Research Article

Yule-Walker Estimation for the Moving-Average Model

Chrysoula Dimitriou-Fakalou

Department of Statistical Science, University College London, London WC1E 6BT, UK

Correspondence should be addressed to Chrysoula Dimitriou-Fakalou, chrysoula@stats.ucl.ac.uk

The standard Yule-Walker equations, as they are known for an autoregression, are generalized to involve the moments of a moving-average process indexed on any number of dimensions. Once observations become available, new moments estimators are set to imitate the theoretical equations. These estimators are not only consistent but also asymptotically normal for any number of indexes. Their variance matrix resembles a standard result from maximum Gaussian likelihood estimation. A simulation study is added to conclude on their efficiency.

1. Introduction

At first, processes taking place on more than one dimension were considered on surfaces or in the three-dimensional space for the sake of spatial dynamics. Next, the time series Autoregressive Moving-Average model could be as well used to clothe the covariance dependence of spatial point processes. Nevertheless, the assumptions of causality and invertibility were based on conventional orderings of the vector indexes, as these were introduced by Whittle [1] and generalized by Guyon [2], and they failed to look natural on the spatial axes. The more general bilateral equations considered by Whittle [1] also related to serious problems during the estimation of the parameters. Besag's [3] automodels as well as Besag's [4] equations for observations on irregular sites moved in a new direction but the weaknesses during the estimation remained.

The inclusion of the time axis has provided more convenience for the spatial analysis that may be summarized in two ways. On the one hand when the observations are collected on irregular sites, the asymptotic properties of the estimators for the spatial dependence parameters can now be established as the number of timings only increases to infinity; for example, Dimitriou-Fakalou [5] has proposed a methodology on how to proceed. On the

other hand when the regular recordings are increasing towards all the spatial as well as the time axes, the unilateral ordering of indexes can now be arranged in a meaningful way and the causal and invertible ARMA model may be preferred as a natural spatiotemporal equation.

Special cases of the ARMA model are the pure autoregressive and moving-average equations and they both relate to useful properties as in the spectral density of interest. The autoregressive process uses a spectral density with "finite denominator", which for Gaussian processes [3] translates into a finite conditional expectation of any value based on all other values. A moving-average process uses a "finite numerator" in the spectral density, resulting in the autocovariance function with nonzero values on a finite set of "lags" only.

For a causal and finite autoregressive equation, the method of moments estimators according to the Yule-Walker equations almost coincides with the least squares or maximum Gaussian likelihood estimators and their consistency as well as their asymptotic normality may be established. Nevertheless, if the estimation for a pure moving-average or ARMA multi-indexed process takes place using its infinite autoregressive representation, a problem known as the "edge-effect" [2] resurrects and the asymptotic normality cannot be guaranteed. A modification of the Gaussian likelihood, which has been written from the autoregressive representation of each observation based on its "previous" ones, has been proposed by Yao and Brockwell [6] to confine the edge-effect with a weak condition of a finite second moment, but this works for the case of processes indexed on two dimensions only.

In this paper, for the multidimensional processes that may be clothed via an invertible and finite moving-average equation, a new method of estimation is proposed imitating the Yule-Walker equations. The finite number of nonzero autocovariances of the moving-average process is used most advantageously to eliminate the edge-effect. As a result, the proposed estimators are not only consistent but also asymptotically normal using relaxed conditions of finite second moments.

2. Theoretical Yule-Walker Equations

For $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$ and d any positive integer, real-valued processes indexed on $\mathbb{Z}^d$ will be considered, such as time series or spatial processes on one line transect ($d = 1$), spatial processes in the plane ($d = 2$) or in the three-dimensional space, as well as the combination of spatiotemporal processes on $\mathbb{Z}^2$, $\mathbb{Z}^3$, or $\mathbb{Z}^4$, respectively. For any $\mathbf{v} \in \mathbb{Z}^d$, an invertible moving-average process $\{Y(\mathbf{v})\}$ will be defined by the following equation:

$$Y(\mathbf{v}) = \varepsilon(\mathbf{v}) + \sum_{n=1}^{q} \theta_{\mathbf{i}_n} \varepsilon(\mathbf{v} - \mathbf{i}_n), \tag{2.1}$$

where $\{\varepsilon(\mathbf{v})\}$ are uncorrelated, zero-mean random variables with positive and finite variance σ^2. In (2.1) a unilateral ordering of the fixed lags $\mathbf{0} < \mathbf{i}_1 < \cdots < \mathbf{i}_q$ has been considered. For $d = 1$, this is the standard ordering of integer numbers and, for $d = 2$, the conventional unilateral ordering has been introduced by Whittle [1]. Guyon [2] has generalized the ordering of two vector indexes on $\mathbb{Z}^d$, $d \geq 3$. As for the invertibility condition, it is only imposed to make sure that it can be written $\varepsilon(\mathbf{v}) = Y(\mathbf{v}) + \sum_{\mathbf{i}>0} \Theta_{\mathbf{i}} Y(\mathbf{v} - \mathbf{i})$, $\sum_{\mathbf{i}>0} |\Theta_{\mathbf{i}}| < \infty$ also using the ordering $\mathbf{i} > \mathbf{0}$. Anderson and Jury [7] have provided a sufficient condition on the invertibility of such filters.

For the complex numbers $z_1, \ldots, z_d$ and the vector $\mathbf{z} = (z_1, \ldots, z_d)$, the polynomial $\theta(\mathbf{z}) \equiv 1 + \sum_{n=1}^{q} \theta_{\mathbf{i}_n} \mathbf{z}^{\mathbf{i}_n}$ is considered, where $\mathbf{z}^{\mathbf{i}} = \mathbf{z}^{(i_1, \ldots, i_d)} \equiv z_1^{i_1}, \ldots, z_d^{i_d}$ for $i_1, \ldots, i_d \in \mathbb{Z}$. Similarly, the backwards operators $B_1, \ldots, B_d$ and the vector backwards operator $\mathbf{B} = (B_1, \ldots, B_d)$ are considered, such that $\mathbf{B}^{\mathbf{i}} \varepsilon(\mathbf{v}) \equiv \varepsilon(\mathbf{v} - \mathbf{i})$. So (2.1) may now be rewritten as

$$Y(\mathbf{v}) = \theta(\mathbf{B})\varepsilon(\mathbf{v}). \tag{2.2}$$

Writing $\mathbf{z}^{-1} \equiv (z_1^{-1}, \ldots, z_d^{-1})$, the autoregressive sequence $\{X(\mathbf{v})\}$ may be defined from $\{\varepsilon(\mathbf{v})\}$ according to the following:

$$\theta\left(\mathbf{B}^{-1}\right) X(\mathbf{v}) \equiv X(\mathbf{v}) + \sum_{n=1}^{q} \theta_{\mathbf{i}_n} X(\mathbf{v} + \mathbf{i}_n) = \varepsilon(\mathbf{v}). \tag{2.3}$$

The auto-regression is unilateral but not causal, in the sense that it can be written that $X(\mathbf{v}) = \varepsilon(\mathbf{v}) + \sum_{\mathbf{i}>0} \Theta_{\mathbf{i}} \varepsilon(\mathbf{v}+\mathbf{i})$, $\sum_{\mathbf{i}>0} |\Theta_{\mathbf{i}}| < \infty$ as a function of the "future" terms $\varepsilon(\mathbf{v}+\mathbf{i})$, $\mathbf{i} \geq \mathbf{0}$. Moreover, writing the two polynomials

$$\gamma(\mathbf{z}) = \theta(\mathbf{z})\theta\left(\mathbf{z}^{-1}\right) \equiv \sum_{\mathbf{i} \in \mathcal{F}} \gamma_{\mathbf{i}} \mathbf{z}^{\mathbf{i}}, \qquad c(\mathbf{z}) = \gamma(\mathbf{z})^{-1} \equiv \sum_{\mathbf{i} \in \mathbb{Z}^d} c_{\mathbf{i}} \mathbf{z}^{\mathbf{i}}, \tag{2.4}$$

it holds that

$$Y(\mathbf{v}) = \gamma(\mathbf{B})X(\mathbf{v}) \quad \text{or} \quad X(\mathbf{v}) = c(\mathbf{B})Y(\mathbf{v}). \tag{2.5}$$

The set $\mathcal{F} \equiv \{\mathbf{i}_n, -\mathbf{i}_n, \mathbf{i}_n - \mathbf{i}_m, \ n, m = 1, \ldots, q\}$ in (2.4) has members the vector "lags", where $\{Y(\mathbf{v})\}$ have nonzero autocorrelations. The set $\{\mathbf{i}_n, \mathbf{i}_n - \mathbf{i}_m, \ n, m = 1, \ldots, q, n > m\}$ that uses the "positive" lags of $\mathcal{F}$ will have cardinality q^*.

The relationship in (2.5) provides the way to write Yule-Walker equations for a moving-average process. The processes $\{Y(\mathbf{v})\}$ and $\{X(\mathbf{v})\}$ are not only "caused" by the same sequence $\{\varepsilon(\mathbf{v})\}$, one from its "past" and the other from its "future" values, but also have "reverse" second-order dependence; that is, the spectral density of the auto-regression is

$$g_X(\omega_1, \ldots, \omega_d) \equiv \frac{\sigma^2}{(2\pi)^d} \frac{1}{\theta(\mathbf{z})\theta(\mathbf{z}^{-1})} = \frac{\sigma^2}{(2\pi)^d} c(\mathbf{z}),$$

$$z_k = e^{-i\omega_k}, \quad \omega_k \in (-\pi, \pi), \ k = 1, \ldots, d, \tag{2.6}$$

where according to (2.4), $E(X(\mathbf{v})X(\mathbf{v} - \mathbf{i})) = \sigma^2 c_{\mathbf{i}}$ and the spectral density of the moving-average process is

$$g_Y(\omega_1, \ldots, \omega_d) \equiv \frac{\sigma^2}{(2\pi)^d} \theta(\mathbf{z})\theta\left(\mathbf{z}^{-1}\right) = \frac{\sigma^2}{(2\pi)^d} \gamma(\mathbf{z}),$$

$$z_k = e^{-i\omega_k}, \quad \omega_k \in (-\pi, \pi), \quad k = 1, \ldots, d, \tag{2.7}$$

which also implies that $E(Y(\mathbf{v})Y(\mathbf{v} - \mathbf{i})) = \sigma^2 \gamma_{\mathbf{i}}$. While the original Yule-Walker equations for $\{X(\mathbf{v})\}$ dictate

$$c_{\mathbf{0}} + \sum_{n=1}^{q} \theta_{\mathbf{i}_n} c_{\mathbf{i}_n} = 1, \tag{2.8}$$

$$c_{\mathbf{i}} + \sum_{n=1}^{q} \theta_{\mathbf{i}_n} c_{\mathbf{i}-\mathbf{i}_n} = 0, \quad \mathbf{i} > \mathbf{0}, \tag{2.9}$$

from (2.9), the following $(q + 1)$ equations may be written for $\mathbf{i} > \mathbf{0}$, that is,

