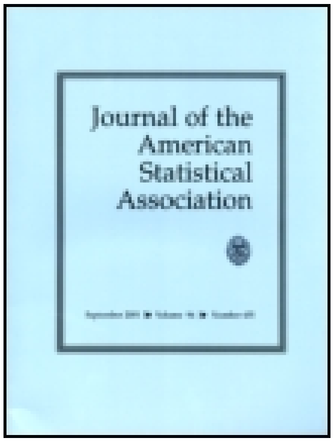

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

Nonparametric Estimation of Probability Density Functions for Irregularly Observed Spatial Data

Zudi Lu & Dag Tjøstheim

Accepted author version posted online: 05 Aug 2014.Published online: 22 Dec 2014.

Nonparametric Estimation of Probability Density Functions for Irregularly Observed Spatial Data

Zudi Lu and Dag Tjøstheim

Nonparametric estimation of probability density functions, both marginal and joint densities, is a very useful tool in statistics. The kernel method is popular and applicable to dependent data, including time series and spatial data. But at least for the joint density, one has had to assume that data are observed at regular time intervals or on a regular grid in space. Though this is not very restrictive in the time series case, it often is in the spatial case. In fact, to a large degree it has precluded applications of nonparametric methods to spatial data because such data often are irregularly positioned over space. In this article, we propose nonparametric kernel estimators for both the marginal and in particular the joint probability density functions for nongridded spatial data. Large sample distributions of the proposed estimators are established under mild conditions, and a new framework of expanding-domain infill asymptotics is suggested to overcome the shortcomings of spatial asymptotics in the existing literature. A practical, reasonable selection of the bandwidths on the basis of cross-validation is also proposed. We demonstrate by both simulations and real data examples of moderate sample size that the proposed methodology is effective and useful in uncovering nonlinear spatial dependence for general, including non-Gaussian, distributions. Supplementary materials for this article are available online.

KEY WORDS: Asymptotic normality; Expanding-domain infill asymptotics; Irregularly positioned spatial data; Marginal and joint probability density functions; Non-Gaussian distribution; Nonlinear spatial dependence; Nonparametric kernel method.

1. INTRODUCTION

Nonparametric estimation methods are well established for time series and have found extensive practical applications; see, for example, Fan and Yao (2003) and Teräsvirta, Tjøstheim, and Granger (2010). In particular this is the case for the kernel method of estimating a density function and a conditional expectation with applications to model identification, diagnostic checking, and forecasting. Nonparametric methods are far less used for spatial random variables. There are several reasons for this, but perhaps the most important one has been the fact that the sampling points are often irregularly positioned in space. For a time series observations are usually taken at, or can be aggregated at, regular time intervals. For spatial series, at least in the absence of computerized pixel-based images, this has not been the case. Because of physical constraints measurement stations cannot usually be put on a regular grid in space. This means that spatial analysis has been almost completely dominated by parametric models; for example, parametric models for covariance functions (or variogram) in kriging (Cressie 1993; Stein 1999). Also, partly for this reason one has to a large degree had to limit oneself to linear models, although there are (Chiles and Delfiner 1999) the so-called conjunctive kriging models that are nonlinear.

One of the main motivations for nonparametric analysis in time series is density estimation, in particular joint density estimation since this in turn can be used in nonlinear modeling in regression and additive modeling. In all of this the regular grid assumption is vital. Consider, for example, the task of estimating the joint density $p_1 = p_{X_t, X_{t-1}}$ of consecutive observations in a time series. Then the kernel estimate is given by

$$\widehat{p}_1(x_1, x_2) = \frac{1}{n}\sum_{1}^{n} K_h(X_t - x_1)K_h(X_{t-1} - x_2),$$

where $K_h(x) = h^{-1}K(x/h)$ with K being a kernel function. To evaluate this asymptotically we need infinitely many pairs (X_t, X_{t-1}) taken exactly one time unit apart. If the observations were taken at irregular time intervals, there would not be enough observation points to estimate the joint density for a specified time difference.

The above problem is much worse for spatial variables. Therefore, the nonparametric theory has so far almost exclusively been developed for the regular grid case (Gao, Lu, and Tjøstheim 2006; Hallin, Lu, and Tran 2001; 2004a, b; Lu et al. 2007; Tran 1990). This literature does not quite have the potential for applications that one would wish for, because spatial measurements on a grid have been an exception.

The purpose of this article is to try to break out of this confinement. We make an attempt to construct an asymptotic theory for nonparametric density estimation, both marginal and joint, for a random field with irregularly placed observations. To our knowledge, our article represents the first step in this direction, but some related attempts have been made for other problems. Hall and Patil (1994) looked at nonparametric estimation of a spatial covariance function using observations whose locations are generated by a probability distribution. Matsuda and Yajima (2009) had a similar model for generating observations irregularly in space and looked at the estimation of the spectral density nonparametrically and parametrically with the emphasis on the latter. The focus by Jenish and Prucha (2009) was also on parametric models.

Zudi Lu, Southampton Statistical Sciences Research Institute, and School of Mathematical Sciences, University of Southampton, SO17 1BJ, UK (E-mail: zudilu@gmail.com); Dag Tjøstheim, Department of Mathematics, The University of Bergen, Bergen 5007, Norway (E-mail: dag.tjostheim@math.uib.no). The authors are grateful to the editor, the associate editor, and two anonymous referees for their very encouraging and constructive comments and suggestions, which have greatly improved the presentation of the article. The authors are also grateful to Prof. Y. Yajima for his kind provision of Tokyo land price dataset, and to Zhenyu Jiang for computing help. This project was partially supported by a Discovery Project grant and a Future Fellowships grant from the Australian Research Council and a Marie Curie Career Integration Grant from European Commission, which are acknowledged.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/jasa.

