Statistica Sinica **18**(2008), 1111-1129

A BLOCKWISE EMPIRICAL LIKELIHOOD FOR SPATIAL LATTICE DATA

Daniel J. Nordman

Iowa State University

Abstract: This article considers an empirical likelihood method for data located on a spatial grid. The method allows inference on spatial parameters, such as means and variograms, without knowledge of the underlying spatial dependence structure. Log-likelihood ratios are shown to have chi-square limits under spatial dependence for calibrating tests and confidence regions, and maximum empirical likelihood estimators permit parameter estimation and testing of spatial moment conditions. A practical Bartlett correction is proposed to improve the coverage accuracy of confidence regions. The spatial empirical likelihood method is investigated through a simulation study and illustrated with a data example.

Key words and phrases: Data blocking, discrete index random fields, estimating equations.

1. Introduction

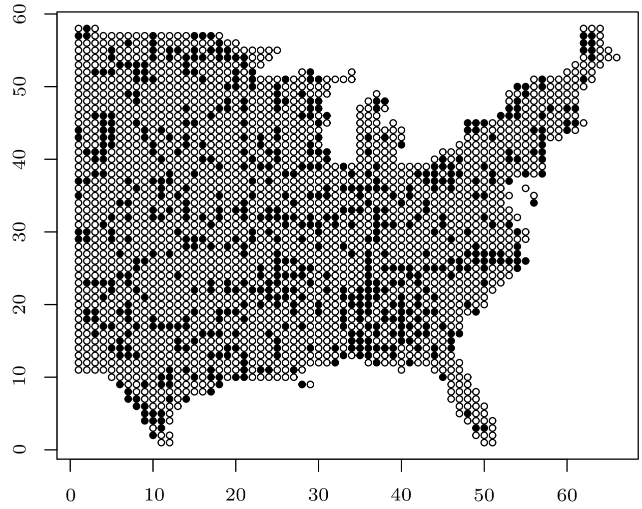
Empirical likelihood (EL), introduced by Owen (1988, 1990), is a statistical method allowing likelihood-based inference without requiring a fully specified parametric model for the data. For independent data, versions of EL are known to share many qualities associated with parametric likelihood, such as limiting chi-square distributions for log-likelihood ratios; see Owen (1988) for means, Hall and La Scala (1990) for smooth mean functions and Qin and Lawless (1994) for parameters satisfying moment restrictions. More recently, attention has focused on formulating EL for dependent time series. For weakly dependent time series, Kitamura (1997) proposed a general EL method based on data blocking techniques, and related "blockwise" versions of EL have been developed for other time series inference: Lin and Zhang (2001) for blockwise Euclidean EL; Chuang and Chan (2002) for autoregressive models; Chen, Härdle and Li (2003) for goodness-of-fit tests; Bravo (2005) for time series regressions; Zhang (2006) for negatively associated series. In econometrics, much research has considered EL for testing moment restrictions and comparisons between EL and generalized method of moments estimators; see, for example, Kitamura, Tripathi and Ahn (2004) and Newey and Smith (2004). Monti (1997) and Nordman and Lahiri

(2006) have considered periodogram-based EL inference for short- and long-memory time series, respectively.

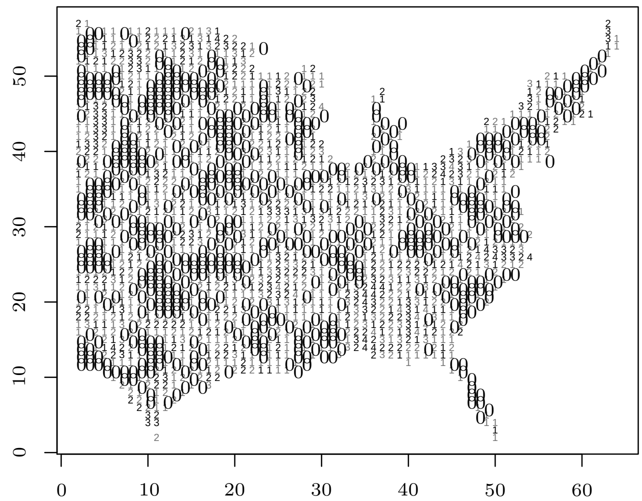
In contrast to time series, the potential application of EL for spatially dependent data has received little consideration. The aim of this paper is to propose an EL method for spatial lattice data and demonstrate that it has some important inference properties in the spatial setting. The method has nonparametric and semiparametric uses and is valid for many spatial processes under weak conditions; this can be appealing when there is uncertainty about an appropriate parametric model. Spatial EL provides a general framework for inference on many spatial parameters through a likelihood function based on estimating equations. Applying the EL method to different spatial problems requires only adjusting the estimating functions that describe the inference scenario. In addition, the spatial EL method does *not* require variance estimation steps to set confidence regions or conduct tests. This feature of spatial EL is particularly important because standard errors can be difficult to obtain for many spatial statistics under an unknown spatial dependence structure. Current nonparametric methods for spatial data, such as spatial subsampling and the spatial block bootstrap, often require direct estimation of the variance of spatial statistics under data dependence (see Sherman and Carlstein (1994), Politis, Romano and Wolf (1999), Lahiri (2003), and references therein).

An example of a situation where spatial EL provides an attractive approach is illustrated in Figure 1(a), which presents a map of high and low cancer mortality rates for the United States. High and low mortality are defined as in Sherman and Carlstein (1994), who fit an autologistic model to assess evidence of clustering among high mortality cases. To estimate the autologistic parameter that describes clustering, these authors employed maximum pseudo-likelihood (Besag (1975)) followed by a spatial subsampling step. In particular, subsampling was used to obtain a standard error for the pseudo-likelihood estimate in order to set a confidence interval for the autologistic parameter through a normal approximation. This example is revisited in Section 6, where the spatial EL method produces a confidence interval for the clustering model parameter automatically, and no separate determination of standard error is required. Intervals from the spatial EL approach indicate spatial clustering, but suggest the evidence for clustering is not as strong as reported by Sherman and Carlstein (1994).

In what follows, a spatial blockwise EL method is developed, based on spatial estimating equations combined with either maximally overlapping or non-overlapping blocks of spatial observations. Data blocking is used as a device to accommodate unknown spatial dependence, similar to the time series blockwise EL of Kitamura (1997). For a broad class of spatial parameters, the spatial EL method yields log-ratios that are asymptotically chi-square, allowing the formulation of tests and confidence regions without knowledge of the data dependence



(a)



(b)

Figure 1. (a) Cancer mortality map, where • and ∘ respectively denote a high, $Z_{\mathbf{s}} = 1$, or low, $Z_{\mathbf{s}} = 0$, mortality rate at site $\mathbf{s} \in \mathcal{R}_n \cap \mathbb{Z}^2$ of the sampling region $\mathcal{R}_n$. (b) Sampling region $\mathcal{R}_{5,n}$ for vectors $Y_{\mathbf{s}}$, $\mathbf{s} \in \mathcal{R}_{5,n} \cap \mathbb{Z}^2$, where $Y_{\mathbf{s}}$ consists of $Z_{\mathbf{s}}$ and its four nearest neighbors $Z_{\mathbf{h}}$, $\|\mathbf{s} - \mathbf{h}\| = 1$; at each site $\mathbf{s} \in \mathcal{R}_{5,n} \cap \mathbb{Z}^2$, the indicated value denotes the sum $S_{\mathbf{s}}$ of the four neighboring indicators $Z_{\mathbf{h}}$ of $Z_{\mathbf{s}}$ where values in dark (light) font denote $Z_{\mathbf{s}} = 1$ ($Z_{\mathbf{s}} = 0$) at site $\mathbf{s}$.

structure. Our EL results include distributions for EL point estimators of spatial parameters as well (i.e., so-called maximum EL estimators). Based on recent results in Chen and Cui (2006, 2007) for independent data, a procedure for a practical Bartlett correction for the spatial EL method is proposed and investigated using simulation. The Bartlett correction makes an adjustment to the log EL ratio that improves coverage accuracy.

The rest of the paper is organized as follows. In Section 2, we describe the spatial sampling and estimating function frameworks, with some examples provided for illustration. We also construct the spatial blockwise EL. The main distributional results for the spatial EL method are presented in Section 3. Section 4 outlines an empirical Bartlett correction. The proposed methodology is assessed through a numerical study in Section 5, and illustrated with the cancer mortality map of the United States in Section 6. Section 7 provides a discussion of EL block selection. Assumptions and detailed proofs for the main results are deferred to an Appendix, available in the online supplement to this manuscript.