$$c_{\mathbf{i}} + \theta_{\mathbf{i}_1} c_{\mathbf{i}-\mathbf{i}_1} + \theta_{\mathbf{i}_2} c_{\mathbf{i}-\mathbf{i}_2} + \cdots + \theta_{\mathbf{i}_q} c_{\mathbf{i}-\mathbf{i}_q} = 0,$$

$$\theta_{\mathbf{i}_1} c_{\mathbf{i}+\mathbf{i}_1} + \theta_{\mathbf{i}_1}^2 c_{\mathbf{i}} + \theta_{\mathbf{i}_1} \theta_{\mathbf{i}_2} c_{\mathbf{i}+\mathbf{i}_1-\mathbf{i}_2} + \cdots + \theta_{\mathbf{i}_1} \theta_{\mathbf{i}_q} c_{\mathbf{i}+\mathbf{i}_1-\mathbf{i}_q} = 0,$$

$$\vdots \tag{2.10}$$

$$\theta_{\mathbf{i}_q} c_{\mathbf{i}+\mathbf{i}_q} + \theta_{\mathbf{i}_q} \theta_{\mathbf{i}_1} c_{\mathbf{i}+\mathbf{i}_q-\mathbf{i}_1} + \theta_{\mathbf{i}_q} \theta_{\mathbf{i}_2} c_{\mathbf{i}+\mathbf{i}_q-\mathbf{i}_2} + \cdots + \theta_{\mathbf{i}_q}^2 c_{\mathbf{i}} = 0$$

and their sum $\sum_{\mathbf{j} \in \mathscr{F}} \gamma_{\mathbf{j}} c_{\mathbf{i}-\mathbf{j}} = \sum_{\mathbf{j} \in \mathscr{F}} \gamma_{\mathbf{j}} c_{\mathbf{j}-\mathbf{i}} = \sum_{-\mathbf{j} \in \mathscr{F}} \gamma_{-\mathbf{j}} c_{-\mathbf{j}-\mathbf{i}} = \sum_{\mathbf{j} \in \mathscr{F}} \gamma_{\mathbf{j}} c_{\mathbf{j}+\mathbf{i}} = 0$, which may be rewritten as

$$\sum_{\mathbf{j}+\mathbf{i} \in \mathscr{F}} c_{\mathbf{j}} \gamma_{\mathbf{j}+\mathbf{i}} = 0, \quad \mathbf{i} \neq \mathbf{0}. \tag{2.11}$$

Further to (2.11), it holds that

$$\sum_{\mathbf{j} \in \mathscr{F}} c_{\mathbf{j}} \gamma_{\mathbf{j}} = 1 \tag{2.12}$$

from (2.8) and (2.9). It is true that (2.11) and (2.12) are equivalent to $\gamma(\mathbf{z})c(\mathbf{z}) = 1$. Like the standard Yule-Walker equations, the general versions (2.11) and (2.12) are constructed from the coefficients $\theta_{\mathbf{i}_1}, \ldots, \theta_{\mathbf{i}_q}$. As the finite number of nonzero autocovariances of $\{Y(\mathbf{v})\}$ is involved now, they will be used as theoretical prototypes to be imitated by their sample analogues for the sake of estimation of the moving-average parameters.

Finally, as the assumption of invertibility of the moving-average equation used for parameterization will be essential, it should be noted that Dimitriou-Fakalou [8] has presented in detail the merits of a causal and invertible spatiotemporal ARMA model. As a special case, the moving-average process $\{Y_t(\mathbf{s})\}$, where t is the time index and $\mathbf{s}$ are spatial indexes, may satisfy the equation $Y_t(\mathbf{s}) = \varepsilon_t(\mathbf{s}) + \sum_{i=1}^{q_t} \sum_{\mathbf{j} \in \mathscr{M}_i} \theta_{i,\mathbf{j}} \varepsilon_{t-i}(\mathbf{s} - \mathbf{j})$, where $\{\varepsilon_t(\mathbf{s})\}$ are zero mean, uncorrelated variables with identical, finite variance, q_t is a finite positive integer, $\mathscr{M}_1, \ldots, \mathscr{M}_{q_t}$ have a finite cardinality, and it holds that

$$\varepsilon_t(\mathbf{s}) = Y_t(\mathbf{s}) + \sum_{i>0} \sum_{\mathbf{j}} \Theta_{i,\mathbf{j}} Y_{t-i}(\mathbf{s} - \mathbf{j}), \qquad \sum_{i>0} \sum_{\mathbf{j}} |\Theta_{i,\mathbf{j}}| < \infty; \tag{2.13}$$

that is, the model is invertible. Note that, in either equation, no unnatural ordering is forced over space as the unilateral conventions "run" over the past timings only. Hence the model both uses a space-time unilateral ordering of indexes and is meaningful. As a result, for the special case of a stationary process with autocovariance function identified to take zero values after a reasonably small number of lags on all spatial and temporal dimensions, the invertible moving-average equation should be preferred and the new results presented next will be useful.

3. Method of Moments Estimators

For $\mathcal{S} \subset \mathbb{Z}^d$ being a set of finite cardinality N, there are available $\{Y(\mathbf{v}), \mathbf{v} \in \mathcal{S}\}$ from (2.2) with true parameters $\boldsymbol{\theta}_0 = (\theta_{\mathbf{i}_1,0}, \dots, \theta_{\mathbf{i}_q,0})^\tau$ that are to be estimated. The following condition will be assumed to be true.

Condition C1. The parameter space $\Theta \subset \mathbb{R}^q$ is a compact set containing the true value $\boldsymbol{\theta}_0$ as an inner point. Further, for any $\boldsymbol{\theta} = (\theta_{\mathbf{i}_1}, \dots, \theta_{\mathbf{i}_q})^\tau \in \Theta$ the moving-average model (2.2) is invertible.

Based on $\mathcal{S}$, a set $\mathcal{F}_{\mathbf{v}}$ will be defined for any $\mathbf{v} \in \mathbb{Z}^d$. More specifically, it will hold that $\mathbf{i} \in \mathcal{F}_{\mathbf{v}}$ only if $\mathbf{v} - \mathbf{i} \in \mathcal{S}$. Next, for the corrected set $\mathcal{S}^*$ with N^* elements, it will hold $\mathbf{v} \in \mathcal{S}^*$ if $\mathbf{v} + \mathbf{i} \in \mathcal{S}$ for every $\mathbf{i} \in \mathcal{F}$. Thus for any $\mathbf{v} \in \mathcal{S}^*$, it holds that $\mathcal{F} \subseteq \mathcal{F}_{\mathbf{v}}$. The reduction from $\mathcal{S}$ to $\mathcal{S}^*$ is essential as it will guarantee later that the edge-effect will fade away; this is because the set $\mathcal{S}^* \subset \mathcal{S}$ includes the locations $\mathbf{v}$ that have all their "neighbors" $\mathbf{v} - \mathbf{i}$ with $E(Y(\mathbf{v})Y(\mathbf{v} - \mathbf{i})) \neq 0$ being available. As the source of the edge-effect is the speed of the absolute bias to zero, $\mathcal{S}^*$ will guarantee that whatever "is missing" from the finite sample will not add at all to the bias of the estimators.

By imitating (2.11) the estimators $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_{\mathbf{i}_1}, \dots, \widehat{\theta}_{\mathbf{i}_q})^\tau$ are set to be the solutions of equations:

$$\sum_{\mathbf{v} \in \mathcal{S}^*} \left\{ \sum_{\mathbf{j} - \mathbf{i}_n \in \mathcal{F}_{\mathbf{v}}} \widehat{c}_{\mathbf{j}} Y(\mathbf{v} + \mathbf{i}_n - \mathbf{j}) \right\} Y(\mathbf{v}) \equiv 0, \quad n = 1, \dots, q, \tag{3.1}$$

where $\widehat{\theta}(\mathbf{z}) \equiv 1 + \sum_{n=1}^{q} \widehat{\theta}_{\mathbf{i}_n} \mathbf{z}^{\mathbf{i}_n}$ and $\widehat{c}(\mathbf{z}) \equiv \sum_{\mathbf{i} \in \mathbb{Z}^d} \widehat{c}_{\mathbf{i}} \mathbf{z}^{\mathbf{i}} \equiv \{\widehat{\theta}(\mathbf{z})\widehat{\theta}(\mathbf{z}^{-1})\}^{-1}$. Using Proposition 3.1, the equations

$$\sum_{\mathbf{v} \in \mathcal{S}^*} \left\{ \sum_{\mathbf{j} - (\mathbf{i}_n - \mathbf{i}_m) \in \mathcal{F}_{\mathbf{v}}} \widehat{c}_{\mathbf{j}} Y(\mathbf{v} + \mathbf{i}_n - \mathbf{i}_m - \mathbf{j}) \right\} Y(\mathbf{v}) \equiv 0, \quad n, m = 1, \dots, q, \ n > m, \tag{3.2}$$

will be used additionally, in order to make sure that the estimators are consistent. The proposition guarantees that the use of q^* equations used as prototypes for estimation can provide a unique way to compute the q parameters of the invertible equation.

Proposition 3.1. *The polynomials* $\theta_0(\mathbf{z}) = 1 + \sum_{n=1}^{q} \theta_{\mathbf{i}_n,0} \mathbf{z}^{\mathbf{i}_n}$ *and* $\theta_0(\mathbf{z})^{-1} = 1 + \sum_{\mathbf{i} > \mathbf{0}} \Theta_{\mathbf{i},0} \mathbf{z}^{\mathbf{i}}$ *with* $\sum_{\mathbf{i} > \mathbf{0}} |\Theta_{\mathbf{i},0}| < \infty$ *as well as* $\gamma_0(\mathbf{z}) = \theta_0(\mathbf{z})\theta_0(\mathbf{z}^{-1}) \equiv \sum_{\mathbf{j} \in \mathcal{F}} \gamma_{\mathbf{j},0} \mathbf{z}^{\mathbf{j}}$ *and* $c_0(\mathbf{z}) = \gamma_0(\mathbf{z})^{-1} \equiv \sum_{\mathbf{j} \in \mathbb{Z}^d} c_{\mathbf{j},0} \mathbf{z}^{\mathbf{j}}$ *are considered. For any* $\boldsymbol{\theta} = (\theta_{\mathbf{i}_1}, \dots, \theta_{\mathbf{i}_q})^\tau \in \Theta$, *such that it holds that* $\theta(\mathbf{z}) = 1 + \sum_{n=1}^{q} \theta_{\mathbf{i}_n} \mathbf{z}^{\mathbf{i}_n}$

and $\theta(\mathbf{z})^{-1} = 1 + \sum_{\mathbf{i}>0} \Theta_{\mathbf{i}} \mathbf{z}^{\mathbf{i}}$ with $\sum_{\mathbf{i}>0} |\Theta_{\mathbf{i}}| < \infty$, as well as $\gamma(\mathbf{z}) = \theta(\mathbf{z})\theta(\mathbf{z}^{-1}) \equiv \sum_{\mathbf{j}\in\mathcal{F}} \gamma_{\mathbf{j}} \mathbf{z}^{\mathbf{j}}$ and $c(\mathbf{z}) = \gamma(\mathbf{z})^{-1} \equiv \sum_{\mathbf{j}\in\mathbb{Z}^d} c_{\mathbf{j}} \mathbf{z}^{\mathbf{j}}$, the unique solution that satisfies the $q^\star$ equations $\sum_{\mathbf{j}\in\mathcal{F}} \gamma_{\mathbf{j},0} c_{\mathbf{j}-\mathbf{i}_n} = 0$, $n = 1, \ldots, q$, and $\sum_{\mathbf{j}\in\mathcal{F}} \gamma_{\mathbf{j},0} c_{\mathbf{j}-(\mathbf{i}_n-\mathbf{i}_m)} = 0$, $n, m = 1, \ldots, q$, $n > m$, is $\theta_{\mathbf{i}_n} \equiv \theta_{\mathbf{i}_n,0}$, $n = 1, \ldots, q$.

