Statistica Sinica **25** (2015), 1399-1420
doi:http://dx.doi.org/10.5705/ss.2013.385

EMPIRICAL LIKELIHOOD FOR IRREGULARLY LOCATED SPATIAL DATA

Matthew Van Hala, Daniel J. Nordman and Zhengyuan Zhu

Iowa State University

Abstract: We develop an empirical likelihood (EL) method for inference over a broad class of spatial data exhibiting stochastic spatial patterns with various levels of infill sampling. Without stringent assumptions about the sampling design or spatial dependence, the EL method (based on estimating functions) produces log-likelihood ratio statistics having chi-square limits for calibrating tests and confidence regions for spatial parameters. Maximum EL estimators are valid for point estimation and formulating tests of spatial structure conditions. The proposed EL approach applies additionally to inference in spatial regression models with irregularly located sampling sites. The method is illustrated with a data example and investigated through simulation for calibrating confidence intervals and goodness-of-fit tests.

Key words and phrases: Blockwise empirical likelihood, infill sampling, spatial regression, stationarity, stochastic sampling.

1. Introduction

Introduced by Owen (1988, 1990), empirical likelihood (EL) is a statistical methodology for likelihood-type inference without an explicit distributional model for the data. For assessing parameter values, EL formulates a nonparametric likelihood function by probability profiling data and produces ratio statistics for inference having some properties analogous to fully parametric likelihood (e.g., chi-square limits). Extending EL to dependent data remains a challenge. In a pivotal work, Kitamura (1997) showed that EL versions for independent and identically distributed (iid) data generally fail for time series in the presence of correlation. As a remedy, Kitamura (1997) proposed a blockwise EL (BEL) for time series based on data blocking techniques to capture the dependence in neighboring observations. This general BEL approach has been shown to be valid over several inference problems with time series (cf., Bravo (2005); Chen and Wong (2009); Wu and Cao (2011)); see Nordman and Lahiri (2013) for a review of EL for time series.

In contrast to time series, EL for spatial data has received less consideration, though some extensions exist. Nordman (2008) developed a BEL version for spatial processes observed on a partial grid in $\mathbb{R}^d$, extending Kitamura's (1997) time series results ($d = 1$). Nordman and Caragea (2008) considered a spatial BEL for estimating variogram model parameters, and Kaiser and Nordman (2012) developed goodness-of-fit tests for spatial Markov models. Recently, Kostov (2013) proposed a smoothed EL method for inference in spatial quantile regressions.

However, all these spatial EL works are limited to spatial *lattice* data, or data collected at regular locations, corresponding to a type of sampling closely connected to equi-spaced time series. While lattice data provide an important form of spatial data, more diverse structures for spatial data are common in applications, typically involving irregularly located spatial observations. Our goal here is to advance EL methodology for inference about this form of spatial data. As a complicating factor with such data, the large-sample distribution of spatial estimators generally depends on an complex interaction of factors (cf., Lahiri (2003)), including the generating mechanism of sampling sites, the unknown correlation of the spatial process, and the type of spatial asymptotic structure. An advantage of EL in this setting is that the method, if appropriately formulated, can provide valid inference without restrictive assumptions or explicit knowledge about any of these factors. Regarding spatial asymptotic structure, previous spatial versions of EL (Nordman (2008)), like most spatial resampling developments, have focused on lattice data at fixed regular distances in a "pure increasing domain" (PID) asymptotic framework (cf., Cressie (1993), Sherman (1996)). This entails that more spatial observations are collected as a spatial sampling region, say $R_n \subset \mathbb{R}^d$, expands in size, similarly to asymptotics for equi-spaced time series, and the region volume $\text{vol}(R_n)$ and number n of spatial observations are asymptotically proportional. In contrast, here we allow for *two* natural types of spatial asymptotic structures with irregularly located data, where $\text{vol}(R_n)$ and n could be proportional or where the sample size n may be larger in magnitude than $\text{vol}(R_n)$, providing a "mixed increasing domain" (MID) framework (cf., Hall and Patil (1994)). The latter design allows for *infill* sampling of any spatial subregion, or many observations sampled in arbitrarily close proximity, as appropriate in some geostatistical and environmental applications. A challenge is that distributional properties and limiting variances of even simple spatial statistics, like sample means, typically change between PID and MID cases (Lahiri (2003)); Matsuda and Yajima (2009)), which makes direct variance estimation difficult, especially because PID/MID frameworks can be hard to distinguish in finite-sample applications. For data with general stochastic locations, we consider a spatial blockwise EL method (SBEL) that shares the common EL property of requiring no direct variance estimation steps and thus applies in a uniform manner under PID and MID spatial structures. In contrast, many model-based and nonparametric inference approaches to irregularly spaced data (cf., Politis, Paparoditis, and Romano (1998); Politis and Sherman (2001)) often assume such data are generated from a homogenous Poisson process, which only allows a uniform distribution for sampling sites with PID structure.

To investigate the SBEL method, we use a spatial sampling design considered by Lahiri and Zhu (2006) for a block bootstrap method. This set-up involves a stationary spatial process observed under a stochastic sampling design where spatial locations are determined by iid random vectors with a potentially non-uniform density and the number n of observations can grow at a rate, equal to or exceeding, the size of a spatial sampling region. SBEL uses general estimating functions and a data-blocking device to develop EL-ratio statistics for spatial parameters determined by the process marginal distribution. The main distributional result shows that the Wilks phenomenon (Wilks (1938)) remains valid, establishing that log-SBEL statistics have chi-square limits under mild conditions, regardless of the concentration of spatial locations or the amount of infill sampling. Maximum SBEL estimators are also considered for developing further SBEL goodness-of-fit tests of spatial moment conditions, which are useful for assessing spatial distributional structures. The SBEL method also applies for estimation of spatial regression models with irregular spatial data.

The rest of the manuscript is organized as follows. Section 2 describes the spatial sampling design for defining the locations of spatial observations. Section 3 then explains the SBEL method for irregularly located spatial data, based on general estimating functions (Section 3.1) and a data blocking technique (Section 3.2). Section 4 states the main distributional results of the paper, describing a Wilks result for SBEL as well as further inference results based on maximum EL estimation. Section 5 provides a simulation study, investigating the SBEL method for confidence intervals (Section 5.1) and goodness-of-fit tests (Section 5.2). A data example in Section 6 illustrates SBEL inference with spatial regression, and Section 7 provides some concluding remarks. Proofs of the main results appear in an appendix here, as well as a supplementary web-appendix.

2. Spatial Data and Sampling Design

We briefly recall the general spatial sampling design of Lahiri and Zhu (2006). Consider an $\mathbb{R}^m$-valued continuously indexed, stationary spatial process $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ observed at n irregularly-spaced sites $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n \in R_n$ within a spatial sampling region $R_n \subset \mathbb{R}^d$; here $d \geq 1$ represents the dimension of spatial locations and each spatial observation $Z(\boldsymbol{s}) \in \mathbb{R}^m$ is a random vector of length m. To describe the sampling region R_n, let R_0 denote a connected subset of $(-1/2, 1/2]^d$ containing the origin to serve as a "template" shape and let $\{\lambda_n\}$ be a positive real sequence such that $\lambda_n \to \infty$ as $n \to \infty$. Then, we assume that the sampling region $R_n = \lambda_n R_0$ is obtain by inflating region R_0 by scaling factor λ_n, providing a common "expanding domain" framework in asymptotic studies of spatial statistics (cf., Cressie (1993); Politis and Sherman (2001); Nordman (2008)). To avoid pathological sampling region shapes, for any positive sequence $a_n \to 0$ as $n \to \infty$, we assume that the number of cubes on the scaled lattice $a_n\mathbb{Z}^d$ which intersect both R_0 and R_0^c is $O(a_n^{-(d-1)})$ as $n \to \infty$, which holds for most sampling region shapes of practical interest (cf., Lahiri and Zhu (2006)).

For defining the locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$ of spatial observations in R_n, suppose $\{\boldsymbol{X}_n : i \geq 1\} \subset R_0$ denotes a sequence of iid $\mathbb{R}^d$-valued random vectors, with a common density function $f(\boldsymbol{x})$ on R_0, that is independent of $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$. Then, the sampling sites $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$ in R_n are taken as $\boldsymbol{s}_i = \lambda_n \boldsymbol{X}_i, i = 1, \ldots, n$, based on a realization of the random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ in R_0. The density function $f(\boldsymbol{x})$ controls the concentration of spatial observations over different parts of R_0 and consequently the sampling region R_n. As $f(\boldsymbol{x})$ may be general, the pattern of sampling locations can be complex and non-uniform over R_n, and no knowledge of $f(\boldsymbol{x})$ is required in the spatial EL method. The sample size n is assumed to grow at a rate equal to, or possibly faster, than volume $\text{vol}(R_n) = \lambda_n^d \text{vol}(R_0)$ of the spatial sampling region, namely $\lim_{n\to\infty} \lambda_n^d/n = c \in [0, \infty)$. The cases $c > 0$ or $c = 0$ correspond to PID or MID asymptotic structures, respectively, where heavy infill sampling $c = 0$ induces strong dependence associated with many samples filling a subregion of R_n.

3. Spatial Blockwise Empirical Likelihood (SBEL)

3.1. General estimating functions

Suppose $Z(\cdot) \in \mathbb{R}^m$ is a spatial process observed at irregularly-spaced sites $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n \in R_n \subset \mathbb{R}^d$, with the spatial sampling design in Section 2. We focus on the case of a stationary process $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ for describing parameter inference and estimating functions; extensions to general spatial regression models are also described (see Example 1 below). Suppose we are interested in inference about a spatial parameter $\theta \in \Theta \subset \mathbb{R}^p$ that can be linked to the spatial data through a system of estimating functions. To this end, let $g(\mathbf{z}; \theta) : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}^r$ be a vector of $r \geq p$ estimating functions satisfying an expectation condition

$$\mathrm{E}\left\{g(Z(\boldsymbol{s}); \theta_0)\right\} = 0_r \tag{3.1}$$

at the true parameter value $\theta_0 \in \mathbb{R}^p$, where 0_r denotes the $\mathbb{R}^r$-zero vector. When $r > p$, the estimating functions are said to be over-identifying for $\theta \in \mathbb{R}^p$. Estimating functions with EL have been considered by various authors (cf., Qin and Lawless (1994) for iid data; Kitamura (1997) and Bravo (2005) for time series; Nordman (2008) for gridded spatial data), and such functions define spatial EL ratio statistics in Section 3.2. Examples of estimating functions are provided in the numerical studies of Section 6, with a further example given next.