2. Spatial Empirical Likelihood Method

To set the stage for development of spatial EL, recall the formulation of EL using a sample $Y_1, \ldots, Y_n$ of independent, identically distributed (i.i.d.) data (e.g., Owen (1990) and Qin and Lawless (1994)). First, a parameter of interest $\theta \in \mathbb{R}^p$ is linked to each observation by creating a function $G_\theta(Y_i)$ of both, using a vector of $r \geq p$ estimating functions $G_\theta(\cdot)$. The estimating functions are chosen so that, at the true parameter value $\theta = \theta_0$, we have an expectation condition $\mathrm{E}\{G_{\theta_0}(Y_i)\} = 0_r$ that identifies θ_0. With such estimating functions in place, an EL function for θ can be constructed by maximizing a product of n probabilities placed on $G_\theta(Y_1), \ldots, G_\theta(Y_n)$ under a linear "expectation" constraint. The resulting EL function for θ has important uses: the function can be maximized for point estimators for θ, or chi-square calibrated to set confidence regions.

An EL for spatial lattice data is similarly based on estimating functions that satisfy a moment condition, but requires modifications to handle spatial dependence. First, we need a spatial sampling region $\mathcal{R}_n \subset \mathbb{R}^d$, $d \geq 1$, on which a spatial process $\{Z_{\mathbf{s}} : \mathbf{s} \in \mathbb{Z}^d\}$ is observed on a grid; here d denotes the dimension of sampling. Then we develop estimating functions involving a spatial parameter θ of interest and the spatial $Z_{\mathbf{s}}$-observations. To provide more generality in the spatial setting, we consider functions $G_\theta(Y_{\mathbf{s}})$, $\mathbf{s} \in \mathbb{Z}^d$, that connect θ to vectors of spatial observations $Y_{\mathbf{s}} = (Z_{\mathbf{s}+\mathbf{h}_1}, \ldots, Z_{\mathbf{s}+\mathbf{h}_m})'$ based on some selection of fixed spatial lags $\mathbf{h}_1, \ldots, \mathbf{h}_m \in \mathbb{Z}^d$; these $Y_{\mathbf{s}}$-observations have their own sampling region $\mathcal{R}_{n,Y}$ based on the region $\mathcal{R}_n$ for the observed spatial process $\{Z_{\mathbf{s}} : \mathbf{s} \in \mathbb{Z}^d\}$. These formulations are made precise in Section 2.1, which also provides some examples. A spatial EL function for θ is then constructed using an estimating function $G_\theta(\cdot)$, along with spatial blocks of $Y_{\mathbf{s}}$-observations, instead of using individual observations, as described in Section 2.2.

For clarity throughout the sequel, a bold font denotes a vector in $\mathbb{R}^d$, e.g., $\mathbf{s}$, $\mathbf{h}$, $\mathbf{i} \in \mathbb{R}^d$.

2.1. Spatial estimating equations

To describe the spatial EL method, we adopt a sampling framework that allows a spatial sampling region $\mathcal{R}_n \subset \mathbb{R}^d$ to grow as the sample size n increases. Using a subset $\mathcal{R}_0 \subset (-1/2, 1/2]^d$ containing an open neighborhood and an increasing positive sequence $\{\lambda_n\}$ of scaling factors, suppose the sampling region $\mathcal{R}_n$ is obtained by inflating the "template" set $\mathcal{R}_0$ by the constant λ_n: $\mathcal{R}_n = \lambda_n \mathcal{R}_0$. This formulation permits a wide variety of shapes for the sampling region $\mathcal{R}_n$, which shape is preserved as the sampling region grows. For spatial subsampling, Sherman and Carlstein (1994), Sherman (1996), and Nordman and Lahiri (2004) use a comparable sampling structure. We assume that a real-valued, strictly stationary process $\{Z_{\mathbf{s}} : \mathbf{s} \in \mathbb{Z}^d\}$ is observed at regular locations on the grid $\mathbb{Z}^d$ inside $\mathcal{R}_n$. Hence, the available data are $\{Z_{\mathbf{s}} : \mathbf{s} \in \mathcal{R}_n \cap \mathbb{Z}^d\}$ observed at n sites $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\} = \mathcal{R}_n \cap \mathbb{Z}^d$, with n as the sample size of the observed $Z_{\mathbf{s}}$.

To describe a finite dimensional parameter $\theta \in \Theta \subset \mathbb{R}^p$ of the spatial process $\{Z_{\mathbf{s}} : \mathbf{s} \in \mathbb{Z}^d\}$ with estimating functions, we collect observations from $\mathcal{R}_n$ into vectors. For a positive integer m, we form an m-dimensional vector $Y_{\mathbf{s}} = (Z_{\mathbf{s}+\mathbf{h}_1}, Z_{\mathbf{s}+\mathbf{h}_2}, \ldots, Z_{\mathbf{s}+\mathbf{h}_m})'$, $\mathbf{s} \in \mathcal{R}_{n,Y} \cap \mathbb{Z}^d$, where $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_m \in \mathbb{Z}^d$ are selected lag vectors, and $\mathcal{R}_{n,Y} = \{\mathbf{s} \in \mathcal{R}_n : \mathbf{s} + \mathbf{h}_1, \ldots, \mathbf{s} + \mathbf{h}_m \in \mathcal{R}_n\}$ denotes the sampling region for the process $\{Y_{\mathbf{s}} : \mathbf{s} \in \mathbb{Z}^d\}$ containing $n_Y \equiv |\mathcal{R}_{n,Y} \cap \mathbb{Z}^d|$ observations. Here and throughout the sequel, $|A|$ represents the size of a finite set A.

As in the i.i.d. data formulation of EL (Qin and Lawless (1994)), suppose information about $\theta \in \Theta \subset \mathbb{R}^p$ exists through $r \geq p$ estimating functions linking θ to a vector form $Y_{\mathbf{s}}, \mathbf{s} \in \mathbb{Z}^d$ of the spatial process $\{Z_{\mathbf{s}} : \mathbf{s} \in \mathbb{Z}^d\}$. With arguments $y = (y_1, \ldots, y_m)' \in \mathbb{R}^m$ and $\theta \in \Theta$, define $G_\theta(y) = (g_{1,\theta}(y), \ldots, g_{r,\theta}(y))' : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}^r$ as a vector of r estimating functions satisfying

$$\mathrm{E}\left\{G_{\theta_0}(Y_{\mathbf{s}})\right\} = 0_r \in \mathbb{R}^r, \quad \mathbf{s} \in \mathbb{Z}^d, \tag{1}$$

at the true and unique parameter value θ_0. When $r > p$, the above functions are said to be "overidentifying" for θ. In Section 2.2, we build an EL function for a spatial parameter θ via the moment condition in (1).

With appropriate choices of vectors $Y_{\mathbf{s}}$ and estimating functions $G_\theta(\cdot)$, EL inference is possible for a large class of spatial parameters, as is illustrated in the following examples.

Example 1. (Poisson counts). Consider a pattern of events in a spatial region that may exhibit spatial randomness (e.g., tree locations in a forest). It is common to partition the region into rectangular plots on a grid, and the number of events occurring in each plot (or quadrat) is considered as a lattice observation $Z_{\mathbf{s}}, \mathbf{s} \in \mathbb{Z}^d$ (e.g., counts of trees in a quadrat), where each count $Z_{\mathbf{s}}$ follows a Poisson distribution with mean $\mathrm{E}\,(Z_{\mathbf{s}}) = \theta$ when the events exhibit complete spatial randomness (Cressie (1993, Chap. 8.2)). For EL inference, we set $Y_{\mathbf{s}} = Z_{\mathbf{s}}$, $\mathbf{s} \in \mathbb{Z}^d$, and use estimating functions $G_\theta(Y_{\mathbf{s}}) = (Z_{\mathbf{s}} - \theta, Z_{\mathbf{s}}^2 - \theta^2 - \theta)'$, based on Poisson moments, so that (1) holds with $p = m = 1, r = 2$. Using EL results in Section 3, it is possible to estimate the mean count θ, or more importantly test if the Poisson assumption (1) holds, without nonparametric variance estimation as used in some previous applications (Sherman (1996)).

Example 2. (Variogram inference). Estimation of the variogram $2\gamma(\mathbf{h}_i) \equiv \mathrm{Var}\,(Z_{\mathbf{s}} - Z_{\mathbf{s}+\mathbf{h}_i}) = \mathrm{E}\,\{(Z_{\mathbf{s}} - Z_{\mathbf{s}+\mathbf{h}_i})^2\}$ of the process $\{Z_{\mathbf{s}} : \mathbf{s} \in \mathbb{Z}^d\}$ at given lags $\mathbf{h}_1, \ldots, \mathbf{h}_p \in \mathbb{Z}^d$ is an important problem. Least squares variogram fitting is commonly proposed in the geostatistical literature; see Lee and Lahiri (2002) and references therein. For EL inference on the variogram $\theta = (2\gamma(\mathbf{h}_1), \ldots, 2\gamma(\mathbf{h}_p))' \in \mathbb{R}^p$, we define a vector function $G_\theta(Y_{\mathbf{s}}) = (g_{1,\theta}(Y_{\mathbf{s}}), \ldots, g_{p,\theta}(Y_{\mathbf{s}}))'$ of the $(p+1)$-dimensional process $Y_{\mathbf{s}} = (Z_{\mathbf{s}}, Z_{\mathbf{s}+\mathbf{h}_1}, \ldots, Z_{\mathbf{s}+\mathbf{h}_p})'$, where $g_{i,\theta}(Y_{\mathbf{s}}) = (Z_{\mathbf{s}} - Z_{\mathbf{s}+\mathbf{h}_i})^2 - 2\gamma(\mathbf{h}_i)$. This selection fulfills (1) with $r = p$, $m = p + 1$.