For mathematical convenience, new variables depending on the sampling set are defined $H_Y(\mathbf{v}) \equiv Y(\mathbf{v})$, $\mathbf{v} \in \mathcal{S}$ and $H_Y(\mathbf{v}) \equiv 0$, $\mathbf{v} \notin \mathcal{S}$, and (3.1) may be rewritten as

$$\sum_{\mathbf{v}\in\mathcal{S}^*} \left\{ \sum_{\mathbf{j}\in\mathbb{Z}^d} c_{\mathbf{j},0} H_Y(\mathbf{v}+\mathbf{i}_n-\mathbf{j}) \right\} Y(\mathbf{v}) - \mathbf{J}_n\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) = 0, \quad n = 1, \ldots, q, \tag{3.3}$$

or

$$\sum_{\mathbf{v}\in\mathcal{S}^*} \{c_0(\mathbf{B}) H_Y(\mathbf{v}+\mathbf{i}_n)\} Y(\mathbf{v}) - \mathbf{J}_n\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) = 0, \quad n = 1, \ldots, q, \tag{3.4}$$

where the zero subindex relates to $\boldsymbol{\theta}_0$. In (3.3) and (3.4), it is $\mathbf{J}_n = (J_{n,1}, \ldots, J_{n,q})$ with elements

$$J_{n,m} = \sum_{\mathbf{v}\in\mathcal{S}^*} \left\{ c_0(\mathbf{B}) \left(\theta_0(\mathbf{B})^{-1} H_Y(\mathbf{v}+\mathbf{i}_n-\mathbf{i}_m) + \theta_0\left(\mathbf{B}^{-1}\right)^{-1} H_Y(\mathbf{v}+\mathbf{i}_n+\mathbf{i}_m) \right) \right\} Y(\mathbf{v})$$
$$+ O_P\left(N\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2\right), \quad n, m = 1, \ldots, q. \tag{3.5}$$

The last term in (3.5) will become more obvious when the properties of the estimators will be established but for now it might be justified by a second-order Taylor expansion, where it holds

$$\frac{\left(\widehat{\theta}_{\mathbf{i}_n} - \theta_{\mathbf{i}_n,0}\right)\left(\widehat{\theta}_{\mathbf{i}_m} - \theta_{\mathbf{i}_m,0}\right)}{\left\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right\|^2} \longrightarrow \lambda_{n,m} \in (0, \infty), \quad n, m = 1, \ldots, q \tag{3.6}$$

due to the fact that $\widehat{\theta}$ will be shown to be consistent estimators in Theorem 4.1. Further, the sums over $\mathcal{S}^*$ involving the second derivatives of the Taylor expansion, when divided by N^* or N ($N^*/N \to 1$), will also tend in probability to the relevant finite expectations when $\{\varepsilon(\mathbf{v})\}$ will be assumed to be independent and identically distributed random variables with a finite variance.

Finally by imitating (2.12), the estimator of the error variance may be defined as

$$\widehat{\sigma}^2 \equiv \frac{\sum_{\mathbf{j}\in\mathcal{F}} \widehat{c}_{\mathbf{j}} \sum_{\mathbf{v}\in\mathcal{S}^*} Y(\mathbf{v})Y(\mathbf{v}-\mathbf{j})}{N^*}. \tag{3.7}$$

Example 3.2. For the simplest one-dimensional case when $q^* = q = 1$, it is demonstrated in detail how to compute the new estimator; some other cases will briefly be considered later. Suppose that are available $\{Y(1), \ldots, Y(N)\}$ from $Y(v) = e(v) + \theta e(v-1)$, where $|\theta| < 1$, $\theta \neq 0$, and $\{e(v)\}$ are uncorrelated random variables with variance unity. The original set $\mathcal{S} \equiv \{1, 2, \ldots, N\}$ will be corrected to $\mathcal{S}^* \equiv \{2, 3, \ldots, N-1\}$ that it is the maximal set, such that

for all $v \in \mathcal{S}^*$ it holds that $E(Y(v)Y(v^*)) = 0$, $v^* \notin \mathcal{S}$. Note that the "lags" on which the autocorrelations of Y are not equal to zero are as in the set $\mathcal{F} = \{1, 0, -1\}$; further, for any $v \in \mathbb{Z}$ it holds that $\mathcal{F}_v = \{v - 1, \dots, v - N\}$ and, thus, for $v \in \mathcal{S}^*$ it is true that $\mathcal{F} \subseteq \mathcal{F}_v$.

The autocovariance of the relevant auto-regression at the "lag" $j \in \mathbb{Z}$ is $(-\theta)^{|j|}/(1-\theta^2)$. As a result, the estimator $\widehat{\theta}$ will be the solution of the following equation:

$$
\begin{aligned}
0 = &-\widehat{\theta} \sum_{v=2}^{N-1} Y(v)^2 + \sum_{v=2}^{N-1} Y(v)Y(v+1) + \widehat{\theta}^2 \sum_{v=2}^{N-1} Y(v)Y(v-1) - \widehat{\theta} \sum_{v=2}^{N-2} Y(v)Y(v+2) \\
&- \widehat{\theta}^3 \sum_{v=3}^{N-1} Y(v)Y(v-2) + \widehat{\theta}^2 \sum_{v=2}^{N-3} Y(v)Y(v+3) + \widehat{\theta}^4 \sum_{v=4}^{N-2} Y(v)Y(v-3) - \cdots \\
&+ \left(-\widehat{\theta}\right)^{N-4} Y(2)Y(N-1) + \left(-\widehat{\theta}\right)^{N-2} Y(N-2)Y(1).
\end{aligned}
\tag{3.8}
$$

Furthermore, when $\{e(v)\}$ are independent and identically distributed random variables, it will hold in probability that $\sum_v Y(v)Y(v-i)/N \to 0$ for $|i| = 2, 3, \dots$ and the estimator may be approximated as a solution of the quadratic equation:

$$
0 \equiv -\widehat{\theta} \sum_{v=2}^{N-1} Y(v)^2 + \sum_{v=2}^{N-1} Y(v)Y(v+1) + \widehat{\theta}^2 \sum_{v=2}^{N-1} Y(v)Y(v-1).
\tag{3.9}
$$

Later, it will be investigated whether the approximation is worthwhile. The estimator reduces to

$$
\widehat{\theta} = \frac{\sum_{v=2}^{N-1} Y(v)^2 \pm \sqrt{\left(\sum_{v=2}^{N-1} Y(v)^2\right)^2 - 4\left(\sum_{v=2}^{N-1} Y(v)Y(v-1)\right)\left(\sum_{v=2}^{N-1} Y(v)Y(v+1)\right)}}{2\sum_{v=2}^{N-1} Y(v)Y(v-1)}
\tag{3.10}
$$

or

$$
\widehat{\theta} = \frac{1 \pm \sqrt{1 - 4\widehat{\rho}_+ \widehat{\rho}_-}}{2\widehat{\rho}_+}
\tag{3.11}
$$

with

$$
\widehat{\rho}_+ = \frac{\sum_{v=2}^{N-1} Y(v)Y(v-1)}{\sum_{v=2}^{N-1} Y(v)^2}, \qquad \widehat{\rho}_- = \frac{\sum_{v=2}^{N-1} Y(v)Y(v+1)}{\sum_{v=2}^{N-1} Y(v)^2}.
\tag{3.12}
$$

For the actual nonzero auto-correlation $\rho = \theta/(1+\theta^2)$, it holds that $|\rho| < 1/2$ and $D = 1-4\rho\rho > 0$. As a result, if $0 < \theta < 1$ and $0 < \rho < 1/2$, then $1/2\rho > 1$, and the value

$$
\frac{1 + \sqrt{1 - 4\rho\rho}}{2\rho} = \frac{1}{2\rho} + \frac{\sqrt{D}}{2\rho}
\tag{3.13}
$$

is bigger than 1. If, on the other hand, $-1 < \theta < 0$, and $-1/2 < \rho < 0$ then $1/2\rho < -1$ and the same value is smaller than -1. Thus, the estimator is approximately

$$\hat{\theta} = \frac{\sum_{v=2}^{N-1} Y(v)^2 - \sqrt{\left(\sum_{v=2}^{N-1} Y(v)^2\right)^2 - 4\left(\sum_{v=2}^{N-1} Y(v)Y(v-1)\right)\left(\sum_{v=2}^{N-1} Y(v)Y(v+1)\right)}}{2\sum_{v=2}^{N-1} Y(v)Y(v-1)}. \quad (3.14)$$

Although the distribution of $\hat{\theta}$ is not known for small sample sizes and $|\hat{\rho}_+|, |\hat{\rho}_-|$ risk to be outside the $(-0.5, 0.5)$ range with some positive probability, both the consistency and the asymptotic normality will be established next. As opposed to what is known for the maximum Gaussian likelihood estimators, the proposed estimators will be asymptotically normal even for processes on more than one dimension. Thus, the new methods might be more useful for higher number of dimensions as well as large sample sizes.

4. Properties of Estimators

The asymptotic normality of standard estimators for the parameters of stationary processes indexed on $\mathbb{Z}^d$, $d \geq 2$, has not been established yet. More specifically as Guyon [2] demonstrated, the bias of the genuine Gaussian likelihood estimators, computed from N regular recordings increasing to infinity on all dimensions and at equal speeds, is of order $N^{-1/d}$. Nevertheless in order to secure the asymptotic normality, the absolute bias when multiplied by $N^{1/2}$ should tend to zero, which is only true for $d = 1$ as in Brockwell and Davis [9].

Regarding the proposed estimators of this paper, their bias will stem from (3.3) as in the expected value of $\sum_{\mathbf{v} \in \mathcal{S}^*} \{\sum_{\mathbf{j} \in \mathbb{Z}^d} c_{\mathbf{j},0}(Y(\mathbf{v}+\mathbf{i}_n-\mathbf{j}) - H_Y(\mathbf{v}+\mathbf{i}_n-\mathbf{j}))Y(\mathbf{v})\}/N^*$, $n = 1, \ldots, q$, which expresses what is "missing" from the sample. Even when the bias is multiplied by $N^{*1/2}$, the random variable $N^{*-1/2} \sum_{\mathbf{v} \in \mathcal{S}^*} \sum_{\mathbf{j}-\mathbf{i}_n \notin \mathcal{F}_{\mathbf{v}}} c_{\mathbf{j},0} Y(\mathbf{v} + \mathbf{i}_n - \mathbf{j})Y(\mathbf{v})$, $n = 1, \ldots, q$ has a zero mean due to the selection $\mathcal{S}^*$ and the fact that the "finite" autocovariance function of a moving-average process is being used. Thus, the edge-effect will not be an obstacle to establish the asymptotic normality of the estimators.

