On estimation of parameters for spatial autoregressive model*

Youri Davydov[1], Vygantas Paulauskas[2]
[1] Université des Sciences et Technologies de Lille, Laboratoire Paul Painlevé
[2] Vilnius University, Department of Mathematics and Informatics, and Institute of Mathematics and Informatics, Vilnius

Abstract

In the paper asymptotic properties of the estimated coefficients of multi-indexed autoregressive model are investigated. Considering Least Squares estimates we use some kind of self-normalization and obtain limit normal law independent of unknown parameters. In the proof of this result we use recent result on the Central Limit Theorem for dependent summands.

Keywords: Autoregressive models; estimation of parameters; random fields.

AMS 2000 Subject Classification: Primary-62M10, 62M40;
Secondary 91B72

*Supported by bilateral Lithuania-France research program Gilibert.

1 Introduction and formulation of results

We consider a real-valued random field $X_{\bar{k}}$ with $\bar{k} \in \mathbb{Z}^d$, satisfying the following relation

$$X_{\bar{k}} = \sum_{\bar{b} \in \Lambda} a_{\bar{b}} X_{\bar{k}-\bar{b}} + \varepsilon_{\bar{k}}, \tag{1}$$

where $\bar{k} = (k_1, \dots, k_d)$, $\bar{b} = (b_1, \dots, b_d)$, $\bar{k} - \bar{b} = (k_i - b_i, i = 1, \dots, d)$, Λ is some subset of $(\mathbb{Z})^d \setminus \{\bar{0}\}$ and $\varepsilon_{\bar{k}}$, $\bar{k} \in \mathbb{Z}^d$, are independent, identically distributed (iid) random variables with finite second moment. It seems that in literature most commonly used name for such fields (of course, with much more general assumptions on random variables $\varepsilon_{\bar{k}}$) is spatial autoregressive process (see, for example, [1] and references therein) and in analogy with standard notation in time series $AR(p), ARMA(p,q)$ we shall denote such process by $SAR(d, \Lambda)$, showing two main parameters d and Λ, or if dimension d is fixed, simply $SAR(\Lambda)$. During last decades such processes, especially in case $d = 2$ were rather deeply investigated, there is a vast literature devoted to SAR processes, starting with fundamental paper by Whittle [18], later most papers appearing mainly in engineering literature (IEEE journals and proceedings). This is due to the fact that most applications (image recognition, segmentation and restoration, textures models, etc.) deal with models with indices on plane. Also one can mention the so-called time-space auto-regression models, which formally can be considered as SAR models, but they are specific in a sense that one coordinate of indices is separated and denotes time, while the others are used to index variables in "space" (or fixed locations, in this case the term "panel data" is used). In such models there is usual lag in time and lag in space, which generally is defined by the so-called weight matrix (see [4] or recent paper [8], where such models are discussed).

While in time series there is the natural notion of "past" and "future", in case of multidimensional index there is no natural such notion, therefore it is not easy to say what sets Λ could be considered as natural. One possible way is to consider requirement $\Lambda \subset (\mathbb{Z}_+)^d \setminus \{\bar{0}\}$, this will lead to the so-called quarter-plane autoregressive models. Another class of examples of SAR processes are so called nonsymmetric half-plane models, in which a set Λ is defined in more complicated way. Mathematical theory of these processes is very well developed (see, for example, [6], [11], [12] and references there). In this paper we shall consider only quarter-plane autoregressive models. Also we shall consider only sets Λ consisting of finite number of elements and we denote $m = m(\Lambda) = |\Lambda|$, where $|\Lambda|$ stands for the number of elements of a set Λ. Here are some examples of sets, satisfying this condition:

$$\Lambda_p = \Big\{ \bar{b} \in \mathbb{Z}^d,\ b_i \geq 0, i = 1, \ldots, d, 0 < \sum_{i=1}^{d} b_i \leq p \Big\},$$

$$\Lambda_{\bar{n}} = \Big\{ \bar{b} \in \mathbb{Z}^d,\ 0 \leq b_i \leq n_i, i = 1, \ldots, d, \sum_{i=1}^{d} b_i > 0 \Big\},$$

In what follows we shall drop the bar in the notation for multi-indices if this will not cause ambiguity.

We consider estimation of the parameters of the model (1), assuming that the vector of these parameters $a = (a_i, i \in \Lambda)$ is such that there exists unique stationary solution of (1) which is given by infinite series

$$X_k = \sum_{i \in \mathbb{Z}_+^d} b_i \varepsilon_{k-i},$$

with $\sum_{i \in \mathbb{Z}_+^d} b_i^2 < \infty$. As far as we know, at present necessary and sufficient conditions for the existence of a stationary solution are known only in case $d = 2$. There are known necessary or sufficient conditions, for example, a sufficient condition is $\sum_{i \in \Lambda} |a_i| < 1$.