Example 3. (Pseudo-likelihood inference). Markov random fields provide an important class of models for spatial lattice data. They allow the conditional distribution of an observation $Z_{\mathbf{s}}$, $\mathbf{s} \in \mathbb{Z}^d$, to be written through a neighborhood structure as

$$f_\theta(z \mid \{Z_{\mathbf{h}} : \mathbf{h} \neq \mathbf{s}\}) = \begin{cases} P_\theta(Z_{\mathbf{s}} = z \mid \{Z_{\mathbf{h}} : \mathbf{h} \in N_{\mathbf{s}}\}) & \text{if } Z_{\mathbf{s}} \text{ is discrete} \\ \text{density } f_\theta(z \mid \{Z_{\mathbf{h}} : \mathbf{h} \in N_{\mathbf{s}}\}) & \text{if } Z_{\mathbf{s}} \text{ is continuous,} \end{cases} \quad z \in \mathbb{R}, \tag{2}$$

where $N_{\mathbf{s}} \subset \mathbb{Z}^d$ denotes a neighborhood of $Z_{\mathbf{s}}$ (Cressie (1993, Chap. 6). Besag (1974) developed models based on conditional distributions from one-parameter exponential families in (2) and estimated them through maximum pseudo-likelihood (Besag (1975)), where the pseudo-likelihood estimator $\hat{\theta}_n^{PL}$ of $\theta \in \Theta \subset \mathbb{R}^p$ solves the score-based system

$$\sum_{\mathbf{s} \in \mathcal{R}_n \cap \mathbb{Z}^d} \frac{\partial \log f_\theta(Z_{\mathbf{s}} \mid \{Z_{\mathbf{h}} : \mathbf{h} \in N_{\mathbf{s}}\})}{\partial \theta} = 0_p \in \mathbb{R}^p.$$

Confidence regions for θ based on a normal approximation for $\hat{\theta}_n^{PL}$ often require estimating the variance Var $(\hat{\theta}_n^{PL})$ of the pseudo-likelihood estimator, a difficult task in general. This issue is relevant when fitting (2) with pseudo-likelihood to examine clustering in the mortality map described in the Introduction. However, the EL method may be generally applied for pseudo-likelihood inference with the advantage that a confidence region for a parameter θ characterizing (2) can be set by simply calibrating an EL function. This will be illustrated for the mortality map example in Section 6.

To set up EL inference for a conditional distribution (2), suppose the neighborhoods $N_{\mathbf{s}}$, $\mathbf{s} \in \mathbb{Z}^d$, have a constant structure, such as "four-nearest neighbor" $N_{\mathbf{s}} \equiv \{\mathbf{s} \pm \mathbf{e} \in \mathbb{Z}^2 : \mathbf{e} = (0,1)', (1,0)'\}$ when $d = 2$. For describing $\theta \in \Theta \subset \mathbb{R}^p$, we choose $r = p$ score-functions $G_\theta(Y_{\mathbf{s}}) = \partial \log f_\theta(Z_{\mathbf{s}} \mid \{Z_{\mathbf{h}} : \mathbf{h} \in N_{\mathbf{s}}\})/\partial\theta$ involving a vector $Y_{\mathbf{s}} = (Z_{\mathbf{s}}, Z_{\mathbf{s}+\mathbf{h}_1}, \ldots, Z_{\mathbf{s}+\mathbf{h}_{|N_{\mathbf{s}}|}})'$, $\mathbf{h}_i \in N_{\mathbf{s}} - \mathbf{s}$, formed by $Z_{\mathbf{s}}$ and its $|N_{\mathbf{s}}|$ neighbors, $\mathbf{s} \in \mathbb{Z}^d$. For Markov random fields based on exponential-family models (2), these functions entail the moment condition (1) for θ.

2.2. Spatial blockwise empirical likelihood construction

Suppose a spatial parameter $\theta \in \Theta \subset \mathbb{R}^p$ is identified through a vector process $Y_{\mathbf{s}}$, $\mathbf{s} \in \mathbb{Z}^d$, and estimating functions $G_\theta(\cdot)$ satisfying (1). Construction of the spatial EL function for θ requires spatial blocks of observed vectors $Y_{\mathbf{s}}$, $\mathbf{s} \in \mathcal{R}_{n,Y} \cap \mathbb{Z}^d$. We consider two possible sources of rectangular blocks within $\mathcal{R}_{n,Y}$, namely, maximally overlapping (OL) and non-overlapping (NOL) blocks. Such blocking schemes are common with other block-based spatial resampling methods, such as the spatial block bootstrap and spatial subsampling (Lahiri (2003)).

Let $\{b_n\}_{n\geq 1}$ be a sequence of positive integers and define general d-dimensional blocks as $\mathcal{B}_{b_n}(\mathbf{i}) \equiv \mathbf{i} + b_n\mathcal{U}$, $\mathbf{i} \in \mathbb{Z}^d$, using the cube $\mathcal{U} = (-1/2, 1/2]^d$. To keep the blocks small relative to the sampling region $\mathcal{R}_{n,Y}$, we suppose b_n grows at a slower rate than the sample size n_Y, and require that

$$b_n^{-1} + \frac{b_n^{2d}}{n_Y} = o(1) \tag{3}$$

as $n \to \infty$. We elaborate on this block condition in Section 7. The integer index set $\mathcal{I}_{b_n}^{OL} = \{\mathbf{i} \in \mathbb{Z}^d : \mathcal{B}_{b_n}(\mathbf{i}) \subset \mathcal{R}_{n,Y}\}$ identifies all integer-translated cubes $b_n\mathcal{U}$ lying completely inside the sampling region $\mathcal{R}_{n,Y}$ for the $Y_{\mathbf{s}}$-observations. From this, the collection of maximally OL blocks is given by $\{\mathcal{B}_{b_n}(\mathbf{i}) : \mathbf{i} \in \mathcal{I}_{b_n}^{OL}\}$; see Figure 2(c). For NOL blocks, the region $\mathcal{R}_{n,Y}$ is divided into disjoint cubes of $Y_{\mathbf{s}}$-observations. Letting $\mathcal{I}_{b_n}^{NOL} = \{b_n\mathbf{k} : \mathbf{k} \in \mathbb{Z}^d, \mathcal{B}_{b_n}(b_n\mathbf{k}) \subset \mathcal{R}_{n,Y}\} \subset \mathbb{Z}^d$ represent the index set of all NOL cubes $\mathcal{B}_{b_n}(b_n\mathbf{k}) = b_n(\mathbf{k} + \mathcal{U})$ lying completely inside $\mathcal{R}_{n,Y}$, the NOL block collection is then $\{\mathcal{B}_{b_n}(\mathbf{i}) : \mathbf{i} \in \mathcal{I}_{b_n}^{NOL}\}$; see Figure 2(b).

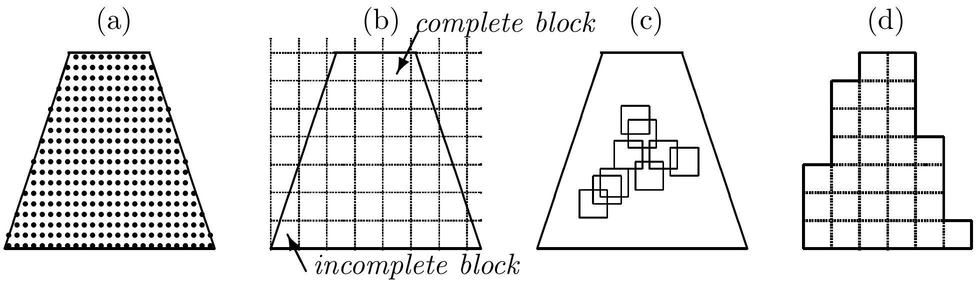


Figure 2. (a) Sampling region $\mathcal{R}_{n,Y}$; (b) NOL complete blocks; (c) OL blocks; (d) Bootstrap region $\mathcal{R}^*_{n,Y}$ formed by the complete blocks in (b). (Bootstrap samples on $\mathcal{R}^*_{n,Y}$ are found by resampling data blocks from (c) and concatenating these into block positions in (d).)

