$b(z;\theta) = \sum_{j_1=-q_{L1}}^{q_{U1}} \cdots \sum_{j_d=-q_{Ld}}^{q_{Ud}} b_j(\theta) \prod_{i=1}^{d} z_i^{j_i}, \quad (1.2)$$

for $j = (j_1, ..., j_d)$, given finite integers $p_{Li} \geq 0$, $p_{Ui} \geq 0$, $q_{Li} \geq 0$, $q_{Ui} \geq 0$ and real-valued functions $a_j(\theta)$, $b_j(\theta)$. We call (1.1) and (1.2) 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

$$ARMA\,(p_{L1}, p_{U1}; ...; p_{Ld}, p_{Ud} : q_{L1}, q_{U1}; ...; q_{Ld}, q_{Ud}) : a(B;\theta_0)(x_t - \mu) = b(B;\theta_0)\varepsilon_t, \quad (1.3)$$

where $\mu = Ex_t$ and

$$E\varepsilon_t = 0,\ E\varepsilon_t^2 = 1,\ E\varepsilon_s\varepsilon_t = 0,\ \text{all } s \neq t,\ t \in \mathbb{Z}^d, \quad (1.4)$$

$$a(z;\theta) \neq 0,\ b(z;\theta) \neq 0, \quad \text{for } |z_i| = 1,\ i = 1, ..., d,\ \theta \in \Theta. \quad (1.5)$$

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

$$f(\lambda;\theta) = (2\pi)^{-d} \left| b\left(E(i\lambda);\theta\right) / a\left(E(i\lambda);\theta\right) \right|^2, \quad \theta \in \Theta, \quad (1.6)$$

with $E(z) = (e^{z_1}, ..., e^{z_d})$. Special cases of (1.3) 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 $a(z;\theta_0) \equiv 1$.

Any of the $p_{Li}, p_{Ui}, q_{Li}, q_{Ui}$ can be positive, so these ARMA structures can be "multilateral", and they provide a flexible approach to modelling. It is necessary that $\theta$ 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 $x_t$. In view of (1.6) this requires in the first place that $\theta$ 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, bearing in mind that we have fixed $E\varepsilon_t^2 = 1$, 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, assuming $\theta$ is identifiable from the $a_j(\theta)$ or $b_j(\theta)$. However, in addition $|a(z;\theta)|^2$, $|b(z;\theta)|^2$ 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 Whittle, 1954), and is thus disregarded. To indicate a more substantive concern, write for $h \geq 1$,

$$a(z;\theta) = \prod_{j=1}^{h} a_j(z;\theta), \quad \text{all } \theta \in \Theta, \tag{1.7}$$

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. van der Waerden, 1953, 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 $\prod_{j=1}^h a_j(z^{\pm 1};\theta)$ with real-valued coefficients are indistinguishable.

When $d = 1$, and $t$ denotes time, the 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 (1.5) 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 $\theta_j$ is the $j$-th element of $\theta$, there is no equivalent bilateral AR$(1,1)$ model.

Unilateral structures have been studied when $d \geq 2$ also. Tjostheim (1978), Korezlioglu and Loubaton (1986) discussed conditions under which $x_t$ has infinite AR and MA representations on a quadrant, so that $x_t$ ($\varepsilon_t$) is expressed in terms of $\varepsilon_s$ ($x_s$) for $s_j \leq t_j$, all $j$. See also Tjostheim (1983), Jiming (1991a). More general representations have also been referred to as "unilateral". Under conditions easily satisfied by (1.3)-(1.5) and in our theorems, $x_t$ has an infinite linear MA representation in $\varepsilon_s$ for $s \leq t$, with square summable coefficients, where $\leq$ 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 Whittle (1954), Helson and Lowdenslager (1958), Guyon (1982), Korezlioglu and Loubaton (1986).

These kind of unilateral representations can be used as a framework for extending to $d \geq 2$ ARMA order-determination methods and AR nonparametric spectral estimation methods developed in case $d = 1$ (see Huang and Anh (1992) and, in the quadrant case, Tjostheim (1983)). They have also been employed in parametric modelling (see e.g. Guyon (1982), Huang (1992), Yao and Brockwell (2002)). However, for $d \geq 2$ a multilateral finite ARMA given by (1.3)-(1.5) cannot necessarily be represented as a unilateral finite ARMA, as demonstrated in a simple example by Whittle (1954), where $d = 2$, $m = 1$, $a(z;\theta) = 1 + \theta^2 - \theta\left(z_1 + z_2 + z_2^{-1}\right)$, $b(z;\theta) \equiv 1$. In this case Whittle (1954) was able to give a closed form expression for the unilateral infinite AR operator which can be shown to be equivalent to $a(z;\theta)\left(1 - \theta z_2\right)\left(1 - \theta z_2^{-1}\right)^{-1}$. This trick can apply somewhat more generally, in particular in the case $d = 2$, $m = 2$, $a(z;\theta) = 1 + \theta_2^2 - \theta_1 z_1 - \theta_2(z_2 + z_2^{-1})$, $b(z;\theta) \equiv 1$, since

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

is unilateral. (The same multilateral model was also considered by Jain (1981), but the unilateral form that he derived, using a different approach, 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 Whittle (1954) 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.

Though we do not assume that the model of interest to the practitioner is multilateral, our approach to asymptotic inference is influenced by this possibility. Following Whittle (1954), lattice multilateral models driven by white noise, such as (1.3), have been discussed by, for example, Ali (1979), Besag (1974), Cliff and Ord (1981), Cressie (1993), Gleeson and McGilchrist (1980), Guyon (1995), Haining (1978), Jiming (1991b), Mardia and Marshall (1984), Moran (1973), Ranneby (1982). The allowance in (1.3) for the $a_j(\theta)$ to $b_j(\theta)$ to depend on a vector $\theta$ 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 Ali (1979)) can be physically natural, and can lead to $a(z;\theta)$ or $b(z;\theta)$ being real-valued, as with (3.1) of Section 3 below. More generally, inequality restrictions, for example asserting that the coefficient of $x_{t+1}$ is no less than that of $x_{t-1}$, are easily enforced in estimation and even when arbitrary are less drastic than choosing the direction of a unilateral model. The structure of Martin (1979), in which $h=d$ in (1.7) and $a_j(z;\theta)$ varies with $z_j$ only, can reduce the identification problem to the familiar one when $d=1$. Isotropic assumptions (see e.g. Stein (1999)) 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.