Suppose, that we observe values of X_k from (1) with $k \in D_n \subset \mathbb{Z}^d$, where D_n is a sequence of sets such that $b_n := |D_n| \to \infty$ as $n \to \infty$. Later on we shall impose additional conditions on D_n. There are several types of estimators of the vector of parameters a, most common are Least Squares (LS), Yule-Walker (YW) type, and Maximum likelihood (ML) estimators. In [15], [16] it was stated that LS and YW estimates are asymptotically the same and for small samples and strongly correlated series YW estimates can be considerably more biased (see also [17] and [9]). Recently Basu and Reinsel claimed (see [2]) that in [15] in the proof there is an error and Gaussian asymptotic distribution of YW estimator contains an asymptotic bias term.

We consider LS estimators of the parameter $a = (a_i, i \in \Lambda)$ of the model (1). Namely, let $\hat{a}_n = (\hat{a}_{n,i}, i \in \Lambda)$ be the value of $a = (a_i, i \in \Lambda)$ which minimizes the function

$$F(a) = \sum_{k \in D_n^*} \big(X_k - \sum_{i \in \Lambda} a_i X_{k-i} \big)^2,$$

where $D_n^* = \cap_{i \in \Lambda \cup \{0\}} \{D_n + i\}$ (the meaning of the set D_n^* is such that if $k \in D_n^*$, then $k \in D_n$ and $k - i \in D_n$ for all $i \in \Lambda$). Thus the estimator $\hat{a}$ is obtained as a solution of the system of linear equations

$$\frac{\partial F(a)}{\partial a_i} = 0, \quad i \in \Lambda,$$

and this system can be written as

$$\sum_{j \in \Lambda} a_j \sum_{k \in D_n^*} X_{k-j} X_{k-i} = \sum_{k \in D_n^*} X_k X_{k-i}, \quad i \in \Lambda. \tag{2}$$

In order to write (2) in usual matrix form we write $a = (a_i, i \in \Lambda)$ as a vector. We recall that $m = |\Lambda|$ and let $\varphi : \Lambda \to \{1, 2, \ldots, m\}$ be one-to-one mapping and if $\varphi(i) = j,\ i \in \Lambda,\ 1 \leq j \leq m$, then $\varphi^{-1}(j) = i$. Using this mapping we can write $a = (a(1), \ldots, a(m))$, where $a(j) = a_{\varphi^{-1}(j)}$, and $\hat{a}_n = (\hat{a}_n(1), \ldots, \hat{a}_n(m))$. Since this mapping will be fixed for all the rest of the paper, we shall use also previously introduced more simple notation $\hat{a}_{n,i}, i \in \Lambda$ for the vector $\hat{a}_n$, understanding that $\hat{a}_{n,i} = \hat{a}_n(\varphi(i))$. This remark applies for other vectors to be introduced later. Denote by $\mathbf{X}_n$ a matrix of the order $m \times m$ with elements $\mathbf{X}_n(k, l),\ k, l = 1, \ldots, m$, where $\mathbf{X}_n(k, l) = \mathbf{X}_{n,i,j}$ with $k = \varphi(i),\ l = \varphi(j)$ and

$$\mathbf{X}_{n,i,j} = \sum_{k \in D_n^*} X_{k-j} X_{k-i}, \quad i, j \in \Lambda.$$

Let Y_n stand for the vector $Y_n(j),\ 1 \leq j \leq m$ with $Y_n(j) = Y_{n,i},\ \varphi(i) = j$, and

$$Y_{n,i} = \sum_{k \in D_n^*} X_k X_{k-i}, \quad i \in \Lambda.$$

Then system (2) can be written as

$$\mathbf{X}_n \hat{a}_n = Y_n. \tag{3}$$

Substituting to the right side of (3) instead of X_k its expression (1), we get

$$\mathbf{X}_n (\hat{a}_n - a) = W_n, \tag{4}$$

where $W_n = \{\sum_{k \in D_n^*} \varepsilon_k X_{k-i}, \quad i \in \Lambda\}$. If we denote $b_n^* = |D_n^*|$, then from (4) it follows

$$(b_n^*)^{1/2} (\hat{a}_n - a) = \left(b_n^{*\,-1} \mathbf{X}_n \right)^{-1} (b_n^*)^{-1/2} W_n.$$