Journal of the American Statistical Association
December 2014, Vol. 109, No. 508, Theory and Methods
DOI: 10.1080/01621459.2014.947376

Traditionally there have been two different approaches to asymptotics for spatial data. The so-called expanding domain asymptotics let the domain of the measurement locations tend to infinity. It is more or less a direct extension of the corresponding device of increasing the number of observations for a time series. It is primarily suited to a situation where the measurements are on a grid. The other approach is the infill asymptotics where the total domain is kept fixed but the density of the measurement points in that domain is allowed to increase indefinitely. In this article we use a mixed domain approach in which both the density of observations and the domain itself are allowed to increase. Using this method we have been able to find conditions under which the density estimates are consistent and asymptotically normal and where error limits and confidence intervals can be found. Besides being of interest on its own, we believe that the theory also opens up for further developments. For instance, one could start looking at nonlinear regression in an additive model context (see Lu et al. 2007) and at nonlinear and nonparametric interpolation, model identification and testing. This would be an alternative to linear and nonlinear kriging, and these ramifications will be the subject of future publications. In addition, as commented by the Associate Editor of this article, an interesting application of our procedure is to estimate the covariance matrix for the longitudinal data, which are also often irregularly sampled; see, for example, Fan, Huang, and Li (2007) and Fan and Wu (2008).

Here is a brief overview of the article: In Section 2, the methodology for the estimation of the marginal and joint probability density functions together with a new framework of expanding-domain infill asymptotics for irregularly positioned spatial data is proposed. Section 3 states the basic assumptions and notation needed for establishing the asymptotic theory of the proposed estimators. Sections 4 and 5 will then develop the asymptotic properties, including asymptotic biases, variances, and distributions, under the new framework of spatial asymptotics, for the marginal and joint probability density estimators. Estimation of the isotropic joint probability density function is further investigated in Section 6. Section 7 provides a practical, reasonable selection of the bandwidths involved in the joint density estimators on the basis of cross-validation combined with asymptotic results, which is important for real applications. Numerical evidence in terms of both simulations and real data examples is presented in Section 8. The proofs of the main theorems are collected in Section 9, with the necessary lemmas provided in an Appendix as online supplementary material.

2. METHODOLOGY

Assume we observe the spatial data $\{Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_N)\}$ from a stationary spatial process $\{Y(\mathbf{s})\}$ on $\mathbb{R}^2$, where $\mathbf{s}_i = (u_i, v_i) \in \mathbb{R}^2$ for $i = 1, 2, \ldots, N$, are allowed to be irregularly positioned. Let $f(x)$ be the marginal probability density function of $Y(\mathbf{s})$, and $f(x, y; \mathbf{s}_0)$ the joint probability density function of $Y(\mathbf{s})$ and $Y(\mathbf{s} + \mathbf{s}_0)$ that characterizes the nonlinear, non-Gaussian spatial dependence, with $\mathbf{s}_0 \neq (0, 0)$. We are interested in the estimation of $f(x)$ and $f(x, y; \mathbf{s}_0)$ based on the observed data.

2.1 Estimating the Marginal and Joint Probability Density Functions

We can easily construct the estimator of the marginal density $f(x)$ of $Y_i = Y(\mathbf{s}_i)$ as follows

$$\hat{f}(x) = \frac{1}{N} \sum_{i=1}^{N} K_h(Y(\mathbf{s}_i) - x), \tag{2.1}$$

where $K_h(x) = h^{-1} K(x/h)$ with $K(\cdots)$ a kernel function on $\mathbb{R}$ and $h = h_N \to 0$ $(N \to \infty)$ is a bandwidth.

However, the joint density $f(x, y; \mathbf{s}_0)$ cannot be estimated as simply as $f(x)$. A formal estimator of $f(x, y; \mathbf{s}_0)$ could be

$$\check{f}(x, y, \mathbf{s}_0) = \frac{1}{n_0} \sum_{j,\ell \in S_0} K_h(Y_j - x) K_h(Y_\ell - y),$$

where $S_0 := \{(j, \ell) : \mathbf{s}_j - \mathbf{s}_\ell = \mathbf{s}_0,\ j, \ell = 1, \ldots, N\}$ and $n_0 = {}^{\#}S_0$, the cardinality of the set S_0. However, in practice, owing to the irregularity over the plane, n_0 could be very small or even equal to 0, and therefore a more intuitively appealing estimator, with more neighboring observations used, could be constructed as

$$\hat{f}(x, y; \mathbf{s}_0) = \frac{\sum_{j,\ell=1}^{N} L_b(\mathbf{s}_j - \mathbf{s}_\ell - \mathbf{s}_0) K_h(Y_j - x) K_h(Y_\ell - y)}{\sum_{j,\ell=1}^{N} L_b(\mathbf{s}_j - \mathbf{s}_\ell - \mathbf{s}_0)}, \tag{2.2}$$

where $L_b(\mathbf{s}) = b^{-2} L(\mathbf{s}/b)$ with $L(\cdot)$ a kernel function on $\mathbb{R}^2$ and $b = b_N \to 0$ $(N \to \infty)$ is a bandwidth.

The spirit of constructing $\hat{f}(x, y; \mathbf{s}_0)$ by using spatial smoothing in (2.2) is similar to that of the nonparametric covariance estimator in Hall and Patil (1994) who constructed a nonparametric estimation of covariance function that characterizes the linear spatial dependence, while, differently, our estimator (2.2) is concerned with the estimation of nonlinear dependence for the spatial process. Clearly, it becomes more challenging to investigate the *irregularly positioned spatial data* than the case of regularly gridded data: for example, as discussed above, since it necessitates the introduction of an additional bandwidth parameter b, it leads to increased theoretical complexity and the practical difficulty of having to choose two bandwidth parameters in the implementation of (2.2).

Our main objective of this article is to establish the asymptotic properties, in Sections 3–5, of the estimators $\hat{f}(x)$ and $\hat{f}(x, y; \mathbf{s}_0)$. We will study the asymptotic properties under a new asymptotic framework that is suggested below.

