Metrika (2008) 68:351–363
DOI 10.1007/s00184-007-0167-y

An empirical likelihood method for spatial regression

Daniel J. Nordman

Received: 2 August 2007 / Published online: 4 January 2008

Abstract Properties of a "blockwise"empirical likelihood for spatial regression with non-stochastic regressors are investigated for spatial data on a lattice. The method enables nonparametric confidence regions for spatial trend parameters to be calibrated, even though non-random regressors introduce *non-stationary* forms of spatial dependence into the "blockwise" construction. Additionally, the regression results are valid in a general framework allowing for a variety of behavior in regressor variables as well as the underlying spatial error process. The same regression method also applies when the regressors are stochastic.

Keywords Blocking · Spatial lattice data · Non-stationarity · Non-stochastic regressors

1 Introduction

Estimation of spatial trend is often an important consideration in spatial analysis, as described by Cressie (1993) (i.e., modeling so-called "large-scale variation"). In this paper, we consider a real-valued spatial regression model given by

$$Z(\mathbf{s}) = X(\mathbf{s})'\beta + \varepsilon(\mathbf{s}), \qquad \mathbf{s} \in \mathbb{Z}^d, \tag{1}$$

where $X(\mathbf{s})$ is a $q \times 1$ vector of non-random regressors, $\beta \in \mathbb{R}^q$ is a vector of regression parameters and $\{\varepsilon(\mathbf{s}) : \mathbf{s} \in \mathbb{Z}^d\}$ is a mean-zero strictly stationary random field on the integer grid $\mathbb{Z}^d$, $d \geq 1$. Here d denotes the dimension of sampling so that the regression model (1) applies to time series ($d = 1$) as a special case. Such models

D. J. Nordman (✉)
Department of Statistics, Iowa State University, Ames, IA 50011, USA
e-mail: dnordman@iastate.edu

with non-random regressors are commonly used, for example, when incorporating the locations of observations in space or time as regression information (e.g., $X(\mathbf{s}) \equiv \mathbf{s}$, $\mathbf{s} \in \mathbb{Z}^d$) or when studying treatment effects in field trials (see Sect. 5.7 of Cressie 1993).

This paper considers an empirical likelihood (EL) method for nonparametric inference on the regression parameter β based on a spatial sample on a grid and develops the asymptotic properties of the EL method in a general framework for the regressors $X(\cdot)$ and error process $\varepsilon(\cdot)$ in (1). This EL method uses data blocks in a spatial version of the blockwise EL proposed by Kitamura (1997) for weakly dependent time series. Previously, Owen (1991) introduced EL regression with independent error terms $\varepsilon(\cdot)$, which was further investigated by Chen (1993) and Bravo (2002) among others in the independent data context. For weakly dependent time series, Bravo (2005) examined a blockwise EL for regression problems with stochastic regressors $X(\cdot)$. In contrast, our results involve *non-random* regressors and a spatially dependent error process $\varepsilon(\cdot)$, which introduces non-stationary forms of spatial dependence through the usual EL estimating functions for regression (see Sect. 2.2). However, the spatial EL is shown to be valid for regression models (1) despite the non-stationarity involved as well as conditions permitting the regressors to become unbounded at possibly non-standard rates. The same method also applies to cases where the regressors $X(\cdot)$ in (1) are stochastic. For a large class of spatial processes, the EL approach allows confidence regions for β without knowledge of the underlying spatial dependence structure and without any explicit spatial variance estimation or studentization, as would be required for confidence regions based on the ordinary least squares estimator of β, for example. EL confidence regions for the regression parameter β can be simply calibrated using the asymptotic chi-squared distribution of a log-EL ratio.

We end this section with a brief literature review on EL for dependent data. Section 2 describes the spatial blockwise EL method and Sect. 3 presents the main EL regression results for spatial data on a grid. Section 4 provides a small data example to illustrate the spatial EL regression method. Proofs of the main results appear in Sect. 5.

Many EL developments continue for independent data (cf. Owen 2001), but recent work has focused on extending EL to dependent time series, especially in econometric applications. In a pivotal work, Kitamura (1997) introduced a EL methodology for time series based on data blocking used for bootstrap and subsampling with time series (Carlstein 1986; Künsch 1989; Politis and Romano 1994). This "blockwise EL" has produced fruitful inference in other time series contexts such as blockwise Euclidean EL (Lin and Zhang 2001), time series regression (Bravo 2005), negatively associated series (Zhang 2006) and long-range dependence (Nordman et al. 2006). Additionally, EL and generalized method of moments estimators have been considered for testing moment restrictions in the econometric literature (cf. Hall and Horowitz 1996; Imbens et al. 1998; Chuang and Chan 2002; Bravo 2004). Using the periodogram (i.e., a data-transformation rather than data blocking), Monti (1997) and Nordman and Lahiri (2006) developed a frequency-domain EL for time series. Recently, Nordman and Caragea (2007) investigated a spatial EL method based on data blocks for variogram estimation, which partly considered de-trending spatial data with a regression model in connection to variogram fitting. The purpose of this article is a full development of EL spatial regression within a broad framework that permits the non-stochastic regressors to exhibit standard and non-standard behaviors. This requires developing an asymptotic theory for spatial EL regression in which maximal spatial regressor values may grow unbounded in non-standard ways.

2 Spatial EL method for regression

2.1 Sampling design

Suppose $\mathcal{R}_n \subset \mathbb{R}^d$, $d \geq 1$, represents a spatial sampling region inside of which we observe the process $Z(\cdot)$, under model (1), at n locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\} \subset \mathbb{Z}^d$. That is, the sampling sites are $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\} = \mathcal{R}_n \cap \mathbb{Z}^d$ and the available (paired) data for inference are $\{Z(\mathbf{s}_i), X(\mathbf{s}_i)\}_{i=1}^n$. We adopt a sampling framework as in Lahiri (2003a, Chapter 12.2) and Nordman and Lahiri (2004) which permits $\mathcal{R}_n$ to expand as the sample size increases (i.e., "increasing domain asymptotics" as termed by Cressie 1993). Let $\mathcal{R}_0 \subset (-1/2, 1/2]^d$ be a Borel subset containing a neighborhood of the origin such that, for any positive sequence $a_n \to 0$, the number of lattice cubes $a_n(\mathbf{i} + [0, 1]^d) \subset \mathbb{R}^d, \mathbf{i} \in \mathbb{Z}^d$ intersecting both closures $\overline{\mathcal{R}_0}$ and $\overline{\mathbb{R}^d \setminus \mathcal{R}_0}$ is $O(a_n^{-d+1})$ as $n \to \infty$. The sampling region $\mathcal{R}_n = \lambda_n \mathcal{R}_0$ is then obtained by inflating the template set $\mathcal{R}_0$ by an increasing sequence of positive scaling factors λ_n, where $\lambda_n \to \infty$ as $n \to \infty$. Because $\mathcal{R}_0$ contains the origin, $\mathcal{R}_n$ has a fixed center and expands in volume, maintaining the same shape, as the sample size n increases. The boundary condition on $\mathcal{R}_0$ entails that, as $n \to \infty$, the total sample size n is larger than the number $O\left(\lambda_n^{d-1}\right)$ of samples near the boundary of $\mathcal{R}_n$; consequently, spatial volume and sample size are equivalent in large samples because $n/\mathrm{vol}(\mathcal{R}_n) \to 1$, where $\mathrm{vol}(\mathcal{R}_n) = \lambda_n^d \mathrm{vol}(\mathcal{R}_0)$ and $\mathrm{vol}(\cdot)$ denotes volume. This assumption is satisfied for most practical regions of interest (e.g., convex or star-shaped sets) and is used only to avoid pathological sampling regions. Note that we may obtain a size n time series stretch (i.e., $d = 1$) by choosing $\mathcal{R}_0 = (-1/2, 1/2]$ and $\lambda_n = n$ to define $\mathcal{R}_n$ as an interval with n integer points (i.e., observation locations), so the results to follow include time series as a special case.