Traditional way (see [15], [16] and other authors) to obtain asymptotic normality of LS estimator $\hat{a}_n$ is to prove the following two relations

$$b_n^{*\,-1} \mathbf{X}_n \xrightarrow{P} \mathbf{X}_0 \tag{5}$$

(or a.s.) and

$$(b_n^*)^{-1/2} W_n \xrightarrow{d} W_0, \tag{6}$$

where $\xrightarrow{d}$ and $\xrightarrow{P}$ stands for the convergence in distribution and in probability, respectively. But in this way the covariance matrix of the limit Gaussian mean zero vector W_0 depends on unknown parameters σ^2 and a. Namely, due to the fact that summands in the sum W_n are stationary and uncorrelated (this follows from independence of random variables $\varepsilon_i,\ i \in \mathbb{Z}^d$) the covariance matrix of a limit normal law is $\sigma^2 R_0$, where $R_0 = \mathrm{cov}(X_{k-i},\ i \in \Lambda)$. Clearly, matrix R_0 is a function (and rather complicated) of coefficients a. In order to get limit distribution independent of unknown parameters we use some kind of self-normalization. By means of observed values $\{X_k, k \in D_n\}$ and constructed LS estimator $\hat{a}_n$ we define empirical residuals

$$\hat{\varepsilon}_k = X_k - \sum_{i \in \Lambda} \hat{a}_{n,i} X_{k-i},\ k \in D_n^*$$

and the estimator for unknown σ^2

$$\hat{\sigma}_n^2 = (b_n^*)^{-1} \sum_{k \in D_n^*} (\hat{\varepsilon}_k - \bar{\varepsilon}_n)^2,$$

where

$$\bar{\varepsilon}_n = (b_n^*)^{-1} \sum_{k \in D_n^*} \hat{\varepsilon}_k.$$

Let $\hat{R}_n = \{\hat{r}_n^{(p,q)},\ p, q = 1, \dots, m\}$, where $\hat{r}_n^{(p,q)} = \hat{r}_{n,\varphi^{-1}(p),\varphi^{-1}(q)}$ and

$$\hat{r}_{(n,i,j)} = (b_n^*)^{-1} \sum_{k \in D_n^*} (X_{k-i} - \bar{X}_{k-i})(X_{k-j} - \bar{X}_{k-j}),\ i, j \in \Lambda,$$

$$\bar{X}_{k-i} = (b_n^*)^{-1} \sum_{k \in D_n^*} X_{k-i}.$$

Finally we introduce conditions on D_n and Λ. For $x \in \mathbb{R}^d$ let $\|x\|_\infty = \max_{1 \le j \le m} |x_j|$ and let $B(r, c)$ stand for the closed ball of radius r and center c with respect to $\|\cdot\|_\infty$, $B(r) := B(r, 0)$. For any $\Gamma \subset \mathbb{Z}^d$ let

$$\partial\Gamma = \{i \in \Gamma : \exists\ j \in \mathbb{Z}^d \text{ such that } j \notin \Gamma \text{ and } \ \|i - j\|_\infty = 1\}$$

and for $a \ge 1$

$$\partial_a \Gamma = \{i \in \mathbb{Z}^d : i + B(a) \cap \Gamma \neq \emptyset \text{ and } i + B(a) \cap \Gamma^c \neq \emptyset\}.$$

For convenience of formulations we make convention $\partial_0\Gamma := \partial\Gamma$. Inner radius of a set $C \subset \mathbb{Z}^d$ is defined as follows

$$\rho(C) = \sup\{r : \exists\ c \text{ such that } B(r,c) \cap \mathbb{Z}^d \subset C\}.$$

Condition (**A**) on D_n: D_n for all n is bounded, $|D_n| \to \infty$, and

$$|D_n|^{-1}|\partial D_n| \to 0, \text{ as } n \to \infty$$

In Proposition 6 we show that if sets D_n for all n are convex and bounded and $|D_n| \to \infty$, then they satisfy condition (**A**).

Now we are able to formulate our main result.

Theorem 1. *Let $X_k,\ k \in D_n$ be observed values from the model* (1)*, for which we assume that $\varepsilon_i, i \in \mathbb{Z}^d$ are iid with mean zero and finite variance σ^2 and the vector of parameters a is such, that* (1) *has a stationary solution. If D_n satisfy condition* (**A**) *and Λ is finite then*

$$\hat{\sigma}_n^{-1} b_n^{1/2} \hat{R}_n^{1/2} (\hat{a}_n - a) \xrightarrow{d} \tilde{W}_0, \tag{7}$$

where $\tilde{W}_0$ is mean zero Gaussian vector with identity covariance matrix.

Comparing with previous results of [15], [16] or [2] the novelty of our result is the limit Gaussian law, not depending on unknown parameters of the model under consideration. Also we consider more general sets D_n and for the proof of the theorem we use existing results in limit theorems for weakly dependent random fields. Namely, we show that rather straight application of CLT result from Dedecker's paper [5] (see Theorem 8 bellow) proves (6) while results from ergodic theory for stationary fields give us (5) and other LLN type relations.