Example 1. For spatial EL inference about a non-stationary spatial process due to trend, consider a spatial regression model

$$Z(\boldsymbol{s}) = w(\boldsymbol{s})'\beta + \varepsilon(\boldsymbol{s}),\ \boldsymbol{s} \in \mathbb{R}^d, \tag{3.2}$$

where $w(\boldsymbol{s}) : \mathbb{R}^d \to \mathbb{R}^q$ is a non-random weight function (here $w(\boldsymbol{s})$ could be a function of available spatial covariates $x_1(\boldsymbol{s}), \ldots, x_j(\boldsymbol{s})$ in addition to the spatial location $\boldsymbol{s} \in \mathbb{R}^d$), $\beta \in \mathbb{R}^q$ is a vector of regression parameters, and $\{\varepsilon(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ is a mean-zero, real-valued stationary process. For EL inference about the regression parameter $\beta \in \mathbb{R}^q$, one may use q estimating functions $g(Z(\boldsymbol{s}); \beta) = w(\boldsymbol{s})[Z(\boldsymbol{s}) - w(\boldsymbol{s})'\beta] \in \mathbb{R}^q$, similarly to Owen (1991), satisfying $\mathrm{E}\{g(Z(\boldsymbol{s}); \beta_0)\} = 0_q$ at the true regression parameter $\beta_0 \in \mathbb{R}^q$. As the variables $\varepsilon(\boldsymbol{s}) = Z(\boldsymbol{s}) - w(\boldsymbol{s})'\beta$ are stationary, one may also develop additional $\tilde{r}$ estimating functions $\tilde{g}(\varepsilon(\boldsymbol{s}); \tilde{\theta}) \in \mathbb{R}^{\tilde{r}}$ for inference about parameters $\tilde{\theta} \in \mathbb{R}^{\tilde{p}}$ associated with the error distribution ($\tilde{r} \geq \tilde{p}$) and consider a set of $r = q + \tilde{r}$ estimating functions

$$g(Z(\boldsymbol{s}); \theta) = \begin{bmatrix} w(\boldsymbol{s})[Z(\boldsymbol{s}) - w(\boldsymbol{s})'\beta] \\ \tilde{g}(Z(\boldsymbol{s}) - w(\boldsymbol{s})'\beta; \tilde{\theta}) \end{bmatrix}$$

satisfying (3.1) for inference about $\theta = (\beta', \tilde{\theta})' \in \mathbb{R}^p$, $p = q + \tilde{p}$ (i.e., both the regression parameters β and additional parameters $\tilde{\theta}$). While we focus on presenting results for the stationary spatial case in the following, the same EL method also remains valid for the spatial regression model above; see Remark 2, Section 4.3, and the data example of Section 6.

3.2. Data blocking and SBEL ratio

The blockwise EL of Kitamura (1997) for time series creates an EL function from data blocks to preserve the dependence in neighboring observations, and the same blocking principle applies to spatial lattice data (Nordman (2008)). However, for irregularly spaced spatial data, caution is required in data blocking. Lahiri and Zhu (2006) have shown that some data blocking schemes, defined by the positions of sampling sites (cf., Politis and Sherman (2001)), are generally invalid for spatial resampling approaches unless sampling locations are uniform. Following Lahiri and Zhu (2006), we adopt a blocking scheme which superimposes an integer grid $\mathbb{Z}^d$ to subsequently define data blocks from observations in $R_n \subset \mathbb{R}^d$. Let $b_n = b$ denote a sequence of positive integers such that $b \to \infty$ as $n \to \infty$ with $b/\lambda_n \to 0$; the latter condition ensures that data blocks are smaller than the spatial sampling region R_n. For $B_n(\boldsymbol{i}) = \boldsymbol{i} + b(0, 1]^d$, $\boldsymbol{i} \in \mathbb{Z}^d$, define a collection of rectangular data blocks as $\{B_n(\boldsymbol{i}) : \boldsymbol{i} \in \mathcal{I}_n\}$, where $\mathcal{I}_n \equiv \{\boldsymbol{j} \in \mathbb{Z}^d : B_n(\boldsymbol{j}) \subset R_n\}$ denotes the index set of all integer-translated blocks lying completely within the region R_n; see Figure 1 for an illustration.

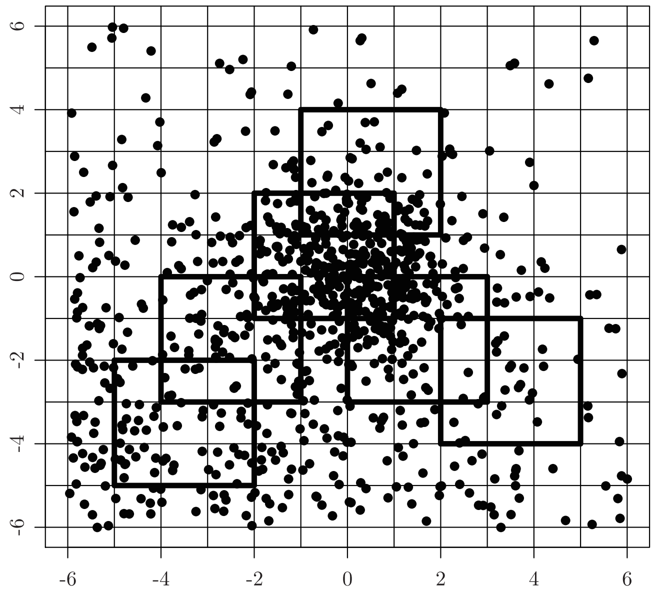

Figure 1. An example of the blocking scheme, showing irregularly located spatial data in a 12×12 sampling region in $\mathbb{R}^2$, along with a superimposed integer grid and overlapping rectangular data blocks ($b = 3$).

For inference on the spatial parameter θ defined by estimating functions applied to the n available spatial observations $\{g(Z(\boldsymbol{s}_j);\theta) \in \mathbb{R}^r : j = 1,\ldots,n\}$, we compute a weighted sum $S_n(\boldsymbol{i};\theta) = b^{-d}\sum_{j=1}^{n} g(Z(\boldsymbol{s}_j);\theta)\mathbb{I}(\boldsymbol{s}_j \in B_n(\boldsymbol{i}))$ of observations for each data block $\boldsymbol{i} \in \mathcal{I}_n$ where $\mathbb{I}(\cdot)$ denotes the indicator function. For convenience, if there are $N = |\mathcal{I}_n|$ data blocks, we re-label the block sums $\{S_n(\boldsymbol{i};\theta) : \boldsymbol{i} \in \mathcal{I}_n\}$ as $S_{nj}(\theta)$, $j = 1,\ldots,N$. For a given parameter value θ, we probability profile block sums to create a SBEL function as

$$L_n(\theta) = \sup\left\{\prod_{j=1}^{N} p_j :\ p_1,\ldots,p_N \geq 0,\ \sum_{j=1}^{N} p_j = 1,\ \sum_{j=1}^{N} p_j S_{nj}(\theta) = 0_r\right\} \tag{3.3}$$

and SBEL ratio as $R_n(\theta) = L_n(\theta)/N^{-N}$. The function $L_n(\theta)$ quantifies the plausibility of a value θ by maximizing a multinomial likelihood from probabilities $\{p_j\}_{j=1}^{N}$ assigned to the block sums $S_{nj}(\theta)$ under an expectation constraint which mimics the moment condition (3.1). Without this expectation constraint in (3.3), the multinomial product is maximized when each $p_j = 1/N$, leading to the ratio $R_n(\theta) \in [0,1]$. When 0_r is in the interior convex hull of $\{S_{nj}(\theta)\}_{j=1}^{N}$, then $L_n(\theta) = \prod_{j=1}^{N} p_{j,\theta} > 0$ holds for probabilities $p_{j,\theta} = N^{-1}[1+t'_{n,\theta}S_{nj}(\theta)]^{-1} \in (0,1)$, $j = 1,\ldots,N$, expressed in terms of a Lagrange multiplier $t_{n,\theta} \in \mathbb{R}^r$ satisfying $0_r = \sum_{j=1}^{N} S_{nj}(\theta)/[N(1 + t'_{n,\theta}S_{nj}(\theta)]$; see Owen (1990) and Qin and Lawless (1994) for computational details with EL.

4. Main Results

4.1. Conditions

We require some spatial dependence conditions on the process $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ prescribed in terms of its strong mixing coefficient. For $\boldsymbol{x} = (x_1, \ldots, x_d)' \in \mathbb{R}^d$, write norms $\|\boldsymbol{x}\| = (\sum_{i=1}^d x_i^2)^{1/2}$ and $\|\boldsymbol{x}\|_1 = \sum_{i=1}^d |x_i|$ and, for a subset $T \subset \mathbb{R}^d$, let $\mathcal{F}_Z(T) = \sigma\langle Z(\boldsymbol{s}) : \boldsymbol{s} \in T\rangle$ denote the σ-algebra generated by $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in T\}$. For subsets $T_1, T_2 \subset \mathbb{R}^d$, define $d(T_1, T_2) = \inf\{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_1 : \boldsymbol{x}_i \in T_i, i = 1, 2\}$ and $\tilde{\alpha}(T_1, T_2) = \sup\{P(A_1 \cap A_2) - P(A_1)P(A_2) : A_i \in \mathcal{F}_Z(T_i), i = 1, 2\}$. Then, the strong mixing coefficient of $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ is defined as

$$\alpha(a, b) = \sup\{\tilde{\alpha}(T_1, T_2) : \ d(T_1, T_2) \geq a, \ T_1, T_2 \in \mathcal{R}(b)\}, \quad a, b > 0,$$

where $\mathcal{R}(b)$ is the collection of all finite disjoint unions of cubes in $\mathbb{R}^d$ with a total volume not exceeding b; the definition of $\alpha(a, b)$ involves subsets $T_1, T_2 \subset \mathbb{R}^d$ of bounded volume to avoiding more restrictive forms of mixing (cf., Lahiri (2003); Lahiri and Zhu (2006)). We also assume that

$$\alpha(a, b) \leq C a^{-\tau_1} b^{\tau_2} \tag{4.1}$$

for some constants $C, \tau_1 > 0$ and $\tau_2 \geq 0$, where $\tau_2 = 0$ for $d = 1$ (bounding the usual mixing coefficient $\alpha(a, \infty)$ for time series $d = 1$). From (4.1), the mixing coefficient decays with the distance a between sets while the strength of spatial dependence is allowed to increase with the volumes of sets T_1, T_2. Doukhan (1994) provides examples of spatial processes satisfying such conditions (e.g., Gibbs, linear, or Markovian fields). We next prescribe assumptions for the SBEL method. Assumption (A1), involving τ_1, τ_2 in (4.1), is a function of an integer k to be specified in theorems to follow. Let $G_{\theta_0}(\boldsymbol{s}) = g(Z(\boldsymbol{s}); \theta_0)$, $\boldsymbol{s} \in \mathbb{R}^d$.