In the following, we let $\mathcal{I}_n$ generically denote either chosen index set $\mathcal{I}^{OL}_{b_n}$ or $\mathcal{I}^{NOL}_{b_n}$ and denote the number of blocks as $N_{\mathcal{I}} = |\mathcal{I}_n|$. Using estimating functions G_θ in (1), we compute a sample mean $M_{\theta,\mathbf{i}} = \sum_{\mathbf{s}\in\mathcal{B}_{b_n}(\mathbf{i})\cap\mathbb{Z}^d} G_\theta(Y_{\mathbf{s}})/b_n^d$, $\mathbf{i}\in\mathcal{I}_n$, for each block $\mathcal{B}_{b_n}(\mathbf{i})$ in the collection, which provides $|\mathcal{B}_{b_n}(\mathbf{i})\cap\mathbb{Z}^d| = b_n^d$ observations of $G_\theta(Y_{\mathbf{s}})$, $\mathbf{s}\in\mathbb{Z}^d$. The EL function $L_n(\theta)$ and EL ratio $R_n(\theta)$ for $\theta\in\Theta$ are then given by

$$L_n(\theta)=\sup\left\{\prod_{\mathbf{i}\in\mathcal{I}_n} p_{\mathbf{i}} \,:\, p_{\mathbf{i}}\geq 0,\ \sum_{\mathbf{i}\in\mathcal{I}_n} p_{\mathbf{i}}=1,\ \sum_{\mathbf{i}\in\mathcal{I}_n} p_{\mathbf{i}} M_{\theta,\mathbf{i}}=0_r\right\},\ R_n(\theta)=\frac{L_n(\theta)}{(\frac{1}{N_{\mathcal{I}}})^{N_{\mathcal{I}}}}. \tag{4}$$

The EL function for $\theta\in\Theta$ involves maximizing a multinomial likelihood created from probabilities assigned to each block sample mean, under an expectation-based linear constraint. Without the expectation constraint in $L_n(\theta)$, the product has a maximum when each $p_{\mathbf{i}} = 1/N_{\mathcal{I}}$, yielding the EL ratio in (4). If $0_r\in\mathbb{R}^r$ is interior to the convex hull of $\{M_{\theta,\mathbf{i}} : \mathbf{i}\in\mathcal{I}_n\}$, then $L_n(\theta)$ represents a positive, constrained maximum and (4) may be written as

$$L_n(\theta)=\prod_{\mathbf{i}\in\mathcal{I}_n} p_{\theta,\mathbf{i}},\ R_n(\theta)=\prod_{\mathbf{i}\in\mathcal{I}_n}\left(1+t_\theta' M_{\theta,\mathbf{i}}\right)^{-1},\quad p_{\theta,\mathbf{i}}=\{N_{\mathcal{I}}(1+t_\theta' M_{\theta,\mathbf{i}})\}^{-1}\in(0,1), \tag{5}$$

where t_θ solves $\sum_{\mathbf{i}\in\mathcal{I}_n} M_{\theta,\mathbf{i}}/(1+t'M_{\theta,\mathbf{i}}) = 0_r$. We define $L_n(\theta) = -\infty$ when the set in (4) is empty. See Owen (1990) and Qin and Lawless (1994) for these computational details on EL.

In the next section, we consider the distribution of the log EL ratio given by

$$\ell_n(\theta) = -2B_n \log R_n(\theta), \qquad B_n = \frac{n_Y}{b_n^d N_{\mathcal{I}}}. \tag{6}$$

The factor B_n is a block adjustment to ensure chi-square limits for (6), and represents the spatial analog of the block correction used for the time series blockwise EL (Kitamura (1997)).

3. Main Results

Distributional results for the spatial EL are established under a set of assumptions referred to as "Assumptions 1-4" in the sequel. We defer technical details on these assumptions to Section 8.1 of the online Appendix. In brief, Assumption 1 provides a condition equivalent to the block growth rate (3). Assumptions 2-4 describe spatial mixing and moment conditions which allow the spatial EL method to be valid for a large class of spatial processes exhibiting weak spatial dependence. All of the EL results to follow apply equally to EL functions $R_n(\theta)$, $\ell_n(\theta)$ constructed of either OL or NOL blocks (i.e., $\mathcal{I}_n = \mathcal{I}_{b_n}^{OL}$ or $\mathcal{I}_{b_n}^{NOL}$).

3.1. Smooth function model

We first establish the distribution of spatial blockwise EL ratios for inference on "smooth function" parameters, as in Hall and La Scala (1990) for i.i.d. data and Kitamura (1997) for mixing time series. Suppose $\theta = \mathrm{E}\{G(Y_{\mathbf{s}})\} \in \Theta \subset \mathbb{R}^p$ represents the mean of a function $G : \mathbb{R}^m \to \mathbb{R}^p$ applied to an m-dimensional vector $Y_{\mathbf{s}}$, $\mathbf{s} \in \mathbb{Z}^d$. EL inference on a more general parameter $\theta^H = H(\theta) \in \mathbb{R}^u$ may be considered using a smooth function $H : \mathbb{R}^p \to \mathbb{R}^u$ of θ. This "smooth function" model permits a wide range of spatial parameters θ^H, including ratios or differences of means θ. For example, $\theta = \{\mathrm{E}\,(Y_{\mathbf{s}}), \mathrm{E}\,(Y_{\mathbf{s}}^2), \mathrm{E}\,(Y_{\mathbf{s}}Y_{\mathbf{s+h}})\}' \in \mathbb{R}^3$ and $H(x_1, x_2, x_3) = (x_3 - x_1^2)/(x_2 - x_1^2) : \mathbb{R}^3 \to \mathbb{R}$ yield a spatial autocorrelation $\theta^H = H(\theta)$ at lag $\mathbf{h} \in \mathbb{Z}^d$. For smooth model inference, we first define an EL ratio $R_n(\theta)$ for θ using functions $G_\theta(Y_{\mathbf{s}}) = G(Y_{\mathbf{s}}) - \theta$, $\mathbf{s} \in \mathbb{Z}^d$ in (5), which satisfy (1) with the same number of parameters and estimating functions $r = p$. An EL ratio and log-ratio for a parameter θ^H are then defined as

$$R_n(\theta^H) \equiv \sup_{\theta \in \Theta : H(\theta) = \theta^H} R_n(\theta), \qquad \ell_n(\theta^H) \equiv -2B_n \log R_n(\theta^H).$$

Theorem 1 provides a nonparametric recasting of Wilks' theorem for spatial data, useful for calibrating confidence regions and tests of spatial "smooth model" parameters based on a chi-square approximation. In the following, χ_ν^2 denotes a chi-square variable with ν degrees of freedom with a lower α-quantile given by $\chi_{\nu;\alpha}^2$ and $\xrightarrow{d}$ denotes distributional convergence.

Theorem 1. (Smooth functions of means) *Suppose* $\mathcal{I}_n = \mathcal{I}_{b_n}^{OL}$ *or* $\mathcal{I}_{b_n}^{NOL}$*;* $\mathrm{E}\{G(Y_{\mathbf{s}})\} = \theta \in \mathbb{R}^p$*; Assumptions* 1-4 *hold with* $r = p$ *estimating functions*

$G_\theta(Y_{\mathbf{s}}) = G(Y_{\mathbf{s}}) - \theta$, $\mathbf{s} \in \mathbb{Z}^d$; $H : \mathbb{R}^p \to \mathbb{R}^u$ is continuously differentiable in a neighborhood of θ_0 and $\theta_0^H = H(\theta_0)$. Then,

$$\ell_n(\theta_0^H) \xrightarrow{d} \chi^2_\nu$$

as $n \to \infty$, where ν denotes the rank of the $u \times p$ matrix $\partial H(\theta)/\partial\theta|_{\theta=\theta_0}$.

See Hall and La Scala (1990) for properties of EL confidence regions for smooth model parameters.

3.2. Maximum empirical likelihood point estimation

We refer to the maximum of $R_n(\theta)$ from (5) as the maximum empirical likelihood estimator (MELE) and denote it by $\hat{\theta}_n$. Using general estimating equations, Qin and Lawless (1994) and Kitamura (1997) considered the distribution of the MELE with independent data and mixing time series, respectively. With spatial data, we show the MELE has properties resembling those available in other EL frameworks.

We first consider establishing the existence, consistency and asymptotic normality of a sequence of maxima of the EL ratio $R_n(\theta)$ from (5), along the lines of the classical arguments of Cramèr (1946). The conditions are mild and have the advantage that they are typically easy to verify. Let $\|\cdot\|$ denote the Euclidean norm in the following.