2.2 Spatial blockwise EL ratio for regression

For EL regression, we use a standard EL moment condition from Owen (1991) that links the trend parameter β to the data. Define an estimating function of $\beta \in \mathbb{R}^q$ as $Y_\beta(\mathbf{s}) = X(\mathbf{s})\{Z(\mathbf{s}) - X(\mathbf{s})'\beta\}$, $\mathbf{s} \in \mathbb{Z}^d$, which satisfies

$$\mathrm{E}Y_{\beta_0}(\mathbf{s}) = 0_q \in \mathbb{R}^q, \quad \mathbf{s} \in \mathbb{Z}^d, \tag{2}$$

at the true parameter value β_0 in (1). The construction of an EL function for $\beta \in \mathbb{R}^q$ involves (2) and blocks of observations $Y_\beta(\cdot)$, where the blocking mechanism helps to capture the unknown spatial dependence structure by keeping neighboring spatial observations together. Similar blocking notions have been key to other nonparametric likelihoods, like the bootstrap and subsampling, to dependent data (cf. Lee and Lahiri 2002; Lahiri 2003a). However, in contrast to previous blockwise EL formulations with time series (cf. Kitamura 1997; Bravo 2005), the spatial process $Y_{\beta_0}(\cdot) = X(\cdot)\varepsilon(\cdot)$ at the true β_0 may exhibit non-stationary forms of dependence when the regressors $X(\cdot)$ are non-stochastic and non-constant.

To define the spatial blocks, let b_n be a sequence of positive integers for block scaling and let $\mathcal{I}_n = \{\mathbf{i} \in \mathbb{Z}^d : \mathcal{B}_n(\mathbf{i}) \subset \mathcal{R}_n\}$ denote the index set of all d-dimensional rectangles $\mathcal{B}_n(\mathbf{i}) \equiv \mathbf{i} + b_n(-1/2, 1/2]^d$, $\mathbf{i} \in \mathbb{Z}^d$, lying within $\mathcal{R}_n$. This provides a collection of maximally overlapping spatial EL blocks as $\{\mathcal{B}_n(\mathbf{i}) : \mathbf{i} \in \mathcal{I}_n\}$. For example, in the time series case $d = 1$, this blocking scheme would produce $n - b_n + 1$ overlapping blocks of length b_n in a time series of length n, i.e., blocks given by $\{(Z_i, \ldots, Z_{i+b_n+1}) : i = 1, \ldots, n - b_n + 1\}$ if $Z_1, \ldots, Z_n$ denotes the time series (using a notational deviation for simplicity, where a time series would be $\{Z(\mathbf{s}) : \mathbf{s} \in \mathcal{R}_n \cap \mathbb{Z}\}$, $\mathcal{R}_n = n(-1/2, 1/2]$ expressed in our notation). Other EL blocking schemes are possible, such as non-overlapping blocks indexed by $\{\mathbf{i} \in \mathbb{Z}^d : \mathcal{B}_n(b_n\mathbf{i}) \subset \mathcal{R}_n\}$, but overlapping produces more data blocks which reduces variability in estimation with block resampling methods (cf. Nordman and Lahiri 2004).

We require that block scaling $b_n \to \infty$ as $n \to \infty$, but with $b_n/\lambda_n \to 0$ to ensure the blocks are small relative to the sampling region $\mathcal{R}_n$; Sect. 3.1 to follow describes block conditions in more detail. Each block $\mathcal{B}_n(\mathbf{i})$, $\mathbf{i} \in \mathcal{I}_n$, in the collection contains $\left|\mathcal{B}_n(\mathbf{i}) \cap \mathbb{Z}^d\right| = b_n^d$ observations of the process $Y_\beta(\cdot)$, $\beta \in \mathbb{R}^q$, with a corresponding block sample mean $\bar{Y}_{\beta,\mathbf{i}} = \sum_{\mathbf{s} \in \mathcal{B}_n(\mathbf{i}) \cap \mathbb{Z}^d} Y_\beta(\mathbf{s})/b_n^d$, $\mathbf{i} \in \mathcal{I}_n$ (letting $|A|$ denote the size of a finite set A).

For inference on the regression parameter, we assess the plausibility of a β-value through the profile EL ratio given by

$$R_n(\beta) = N_{\mathcal{I}}^{N_{\mathcal{I}}} \cdot \sup \left\{ \prod_{\mathbf{i} \in \mathcal{I}_n} p_{\mathbf{i}} \; : \; p_{\mathbf{i}} \geq 0, \; \sum_{\mathbf{i} \in \mathcal{I}_n} p_{\mathbf{i}} = 1, \; \sum_{\mathbf{i} \in \mathcal{I}_n} p_{\mathbf{i}} \bar{Y}_{\beta,\mathbf{i}} = 0_q \right\}, \quad \beta \in \mathbb{R}^q, \tag{3}$$

where $N_{\mathcal{I}} = |\mathcal{I}_n|$ denotes the number of overlapping blocks available. Note that 1 is the largest possible value of $R_n(\beta)$, occurring when each $p_{\mathbf{i}} = 1/N_{\mathcal{I}}$. If the zero vector $0_q \in \mathbb{R}^q$ is interior to the convex hull of $\{\bar{Y}_{\beta,\mathbf{i}} : \mathbf{i} \in \mathcal{I}_n\}$ for a given $\beta \in \mathbb{R}^q$, then (3) achieves a positive maximum and may be written as

$$R_n(\beta) = \prod_{\mathbf{i} \in \mathcal{I}_n} p_{\beta,\mathbf{i}} N_{\mathcal{I}}, \qquad p_{\beta,\mathbf{i}} = \left\{N_{\mathcal{I}}(1 + t_\beta' \bar{Y}_{\beta,\mathbf{i}})\right\}^{-1} \in (0, 1), \tag{4}$$

where $t_\beta \in \mathbb{R}^q$ satisfies $0_q = \sum_{\mathbf{i} \in \mathcal{I}_n} \bar{Y}_{\beta,\mathbf{i}}/(1 + t_\beta' \bar{Y}_{\beta,\mathbf{i}})$; see Owen (1991) and Qin and Lawless (1994) for these and further computational details with EL with independent data, which are applicable here.