2 Auxiliary results

In this section we collect results which are main ingredients in the proof of our theorem. We start with some results from ergodic theory (see, for example [10]).

Let $\xi = \{\xi_i,\ i \in \mathbb{Z}^d\}$ be a stationary random field with values in $\mathbb{R}$ (Theorem 2 and its corollary holds for random fields with values in $\mathbb{R}^m$, but for our purposes it is sufficient to take $m = 1$). Let $\tau_j,\ j \in \mathbb{Z}^d$, stand for the translation operator, defined on $(\mathbb{R})^{\mathbb{Z}^d}$: $\tau_j(\xi) = \{\xi_{i+j},\ i \in \mathbb{Z}^d\}$. Denote by $\mathcal{T}$ σ-algebra of invariant sets: a set $A \in \mathcal{T}$ if and only if $\tau_j(A) = A$ for all $j \in \mathbb{Z}^d$. The σ-algebra of invariant sets of a random field ξ is defined as $\xi^{-1}(\mathcal{T})$. The random field ξ is called ergodic if its σ-algebra of invariant sets $\xi^{-1}(\mathcal{T})$ is trivial. For example, if $\{\xi_i,\ i \in \mathbb{Z}^d\}$ are i.i.d. random variables, then ξ is stationary ergodic random field. If $\eta = \{\eta_i = g(\tau_i(\xi)),\ i \in \mathbb{Z}^d\}$, where g – any measurable mapping from $(\mathbb{R})^{\mathbb{Z}^d}$ to any measurable space and if ξ is ergodic, then η is ergodic too. Main result which will be used in our proof can be formulated as follows.

Theorem 2. (Corollary 14.A5 p.304 in [7]) *Let $\xi = \{\xi_i,\ i \in \mathbb{Z}^d\}$ be a stationary random field with values in $\mathbb{R}$ and $E|\xi_i| < \infty$. Let $D_n \subset \mathbb{Z}^d, n \geq 1$, be a sequence of bounded (or finite, which is the same) sets such that $b_n = |D_n| \to \infty$ and for any fixed $i \in \mathbb{Z}^d$*

$$\frac{|(D_n + i)\Delta D_n|}{|D_n|} \to 0, \tag{8}$$

as $n \to \infty$. Then

$$\bar{\xi}_n := b_n^{-1} \sum_{i \in D_n} \xi_i \xrightarrow{L_1} E(\xi_0 |\, \xi^{-1}(\mathcal{T})). \tag{9}$$

Corollary 3. *Suppose that $\xi = \{\xi_i,\ i \in \mathbb{Z}^d\}$ is a stationary ergodic random field with values in $\mathbb{R}$ and $g : (\mathbb{R})^{\mathbb{Z}^d} \to \mathbb{R}$ is integrable. If sequence $D_n,\ n \geq 1$ is as in Theorem 2, then*

$$b_n^{-1} \sum_{i \in D_n} g(\tau_i(\xi)) \xrightarrow{L_1} Eg(\xi). \tag{10}$$

Remark 4. Although in [7] these results for simplicity were formulated for increasing sequence of cubes, but it is easy to see (and it is mentioned at the end of the corresponding chapter in the book) that one can take sets D_n, satisfying condition (8).

Now we shall formulate some useful properties of sets D_n, satisfying condition (**A**).

Proposition 5. *Suppose that sets D_n satisfy condition* (**A**). *Then*

a) for any $m \geq 1$

$$\partial_m(D_n) \subset \cup_{|i| \leq 1} \partial_{m-1}(D_n + i), \tag{11}$$

therefore there exists a constant $C(d, m)$, depending on d and m, but not on n, such that

$$|\partial_m(D_n)| \leq \mathrm{C}(d, m)|\partial D_n|; \tag{12}$$

b) the relation $\rho(D_n) \to \infty$ holds;

c) the relation (8) holds.