Consider estimation of $\theta_0$ for $x_t$ observed on the rectangular lattice $\mathbb{N}=\{t:-n_{Li}\leq t_i\leq n_{Ui},\ i=1,...,d\}$, for $n_{Ui},n_{Li}\geq 0$, $i=1,...,d$. 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 $n$ sequences in $\mathbb{Z}_+ = \{1, 2, ...\}$. Domains of observation are often more realistically viewed as bounded, where "infill" asymptotics (see Cressie (1993), Stein (1999)) may have more appeal. This would also require either modelling $x_t$ continuously across the domain, or making the model $n$-dependent; our goal is to provide some justification for useful rules of inference in finite samples, rather than explore issues of interpolation. Introduce assumption

A1. *For all sufficiently large $n$, there exist $\xi > 0$, $c_1 > 0$ such that*

$$n_i(n) \geq c_1 n^{\xi}, \; i = 1, ..., d. \tag{1.8}$$

The inequality between arithmetic and geometric means indicates that

$$\sum_{i=1}^{d} n_i^{-1}(n) \geq d n^{-1/d}, \tag{1.9}$$

so that $\xi \leq 1/d$, 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 usual random fields prescription $n_{Li} \equiv 0$ is included. It might sometimes seem artificial to suppose that further sampling is only possible in particular directions, and multilateral increase seems a more natural asymptotic regime when multilateral modelling is attempted.

We say that an estimate $\hat{\theta}$ of $\theta_0$ satisfies Property E if $n^{1/2}(\hat{\theta} - \theta_0)$ converges in distribution to a $\mathcal{N}(0, \Phi^{-1}\Psi\Phi^{-1})$ variate, where $\Phi$ and $\Psi$ are non-singular matrices given by

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

the prime denoting transposition and $\kappa$ as defined in assumption

A2. *$x_t$ has representation*

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

*where the $\varepsilon_j$ satisfy (1.4) and are also independent and identically distributed with finite fourth cumulant, denoted $\kappa$, and $\sum_j$ denotes $\sum_{j\in\mathbb{Z}^d}$.*

If $x_t$ has an ARMA representation (1.3), (1.5), with the $\varepsilon_t$ as in (1.4) and A2, the rest of A2 holds because $b(E(i\lambda);\theta_0)/a(E(i\lambda);\theta_0)$ is an analytic function of $\lambda$, and thus has absolutely convergent multiple Fourier series. Mixing conditions have been popular in asymptotic theory for random fields; though $\alpha$- and $\beta$-mixing can sometimes be checked, they are likely to strengthen the moment condition in A2, and given our focus on linear modelling, we prefer to strengthen assumptions on the white noise innovations, as in A2.

Inspection of much real data suggests trending in mean and/or variance across the domain, as Cressie (1993) has argued. Accounting for spatial correlation by means of a parametric model is relevant to efficient trend estimation, especially when data are in limited supply, and aspects of the methods and theory of stationary multilateral models are extendable to many such nonstationary ones. Nonstationarity analogous to unit roots in time series leads to a different type of theory, see Künsch (1987), Bhattacharya, Richardson and Franklin (1997), Baran, Pap and van Zuijlen (2002).

The spatial literature has discussed the original, continuous-frequency, form of estimate proposed by Whittle, in relation to Property E. Define

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

for an 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 $1 - n_i \leq j \leq n_i - 1$, $i = 1, ..., d$, $\sum_{t(j)}$ is a sum over $-n_{Li} \leq t_i$, $t_i + j_i \leq n_{Ui}$, $i = 1, ..., d$, 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. With a finite AR model, $Q_{C1}(\theta; I)$ and its derivatives in $\theta$ are easily analytically evaluated as a linear combination 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 an element of $\theta$ functionally unrelated to the remainder, but in multilateral models it in general depends on the whole of $\theta$, and does not have a neat closed form; even in quite simple models, Whittle (1954) found only infinite series representations, and individual terms of this can be complicated. Yao and Brockwell (2002) 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 e.g. Ali, 1979).

A statistical drawback of $\hat{\theta}_C(I)$ noted by Guyon (1982) is the edge effect: for fixed $j$, as the $n_i \to \infty$ the bias of $c_j$ for $\gamma_j = cov(x_0, x_j)$ is of order $\sum_{i=1}^d n_i^{-1}$, which by (1.9) is of order no less than $n^{-1/d}$. As (1.8) suggests, $\hat{\theta}_C(I)$ is $n^{\xi}$-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 $d \geq 3$ $\hat{\theta}_C(I)$ is never $n^{\frac{1}{2}}$-consistent; thus for $d \geq 2$ $\hat{\theta}_C(I)$ lacks Property E.

The computational drawbacks of $\hat{\theta}_C(I)$ can be avoided by extending the discrete form of Whittle estimate considered by Hannan (1973) in the time series case $d = 1$.

Define

$$\begin{aligned} Q_{D1}(\theta) &= \frac{1}{n}\sum_{j\in\mathbb{N}} \log f(\omega_j;\theta), \quad Q_{D2}(\theta;h) = \frac{1}{n}\sum_{j\in\mathbb{N}} \frac{h(\omega_j)}{f(\omega_j;\theta)}, \\ Q_D(\theta;h) &= Q_{D1}(\theta) + Q_{D2}(\theta;h), \quad \hat{\theta}_D(h) = \arg\min_{\Theta} Q_D(\theta;h), \end{aligned}$$

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 $\omega_j$ 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 produce asymptotic bias. $Q_D$ is motivated by models in which $f(\lambda;\theta)$ has a simple closed form. This is not always the case; for example Whittle (1954, 1963), Mardia and Marshall (1984), Stein (1999) 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 $f(\lambda;\theta)$; the infinite series can be truncated but at cost of asymptotic bias unless the truncation rule is suitably $n$-dependent. However, in view of (1.5), $Q_D$ is convenient in case of, for example, multilateral ARMA models, as well as ARMA-signal-plus-ARMA-noise ones, also motivated by Whittle (1954). Unlike when $d = 1$, these signal-plus-noise processes do not necessarily have a finite ARMA representation, because a non-negative multivariate trigonometric polynomial cannot necessarily be factored (see Kashyap, 1984). Likewise Rosanov (1967) motivated reciprocals of such polynomials as models for $f(\lambda)$ without requiring an AR representation. Kent and Mardia (1996) discussed an objective function based on a matrix which would be the covariance 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.

The same edge-effect bias is found in $\hat{\theta}_D(I)$ as in $\hat{\theta}_C(I)$, with respect to which

Guyon (1982) 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, Guyon (1982) showed that $\hat{\theta}_C(I_*)$ satisfies Property E, thereby avoiding edge-effect bias under a short range dependence condition similar to A2; Heyde and Gay (1993) similarly covered long range dependent models. Dahlhaus and Künsch (1987) 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 $d = m = 1$, $x_t = \theta x_{t-1} + \varepsilon_t$, $|\theta| < 1$, 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 $(1 - \theta^2)^{-2}$ in the Gaussian case), $\hat{\theta}_D(I_*)$ is not even consistent. 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 Hannan (1973) in case $d = 1$. These observations may explain the large numerical discrepancy between $\hat{\theta}_C(I_*)$ and $\hat{\theta}_D(I_*)$ found by Mardia and Marshall (1984).