2.2 Framework of Domain-Expanding Infill (DEI) Asymptotics

As mentioned in the introduction, there are two distinct asymptotics in spatial statistics: expanding domain asymptotics, where more data are collected by increasing the domain with the distance between neighboring observations remaining at least roughly constant (see, e.g., Dalenius, Hájek, and Zubrzycki 1961; Matérn 1960; Quenouille 1949), and fixed-domain or infill asymptotics, where more data are collected by sampling more densely in a fixed domain with the distance between neighboring observations tending to zero (see, e.g., Novak 1988 and Traub, Wozniakowski, and Wasilkowski 1988). See also Cressie

(1993) and Stein (1999). Asymptotic properties of the estimators are quite different under the two asymptotic schemes. A notable difference is in the consistency versus inconsistency of the estimators under the increasing-domain and fixed-domain asymptotics, respectively. There is a rich literature under both frameworks in spatial statistics and econometrics; see Zhang (2004), Zhang and Zimmerman (2005), Lu, Tjøstheim, and Yao (2008), Du, Zhang, and Mandrekar (2009), Lu et al. (2009), and the references therein for recent reviews on fixed-domain asymptotics, and Hallin, Lu, and Tran (2001, 2004a,b), Gao, Lu, and Tjøstheim (2006), Lu et al. (2007), Robinson (2008, 2009, 2011), etc., for the recent developments under expanding domain asymptotics of nonlinear and/or non-Gaussian spatial processes.

Different from the above references, in this article we will study the asymptotic properties of the proposed joint probability density estimators under a framework that combines the two frameworks of the asymptotics. We will call it a framework of *domain-expanding infill* asymptotics, or simply DEI asymptotics, which is defined as

$$\begin{aligned}\delta_N &= \max_{1\le j\le N}\delta_{j,N} \to 0, \text{ with } \delta_{j,N}\\ &= \min\{\|\mathbf{s}_i - \mathbf{s}_j\| : 1 \le i \le N, i \ne j\},\end{aligned} \tag{2.3}$$

that is, the distance between neighboring observations all tends to 0, as $N \to \infty$, and

$$\begin{aligned}\Delta_N &= \min_{1\le j\le N} \Delta_{j,N} \to \infty, \text{ with } \Delta_{j,N}\\ &= \max\{\|\mathbf{s}_i - \mathbf{s}_j\| : 1 \le i \le N, i \ne j\},\end{aligned} \tag{2.4}$$

that is, the domain at each location is expanding to ∞, as $N \to \infty$, where $\|\cdot\|$ is the usual Euclidean norm in $\mathbb{R}^2$. To avoid confusion, we will call the framework of traditional infill asymptotics *domain-fixed infill* (DFI) asymptotics, which in our terms means that (2.3) is true but $\max_{1\le j\le N}\Delta_{j,N} \le C_0 < \infty$ for all positive integer N, while the framework of traditional *domain-expanding* (DE) asymptotics means that (2.4) holds true but $\min_{1\le j\le N}\delta_{j,N} \ge c_0 > 0$ for all N. For convenience, we will call δ_N an *infilling distance* and Δ_N an *expanding distance* of the spatial sites, in the sequel.

The reasons why we adopt DEI asymptotics are mainly from the following considerations: First, DEI asymptotics reconciles the gap between DFI and DE asymptotics. It overcomes the drawbacks of DFI and DE asymptotics in that DFI cannot guarantee the estimators to be consistent as showed by Lahiri (1996) and Zhang (2004), while DE asymptotics for which the distance between neighboring observations does not tend to zero appears less attractive in such applications as spatial interpolation (kriging) as argued in Stein (1999, p. 62). Second, DEI asymptotics combines both the advantages of DE and DFI frameworks, and is therefore appealing in both furnishing consistent estimators ensured by the expanding aspect and having the potential of application to spatial prediction (kriging) requiring the infill aspect. Third, which kind of asymptotics need be applied in constructing confidence intervals for example, may depend on the viewpoint of the user. But in many situations the DEI framework may be natural as a result of the data structure. For example, socio-economic data are often collected over individual cities and suburbs: on one hand, more cities and suburbs are taken into account, meaning an expanding domain is adopted; on the other hand, within each city or suburb, more observations may be collected, requiring an infill asymptotic simultaneously.

It may also be worth pointing out that our defined domain-expanding infill asymptotic by (2.3) and (2.4) is different from an alternative version of mixed asymptotic adopted by Hall and Patil (1994); see also Matsuda and Yajima (2009), who assumed in the notation of this article, that $\mathbf{s}_i = (A_1u_{i,1}, A_2u_{i,2})$, where $\mathbf{u}_i = (u_{i,1}, u_{i,2})$, $i = 1, 2, \dots, N$, are independently and identically distributed random vectors with a probability density function $f_{\mathbf{U}}(\mathbf{u})$ of a compact support in $[0, 1]^2$, and $A_1 = A_{1,N} \to \infty$, $A_2 = A_{2,N} \to \infty$ as $N \to \infty$. We do not impose such a particular structure on the spatial locations, because such a location structure is sample size dependent via A_1 and A_2 and appears not suitable for defining a fixed location where spatial prediction (kriging) needs to be made.

3. ASSUMPTIONS AND NOTATION

For the sake of convenience, we are summarizing here the main assumptions we are making on the random field $\{Y(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^2\}$ and the kernels $K(\cdot)$ and $L(\cdot)$ to be used in the estimation method. Assumptions (I)–(II) are related to the random field itself.

For any collection of sites $\mathcal{S} \subset \mathbb{R}^2$, denote by $\mathcal{B}(\mathcal{S})$ the Borel σ-field generated by $\{Y(\mathbf{s})|\ \mathbf{s} \in \mathcal{S}\}$, and for each couple $\mathcal{S}', \mathcal{S}''$, let $d(\mathcal{S}', \mathcal{S}'') := \min\{\|\mathbf{s}' - \mathbf{s}''\| \,|\, \mathbf{s}' \in \mathcal{S}', \ , \mathbf{s}'' \in \mathcal{S}''\}$ be the distance between $\mathcal{S}'$ and $\mathcal{S}''$, where $\|\mathbf{s}\| := (u^2 + v^2)^{1/2}$, for $\mathbf{s} = (u, v) \in \mathbb{R}^2$, stands for the Euclidean norm. Finally, write $\mathrm{Card}(\mathcal{S})$ for the cardinality of $\mathcal{S}$.