Proof. We want to prove (11) and at first consider $m \geq 2$. Let $x \in \partial_m(D_n)$, this means that

$$i + B(m) \cap D_n \neq \emptyset \text{ and } i + B(m) \cap D_n^c \neq \emptyset. \tag{13}$$

Suppose, that x does not belong to the union of sets on the right-hand side of (11), that is , for all i, $|i| \leq 1$

$$x \notin \partial_{m-1}(D_n + i).$$

It is easy to see that this is equivalent to one of the following two relations:

$$(x - i) + B(m-1) \subset D_n \text{ or } \ (x - i) + B(m-1) \subset D_n^c.$$

Since $|i| \leq 1$ and $m - 1 \geq 1$, then

$$\bigcap_{|i| \leq 1} \{(x - i) + B(m-1)\} \neq \emptyset,$$

therefore we have that either

$$(x - i) + B(m-1) \subset D_n \text{ for all } \ i, \ |i| \leq 1,$$

or

$$(x - i) + B(m-1) \subset D_n^c \text{ for all } \ i, \ |i| \leq 1.$$

In the first case we get that $x + B(m) \subset D_n$, in the second one - $x + B(m) \subset D_n^c$, and both cases contradict (13). The obtained contradiction proves (11) in case $m \geq 2$. The case $m = 1$ must be considered separately, since definition of ∂D_n is not obtained from definition of $\partial_m D_n$, taking $m = 0$. But the changes in the proof are simple, so we do not provide the proof of this case.

Now we shall prove *b*). Suppose that sets D_n satisfy condition (**A**), but $\rho(D_n) \leq c$ for all n with some finite constant c. Let us denote $c_1 = c + 1$ and take $x_1 \in D_n$ and $B(c_1, x_1)$. Since $\rho(D_n) \leq c$ there is an element $y_1 \in \partial D_n \cap B(c_1, x_1)$. Now take $B(2c_1 + 1, x_1)$ and without loss of generality we may suppose that $|D_n|$ is sufficiently large, that $D_n \cap B(2c_1 + 1, x_1) \neq \emptyset$. Then we take $x_2 \in D_n \setminus B(2c_1 + 1, x_1)$ and we find $y_2 \in \partial D_n \cap B(c_1, x_2)$, and $y_2 \neq y_1$, since $B(c_1, x_1) \cap B(c_1, x_2) = \emptyset$. Continuing this process after some steps (but this process will stop, since D_n is finite) we get a set of elements $y_1, y_2, \dots, y_{k_n}$, which are all different and belong to ∂D_n, therefore we have

$$|\partial D_n| \geq k_n.$$

On the other hand, since the process had stopped, we have $\bigcup_{i=1}^{k_n} B(2c_1 + 1, x_i) \supset D_n$, therefore

$$|D_n| \leq k_n (2(2c_1+1))^d.$$

From the last two estimates we get

$$\frac{|\partial D_n|}{|D_n|} \geq \frac{1}{(2(2c_1+1))^d},$$

but this contradicts condition (**A**).

It remains to prove c). Let i be fixed, $||i|| = m$, we shall prove that

$$(D_n + i)\Delta D_n \subset \partial_m D_n \cup \partial_m (D_n + i). \tag{14}$$

Let $x \in (D_n + i)\Delta D_n$ and suppose that $x \in D_n$, but $x \notin D_n + i$. From these two relations it follows that

$$x + B(m) \cap D_n \neq \emptyset \text{ and } x + B(m) \cap D_n^c \neq \emptyset$$

and, therefore, $x \in \partial_m D_n$. The case $x \in D_n + i$, but $x \notin D_n$ can be considered in a similar way, we get in this case $x \in \partial_m (D_n + i)$, and relation (14) is proved. From this relation, inequality (12) and condition ((**A**) we get relation (8). The proposition is proved.

Proposition 6. *If sets $D_n, n \geq 1$ are convex (the convexity of a set $D \subset \mathbb{Z}^d$ is understood as follows: there exists a convex set $V \subset \mathbb{R}^d$ such that $D = V \cap \mathbb{Z}^d$) and $|D_n| \to \infty$, then for any fixed $l \geq 1$*

$$(b_n)^{-1} |\{\partial_l D_n\}| \to 0, \tag{15}$$

and, in particular, sets D_n satisfy (**A**).

The proof of this result is based on classical Gauss argument, and the result itself most probably is known. But since it is easier to proof it than to find appropriate reference, we provide the sketch of the proof. It is clear that the convex body V_n which is in definition of convexity of D_n is simply convex hull of points from D_n. Then

$$|\partial_l D_n| = \text{vol}(\cup_{i \in \partial_l D_n} B(i, 1/2)) \leq \text{vol}((\partial V_n)^{2l\sqrt{d}}),$$

here A^ϵ stands for usual (with respect to Euclidean distance) ϵ-neighborhood of a set $A \subset \mathbb{R}^d$, ∂A and vol(A) denote a boundary and d-dimensional volume, respectively, of a set $A \subset \mathbb{R}^d$. From [13] we have

$$\text{vol}(V_n) \leq C(d) b_n.$$

(In [13] one can find explicit expression of a constant $C(d)$.) Therefore

$$(b_n)^{-1}|\partial_l D_n| \le C(d)\frac{\mathrm{vol}((\partial V_n)^{2l\sqrt{d}})}{\mathrm{vol}(V_n)}.$$