Yao and Brockwell (2002) handled the edge effect in their Gaussian pseudo-likelihood by trimming out observations near the edges, thereby retaining the non-negativity of the objective function. Dahlhaus and Künsch (1987) 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 $\theta$) are always non-negative. They showed that, for $d \le 3$ 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. No doubt the same desirable properties hold for $\hat{\theta}_D(I_T)$. It seems from their proof that Dahlhaus and Künsch's requirement that the $n_i$ increase at the same rate ($\xi = 1/d$ for $d \leq 3$) can be relaxed to taking $\xi \geq \frac{1}{4}$ in (1.8), and perhaps their result can be further improved, covering also $d \geq 4$, if a smoother taper is employed, though this is liable to make the choice of bandwidth a more delicate issue, and the need to choose both a taper and a bandwidth introduces some ambiguity for the practitioner.

We propose an estimate of $\theta_0$ that enjoys some computational advantages of discrete-frequency Whittle and achieves Property E, without tapering, in a quite general class of processes that includes ARMA ones and ones in which autocorrelation falls off more slowly, while falling short of long range dependence. All $d$ are covered, with arbitrary relative rates of increase of the $n_i$ subject to A1. The function $Q_D(\theta; I)$ is first numerically optimized, and then finitely many iterations based on a suitably modified objective function are carried out. The strategy is described in the following section, along with regularity conditions and statement of asymptotic properties, with a small Monte Carlo study of finite sample performance reported in Section 3. Section 2 also proposes a consistent, guaranteed non-negative definite, estimate of the limiting covariance matrix $\Phi^{-1}\Psi\Phi^{-1}$ when $x_t$ can be non-Gaussian. Proofs are included in Sections 4 and 5. Though we are motivated in part by multilateral representations, our work also offers something new for inference on unilateral ones.

## 2. MAIN RESULTS

We introduce first a truncated version of $I_*(\lambda)$,

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

where $g(x)$ satisfies assumption

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

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

*and for all $x > 0$*

$$g(x) \leq c_2 x, \quad \text{some } c_2 < 1.$$

When averaged over the $\omega_j$, $I_g$ is immune to the aliasing problems affecting $I_*$. 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 that is 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. The aliasing can alternatively be avoided without truncating but instead evaluating $I_*$ over a finer grid of frequencies, 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 by Dahlhaus and Künsch (1987) and we do not discuss $\hat{\theta}_D(I_g)$. Theorem 5 of Robinson (1988) suggests that finitely many Newton iterations, based on $Q_D(\theta; I_g)$ and commencing from an $n^\zeta$-consistent estimate, for any $\zeta \in (0, \frac{1}{2}]$, will produce an estimate with Property E. His results built on development by Hosoya and Taniguchi (1982) and others of LeCam's (1956) observation that a single Newton step can convert an $n^{\frac{1}{2}}$-consistent estimate into an asymptotically efficient one. The choice of initial estimate is addressed subsequently.

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 possible recursions. For $\ell = 1, 2$, given an initial estimate $\hat{\theta}_{[1]}^{(\ell)}$ of $\theta_0$, define

$$\hat{\theta}_{[u+1]}^{(1)} = \hat{\theta}_{[u]}^{(1)} + R\left(\hat{\theta}_{[1]}^{(1)}\right)^{-1} r\left(\hat{\theta}_{[u]}^{(1)}\right), \quad u \geq 1, \tag{2.1}$$

$$\hat{\theta}_{[u+1]}^{(2)} = \hat{\theta}_{[u]}^{(2)} + R\left(\hat{\theta}_{[u]}^{(2)}\right)^{-1} r\left(\hat{\theta}_{[u]}^{(2)}\right), \quad u \geq 1. \tag{2.2}$$

Thus, $\left\{\hat{\theta}_{[u]}^{(1)}\right\}$ 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 iteration famously numerically converges faster, in a suitable neighbourhood of the target, and Robinson (1988) showed that this can be matched by a faster statistical convergence. However, he 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, it is Property E that 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 $\xi$ 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 $\theta_0$, $f(\lambda; \theta)$ is positive and thrice boundedly differentiable in $\theta$; $f(\lambda; \theta)$ and its first three derivatives in $\theta$* are *continuous in $\lambda$ at $\theta = \theta_0$.*

A6. $\Phi$ *is positive definite.*

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

Assumption A4 controls the bias. For ARMA models (1.3), $f(\lambda)$ is analytic so the $\gamma_j$ 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 after

**Theorem 1** *Under Assumptions A1-A7:*

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

$$u > (2\zeta)^{-1}; \tag{2.3}$$

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

$$u > \frac{\ell n \zeta}{\ell n(\frac{1}{2})}. \tag{2.4}$$

The proof is left to Section 4. It follows from the inequality $x^x > (\frac{1}{2})^{\frac{1}{2}}$ for $0 < x < \frac{1}{2}$ that (2.1) requires at least as many iterations as (2.2), reflecting the anticipated benefit of updating $R$ in (2.2).

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 $j \in \mathbb{N}$. Indeed, in the AR case of (1.3) with $a(z;\theta)$ linear in $\theta$, $Q_D(\theta; I)$ 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 first introduce the following additional assumptions.

A8. $\Theta$ *is a compact subset of* $R^m$*.*

A9. $\theta_0$ *is an interior point of* $\Theta$*.*

A10. $f(\lambda;\theta) \neq f(\lambda;\theta_0)$, $\theta \in \Theta - \{\theta_0\}$, *for all* $\lambda$ *in a subset of* $\Pi^d$ *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,*

$$\hat{\theta}_D(I) - \theta_0 = O_p(n^{-\xi}), \ \ as\ n \to \infty. \tag{2.5}$$

The rate in (2.5) was anticipated in Section 1, but in view of A7 it seems desirable to state formal justification, especially as we later discuss a modified estimate. Nevertheless, Theorem 2 relates closely to results of Guyon (1982), Kent and Mardia (1996) pertaining to $\hat{\theta}_C(I)$ for unilateral models with $d > 1$, and Hannan (1973) for $\hat{\theta}_D(I)$ when $d = 1$, so we only comment briefly on the proof. Consistency, with no rate, may be established much as in Hannan's proof, 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 $\theta_0$, 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 (4.17) of Section 4). Note that A11 is milder than A4, and could be relaxed at cost of a slower rate than in (2.5), and possibly an increase in the number of recursions needed to achieve Property E.