A series of ARMA model-based arguments have been used before to deal with the edge-effect by Yao and Brockwell [6] for two-dimensional processes as well as a weak condition of finite second moments. Guyon [2], on the other hand, used the periodogram as in the Gaussian likelihood of observations proposed by Whittle [1] and required that the fourth moments would be finite. Dahlhaus and Künsch [10] improved certain weaknesses of Guyon's [2] modification on the likelihood but paid the price of having to reduce the dimensionality to secure their results. The proposed estimators of this paper will also use the weak condition of finite second moments, which is a great advantage, since they will refer to moving-average processes on any number of dimensions.

Condition C2. (i) For a set $\mathcal{S} \equiv \mathcal{S}_N \subset \mathbb{Z}^d$ of cardinality N, it is written $N \rightarrow \infty$ if the length M of the minimal hypercube including $\mathcal{S}$, say $\mathcal{S} \subseteq \mathcal{C}_M$, and the length m of the maximal hypercube included in $\mathcal{S}$, say $\mathcal{C}_m \subseteq \mathcal{S}$, are such that $M, m \rightarrow \infty$. (ii) Further, as $M, m \rightarrow \infty$ it holds that M/m is bounded away from ∞.

Theorem 4.1. *If $\{\varepsilon(\mathbf{v})\} \sim IID(0, \sigma^2)$, then under Conditions C1 and C2(i) and as $N \to \infty$, it holds that*

$$\hat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0, \qquad \hat{\sigma}^2 \xrightarrow{P} \sigma^2. \tag{4.1}$$

From (3.3) and (3.4), all the q equations together may be written as

$$N^{1/2}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) = \left(\frac{\mathbf{J}}{N}\right)^{-1}\left(N^{-1/2}\sum_{\mathbf{v} \in \mathcal{S}^*} \mathbf{H}_Y(\mathbf{v})\right), \tag{4.2}$$

where $\mathbf{J}^{\tau} = (\mathbf{J}_1^{\tau}, \dots, \mathbf{J}_q^{\tau})$ and

$$\mathbf{H}_Y(\mathbf{v}) \equiv \begin{pmatrix} c_0(\mathbf{B})H_Y(\mathbf{v} + \mathbf{i}_1) \\ \vdots \\ c_0(\mathbf{B})H_Y(\mathbf{v} + \mathbf{i}_q) \end{pmatrix} Y(\mathbf{v}), \quad \mathbf{v} \in \mathbb{Z}^d, \tag{4.3}$$

which depends on the sampling set $\mathcal{S}$. The next proposition reveals what happens to the part $(\mathbf{J}/N)$ in (4.2). Then Theorem 4.3 establishes the asymptotic normality of the estimators.

Proposition 4.2. *Let the polynomial $\theta_0(\mathbf{z})^{-1} = (1 + \sum_{n=1}^{q} \theta_{i_n,0}\mathbf{z}^{\mathbf{i}_n})^{-1} \equiv \sum_{\mathbf{i} \geq 0} \Theta_{\mathbf{i},0}\mathbf{z}^{\mathbf{i}}, \ \Theta_{\mathbf{0},0}$. If $\{\varepsilon(\mathbf{v})\} \sim IID(0, \sigma^2)$, then under Conditions C1 and C2(i) and as $N \to \infty$, it holds that*

$$\frac{\mathbf{J}}{N} \xrightarrow{P} \sigma^2\Theta_0 \equiv \sigma^2 \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \Theta_{\mathbf{i}_2-\mathbf{i}_1,0} & 1 & 0 & \cdots & 0 \\ \vdots & & & \ddots & \\ \Theta_{\mathbf{i}_q-\mathbf{i}_1,0} & \Theta_{\mathbf{i}_q-\mathbf{i}_2,0} & \Theta_{\mathbf{i}_q-\mathbf{i}_3,0} & & 1 \end{pmatrix}. \tag{4.4}$$

Theorem 4.3. *For $\{W(\mathbf{v})\} \sim IID(0, 1)$, the auto-regression $\{\eta(\mathbf{v})\}$ is defined by $\theta_0(\mathbf{B})\eta(\mathbf{v}) \equiv W(\mathbf{v})$. Also the vector $\boldsymbol{\xi} \equiv (\eta(-\mathbf{i}_1), \dots, \eta(-\mathbf{i}_q))^{\tau}$ and the variance matrix*

$$\mathbf{W}_q^* \equiv \mathrm{Var}\left(\boldsymbol{\xi} \mid W(-\mathbf{i}_1 - \mathbf{i}, \ \mathbf{i} > \mathbf{0}, \ \mathbf{i} \neq \mathbf{i}_2 - \mathbf{i}_1, \dots, \mathbf{i}_q - \mathbf{i}_1)\right) \tag{4.5}$$

are considered. If $\{\varepsilon(\mathbf{v})\} \sim IID(0, \sigma^2)$, then under Conditions C1 and C2 and as $N \to \infty$, it holds that

$$N^{1/2}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{D} N\left(\mathbf{0}, \mathbf{W}_q^{*-1}\right). \tag{4.6}$$

5. Empirical Results

In this section, an empirical comparison of the proposed estimators to maximum Gaussian likelihood estimators for moving-average processes is presented in detail. The theoretical

Table 1: Approximate bias, variance, and Mean Squared Error of the Method of Moments and the Maximum Likelihood Estimator for the MA(1) parameter from 1000 replications.

	$\theta = 0.5$			$\theta = -0.8$		
$N = 30$	Bias	Var	MSE	Bias	Var	MSE
MME	−0.403345	0.00020972	0.162897	0.0931052	0.00097912	0.00964771
MLE	0.01795	0.00015984	0.00048204	0.020949	0.00062347	0.00104331
$N = 100$	Bias	Var	MSE	Bias	Var	MSE
MME	−0.123812	0.00021444	0.0155439	0.0217928	0.00097912	0.00145404
MLE	0.00521	0.00031329	0.00034043	0.00599	0.00063936	0.00067524
$N = 200$	Bias	Var	MSE	Bias	Var	MSE
MME	−0.0422653	0.00021173	0.00199808	0.0123275	0.00097912	0.00113109
MLE	−0.00116	0.00022068	0.00022202	0.00275	0.00062348	0.00063104
$N = 300$	Bias	Var	MSE	Bias	Var	MSE
MME	−0.00708814	0.00023768	0.00028792	−0.00166597	0.00097912	0.00098189
MLE	0.00313	0.00029131	0.00036115	0.00386	0.00065544	0.00067034

foundations in favour of the new estimators have been provided already when their asymptotic normality on any number of dimensions has been established based on finite second moments only. As a result, the speed of the bias to zero is not expected to cause them to perform worse than maximum likelihood estimators, especially when the dimensionality is large.

On the other hand, Theorem 4.3 attributes to the new estimators the variance matrix $\mathbf{W}_q^{*-1}$ when, according to Hannan [11], efficient estimation results in $\mathbf{W}_q^{-1}$ with $\mathbf{W}_q \equiv \mathrm{Var}(\boldsymbol{\xi})$ and the same notation as in Theorem 4.3. $\mathbf{W}_q$ and $\mathbf{W}_q^*$ are defined similarly but they are not, in general, equal. It seems that as the number of moving-average parameters of the process increases, the two types elements get closer. Nevertheless, as the pure moving-average model will be preferred when it is parsimonious, a decision is made here whether the new estimators are efficient then.

The investigation has started with the one-dimensional case by generating $\{Y(1), \ldots, Y(N)\}$ from the model with one parameter only. The moments estimator has been approximated as in the example earlier, while the minimizer of $\sum_{v \in \mathcal{S}^*} e_*^2(v)$ with respect to the parameter θ has been considered to be the Gaussian likelihood estimator. The true values $\theta = 0.5, -0.8$ have been considered and the sample size has been set equal to $N = 30, 100, 200, 300$. Very encouraging results for the efficiency of the proposed estimator are presented in Table 1 as even when the sample size is still small, extreme differences in the variances of the two types estimators cannot be detected. It is the bias of the moments estimator that seems to be the only reason why it might be outperformed by the likelihood estimator in terms of the Mean Squared Error. Nevertheless, the speed with which the bias tends to zero is much faster as one would expect and, eventually, the new estimator for $\theta = 0.5$ performs better altogether.

Next for the two-dimensional case, $\{Y(u, v), \ u, v = 1, \ldots, n\}$ have been generated from

$$Y(u, v) = e(u, v) + \beta e(u - 1, v) + \gamma e(u, v - 1), \qquad |\beta| + |\gamma| < 1, \qquad (5.1)$$

with $\{e(u,v),\ u,v \in \mathbb{Z}\}$ being independent, Gaussian random variables with zero mean and variance unity. Using similar arguments as in the example, the moments estimators of β and γ might be derived by solving simultaneously

$$
\begin{aligned}
0 &\equiv \sum Y(u,v)Y(u+1,v)\widehat{c}_{(0,0)} + \sum Y^2(u,v)\widehat{c}_{(-1,0)} + \sum Y(u,v)Y(u-1,v)\widehat{c}_{(-2,0)} \\
&\quad + \sum Y(u,v)Y(u,v+1)\widehat{c}_{(-1,1)} + \sum Y(u,v)Y(u,v-1)\widehat{c}_{(-1,-1)} \\
&\quad + \sum Y(u,v)Y(u+1,v-1)\widehat{c}_{(0,-1)} + \sum Y(u,v)Y(u-1,v+1)\widehat{c}_{(-2,1)}, \\
0 &\equiv \sum Y(u,v)Y(u,v+1)\widehat{c}_{(0,0)} + \sum Y^2(u,v)\widehat{c}_{(0,-1)} + \sum Y(u,v)Y(u,v-1)\widehat{c}_{(0,-2)} \\
&\quad + \sum Y(u,v)Y(u+1,v)\widehat{c}_{(1,-1)} + \sum Y(u,v)Y(u-1,v)\widehat{c}_{(-1,-1)} \\
&\quad + \sum Y(u,v)Y(u+1,v-1)\widehat{c}_{(1,-2)} + \sum Y(u,v)Y(u-1,v+1)\widehat{c}_{(-1,0)},
\end{aligned}
\tag{5.2}
$$

as well as the third equation

$$
\begin{aligned}
0 &\equiv \sum Y(u,v)Y(u+1,v-1)\widehat{c}_{(0,0)} + \sum Y^2(u,v)\widehat{c}_{(-1,1)} + \sum Y(u,v)Y(u-1,v+1)\widehat{c}_{(-2,2)} \\
&\quad + \sum Y(u,v)Y(u+1,v)\widehat{c}_{(0,1)} + \sum Y(u,v)Y(u-1,v)\widehat{c}_{(-2,1)} \\
&\quad + \sum Y(u,v)Y(u,v+1)\widehat{c}_{(-1,2)} + \sum Y(u,v)Y(u,v-1)\widehat{c}_{(-1,0)},
\end{aligned}
\tag{5.3}
$$

which is needed to secure the consistency (Proposition 3.1). All of the sums above extend over $\mathcal{S}^* \equiv \{2,\dots,n-1\} \times \{2,\dots,n-1\}$. The estimators introduced here are $\widehat{\beta}$, $\widehat{\gamma}$, and in the three equations, it has been considered that $\widehat{c}_{(j,k)} \equiv \mathrm{Corr}(X(u,v), X(u-j,v-k))$, where

$$
X(u,v) + \widehat{\beta}X(u-1,v) + \widehat{\gamma}X(u,v-1) = e^*(u,v)
\tag{5.4}
$$

and $\{e^*(u,v)\}$ are zero mean, uncorrelated random variables all with variance unity. Under the condition $|\widehat{\beta}| + |\widehat{\gamma}| < 1$, a special form of "causality" $X(u,v) \equiv \sum_{j,k=0}^{\infty} \widehat{\Theta}_{(j,k)} e^*(u-j, v-k)$, $\sum_{j,k=0}^{\infty} |\widehat{\Theta}_{(j,k)}| < \infty$ is implied and the standard Yule-Walker equations may be used to easily write the needed autocorrelations $\widehat{c}$, as functions of $\widehat{\beta}$ and $\widehat{\gamma}$. Again, the maximum likelihood estimators have been approximated by the minimizers of $\sum_{u,v} (Y(u,v) - \beta e(u-1,v) - \gamma e(u,v-1))^2$ with respect to β and γ.