3 Main distributional results

3.1 Assumptions with non-stochastic regressors

To describe the main results on the EL regression method, we require some assumptions on the process (1). For a vector $\mathbf{x} = (x_1, \ldots, x_d)' \in \mathbb{R}^d$, let $\|\mathbf{x}\|$ and $\|\mathbf{x}\|_\infty =$

$\max_{1\le i\le d}|x_i|$ denote the Euclidean and l^∞ norms of $\mathbf{x}$, respectively, and define the distance between two sets $E_1,\ E_2\subset\mathbb{R}^d$ as: $\mathrm{dis}(E_1,\ E_2)=\inf\{\|\mathbf{x}-\mathbf{y}\|_\infty:\mathbf{x}\in E_1,\mathbf{y}\in E_2\}$. Let $\mathcal{F}(T)$ denote the σ-field generated by the random variables $\{\varepsilon(\mathbf{s}):\mathbf{s}\in T\}$, $T\subset\mathbb{Z}^d$. We define the strong mixing coefficient for the strictly stationary random field $\varepsilon(\cdot)$ as a function of set distance $v>0$ and size $w>0$,

$$\alpha(v,w)=\sup\left\{\tilde{\alpha}(T_1,T_2)\ :\ T_i\subset\mathbb{Z}^d,\ |T_i|\le w,\ i=1,2;\ \mathrm{dis}(T_1,T_2)\ge v\right\},\qquad(5)$$

where $\tilde{\alpha}(T_1,T_2)=\sup\{|P(A\cap B)-P(A)P(B)|\ :\ A\in\mathcal{F}(T_1),B\in\mathcal{F}(T_2)\}$ for $T_1,\ T_2\subset\mathbb{Z}^d$. In spatial settings $d\ge 2$, defining the mixing coefficient (5) with bounded sets T_1,T_2 is important to avoid more restrictive forms of mixing (Bradley 1989). We shall assume that $\alpha(v,w)\le\alpha_1(v)g(w)$, $v,w>0$ holds in Assumption A.1 (to follow) for some non-increasing $\alpha_1(\cdot)$ and non-decreasing $g(\cdot)$, where both functions are non-negative. Let β_0 denote the unique parameter value satisfying (2) and define a positive definite matrix $A_n=\sum_{i=1}^n X(\mathbf{s}_i)X(\mathbf{s}_i)'$ with $M_n=\max_{1\le i\le n}\|A_n^{-1/2}X(\mathbf{s}_i)\|$. In the following, recall d denotes the dimension of sampling.

Assumptions

A.1 For some $\delta>0$, $\mathrm{E}|\varepsilon(\mathbf{s})|^{6+\delta}<\infty$ holds; there exists $\tau>5d(6+\delta)/\delta$ such that $\alpha_1(v)\le Cv^{-\tau}$, $v\ge 1$; and $M_n^2 n=o\big(\lambda_n^{\frac{\tau-d}{4\tau}}\big)$ as $n\to\infty$. Additionally, $g(w)=o(w^{\frac{\tau-d}{4d}})$ as $w\to\infty$ when $d\ge 2$ and, for the time series case $d=1$, $g(\cdot)$ is bounded.

A.2 $\lim_{n\to\infty}\mathrm{Var}\{\sum_{i=1}^n A_n^{-1/2}X(\mathbf{s}_i)\varepsilon(\mathbf{s}_i)\}\equiv\Sigma_{\beta_0}$ is positive definite.

A.3 $b_n^{-1}+b_n^{2d}/n=o(1)$ as $n\to\infty$.

A.4 $P\{R_n(\beta_0)>0\}\to 1$ as $n\to\infty$.

Mixing conditions in A.1 are mild and embody weak spatial dependence, where bounds on (5) entail that the spatial dependence decreases as a function of distance between observations but may increase as a function of set sizes. These resemble assumptions from Lahiri (2003a, Theorem 12.6) or Lee and Lahiri (2002, condition 7) for spatial resampling methods. The growth rate on $M_n^2 n$ in assumption A.1 permits a central limit theorem for regression sums like $\sum_{i=1}^n A_n^{-1/2}X(\mathbf{s}_i)\varepsilon(\mathbf{s}_i)$ needed in the EL theoretical development; in particular, these orders of $g(\cdot)$ and $M_n^2 n$ are described by Lahiri (2003b) (Proposition 4.2.ii) for limit theorems under spatial regression. Of note, this regression framework for EL allows standard regressors for which $M_n^2 n$ is typically bounded (e.g., constant regressors $X(\cdot)=1$ for defining a real-valued mean $\beta\in\mathbb{R}$ in (1)), but the sequence $M_n^2 n$ may be *unbounded* to allow for a variety of behavior in the non-random regressors $X(\cdot)$. In A.2, Σ_{β_0} represents the limiting covariance matrix of $A_n^{1/2}(\hat{\beta}_{n,OLS}-\beta_0)$ corresponding to a normalized OLS estimator $\hat{\beta}_{n,OLS}=A_n^{-1}\sum_{i=1}^n X(\mathbf{s}_i)Z(\mathbf{s}_i)$ of β. The block condition in A.3 stipulates that the squared number $(b_n^d)^2$ of observations in a spatial block should be smaller order than the spatial sample size n, similar to the EL block size condition of Kitamura (1997) for time series $d = 1$. The probability statement in A.4 implies that the EL ratio can be positively computed at β_0, which is required for taking the log of $R_n(\beta_0)$ in the main EL result of Sect. 3.2. This is a mild condition which is typical in EL regression contexts (cf. Condition 3.3a of Owen (1991) for independent data). For some special regressors, such as constant regressors $X(\cdot) = 1$ mentioned above, the probability condition in A.4 can also be shown to follow from the remaining assumptions A.1–A.3.

3.2 Spatial EL result with non-stochastic regressors

The main result of the paper is a nonparametric recasting of Wilks' theorem for the spatial EL with non-random regressors, which establishes the asymptotic chi-squared distribution of the log-EL ratio, $-2\log R_n(\beta_0)$, at the true parameter $\beta_0 \in \mathbb{R}^q$. Despite the general non-stationarity of the estimating functions (2) with non-random regressors $X(\cdot)$ as well as various growth behaviors allowed in the regressors, the EL method automatically performs studentization within its mechanics so that no spatial variance estimation is required and confidence regions for β may be simply calibrated with the log-EL ratio. For $\eta > 0$, define a confidence set for the regression parameter as $C_n(\eta) \equiv \{\beta \in \mathbb{R}^q : -2\log R_n(\beta)/b_n^d \le \eta\}$.

Theorem 1 *Let $\eta > 0$ and $t \in \mathbb{R}^q$. Under* (2) *and Assumptions* A.1–A.4*, as $n \to \infty$*

(i) $-2\log R_n(\beta_0)/b_n^d \xrightarrow{d} \chi_q^2$, *chi-squared with q degrees of freedom.*
(ii) *$C_n(\eta)$ is a connected set in $\mathbb{R}^q$, without voids.*
(iii) $-2\log R_n(\beta_0 + A_n^{-1/2}\Sigma_{\beta_0}^{1/2}t)/b_n^d \xrightarrow{d} \chi_q^2(\|t\|^2)$ *with non-centrality parameter* $\|t\|^2$.

The scalar b_n^{-d} in Theorem 1(i) adjusts the log-EL ratio for the spatial blocking used; recall b_n^d represents the sample size or volume of a spatial block. For time series $d = 1$, this adjustment corresponds to the reciprocal of a block length b_n and asymptotically matches the blockwise EL correction of Kitamura (1997, pp. 2087) with overlapping blocks. (Rephrased in terms of Kitamura's notation, $Q \equiv n - b_n + 1$ is the number of maximally overlapping blocks of length $M \equiv b_n$ in a time series of length $N \equiv n$ (see Sect. 2.2 here) so that Kitamura's block correction $(MQ)^{-1}N$ is equivalent to $M^{-1} \equiv b_n^{-1}$ for such overlapping blocks because $N/Q \to 1$.)