Assumption (I) (spatial process):

(i) $\{Y(\mathbf{s}), \mathbf{s} \in \mathbb{R}^2\}$ is a strictly stationary spatial process, satisfying the α-mixing property that there exist a function φ such that $\varphi(t) \downarrow 0$ as $t \to \infty$, and a function $\psi : \mathcal{N}^2 \to \mathbb{R}^+$ symmetric and increasing in each of its two arguments, such that

$$\begin{aligned}&\alpha(\mathcal{B}(\mathcal{S}'), \mathcal{B}(\mathcal{S}''))\\ &\quad := \sup\{|\mathrm{P}(AB) - \mathrm{P}(A)\mathrm{P}(B)|, A \in \mathcal{B}(\mathcal{S}'),\ B \in \mathcal{B}(\mathcal{S}'')\}\\ &\quad \le \psi(\mathrm{Card}(\mathcal{S}'), \mathrm{Card}(\mathcal{S}''))\varphi(d(\mathcal{S}', \mathcal{S}'')),\end{aligned} \tag{3.5}$$

for any $\mathcal{S}', \mathcal{S}'' \subset \mathbb{R}^2$. The function φ moreover is such that

$$\lim_{m\to\infty} m^{\gamma} \sum_{j=m}^{\infty} j^2\{\varphi(j)\}^{\kappa/(2+\kappa)} = 0$$

for some constant $\gamma > \max\{1, 2\kappa/(2+\kappa)\}$ and some $\kappa > 0$.

(ii) Denote by $f(x, y; \mathbf{s}_0)$ the joint density function of $Y(\mathbf{s})$ and $Y(\mathbf{s} + \mathbf{s}_0)$, where $\mathbf{s}_0 \ne (0, 0)$. Here $f(x, y; \mathbf{s})$ is continuous as a function of (x, y) uniformly with respect to $\mathbf{s} \in \mathbb{R}^2$, and has second-order partial derivatives with respect to x, y, and $\mathbf{s}$, which are continuous.

(iii) The marginal and joint probability density functions $f(x)$ and $f_{i,j}(x, y)$ for Y_i and (Y_i, Y_j) satisfy $|f_{i,j}(x, y) - f(x)f(y)| \le C$ uniformly for $i \ne j$ and $(x, y) \in \mathbb{R}^2$, where C is a generic positive constant; further, the joint probability density functions $f_{i,j,k}(x, y, z)$ and $f_{i,j,k,\ell}(x, y, z, w)$ for (Y_i, Y_j, Y_k) and

(Y_i, Y_j, Y_k, Y_ℓ), respectively, are bounded uniformly with respect to $i \neq j \neq k \neq \ell$.

Assumption (II) (sampling sites): The observations are positioned at $\{\mathbf{s}_i,\ i = 1, 2, \ldots, N\} \subset \mathbb{R}^2$, for which (2.3) and (2.4) hold true with $\min_{1\le j\le N} \delta_{j,N}/\delta_N \ge c_1 > 0$ and $\max_{1\le j\le N} \Delta_{j,N}/\Delta_N \le C_1 < \infty$ for all N, and there exists a continuous sampling intensity function (i.e., density function) $f_{\mathbf{S}}$ defined on $\mathbb{R}^2$ such that:

(i) for any measurable set $A \subset \mathbb{R}^2$, $N^{-1}\sum_{i=1}^N I(\mathbf{s}_i \in A) \to \int_A f_{\mathbf{S}}(\mathbf{s})d\mathbf{s}$ as $N \to \infty$,
(ii) $f_{\mathbf{S}}(\mathbf{s})$ is bounded and has second derivatives which are continuous on $\mathbb{R}^2$, and
(iii) $A_0(\mathbf{s}_0) = \int_{\mathcal{R}^2} f_{\mathbf{S}}(\mathbf{s}_2 + \mathbf{s}_0) f_{\mathbf{S}}(\mathbf{s}_2)d\mathbf{s}_2 > 0$, and $\mathbf{A}_1(\mathbf{s}_0) = \int_{\mathcal{R}^2} \frac{\partial f_{\mathbf{S}}(\mathbf{s}_2+\mathbf{s}_0)}{\partial \mathbf{s}} f_{\mathbf{S}}(\mathbf{s}_2)d\mathbf{s}_2$ and $A_2(\mathbf{s}_0) = \int_{\mathcal{R}^2} f_{\mathbf{S}}^2(\mathbf{s}_2 + \mathbf{s}_0) f_{\mathbf{S}}(\mathbf{s}_2)d\mathbf{s}_2$ exist.

Assumption (III) (kernel functions):

(i) The kernel function $K(\cdot)$ satisfies that $\int K(u)du = 1$, $\int uK(u)du = 0$, and $\mu_{K,2} = \int u^2K(u)du < \infty$, $\nu_K = \int K^2(u)du < \infty$.
(ii) The kernel function $L(\cdot)$ has a bounded support, such that $\int_{\mathcal{R}^2} L(\mathbf{s})d\mathbf{s} = 1$, $\int_{\mathcal{R}^2} \mathbf{s}L(\mathbf{s})d\mathbf{s} = \mathbf{0}$, and $b\mu_{L,2} = \int_{\mathcal{R}^2} \mathbf{s}\mathbf{s}^\tau L(\mathbf{s})d\mathbf{s} < \infty$, $\nu_L = \int_{\mathcal{R}^2} L^2(\mathbf{s})d\mathbf{s} < \infty$.

Assumption (IV) (bandwidths):

(a) As $N \to \infty$, (i) $h \to 0$, (ii) $Nh \to \infty$, and (iii) $\delta_N^{-2(1+2/\gamma)} h^{1-2\kappa/\{(2+\kappa)\gamma\}} \to 0$.
(b) As $N \to \infty$, (i) $h \to 0$, (ii) $b \to 0$, (iii) $Nhb \to \infty$, $Nhb^2 \to 0$, (iv) $N(b/\delta_N)^2 h^{-2\kappa/(2+\kappa)} \to \infty$, $N(b/\delta_N)^2 h^{2(\gamma-2\kappa/(2+\kappa))/(2+\gamma)} \to 0$. Here in both (a) and (b), $\gamma > 2\kappa/(2+\kappa)$ and $\kappa > 0$ were defined in Assumption (I)(i).