Assumptions:

(A1) For some $\delta > 0$ and integer $k \geq 1$, $\mathrm{E}\left\{\|G_{\theta_0}(\boldsymbol{s})\|^{2k+\delta}\right\} < \infty$ and

$$\tau_1 > \frac{(2k-1)(2k+\delta)}{\delta}, \qquad \tau_2 < \frac{\tau_1 - d}{4d} \quad (\text{for } d \geq 2).$$

(A2) $\Sigma_0 \equiv \int \sigma(\boldsymbol{x}) d\boldsymbol{x}$ is positive definite, $\sigma(\boldsymbol{x}) = \mathrm{Cov}\left[G_{\theta_0}(\boldsymbol{0}), G_{\theta_0}(\boldsymbol{x})\right], \boldsymbol{x} \in \mathbb{R}^d$.

(A3) The density $f(\cdot)$ of $\boldsymbol{X}_1$ is positive and continuous on the closure $\overline{R}_0$.

(A4) $\lim_{n\to\infty} n^{\epsilon}/\lambda_n = 0$ for some $\epsilon > 0$ and $\lim_{n\to\infty} \lambda_n^d/n = c \in [0, \infty)$.

These assumptions essentially match a subset of those of Lahiri and Zhu (2006) for establishing a block bootstrap with similar spatial data, where Assumption (A1) prescribes mild conditions on the strength of the spatial dependence. Assumption (A3) allows the density $f(\cdot)$ for spatial locations to be quite general and non-uniform. Assumption (A4) permits various possibilities for the amount of infill sampling, as described in Sections 1−2, depending on the number n of spatial observations relative to the volume $\mathrm{vol}(R_n) = \lambda_n^d \mathrm{vol}(R_0)$ of the sampling region. Finally, Assumption (A2) regards the limiting variance of a scaled sample mean of $\{G_{\theta_0}(\boldsymbol{s}_i)\}_{i=1}^n$, having a form $c\sigma(\mathbf{0}) + \Sigma_0$ that changes in PID/MID cases; see also Lemma 2(a) in the Appendix.

4.2. Basic wilks result for the spatial EL method

We give a fundamental Wilks result for the SBEL method, showing the log-EL ratio at the true parameter θ_0 has a chi-square limit. As a function of the spatial data $\{Z(\boldsymbol{s}_i)\}_{i=1}^n$ ($\boldsymbol{s}_i = \lambda_n \boldsymbol{X}_i$), the log-EL ratio technically has a conditional distribution given a realization of the sampling design random vectors $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$. Let $P_{\cdot|\boldsymbol{X}}(\cdot)$ denote conditional probability with respect to the σ-algebra $\sigma\langle\{\boldsymbol{X}_i : i \geq 1\}\rangle$ generated by $\{\boldsymbol{X}_i : i \geq 1\}$ (cf., Lahiri and Zhu (2006)) and let $P_{\boldsymbol{X}}$ denote the joint distribution of the random vectors $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$. Theorem 1 states that the limiting (conditional) distribution of the log-EL ratio is chi-square no matter what the spatial locations may be as determined by $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ under the sampling design.

Theorem 1. *Suppose Assumptions* (A1)−(A4) *hold with* $k = 3$ *in* (A1)*, and that* $b \to \infty$ *with* $b^2/\lambda_n = O(1)$ *as* $n \to \infty$*. Then, as* $n \to \infty$*,*

$$\sup_{y \in \mathbb{R}} \left| P_{\cdot|\boldsymbol{X}}\left(-2b^{-d} \log R_n(\theta_0) \leq y\right) - P(\chi_r^2 \leq y) \right| \to 0.$$

Equivalently, $-2b^{-d} \log R_n(\theta_0) \xrightarrow{d} \chi_r^2$, for any given $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ with probability 1 ($P_{\boldsymbol{X}}$). Regardless of the concentration of sampling sites or the size of an infill component, the SBEL method has a chi-square limit for calibrating confidence regions and tests. Here b^{-d} denotes a scalar correction to the EL ratio, which is needed to adjust for overlapping spatial blocks of data. Similar block adjustments exist for block-based EL statistics for time series (Kitamura (1997)) or gridded spatial data (Nordman (2008)).

4.3. Maximum spatial EL estimation and hypothesis testing

Analogously to parametric likelihood, the SBEL ratio $R_n(\theta)$ from (3.3) can be maximized over the parameter space Θ to produce a point estimator for $\theta \in \mathbb{R}^p$ characterized by the estimating functions $g(\cdot;\theta) \in \mathbb{R}^r$; we denote the result as the maximum empirical likelihood estimator (MELE) as $\hat{\theta}_n \in \mathbb{R}^p$ and the Lagrange multiplier associated with the MELE as $t_{n,\hat{\theta}_n} \in \mathbb{R}^r$. We next establish that the MELE has a normal limit under the stochastic spatial sampling design and, more importantly, that log-ratio statistics have chi-square limits. For independent, time series, and gridded spatial data, respectively, Qin and Lawless (1994), Kitamura (1997), and Nordman (2008) showed similar EL results.

Let $\Sigma_\infty = c\sigma(\mathbf{0}) + \Sigma_0$ for Σ_0 and $\sigma(\cdot)$ defined in Assumption (A2) and $c = \lim_{n\to\infty} \lambda_n^d/n \in [0,\infty)$ defined in Assumption (A4).

Theorem 2. *With Assumptions* (A1)$-$(A4)*, suppose that, in a neighborhood of* $\theta_0 \in \mathbb{R}^p$*, first partial derivatives* $\partial g(\cdot;\theta)/\partial\theta$ *exist and satisfy a Lipschitz condition of order* $\gamma > 0$*; assume, for* $\tilde{G}(\mathbf{0}) \equiv \partial g(Z(\mathbf{0});\theta_0)/\partial\theta \in \mathbb{R}^{r\times p}$*, that* $\mathrm{E}\{\|\tilde{G}(\mathbf{0})\|^{2+\delta}\} < \infty$ *for* $\delta > 0$ *in (A1) and that* $D_{\theta_0} \equiv \mathrm{E}\{\tilde{G}(\mathbf{0})\}$ *has full column rank p. Then, for any given* $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ *with probability 1* ($P_{\boldsymbol{X}}$)*: as* $n \to \infty$*,*

(i) *for* $V(\theta_0) = [D(\theta_0)'\Sigma_\infty D(\theta_0)]^{-1}$, $U(\theta_0) = \Sigma_\infty^{-1} - \Sigma_\infty^{-1} D(\theta_0)V(\theta_0)D(\theta_0)'\Sigma_\infty^{-1}$,

$$\lambda_n^{d/2}\begin{bmatrix} (\hat{\theta}_n - \theta_0) \\ b^{-d}\lambda_n^{-d} n t_{n,\hat{\theta}_n} \end{bmatrix} \xrightarrow{d} N\left[\begin{pmatrix} 0_p \\ 0_r \end{pmatrix}, \begin{pmatrix} V(\theta_0) & 0_{p\times r} \\ 0_{r\times p} & U(\theta_0) \end{pmatrix}\right];$$

(ii) *under* $H_0 : \theta = \theta_0$, $-2b^{-d}\log[R_n(\theta_0)/R_n(\hat{\theta}_n)] \xrightarrow{d} \chi^2_p$;

(iii) *under* H_0 : *the moment condition* (3.1) *holds for some* $\theta_0 \in \Theta$*,*
$-2b^{-d}\log R_n(\hat{\theta}_n) \xrightarrow{d} \chi^2_{r-p}$;

(iv) *the spatial EL test statistics in* (ii) *and* (iii) *are asymptotically independent.*

Remark 1. Under these regularity conditions, the proof of Theorem 2 shows that, in $P_{\cdot|\boldsymbol{X}}$-probability, a sequence of maximizers θ_n of $R_n(\theta)$ is guaranteed to exist on $\{\theta \in \Theta : \|\theta - \theta_0\| \le \lambda_n^{-d/2}\log n\}$ for any given $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ with probability 1 ($P_{\boldsymbol{X}}$). The normal limit in Theorem 2(i) is difficult to use directly, because the limiting variance is complicated and changes depending on PID/MID cases. However, this result establishes that the MELE $\hat{\theta}_n$ is consistent for the true θ_0 at a rate $\lambda_n^{-d/2}$. Consequently, in the PID case $c > 0$, $\hat{\theta}_n$ is consistent at a rate $n^{-1/2}$ in terms of the spatial sample size n, $n \propto \lambda_n^d$, but exhibits a slower rate, $n^{-1/2} \ll \lambda_n^{-d/2}$, in the MID case $c = 0$ associated with stronger dependence among sampled spatial observations.

Under the stochastic sampling design, Theorem 2(ii)$-$(iii) entails that log-ratio EL tests based on the MELE are valid for evaluating parameter hypotheses, as well as assessing whether spatial moment conditions (3.1) hold. The latter can be useful for assessing spatial structures, as illustrated in Section 5.2. Again, the limits and forms of these test statistics, based on a block correction factor b^{-d}, resemble those for time series and spatial lattice data (Kitamura (1997); Nordman (2008)) though the asymptotic sampling structure differs here. The distributional results in Theorem 2 also imply that the SBEL method is valid for testing subsets of the parameter vector θ (by probability profiling $R_n(\theta)$ as with regular likelihood) as well as testing under parameter constraints; see Nordman (2008) for similar tests with spatial lattice data.