When the $n_i$ increase at the same rate, $\xi = 1/d$, and Table 1 indicates the minimal values of $u$, $u(1)$ and $u(2)$, satisfying (2.3) and (2.4) when $\hat{\theta}^{(\ell)}_{[1]} = \hat{\theta}_D(I)$ for $\ell = 1, 2$. For the practically most typical $d$, $\hat{\theta}^{(1)}_{[u]}$ dominates on computational grounds. On the other hand 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. If relative rates of the $n_i$, or at least $\zeta$, are not assumed known, then the $u(\ell)$ are unknown, albeit finite.

**Table 1:**

Minimum values $u(\ell)$, $\ell = 1, 2$, of $u$ satisfying (2.3) and (2.4) 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 is regarded as becoming suitably finer as $n$ increases. Robinson (1988) 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 $n^{\psi}$-consistent estimate, for $\psi \leq \frac{1}{4}$. To develop a corresponding approximation to $\hat{\theta}_D(I)$, define by $G_n$ a set of points that is regularly-spaced throughout $\Theta$, and such that $\#\{\theta : \theta \in G_n\} \geq c_3 n^{m\psi}$, $c_3 > 0$, 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,*

$$\hat{\theta}_D^{(s)}(I) - \theta_0 = O_p(n^{-\psi}), \quad \text{as } n \to \infty, \tag{2.6}$$

*for* $\psi \leq \xi/2$.

We omit the proof because it largely applies Theorem 8 of Robinson (1988), whose conditions are checkable much as would be done in proving Theorem 2. His conditions would require 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 (2.6) that emerges by following his method of proof.

The strategy justified in Theorems 1 and 3 stresses statistical and computational considerations to demonstrate that Property E can be achieved 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 (2.1) or (2.2) to achieve satisfactory numerical convergence, would obviously be desirable.

When $x_t$ is Gaussian, estimates satisfying Property E are asymptotically efficient, and have limiting variance matrix $2\Phi^{-1}$, since $\kappa = 0$. Then Theorem 1 can be applied in approximate inference on $\theta_0$ by consistently estimating $\Phi$ by $\hat{\Phi} = R(\hat{\theta})$, where $\hat{\theta}$ is any consistent estimate of $\theta_0$. More generally, if we can partition $\theta$ in the ratio $m_a : m_b$ as $\theta = (\theta_a', \theta_b')'$, and correspondingly $\partial(\lambda; \theta) = (\partial_a(\lambda; \theta)', \partial_b(\lambda; \theta)')'$, 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 $\Phi$ (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 $(2\pi)^{-d} \int_{\Pi^d} \log f(\lambda; \theta) d\lambda = \log \theta_b$, but not in non-standard parameterizations, such as signal-plus-noise and multilateral models, as the discussion of $Q_{C1}(\theta)$ in Section 1 suggests. Here, asymptotic inference requires consistently estimating $\Psi$, for which several approaches have been suggested in case $d = 1$.

For unilateral models, Hannan, Dunsmuir and Deistler (1980) proposed a consistent estimate of $\Psi$, involving time-domain filtering, that is advantageously guaranteed to be non-negative definite (nnd), but seems difficult to extend to multilateral spatial models. Taniguchi's (1982) frequency-domain proposal, 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 $\rho$ is a continuous function on $\Pi^2$, does seem to be extendable to our context, 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 $\Psi$ does not seem to be necessarily nnd. Chiu (1988) 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 $\rho$ now a continuous function on $\Pi$, 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 $\left\{ n^{-1} \sum_{j\in\mathbb{N}} \rho(\omega_j) I(\omega_j) \right\}^2 \to_p \left\{ (2\pi)^{-1} \int_{\Pi} \rho(\lambda) f(\lambda) d\lambda \right\}^2$.

We propose an alternative approach, that would be useful also in time series problems and applies also to long range dependent processes. Since $\Phi$ is consistently estimated by $\hat{\Phi}$, and $\Xi$ by $\hat{\Xi} = n^{-1} \sum_{j\in\mathbb{N}} \partial(\omega_j;\hat{\theta})$, it suffices, according to the form of $\Psi$ (which is due to the linearity assumption A2), to estimate $\kappa$. Given $\hat{\varepsilon}_t$, $t \in \mathbb{N}$, introduce

$$\hat{\mu}_2 = n^{-1} \sum_{t\in\mathbb{N}} \hat{\varepsilon}_t^2, \quad \hat{\mu}_4 = n^{-1} \sum_{t\in\mathbb{N}} \hat{\varepsilon}_t^4. \tag{2.7}$$

The simplest estimate of $\kappa$ is $\tilde{\kappa} = \hat{\mu}_4 - 3$, but $2\hat{\Phi} + \tilde{\kappa}\hat{\Xi}\hat{\Xi}'$ is not necessarily nnd. 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 nnd, 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 $\Psi$ if $\hat{\mu}_2$ and $\hat{\mu}_4$ are consistent for $E\varepsilon_0^2$ and $E\varepsilon_0^4$ (explaining the introduction of $\hat{\mu}_2$ despite $E\varepsilon_0^2 = 1$ 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 (1.1) 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 Robinson (1987) (intended for unilateral models with $d = 1$) which assumes we know a function $\alpha(z;\theta)$ of $z$ and $\theta$ such that $f(\lambda;\theta) = (2\pi)^{-d} \left|\alpha\left(E(i\lambda);\theta\right)\right|^{-2}$; for example in the ARMA model (1.3), $\alpha(z;\theta) = a(z;\theta)/b(z;\theta)$. Define $w(\lambda) = \{(2\pi)^d n\}^{-\frac{1}{2}} \sum_{t\in\mathbb{N}} x_t e^{it.\lambda}$ and

$$\hat{\varepsilon}_t^{(2)} = (2\pi)^{d/2} n^{-\frac{1}{2}} \sum_{j\in\mathbb{N}} \alpha\left(E(i\omega_j);\hat{\theta}\right) w(\omega_j) e^{-it\cdot\omega_j}, \quad t \in \mathbb{N}. \tag{2.8}$$

When expressed in the time domain, (2.8) effectively treats $x_t$ on $\mathbb{Z}^d$ as a circulant, with observations on $\mathbb{N}$ repeated periodically. This violates our assumptions, but we show that, as with $\hat{\theta}_D(I)$, the consequent error is asymptotically negligible, and (2.8) is computationally advantageous when $\alpha$ is a simple function, as in ARMA models, and in making double use of the fast Fourier transform. Robinson (1987) 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* $\lambda \in \Pi^d, \alpha\left(E(i\lambda);\theta\right)$ *is boundedly differentiable in a neighbourhood of* $\theta_0$*, it is nonzero 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* $\varepsilon_t$ *are independent with zero mean, unit variance and uniformly bounded fourth moment.*

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

Unlike in the estimation of $\theta_0$, assumption A12 implies knowledge of a factorization of $f(\lambda;\theta)$. However, 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 $\theta$, as well as for many processes with long range dependence; there, the summability of $\beta_j$ 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 $\theta_0$, though the $\hat{\theta}_{[1]}^{(\ell)}$ in A7 also satisfy A13 if $\zeta = 1/d$ for $d \leq 3$. Proof details of the following theorem are left to Section 5.