Assumption (I) concerns the conditions on the spatial data-generating process, which are standard in the context of the problem under study. For example, Assumption (I)(i), the α-mixing condition, is similar to (A4) of Hallin, Lu, and Tran (2004b). This is a technical assumption used in both nonlinear time series and spatial literature to characterize the data dependence. In the serial case, many stochastic processes and time series are known to be strongly mixing. Withers (1981) obtained various conditions for linear processes to be strongly mixing. Under certain weak assumptions, autoregressive and more general nonlinear time-series models are strongly mixing with exponential mixing rates; see Pham and Tran (1985), Pham (1986), Tjøstheim (1990), and Lu (1998). Guyon (1987), as well as Guyon (1995), has shown that the results of Withers under certain conditions extend to linear random fields, of the form $X_{\mathbf{n}} = \sum_{\mathbf{j}\in\mathbb{Z}^2} g_{\mathbf{j}} Z_{\mathbf{n}-\mathbf{j}}$, with $\mathbf{n} \in \mathbb{Z}^2$, over gridded space, where the $Z_{\mathbf{j}}$'s are independent random variables. In addition, Assumptions (M, iii) and (J, iii), given in the next sections, on the functions $\psi(\cdot)$ and $\varphi(\cdot)$ in (3.5) are very mild, which allow for Assumptions (A4′) and (A4″) of Hallin, Lu, and Tran (2004b) that are the same as the mixing conditions used by Neaderhouser (1980) and Takahata (1983), respectively, and are weaker than the uniform strong mixing condition considered by Nakhapetyan (1980). They are satisfied by many spatial models, as shown by Neaderhouser (1980), Rosenblatt (1985), and Guyon (1987). This α-mixing assumption can be relaxed to a more general Near Epoch Dependence, having the α-mixing and the data process of Hallin, Lu, and Tran (2001, 2004a) as special cases, but it would involve much lengthier proofs (e.g., in time series case, see Lu 2001 and Lu and Linton 2007).

Assumption (I)(ii) is a standard condition on the probability density function to be estimated. Assumption (I)(iii) is a mild technical condition with uniform boundedness imposed on the joint probability density functions to ensure the uniform consistency with respect to the different spatial sites. Here the first part of Assumption (I)(iii) is standard in this context; it has been used, for instance, by Masry (1986) in the time series case, and by Tran (1990) and Hallin, Lu, Tran (2004b) in the gridded spatial context. If the random field $Y_i = Y(\mathbf{s}_i)$ consists of independent observations, then $|f_{i,j}(x, y) - f(x)f(y)|$ vanishes as soon as i and j are distinct. Note that the first part of Assumption (I)(iii) also allows for an unbounded marginal density.

Assumption (II) provides the conditions on the spatial sites where observations are irregularly positioned, under our framework of DEI asymptotics, among which (II)(i) is the same as A3 of Lu, Tjøstheim, and Yao (2008) and Condition 8 of Lu et al. (2009, p. 878), and (II)(ii, iii) are also the mild conditions required in the derivation of asymptotic bias and variance of the concerned estimators. Assumption (III) specifies the conditions on the kernel functions used, which are very standard in nonparametric kernel estimation. Assumption (IV) part (a) lists the conditions on the bandwidths used for the marginal density estimator (2.1), where (i) and (ii) are standard as in traditional marginal density estimator and (iii) is a new condition related to δ_N; while part (b) for the joint density estimator (2.2), where (i) and (ii) are standard, (iii) is mild and (iv) is a bit involved relating to the δ_N but is still easily satisfied. For example, if $b = CN^{-1/2}$, where $0 < C < \infty$ is a generic constant independent of N, then $Nhb^2 = O(h) \to 0$ and $N(b/\delta_N)^2h^{-2\kappa/(2+\kappa)} = O(1)(\delta_N)^{-2}h^{-2\kappa/(2+\kappa)} \to \infty$ hold naturally, while $Nhb \to \infty$ and $N(b/\delta_N)^2h^{2(\gamma-2\kappa/(2+\kappa))/(2+\gamma)} \to 0$ hold true under simple conditions that $Nh^2 \to \infty$ and $h^{(\gamma-2\kappa/(2+\kappa))/(2+\gamma)}/\delta_N \to 0$ for $\gamma > 2\kappa/(2+\kappa)$ and $\kappa > 0$.

We finally let $\mathrm{tr}(A)$ stand for the trace of a square matrix A.

4. ASYMPTOTICS FOR THE MARGINAL DENSITY FUNCTION ESTIMATOR

In this section, we are first concerned with the asymptotics for the estimator $\hat{f}(x)$, defined in (2.1), of the marginal distribution $f(x)$, which takes the same form as in the case of the regularly positioned spatial data in Tran (1990) and Hallin, Lu, and Tran (2001, 2004a).

First, it is quite obvious by the stationarity of Y_i that $\hat{f}(x)$ has a dominating bias term $B(\hat{f}(x)) = \frac{1}{2}\ddot{f}(x)h^2\int u^2K(u)du$ because

$$\begin{aligned} E\hat{f}(x) &= EK_h(Y_j - x) = h^{-1}\int K((u-x)/h)f(u)du \\ &= \int K(u)f(x+hu)du = \int K(u)[f(x) + \dot{f}(x)hu \\ &\quad + \frac{1}{2}\ddot{f}(x+\zeta hu)(hu)^2]du \\ &= f(x) + \frac{1}{2}\ddot{f}(x)h^2\int u^2K(u)du(1+o(1)), \end{aligned} \tag{4.6}$$

by using Assumption (III, i) on the kernel $K(\cdot)$, where $|\zeta| < 1$. Then, under Assumption (IV)(a), we can derive the asymptotic variance for $\hat{f}(x)$ as follows

$$\mathrm{var}(\hat{f}(x)) = \frac{1}{Nh} f(x) \int K^2(u)du(1 + o(1)). \tag{4.7}$$

Moreover, for marginal density estimation we state the following additional assumption on the bandwidth h relating to the infilling distance δ_N and the mixing coefficient defined in (3.5).

Assumption (M): Let $c_N = \{\delta_N^2 h^{\kappa/(2+\kappa)}\}^{-1/\gamma}$, which tends to ∞ as $N \to \infty$. (i) $\overline{\lim}_{N\to\infty} Nh^2 > 0$, $Nh^5 = O(1)$; (ii) $\underline{\lim}_{m\to\infty} m^{4+3\gamma} \sum_{t=m}^{\infty} t^2[\varphi(t)]^{\kappa/(2+\kappa)} < \infty$; and (iii) $N\psi(1, N)\varphi(c_N) \to 0$, as $N \to \infty$, where φ and ψ are defined in (3.5).