Theorem 2. (General estimating equations) *Assume $\mathcal{I}_n = \mathcal{I}_{b_n}^{OL}$ or $\mathcal{I}_{b_n}^{NOL}$, Assumptions 1–4 and (1) hold. In addition, suppose in a neighborhood of θ_0, $\partial G_\theta(\cdot)/\partial\theta$, $\partial^2 G_\theta(\cdot)/\partial\theta\partial\theta'$ are continuous in θ and $\|\partial G_\theta(\cdot)/\partial\theta\|, \|\partial^2 G_\theta(\cdot)/\partial\theta\partial\theta'\|$ are bounded by a nonnegative, real-valued $J(\cdot)$ with $\mathrm{E}\{J^3(Y_{\mathbf{s}})\} < \infty$; and $D_{\theta_0} \equiv \mathrm{E}\{\partial G_\theta(Y_{\mathbf{s}})/\partial\theta|_{\theta=\theta_0}\}$ has full column rank p. Then, as $n \to \infty$, $P\big(R_n(\theta)$ is continuously differentiable on $\|\theta - \theta_0\| \le n_Y^{-5/12}\big) \to 1$; there exists a sequence $\{\hat{\theta}_n\}$ such that $P\Big(R_n(\hat{\theta}_n) = \max_{\|\theta-\theta_0\| \le n_Y^{-5/12}} R_n(\theta)$ & $\|\hat{\theta}_n - \theta_0\| < n_Y^{-5/12}\Big) \to 1$; and*

$$n_Y^{\frac{1}{2}} \begin{pmatrix} \hat{\theta}_n - \theta_0 \\ b_n^{-d} t_{\hat{\theta}_n} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix} 0_p \\ 0_r \end{pmatrix}, \begin{bmatrix} V_{\theta_0} & 0 \\ 0 & U_{\theta_0} \end{bmatrix} \right),$$

where $V_{\theta_0} = \big(D'_{\theta_0}\Sigma_{\theta_0}^{-1}D_{\theta_0}\big)^{-1}$, $U_{\theta_0} = \Sigma_{\theta_0}^{-1} - \Sigma_{\theta_0}^{-1}D_{\theta_0}V_{\theta_0}D'_{\theta_0}\Sigma_{\theta_0}^{-1}$.

Remark. For an i.i.d. sample of size n, Qin and Lawless (1994) established a related result for a ball $\|\theta - \theta_0\| \le n^{-1/3}$. We could replace $n_Y^{-5/12}$ with $n_Y^{-1/3}$, to allow a larger ball in Theorem 2, by strengthening moment assumptions (i.e., $\mathrm{E}\left(\|G_\theta(Y_{\mathbf{s}})\|^{12+\delta}\right) < \infty$ in Assumption 1 of Section 8.1 of the online Appendix).

However, regardless of the ball radius $n_Y^{-5/12}$ or $n_Y^{-1/3}$, the maximizer of the EL function on each ball satisfies $\|\hat{\theta}_n - \theta_0\| = O_p(n_Y^{-1/2})$, and thereby maximizers on the different balls must ultimately be equal.

Theorem 2 establishes the existence of a local maximizer of the spatial EL function. When the likelihood $R_n(\theta)$ has a single maximum with probability approaching 1, by the concavity of $R_n(\theta)$ for example, then the sequence $\{\hat{\theta}_n\}$ corresponds to a global MELE. Under stronger conditions, as in Kitamura (1997), a global maximum on Θ can be shown to satisfy $\|\hat{\theta}_n - \theta_0\| = O_p(n_Y^{-1/2})$, thereby coinciding with the sequence in Theorem 2. However, Theorem 2 conditions are often sufficient, for many estimating functions, to ensure that the sequence $\{\hat{\theta}_n\}$ in Theorem 2 corresponds to global maximizers without more restrictive assumptions, such as compactness of the parameter space Θ. For example, this is true with estimating functions of the common form $G_\theta(Y_{\mathbf{s}}) = G(Y_{\mathbf{s}}) - \gamma(\theta)$ for some $G : \mathbb{R}^m \to \mathbb{R}^r$ and differentiable $\gamma : \Theta \to \mathbb{R}^r$ with $\|\gamma(\theta) - \gamma(\theta_0)\|$ increasing in $\|\theta - \theta_0\|$; see Example 1 of Section 2.1 for illustration.

3.3. Empirical likelihood tests of hypotheses

As in the EL frameworks of Qin and Lawless (1994) and Kitamura (1997), the spatial EL method allows test statistics based on $\hat{\theta}_n$ for both spatial parameter and moment hypotheses. The distribution of the log-EL ratio $r_n(\theta) \equiv \ell_n(\theta) - \ell_n(\hat{\theta}_n)$ at $\theta = \theta_0$ is useful for simple hypothesis tests or for calibrating approximate $100(1-\alpha)\%$ EL confidence regions for θ as $\{\theta \in \Theta : r_n(\theta) \le \chi^2_{p;1-\alpha}\}$. For testing the null hypothesis that the moment condition (1) holds for the estimating functions, the log-ratio statistic $\ell_n(\hat{\theta}_n)$ may be applied. Theorem 3 provides the limiting chi-square distributions of these EL log-ratio statistics.

In Theorem 3, we show additionally that the profile spatial EL ratio statistics can be developed to conduct tests and set confidence regions in the presence of nuisance parameters; see Qin and Lawless (1994) for the i.i.d. data case. Let $\theta = (\theta_1', \theta_2')'$, where θ_1 denotes the $q \times 1$ parameter of interest and θ_2 denotes a $(p-q) \times 1$ nuisance vector. For fixed θ_1, suppose that $\hat{\theta}_2^{(\theta_1)}$ maximizes the EL function $R_n(\theta_1, \theta_2)$ with respect to θ_2 and define the profile log-EL ratio $\ell_n(\theta_1) \equiv -2B_n \log R_n(\theta_1, \hat{\theta}_2^{(\theta_1)})$ for θ_1.

Theorem 3. *Under the assumptions of Theorem 2 with the sequence $\{\hat{\theta}_n\}$, as $n \to \infty$,*

(i) $r_n(\theta_0) = \ell_n(\theta_0) - \ell_n(\hat{\theta}_n) \xrightarrow{d} \chi^2_p$ and $\ell_n(\hat{\theta}_n) \xrightarrow{d} \chi^2_{r-p}$.

(ii) *If H_0: $\theta_1 = \theta_1^0$ holds, then $r_n(\theta_1^0) = \ell_n(\theta_1^0) - \ell_n(\hat{\theta}_{1n}) \xrightarrow{d} \chi^2_q$, where $\hat{\theta}_n = (\hat{\theta}_{1n}, \hat{\theta}_{2n})'$.*

We examine the performance of the spatial EL in subsequent sections. EL inference for spatial parameters under constraints is also possible, as considered by Qin and Lawless (1995) and Kitamura (1997) for i.i.d. and time series data; see Section 8.4 of the online Appendix for this.

4. A Bartlett Correction Procedure

A Bartlett correction is often an important property for EL methods. This involves making a mean adjustment to the EL log-ratio in order to improve the limiting chi-square approximation, and to enhance the coverage accuracy of EL confidence regions. For EL with independent data, a Bartlett correction has been established by DiCiccio, Hall and Romano (1991) for smooth function means, and by Chen and Cui (2006, 2007) under general estimating equations and nuisance parameters; see Chen and Cui (2007) for additional references with i.i.d. data. With weakly dependent time series, Kitamura (1997) and Monti (1997) considered Bartlett corrections for blockwise EL with mean parameters and a periodogram-type EL, respectively.