Remark 2. The SBEL method applies for inference about spatial regression models (3.2). SBEL log-ratio statistics $-2b^{-d}\log R_n(\theta)$ follow from the estimating functions in Example 1 and the construction of Section 3.2. In this general spatial regression setting, chi-square limits for log-ratio statistics in Theorem 1 and Theorem 2(ii)−(iii) remain valid under some additional regularity conditions on the regressor weights $w(\cdot)$ in (3.2) (e.g., conditions C1, C2, C6 in Lahiri and Zhu (2006)). Section 6 provides a numerical illustration.

5. Simulation Results

We present simulation studies of the SBEL method for confidence interval (CI) estimation in Section 5.1 and goodness-of-fit (or moment) testing in Section 5.2.

5.1. Confidence intervals

We considered the performance of SBEL for constructing CIs for the mean $\mathrm{E}\{Z(\mathbf{s})\} = \mu$ of a real-valued stationary process $\{Z(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^2\}$ based on combinations of the following factors: sampling region size, sample size, stochastic sampling design, spatial dependence strength, and block size. Mean-zero Gaussian random fields on $\mathbb{R}^2$ were generated with an exponential covariance structure $\mathrm{Cov}\,(Z_{\mathbf{s}}, Z_{\mathbf{s}+\mathbf{h}}) = \exp\,(-\|\mathbf{h}\|/r)$ with dependence/range parameter $r > 0$. We considered sampling regions as $R_n = [-\lambda_n/2, \lambda_n/2]^2$ for $\lambda_n = 12, 24, 36, 48$, sample sizes of $n = 100$ or 900, with uniformly or non-uniformly distributed spatial locations on R_n. Non-uniform sites $\boldsymbol{s}_i = \lambda_n \boldsymbol{X}_i$, $i = 1, \ldots, n$ were generated as in Lahiri and Zhu (2006) by iid draws $\boldsymbol{X}_i \sim 0.5BN_1 + 0.5BN_2$ from a bivariate normal mixture (truncated to $[-1, 1]^2$), where BN_i represents a bivariate normal with mean $1 - (0.5^{i-1}, 0.5^{i-1})$ and identity covariance $\mathbf{I}_2$. Figure 2 shows an example of uniform and non-uniform locations.

For comparison against SBEL for estimating the process mean, we computed normal approximation CIs based on the sample mean $\bar{Z}_n = \sum_{i=1}^n Z(\boldsymbol{s}_i)/n$ (cf., Lahiri (2003)) as well as intervals based on a spatial block bootstrap method. For a given block size b, $M = 500$ spatial bootstrap data sets $\mathcal{Z}_1^*, \ldots, \mathcal{Z}_M^*$ were generated from a given simulated data set $\mathcal{Z} = \{Z(\boldsymbol{s}_i) : i = 1, \ldots, n\}$ using the bootstrap method of Lahiri and Zhu (2006), producing block bootstrap sample means, say $\bar{Z}_1^*, \ldots, \bar{Z}_M^*$, based on resulting bootstrap sample sizes, say, $n_1^*, \ldots, n_M^*$. Letting $\hat{V}$ denote the sample variance of $\{\bar{Z}_j^*\}_{j=1}^M$ as a bootstrap estimate of the variance of $\bar{Z}_n$, a $100(1-\alpha)\%$ normal approximation CI for μ is $\bar{Z}_n \pm z_{\alpha/2}\hat{V}^{1/2}$, using a lower $\alpha/2$ standard normal quantile $z_{\alpha/2}$. By approximating the distribution of

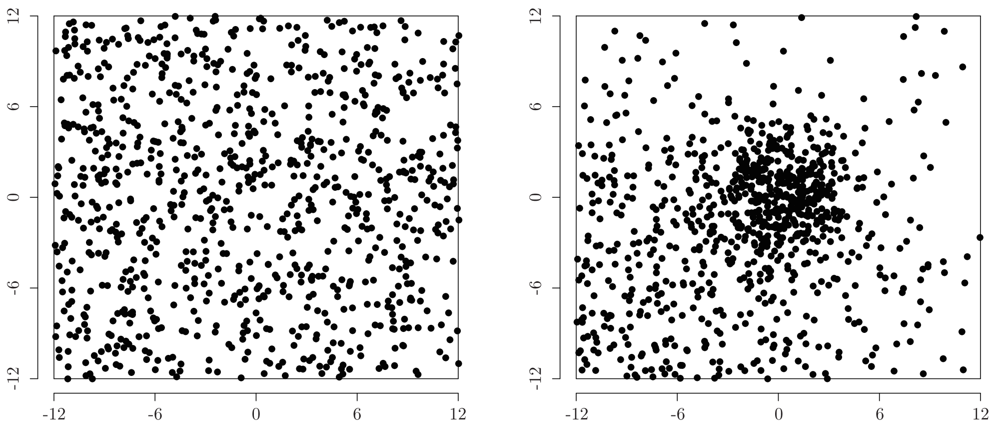

Figure 2. Example of uniform (left) and non-uniform (right) locations on a 24×24 grid with $n = 900$ points.

$(\bar{Z}_n - \mu)$ with the bootstrap counterpart $(\bar{Z}_i^* - \tilde{\mu})$ for $\tilde{\mu} = \sum_{i=1}^M n_i^* \bar{Z}_i^* / \sum_{i=1}^M n_i^*$, a $100(1-\alpha)\%$ block bootstrap CI for μ is given by $\left(\bar{Z}_n - q_{M,1-\alpha/2}, \bar{Z}_n - q_{M,\alpha/2}\right)$, where $q_{M,\gamma}$ denotes the γ quantile of the bootstrap values $\{\bar{Z}_i^* - \tilde{\mu}\}_{i=1}^M$. We used block sizes $b = 2, 4, 6$ for the 12×12 region, $b = 4, 6, 8$ for the 24×24 region, $b = 6, 9, 12$ for the 36×36 region, and $b = 6, 12, 16$ for the 48×48 region; these sizes roughly correspond to choices $C\mathrm{vol}(R_n)^{1/4}$ for $C = 0.5, 1, 1.5$, having optimal order known for some instances of block resampling (Nordman and Lahiri (2004)).

We implemented the SBEL method for constructing CIs for μ based a estimating function $g(Z(\boldsymbol{s}); \mu) = Z(\boldsymbol{s}) - \mu$ and two different distributional calibrations. We used the standard chi-square calibration for the SBEL log-ratio statistic (Theorem 1), producing a $100(1-\alpha)\%$ CI for μ as $\left\{\mu : -2b^{-2} \log R_n(\mu) \le \chi^2_{1,1-\alpha}\right\}$ with the same block sizes b used for the block bootstrap with each sampling region. As a second approach, we used the block boostrap to calibrate SBEL CIs. With the same bootstrap data sets $\mathcal{Z}_1^*, \ldots, \mathcal{Z}_M^*$ described above, let r_i^* denote the SBEL ratio from i^{th} bootstrap sample evaluated at $\bar{Z}_n$, $i = 1, \ldots, M$, which serves to approximate $R_n(\mu_0)$ at the true mean. Then, a $100(1-\alpha)\%$ SBEL interval with bootstrap calibration is $\left\{\mu : -2b^{-2} \log R_n(\mu) \le t_{M,1-\alpha}\right\}$, where $t_{M,1-\alpha}$ is the $1-\alpha$ sample quantile of $\left\{-2b^{-2} \log r_i^*\right\}_{i=1}^M$.

Based on 1,000 simulations, Table 1 shows empirical coverages for 90% CIs for SBEL, normal approximation, and block bootstrap methods with processes having dependence parameter $r = 1$. We make the following observations:

1. For any region size, block size, or sample size, coverage is better with uniform sites than for non-uniform sites. With non-uniform sites, there are concentrated pockets of sampling sites, which induces stronger dependence between

spatial observations, inducing lower coverage rates. This effect is most pronounced for smaller sampling regions (e.g., 12×12) with the larger sample size ($n = 900$), which corresponds to infill sampling and compounds the dependence between observations. With uniformly distributed sampling sites, the SBEL method exhibits coverages quite close to nominal coverage except for the smallest sampling region 12×12.

2. Coverage improves as grid size increases for all sample designs and sizes.
3. The coverage performance is markedly better for the SBEL method than the normal approximation or block bootstrap methods across all combinations of region size, block size, sample size, and stochastic sampling design.
4. For small sampling regions, the SBEL method with bootstrap calibration performs better than the chi-squared calibration; as the sampling regions increase, the two SBEL calibrations perform similarly.

A supplementary web-appendix provides empirical coverages for 90% CIs with values of the dependence parameter $r = 1/3$ or 3 for 12×12 and 24×24 sampling regions. As expected compared to the $r = 1$ case, coverage is generally better with weaker dependence ($r = 1/3$) and worse with stronger dependence ($r = 3$). In fact, non-uniform sites and a larger range $r = 3$ lead to "hot spots" of strongly correlated spatial observations (due to their close proximities), where all methods consequently exhibit severe under-coverage for the region sizes considered; larger sampling regions 36×36 and 48×48 are needed for better coverage in this situation. However, in all cases, the SBEL method again has better coverage accuracy than normal approximation and block bootstrap methods, and the bootstrap-based calibration for SBEL often performs better than the chi-square calibration with small sampling regions.

5.2. Goodness-of-fit tests

We also conducted a simulation study to examine the performance of the SBEL method for assessing goodness-of-fit of a specified marginal distribution for the data. We generated realizations of real-valued spatial processes $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^2\}$ having different marginal distributions (described below), and applied the SBEL method to assess if the data were marginally normally distributed. As a marginal normal distribution would be characterized by two unknown parameters $\theta = (\theta_1, \theta_2)'$ (mean and variance), we considered estimating functions g having more than two estimating functions for prescribing θ and satisfying the moment condition (3.1) when the data are indeed normally distributed. For comparison,

Table 1. Empirical coverage of 90% intervals for the mean for various methods: SBEL with chi-square calibration (ELC), SBEL with bootstrap-based calibration (ELB), normal approximation (Nor), and block bootstrap (Boot).