One can show that the relation

$$\frac{\mathrm{vol}((\partial V_n)^{2l\sqrt{d}})}{\mathrm{vol}(V_n)} \to 0, \quad \text{as } n \to \infty,$$

will follow if we prove that

$$m(\epsilon) = \sup \frac{\mathrm{vol}((\partial V)^{\epsilon})}{\mathrm{vol}(V)} \to 0, \quad \text{as } \epsilon \to 0, \tag{16}$$

where sup is taken over all convex bodies containing unit ball with a center at 0. We have (for polytopes this inequality can be proved directly, general case is proved by approximation of V by polytopes)

$$\mathrm{vol}((\partial V)^{\epsilon}) \le 2\mathrm{vol}(V^{\epsilon} \setminus V)$$

and

$$V^{\epsilon} \subset V + \epsilon V = (1+\epsilon)V,$$

where sum of sets is in the sense of Minkovski. Therefore

$$\mathrm{vol}(V^{\epsilon}\setminus V) = \mathrm{vol}(V^{\epsilon}) - \mathrm{vol}(V) \le \mathrm{vol}((1+\epsilon)V) - \mathrm{vol}(V) = ((1+\epsilon)^d - 1)\mathrm{vol}(V),$$

and we easily get (16), what finishes the proof of the proposition.

Since we intend to apply results of Theorem 2 and Corollary 3 to statistics $\hat{\sigma}_n^2$, $\hat{R}_n$, and $\mathbf{X}_n$ which are expressed by sums over the sets D_n^*, we must show that these sets under assumptions made for D_n satisfy conditions of Theorem 2.

Proposition 7. *If the sets D_n satisfy $(\mathbf{A})$ and Λ is finite set then the sets D_n^* satisfy $(\mathbf{A})$, and $b_n \sim b_n^*$, where $b_n^* = |D_n^*|$.*

Proof. First we prove that

$$r_n^* := \rho(D_n^*) \to \infty. \tag{17}$$

From Proposition 5 we have that $r_n = \rho(D_n) \to \infty$. Let $c \in \mathbb{R}^d$ be such that $B(r_n, c) \cap \mathbb{Z}^d \subset D_n$ (existence of such c follows from the definition of inner radius). Without loss of generality we can assume that n is sufficiently large and such that $r_n \ge m\sqrt{d}$ (we recall that $m = |\Lambda|$). Then for any $i \in \Lambda$ we have $B(r_n - m\sqrt{d}, c) \cap \mathbb{Z}^d \subset D_n + i$, therefore $B(c, r_n - m\sqrt{d}) \cap \mathbb{Z}^d \subset D_n^*$. This means that $r_n^* \ge r_n - m\sqrt{d}$ and (17) follows. Now we prove that

$$b_n^*/b_n \to 1. \tag{18}$$

To this aim we show that $D_n \setminus D_n^* \subset \partial_m D_n$. Let $j \in D_n$ and $j \notin D_n^*$. Then there exists $i \in \Lambda$ such that $j \notin D_n + i$, or that $j - i \notin D_n$. Therefore, $j \in \partial_m D_n$, since cube $j + B(m)$ contains a point $j \in D_n$ and a point $j - i \notin D_n$. Thus we have

$$b_n^* \le b_n \le b_n^* + |\partial_m D_n|. \tag{19}$$

Now we use (12) from Proposition 5 and we get

$$\frac{|\partial_m D_n|}{b_n} \to 0 \tag{20}$$

and (18) follows. The last relation to be proved is

$$|\partial D_n^*|(b_n^*)^{-1} \to 0. \tag{21}$$

From the definition of the set D_n^* we have

$$|\partial D_n^*| \le \sum_{i \in \Lambda} |\partial(D_n + i)| = (m+1)|\partial D_n|.$$

This inequality and (19) imply

$$\frac{|\partial D_n^*|}{b_n^*} \le \frac{m+1}{1 - |\partial_m D_n| b_n^{-1}} \frac{|\partial D_n|}{b_n}.$$

Taking into account (20) and condition (**A**) we get (21). The proposition is proved.