While a formal justification of a Bartlett correction in the spatial setting is difficult, a practical Bartlett correction for the spatial EL may be proposed using a spatial block bootstrap. Let $r_n(\theta) = \ell_n(\theta) - \ell_n(\hat{\theta}_n)$, $\theta \in \Theta \subset \mathbb{R}^p$, denote the log EL ratio from Section 3.3 based on the MELE $\hat{\theta}_n$ and (6). By Theorems 2−3, we have $r_n(\theta_0) \xrightarrow{d} \chi^2_p$ and $\hat{\theta}_n$ is consistent for θ_0, so that a bootstrap Bartlett correction factor may be calculated as follows. Pick some large $M \in \mathbb{N}$. For $i = 1, \ldots, M$, independently generate a block bootstrap rendition, say $\mathcal{Y}_n^{*i}$, of the original vectorized spatial data $\mathcal{Y}_n \equiv \{Y_{\mathbf{s}} : \mathbf{s} \in \mathcal{R}_{n,Y} \cap \mathbb{Z}^d\}$ and compute $r_n^{*i}(\hat{\theta}_n) = \ell_n^{*i}(\hat{\theta}_n) - \ell_n^{*i}(\hat{\theta}_n^{*i})$, where ℓ_n^{*i} and $\hat{\theta}_n^{*i}$ are the log EL ratio and MELE analogs based on $\mathcal{Y}_n^{*i}$. We then compute $\bar{r}_n^* = M^{-1}\sum_{i=1}^M r_n^{*i}(\hat{\theta}_n)$ to estimate $\mathrm{E}\{r_n(\theta_0)\}$ and set a Bartlett-corrected $100(1-\alpha)\%$ confidence region as $\{\theta : (p/\bar{r}_n^*) r_n(\theta) \leq \chi^2_{p,1-\alpha}\}$. If $\theta = (\theta_1', \theta_2')'$ with interest on $\theta_1 \in \mathbb{R}^q$, treating $\theta_2 \in \mathbb{R}^{p-q}$ as nuisance parameter as in Theorem 3, we take a Bartlett-corrected confidence region for θ_1 as $\{\theta_1 : (q/\bar{r}_n^*) r_n(\theta_1) \leq \chi^2_{q,1-\alpha}\}$ with respect to $r_n(\theta_1) = \ell_n(\theta_1) - \ell_n(\hat{\theta}_{1n})$ and $\bar{r}_n^*$ based on $r_n^{*i}(\hat{\theta}_{1n}) = \ell_n^{*i}(\hat{\theta}_{1n}) - \ell_n^{*i}(\hat{\theta}_{1n}^{*i})$. Under the smooth function model in Theorem 1, the same algorithm applies for the EL ratio $r_n(\theta_0^H) \equiv \ell_n(\theta_0^H)$, with $\ell_n(\hat{\theta}_n) = 0$ in this case.

As an alternative to the Bartlett correction, another option would be to calibrate confidence regions for the log El ratio $r_n(\theta)$ using sample quantiles from the M bootstrap replicates $r_n^{*i}(\hat{\theta}_n)$. The Bartlett correction involves estimating the mean of $r_n(\theta_0)$ at the true parameter θ_0 while the bootstrap calibration aims to approximate extreme quantiles of the distribution of $r_n(\theta_0)$. Intuitively, mean estimation is a more robust task and may possibly require fewer bootstrap replicates M for adequate estimates. Simulation studies with independent data in Chen and Cui (2007) appear to suggest this as well. For this reason, we concentrate our numerical studies in Section 5 on the Bartlett correction.

For completeness, we describe a spatial block bootstrap method for generating a bootstrap version of $\mathcal{Y}_n$ on $\mathcal{R}_{n,Y}$ in Section 8.5 of the online Appendix. The bootstrap involves spatial blocks determined by a block scaling factor $b_{n,bt}$, satisfying $b_{n,bt}^{-1} + b_{n,bt}^d / n_Y = o(1)$. The bootstrap scaling $b_{n,bt}$ may differ from the EL block scaling b_n and might be expected to be larger than b_n in many cases.

5. Numerical Study

We conducted a simulation study to compare OL and NOL versions of the blockwise EL method, and to examine the Bartlett correction algorithm for inference on the mean $\mathrm{E}(Z_{\mathbf{s}}) = \theta$ of a real-valued spatial process $Z_{\mathbf{s}}$, $\mathbf{s} \in \mathbb{Z}^2$, on the integer grid. Sampling regions $\mathcal{R}_n = \lambda_n(-1/2, 1/2]^2 \subset \mathbb{R}^2$ of different sizes were considered with $\lambda_n = 10, 20, 30$; a fourth region was taken as $\mathcal{R}_n = (-5, 5] \times (-15, 15]$. We used the circulant embedding method of Chan and Wood (1997) to generate real-valued mean-zero Gaussian random fields on $\mathbb{Z}^2$ with an Exponential or Gaussian covariance structure:

$$\mathbf{h} = (h_1, h_2)' \in \mathbb{Z}^2, \mathrm{Cov}\,(Z_{\mathbf{s}}, Z_{\mathbf{s}+\mathbf{h}}) = \begin{cases} \exp\big[-\beta_1|h_1| - \beta_2|h_2|\big] & \text{model E}(\beta_1,\beta_2) \\ \exp\big[-\beta_1|h_1|^2 - \beta_2|h_2|^2\big] & \text{model G}(\beta_1,\beta_2), \end{cases}$$

with values $(\beta_1, \beta_2) = (0.8, 0.8)$ and $(0.4, 0.2)$. Using $Y_{\mathbf{s}} = Z_{\mathbf{s}}$ and $G_\theta(Y_{\mathbf{s}}) = Z_{\mathbf{s}} - \theta$ in (1), we calculated approximate two-sided 90% EL intervals for θ as $\{\theta : r_n(\theta) \leq \chi^2_{1;0.9}\}$ using OL/NOL blocks of length $b_n = Cn^{1/5}$, $C = 1, 2$, where $n = |\mathcal{R}_n \cap \mathbb{Z}^2|$ and $r_n(\theta) = \ell_n(\theta)$; note $\ell_n(\hat{\theta}_n) = 0$ here for the mean and $n_Y = n$. This order of the EL block factor was intuitively chosen to be smaller than the optimal order $O(n^{1/(d+2)})$ known for spatial subsampling variance estimation when $d = 2$ (Sherman and Carlstein (1994); EL block scaling b_n is discussed further in Section 7. Using the algorithm from Section 4, Bartlett-corrected EL intervals were also computed using $M = 1,000$ Monte Carlo approximations and bootstrap block sizes $b_{n,bt} = n^{1/4}, n^{1/3}$. Additionally, for comparison to EL intervals, normal approximation intervals for θ were taken as $\bar{Z}_n \pm 1.645 S_n$ using the sample mean $\bar{Z}_n$ over $\mathcal{R}_n$ and a spatial subsampling variance estimator S_n^2 of $\mathrm{Var}\,(\bar{Z}_n)$ based on a plug-in estimate of its optimal block size, with pilot block sizes $n^{1/(2+2i)}$, $i = 1, 2$, see Nordman and Lahiri (2004). Table 1 provides summaries of the coverage accuracies and interval lengths for the EL method based on 1,000 simulation runs for each sampling region and covariance structure; Table 2 provides the same for the subsampling-based intervals. The Bartlett correction appears to provide large improvements in the EL intervals across a variety of dependence structures. From the results, we make the following observations:

Table 1. Coverage probabilities for approximate two-sided 90% EL confidence intervals for the process mean, with expected interval lengths, based on OL/NOL blocks of length b_n; UC, BC_3, BC_4 denote uncorrected and Bartlett-corrected intervals based on bootstrap blocks $b_{n,bt} = n^{1/3}, n^{1/4}$, respectively, and $n_1 \times n_2$ denotes the size of the sampling region with $n = n_1 n_2$.

	$b_n = n^{1/5}$						$b_n = 2n^{1/5}$					
	NOL			OL			NOL			OL		
E(0.4,0.2)	UC	BC_3	BC_4	UC	BC_3	BC_4	UC	BC_3	BC_4	UC	BC_3	BC_4
10×10	39.6 0.57	68.1 1.11	76.6 1.54	38.7 0.57	60.9 0.96	65.7 1.12	47.9 0.63	68.8 1.08	68.8 1.08	42.4 0.58	76.5 1.31	74.4 1.28
20×20	41.8 0.35	69.3 0.65	77.2 0.90	42.0 0.36	61.4 0.54	73.0 0.72	39.6 0.42	75.1 0.98	62.3 0.74	51.0 0.51	79.3 0.91	74.2 0.82
30×30	50.9 0.30	72.6 0.48	77.6 0.55	50.0 0.30	68.0 0.44	79.5 0.59	55.2 0.40	95.7 1.21	95.4 1.25	58.8 0.41	78.6 0.63	77.1 0.63
10×30	42.3 0.40	75.6 0.91	80.2 1.28	40.4 0.40	58.3 0.61	69.9 0.83	48 0.49	80.3 1.21	80.3 1.21	50.1 0.53	80.0 1.02	90.5 1.47
E(0.8,0.8)												
10×10	64.6 0.49	88.8 0.98	84.9 1.06	63.8 0.50	81.5 0.76	78.7 0.76	62.8 0.47	83.6 0.81	83.6 0.81	55.2 0.46	91.3 1.06	89.0 1.00
20×20	69.4 0.28	86.9 0.42	83.6 0.49	71.4 0.28	83.6 0.37	85.4 0.41	53.2 0.28	82.7 0.62	71.6 0.48	70.8 0.32	86.3 0.52	81.1 0.46
30×30	74.9 0.21	87.8 0.28	85.4 0.27	73.6 0.21	85.6 0.27	88.1 0.29	73.7 0.22	98.4 0.61	98.0 0.60	76.7 0.24	89.3 0.33	87.9 0.31
10×30	71.9 0.32	92.2 0.56	87.8 0.63	70.5 0.32	82.3 0.43	84.5 0.47	62.7 0.33	91.3 0.81	91.0 0.81	70.5 0.36	89.2 0.61	92.4 0.82
G(0.4,0.2)												
10×10	64.3 0.63	89.7 1.27	90.4 1.55	62.6 0.64	86.0 1.08	83.6 1.09	62.5 0.60	80.1 1.02	80.1 1.02	56.1 0.58	87.9 1.30	86.4 1.26
20×20	66.7 0.35	89.0 0.57	86.1 0.65	67.6 0.35	83.4 0.49	85.0 0.55	54.3 0.34	86.1 0.80	77.5 0.63	70.1 0.41	91.9 0.70	85.8 0.60
30×30	76.3 0.27	90.4 0.38	87.5 0.36	76.6 0.27	87.1 0.35	89.6 0.39	71.4 0.29	98.6 0.81	98.3 0.78	74.9 0.30	89.8 0.43	86.5 0.40
10×30	70.6 0.40	92.9 0.77	87.6 0.85	69.1 0.40	83.6 0.57	85.2 0.63	59.8 0.42	91.3 1.04	91.0 1.04	69.0 0.46	92.4 0.84	94.9 1.15