For a wide range of spatial processes and regressors under Theorem 1, an approximate $100(1-\alpha)\%$ EL confidence region for β can be calibrated as $C_n(\chi^2_{q;1-\alpha})$ based on a lower chi-squared quantile $\eta = \chi^2_{q;1-\alpha}$. The size of the confidence set $C_n(\chi^2_{q;1-\alpha})$, as measured by $\sup\{\|\beta - \beta_0\| : \beta \in C_n(\chi^2_{q;1-\alpha})\}$, is determined by the growth of the largest eigenvalue of $A_n^{-1/2}$ (for example, $A_n^{-1/2} = n^{-1/2}$ when considering inference on a stationary process mean $\mathrm{E}Z(\mathbf{s}) = \beta \in \mathbb{R}$ in (1) corresponding to regressors $X(\mathbf{s}) = 1$). Theorem 1(ii) indicates some geometrical properties for confidence regions produced by the spatial EL method. For independent data, Hall and La Scala (1990) have shown that EL confidence regions are convex for population means and must be connected for parameters that are smooth functions of means.

Theorem 1 indicates that EL regions for regression parameters must be connected sets as well.

We next provide some comments on block scaling b_n and selection in the EL method. Under Theorem 1, potential choices for the EL block factor may involve $b_n = Cn^{\theta}$ or $C[\text{vol}(\mathcal{R}_n)]^{\theta}$ for some $C > 0$ and $0 < \theta < 1/(2d)$. However, optimal block sizes for EL coverage accuracy are presently unknown even for time series blockwise EL methods, making the best blocks in the spatial setting difficult to determine here. To suggest choices of block scaling b_n in practice, one pragmatic approach may be based on the so-called "minimum volatility" method, described by Politis et al. (1999, Sect. 9.3) for subsampling time series. The idea is that, while some scaling b_n may produce over- or under-coverage in EL confidence regions, we might expect approximately correct EL inference over a range of b_n-values where the method appears stable and produces confidence regions that change very little volume as a function of the block scaling. Hence, we may compute EL confidence regions over a series of b_n and, by visual inspection, choose a block size where the EL confidence regions appear relatively stable. This minimum volatility method is illustrated in more detail with a data example involving spatial EL regression in Sect. 4.

3.3 Extension to stochastic regressors

We remark that the same EL construction from Sect. 2.2 also provides valid inference on β with random regressors $X(\cdot)$ in the spatial regression model (1). Namely, suppose that the random field given by $\{U(\mathbf{s}) = [X(\mathbf{s}), \varepsilon(\mathbf{s})] : \mathbf{s} \in \mathbb{Z}^d\}$ is strictly stationary and the orthogonality condition (2) holds. Then the following Theorem 2 shows that the spatial EL method with non-random regressors applies equally to the stochastic regressor case with spatial data. To state the result, we define slightly modified assumptions A.1$'$ and A.2$'$ by replacing $\varepsilon(\cdot)$ with $U(\cdot)$ in the mixing/moment conditions, by dropping the condition on $M_n^2 n$ in A.1, and by re-defining A_n as the sample size n in A.2 (i.e., setting $A_n \equiv n$).

Theorem 2 *Suppose* (2) *and Assumptions* A.1$'$, A.2$'$, A.3 *hold. Then, the distributional results in Theorem* 1 (i) *and* (iii) *remain valid, setting* $A_n^{-1/2} \equiv n^{-1/2}$ *in* (iii). *In addition, if* $\text{Var}[X(\mathbf{s})]$ *is positive definite, then Theorem* 1 (ii) *holds as well.*

Theorem 2 provides an extension of some blockwise EL regression results in Bravo (2005), based on random regressors with weakly dependent time series ($d = 1$), to the spatial setting.

4 Data example

Here we describe the spatial EL method in a regression example involving simulated data. On a circular sampling region $\mathcal{R}_n = \{\mathbf{s} \in \mathbb{R}^2 : \|\mathbf{s}\| \leq 18\}$ in the plane, a data realization $\{\varepsilon(\mathbf{s}) : \mathbf{s} \in \mathbb{Z}^2 \cap \mathcal{R}_n\}$ was generated from a real-valued mean-zero Gaussian field with covariance structure $\text{Cov}\{\varepsilon(\mathbf{s}), \varepsilon(\mathbf{s}+\mathbf{h})\} = 3\exp(-\|\mathbf{h}\|^2)$, $\mathbf{h} \in \mathbb{Z}^2$, using the circulant embedding method of Chan and Wood (1997); this region then contains 1009

Table 1 90% EL confidence intervals for $\beta = 2$ based on one spatial data realization with regressors $X(\cdot)$, for various spatial block scaling b_n; intervals include the lengths (Len) and an endpoint standard deviation $S(b_n)$ criterion

b_n	Bounded Regressors $X(\mathbf{s}) = 1$ One realization				Unbounded Regressors $X(\mathbf{s}) = \|\mathbf{s}\|$ One realization			
	90% interval	Len	$S(b_n)$	CP (%)	90% interval	Len	$S(b_n)$	CP (%)
1	(1.837, 2.013)	0.176	2.608	65.8	(1.977, 1.992)	0.015	0.281	67.6
2	(1.798, 2.047)	0.249	2.549	81.8	(1.971, 1.993)	0.022	0.305	82.5
3	(1.785, 2.062)	0.277	2.479	85.7	(1.967, 1.993)	0.026	0.319	87.3
4	(1.778, 2.068)	0.290	1.171	87.9	(1.965, 1.994)	0.029	0.220	89.6
5	(1.770, 2.068)	0.298	0.719	89.6	(1.963, 1.994)	0.031	0.229	91.1
6	(1.761, 2.069)	0.308	0.568	90.9	(1.961, 1.994)	0.033	0.272	92.2
7	(1.763, 2.074)	0.311	0.461	91.9	(1.958, 1.996)	0.038	0.331	93.6
8	(1.764, 2.078)	0.314	0.968	92.3	(1.956, 1.998)	0.042	0.391	94.4
9	(1.769, 2.080)	0.311	0.995	92.3	(1.954, 2.001)	0.047	0.325	94.9
10	(1.736, 2.086)	0.350	1.097	93.7	(1.952, 2.005)	0.053	0.276	96.5

Coverage probabilities of 90% EL intervals (CP) also appear for each b_n based 1000 simulations

$= |\mathbb{Z}^2 \cap \mathcal{R}_n|$ sampling sites. Spatial regression observations $Z(\mathbf{s}) = \beta X(\mathbf{s}) + \varepsilon(\mathbf{s})$, $\mathbf{s} \in \mathbb{Z}^2 \cap \mathcal{R}_n$ were formed by taking $\beta = 2 \in \mathbb{R}$ in (1) with either bounded $X(\mathbf{s}) = 1$ or unbounded $X(\mathbf{s}) = \|\mathbf{s}\|$ regressors. EL inference on the regression parameter was considered by applying EL log-ratio in Theorem 1(i) to calibrate confidence intervals. For one spatial realization with each regressor type, Table 1 displays the resulting approximate 90% EL confidence intervals for $\beta = 2$, over a variety of block factors b_n.

We may illustrate the minimum volatility method described in Sect. 3.2 for picking block scaling b_n with this data example. By visual inspection in Table 1, interval lengths appear to be fairly stable for $b_n \in \{6, 7, 8\}$ with bounded regressors, or for $b_n \in \{4, 5, 6\}$ for unbounded regressors, so that the "minimum volatility" principle would suggest these block choices as appropriate for the EL method according to this data realization. For a quantitative measure of volatility, we may compute and minimize a standard deviation criterion using upper and lower endpoints of EL intervals, say (L_{b_n}, U_{b_n}), based on block scaling b_n (see Politis et al. 1999, Sect. 9.3). That is, for each integer $b_n \geq 1$, we compute $S(b_n) \equiv [S_L(b_n) + S_U(b_n)]/2$, where $S_L(b_n)$ is the standard deviation of lower endpoints $\{L_i : |i - b_n| \leq k\}$ for some k (e.g., $k = 1, 2$) defining intervals with scaling in a neighborhood of b_n and $S_U(b_n)$ is analogously defined with upper endpoints. We then select the block factor b_n with a minimal $S(b_n)$ value. For this data example, applying this criterion in Table 1 with $k = 2$ indicates block choices $b_n = 7$ for bounded regressors and $b_n = 4$ for unbounded regressors, which supports blocks indicated by visual inspection of interval lengths. For each factor b_n, Table 1 also gives coverage probabilities for 90% EL confidence intervals based on 1000 data simulations. The coverage probabilities indicate that the block choices suggested by "minimum volatility" in this data example are reasonable.