The next result which we shall need is the CLT for stationary random fields. It turns out that for our purposes the most appropriate is the following rather general theorem from [5]. In order to formulate this result we introduce some notations from [5]. The lexicographic order on $\mathbb{Z}^d$ is defined as follows: if $i, j \in \mathbb{Z}^d$ then $i <_{lex} j$ means that either $i_1 < j_1$ or for some $p \in \{2, \ldots, d\}$, $i_p < j_p$ and $i_m = j_m$ for all $1 \le m < p$. Let us define sets $V_i^k, \;\; k \in \mathbb{N}, \; i \in \mathbb{Z}^d$, by the following relations

$$V_i^1 = \{j \in \mathbb{Z}^d : \; j <_{lex} i\},$$

and for $k \geq 2$

$$V_i^k = V_i^1 \cap \{j \in \mathbb{Z}^d : ||i - j||_\infty \geq k\}.$$

Let $\xi = \{\xi_i, \ i \in \mathbb{Z}^d\}$ be a real valued strictly stationary random field, $E\xi_0 = 0, \quad E\xi_0^2 < \infty$. For any $\Gamma \subset \mathbb{Z}^d$ let $\mathcal{F}_\Gamma = \sigma(\xi_j, \ j \in \Gamma)$ and

$$S_\Gamma = \sum_{i \in \Gamma} \xi_i.$$

We assume that $\Gamma_n, \ n \geq 1$, is a sequence of finite subsets of $\mathbb{Z}^d$ satisfying the following conditions:

$$\lim_{n \to \infty} |\Gamma_n| = +\infty \quad \text{and} \quad \lim_{n \to \infty} |\Gamma_n|^{-1} |\partial \Gamma_n| = 0. \tag{22}$$

Theorem 8. (Theorem 1 in [5]). *Assume that sequence $\Gamma_n, \ n \geq 1$, of finite subsets of $\mathbb{Z}^d$ satisfies the condition (22) and a random field introduced above is ergodic and satisfies condition*

$$\sum_{k \in V_0^1} E|\xi_k E(\xi_0 | \mathcal{F}_{V_0^{||k||}})| < \infty. \tag{23}$$

Then the series $\eta = \sum_{k \in \mathbb{Z}^d} E(\xi_0 \xi_k)$ converges and random variable $|\Gamma_n|^{-1/2} S_{\Gamma_n}$ converges in distribution to $N(0, \eta)$.

Remark 9. We formulated result from [5] under additional assumption of ergodicity in order to avoid the notion of the so-called $\mathcal{U}$-stable convergence, which is needed in general case when η is random.

3 Proof of Theorem 1

Relation (7) will follow from the relation

$$(\hat{\sigma}_n^2, \hat{R}_n, {b_n^*}^{-1} \mathbf{X}_n \ , (b_n^*)^{-1/2} W_n) \xrightarrow{d} (\sigma^2, R_0, \mathbf{X}_0, W_0) \tag{24}$$

by continuous mapping theorem. In its turn the relation (24) will follow if we prove

$$\hat{\sigma}_n^2 \xrightarrow{P} \sigma^2, \tag{25}$$

$$\hat{R}_n, \xrightarrow{P} R_0, \tag{26}$$

and relations (5) and (6) and then use the well-known result from [3], stating that if $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{P} a$, then $(X_n, Y_n) \xrightarrow{d} (X, a)$. Although for our main result it is sufficient to prove weak consistency of estimators in relations (5), (25), and (26), here it is necessary to mention, that, under little bit stronger conditions on sets D_n and using results from [10] or [7], it is possible to prove strong consistency of these estimators. Let us note, that $X = (X_t, t \in \mathbb{Z}^d)$ is a stationary solution of (1), therefore it can be written as a rather simple functional of $\varepsilon = (\varepsilon_i, i \in \mathbb{Z}^d)$ with $\varepsilon_i,\ i \in \mathbb{Z}^d$, being iid random variables. Thus, X is ergodic and we can apply Theorem 2 or Corollary 3 with appropriately chosen function g. Since statistics $\hat{\sigma}_n^2, \hat{R}_n$, and $\mathbf{X}_n$ are expressed by sums over the sets D_n^* we use Proposition 7, allowing to apply results of Theorem 2 with sets D_n^* instead of D_n.

Having relations (25), (26), and (5), it remains to prove (6). For this aim we shall show that Theorem 8 is applicable. As usual, to prove CLT in $\mathbb{R}^m$ one can use reduction to one-dimensional CLT by Cramer-Wold device, that is, one must prove that for any $\alpha \in \mathbb{R}^m,\ \alpha = (\alpha_1, \dots, \alpha_m),\ \alpha_1^2 + \cdots + \alpha_m^2 = 1$ random variables $((b_n^*)^{-1/2} W_n, \alpha)$ converge to univariate normal law. From this relation it will follow

$$(b_n^*)^{-1/2} W_n \xrightarrow{d} N(0, \sigma^2 R_0).$$

Let us denote

$$Z_k = \sum_{i=1}^m \alpha_i X_{k-\varphi^{-1}(i)} \varepsilon_k, \quad k \in \mathbb{Z}^d.$$