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

$$\hat{\mu}_2^{(i)} \to_p E\varepsilon_0^2, \quad \hat{\mu}_4^{(i)} \to_p E\varepsilon_0^4, \quad i = 1, 2. \tag{2.9}$$

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

$$2\Phi^{-1} + \left(\hat{\mu}_4^{(i)} - \hat{\mu}_2^{(i)2} - 2\right)\left(\hat{\Phi}^{-1}\hat{\Xi}\right)\left(\hat{\Phi}^{-1}\hat{\Xi}\right)', \quad i = 1, 2, \tag{2.10}$$

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

## 3. MONTE CARLO STUDY OF FINITE-SAMPLE BEHAVIOUR

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

$$x_t = \sigma_0\varepsilon_t + \rho_0\sigma_0 \sum_{\substack{j_1=-1 \\ j\neq(0,\ldots,0)}}^{1} \cdots \sum_{j_d=-1}^{1} \varepsilon_{t-j}. \tag{3.1}$$

This is an MA $(1,1;...;1,1)$ representation defined as in Section 1 with $a(z;\theta) \equiv 1$, $b_j(\theta) = \sigma$ for $j = (0,...,0)$, $b_j(\theta) = \sigma\rho$ for $j = (\pm 1,...,\pm 1)$, and $b_j(\theta) \equiv 0$ otherwise, taking $\theta = (\rho, \theta)'$. Haining (1978) discussed a similar model. We deduce that

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

where

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

An "invertibility" condition satisfying (1.5) is

$$|\rho_0| < \left(3^d - 1\right)^{-1}. \tag{3.2}$$

For given $n^*$, we generated NID$(0,1)$ $\varepsilon_t$ for $t_\ell = 0, \pm 1, ..., \pm(n^* + 1)$, $\ell = 1, ..., d$, and then $x_t$ $t \in \mathbb{N} = \{t : t_\ell = 0, \pm 1, ..., \pm n^*, \ell = 1, ..., d\}$, using (3.1). 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 (3.2) is satisfied.

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

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

where

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

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

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

indicating equally-spaced points over the set (3.2). 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 $\sigma$ we can get the $n^{1/(2d)}$-consistency of $\hat{\theta}_D^{(s)}(I)$ in the statement of Theorem 3 by searching over $G_n^{(d)}$.

Both sequences of iterations (2.1) and (2.2) were pursued. Property E is first achieved by $\hat{\rho}_{[3]}^{(1)}$ and $\hat{\rho}_{[3]}^{(2)}$ for $d = 2$, and by $\hat{\rho}_{[4]}^{(1)}$ and $\hat{\rho}_{[3]}^{(2)}$ 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 (2.1) and (2.2) 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 $\rho$ 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 (2.1) and (2.2) when $d = 3$ and $n = 343$.

**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) |

**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) |

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

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

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 (2.1) and (2.2). 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.

**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) |
| $\hat{\rho}_{[5]}^{(1)}$ | .0022 (.0104) | .0044 (.0129) | .0005 (.0066) | .0006 (.0060) |
| $\hat{\rho}_{[4]}^{(2)}$ | .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) |
| $\hat{\rho}_{[5]}^{(1)}$ | -.0024 (.0125) | .0020 (.0155) | .0010 (.0072) | .0004 (.0072) |
| $\hat{\rho}_{[4]}^{(2)}$ | -.0031 (.0128) | .0028 (.0167) | .0011 (.0075) | .0005 (.0071) |

## 4. 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

$$\hat{\theta}^{(\ell)}_{[u]} - \hat{\theta} = o_p(n^{-\frac{1}{2}}), \quad \ell = 1, 2, \tag{4.1}$$

when $u$ satisfies (2.3) for $\ell = 1$ and (2.4) for $\ell = 2$.

The first statement will follow on showing

$$n^{\frac{1}{2}} r(\theta_0) \to_d \mathcal{N}(0, \Psi) \tag{4.2}$$

and

$$R(\theta_0) \to_p \Phi. \tag{4.3}$$

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

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

where $I_m$ is the $m$-rowed identity matrix and $\tilde{S}^{(\ell)}_{[u]}$ is the matrix obtained by evaluating each row of $S(\theta) = (\partial/\partial\theta')\, r(\theta)$ at a point on the line segment between $\hat{\theta}^{(\ell)}_{[u]}$ and $\theta_0$. On showing

$$\begin{aligned}
\left\|R\left(\tilde{\theta}^{(\ell)}_{[u]}\right)^{-1} - R(\theta_0)^{-1}\right\| &= O_p\left(\left\|\tilde{\theta}^{(\ell)}_{[u]} - \theta_0\right\|\right), && (4.4) \\
\left\|I_m + R\left(\tilde{\theta}^{(\ell)}_{[u]}\right) \tilde{S}^{(\ell)}_{[u]}\right\| &= O_p\left(\left\|\tilde{\theta}^{(\ell)}_{[u]} - \theta_0\right\| + n^{-\frac{1}{2}}\right), && (4.5)
\end{aligned}$$

where $\|A\| = \{tr(AA')\}^{\frac{1}{2}}$, 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 Robinson (1988) we have the solutions

$$\hat{\theta}^{(1)}_{[u+1]} - \hat{\theta}_0 = O_p\left(\left\|\hat{\theta}^{(1)}_{[1]} - \theta_0\right\|^{u+1}\right) + o_p(n^{-\frac{1}{2}}) = O_p\left(n^{-(u+1)\xi}\right) + o_p(n^{-\frac{1}{2}}),$$

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

whence (4.1) holds under (2.3) and (2.4) respectively.