Table 2 verifies once more the conclusions drawn from the one-dimensional case. It is safe to consider that the new estimators are efficient and this is very apparent in the case that the parameters are in absolute value further away from zero ($\beta = 0.5$ and $\gamma = 0.45$). The striking case with the small sample size of around 50 points observed in the plane reveals that the variances of the moments estimators may even happen to be smaller. On the other hand, the bias heavily affects the results for the MSE for small sample sizes. Nevertheless as the sample size increases, the absolute bias of the likelihood estimators does not seem to decrease at all versus the bias of the proposed estimators that speedily reaches the zero value. As a result, the new estimators eventually equalize the MSE performance of the standard estimators.

Table 2: Approximate bias, variance, and mean squared error of the method of moments and the maximum likelihood estimators (based on $N = n^2$ observations) for the two-dimensional MA parameters β and γ from 1000 replications.

	$\beta = 0.5$			$\gamma = 0.45$		
$n = 7$	Bias	Var	MSE	Bias	Var	MSE
MME	0.02842	0.00010879	0.00091649	−0.07858	0.00001209	0.0061869
MLE	0.0006	0.0002023	0.00020266	−0.00604	0.00006244	0.00009892
$n = 30$	Bias	Var	MSE	Bias	Var	MSE
MME	0.0149	0.00037173	0.00059374	−0.00746	0.0002023	0.00025795
MLE	−0.00184	0.0003022	0.00030558	0.00066	0.00016793	0.00016837
$n = 100$	Bias	Var	MSE	Bias	Var	MSE
MME	0.00576	0.00025984	0.00029302	0.0065	0.00023986	0.00028211
MLE	−0.00044	0.00022068	0.00022087	−0.00024	0.0002023	0.00020236
	$\beta = 0.15$			$\gamma = -0.3$		
$n = 30$	Bias	Var	MSE	Bias	Var	MSE
MME	0.02638	0.00004406	0.00073996	0.03274	0.00001688	0.00108879
MLE	0.00012	0.00002248	0.00002249	0.00004	0.00010879	0.00010879
$n = 100$	Bias	Var	MSE	Bias	Var	MSE
MME	−0.0422653	0.00021173	0.00199808	0.0123275	0.00097912	0.00113109
MLE	−0.00058	0.00002248	0.00002281	0.00026	0.00000402	0.00008408

Appendix

A. Outline Proofs

A.1. Proof of Proposition 3.1

Under the invertibility condition for the polynomials $\theta(\mathbf{z})$, there is a one-to-one correspondence between the q coefficients $\theta_{\mathbf{i}_1}, \ldots, \theta_{\mathbf{i}_q}$ and the $q^\star$ autocorrelations $\rho_{\mathbf{i}_1}, \ldots, \rho_{\mathbf{i}_q}, \rho_{\mathbf{i}_2 - \mathbf{i}_1}, \ldots, \rho_{\mathbf{i}_q - \mathbf{i}_1}$, where $\rho_{\mathbf{j}} = \gamma_{\mathbf{j}}/\gamma_{\mathbf{0}}$, $\mathbf{j} \in \mathbb{Z}^d$. It is true that the coefficients $\theta_{\mathbf{i}_n, 0}$, $n = 1, \ldots, q$, generate the numbers $c_{\mathbf{j}, 0}$, $\mathbf{j} \in \mathbb{Z}^d$, and they are a solution to the $q^\star$ equations of interest. If there is another solution, say $\theta_{\mathbf{i}_n, 1}$, $n = 1, \ldots, q$ generating $c_{\mathbf{j}, 1}$, $\mathbf{j} \in \mathbb{Z}^d$, then it must hold that

$$\sum_{\mathbf{j} \in \mathcal{F} - \{\mathbf{0}\}} \rho_{\mathbf{j}, 0} c_{\mathbf{j} - \mathbf{i}_n, 1} = -c_{\mathbf{i}_n, 1}, \quad n = 1, \ldots, q,$$

$$\sum_{\mathbf{j} \in \mathcal{F} - \{\mathbf{0}\}} \rho_{\mathbf{j}, 0} c_{\mathbf{j} - (\mathbf{i}_n - \mathbf{i}_m), 1} = -c_{\mathbf{i}_m - \mathbf{i}_n, 1}, \quad n, m = 1, \ldots, q, \ n > m.$$

(A.1)

On the other hand, the general Yule-Walker equations for this solution imply that

$$\sum_{\mathbf{j} \in \mathcal{F} - \{\mathbf{0}\}} \rho_{\mathbf{j}, 1} c_{\mathbf{j} - \mathbf{i}_n, 1} = -c_{\mathbf{i}_n, 1}, \quad n = 1, \ldots, q,$$

$$\sum_{\mathbf{j} \in \mathcal{F} - \{\mathbf{0}\}} \rho_{\mathbf{j}, 1} c_{\mathbf{j} - (\mathbf{i}_n - \mathbf{i}_m), 1} = -c_{\mathbf{i}_m - \mathbf{i}_n, 1}, \quad n, m = 1, \ldots, q, \ n > m.$$

(A.2)

Thus, the $q^\star$ linear equations with $q^\star$ unknowns are derived as

$$\sum_{\mathbf{j}\in\mathcal{F}-\{\mathbf{0}\}} (\rho_{\mathbf{j},0} - \rho_{\mathbf{j},1})c_{\mathbf{j}-\mathbf{i}_n,1} = 0, \quad n = 1,\dots,q,$$

$$\sum_{\mathbf{j}\in\mathcal{F}-\{\mathbf{0}\}} (\rho_{\mathbf{j},0} - \rho_{\mathbf{j},1})c_{\mathbf{j}-(\mathbf{i}_n-\mathbf{i}_m),1} = 0, \quad n, m = 1,\dots,q, \; n > m,$$

(A.3)

with a unique solution as it is explained next; the autocovariances $c_{\mathbf{j},1}$, $\mathbf{j} \in \mathbb{Z}^d$ refer to a (weakly) stationary process, say $\{X_1(\mathbf{v})\}$, with spectral density bounded away from 0 and ∞ under Condition C1 . Consequently, the variance matrix

$$[c_{\mathbf{l}_n+\mathbf{l}_m,1} + c_{\mathbf{l}_n-\mathbf{l}_m,1}]_{n,m=1}^{q^\star} = \frac{C_1}{2}, \quad \mathbf{l}_n \in \{\mathbf{i}_r, \mathbf{i}_r - \mathbf{i}_m, \; r, m = 1,\dots,q, \; r > m\}, \; n = 1,\dots,q^\star, \quad \text{(A.4)}$$

where $C_1 \equiv [\mathrm{Cov}(X_1(\mathbf{v}+\mathbf{l}_n) + X_1(\mathbf{v}-\mathbf{l}_n), \; X_1(\mathbf{v}+\mathbf{l}_m) + X_1(\mathbf{v}-\mathbf{l}_m))]_{n,m=1}^{q^\star}$, is nonsingular and there is a unique solution to the equations above, that is, $\rho_{\mathbf{j},1} = \rho_{\mathbf{j},0}$, $\mathbf{j} \in \mathcal{F}$.

A.2. Proof of Theorem 4.1

For $n = 1,\dots,q$, (3.1) may be rewritten as

$$\frac{\sum_{\mathbf{v}\in S^*} \sum_{\mathbf{j}-\mathbf{i}_n\in\mathcal{F}_\mathbf{v}} \widehat{c}_{\mathbf{j}} Y(\mathbf{v}+\mathbf{i}_n-\mathbf{j})Y(\mathbf{v})}{N^*}$$

$$= \frac{\sum_{\mathbf{j}-\mathbf{i}_n\in\mathbb{Z}^d} \widehat{c}_{\mathbf{j}} \sum_{\mathbf{v}\in S^*} Y(\mathbf{v}+\mathbf{i}_n-\mathbf{j})Y(\mathbf{v})}{N^*} - \frac{\sum_{\mathbf{v}\in S^*} \sum_{\mathbf{j}-\mathbf{i}_n\notin\mathcal{F}_\mathbf{v}} \widehat{c}_{\mathbf{j}} Y(\mathbf{v}+\mathbf{i}_n-\mathbf{j})Y(\mathbf{v})}{N^*} = 0.$$

(A.5)

Under the assumption that $\{\varepsilon(\mathbf{v})\}$ are independent and identically distributed, it holds as $N \to \infty$ that $\sum_{\mathbf{v}\in S^*} Y(\mathbf{v}+\mathbf{i}_n-\mathbf{j})Y(\mathbf{v})/N^* \xrightarrow{P} E\{Y(\mathbf{v}+\mathbf{i}_n-\mathbf{j})Y(\mathbf{v})\} = \sigma^2\gamma_{\mathbf{j}-\mathbf{i}_n,0}$, according to Proposition 6.3.10 or the Weak Law of Large Numbers and Proposition 7.3.5 of Brockwell and Davis [9], extended to include the cases $d \geq 2$. Then for $n = 1,\dots,q$ and the first of two terms in (A.5), it may be written as

$$\frac{\sum_{\mathbf{j}-\mathbf{i}_n\in\mathbb{Z}^d} \widehat{c}_{\mathbf{j}} \sum_{\mathbf{v}\in S^*} Y(\mathbf{v}+\mathbf{i}_n-\mathbf{j})Y(\mathbf{v})}{N^*} - \sigma^2 \sum_{\mathbf{j}-\mathbf{i}_n\in\mathbb{Z}^d} \widehat{c}_{\mathbf{j}}\gamma_{\mathbf{j}-\mathbf{i}_n,0} \xrightarrow{P} 0$$