Table 2. Coverage probabilities and expected interval lengths for approximate two-sided 90% confidence intervals for the process mean based on a normal approximation with a subsampling variance estimator. Sampling region $\mathcal{R}_n$ sizes noted by $n_1 \times n_2$.

(β_1, β_2)	E(β_1, β_2)				G(β_1, β_2)			
	10×10	20×20	30×30	10×30	10×10	20×20	30×30	10×30
(0.4,0.2)	44.2 0.59	50.9 0.40	64.0 0.38	53.9 0.50	70.4 0.64	71.9 0.36	81.4 0.28	76.0 0.45
(0.8,0.8)	69.0 0.50	77.9 0.30	79.7 0.22	76.7 0.34	80.4 0.49	81.9 0.27	86.0 0.19	82.5 0.31

1. The coverage accuracies of the intervals often improved, and the interval lengths decreased, as the strength of underlying spatial dependence decreased and the size of the sampling region increased.
2. Coverage probabilities of uncorrected EL and normal approximation intervals were similar, and often far below the nominal level. This agrees with other simulation results for EL with independent data, in which uncorrected EL intervals often appeared too narrow (DiCiccio, Hall and Romano (1991) and Chen and Cui (2007)).
3. Bartlett-corrected EL intervals based on NOL and OL blocks were generally competitive and had coverage accuracies that were much closer to the nominal level than uncorrected intervals. The NOL block version typically performed better with shorter blocks b_n.
4. Under spatial dependence E(0.4,0.2), the Bartlett-corrected EL intervals were most sensitive to the EL and bootstrap block sizes. In this case, larger blocks seemed preferable to capture the stronger dependence structure.

Repeating the simulation with $M = 500$ or 250 bootstrap renditions did not change the results significantly, suggesting an adequate Bartlett correction may also be possible with fewer spatial bootstrap replicates.

6. Data Example: Cancer mortality map

The spatial EL method was applied to the cancer mortality map shown in Figure 1(a), constructed using mortality rates from liver and gallbladder cancer in white males during 1950-1959. Sherman and Carlstein (1994) considered these data for applying subsampling. We use their division of high and low mortality rates for illustration purposes, recognizing that the map's binary nature discards useful information relevant to the underlying scientific problem. The sampling region $\mathcal{R}_n$ in Figure 1(a) contains 2298 sites on a portion of the integer grid $(0, 66] \times (0, 58] \cap \mathbb{Z}^2$. For a given site $\mathbf{s} \in \mathbb{Z}^2$, we code $Z_{\mathbf{s}} = 0$ or 1 to indicate a low or high mortality rate, and let $S_{\mathbf{s}} = \sum_{\mathbf{h} \in N_{\mathbf{s}}} Z_{\mathbf{h}}$ denote the sum of indicators $Z_{\mathbf{h}}$ over the four nearest-neighbors $N_{\mathbf{s}} = \{\mathbf{h} \in \mathbb{Z}^2 : \|\mathbf{s} - \mathbf{h}\| = 1\}$ of site $\mathbf{s}$.

To test whether incidences of high cancer mortality exhibit clumping, Sherman and Carlstein proposed examining the spatial dependence parameter β of an autologistic model of the type introduced by Besag (1974). That is, suppose the binary process $Z_{\mathbf{s}}$, $\mathbf{s} \in \mathbb{Z}^2$ was generated by the conditional model, with parameters $\theta = (\alpha, \beta)'$, written as

$$f_\theta(z|\{Z_{\mathbf{h}} : \mathbf{h} \neq \mathbf{s}\}) = P_\theta\big(Z_{\mathbf{s}} = z \mid \{Z_{\mathbf{h}} \in N_{\mathbf{s}}\}\big) = \frac{\exp\left[z(\alpha + \beta S_{\mathbf{s}})\right]}{1 + \exp\left[\alpha + \beta S_{\mathbf{s}}\right]}, \quad z = 0, 1. \tag{7}$$

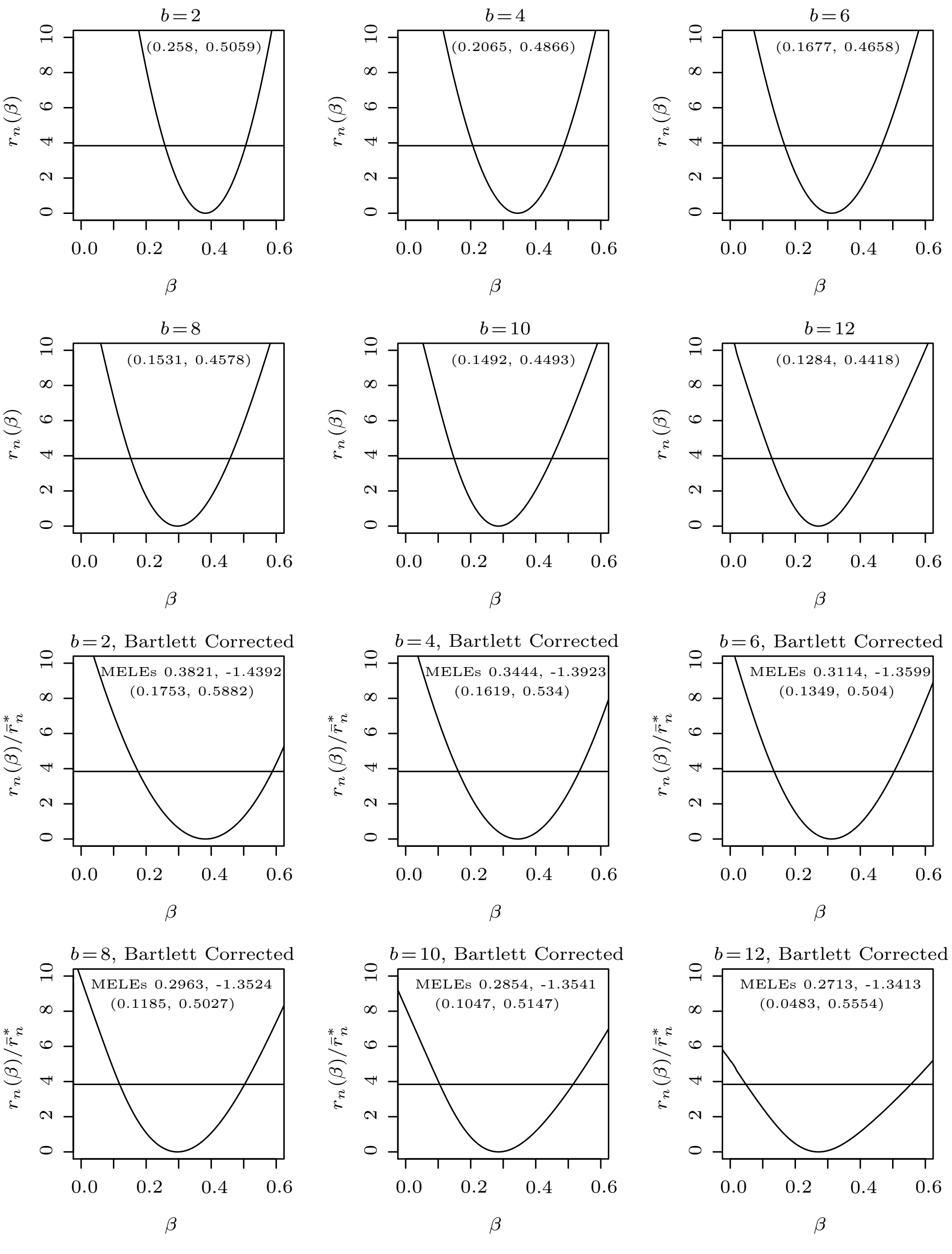


Figure 3. Spatial log-EL ratio $r_n(\beta)$ for β, and a Bartlett-corrected version $r_n(\beta)/\bar{r}_n^*$ for various block lengths b_n. Horizontal lines indicate the chi-square quantile $\chi^2_{1;0.95}$, and approximate 95% confidence intervals for β appear in brackets; MELEs $\hat{\beta}_n$, $\hat{\alpha}_n$ are given for each b_n.