The proof of (4.4) involves standard application of the mean value theorem, given A5, A6 and (4.3), which follows immediately from continuity of $\partial(\lambda;\theta_0)$. The proof of (4.5) 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 (4.2), 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 we 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 $\mu$ under A2. Now

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

This is bounded by

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

under A1 and A4, $K$ being a generic, positive constant. Thence $r_2 = o(n^{-\frac{1}{2}})$ and it suffices to establish (4.2) 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 $\ell = (\ell_1, ..., \ell_d)$, $A_L = \{\ell : |\ell_i| \leq L, i = 1, ..., d\}$ 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

$$\sup_\lambda |\tau(\lambda) - \tau_L(\lambda)| < \eta_1. \tag{4.6}$$

Writing

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

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

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

$$= \{(2\pi)^d n\}^{-2} \sum_{j \in \mathbb{N}} \sum_{k \in \mathbb{N}} \tilde{\tau}_L(\omega_j) \tilde{\tau}_L(\omega_k) \left\{ \sum_u{}'' \sum_v{}'' cov(c_u^*, c_v^*) e^{i(v.\omega_k - u.\omega_j)} \right\}, \tag{4.7}$$

where $\tilde{\tau}_L(\lambda) = \tau(\lambda) - \tau_L(\lambda)$ and $\sum_u'' = \sum \cdots \sum_{|u_i| \leq g(n_i)}$, $i = 1, ..., d$. The proof that (4.7) $= o(n^{-1})$ 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 (4.7) is

$$\begin{aligned}
& \sum_u{}'' \sum_v{}'' [n(u)n(v)]^{-1} \sum_{s(u)} \sum_{t(v)} \{\gamma_{t-s-u}\gamma_{t+v-s} + \gamma_{t-s}\gamma_{t-s+v-u} \\
& + cum\,(x_s, x_{s+u}, x_t, x_{t+v})\} e^{i(v.\omega_k - u.\omega_j)} \\
= \; & \sum_u{}'' \sum_v{}'' [n(u)n(v)]^{-1} \sum_{s(u)} \sum_{t(v)} \left[ \int_{\Pi^d} \int_{\Pi^d} f(\lambda) f(\chi) \right. \\
& \times \left\{ e^{i(t-s-u).\lambda - i(t+v-s).\chi} + e^{i(t-s).\lambda - i(t-s+v-u).\chi} \right\} d\lambda d\chi \\
& \left. + \kappa \sum_\ell \beta_{s-\ell} \beta_{s+u-\ell} \beta_{t-\ell} \beta_{t+v-\ell} \right] e^{i(v.\omega_k - u.\omega_j)}.
\end{aligned} \tag{4.8}$$

The contribution to (4.7) from the first term in braces in (4.8) is

$$
\begin{aligned}
& \{(2\pi)^d n\}^{-2} \int_{\Pi^d} \int_{\Pi^d} \sum_{j\in\mathbb{N}} \sum_{k\in\mathbb{N}} \tilde{\tau}_L(\omega_j) \tilde{\tau}_L(\omega_k) \sum_u{}'' \sum_v{}'' \{n(u)n(v)\}^{-1} \\
& \times e^{-iu.(\lambda+\omega_j)-iv.(\chi-\omega_k)} \sum_{s(u)} \sum_{t(v)} e^{i(t-s).(\lambda-\chi)} f(\lambda) f(\chi) d\lambda d\chi \\
= \; & \{(2\pi)^d n\}^{-2} \int_{\Pi^d} \int_{\Pi^d} \left\{ \sum_{j\in\mathbb{N}} \tilde{\tau}_L(\omega_j) \sum_u{}'' n(u)^{-1} e^{-iu.(\lambda+\omega_j)} \sum_{s(u)} e^{is.(\chi-\lambda)} \right\} \\
& \times \left\{ \sum_{k\in\mathbb{N}} \tilde{\tau}_L(\omega_k) \sum_v{}'' n(v)^{-1} e^{iv.(\omega_k-\chi)} \sum_{t(v)} e^{it.(\lambda-\chi)} \right\} f(\lambda) f(\chi) d\lambda d\chi .
\end{aligned}
$$

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

$$
\begin{aligned}
& \{(2\pi)^d n\}^{-2} \left\{ \int_{\Pi^d} \int_{\Pi^d} \left\| \sum_{j\in\mathbb{N}} \tilde{\tau}_L(\omega_j) \sum_u{}'' n(u)^{-1} e^{-iu.(\lambda+\omega_j)} \sum_{s(u)} e^{is.(\chi-\lambda)} \right\|^2 d\lambda d\chi \right. \\
& \left. \times \int_{\Pi^d} \int_{\Pi^d} \left\| \sum_{k\in\mathbb{N}} \tilde{\tau}_L(\omega_k) \sum_v{}'' n(v)^{-1} e^{iv.(\omega_k-\chi)} \sum_{t(v)} e^{it.(\lambda-\chi)} \right\|^2 d\lambda d\chi \right\}^{\frac{1}{2}} \\
= \; & n^{-2} \sum_u{}'' n(u)^{-1} \left\| \sum_{j\in\mathbb{N}} \tilde{\tau}_L(\omega_j) e^{-iu.\omega_j} \right\|^2
\end{aligned}
$$

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

$$
n^{-3} \sum_u{}'' \left\| \sum_{j\in\mathbb{N}} \tilde{\tau}_L(\omega_j) e^{-iu.\omega_j} \right\|^2 \le n^{-3} \sum_u{}''' \left\| \sum_{j\in\mathbb{N}} \tilde{\tau}_L(\omega_j) e^{-iu.\omega_j} \right\|^2 \tag{4.9}
$$

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

$$
\sum_{u_\ell=1-n_\ell}^{0} e^{2\pi i(k_\ell-j_\ell)/n_\ell} = \sum_{u_\ell=1}^{n_\ell} e^{2\pi i(k_\ell-j_\ell)n_\ell} = n_\ell 1(j_\ell = k_\ell) \tag{4.10}
$$

for $1 \le j_\ell, k_\ell \le n_\ell$, it follows that the bound in (4.9) is

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

The contribution to (4.7) from the second term in braces in (4.8) is readily found to be of the same order. The contribution to (4.7) from the fourth cumulant term in (4.8) is bounded by

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

as before.

We now wish to show that for fixed $L$

$$
n^{\frac{1}{2}} r_{1L} \to_d \mathcal{N}\left(0, \Psi_L\right), \tag{4.11}
$$

where

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

Using (4.10),

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

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,

$$
c_\ell^* - \gamma_\ell = n(\ell)^{-1} \sum_{j} \sum_{k} \beta_j \beta_k \sum_{t(\ell)} \left\{ \varepsilon_{t-j}\varepsilon_{t+\ell-k} - 1(j = k - \ell) \right\}. \tag{4.12}
$$

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

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

The difference between (4.12) and

$$\begin{aligned} q_{\ell,M} &= n(\ell)^{-1} \sum_{j,j+\ell \in A_M} \beta_j \beta_{j+\ell} \sum_{t(\ell)} \left(\varepsilon_{t-j}^2 - 1\right) \\ &+ n(\ell)^{-1} \sum_{j \in A_M} \sum_{\substack{k \in A_M \\ k \neq j+\ell}} \beta_j \beta_k \sum_{t(\ell)} \varepsilon_{t-j} \varepsilon_{t+\ell-k} \end{aligned} \tag{4.13}$$

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

$$n^{\frac{1}{2}} \sum_{\ell \in A_L} a_\ell q_{\ell,M} \tag{4.14}$$

for arbitrary $a_\ell$, 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 (see Jiming (1991b)) but on the basis of unidirectional increase, i.e. with only the $n_{Ui}$ increasing. For $k \geq 1$, 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