			Uniform Sites				Non-Uniform Sites			
			Method				Method			
Points	Grid Size	b	ELC	ELB	Nor	Boot	ELC	ELB	Nor	Boot
$n = 100$	12×12	2	74.3	77.4	69.6	68.9	60.0	67.2	56.7	58.6
		4	82.0	87.1	74.3	68.8	61.6	67.3	53.3	50.4
		6	82.7	89.4	67.7	64.5	53.2	64.2	38.1	35.9
	24×24	4	86.1	86.5	81.0	82.8	76.2	80.0	71.4	68.0
		6	88.6	90.5	82.6	80.3	74.8	86.8	68.9	65.7
		8	89.1	90.2	83.0	80.0	75.2	75.9	60.9	58.0
	36×36	6	89.4	91.3	87.3	87.5	82.0	83.5	77.4	78.0
		9	90.8	89.9	85.3	87.7	82.6	81.6	71.6	72.3
		12	92.7	91.6	83.1	82.9	82.5	78.7	63.3	65.3
	48×48	6	88.6	91.6	88.3	87.1	85.2	85.3	80.8	83.1
		12	90.9	91.3	87.8	85.2	85.2	84.6	77.3	72.7
		16	95.0	91.8	83.0	82.4	85.4	79.5	70.7	67.9
$n = 900$	12×12	2	65.3	69.7	62.7	60.7	53.6	62.5	53.7	50.9
		4	90.5	86.8	70.8	72.6	56.9	64.0	52.4	47.2
		6	81.1	90.0	67.2	67.0	50.6	63.9	32.4	35.3
	24×24	4	80.9	85.0	77.2	77.1	68.1	76.0	65.5	63.1
		6	85.9	89.7	77.7	76.3	67.5	71.5	62.8	59.2
		8	88.0	90.3	78.8	78.2	70.7	69.8	58.2	52.0
	36×36	6	84.0	91.0	82.8	80.1	72.9	79.8	71.2	69.5
		9	89.8	91.0	82.5	81.7	74.4	77.1	66.2	66.7
		12	91.8	90.9	80.3	78.8	77.4	72.4	61.6	57.8
	48×48	6	87.5	89.4	84.0	84.8	79.8	85.2	78.0	73.0
		12	89.2	91.8	84.8	84.3	77.0	76.1	71.9	67.2
		16	92.8	92.1	83.8	82.2	79.5	73.0	63.8	64.6

we considered three different sets of estimating functions $g(Z(\boldsymbol{s});\theta)$ as

$$\text{Set 1: } \Big[Z(\boldsymbol{s}) - \theta_1,\, (Z(\boldsymbol{s}) - \theta_1)^2 - \theta_2,\, (Z(\boldsymbol{s}) - \theta_1)^3\Big]',$$
$$\text{Set 2: } \Big[Z(\boldsymbol{s}) - \theta_1,\, (Z(\boldsymbol{s}) - \theta_1)^2 - \theta_2,\, (Z(\boldsymbol{s}) - \theta_1)^3,\, (Z(\boldsymbol{s}) - \theta_1)^4 - 3\theta_2^2\Big]',$$
$$\text{Set 3: } \Big[Z(\boldsymbol{s}) - \theta_1,\, (Z(\boldsymbol{s}) - \theta_1)^2 - \theta_2,\, \Phi\left(\tfrac{Z(\boldsymbol{s})-\theta_1}{\theta_2^{1/2}}\right) - 0.5\Big]',$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function (cdf). To assess normality based on a given set of functions $g(\cdot;\theta)$, we tested the hypothesis that the moment condition holds (3.1) using the log-ratio statistic $-2b^{-2}\log R_n(\hat{\theta}_n)$ which has a chi-square limit under Theorem 2(iii) (with 1 df for function sets 1 and 3, and 2 df for function set 2).

To generate data, we simulated dependent (marginally standard normal) normal observations $\tilde{Z}(\boldsymbol{s}_1), \ldots, \tilde{Z}(\boldsymbol{s}_n)$ as in Section 5.1, using 12×12 and 24×24 sampling regions, uniform and non-uniform locations for $n = 100$ sampling sites, and dependence parameter $r = 1$. Spatial realizations were then given by $Z(\boldsymbol{s}_i) = F[\Phi^{-1}[\tilde{Z}(\boldsymbol{s}_i)]]$, $i = 1, \ldots, n$ based on a probability integral transform using a proposal cdf $F(\cdot)$ to determine the marginal distribution structure. Choices of the cdf $F(\cdot)$ were taken as standard normal (corresponding to the null hypothesis), log-normal with $\sigma = 0.5$, chi-square with 1 or 20 df, or t-distribution with 2 or 20 df. For each sampling region, stochastic design, set of estimating functions, and marginal cdf $F(\cdot)$, 1,000 data sets were simulated for producing an empirical power function for the SBEL goodness-of-fit-tests, as shown in Figure 3 for 12×12 regions; a qualitatively similar figure is presented in the supplementary web-appendix for 24×24 regions. From these figures, one observes:

1. At most combinations of region size and stochastic sampling design, empirical sizes are close to nominal sizes for estimating functions sets 1 and 3 (as judged by results for normal data following the null hypothesis); the results are fairly insensitive to the block choices. Empirical size can be higher than nominal for estimating function set 2 and more sensitive to the block b, especially on the 12×12 region, so that a greater number of estimating functions need to be more cautiously applied on smaller sampling regions.
2. Power functions are best for estimating functions sets 1 and 2 based on over-identifying estimating functions involving higher distributional moments, and lowest for set 3 involving over-identification of parameters with a probability transform condition.
3. These sets of estimating functions perform very well in rejecting non-normal data that are clearly skewed or heavy-tailed (e.g. log-normal, χ^2_1, t_2) with small sample sizes $n = 100$, but then exhibit more difficulty in distinguishing those cases which are more closely normal (e.g., χ^2_{20}, t_{20}).

Simulations indicate that, even with small sample sizes, the SBEL method can provide an effective tool for goodness-of-fit assessments of spatial distributions.

6. Data Illustration

We illustrate the SBEL method applied to spatial regression (cf., Remark 2, Section 4.2) using a temperature data set from the National Oceanic and Atmospheric Administration's National Climatic Data Center. The data consist of the average January temperature between 1981 and 2010 (degrees Fahrenheit), latitude, longitude, and elevation (hundreds of feet above sea level) at 557 locations across the Midwest. The locations have latitude between 37 and 45 degrees

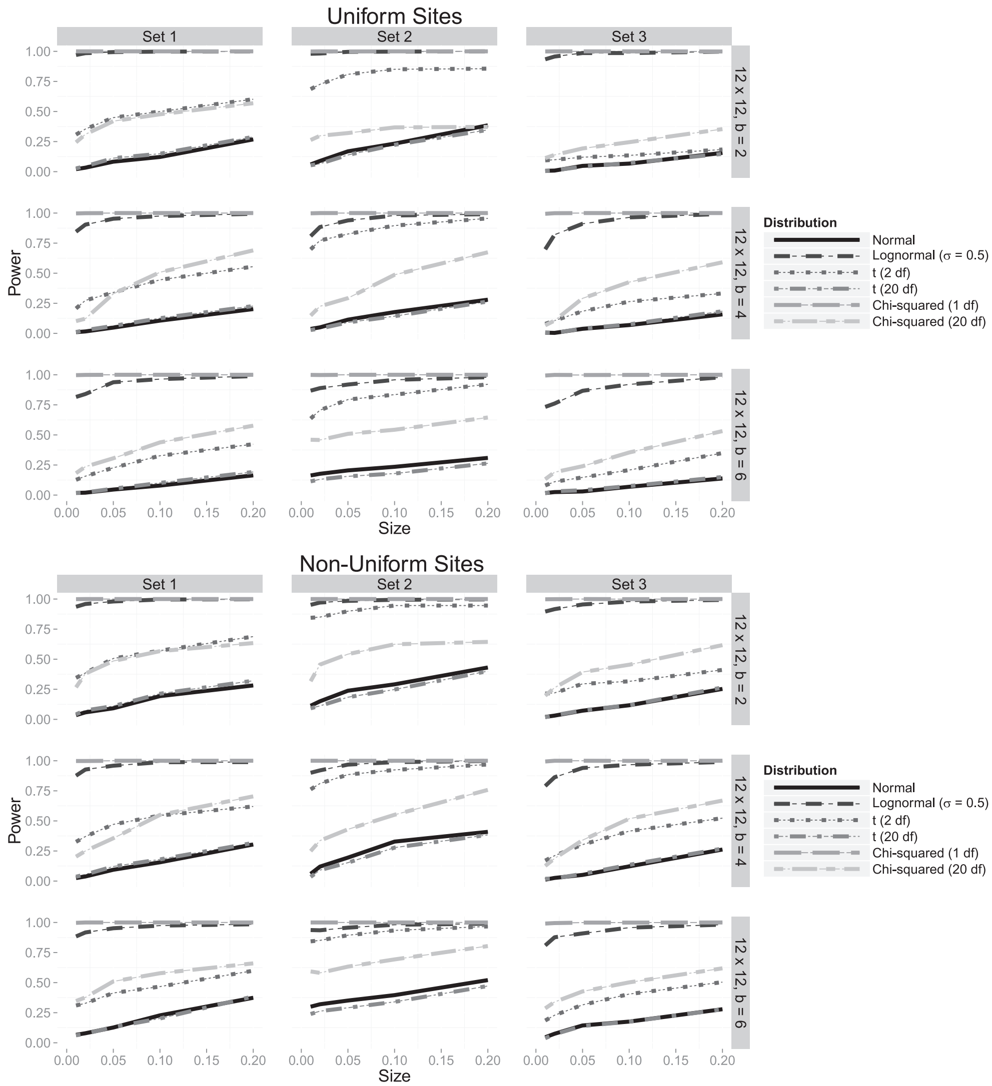

Figure 3. Empirical power functions for SBEL goodness-of-fit tests of normality using three sets of estimating functions and block sizes $b = 2$, 4, and 6 on a 12×12 region, sample size $n = 100$, and uniform and non-uniform locations; data are marginally normal, log-normal, t_2, t_{20}, χ^2_1, and χ^2_{20}.

north and longitude between 89 and 97 degrees west, and Figure 4 maps the region shaded by average January temperature.