(A.6)

or

$$\frac{\sum_{\mathbf{j}-\mathbf{i}_n\in\mathbb{Z}^d} \widehat{c}_{\mathbf{j}} \sum_{\mathbf{v}\in S^*} Y(\mathbf{v}+\mathbf{i}_n-\mathbf{j})Y(\mathbf{v})}{N^*} - \sigma^2 \sum_{\mathbf{j}-\mathbf{i}_n\in\mathcal{F}} \widehat{c}_{\mathbf{j}}\gamma_{\mathbf{j}-\mathbf{i}_n,0} \xrightarrow{P} 0.$$

(A.7)

For the next term, it may be written as

$$
E\left|\frac{\sum_{\mathbf{v}\in\mathcal{S}^*}\sum_{\mathbf{j}-\mathbf{i}_n\notin\mathcal{F}_{\mathbf{v}}}\widehat{c}_{\mathbf{j}}Y(\mathbf{v}+\mathbf{i}_n-\mathbf{j})Y(\mathbf{v})}{N^*}\right| \leq \frac{\sum_{\mathbf{v}\in\mathcal{S}^*}\sum_{\mathbf{j}-\mathbf{i}_n\notin\mathcal{F}_{\mathbf{v}}}E\left|\widehat{c}_{\mathbf{j}}Y(\mathbf{v}+\mathbf{i}_n-\mathbf{j})Y(\mathbf{v})\right|}{N^*}
$$

$$
\leq \frac{\sum_{\mathbf{v}\in\mathcal{S}^*}\sum_{\mathbf{j}-\mathbf{i}_n\notin\mathcal{F}_{\mathbf{v}}}E\left(\widehat{c}_{\mathbf{j}}^2\right)^{1/2}E\left(Y(\mathbf{v}+\mathbf{i}_n-\mathbf{j})^2Y(\mathbf{v})^2\right)^{1/2}}{N^*}
$$

$$
= \frac{E\left(Y(\mathbf{v})^2\right)\sum_{\mathbf{v}\in\mathcal{S}^*}\sum_{\mathbf{j}-\mathbf{i}_n\notin\mathcal{F}_{\mathbf{v}}}E\left(\widehat{c}_{\mathbf{j}}^2\right)^{1/2}}{N^*},
$$

$$(A.8)$$

due to the Cauchy-Schwartz inequality and the independence of the random variables $Y(\mathbf{v}), Y(\mathbf{v}-\mathbf{j})$, $\mathbf{j}\notin\mathcal{F}$. Now for any $\boldsymbol{\theta}\in\Theta$, it holds that $c_{\mathbf{j}}\equiv c_{\mathbf{j}}(\boldsymbol{\theta})$ is the corresponding autocovariance function of a causal auto-regression. This guarantees that the autocovariance function decays at an exponential rate and there are constants $C(\boldsymbol{\theta})>0$ and $\alpha(\boldsymbol{\theta})\in(0,1)$, such that $c_{\mathbf{j}}(\boldsymbol{\theta})^2\leq C(\boldsymbol{\theta})\alpha(\boldsymbol{\theta})^{\sum_{k=1}^d|j_k|}$, $\mathbf{j}=(j_1,\ldots,j_d)$. Similarly for the estimator $\widehat{\boldsymbol{\theta}}$, it holds that

$$
\widehat{c}_{\mathbf{j}}^2 \leq C\left(\widehat{\boldsymbol{\theta}}\right)\alpha\left(\widehat{\boldsymbol{\theta}}\right) \leq \sup_{\boldsymbol{\theta}\in\Theta}C(\boldsymbol{\theta})\alpha(\boldsymbol{\theta})^{\sum_{k=1}^d|j_k|} \leq \sup_{\boldsymbol{\theta}\in\Theta}C(\boldsymbol{\theta})\left\{\sup_{\boldsymbol{\theta}\in\Theta}\alpha(\boldsymbol{\theta})\right\}^{\sum_{k=1}^d|j_k|}
$$

$$(A.9)$$

with probability one and $E(\widehat{c}_{\mathbf{j}}^2) \leq \sup_{\boldsymbol{\theta}\in\Theta}C(\boldsymbol{\theta})\{\sup_{\boldsymbol{\theta}\in\Theta}\alpha(\boldsymbol{\theta})\}^{\sum_{k=1}^d|j_k|}$. For the case of observations on a hyperrectangle and under Condition C2(ii) , it can be easily verified that $\sum_{\mathbf{v}\in\mathcal{S}^*}\sum_{\mathbf{j}-\mathbf{i}_n\notin\mathcal{F}_{\mathbf{v}}}E(\widehat{c}_{\mathbf{j}}^2)=O(N^{(d-1)/d})$, for example, the arguments of Yao and Brockwell [6] when $d=2$. In general, it can be written as Condition C2(i) that

$$
\frac{\sum_{\mathbf{v}\in\mathcal{S}^*}\sum_{\mathbf{j}-\mathbf{i}_n\notin\mathcal{F}_{\mathbf{v}}}E\left(\widehat{c}_{\mathbf{j}}^2\right)}{N^*} \longrightarrow 0, \qquad \frac{\sum_{\mathbf{v}\in\mathcal{S}^*}\sum_{\mathbf{j}-\mathbf{i}_n\notin\mathcal{F}_{\mathbf{v}}}\widehat{c}_{\mathbf{j}}Y(\mathbf{v}+\mathbf{i}_n-\mathbf{j})Y(\mathbf{v})}{N^*} \xrightarrow{P} 0. \qquad (A.10)
$$

After combining the results for the two terms in (A.5), it can be written that

$$
\frac{\sum_{\mathbf{v}\in\mathcal{S}^*}\sum_{\mathbf{j}-\mathbf{i}_n\in\mathcal{F}_{\mathbf{v}}}\widehat{c}_{\mathbf{j}}Y(\mathbf{v}+\mathbf{i}_n-\mathbf{j})Y(\mathbf{v})}{N^*} - \sigma^2\sum_{\mathbf{j}-\mathbf{i}_n\in\mathcal{F}}\widehat{c}_{\mathbf{j}}\gamma_{\mathbf{j}-\mathbf{i}_n,0} \xrightarrow{P} 0, \quad n=1,\ldots,q, \qquad (A.11)
$$

where the first term has been set equal to zero. Thus, $\sum_{\mathbf{j}-\mathbf{i}_n\in\mathcal{F}}\widehat{c}_{\mathbf{j}}\gamma_{\mathbf{j}-\mathbf{i}_n,0} \xrightarrow{P} 0$ exactly like the theoretical analogue $\sum_{\mathbf{j}-\mathbf{i}_n\in\mathcal{F}}c_{\mathbf{j},0}\gamma_{\mathbf{j}-\mathbf{i}_n,0}=0$ implies. Since q^* instead of q equations have been used, there is a unique solution $\boldsymbol{\theta}_0$, according to Proposition 3.1 and $\widehat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$ as $N\to\infty$ and C2(i) holds. Finally, the consistency for $\widehat{\boldsymbol{\theta}}$ implies, according to (3.7), that

$$
\widehat{\sigma}^2 \xrightarrow{P} \sigma^2\sum_{\mathbf{j}\in\mathcal{F}}c_{\mathbf{j},0}\gamma_{\mathbf{j},0} = \sigma^2 \quad \text{since } \sum_{\mathbf{j}\in\mathcal{F}}c_{\mathbf{j},0}\gamma_{\mathbf{j},0}=1. \qquad (A.12)
$$

A.3. Proof of Proposition 4.2

According to (3.5), for the (n, m)th element of $\mathbf{J}/N$, $n, m = 1, \ldots, q$, it suffices to look at

$$\frac{\sum_{\mathbf{v} \in \mathcal{S}^*} \left\{ c_0(\mathbf{B}) \left(\theta_0(\mathbf{B})^{-1} H_Y(\mathbf{v} + \mathbf{i}_n - \mathbf{i}_m) + \theta_0(\mathbf{B}^{-1})^{-1} H_Y(\mathbf{v} + \mathbf{i}_n + \mathbf{i}_m) \right) \right\} Y(\mathbf{v})}{N} + o_P(1), \quad (A.1)$$

where the last term tends to zero in probability, thanks to the consistency of the estimators from the use of all the q^* equations. For the polynomial

$$d_0(\mathbf{z}) \equiv \theta_0(\mathbf{z})^{-1} c_0(\mathbf{z}) = \theta_0(\mathbf{z})^{-1} c_0\left(\mathbf{z}^{-1} \right) \equiv \sum_{\mathbf{i} \in \mathbb{Z}^d} d_{\mathbf{i},0} \mathbf{z}^{\mathbf{i}}, \quad (A.2)$$

the second term in (A.1) can be written as

$$\frac{\sum_{\mathbf{v} \in \mathcal{S}^*} \left(d_0(\mathbf{B}^{-1}) H_Y(\mathbf{v} + \mathbf{i}_n + \mathbf{i}_m) \right) Y(\mathbf{v})}{N} = \frac{\sum_{\mathbf{v} \in \mathcal{S}^*} \left(d_0(\mathbf{B}^{-1}) Y(\mathbf{v} + \mathbf{i}_n + \mathbf{i}_m) \right) Y(\mathbf{v})}{N} + o_P(1). \quad (A.3)$$

This comes straight from the fact that

$$
\begin{aligned}
E \left| \frac{1}{N} \sum_{\mathbf{v} \in \mathcal{S}^*} \left\{ \sum_{-\mathbf{i}_n - \mathbf{i}_m - \mathbf{i} \notin \mathcal{F}_\mathbf{v}} d_{\mathbf{i},0} Y(\mathbf{v} + \mathbf{i}_n + \mathbf{i}_m + \mathbf{i}) \right\} Y(\mathbf{v}) \right| \\
\leq \frac{1}{N} \sum_{\mathbf{v} \in \mathcal{S}^*} \sum_{-\mathbf{i}_n - \mathbf{i}_m - \mathbf{i} \notin \mathcal{F}_\mathbf{v}} |d_{\mathbf{i},0}| E |Y(\mathbf{v} + \mathbf{i}_n + \mathbf{i}_m + \mathbf{i}) Y(\mathbf{v})| \quad (A.4) \\
= (E|Y(\mathbf{v})|)^2 \frac{1}{N} \sum_{\mathbf{v} \in \mathcal{S}^*} \sum_{-\mathbf{i}_n - \mathbf{i}_m - \mathbf{i} \notin \mathcal{F}_\mathbf{v}} |d_{\mathbf{i},0}| \longrightarrow 0,
\end{aligned}
$$

as $N \to \infty$ and Condition C2(i) holds. The limit comes from the same argument, as for the proof of the consistency for the estimators. For example, if condition C2(ii) is true, it may be written $\sum_{\mathbf{v} \in \mathcal{S}^*} \sum_{\mathbf{i} \notin \mathcal{F}_\mathbf{v}} |d_{\mathbf{i},0}| = O(N^{(d-1)/d})$, since for any $\mathbf{i} = (i_1, \ldots, i_d) \in \mathbb{Z}^d$, it holds that $|d_{\mathbf{i},0}| \leq C \alpha^{\sum_{k=1}^{d} |i_k|}$ for constants $C > 0$ and $\alpha \in (0, 1)$. Similar action might be taken for the first term in (A.1).