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

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 $\delta_n(s)$, $1 \le s \le n$, of iid variates with zero mean, variance 1 and fourth cumulant $\kappa$, such that

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

Considering now the contribution to (4.14) from the "squared" terms $\varepsilon_{t-j}^2$ in $q_{\ell,M}$,

$$\sum_{t(\ell)} \left(\varepsilon_{t-j}^2 - 1\right) \tag{4.15}$$

differs from

$$\sum_{t \in \mathbb{N}} \left(\varepsilon_t^2 - 1\right) \tag{4.16}$$

by

$$O\left(\sum_{i=1}^{d} \prod_{j=1, j\neq i}^{d} n_j\right) = O\left(n \sum_{i=1}^{d} n_i^{-1}\right) = O\left(n^{1-\xi}\right) \tag{4.17}$$

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 (4.15) and (4.16) is $O_p\left(n^{(1-\xi)/2}\right)$. As for product terms, note that in

$$\sum_{t(\ell)} \varepsilon_{t-j}\varepsilon_{t+\ell-k} \tag{4.18}$$

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 $s$, (4.18) 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 (4.14) differs by $o_p(1)$ from $n^{-\frac{1}{2}} \sum_{s=1}^{n} u_n(s)$, where

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

The $u_n(s)$ thus comprise a martingale difference array. Denote by $F_{s,n}$ the $\sigma$-field of events generated by $\delta_n(t)$, $t \leq s$. It follows from Scott (1973), Hall and Heyde (1980, Chapter 2), that if

$$\lim_{n\to\infty} n^{-1} \sum_{s=1}^{n} E u_n^2(s) \tag{4.19}$$

is positive and finite and

$$n^{-1} \sum_{s=1}^{n} E\left\{u_n^2(s) 1\left(|u_n(s)| \geq \eta_3 n^{\frac{1}{2}}\right)\right\} \rightarrow 0, \quad \text{all } \eta_3 > 0, \tag{4.20}$$

$$n^{-1} \sum_{s=1}^{n} \left[E\left\{u_n^2(s)\middle| F_{s-1,n}\right\} - E u_n^2(s)\right] \rightarrow_p 0, \tag{4.21}$$

then

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

where $\sigma^2$ is given by (4.19).

To prove (4.20) 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 $\beta_j$. For $i = 2$ it follows from the same facts after applying Cauchy and elementary inequalities.

Next consider (4.21), which is equivalent to

$$\begin{aligned} & n^{-1}\sum_{s=1}^{n}\left[\left\{\sum_{\ell\in A_L} a_\ell \frac{n}{n(\ell)} \sum_{j\in A_M} \sum_{\substack{k\in A_M \\ k\neq j+\ell}} \beta_j\beta_k\delta_n\left(s - r_{sn}(j,k,\ell)\right)\right\}^2\right. \\ & \left. -E\left\{\sum_{\ell\in A_L} a_\ell \frac{n}{n(\ell)} \sum_{j\in A_M} \sum_{\substack{k\in A_M \\ k\neq j+\ell}} \beta_j\beta_k\delta_n\left(s - r_{sn}(j,k,\ell)\right)\right\}^2\right] \\ & +2E\varepsilon_0^3 n^{-1}\sum_{s=1}^{n}\left\{\sum_{\ell\in A_L} a_\ell \frac{n}{n(\ell)} \sum_{j,j+\ell\in A_M} \beta_j\beta_{j+\ell}\right\} \qquad (4.22) \\ & \times\left\{\sum_{\ell\in A_L} a_\ell \frac{n}{n(\ell)} \sum_{j\in A_M} \sum_{\substack{k\in A_M \\ k\neq j+\ell}} \beta_j\beta_k\delta_n\left(s - r_{sn}(j,k,\ell)\right)\right\} \to_p 0 \end{aligned}$$

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

$$n^{-1}\sum_{s=1}^{n}\left\{\delta_n(s - r_{sn1})\delta_n(s - r_{sn2}) - E\delta_n(s - r_{sn1})\delta_n(s - r_{sn2})\right\} \qquad (4.23)$$

where $r_{sni} = r_{sn}\left(j_{(i)}, k_{(i)}, \ell_{(i)}\right)$. Now (4.23) has mean zero and variance

$$\begin{aligned} & n^{-2}\sum_{s=1}^{n}\sum_{t=1}^{n}[E\delta_n(s - r_{sn1})\delta_n(t - r_{tn1})E\delta_n(s - r_{sn2})\delta_n(t - r_{tn2}) \\ & +E\delta_n(s - r_{sn1})\delta_n(t - r_{tn2})E\delta_n(s - r_{sn2})\delta_n(t - r_{tn1}) \qquad (4.24) \\ & +cum\left\{\delta_n(s - r_{sn1}), \delta_n(t - r_{tn1}), \delta_n(s - r_{sn2}), \delta_n(t - r_{tn2})\right\}]. \end{aligned}$$

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 (4.17), 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 $|s - t| \leq n^{1-\xi/2}$ and one terms for which $|s - t| > n^{1-\xi/2}$ the first component contributes $O(n^{-\xi/2})$ to (4.24), and the second, zero. Since only finitely many terms of form of (4.23) are involved, and because clearly $n^{-1} \sum_{i=1}^n \delta_n (s - r_{sn}(j, k, \ell)) = O_p(n^{-\frac{1}{2}})$, (4.22) is established.

We can evaluate (4.19) as

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

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 (4.11) via Bernstein's lemma. From (4.6), $\Psi_L \to \Psi$ as $L \to \infty$, so we then likewise deduce (4.2). □

## 5. PROOF OF THEOREM 4

Given (2.9), we have already justified the claims about (2.10), and for (2.9) 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 $x^4 - y^4 = (x-y)(x^3 + x^2 y + xy^2 + y^3)$ and Hölder's inequality the first term is $o_p(1)$ if

$$n^{-1} \sum_{t \in \mathbb{N}} \left( \hat{\varepsilon}_t^{(2)} - \varepsilon_t \right)^4 \to_p 0. \tag{5.1}$$

Write

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

where

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

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

By direct calculation, using (4.10) 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 $\beta_j$ possibly being only square-summable. Nevertheless,

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

Thus

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

It follows that

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

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

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

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

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

where $\mathbb{N}_2=\{j:|j_\ell|\leq 2n_\ell,\ell=1,...,d\}$. 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

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

by summability of $\alpha_j$.