We comment additionally that, by Theorem 1(iii), the length of an EL interval is influenced here by $A_n^{-1/2} = (\sum_{\mathbf{s}\in\mathcal{R}_n\cap\mathbb{Z}^2} X(\mathbf{s})^2)^{-1/2}$. In Table 1, this explains the shorter intervals for the example with unbounded regressors $X(\mathbf{s}) = \|\mathbf{s}\|$ compared to the case $X(\mathbf{s}) = 1$ that corresponds to inference on the spatial mean $\mathrm{E}Z(\mathbf{s}) = 2$. Note also that coverage probabilities are particulary poor for the unblocked ($b_n = 1$) EL version with spatial data, which builds the EL function from individual observations and so fails to capture the data dependence without the blocking mechanism.

5 Appendix: Proofs of main results

Under the regression model (1), it holds for any $T \subset \mathbb{Z}^d$ and non-stochastic regressors $X(\mathbf{s}), \mathbf{s} \in T$ that

$$\left\|\sum_{\mathbf{s}\in T} X(\mathbf{s})\varepsilon(\mathbf{s})\right\|^6 \le C|T|^3 \sup_{\mathbf{s}\in T} \|X(\mathbf{s})\|^6 \qquad (6)$$

by A.1 and moment bounds in Doukhan (1994, Theorem 1, Sect. 1.4.1). For bounded $T_1, T_2 \subset \mathbb{Z}^d$ with $\mathrm{dis}(T_1, T_2) > 0$, if a random variable Y_i is measurable with respect to $\mathcal{F}(T_i)$, $i = 1, 2$, then A.1 and mixing bounds in Doukhan (1994, Sect. 1.2.2) yield

$$|\mathrm{Cov}(Y_1, Y_2)| \le 8\left\{\mathrm{E}|Y_1|^{\frac{6+\delta}{3}}\mathrm{E}|Y_2|^{\frac{6+\delta}{3}}\right\}^{\frac{3}{6+\delta}} \alpha\left(\mathrm{dis}(T_1, T_2); \max_{i=1,2}|T_i|\right)^{\frac{\delta}{6+\delta}}. \qquad (7)$$

Limits in order symbols are taken as $n \to \infty$ and, for two positive sequences, we write $s_n \sim t_n$ if $s_n/t_n \to 1$. We let C denote a generic constant which does not depend on n or any $\mathbb{Z}^d$ lattice points and we use $\mathbf{0}$ for the $\mathbb{R}^d$-zero vector. Lemma 1 provides tools for proving the spatial EL result in Theorem 1 with non-stochastic regressors. Recall the data under (1) are located at spatial sites $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\} = \mathcal{R}_n \cap \mathbb{Z}^d \subset \mathbb{R}^d$, $d \ge 1$ while $\mathcal{I}_n$ denotes the index set of EL blocks and $\bar{Y}_{\beta_0,\mathbf{i}}$, $\mathbf{i} \in \mathcal{I}_n$ denotes a block sample mean satisfying $\mathrm{E}\bar{Y}_{\beta_0,\mathbf{i}} = 0_q$ at the true regression parameter β_0 by (2).

Lemma 1 *Let* $W_\mathbf{i} = A_n^{-1/2}\bar{Y}_{\beta_0,\mathbf{i}}$, $\mathbf{i} \in \mathcal{I}_n$. *Under* (2) *and Assumptions* A.1–A.4,
(*i*) $n = |\mathcal{R}_n \cap \mathbb{Z}^d|$ *and* $N_\mathcal{I} = |\mathcal{I}_n| \sim \mathrm{vol}(\mathcal{R}_n) = \lambda_n^d \mathrm{vol}(\mathcal{R}_0)$, *where* $\mathrm{vol}(\cdot)$ *denotes volume*; (ii) $W_n \equiv \sum_{\mathbf{i}\in\mathcal{I}_n} W_\mathbf{i} \xrightarrow{d} \mathcal{N}(0_q, \Sigma_{\beta_0})$, *a normal limit law*; (iii) $\max_{\mathbf{i}\in\mathcal{I}_n} b_n^d\|W_\mathbf{i}\| \xrightarrow{p} 0$; (iv) $\widehat{\Sigma}_n \equiv b_n^d \sum_{\mathbf{i}\in\mathcal{I}_n} W_\mathbf{i} W_\mathbf{i}' \xrightarrow{p} \Sigma_{\beta_0}$; *(v)* $Q_n \equiv \sum_{\mathbf{i}\in\mathcal{I}_n} Q_\mathbf{i}$ *is positive definite for large n, where* $Q_\mathbf{i} = \sum_{\mathbf{s}\in\mathcal{B}_n(\mathbf{i})\cap\mathbb{Z}^d} X(\mathbf{s})X(\mathbf{s})'/b_n^d$.

Proof Part (i) follows from the $\mathcal{R}_0$-boundary condition in Sect. 2.1; see Nordman and Lahiri (2004). For part (ii), note $S_n \equiv \sum_{j=1}^n A_n^{-1/2}X(\mathbf{s}_j)\varepsilon(\mathbf{s}_j) \xrightarrow{d} \mathcal{N}(0_q, \Sigma_{\beta_0})$ under A.1–A.2 by Lahiri (2003b, Theorem 4.3.ii) and we may write $S_n - W_n = \sum_{j=1}^n\{1 - b_n^{-d}w(\mathbf{s}_j)\}A_n^{-1/2}X(\mathbf{s}_j)\varepsilon(\mathbf{s}_j)$ where $w(\mathbf{s}_j) \equiv |\{\mathbf{i} \in \mathcal{I}_n : \mathbf{s}_j \in \mathcal{B}_n(\mathbf{i})\}|$ is the number of blocks containing site $\mathbf{s}_j$, $1 \le j \le n$. For a special set $L_n = \{\mathbf{s}_j : 1 \le j \le n, (\mathbf{s}_j + b_n(-2, 2]^d) \subset \mathcal{R}_n\}$ of sites in $\mathcal{R}_n$, note $w(\mathbf{s}_j) = b_n^d$ holds for $\mathbf{s}_j \in L_n$ implying $|\{j : w(\mathbf{s}_j) \ne b_n^d\}| \le |\{j : \mathbf{s}_j \notin L_n\}|$, where by the $\mathcal{R}_0$-boundary condition

$$|\{j : \mathbf{s}_j \notin L_n\}| \le |\{j : (\mathbf{s}_j + b_n(-2, 2]^d) \cap \overline{\mathbb{R}^d \setminus \mathcal{R}_n} \ne \emptyset\}| \le Cb_n\lambda_n^{d-1}. \qquad (8)$$

From (6), (8), A.1 and A.3, we have $\mathrm{E}\|S_n - W_n\|^6 \le CM_n^6 |\{j : w(\mathbf{s}_j) \ne b_n^d\}|^3 = o(1)$. Hence, $S_n - W_n = o_p(1)$ and part (ii) follows by Slutsky's theorem.