We applied the SBEL method to fit a regression model (3.2) with latitude, longitude, and elevation as explanatory variables for temperature. We imposed an 80×80 grid, as the sampling region is 8 degrees longitude by 8 degrees

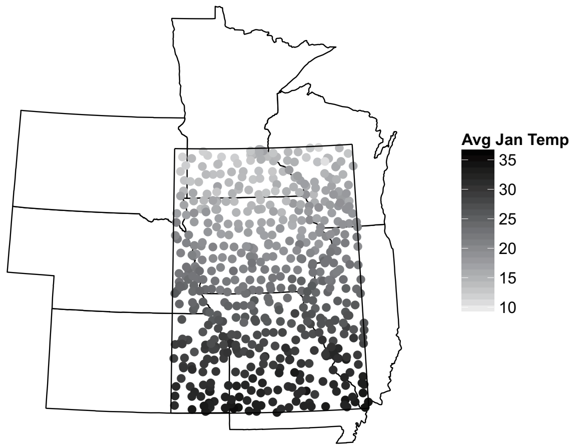

Figure 4. Region for average January temperature data.

latitude with locations measured in tenths of a degree. To choose a block size b, we used the "minimal volatility" technique of Politis, Romano, and Wolf (1999, Sec. 9.3.2), which suggested $b = 27$; see the supplementary web-appendix for details. Table 2 shows subsequent 90% CIs for the regression parameters from the SBEL method as well as intervals based on standard multiple regression with an independence assumption and maximum likelihood estimation using a parametric Gaussian random field model with an exponential covariogram (e.g., based on "likfit" in the geoR package); point estimates from the three methods are also provided. The longitude and intercept estimates for the SBEL method are between those from the multiple regression and Gaussian models, but the SBEL method suggests latitude and elevation have a larger effect on temperature than the other methods. SBEL involves no explicit distributional assumptions about the form of the spatial dependence, and is thereby less sensitive to model misspecification. Consequently, the SBEL method tends to produce the widest CIs, but these are in closer agreement to the parametric model-based intervals than those from the independence assumption (which are comparatively narrow). For instance, the SBEL and Gaussian methods plausibly suggest longitude is not important as a predictor of average temperature. Unlike the other methods, SBEL CIs are not symmetric and, for example, suggest more uncertainty in the lower limit of the elevation parameter. Adding another estimating function based on a third moment condition, a SBEL test for normality of residuals yields a p-value of 0.15.

Table 2. Point estimates and 90% intervals for spatial regression parameters [Longitude (Long), Latitude (Lat), Elevation (Elev), Intercept (Inter)] based on multiple regression (independence assumption), parametric Gaussian maximum likelihood, and SBEL.

	Multiple Regression		Gaussian Likelihood		SBEL	
Long.	-0.011	(-0.016, -0.007)	-0.016	(-0.032, 0.001)	-0.014	(-0.036, 0.011)
Lat.	-0.255	(-0.259, -0.252)	-0.241	(-0.257, -0.225)	-0.270	(-0.284, -0.249)
Elev.	-0.902	(-1.042, -0.763)	-0.943	(-1.100, -0.7850)	-1.038	(-1.790, -0.543)
Inter.	25.59	(25.18 , 26.00)	26.02	(25.27 , 26.76)	25.68	(24.40 , 27.58)

7. Conclusions

We introduced a spatial blockwise empirical likelihood (SBEL) for irregularly located spatial data. Data blocks serve to locally capture the underlying spatial dependence, without explicit assumptions about the data distribution or the general distribution of sampling sites. As a theoretical challenge compared to previous EL for lattice data (Nordman (2008)), the sampling designs considered allow for two fundamental types of spatial asymptotics, determined by the growth rate of the spatial sample size relative to the volume of a sampling region. SBEL applies for both structures with no direct steps of variance estimation, which is an advantage in that large-sample distributional properties of spatial statistics typically change with each asymptotic framework (Lahiri (2003)). Log SBEL-ratios were shown to have chi-square limits for tests and confidence regions of spatial parameters, thus extending block-based EL methods for time series or gridded spatial data (Kitamura (1997); Nordman (2008)). We illustrated SBEL for spatial regression and goodness-of-fit testing, and simulations indicated that SBEL can outperform other resampling procedures. Further research possibilities include extending SBEL for irregularly located spatial data into inference about the correlation structure of a spatial process, which may require local kernel smoothing steps. Such extensions would allow EL tests of spatial isotropy or separability, variogram estimation, and assessments of spatial Markov structures.

Acknowledgements

The authors are grateful to two reviewers and an associate editor for suggestions which greatly clarified and improved the manuscript. Research was partially supported by NSF DMS-0906588.

A1. Proofs

We establish Theorem 1 (spatial EL Wilks theorem) which is a critical component for Theorem 2; remaining proofs are provided in the supplementary web-appendix. We require some additional notation. Let $\mathrm{E}_{\boldsymbol{X}}$ denote expectation with respect to the joint distribution $P_{\boldsymbol{X}}$ of $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ on the probability space $(\Omega, \mathcal{F}, P)$, where both $\{\boldsymbol{X}_i : i \geq 1\}$ and $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ are defined and independent. Let $P_{\cdot|\boldsymbol{X}}$ and $\mathrm{E}_{\cdot|\boldsymbol{X}}$ denote conditional probability and conditional expectation given $\{\boldsymbol{X}_i : i \geq 1\}$. For subset $A \subset \mathbb{R}^d$, let $|A|$ denote the cardinality of A if A is countable and, otherwise, let $|A| = \mathrm{vol}(A)$ denote the volume (Lebesgue measure) of A. In the following, C and $C(\cdot)$ denote generic constants not depending on n, b, or a given realization $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$. For $\theta \in \Theta \subset \mathbb{R}^p$, recall the EL function (3.3) involves sums $S_n(\boldsymbol{i}; \theta)$ over data blocks $B_n(\boldsymbol{i}) = \boldsymbol{i} + b(0,1]^d$, $\boldsymbol{i} \in \mathcal{I}_n \equiv \{\boldsymbol{j} \in \mathbb{Z}^d : B_n(\boldsymbol{j}) \subset R_n\}$, and define

$$A_n(\boldsymbol{i}; \theta) \equiv \lambda_n^{d/2} n^{-1} S_n(\boldsymbol{i}; \theta) = \lambda_n^{d/2} n^{-1} b^{-d} \sum_{j=1}^{n} g(Z(\boldsymbol{s}_j); \theta) \mathbb{I}(\boldsymbol{s}_j \in B_n(\boldsymbol{i})) \quad \text{(A.1)}$$

for $\boldsymbol{i} \in \mathcal{I}_n$ as well as sums of these $A_n(\theta) = \sum_{\boldsymbol{i} \in \mathcal{I}_n} A_n(\boldsymbol{i}; \theta)$.

We require some preliminary results established in Lemmas A.1−A.2 below. Lemma A.1 provides moment bounds on sums of random variables over sampling locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n \in R_n$. Let $m_n = \max\{\log n, \lambda_n^{-d} n\}$, noting that $m_n = \log n$ eventually if $\lambda_n^d / n \to c \in (0, \infty)$ and $m_n = \lambda_n^{-d} n$ eventually if $c = 0$ under Assumption (A4).

Lemma A.1. *Suppose $b^{-1} + b/\lambda_n^{1-\epsilon} = o(1)$ for some $\epsilon \in (0,1)$. Let $h : \mathbb{R}^m \to \mathbb{R}$ be a Borel-measurable function such that $\mathrm{E}\{h(Z(\mathbf{0}))\} = 0$ and that, for some $\delta > 0$ and integer $k \geq 1$, it holds that $\mathrm{E}\{|h(Z(\mathbf{0}))|^{2k+\delta}\} < \infty$ with τ_1, τ_2 as in Assumption* (A1)*. Let $\{w_{in} : 1 \leq i \leq n\}$ be $\sigma\langle \boldsymbol{X}_i : i \geq 1\rangle$-measurable variables with $W_n = \max_{1 \leq i \leq n} |w_{in}| < \infty$. Then, with probability* 1 *(w.p.1)* $(P_{\boldsymbol{X}})$*, as $n \to \infty$,*

(i) $\displaystyle\sum_{\boldsymbol{i} \in \mathcal{I}_n} \mathrm{E}_{\cdot|\boldsymbol{X}} \left\{ \Big| \sum_{j=1}^{n} w_{nj} h(Z(\boldsymbol{s}_j)) \mathbb{I}(\boldsymbol{s}_j \in B_n(\boldsymbol{i})) \Big|^{2k} \right\} = O(\lambda_n^d b^{dk} [W_n m_n]^{2k}).$

(ii) *For any $D_n \subset R_n$ and defining $J_{D_n} = \{\boldsymbol{j} \in \mathbb{Z}^d : D_n \cap (\boldsymbol{j} + [0,1)^d) \neq \emptyset\}$,*

$$\mathrm{E}_{\cdot|\boldsymbol{X}} \left\{ \Big| \sum_{j=1}^{n} w_{nj} h(Z(\boldsymbol{s}_j)) \mathbb{I}(\boldsymbol{s}_j \in D_n) \Big|^{2k} \right\} = O(|J_{D_n}|^k [W_n m_n]^{2k}).$$

Proof of Lemma A.1. This follows by modifying Lemma 2 of Lahiri and Zhu (2006) (setting $\mathcal{K}_n = \{\mathbf{0}\}$ and noting $\mathcal{I}_n \subset \ell_n$ for "$\mathcal{K}_n$ and ℓ_n" in their notation).