We want to prove CLT for

$$S_n = (b_n^*)^{-1/2} \sum_{k \in D_n^*} Z_k$$

and for this aim we use Theorem 8 with a sequence of sets D_n^* as Γ_n and random field $Z = (Z_k)$ as ξ. Conditions for D_n^* follow from Proposition 7. Since random field Z is ergodic, σ-algebra of invariant sets $Z^{-1}(\mathcal{T})$ is trivial, η is non-random (and it is easy to check that $\eta = E(\sum_{i=1}^m \alpha_i X_{-\varphi^{-1}(i)})^2$ is a variance of the univariate limit normal law). Therefore it remains to show that the main condition (23) of Theorem 8 is satisfied, namely, that the relation

$$\sum_{k \in V_0^1} E|Z_k E(Z_0 | \mathcal{F}_{V_0^{||k||}})| < \infty \tag{27}$$

holds. We remind that

$$Z_0 = \sum_{i \in \Lambda} \alpha_i X_{-i} \varepsilon_0,$$

(here we used notation $\alpha = (\alpha_i,\ i \in \Lambda)$). Let us denote $\mathcal{F}_0 = \sigma(\varepsilon_i,\ i \neq 0)$. It is easy to see that for any $i \in \Lambda$ random variable X_{-i} is measurable with respect to $\mathcal{F}_0$, therefore $E(Z_0|\mathcal{F}_0) = 0$. Now let us note that $\mathcal{F}_{V_0^{||k||}} \subset \mathcal{F}_0$ for any $k \in V_0^1$, therefore

$$E\Big(Z_0|\mathcal{F}_{V_0^{||k||}}\Big) = E\Big(E(Z_0|\mathcal{F}_0)|\mathcal{F}_{V_0^{||k||}}\Big) = 0$$

and (27) trivially follows. This ends the proof of Theorem 1.

Acknowledgment. The authors express their gratitude to Prof. J. Dedecker whose valuable remarks and suggestions helped to improve the paper.

References

[1] ANSELIN, L., (1988). Spatial Econometrics: Methods and Models. *Kluwer, Academic Publishers*, Dordrecht.

[2] BASU, S. and REINSEL, G. C. (1992). A note on properties of spatial Yule-Walker estimators. *J. Statist. Comput. Simul.,* **41**, 243–255.

[3] BILINGSLEY, P. (1968). Convergence of Probability Measures. *Wiley*, New York.

[4] CLIFF, A. D. and ORD, J. K. (1973). Spatial Autocorrelation. *Pion,* London.

[5] DEDECKER, J. (1998). A central limit theorem for stationary random fields. *Probab. Theory Related Fields*, **110**, 397–426.

[6] FRANCIS, J. M., MEIRI, A. Z. and PORAT, B. (1995). A Wold-like decomposition of two-dimensional discrete homogeneous fields. *Ann. of Appl. Probab.*, **5**, 1, 248–260.

[7] GEORGII, H-O. (1988). Gibbs measures and phase transitions. *de Gruyter*, Berlin, New York.

[8] GIACOMINI, R. and GRANGER, W. J. (2004). Aggregation of space-time processes. *J. of Econometrics*, **118**, 7–26.

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

[10] GUYON, X. (1995). Random Fields on a Network. Modeling, Statistics, and Applications. *Springer*, New York , Berlin.

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

[12] HELSON, H. and LOWDENSLAGER, D. (1962). Prediction theory and Fourier series in several variables II. *Acta Math.*, **106**, 175–213.

[13] PIKHURKO, O. (2001). Lattice points in lattice polytopes. *Mathematika*, **48**, 15-24.

[14] RÉNYI, A. (1963). On stable sequences of events. *Sankhya, Ser A*, **25**, 189–206.

[15] Tjostheim, D. (1978). Statistical spatial series modeling. *Adv. Appl. Prob.*, **10**, 130–154.

[16] Tjostheim, D. (1983). Statistical spatial series modeling II: some further results on unilateral lattice processes. *Adv. Appl. Prob.*, **15**, 562–584.

[17] Tjostheim, D. and Paulsen, J. (1983). Bias of some commonly used time series estimates *Biometrika*, **70**, 389–399.

[18] Whittle, P. (1954). On stationary processes in the plane. *Biometrica*, **41**, 434–449.

Laboratoire Paul Painlevé, UMR 8524,
Statistique et Probabilités,
Université des Sciences et Technologies de Lille,
59655 Villeneuve d'Ascq Cedex, France.
E-mail: `Youri.Davydov@univ-lille1.fr`

Department of Mathematics and Informatics,
Vilnius University
Naugarduko 24, 03225 Vilnius, Lithuania
E-mail: `vygantas.paulauskas@maf.vu.lt`