For part (iii), note $\mathrm{E}(\max_{\mathbf{i}\in\mathcal{I}_n} \|W_{\mathbf{i}}\|) \le (\sum_{\mathbf{i}\in\mathcal{I}_n} \mathrm{E}\|W_{\mathbf{i}}\|^6)^{1/6} \le CN_{\mathcal{I}}^{1/6} b_n^{-d/2} M_n = o(b_n^{-d})$ by (6), Lemma 1(i) and $(b_n^2/\lambda_n)^{d/4}$, $\lambda_n^{d/4+d/6} M_n = o(1)$.

Part (v) follows from $\|A_n^{-1/2} Q_n A_n^{-1/2} - I_{q\times q}\| \le M_n^2 |\{j : w(\mathbf{s}_j) \ne b_n^d\}| = o(1)$ by (8) and A.3, letting $I_{q\times q}$ denote the identity matrix.

We establish part (iv) by showing $\|\mathrm{E}\widehat{\Sigma}_n - \Sigma_{\beta_0}\| = o(1)$ first and then $\mathrm{Var}(\widehat{\Sigma}_n) = o(1)$. Under A.2, $\|\Sigma_{\beta_0} - \mathrm{Var}(S_n)\| = o(1)$ holds, where $\mathrm{Var}(S_n) = \sum_{\mathbf{h}\in\mathbb{Z}^d} J_n(\mathbf{h})\sigma(\mathbf{h})$, letting $\sigma(\mathbf{h}) = \mathrm{Cov}\{\varepsilon(\mathbf{0}), \varepsilon(\mathbf{h})\}$ and

$$J_n(\mathbf{h}) = \sum_{1\le j\le n, \mathbf{s}_j \in \mathcal{R}_n \cap (\mathcal{R}_n - \mathbf{h})} A_n^{-1/2} X(\mathbf{s}_j) X(\mathbf{s}_j + \mathbf{h})' A_n^{-1/2}.$$

Using $\sup_{n\ge 1, \mathbf{h}\in\mathbb{Z}^d} \|J_n(\mathbf{h})\| \le C$ by the Cauchy-Schwarz inequality along with

$$\mathbf{h} \ne \mathbf{0} \in \mathbb{Z}^d, \quad |\sigma(\mathbf{h})| \le C\|\mathbf{h}\|_\infty^{-\frac{\tau\delta}{6+\delta}}, \quad |\{\mathbf{h} \in \mathbb{Z}^d : \|\mathbf{h}\|_\infty = m\}| \le Cm^{d-1}, \quad m \ge 1, \tag{9}$$

by (7), we have

$$\begin{aligned}\left\|\mathrm{Var}(S_n) - \sum_{\mathbf{h}\in\mathbb{Z}^d, \|\mathbf{h}\|_\infty \le b_n} J_n(\mathbf{h})\sigma(\mathbf{h})\right\| &\le \sum_{\mathbf{h}\in\mathbb{Z}^d, \|\mathbf{h}\|_\infty > b_n} \|J_n(\mathbf{h})\| |\sigma(\mathbf{h})| \\ &\le \sum_{m=b_n+1}^{\infty} Cm^{d-1-\frac{\tau\delta}{6+\delta}} = o(1).\end{aligned}$$

Setting $\tilde{J}_n(\mathbf{h}) = \sum_{\mathbf{s}_j \in L_n} A_n^{-1/2} X(\mathbf{s}_j) X(\mathbf{s}_j + \mathbf{h})' A_n^{-1/2}$, $\mathbf{h} \in \mathbb{Z}^d$, similar arguments yield

$$\left\|\sum_{\mathbf{h}\in\mathbb{Z}^d, \|\mathbf{h}\|_\infty \le b_n} \{\tilde{J}_n(\mathbf{h}) - J_n(\mathbf{h})\}\sigma(\mathbf{h})\right\| = o(1) \tag{10}$$

since $\|\tilde{J}_n(\mathbf{h}) - J_n(\mathbf{h})\| \le CM_n^2 b_n \lambda_n^{d-1} = o(1)$ for $\|\mathbf{h}\|_\infty \le b_n$ by (8), A.1, A.3.

Toward establishing $\|\mathrm{E}\widehat{\Sigma}_n - \Sigma_{\beta_0}\| = o(1)$, the arguments in the previous paragraph give $\|\Sigma_{\beta_0} - \sum_{\|\mathbf{h}\|_\infty \le b_n} \tilde{J}_n(\mathbf{h})\sigma(\mathbf{h})\| = o(1)$ and we aim now to show $\|\mathrm{E}\widehat{\Sigma}_n - \sum_{\|\mathbf{h}\|_\infty \le b_n} \tilde{J}_n(\mathbf{h})\sigma(\mathbf{h})\| = o(1)$. For $\mathbf{h} = (h_1, \ldots, h_d)' \in \mathbb{Z}^d$, define $K_n(\mathbf{h})$, $\tilde{K}_n(\mathbf{h})$ analogous to $J_n(\mathbf{h})$, $\tilde{J}_n(\mathbf{h})$ but with a summand $b_n^{-d} w(\mathbf{s}_j, \mathbf{h}) A_n^{-1/2} X(\mathbf{s}_j) X(\mathbf{s}_j + \mathbf{h})' A_n^{-1/2}$, where $w(\mathbf{s}_j, \mathbf{h}) \equiv |\{\mathbf{i} \in \mathcal{I}_n : \mathbf{s}_j \in \mathcal{B}_n(\mathbf{i}) \cap [\mathcal{B}_n(\mathbf{i}) - \mathbf{h}]\}| \le b_n^d$ is the number of blocks containing both $\mathbf{s}_j, \mathbf{s}_j + \mathbf{h}$. Then, $\mathrm{E}\widehat{\Sigma}_n = \sum_{\|\mathbf{h}\|_\infty \le b_n} K_n(\mathbf{h})\sigma(\mathbf{h})$ and so $\|\mathrm{E}\widehat{\Sigma}_n - \sum_{\|\mathbf{h}\|_\infty \le b_n} \tilde{K}_n(\mathbf{h})\sigma(\mathbf{h})\| = o(1)$ follows as in (10). For $\mathbf{s}_j \in L_n$ and $\|\mathbf{h}\|_\infty \le b_n$, it holds that $w(\mathbf{s}_j, \mathbf{h}) = \prod_{j=1}^d (b_n - |h_j|)$ so that $\tilde{K}_n(\mathbf{h}) = b_n^{-d} \tilde{J}_n(\mathbf{h}) \prod_{j=1}^d (b_n - |h_j|)$ when $\|\mathbf{h}\|_\infty \le b_n$. By $|1 - b_n^{-d} \prod_{j=1}^d (b_n - |h_j|)| \le Cb_n^{-1} \|\mathbf{h}\|_\infty$ for $\mathbf{h} \in \mathbb{Z}^d$,

$\|\mathbf{h}\|_\infty \le b_n$ along with $\sup_{n\ge 1,\mathbf{h}\in\mathbb{Z}^d} \|\tilde{J}_n(\mathbf{h})\| \le C$ and (9), we have

$$\left\| \sum_{\mathbf{h}\in\mathbb{Z}^d, \|\mathbf{h}\|_\infty \le b_n} \{\tilde{J}_n(\mathbf{h}) - \tilde{K}_n(\mathbf{h})\}\sigma(\mathbf{h}) \right\| \le b_n^{-1} C \sum_{m=1}^{b_n} m^{d-\tau\delta/(6+\delta)} = o(1).$$

Hence, $\|\mathrm{E}\widehat{\Sigma}_n - \sum_{\|\mathbf{h}\|_\infty \le b_n} \tilde{J}_n(\mathbf{h})\sigma(\mathbf{h})\| = o(1)$ and we now have $\|\mathrm{E}\widehat{\Sigma}_n - \Sigma_{\beta_0}\| = o(1)$.