Note that (i) of Assumption (M) is standard and mild for this kind of nonparametric marginal density estimation. Assumption (M, ii) is a strengthened version of Assumption (I, i) on the mixing coefficient $\varphi(\cdot)$. The function $\psi(\cdot, \cdot)$ in Assumption (M, iii) can take the forms such as $\psi(n', n'') \le \min(n', n'')$ and $\psi(n', n'') \le C(n' + n'' + 1)^\tau$ for some $C > 0$ and $\tau > 1$, in the mixing conditions used by Neaderhouser (1980) and Takahata (1983), respectively. Under these forms of $\psi(\cdot, \cdot)$, we can easily derive some specific conditions on the mixing coefficient $\varphi(\cdot)$ to ensure Assumption (M, iii). For example, if $\psi(n', n'') \le \min(n', n'')$, then Assumption (M, iii) is guaranteed by $N\varphi(c_N) \to 0$ as $N \to \infty$, while if $\psi(n', n'') \le C(n' + n'' + 1)^\tau$, Assumption (M, iii) is ensured by $N^{\tau+1}\varphi(c_N) \to 0$ as $N \to \infty$, which only requires $\varphi(c_N)$ decreasing to zero at some reasonably fast convergence rates.

We can then state the following theorem.

Theorem 1. Under Assumptions (I), (II)(i, ii), (III)(i), (IV)(a), and (M), we have

$$(Nh)^{1/2}\left(\hat{f}(x) - f(x) - \frac{1}{2}\mu_{K,2}\ddot{f}(x)h^2\right) \Rightarrow N(0, \nu_K f(x)).$$

We will only sketch the proof for (4.7) and Theorem 1 in Section 9.1 as the detailed proof for this theorem is similar to that of the asymptotic normality for the joint density estimator in Theorem 4 in Section 5.

5. ASYMPTOTICS FOR THE JOINT DENSITY FUNCTION ESTIMATOR

5.1 Asymptotic Bias

We first state a theorem concerning the asymptotic bias of $\hat{f}(x, y; \mathbf{s}_0)$.

Theorem 2. Under Assumptions (I), (II), (III), and (IV)(b), the bias of the estimator $\hat{f}$ defined in (2.2) satisfies

$$B(\hat{f}(x, y; \mathbf{s}_0)) = \frac{1}{2}\{B_1(x, y, \mathbf{s}_0)b^2 + B_2(x, y, \mathbf{s}_0)h^2\}(1 + o(1)),$$

as $N \to \infty$, where

$$B_1(x, y, \mathbf{s}_0) = 2\frac{\partial f(x, y, \mathbf{s}_0)}{\partial \mathbf{s}^\tau}\boldsymbol{\mu}_{L,2}\frac{\mathbf{A}_1(\mathbf{s}_0)}{A_0(\mathbf{s}_0)} + \mathrm{tr}\left(\frac{\partial^2 f(x, y, \mathbf{s}_0)}{\partial \mathbf{s}\mathbf{s}^\tau}\boldsymbol{\mu}_{L,2}\right),$$

and

$$B_2(x, y, \mathbf{s}_0) = \left[\frac{\partial^2 f(x, y, \mathbf{s}_0)}{\partial x^2} + \frac{\partial^2 f(x, y, \mathbf{s}_0)}{\partial y^2}\right]\mu_{K,2}.$$

Remark. The asymptotic bias in Theorem 2 looks a bit cumbersome. It consists of the biases from the two kernel smoothings over the values of $Y_i = Y(\mathbf{s}_i)$ and over the values of $\mathbf{s}_i$, respectively. However, note that we can take $b/h \to 0$ as $N \to \infty$, under which Theorem 2 reduces to the following corollary, where the bias expression is simple, only owing to the smoothing over the values of $Y_i = Y(\mathbf{s}_i)$.

Corollary 1. Under the conditions of Theorem 2, if furthermore $b/h \to 0$ as $N \to \infty$, then the bias of the estimator $\hat{f}$ defined in (2.2) satisfies

$$B(\hat{f}(x, y; \mathbf{s}_0)) = \frac{1}{2}B_2(x, y, \mathbf{s}_0)h^2(1 + o(1)),$$

as $N \to \infty$.

5.2 Asymptotic Variance

The next theorem establishes the asymptotic variance of $\hat{f}(x, y; \mathbf{s}_0)$.

Theorem 3. Under Assumptions (I), (II), (III), and (IV)(b), the variance of the estimator $\hat{f}$ defined in (2.2) satisfies

$$\mathrm{var}(\hat{f}) = (Nbh)^{-2}V_1(x, y, \mathbf{s}_0)(1 + o(1)),$$

as $N \to \infty$, where

$$V_1(x, y, \mathbf{s}_0) = \frac{f(x, y, \mathbf{s}_0)(\nu_K)^2\nu_L}{A_0(\mathbf{s}_0)}.$$

We remark that for a kernel joint density estimator of (Y_i, Y_j) without using spatial smoothing, its asymptotic variance shall be of order $O\{(Nh^2)^{-1}\}$ as in the usual time series case. Comparing this with the asymptotic variance in Theorem 3, it appears reasonable to take $b = CN^{-1/2}$ with $0 < C < \infty$ independent of N, which again reconfirms $b/h \to 0$ in Corollary 1 if $Nh^2 \to \infty$.

5.3 Asymptotic Normality

Before stating the asymptotic normality of our joint density estimators, we also need

Assumption (J): Let $c_N = \{Nb^2\delta_N^{-2}h^{-2\kappa/(2+\kappa)}\}^{1/\gamma}$, which tends to ∞ as $N \to \infty$. (i) $c_N(Nbh\delta_N)^{-1} \to 0$; (ii) $\overline{\lim}_{m\to\infty} m^{4+3\gamma}\sum_{t=m}^{\infty} t^2[\varphi(t)]^{\kappa/(2+\kappa)} < \infty$; and (iii) $N\psi(1, N)\varphi(c_N) \to 0$, as $N \to \infty$, where φ and ψ are defined in (3.5).