The next lemma establishes the behavior of weighed-block sums $A_n(\boldsymbol{i}; \theta_0) \in \mathbb{R}^r$, $\boldsymbol{i} \in \mathcal{I}_n$, from (A.1) at the true parameter $\theta_0 \in \mathbb{R}^r$ where $\mathrm{E}_{\cdot|\boldsymbol{X}}\{A_n(\boldsymbol{i}; \theta_0)\} = 0_r$. Let $\boldsymbol{Z} \sim N(0_r, \Sigma_\infty)$ denote a normal variable in $\mathbb{R}^r$ for $\Sigma_\infty \equiv c\sigma(\mathbf{0}) + \boldsymbol{\Sigma}_0$, with $\boldsymbol{\Sigma}_0$ and $\sigma(\cdot)$ defined in Assumption (A2) and $\lambda_n^d / n \to c \in [0, \infty)$ under Assumption (A4).

Lemma A.2. *Let* $A_n(\theta_0)=\sum_{\boldsymbol{i}\in\mathcal{I}_n}A_n(\boldsymbol{i};\theta_0)$, $\hat{\boldsymbol{\Sigma}}_n(\theta_0)=b^d\sum_{\boldsymbol{i}\in\mathcal{I}_n}A_n(\boldsymbol{i};\theta_0)A_n(\boldsymbol{i};\theta_0)'$, *and* $Z_n(\theta_0)\equiv\max\{\|A_n(\boldsymbol{i};\theta_0)\|:\boldsymbol{i}\in\mathcal{I}_n\}$. *Under the assumptions of Theorem* 1, *the following hold w.p.*1 $(P_{\boldsymbol{X}})$*:* (a) $A_n(\theta_0)\xrightarrow{d}\boldsymbol{Z}$ *in* $P_{\cdot|\boldsymbol{X}}$*-probability;* (b) $(\log n)b^dZ_n(\theta_0)\to 0$ *in* $P_{\cdot|\boldsymbol{X}}$*-probability;* (c)$\|\hat{\Sigma}_n(\theta_0)-\Sigma_\infty\|\to 0$ *in* $P_{\cdot|\boldsymbol{X}}$*-probability;* (d) $P_{\cdot|\boldsymbol{X}}(R_n(\theta_0)>0)\to 1$. (e) *Additionally,* $|\mathcal{I}_n|/|R_n|\to 1$ *where* $|R_n|=\lambda_n^d|R_0|$.

Proof of Lemma A.2. We begin with part (e) involving a non-stochastic sequence. Let $\mathcal{U}=(0,1]^d$ and define $a_n(y)=|\{\boldsymbol{i}\in y\mathbb{Z}^d:(\boldsymbol{i}+y\mathcal{U})\cap R_0\neq\emptyset,(\boldsymbol{i}+y\mathcal{U})\cap R_0^c\neq\emptyset\}|$ for $y>0$. Using the R_0 boundary condition, $|\mathcal{I}_n|\le|R_n|+a_n(\lambda_n^{-1})=|R_n|+O(\lambda_n^{d-1})$. Additionally, $|\mathcal{I}_n|\ge|R_n|-|\{\boldsymbol{i}\in\mathbb{Z}^d:(\boldsymbol{i}+b\mathcal{U})\cap R_n\neq\emptyset,(\boldsymbol{i}+b\mathcal{U})\cap R_n^c\neq\emptyset\}|\ge|R_n|-b^da_n(b/\lambda_n)=|R_n|-O(b\lambda_n^{d-1})$. As $b/\lambda_n\to 0$, we have $|\mathcal{I}_n|/|R_n|\to 1$.

To establish part (a), under the mixing/moment Assumptions (A1)$-$(A4), it follows that $\bar{G}_n(\theta_0)=\lambda_n^{d/2}n^{-1}\sum_{i=1}^n g(Z(\boldsymbol{s}_i);\theta_0)\xrightarrow{d}Z$ w.p.1 $(P_{\boldsymbol{X}})$ by Theorem 3.2 of Lahiri (2003). Then, $\bar{G}_n(\theta_0)-A_n(\theta_0)=\lambda_n^{d/2}n^{-1}\sum_{j=1}^n w_{jn}g(Z(\boldsymbol{s}_j);\theta_0)$ where $w_{jn}=1-b^{-d}\sum_{\boldsymbol{i}\in\mathcal{I}_n}\mathbb{I}(\boldsymbol{s}_j\in B_n(\boldsymbol{i}))\in[0,1]$ for $1\le j\le n$. Let $D_n=\{\boldsymbol{x}\in R_n:(\boldsymbol{x}+b[-1,1)^d)\cap R_n^c\neq\emptyset\}$. If $\boldsymbol{s}_j\in R_n\setminus D_n$ for some $1\le j\le n$, then $w_{jn}=0$. Hence, by Lemma A.1(ii) (with J_{D_n} defined there with respect to D_n),

$$\mathrm{E}_{\cdot|\boldsymbol{X}}\{\|\bar{G}_n(\theta_0)-A_n(\theta_0)\|^2\}=O(\lambda_n^dn^{-2}m_n^2|J_{D_n}|)=o(1)\quad\text{w.p.1 }(P_{\boldsymbol{X}})$$

using $|J_{D_n}|=O(b^da_n(3b/\lambda_n))=O(b\lambda^{d-1})$. Hence, $\bar{G}_n(\theta_0)-A_n(\theta_0)\to 0$ in $P_{\cdot|\boldsymbol{X}}$-probability and so $A_n(\theta_0)\xrightarrow{d}Z$ w.p.1 $(P_{\boldsymbol{X}})$ by Slutsky's theorem.

For part (b), by Jensen's inequality and Lemma A.1(i) with $k=3$, we have

$$\begin{aligned}(\log n)b^d\mathrm{E}_{\cdot|\boldsymbol{X}}\{Z_n(\theta_0)\}&\le(\log n)b^d\Big(\sum_{\boldsymbol{i}\in\mathcal{I}_n}\mathrm{E}_{\cdot|\boldsymbol{X}}\{\|A_n(\boldsymbol{i};\theta_0)\|^6\}\Big)^{1/6}\\&=(\log n)b^d(b^{-d}\lambda^{d/2}n^{-1})\left(\lambda_n^{d/6}b^{d/2}m_n\right)=o(1)\end{aligned}$$

w.p.1 $(P_{\boldsymbol{X}})$ using $|\mathcal{I}_n|=O(\lambda_n^d)$ and the growth assumption $b^2/\lambda_n=O(1)$.

Part (c) of Lemma A.2 follows by modifying the proof of Lahiri and Zhu (2006, Lemma 3). For $\tilde{\Sigma}_n(\theta_0)=\mathrm{E}_{\cdot|\boldsymbol{X}}\{\hat{\Sigma}_n(\theta_0)\}$,

$$\mathrm{E}_{\cdot|\boldsymbol{X}}\{\|\hat{\Sigma}_n(\theta_0)-\tilde{\Sigma}_n(\theta_0)\|^2\}\le O(b^{2d}(b^{-4d}\lambda_n^{2d}n^{-4})\lambda^db^d(m_n)^4b^{2d})=o(1)\text{ w.p.1 }(P_{\boldsymbol{X}})$$

using Lemma A.2(i) with $k=2$ and the re-grouping argument in Lahiri and Zhu (2006, p. 1810). Then, $\mathrm{E}_{\boldsymbol{X}}\{\tilde{\Sigma}_n(\theta_0)\}\to\Sigma_\infty$ by a similar argument in Lahiri and Zhu (2006, p. 1810); namely, $\mathrm{E}_{\boldsymbol{X}}\{\tilde{\Sigma}_n(\theta_0)\}$ here is equivalent to the quantity $b^d|\mathcal{K}_{1n}|^{-1}|\ell_n|\mathrm{E}_{\boldsymbol{X}}\{\tilde{\Sigma}_{1n}\}$ in the notation "$|\mathcal{K}_{1n}|,|\ell_n|,\mathrm{E}_{\boldsymbol{X}}\tilde{\Sigma}_{1n}$" of their proof, where $b^d|\mathcal{K}_{1n}|^{-1}|\ell_n|\to 1$ holds and they show $\mathrm{E}_{\boldsymbol{X}}\{\tilde{\Sigma}_{1n}\}\to\Sigma_\infty$. It then follows that $\tilde{\Sigma}_n(\theta_0)-\mathrm{E}_{\boldsymbol{X}}\{\tilde{\Sigma}_n(\theta_0)\}\to 0$ w.p.1 $(P_{\boldsymbol{X}})$ by writing the difference as a U-statistic of order 2 in $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ as using arguments as in Lahiri (2003, Lemma 5.2). Consequently, $\hat{\Sigma}_n(\theta_0) \to \Sigma_\infty$ in $P_{\cdot|\boldsymbol{X}}$-probability w.p.1 ($P_{\boldsymbol{X}}$), establishing part (c).

To show part (d), note that $R_n(\theta_0) > 0$ holds if 0_r is interior to the convex hull of $\{A_n(\boldsymbol{i};\theta_0) : \boldsymbol{i} \in \mathcal{I}_n\}$. Hence, it suffices to show that the $P_{\cdot|\boldsymbol{X}}$-probability of this latter event converges to 1 (w.p.1 $P_{\boldsymbol{X}}$). The argument uses the supporting/separating hyperplane theorem but is rather involved under the stochastic sampling design. We defer details to the web-appendix.

Proof of Theorem 1. On the common probability space $(\Omega, \mathcal{F}, P)$, there exists $A \in \mathcal{F}$ with $P(A) = 1$ such that all events in Lemma A.2(a)−(d) hold simultaneously conditioned on $\boldsymbol{X}_1 \equiv \boldsymbol{X}_1(\omega), \boldsymbol{X}_2 \equiv \boldsymbol{X}_2(\omega), \ldots$ for any $\omega \in A$. For simplicity, we fix $\omega \in A$ and show distributional convergence of the log-EL ratio, in $P_{\cdot|\boldsymbol{X}}$-probability, conditioned on a pointwise sequence $\{\boldsymbol{X}_i(\omega)\}$; then $P_{\cdot|\boldsymbol{X}}$ is the only probability measure needed in the proof and we let $o_p(\cdot)$ and $O_p(\cdot)$ denote probabilistic orders in $P_{\cdot|\boldsymbol{X}}$-probability.