For the auto-regression $\{X(\mathbf{v})\}$ as it was defined in (2.3), it can be seen immediately that $Y(\mathbf{v})$ is uncorrelated to $X(\mathbf{v} + \mathbf{i})$, $\mathbf{i} > \mathbf{0}$, since these are linear functions of "future" error terms only. In general, it can be written according to (2.5) that

$$E(Y(\mathbf{v}) X(\mathbf{v} - \mathbf{i})) = \sum_{\mathbf{j} \in \mathcal{F}} \gamma_{\mathbf{j},0} E(X(\mathbf{v} - \mathbf{j}) X(\mathbf{v} - \mathbf{i})) = \sigma^2 \sum_{\mathbf{j} \in \mathcal{F}} \gamma_{\mathbf{j},0} c_{\mathbf{j} - \mathbf{i},0} \quad (A.5)$$

and back to the general Yule-Walker equations. Thus,

$$E(Y(\mathbf{v}) X(\mathbf{v})) = \sigma^2, \qquad E(Y(\mathbf{v}) X(\mathbf{v} - \mathbf{i})) = 0, \quad \mathbf{i} \neq \mathbf{0}, \quad (A.6)$$

and $Y(\mathbf{v})$ is uncorrelated to $X(\mathbf{v} - \mathbf{i})$, $\mathbf{i} \neq \mathbf{0}$. As a result, it holds for $n, m = 1, \ldots, q$, that

$$E\left(\left(\theta_0\left(\mathbf{B}^{-1}\right)^{-1}c_0(\mathbf{B})Y(\mathbf{v} + \mathbf{i}_n + \mathbf{i}_m)\right)Y(\mathbf{v})\right) = E\left(\left(\theta_0\left(\mathbf{B}^{-1}\right)^{-1}X(\mathbf{v} + \mathbf{i}_n + \mathbf{i}_m)\right)Y(\mathbf{v})\right) = 0, \tag{A.7}$$

and, for $n \geq m$, that

$$E\left(\left(\theta_0(\mathbf{B})^{-1}c_0(\mathbf{B})Y(\mathbf{v} + \mathbf{i}_n - \mathbf{i}_m)\right)Y(\mathbf{v})\right) = E\left(\left(\theta_0(\mathbf{B})^{-1}X(\mathbf{v} + \mathbf{i}_n - \mathbf{i}_m)\right)Y(\mathbf{v})\right) = \sigma^2\Theta_{\mathbf{i}_n - \mathbf{i}_m, 0}. \tag{A.8}$$

The proof is completed when it is seen that both $Y(\mathbf{v})$ and $X(\mathbf{v})$ are linear functions of members from the sequence $\{\varepsilon(\mathbf{v})\}$ and, thus,

$$\frac{\sum_{\mathbf{v} \in \mathcal{S}^*}\left(\theta_0\left(\mathbf{B}^{-1}\right)^{-1}c_0(\mathbf{B})Y(\mathbf{v} + \mathbf{i}_n + \mathbf{i}_m)\right)Y(\mathbf{v})}{N} \xrightarrow{P} E\left(\left(\theta_0\left(\mathbf{B}^{-1}\right)^{-1}X(\mathbf{v} + \mathbf{i}_n + \mathbf{i}_m)\right)Y(\mathbf{v})\right),$$

$$\frac{\sum_{\mathbf{v} \in \mathcal{S}^*}\left(\theta_0(\mathbf{B})^{-1}c_0(\mathbf{B})Y(\mathbf{v} + \mathbf{i}_n - \mathbf{i}_m)\right)Y(\mathbf{v})}{N} \xrightarrow{P} E\left(\left(\theta_0(\mathbf{B})^{-1}X(\mathbf{v} + \mathbf{i}_n - \mathbf{i}_m)\right)Y(\mathbf{v})\right). \tag{A.9}$$

A.4. Proof of Theorem 4.3

From (3.3), (4.2), and (4.3) and for $n = 1, \ldots, q$, it can be written that

$$N^{-1/2}\sum_{\mathbf{v} \in \mathcal{S}^*}\sum_{\mathbf{j} \in \mathbb{Z}^d}c_{\mathbf{j},0}H_Y(\mathbf{v} + \mathbf{i}_n - \mathbf{j})Y(\mathbf{v}) = N^{-1/2}\sum_{\mathbf{v} \in \mathcal{S}^*}\sum_{\mathbf{j} \in \mathbb{Z}^d}c_{\mathbf{j},0}Y(\mathbf{v} + \mathbf{i}_n - \mathbf{j})Y(\mathbf{v}) + o_P(1). \tag{A.1}$$

The convergence in probability to zero of the remainder might be justified by the fact that its expected value is equal to zero due to the selection $\mathcal{S}^*$, and that its variance is

$$\mathrm{Var}\left(N^{-1/2}\sum_{\mathbf{v} \in \mathcal{S}^*}\sum_{\mathbf{j} - \mathbf{i}_n \notin \mathcal{F}_{\mathbf{v}}}c_{\mathbf{j},0}Y(\mathbf{v} + \mathbf{i}_n - \mathbf{j})Y(\mathbf{v})\right) \equiv \frac{\mathrm{Var}\left(\sum_{\mathbf{v} \in \mathcal{S}^*}\tilde{u}_n(\mathbf{v})\right)}{N}, \tag{A.2}$$

where

$$\tilde{u}_n(\mathbf{v}) \equiv \sum_{\mathbf{j} - \mathbf{i}_n \notin \mathcal{F}_{\mathbf{v}}}c_{\mathbf{j},0}Y(\mathbf{v} + \mathbf{i}_n - \mathbf{j})Y(\mathbf{v}), \quad n = 1, \ldots, q, \ \mathbf{v} \in \mathbb{Z}^d, \tag{A.3}$$

given the sampling set $\mathcal{S}$. First, $\{\varepsilon(\mathbf{v})\}$ are independent and identically distributed and then

$$E\left(\tilde{u}_n(\mathbf{v})^2\right) = E\left(Y(\mathbf{v})^2\right)E\left(\left(\sum_{\mathbf{j} - \mathbf{i}_n \notin \mathcal{F}_{\mathbf{v}}}c_{\mathbf{j},0}Y(\mathbf{v} + \mathbf{i}_n - \mathbf{j})\right)^2\right) \tag{A.4}$$

without the assumption of a finite third or fourth moment. Under Condition C2(ii), it can be written that

$$\sum_{\mathbf{v}\in\mathcal{S}^*}\mathrm{Var}(\tilde{u}_n(\mathbf{v})) = \sum_{\mathbf{v}\in\mathcal{S}^*}E\left(\tilde{u}_n(\mathbf{v})^2\right) = O\left(N^{(d-1)/d}\right), \tag{A.5}$$

and a similar argument may be used for the cross-terms due to the Cauchy-Schwartz inequality. For the case $d = 2$ and recordings on a rectangle, there is a justification for that by Yao and Brockwell [6]. Thus, $\mathrm{Var}(N^{-1/2}\sum_{\mathbf{v}\in\mathcal{S}^*}\tilde{u}_n(\mathbf{v})) \to 0$ as $N \to \infty$ and Condition C2 holds, which guarantees the convergence in probability to zero.

Equation (4.2) may now be rewritten as

$$N^{1/2}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) = \left(\frac{\mathbf{J}}{N}\right)^{-1}\left(N^{-1/2}\sum_{\mathbf{v}\in\mathcal{S}^*}\mathbf{U}(\mathbf{v})\right) + o_P(1), \tag{A.6}$$

where $\mathbf{U}(\mathbf{v})^\tau \equiv (X(\mathbf{v}+\mathbf{i}_1),\dots,X(\mathbf{v}+\mathbf{i}_q))Y(\mathbf{v})$, $\mathbf{v}\in\mathbb{Z}^d$. It holds that $Y(\mathbf{v})$ is a linear function of $\varepsilon(\mathbf{v}-\mathbf{i})$, $\mathbf{i}=\mathbf{0}$, $\mathbf{i}_1,\dots,\mathbf{i}_q$, and $X(\mathbf{v})$ is a function of $\varepsilon(\mathbf{v}+\mathbf{i})$, $\mathbf{i}\geq\mathbf{0}$. Then for $n,m=1,\dots,q$, it can be written for $\mathbf{j}\geq\mathbf{0}$ that

$$E(X(\mathbf{v}+\mathbf{i}_n)Y(\mathbf{v})X(\mathbf{v}+\mathbf{i}_m+\mathbf{j})Y(\mathbf{v}+\mathbf{j}))$$

$$= E(E(X(\mathbf{v}+\mathbf{i}_n)Y(\mathbf{v})X(\mathbf{v}+\mathbf{i}_m+\mathbf{j})Y(\mathbf{v}+\mathbf{j})\mid\varepsilon(\mathbf{v}+\mathbf{i}_m+\mathbf{j}+\mathbf{i}),\ \mathbf{i}\geq\mathbf{0})) \tag{A.7}$$

$$= E(Y(\mathbf{v})Y(\mathbf{v}+\mathbf{j}))E(X(\mathbf{v}+\mathbf{i}_n)X(\mathbf{v}+\mathbf{i}_m+\mathbf{j})) = \sigma^4\gamma_{\mathbf{j},0}c_{\mathbf{j}+\mathbf{i}_m-\mathbf{i}_n,0}.$$

Thus, for $\mathbf{X}(\mathbf{v})\equiv(X(\mathbf{v}+\mathbf{i}_1),\dots,X(\mathbf{v}+\mathbf{i}_q))^\tau$, $\mathbf{C}_{\mathbf{j},0}\equiv E(\mathbf{X}(\mathbf{v})\mathbf{X}(\mathbf{v}+\mathbf{j})^\tau)/\sigma^2$, it can be written that $E(\mathbf{U}(\mathbf{v}))=\mathbf{0}$ and

$$\mathrm{Cov}(\mathbf{U}(\mathbf{v}),\mathbf{U}(\mathbf{v}+\mathbf{j}))\equiv E\left(\mathbf{U}(\mathbf{v})\mathbf{U}(\mathbf{v}+\mathbf{j})^\tau\right) = \sigma^4\gamma_{\mathbf{j},0}\mathbf{C}_{\mathbf{j},0}, \quad \mathbf{j}\geq\mathbf{0}. \tag{A.8}$$

Now, for any positive integer K, the set

$$\mathcal{B}_K \equiv \{(i_1,i_2,\dots,i_d) : i_1 = 1,\dots,K,\ i_k = 0,\pm1,\dots,\pm K,\ k=2,\dots,d\}$$

$$\cup\left\{(0,i_2,\dots,i_d) : j_2 = 1,\dots,K,\ i_k = 0,\pm1,\dots,\pm K,\ k=3,\dots,d\right\} \tag{A.9}$$

$$\cup\cdots\cup\{(0,0,\dots,i_d) : i_d = 1,\dots,K\}$$

is defined. According to the infinite moving-average representation of $X(\mathbf{v})$, for fixed K the new process $\{X^{(K)}(\mathbf{v})\}$ may be defined from

$$X^{(K)}(\mathbf{v}) \equiv \varepsilon(\mathbf{v}) + \sum_{\mathbf{i}\in\mathcal{B}_K}\Theta_{\mathbf{i},0}\varepsilon(\mathbf{v}+\mathbf{i}). \tag{A.10}$$