We remark that Assumption (IV, b, iv) in the above guarantees $c_N \to \infty$ in Assumption (J). Assumption (J, i) implies that $Nbh\delta_N \to \infty$ and hence the infilling distance δ_N should not decrease to zero too fast with regard to the sample size, which is understandable. Assumption (J, ii) and Assumption (J, iii) are of the same forms as (ii) and (iii) of Assumption (M) for marginal density estimator.

We now can state our main theorem on asymptotic normality as follows.

Theorem 4. Under Assumptions (I), (II), (III), (IV) and (J), we have

$$(Nhb)(\hat{f}(x, y; \mathbf{s}_0) - f(x, y; \mathbf{s}_0) - \frac{1}{2}\{B_1(x, y, \mathbf{s}_0)b^2 + B_2(x, y, \mathbf{s}_0)h^2\}) \Rightarrow N(0, V_1(x, y, \mathbf{s}_0)),$$

as $N \to \infty$, where $\Rightarrow$ stands for the convergence in distribution, $B_1(x, y, \mathbf{s}_0)$ and $B_2(x, y, \mathbf{s}_0)$ are defined in Theorem 2, and $V_1(x, y, \mathbf{s}_0)$ in Theorem 3.

The proof of the asymptotic normality of our estimators in Theorem 4 relies on Theorems 2 and 3 in the above and follows directly from the following proposition.

Proposition 1. Suppose that Assumptions (I), (II), (III), (IV)(b), and (J) hold. Then $(Nhb)(\hat{f}(x, y; \mathbf{s}_0) - \mathrm{E}\hat{f}(x, y; \mathbf{s}_0)]/\sigma)$ is asymptotically standard normal as $N \to \infty$, where $\sigma = \sqrt{V_1(x, y, \mathbf{s}_0)}$, with $V_1(x, y, \mathbf{s}_0)$ defined in Theorem 3.

To prove the asymptotic normality in this proposition, as usual we will need to express $\hat{f}(x, y; \mathbf{s}_0) - \mathrm{E}\hat{f}(x, y; \mathbf{s}_0$ as a finite summation, S_N, of a dependent process as indicated in (9.36) below. However, we note that in the double summations that arise in the definition of $\hat{f}(x, y; \mathbf{s}_0)$ in (2.2) the terms are irregularly positioned in space and this makes it challenging and much more involved to derive the asymptotic normality. In the case where the observations of a stochastic process are regularly positioned in time or space, one can often use Bernstein's method by decomposing the finite summation S_N into the summations of smaller pieces involving "large" and "small" blocks, respectively; see, for example, Tran (1990) and Hallin, Lu, and Tran (2004b) in the spatial case with spatial lattice observations. However, unlike the case in Hallin, Lu, and Tran (2004b, sec. 5.3), we are dealing with the observations that are irregularly positioned in space, where it is difficult to construct "large" and "small" blocks.

We are here adopting an alternative device due to Stein (1972) and Bolthausen (1982) to derive a central limit theorem, in which one does not need to partition the summation over all spatial sites into the "large" and "small" blocks according to the spatial site indexes. This device is particularly useful for the irregularly positioned observations. It has been recently used by Jenish and Prucha (2009), who are concerned with establishing a central limit theorem of a single summation for an α-mixing spatial process under an increasing-domain only asymptotics. Differently, we are developing the asymptotics, by using this tool, in a context that is more involved with double summations in (2.2) under a more complex framework of infill asymptotics together with an increasing domain. This tool is also different from that used in Hall and Patil (1994), who use U-statistic technique by assuming spatial sites being iid randomly distributed over space. We will fully use the bounded support of the kernel function $L(\cdot)$ used in spatial smoothing to deal with the double summations in (2.2). Proofs are given in Section 9.4.

6. ESTIMATION OF THE ISOTROPIC JOINT PROBABILITY DENSITY FUNCTION

In the above sections, we were concerned with the estimation of $f(x, y; \mathbf{s}_0)$, which allows anisotropy in $\mathbf{s}_0$. However, if this joint probability density is isotropic, that is, $f(x, y; \mathbf{s}_0)$ is only dependent on $\|\mathbf{s}_0\|$, denoted as $f(x, y; \|\mathbf{s}_0\|)$, then we can construct an improved estimator as follows:

$$\hat{f}(x, y; \|\mathbf{s}_0\|) = \frac{\sum_{j,\ell=1}^{N} L_b(\|\mathbf{s}_j - \mathbf{s}_\ell\| - \|\mathbf{s}_0\|)K_h(Y_j - x)K_h(Y_\ell - y)}{\sum_{j,\ell=1}^{N} L_b(\|\mathbf{s}_j - \mathbf{s}_\ell\| - \|\mathbf{s}_0\|)}, \tag{6.8}$$

where $L_b(x) = b^{-1}L(x/b)$ with $L(\cdot)$ a kernel function on $\mathbb{R}^1$ and $b = b_N \to 0$ $(N \to \infty)$ a bandwidth.

Here are some adaptations of the assumptions in Section 3.

Assumption (I′): Assumption (I) holds true with (ii) replaced by (ii′): Denote by $f(x, y; \|\mathbf{s}_0\|)$ the joint density function of $Y(\mathbf{s})$ and $Y(\mathbf{s} + \mathbf{s}_0)$, where $\mathbf{s}_0 \neq (0, 0)$. $f(x, y; s)$ is continuous as a function of (x, y) uniformly with respect to $s \in \mathbb{R}$, and has second-order partial derivatives with respect to x, y and s, which are all continuous.

Assumption (II′): Assumption (II) holds true with (iii) replaced by (iii′):

$$A_0^*(\|\mathbf{s}_0\|) = \|\mathbf{s}_0\| \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\int_0^{2\pi} f_{\mathbf{S}}(u + \|\mathbf{s}_0\|\cos(\vartheta), v + \|\mathbf{s}_0\|\sin(\vartheta))d\vartheta\right] f_{\mathbf{S}}(u, v)\,du\,dv > 0$$

and

$$A_1^*(\|\mathbf{s}_0\|) = \|\mathbf{s}_0\| \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\int_0^{2\pi} (\cos(\vartheta), \sin(\vartheta)) \times \frac{\partial f_{\mathbf{S}}(u + \|\mathbf{s}_0\|\cos(\vartheta), v + \|\mathbf{s}_0\|\sin(\vartheta))}{\partial \mathbf{s}} d\vartheta\right] \times f_{\mathbf{S}}(u, v)du\,dv$$

exist.