Positive values of β suggest a tendency for clustering, while $\beta = 0$ implies no clustering among sites. Sherman and Carlstein set a normal-theory confidence interval for β based on the pseudo-likelihood estimate $\hat{\beta}_n^{PL}$ and a spatial subsampling variance estimate for Var $(\hat{\beta}^{PL})$.

The spatial EL may be applied to investigate evidence of clumping without a variance estimation step. For this, we use pseudo-likelihood-type estimating functions as described in Example 3 of Section 2.1. For $\theta = (\alpha, \beta)'$ in (7), we consider the vector process $Y_{\mathbf{s}}$ of dimension $m = 5$, formed by $Z_{\mathbf{s}}$ and its four nearest neighbors $Z_{\mathbf{h}}$, $\mathbf{h} \in N_{\mathbf{s}}$, along with $r = p = 2$ estimating functions $G_\theta(Y_{\mathbf{s}}) = \partial \log f_\theta(Z_{\mathbf{s}} \mid \{Z_{\mathbf{h}} : \mathbf{h} \in N_{\mathbf{s}}\})/\partial\theta$ based on (7). Figure 1(b) shows the sampling region $\mathcal{R}_{5,n}$ of these $Y_{\mathbf{s}}$-observations. Treating α as a nuisance parameter, we obtain a profile log-EL ratio $r_n(\beta) = \ell_n(\beta) - \ell_n(\hat{\beta}_n)$ for each β value, where $\ell_n(\beta) = \ell_n(\beta, \hat{\alpha}_n^{(\beta)})$, $\hat{\alpha}_n^{(\beta)} = \arg\max_\alpha R_n(\beta, \alpha)$, and $\hat{\beta}_n$ is the MELE for β. For various block choices b_n, we computed the MELEs $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n)'$ and, by Theorem 3, calibrated approximate 95% confidence intervals for β based on a χ_1^2 distribution for $r_n(\beta)$. Figure 3 shows the log-EL ratio $r_n(\beta)$, MELEs, and corresponding approximate 95% confidence interval for β with and without Bartlett corrections for each block size used. The Bartlett correction factor $\bar{r}_n^*$ was computed based on $M = 1,000$ bootstrap renditions of $\mathcal{R}_{5,m}$ and a block factor $b_{n,bt} = 6$.

As in the simulation study of Section 5, Bartlett-corrected EL intervals for β are notably wider than their uncorrected counterparts. EL intervals for β suggest clustering but these are shifted much closer to zero compared to Sherman and Carlstein's subsampling-based 95% confidence interval $(0.2185, 0.6183)$ (after re-parameterization there). In comparison, the EL method gives a slightly moderated interpretation of clustering. Of additional note, the behavior of EL intervals in Figure 3 also suggests a visual way of selecting a block size for the EL method; this is described in the next section.

7. Spatial Empirical Likelihood Block Scaling

The spatial EL proposed in this article involves a block condition (3), stating that the spatial sample size n_Y for $\mathcal{R}_{n,Y}$ must be of larger order than the squared number of observations in a spatial block b_n^{2d}. This appears also to be necessary for the results presented previously. To see why, note that, from Theorem 2, the exact order of the EL Lagrange multiplier $t_{\hat{\theta}_n}$ is $O_p(b_n^d/n_Y^{1/2})$, which is also the order of t_{θ_0} at the true parameter θ_0. Under the EL moment condition (1), we expect t_{θ_0} to converge to zero in probability (requiring $b_n^d/n_Y^{1/2} \to 0$) as the sample size increases, so that the EL block probabilities $p_{\theta_0,\mathbf{i}}$ from (5) become close to the probabilities $1/N_{\mathcal{I}}$ maximizing the EL function. Hence, (3) may represent the weakest possible requirement on the blocks.

Potential EL block scaling in $\mathbb{R}^d$ can involve $b_n = Cn_Y^{\kappa}$, for some $C > 0$ and $0 < \kappa < 1/(2d)$, although the best EL block orders for coverage accuracy are presently unknown for any d. With some time series block resampling methods, MSE-optimal blocks for distribution estimation are usually smaller than optimal blocks for variance estimation (Lahiri (2003)). This motivated the choice $b_n = Cn_Y^{1/5}$ in the simulation study of Section 5 so as to be smaller than the optimal block order $O(n_Y^{1/4})$ known for subsampling variance estimation when $d = 2$ (Sherman (1996)). This order choice of $\kappa = 1/5$ is also a compromise between the optimal block orders $\kappa \in [1/4, 1/6]$ for some $\mathbb{R}^2$-subsampling distribution estimators studied by Garcia-Soidan and Hall (1997).

In practice, EL block sizes might be chosen by the "minimum volatility" method, described by Politis, Romano and Wolf (1999, Sec. 9.3.2) for time series subsampling. The method is heuristic and based on the idea that, while some block sizes b_n may be too large or small, we might expect to find a range of b_n-values yielding approximately correct inference. In this range, confidence regions should be stable as a function of the block size. Hence, by creating EL confidence regions over a range of block sizes, an appropriate block size could be chosen by visual inspection. For illustration, we consider the EL confidence intervals in Figure 3 from the mortality map example. The apparent stability of these intervals over $b_n = 6, 8, 10$ seems to indicate that these block choices are reasonable for applying the EL method.

Acknowledgements

The author wishes to thank an associate editor and two referees for constructive comments that improved an earlier version of the paper, as well as Mark Kaiser for helpful discussions.

References

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. Roy. Statist. Soc. Ser. B* **36**, 192-236.

Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician* **24**, 179-195.

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

Chuang, C. and Chan, N. H. (2002). Empirical likelihood for autoregressive models, with applications to unstable time series. *Statist. Sinica* **12**, 387-407.

Chan, G. and Wood, A. T. A. (1997). An algorithm for simulating stationary Gaussian random fields. *Appl. Statist.* **46**, 171-181.

Chen, S. X. and Cui, H.-J. (2006). On Bartlett correction of empirical likelihood in the presence of nuisance parameters. *Biometrika* **93**, 215-220.

Chen, S. X. and Cui, H.-J. (2007). On the second order properties of empirical likelihood with moment restrictions. *J. Econometrics* (to appear)

Chen, S. X., Härdle, W. and Li, M. (2003). An empirical likelihood goodness-of-fit test for time series. *J. Roy. Statist. Soc. Ser. B* **65**, 663-678.

Cramèr, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press, New Jersey.

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

DiCiccio, T., Hall, P., and Romano, J. P. (1991). Empirical likelihood is Bartlett-correctable. *Ann. Statist.* **19**, 1053-1061.

Garcia-Soidan, P. H. and Hall, P. (1997). In sample reuse methods for spatial data. *Biometrics* **53**, 273-281.

Hall, P. and La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *Internat. Statist. Rev.* **58**, 109-127.

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

Kitamura, Y., Tripathi, G. and Ahn, H. (2004). Empirical likelihood-based inference in conditional moment restriction models. *Econometrica* **72**, 1667-1714.

Lahiri, S. N. (2003). *Resampling Methods for Dependent Data.* Springer, New York.

Lin, L. and Zhang, R. (2001). Blockwise empirical Euclidean likelihood for weakly dependent processes. *Statist. Probab. Lett.* **53**, 143-152.

Lee, Y. D. and Lahiri, S. N. (2002). Least squares variogram fitting by spatial subsampling. *J. Roy. Statist. Soc. Ser. B* **64**, 837-854.

Monti, A. C. (1997). Empirical likelihood confidence regions in time series models. *Biometrika* **84**, 395-405.

Newey, W. K. and Smith, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* **72**, 219-255.

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. (2006). A frequency domain empirical likelihood for short- and long-range dependence. *Ann. Statist.* **34**, 3019-3050.

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.

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

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

Qin, J. and Lawless, J. (1995). Estimating equations, empirical likelihood and constraints on parameters. *Canad. J. Statist.* **23**, 145-159.

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

Sherman, M. and Carlstein, E. (1994). Nonparametric estimation of the moments of a general statistic computed from spatial data. *J. Amer. Statist. Assoc.* **89**, 496-500.

Zhang, J. (2006). Empirical likelihood for NA series. *Statist. Probab. Lett.* **76**, 153-160.

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

E-mail: dnordman@iastate.edu

(Received May 2006; accepted April 2007)