Finally,

$$n^{-1}\sum_{t\in\mathbb{N}}E_t^4 \leq n^{-1}\left(\sum_{t\in\mathbb{N}}E_t^2\right)^2 \tag{5.2}$$

and from (4.10)

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

with probability approaching 1 as $n \to \infty$, in view of A12 and A13. Then (5.2) $= O_p(n^{1-4\zeta}) = o_p(1)$ for $\zeta > \frac{1}{4}$. This completes the proof of (5.1). $\square$

## REFERENCES


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

[2] BARAN, S., PAP, G. and van ZUIJLEN, M. (2002). Asymptotic inference for an unstable spatial AR model. Preprint.

[3] BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with Discussion). *J.R. Statist. Soc.* Ser. B **36**, 192-236.

[4] BHATTACHARYA, B.B., RICHARDSON, G.D. and FRANKLIN, L.A. (1997). Asymptotic inference for near unit roots in spatial autoregression. *Ann. Statist.* **25**, 1709-1724.

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

[6] CLIFF, A.D. and ORD, J.K. (1981). *Spatial Processes: Models & Applications.* Pion, London.

[7] CRESSIE, N. (1993). *Statistics for Spatial data*, 2nd ed. Wiley, New York.

[8] DAHLHAUS, R. and KÜNSCH, H. (1987). Edge effects and efficient parameter estimation for stationary random fields. *Biometrika* **74**, 877-882.

[9] GLEESON, A.C. and McGILCHRIST, C.A. (1980). Bilateral processes on a rectangular lattice. *Austral. J. Statist.* **22**, 197-206.

[10] GUYON, X. (1982). Parameter estimation for a stationary process on a $d$-dimensional lattice. *Biometrika* **69**, 95-106.

[11] GUYON, X. (1995). *Random Fields on a Network.* Springer-Verlag, New York.

[12] HAINING, R. (1978). The moving average model for spatial interaction. *Trans Inst. Brit. Geographers* **3**, 202-225.

[13] HALL, R. and HEYDE, C.C. (1980). *Martingale Limit Theory and its Applications.* Academic Press, New York.

[14] HANNAN, E.J. (1973). The asymptotic theory of linear time series models. *J. Appl. Probab.* **10**, 130-145.

[15] HANNAN, E.J., DUNSMUIR, W.T.M. and DEISTLER, M. (1980). Estimation of vector ARMAX models. *J. Multivariate Anal.* **10**, 275-295.

[16] HELSON, H. and LOWDENSLAGER, D. (1958). Prediction theory and Fourier series in several variables. *Acta Math.* **99**, 165-202.

[17] HEYDE, C.C. and GAY, R. (1993). Smoothed periodogram asymptotics and estimation for processes and fields with possible long range dependence. *Stoch. Proc. Appl.* **45**, 169-182.

[18] HOSOYA, Y. and TANIGUCHI, M. (1982). A central limit theorem for stationary processes and the parametric estimation of linear processes. *Ann. Statist.* **10**, 132-153.

[19] HUANG, D. (1992). Central limit theorem for two-parameter martingale differences with application to stationary random fields. *Science in China* **35**, 413-425.

[20] HUANG, D. and ANH, V.V. (1992). Estimation of spatial ARMA models. *Austral. J. Statist.* **34**, 513-530.

[21] JAIN, A.K. (1981). Advances in mathematical models for image processing. *Proc. IEEE* **69**, 502-528.

[22] JIMING, J. (1991a). Parameter estimation of spatial AR model. *Chin. Ann. Math.* **12B**, 432-444.

[23] JIMING, J. (1991b). Uniform convergence rate of sample ACV and ACR for linear spatial series under more general martingale condition. *Adv. Math.* **20**, 39-50.

[24] KASHYAP, R.L. (1984). Characterization and estimation of two-dimensional ARMA models. *IEEE Trans. Inform. Th.* **IT-30**, 736-745.

[25] KENT, J.T. and MARDIA, K.V. (1996). Spectral and circulant approximations to the likelihood for stationary Gaussian random fields. *J. Statist. Plann. Inf.* **50**, 379-394.

[26] KOREZLIOGLU, H. and LOUBATON, P. (1986). Spectral factorization of wide sense stationary processes on $\mathbb{Z}^2$. *J. Multivariate Anal.* **19**, 24-47.

[27] KÜNSCH, H. (1987). Intrinsic autoregressions and related models on the two-dimensional lattice. *Biometrika* **74**, 517-524.

[28] LECAM, L. (1956). On the asymptotic theory of estimation and testing hypotheses. *Proc. Third Berkeley Symp. Math. Statist. and Probab.* **1**, 129-156. University of California Press, Berkeley.

[29] MARDIA, K.V. and MARSHALL, R.J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71**, 135-146.

[30] MARTIN, R.J. (1979). A subclass of lattice processes applied to a problem in planar sampling. *Biometrika* **66**, 209-217.

[31] MORAN, P.A.P. (1973). Necessary conditions for Markovian processes on a lattice. *J. Appl. Probab.* **10**, 605-612.

[32] RANNEBY, B. (1982). Stochastic models of variation in time and space. In *Statistics in Theory and Practice: Essays in Honour of Bertil Matérn* (B. Ranneby, ed.), 227-245. Swedish University of Agricultural Sciences, Umea

[33] ROBINSON, P.M. (1987). Time series residuals with application to probability density estimation. *J. Time Series Anal.* **8**, 329-344.

[34] ROBINSON, P.M. (1988). The stochastic difference between econometric statistics. *Econometrica* **56**, 531-548.

[35] ROSANOV, Yu A. (1967). On Gaussian fields with given conditional distributions. *Probab. Th. Appl.* **12**, 381-391.

[36] SCOTT, D.J. (1973). Central limit theorems for martingales using a Skorokhod representation approach. *Adv. Appl. Probab.* **5**, 119-137.

[37] STEIN, M.L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging.* Springer-Verlag, New York.

[38] TANIGUCHI, M. (1982). On estimation of the integrals of the fourth order cumulant spectral density. *Biometrika* **69**, 117-122.

[39] TJOSTHEIM, D. (1978). Statistical spatial series modelling. *Adv. Appl. Probab.* **10**, 130-154. Correction (1984), *Adv. Appl. Probab.* **16**, 120.

[40] TJOSTHEIM, D. (1983). Statistical spatial series modelling II: some further results on unilateral lattice processes. *Adv. Appl. Probab.* **15**, 562-584.

[41] VAN DER WAERDEN, B.K. (1953). *Modern Algebra*, Vol. 1. Ungar, New York.

[42] WHITTLE, P. (1954). On stationary processes in the plane. *Biometrika* **41**, 434-449.

[43] WHITTLE, P. (1963). Topographic correlation, power-law covariance functions, and diffusion. *Biometrika* **49**, 305-314.

[44] YAO, Q. and BROCKWELL, P. (2002). Gaussian maximum likelihood estimation for ARMA models II: spatial processes. Preprint.

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

p.m.robinson@lse.ac.uk