To show $\mathrm{Var}(\widehat{\Sigma}_n) = o(1)$, write $\sigma_n(\mathbf{i},\mathbf{h}) = \mathrm{Cov}(W_\mathbf{i} W_\mathbf{i}', W_{\mathbf{i}+\mathbf{h}} W_{\mathbf{i}+\mathbf{h}}')$ and expand

$$\mathrm{Var}(\widehat{\Sigma}_n) = b_n^{2d} \sum_{\mathbf{h}\in\mathbb{Z}^d} \sum_{\mathbf{i}\in\mathcal{I}_n\cap(\mathcal{I}_n-\mathbf{h})} \sigma_n(\mathbf{i},\mathbf{h}) \equiv T_{1n} + T_{2n}$$

as two sums T_{1n}, T_{2n} over $\|\mathbf{h}\|_\infty \le 2b_n$ or $> 2b_n$, respectively. For any $\mathbf{i} \in \mathcal{I}_n, \mathbf{h} \in \mathbb{Z}^d$, it follows that $\|\sigma_n(\mathbf{i},\mathbf{h})\| \le C(\mathrm{E}\|W_\mathbf{i}\|^4 \mathrm{E}\|W_{\mathbf{i}+\mathbf{h}}\|^4)^{1/2} \le C b_n^{-2d} M_n^4$ by Holder's inequality, (6), (7); hence, $\|T_{1n}\| \le C\lambda_n^d b_n^d M_n^4 = o(1)$ by A.1, A.2, $|\mathcal{I}_n| \le \lambda_n^d$ and $|\{\mathbf{h} \in \mathbb{Z}^d : \|\mathbf{h}\|_\infty \le 2b_n\}| \le C b_n^d$. For $\|\mathbf{h}\|_\infty > 2b_n$, $\mathrm{dis}\{\mathcal{B}_n(\mathbf{0}), \mathcal{B}_n(\mathbf{h})\} = \|\mathbf{h}\|_\infty - b_n \ge \|\mathbf{h}\|_\infty/2$ holds implying that

$$\|\sigma_n(\mathbf{i},\mathbf{h})\| \le C(\mathbf{i},\mathbf{h})\alpha(\|\mathbf{h}\|_\infty/2, b_n^d)^{\frac{\delta}{6+\delta}} \le C b_n^{-2d} M_n^4 \|\mathbf{h}\|_\infty^{-\frac{\tau\delta}{6+\delta}} g(b_n^d)^{\frac{\delta}{6+\delta}}, \ \mathbf{i} \in \mathcal{I}_n,$$

by A.1 and (6)–(7), where $C(\mathbf{i},\mathbf{h}) \equiv (\mathrm{E}\|W_\mathbf{i}\|^{\frac{2(6+\delta)}{3}} \mathrm{E}\|W_{\mathbf{i}+\mathbf{h}}\|^{\frac{2(6+\delta)}{3}})^{\frac{3}{6+\delta}}$. By $|\mathcal{I}_n| \le \lambda_n^d$, $|\mathcal{I}_n \cap (\mathcal{I}_n - \mathbf{h})| = 0$ for $\|\mathbf{h}\|_\infty > \lambda_n$ and $|\{\mathbf{h} \in \mathbb{Z}^d : \|\mathbf{h}\|_\infty = m \ge 1\}| \le C m^{d-1}$, we find

$$\begin{aligned} \|T_{2n}\| &\le C g(b_n^d)^{\frac{\delta}{6+\delta}} \lambda_n^d M_n^4 \sum_{m=2b_n+1}^{\lambda_n} m^{d-1-\frac{\tau\delta}{6+\delta}} \\ &\le \begin{cases} C\lambda_n^{d+1} M_n^4 b_n^{d-1-\frac{(3\tau+d)\delta}{24+4\delta}} = o(1) & d > 1 \\ o(\lambda_n^d M_n^4) = o(1) & d = 1, \end{cases} \end{aligned}$$

by A.1 and A.3. Then, $\mathrm{Var}(\widehat{\Sigma}_n) = o(1)$ follows, proving Lemma 1(iv). □

Proof of Theorem 1 We begin with Theorem 1(i), using notation in Lemma 1. By A.4 and (4), we may write $R_n(\beta_0) = \prod_{\mathbf{i}\in\mathcal{I}_n}(1+\gamma_{\theta_0,\mathbf{i}})^{-1}$ with $\gamma_{\beta_0,\mathbf{i}} = \tilde{t}_{\beta_0}' W_\mathbf{i} > -1$, where $\tilde{t}_{\beta_0} = A_n^{1/2} t_{\beta_0} \in \mathbb{R}^q$ satisfies

$$\sum_{\mathbf{i}\in\mathcal{I}} \frac{W_\mathbf{i}}{1+\gamma_{\beta_0,\mathbf{i}}} = 0_q. \tag{11}$$

By Lemma 1(iii), it holds that $Z_n \equiv \max_{\mathbf{i}\in\mathcal{I}_n} \|W_\mathbf{i}\| = o_p(b_n^{-d})$. From this combined $W_n = O_p(1)$ by Lemma 1(ii), we deduce $\|\tilde{t}_{\beta_0}\| = O_p(b_n^d)$ from (11) following Owen (1990, pp. 101) so that $\max_{\mathbf{i}\in\mathcal{I}_n} |\gamma_{\beta_0,\mathbf{i}}| \le \|\tilde{t}_{\beta_0}\| Z_n = o_p(1)$.

We may algebraically solve (11) to find

$$\tilde{t}_{\beta_0} = b_n^d \widehat{\Sigma}_n^{-1} W_n + \phi_n, \qquad \phi_n \equiv b_n^d \widehat{\Sigma}_n^{-1} \sum_{\mathbf{i}\in\mathcal{I}_n} W_\mathbf{i} \gamma_{\beta_0,\mathbf{i}}^2/(1+\gamma_{\beta_0,\mathbf{i}}),$$

where $\|\phi_n\| \leq b_n^d Z_n \|\tilde{t}_{\beta_0}\|^2 \|\widehat{\Sigma}_n^{-1}\| \|\widehat{\Sigma}_n\|/(1 - \|\tilde{t}_{\beta_0}\| Z_n) = o_p(b_n^d)$. Applying a Taylor expansion gives $\log(1 + \gamma_{\beta_0,\mathbf{i}}) = \gamma_{\beta_0,\mathbf{i}} - \gamma_{\beta_0,\mathbf{i}}^2/2 + \Delta_{\mathbf{i}}$ for each $\mathbf{i} \in \mathcal{I}_n$ so that

$$-2b_n^{-d} \log R_n(\beta_0) = W_n' \widehat{\Sigma}_n^{-1} W_n - b_n^{-2d} \phi_n' \widehat{\Sigma}_n \phi_n + 2b_n^{-d} \sum_{\mathbf{i} \in \mathcal{I}_n} \Delta_{\mathbf{i}}.$$

Theorem 1(i) follows by Slutsky's Theorem using $W_n' \widehat{\Sigma}_n^{-1} W_n \xrightarrow{d} \chi_q^2$ by Lemma 1, $b_n^{-2d} \phi_n' \widehat{\Sigma}_n \phi_n = o_p(1)$ and $b_n^{-d} |\sum_{\mathbf{i} \in \mathcal{I}_n} \Delta_{\mathbf{i}}| \leq b_n^{-2d} Z_n \|\tilde{t}_{\beta_0}\|^3 \|\widehat{\Sigma}_n\|/(1 - \|\tilde{t}_{\beta_0}\| Z_n)^3 = o_p(1)$.