When $R_n(\theta_0) > 0$, which holds with arbitrarily large $P_{\cdot|\boldsymbol{X}}$-probability for large n by Lemma A.2(d), we may write $R_n(\theta_0) = \prod_{\boldsymbol{i}\in\mathcal{I}_n}(1+\gamma_{\boldsymbol{i},\theta_0})^{-1}$, with $\gamma_{\boldsymbol{i},\theta_0} = \tilde{t}'_{n,\theta_0} A_n(\boldsymbol{i};\theta_0) > 0$, $\boldsymbol{i} \in \mathcal{I}_n$, and a Lagrange multiplier $\tilde{t}_{n,\theta_0} \in \mathbb{R}^r$ fulfilling

$$0_r = \sum_{\boldsymbol{i}\in\mathcal{I}_n} \frac{A_n(\boldsymbol{i};\theta_0)}{1+\gamma_{\boldsymbol{i},\theta_0}} = A_n(\theta_0) - \sum_{\boldsymbol{i}\in\mathcal{I}_n} \frac{A_n(\boldsymbol{i};\theta_0)A_n(\boldsymbol{i};\theta_0)'\tilde{t}_{n,\theta_0}}{1+\gamma_{\boldsymbol{i},\theta_0}} \tag{A.2}$$

for $A_n(\theta_0) = \sum_{\boldsymbol{i}\in\mathcal{I}_n} A_n(\boldsymbol{i};\theta_0)$; $\tilde{t}_{n,\theta_0}$ is related to t_{n,θ_0} mentioned in (3.3) by $\tilde{t}_{n,\theta_0} = \lambda_n^{-d/2} n t_{n,\theta_0}$. Writing $\tilde{t}_{n,\theta_0} = \|\tilde{t}_{n,\theta_0}\| v_n$ for some $v_n \in \mathbb{R}^r$, $\|v_n\| = 1$ and multiplying (A.2) by $-v_n$, we have $\|A_n(\theta_0)\| \geq [1+\|b^{-d}\tilde{t}_{n,\theta_0}\| b^d Z_n(\theta_0)]^{-1} \|b^{-d}\tilde{t}_{n,\theta_0}\| v_n' \hat{\Sigma}_n(\theta_0) v_n$ for $\hat{\boldsymbol{\Sigma}}_n(\theta_0) = b^d \sum_{\boldsymbol{i}\in\mathcal{I}_n} A_n(\boldsymbol{i};\theta_0)A_n(\boldsymbol{i};\theta_0)'$ and $Z_n(\theta_0) \equiv \max\{\|A_n(\boldsymbol{i};\theta_0)\| : \boldsymbol{i} \in \mathcal{I}_n\}$. By Lemma A.2(a)−(c) and letting $\sigma^2_\infty > 0$ denote the smallest eigenvalue of Σ_∞, we then have that $\|A_n(\theta_0)\| \geq \|b^{-d}\tilde{t}_{n,\theta_0}\|[\sigma^2_\infty + o_p(1)]$ holds with arbitrarily large $P_{\cdot|\boldsymbol{X}}$-probability as $n \to \infty$, or that $b^{-d}\tilde{t}_{n,\theta_0} = O_p(1)$. By Lemma A.2(b), we then have that $\max_{\boldsymbol{i}\in\mathcal{I}_n} |\gamma_{\boldsymbol{i},\theta_0}| \leq Z_n(\theta_0)\|\tilde{t}_{n,\theta_0}\| = o_p(1)$. With probability approaching 1 as $n \to \infty$, we may expand (A.2) using Lemma A.2(c) to obtain

$$b^{-d}\tilde{t}_{n,\theta_0} = \hat{\Sigma}_n(\theta_0)^{-1}[A_n(\theta_0) + \beta_n(\theta_0)], \tag{A.3}$$

for $\beta_n(\theta_0) \equiv \sum_{\boldsymbol{i}\in\mathcal{I}_n} \gamma^2_{\boldsymbol{i},\theta_0} A_n(\boldsymbol{i};\theta_0)/(1+\gamma_{\boldsymbol{i},\theta_0})$ and bound

$$\|\beta_n(\theta_0)\| \leq \frac{Z_n(\theta_0)\|\tilde{t}_{n,\theta_0}\|^2 b^{-d}\mathrm{trace}[\hat{\Sigma}_n(\theta_0)]}{1 - \|\tilde{t}_{n,\theta_0}\| Z_n(\theta_0)} = o_p(1).$$

When $\max_{\boldsymbol{i}\in\mathcal{I}_n} |\gamma_{\boldsymbol{i},\theta_0}| \leq Z_n(\theta_0)\|\tilde{t}_{n,\theta_0}\| < 1$, Taylor expansion produces $\log(1 + \gamma_{\boldsymbol{i},\theta_0}) = \gamma_{\boldsymbol{i},\theta_0} - \gamma^2_{\boldsymbol{i},\theta_0}/2 + \eta_{\boldsymbol{i},\theta_0}$ with $|\eta_{\boldsymbol{i},\theta_0}| \leq \|\tilde{t}_{n,\theta_0}\|^3 Z_n(\theta_0)\|A_n(\boldsymbol{i};\theta_0)\|^2/[1 -$

$\|\tilde{t}_{n,\theta_0}\| Z_n(\theta_0)]^3$, $\boldsymbol{i} \in \mathcal{I}_n$. Note that

$$\sum_{\boldsymbol{i} \in \mathcal{I}_n} |\eta_{\boldsymbol{i},\theta_0}| \leq \frac{\|\tilde{t}_{n,\theta_0}\|^3 Z_n(\theta_0) b^{-d} \mathrm{trace}[\hat{\Sigma}_n(\theta_0)]}{[1 - \|\tilde{t}_{n,\theta_0}\| Z_n(\theta_0)]^3} = o_p(b^d).$$

Using this with (A.3) and $\beta_n(\theta_0) = o_p(1)$, we may expand $-2b^{-d} \log R_n(\theta_0) = -2b^{-d} \sum_{\boldsymbol{i} \in \mathcal{I}_n} \log(1 + \gamma_{\boldsymbol{i},\theta_0})$ as

$$\begin{aligned} &A_n(\theta_0)' \hat{\Sigma}_n(\theta_0)^{-1} A_n(\theta_0) - \beta_n(\theta_0)' \hat{\Sigma}_n(\theta_0)^{-1} \beta_n(\theta_0) + b^{-d} 2 \sum_{\boldsymbol{i} \in \mathcal{I}_n} \eta_{\boldsymbol{i},\theta_0} \\ &\quad = A_n(\theta_0)' \hat{\Sigma}_n(\theta_0)^{-1} A_n(\theta_0) + o_p(1). \end{aligned}$$

Lemma A.2(a) and (c) with Slutsky's theorem complete the proof.

References

Bravo, F. (2005). Blockwise empirical entropy tests for time series regressions. *J. Time Series Anal.* **26**, 185-210.

Chen, S. X. and Wong, C. M. (2009). Smoothed block empirical likelihood for quantiles of weakly dependent processes. *Statist. Sinica* **19**, 71-81.

Cressie, N. (1993). *Statistics for Spatial Data*. Wiley, New York.

Doukhan, P. (1994). *Mixing: Properties and Examples.* Lecture Notes in Statistics, **85**. Springer-Verlag, New York.

Hall, P. and Patil, P. (1994). Properties of nonparametric estimators of autocovariance for stationary random fields. *Probab. Theory Related Fields* **99**, 399-424.

Kaiser, M. S. and Nordman, D. J. (2012). Empirical likelihood for spatial model assessment. *Statist. and Its Interface* **5** 303-318.

Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *Ann. Statist.* **25**, 2084-2102.

Kostov, P. (2013). Empirical likelihood estimation of the spatial quantile regression. *J. Geogr. Syst.* **15**, 51-69.

Lahiri, S. N. (2003). Central limit theorems for weighted sums of a spatial process under a class of stochastic and fixed designs. *Sankhyā Ser. A* **65**, 356-388.

Lahiri, S. N. and Zhu, J. (2006). Resampling methods for spatial regression models under a class of stochastic designs. *Ann. Statist.* **34**, 1774-1813.

Matsuda, Y. and Yajima, Y. (2009). Fourier analysis of irregularly spaced data on R^d. *J. Roy. Statist. Soc. Ser. B* **71**, 191-217.

Nordman, D. J. (2008). A blockwise empirical likelihood for spatial lattice data. *Statist. Sinica* **18**, 1111-1129.

Nordman, D. J. and Caragea, P. C. (2008). Point and interval estimation of variogram models using spatial empirical likelihood. *J. Amer. Statist. Assoc.* **103**, 350-361.

Nordman, D. J. and Lahiri, S. N. (2004). On optimal spatial subsample size for variance estimation. *Ann. Statist.* **32**, 1981-2027.

Nordman, D. J. and Lahiri, S. N. (2013). A review of empirical likelihood methods for time series. *J. Statist. Plann. Inference*, invited paper.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237-249.

Owen, A. B. (1990). Empirical likelihood confidence regions. *Ann. Statist.* **18**, 90-120.

Owen, A. B. (1991). Empirical likelihood for linear models. *Ann. Statist.* **19**, 1725-1747.

Politis, D. N., Paparoditis, E. and Romano, J. P. (1998). Large sample inference for irregularly spaced dependent observations based on subsampling. *Sankhyā Ser. A* **60**, 274-292.

Politis, D. N. and Sherman, M. (2001). Moment estimation for statistics from marked point processes. *J. Roy. Statist. Soc. Ser. B* **63**, 261-275.

Politis, D. N., Romano, J. P. and Wolf, M. (1999). *Subsampling*. Springer-Verlag, New York.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**, 300-325.

Sherman, M. (1996). Variance estimation for statistics computed from spatial lattice data. *J. Roy. Statist. Soc. Ser. B* **58**, 509-523.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.* **9**, 60-62.

Wu, R. and Cao, J. (2011). Blockwise empirical likelihood for time series of counts. *J. Multivariate Anal.* **102**, 661-673.

Department of Statistics, Iowa State University, Ames, IA 50011, U.S.A.

E-mail: mvanhala@iastate.edu

Department of Statistics, Iowa State University, Ames, IA 50011, U.S.A.

E-mail: dnordman@iastate.edu

Department of Statistics, Iowa State University, Ames, IA 50011, U.S.A.

E-mail: zhuz@iastate.edu

(Received December 2013; accepted September 2014)