Similarly, $\mathbf{U}^{(K)}(\mathbf{v})^\tau \equiv (X^{(K)}(\mathbf{v} + \mathbf{i}_1), \ldots, X^{(K)}(\mathbf{v} + \mathbf{i}_q))Y(\mathbf{v})$, $\mathbf{v} \in \mathbb{Z}^d$, and

$$\mathbf{X}^{(K)}(\mathbf{v}) \equiv \left(X^{(K)}(\mathbf{v} + \mathbf{i}_1), \ldots, X^{(K)}(\mathbf{v} + \mathbf{i}_q) \right)^\tau, \qquad \mathbf{C}_{\mathbf{j},0}^{(K)} \equiv \frac{E\left(\mathbf{X}^{(K)}(\mathbf{v})\mathbf{X}^{(K)}(\mathbf{v} + \mathbf{j})^\tau\right)}{\sigma^2} \qquad (A.11)$$

are defined here. For the same reasons as before, it holds that $E(\mathbf{U}^{(K)}(\mathbf{v})) = \mathbf{0}$ and that

$$\text{Cov}\left(\mathbf{U}^{(K)}(\mathbf{v}), \mathbf{U}^{(K)}(\mathbf{v} + \mathbf{j})\right) \equiv E\left(\mathbf{U}^{(K)}(\mathbf{v})\mathbf{U}^{(K)}(\mathbf{v} + \mathbf{j})^\tau\right) = \sigma^4\gamma_{\mathbf{j},0}\mathbf{C}_{\mathbf{j},0}^{(K)}, \quad \mathbf{j} \geq \mathbf{0}. \qquad (A.12)$$

For any vector $\boldsymbol{\lambda} \in \mathbb{R}^q$, it holds that $\{\boldsymbol{\lambda}^\tau\mathbf{U}^{(K)}(\mathbf{v})\}$ is a strictly stationary and K^*-dependent process, for a positive and finite integer K^*. The definition of K-dependent processes as well as a theorem for the asymptotic normality for strictly stationary and K-dependent processes on $\mathbb{Z}^d$ might be given similarly to the one-dimensional case by Brockwell and Davis [9]. Then

$$N^{-1/2} \sum_{\mathbf{v} \in \mathcal{S}^*} \boldsymbol{\lambda}^\tau\mathbf{U}^{(K)}(\mathbf{v}) \xrightarrow{D} N\left(0, \sigma^4\boldsymbol{\lambda}^\tau\mathbf{M}_K\boldsymbol{\lambda}\right), \qquad (A.13)$$

as $N \to \infty$ and under Condition C2, where $\mathbf{M}_K \equiv \gamma_{\mathbf{0},0}\mathbf{C}_{\mathbf{0},0}^{(K)} + \sum_{\mathbf{j} \in \mathcal{F}-\{\mathbf{0}\}} \gamma_{\mathbf{j},0}(\mathbf{C}_{\mathbf{j},0}^{(K)} + \mathbf{C}_{\mathbf{j},0}^{(K)\tau})$. Similarly, for

$$\mathbf{M} \equiv \gamma_{\mathbf{0},0}\mathbf{C}_{\mathbf{0},0} + \sum_{\mathbf{j} \in \mathcal{F}-\{\mathbf{0}\}} \gamma_{\mathbf{j},0}\left(\mathbf{C}_{\mathbf{j},0} + \mathbf{C}_{\mathbf{j},0}^\tau\right), \qquad (A.14)$$

it holds as $K \to \infty$ that $\boldsymbol{\lambda}^\tau\mathbf{M}_K\boldsymbol{\lambda} \to \boldsymbol{\lambda}^\tau\mathbf{M}\boldsymbol{\lambda}$. Using Chebychev's inequality, it may be verified that

$$P\left(\left| N^{-1/2} \sum_{\mathbf{v} \in \mathcal{S}^*} \boldsymbol{\lambda}^\tau\mathbf{U}(\mathbf{v}) - N^{-1/2} \sum_{\mathbf{v} \in \mathcal{S}^*} \boldsymbol{\lambda}^\tau\mathbf{U}^{(K)}(\mathbf{v}) \right| > \epsilon \right)$$
$$\leq \left(\frac{1}{\epsilon^2}\right)\left(\frac{N^*}{N}\right)\boldsymbol{\lambda}^\tau \text{Var}\left(\mathbf{U}(\mathbf{v}) - \mathbf{U}^{(K)}(\mathbf{v})\right)\boldsymbol{\lambda} \longrightarrow 0, \qquad (A.15)$$

as $K \to \infty$ and, thus, it holds that

$$N^{-1/2} \sum_{\mathbf{v} \in \mathcal{S}^*} \boldsymbol{\lambda}^\tau\mathbf{U}(\mathbf{v}) \xrightarrow{D} N\left(0, \sigma^4\boldsymbol{\lambda}^\tau\mathbf{M}\boldsymbol{\lambda}\right) \quad \text{or} \quad N^{-1/2} \sum_{\mathbf{v} \in \mathcal{S}^*} \mathbf{U}(\mathbf{v}) \xrightarrow{D} N\left(\mathbf{0}, \sigma^4\mathbf{M}\right), \qquad (A.16)$$

as $N \to \infty$ and under Condition C2. According to (A.14), the (n, m)th element of $\mathbf{M}$ is equal to

$$
\begin{aligned}
\gamma_{0,0} c_{0+\mathbf{i}_m-\mathbf{i}_n,0} &+ \sum_{\mathbf{j}\in\mathcal{F},\mathbf{j}>0} \left(\gamma_{\mathbf{j},0} c_{\mathbf{j}+\mathbf{i}_m-\mathbf{i}_n,0} + \gamma_{\mathbf{j},0} c_{\mathbf{j}+\mathbf{i}_n-\mathbf{i}_m,0} \right) \\
&= \gamma_{0,0} c_{0+\mathbf{i}_m-\mathbf{i}_n,0} + \sum_{\mathbf{j}\in\mathcal{F},\mathbf{j}>0} \gamma_{\mathbf{j},0} c_{\mathbf{j}+\mathbf{i}_m-\mathbf{i}_n,0} + \sum_{\mathbf{j}\in\mathcal{F},\mathbf{j}<0} \gamma_{\mathbf{j},0} c_{-\mathbf{j}+\mathbf{i}_n-\mathbf{i}_m,0} \\
&= \gamma_{0,0} c_{0+\mathbf{i}_m-\mathbf{i}_n,0} + \sum_{\mathbf{j}\in\mathcal{F},\mathbf{j}>0} \gamma_{\mathbf{j},0} c_{\mathbf{j}+\mathbf{i}_m-\mathbf{i}_n,0} + \sum_{\mathbf{j}\in\mathcal{F},\mathbf{j}<0} \gamma_{\mathbf{j},0} c_{\mathbf{j}+\mathbf{i}_m-\mathbf{i}_n,0} \\
&= \sum_{\mathbf{j}\in\mathcal{F}} \gamma_{\mathbf{j},0} c_{\mathbf{j}+\mathbf{i}_m-\mathbf{i}_n,0}
\end{aligned}
\tag{A.17}
$$

and back to the general Yule-Walker equations with $\mathbf{M} \equiv \mathbf{I}_q$, that is, the identity matrix of order q. Equation (A.16) may be rewritten

$$
N^{-1/2} \sum_{\mathbf{v}\in S^*} \mathbf{U}(\mathbf{v}) \xrightarrow{D} N\left(\mathbf{0}, \sigma^4 \mathbf{I}_q\right),
\tag{A.18}
$$

as $N \to \infty$ and under Condition C2. Combining (A.6) and (A.18) with Proposition 4.2, it can be written that

$$
N^{1/2}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{D} N\left(\mathbf{0}, \left(\boldsymbol{\Theta}_0^\tau \boldsymbol{\Theta}_0\right)^{-1}\right),
\tag{A.19}
$$

as $N \to \infty$ and under Condition C2. The proof will be completed when it is shown that $\boldsymbol{\Theta}_0^\tau\boldsymbol{\Theta}_0 = \mathbf{W}_q^*$. Indeed, for the vector $\mathbf{W} = (W(-\mathbf{i}_1), \ldots, W(-\mathbf{i}_q))^\tau$ and for $\boldsymbol{\xi} = \boldsymbol{\Theta}_0^\tau \mathbf{W} + \mathbf{R}$ with $\mathbf{R}$ being a $(q \times 1)$ random vector that is independent of $\mathbf{W}$, since it is a linear function of $W(-\mathbf{i}_1 - \mathbf{i})$, $\mathbf{i} > 0$, $\mathbf{i} \neq \mathbf{i}_2 - \mathbf{i}_1, \ldots, \mathbf{i}_q - \mathbf{i}_1$, the required result might be obtained then, as

$$
\mathrm{Var}(\mathbf{W} \mid \mathbf{R}) \equiv \mathrm{Var}\left(\mathbf{W} \mid W(-\mathbf{i}_1 - \mathbf{i}),\ \mathbf{i} > 0,\ \mathbf{i} \neq \mathbf{i}_2 - \mathbf{i}_1, \ldots, \mathbf{i}_q - \mathbf{i}_1\right) = \mathrm{Var}(\mathbf{W}) = \mathbf{I}_q.
\tag{A.20}
$$

Note that the decomposition $\mathbf{W}_q^* = \boldsymbol{\Theta}_0^\tau \boldsymbol{\Theta}_0$, as it is justified in the end of proof of Theorem 4.3, guarantees that $\mathbf{W}_q^*$ has a nonzero determinant.

Acknowledgments

References

[1] P. Whittle, "On stationary processes in the plane," *Biometrika*, vol. 41, pp. 434–449, 1954.

[2] X. Guyon, "Parameter estimation for a stationary process on a d-dimensional lattice," *Biometrika*, vol. 69, no. 1, pp. 95–105, 1982.

[3] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society*, vol. 36, pp. 192–236, 1974.

[4] J. Besag, "Statistical analysis of non-lattice data," *Statistician*, vol. 24, pp. 179–195, 1975.

[5] C. Dimitriou-Fakalou, "Modelling data observed irregularly over space and regularly in time," *Statistical Methodology*, vol. 6, no. 2, pp. 120–132, 2009.

[6] Q. Yao and P. J. Brockwell, "Gaussian maximum likelihood estimation for ARMA models II: spatial processes," *Bernoulli*, vol. 12, no. 3, pp. 403–429, 2006.

[7] B. D. O. Anderson and E. I. Jury, "Stability of multidimensional digital filters," *IEEE Transactions on Circuits and Systems*, vol. 21, no. 2, pp. 300–304, 1974.

[8] C. Dimitriou-Fakalou, "Modified Gaussian likelihood estimators for ARMA models on $\mathbb{ZZ}^d$," *Stochastic Processes and Their Applications*, vol. 119, no. 12, pp. 4149–4175, 2009.

[9] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, Springer Series in Statistics, Springer, New York, NY, USA, 2nd edition, 1991.

[10] R. Dahlhaus and H. Künsch, "Edge effects and efficient parameter estimation for stationary random fields," *Biometrika*, vol. 74, no. 4, pp. 877–882, 1987.

[11] E. J. Hannan, "The asymptotic theory of linear time-series models," *Journal of Applied Probability*, vol. 10, pp. 130–145, 1973.