To prove Theorem 1(ii), fix a large n for which $Q_n = \sum_{\mathbf{i} \in \mathcal{I}_n} Q_{\mathbf{i}}$ is positive definite by Lemma 1(v). For $\nu \in (0, 1]$, let $\mathcal{P}_\nu$ denote the simplex of probabilities $\{p_{\mathbf{i}}\} \equiv \{p_{\mathbf{i}} : \mathbf{i} \in \mathcal{I}_n\}$ satisfying $\sum_{\mathbf{i} \in \mathcal{I}_n} p_{\mathbf{i}} = 1$, $\prod_{\mathbf{i} \in \mathcal{I}_n} N_{\mathcal{I}} p_{\mathbf{i}} \geq \nu$. Then, for any $\{p_{\mathbf{i}}\} \in \mathcal{P}_\nu$, the positivity of Q_n and the fact that, for each $\mathbf{i} \in \mathcal{I}_n$, $Q_{\mathbf{i}}$ is non-negative definite and $p_{\mathbf{i}} > 0$ imply the matrix $\sum_{\mathbf{i} \in \mathcal{I}_n} p_{\mathbf{i}} Q_{\mathbf{i}}$ must be positive and we may define $\beta_{\{p_{\mathbf{i}}\}} \equiv (\sum_{\mathbf{i} \in \mathcal{I}_n} p_{\mathbf{i}} Q_{\mathbf{i}})^{-1} \sum_{\mathbf{i} \in \mathcal{I}_n} p_{\mathbf{i}} X Z_{\mathbf{i}}$ for $X Z_{\mathbf{i}} \equiv \sum_{\mathbf{s} \in \mathcal{B}_n(\mathbf{i}) \cap \mathbb{Z}^d} X(\mathbf{s}) Z(\mathbf{s})/b_n^d$, $\mathbf{i} \in \mathcal{I}_n$. Theorem 1(ii) then follows from the set equality $\{\beta \in \mathbb{R}^q : R_n(\beta) \geq \nu\} = \{\beta_{\{p_{\mathbf{i}}\}} \in \mathbb{R}^q : \{p_{\mathbf{i}}\} \in \mathcal{P}_\nu\}$ and the fact that the latter set is connected in $\mathbb{R}^q$ since $\beta_{\{p_{\mathbf{i}}\}}$ is a continuous function on the compact, convex set $\mathcal{P}_\nu$.

Theorem 1(iii) follows from simple modifications to Theorem 1(i) arguments. □

Proof of Theorem 2 In the case that the regressors $X(\cdot)$ are stochastic and stationary, Lemma 1 remains valid upon re-defining $A_n \equiv N_{\mathcal{I}}$ there (i.e., $W_{\mathbf{i}} = N_{\mathcal{I}}^{-1/2} \bar{Y}_{\beta_0,\mathbf{i}}$). With stochastic regressors, the same proof for Lemma 1(ii)–(iv) applies under the convention that we set $M_n \equiv N_{\mathcal{I}}^{-1/2}$ in the arguments, drop "$\sup_{\mathbf{s} \in T} \|X(\mathbf{s})\|^6$" in (6), and note $\mathrm{E}\widehat{\Sigma}_n = b_n^d \mathrm{Var}(\bar{Y}_{\beta_0,\mathbf{0}}) \to \Sigma_{\beta_0}$ using stationarity. Lemma 1(v) still holds because $Q_n/N_{\mathcal{I}} \xrightarrow{p} \mathrm{E}X(\mathbf{0})X(\mathbf{0})'$, which is positive definite. Then, Theorem 2 follows from the proof of Theorem 1, noting A.4 can be shown to hold for stationary regressors.

□

References

Bradley RC (1989) A caution on mixing conditions for random fields. Stat Probab Lett 8:489–491

Bravo F (2002) Testing linear restrictions in linear models with empirical likelihood. Econom J Online 5:104–130

Bravo F (2004) Empirical likelihood based inference with applications to some econometric models. Econom Theory 20:231–264

Bravo F (2005) Blockwise empirical entropy tests for time series regressions. J Time Ser Anal 26:185–210

Carlstein E (1986) The use of subseries methods for estimating the variance of a general statistic from a stationary time series. Ann Stat 14:1171–1179

Chan G, Wood ATA (1997) An algorithm for simulating stationary Gaussian random fields. Appl Stat 46:171–181

Chen SX (1993) On the coverage accuracy of empirical likelihood regions for linear regression models. Ann Inst Stat Math 45:621–637

Chuang C, Chan NH (2002) Empirical likelihood for autoregressive models, with applications to unstable time series. Stat Sin 12:387–407

Cressie N (1993) Statistics for Spatial Data, 2nd edn. Wiley, New York

Doukhan P (1994) Mixing: properties and examples. Lecture notes in Statistics. vol 85, Springer, New York

Hall P, Horowitz JL (1996) Bootstrap critical values for tests based on generalized-method-of-moments estimators. Econometrica 64:891–916
Hall P, La Scala B (1990) Methodology and algorithms of empirical likelihood. Int Stat Rev 58:109–127
Imbens GW, Spady RH, Johnson P (1998) Information theoretic approaches to inference in moment condition models. Econometrika 66:333–357
Kitamura Y (1997) Empirical likelihood methods with weakly dependent processes. Ann Stat 25:2084–2102
Künsch HR (1989) The jackknife and the bootstrap for general stationary observations. Ann Stat 17:1217–1241
Lahiri SN (2003a) Resampling methods for dependent data. Springer, New York
Lahiri SN (2003b) Central limit theorems for weighted sums of a spatial process under a class of stochastic and fixed designs. Sankhya Ser A 65:356–388
Lee YD, Lahiri SN (2002) Least squares variogram fitting by spatial subsampling. J R Stat Soc Ser B 64:837–854
Lin L, Zhang R (2001) Blockwise empirical Euclidean likelihood for weakly dependent processes. Stat Probab Lett 53:143–152
Monti AC (1997) Empirical likelihood confidence regions in time series models. Biometrika 84:395–405
Nordman DJ, Caragea PC (2007) Joint and interval estimation of variogram models using spatial empirical likelihood. J Am Stat Assoc (in press)
Nordman DJ, Lahiri SN (2004) On optimal spatial subsample size for variance estimation. Ann Stat 32:1981–2027
Nordman DJ, Lahiri SN (2006) A frequency domain empirical likelihood for short- and long-range dependence. Ann Stat 34:3019–3050
Nordman DJ, Sibbertsen P, Lahiri SN (2006) Empirical likelihood for the mean under long-range dependence. J Time Ser Anal (in press)
Owen AB (1990) Empirical likelihood confidence regions. Ann Stat 18:90–120
Owen AB (1991) Empirical likelihood for linear models. Ann Stat 19:1725–1747
Owen AB (2001) Empirical likelihood. Chapman & Hall, London
Politis DN, Romano JP (1994) Large sample confidence regions based on subsamples under minimal assumptions. Ann Stat 22:2031–2050
Politis DN, Romano JP, Wolf M (1999) Subsampling. Springer, New York
Qin J, Lawless J (1994) Empirical likelihood and general estimating equations. Ann Stat 22:300–325
Zhang J (2006) Empirical likelihood for NA series. Stat Probab Lett 76:153–160