Assumption (III′): Assumption (III) holds true with (ii) replaced by (ii′): The kernel function $L(\cdot)$ satisfies that $\int_{\mathcal{R}} L(s)ds = 1$, $\int_{\mathcal{R}} sL(s)ds = \mathbf{0}$, and $\mu_{L,2} = \int_{\mathcal{R}} s^2L(s)ds < \infty$, $\nu_L = \int_{\mathcal{R}} L^2(s)ds < \infty$.

Assumption (IV′): As $N \to \infty$, (i) $h \to 0$, (ii) $b \to 0$, (iii) $Nhb^{1/2} \to \infty$, $Nhb \to 0$, (iv) $Nb/\delta_N^2 h^{-2\kappa/(2+\kappa)} \to \infty$, $Nb/\delta_N^2 h^{2(\gamma - 2\kappa/(2+\kappa))/(2+\gamma)} \to 0$, where $\gamma > 2\kappa/(2+\kappa)$ and $\kappa > 0$ were defined in Assumption (I)(i).

Theorem 5. Under Assumptions (I′), (II′), (III′), and (IV′), the bias of the estimator $\hat{f}$ defined in (6.8) satisfies

$$B(\hat{f}(x, y; \|\mathbf{s}_0\|)) = \frac{1}{2}\{B_1^*(x, y, \|\mathbf{s}_0\|)b^2 + B_2^*(x, y, \|\mathbf{s}_0\|)h^2\}(1 + o(1)),$$

as $N \to \infty$, where

$$B_1^*(x, y, \|\mathbf{s}_0\|) = 2\frac{\partial f(x, y; \|\mathbf{s}_0\|)}{\partial s}\mu_{L,2}\frac{\{A_1^*(\|\mathbf{s}_0\|) + A_0^*(\|\mathbf{s}_0\|)/\|\mathbf{s}_0\|\}}{A_0^*(\|\mathbf{s}_0\|)} + \left(\frac{\partial^2 f(x, y; \|\mathbf{s}_0\|)}{\partial s^2}\mu_{L,2}\right),$$

and

$$B_2^*(x, y, \|\mathbf{s}_0\|) = \left[\frac{\partial^2 f(x, y; \|\mathbf{s}_0\|)}{\partial x^2} + \frac{\partial^2 f(x, y; \|\mathbf{s}_0\|)}{\partial y^2}\right]\mu_{K,2}.$$

Theorem 6. Under Assumptions (I$'$), (II$'$), (III$'$), and (IV$'$), the variance of the estimator $\hat{f}$ defined in (2.2) satisfies

$$\mathrm{var}(\hat{f}) = (N^2bh^2)^{-1}V_1^*(x, y, \|\mathbf{s}_0\|)(1 + o(1)),$$

as $N \to \infty$, where

$$V_1^*(x, y, \|\mathbf{s}_0\|) = \frac{f(x, y; \|\mathbf{s}_0\|)(\nu_K)^2\nu_L}{A_0^*(\|\mathbf{s}_0\|)}.$$

We need the following further assumption for asymptotic normality.

Assumption (V$'$): Let $c_N = \{Nb\delta_N^{-2}h^{-2\kappa/(2+\kappa)}\}^{1/\gamma}$, which tends to ∞ as $N \to \infty$. (i) $c_N(Nb^{1/2}h\delta_N)^{-1} \to 0$; (ii) $\lim_{m\to\infty} m^{4+3\gamma}\sum_{t=m}^{\infty} t^2[\varphi(t)]^{\kappa/(2+\kappa)} = 0$; and (iii) $N\psi(1, N)\varphi(c_N) \to 0$, as $N \to \infty$.

Theorem 7. Under Assumptions (I$'$), (II$'$), (III$'$), (IV$'$), and (V$'$), we have

$$\left(Nhb^{1/2}\right)\left(\hat{f}(x, y; \|\mathbf{s}_0\|) - f(x, y; \|\mathbf{s}_0\|) - \frac{1}{2}\{B_1^*(x, y, \|\mathbf{s}_0\|)b^2 + B_2^*(x, y, \|\mathbf{s}_0\|)h^2\}\right) \Rightarrow N(0, V_1^*(x, y, \|\mathbf{s}_0\|)),$$

as $N \to \infty$, where $B_1^*(x, y, \mathbf{s}_0)$ and $B_2^*(x, y, \mathbf{s}_0)$ are defined in Theorem 5, and $V_1^*(x, y, \mathbf{s}_0)$ in Theorem 6.

7. ISSUES OF BANDWIDTH SELECTION

In this section we suggest a practical rule for bandwidth selection.

We are applying the cross-validation principle (see, Stone 1974) to select the bandwidths b and h in $\hat{f}_{b,h}(x, y; \mathbf{s}_0) := \hat{f}(x, y; \mathbf{s}_0)$ defined in (2.2). To exclude the extreme values at which the density is poorly estimated, we look at the estimators of the density function at the points, $(x, y) \in [a, A] \times [a, A]$, where a and A are chosen appropriately, say as 5% and 95% sample quantiles of Y_i's, respectively. As done in the literature for marginal kernel density estimation (see Li and Racine 2007, sec. 1.3), we examine the main terms related to the bandwidths in the integrated squared error

$$\begin{aligned}\mathrm{ISE}(b, h; \mathbf{s}_0) &= \int_a^A\int_a^A \left[\hat{f}_{b,h}(x, y; \mathbf{s}_0) - f(x, y; \mathbf{s}_0)\right]^2 dxdy \\ &= \int_a^A\int_a^A \left[\hat{f}_{b,h}^2(x, y; \mathbf{s}_0) - 2\hat{f}_{b,h}(x, y; \mathbf{s}_0)f(x, y; \mathbf{s}_0) + f^2(x, y; \mathbf{s}_0)\right]dxdy \qquad (7.9)\end{aligned}$$

by considering a leave-two-out CV as follows:

$$\begin{aligned}\mathrm{CV}(b, h; \mathbf{s}_0) = \int_a^A\int_a^A \hat{f}_{b,h}^2(x, y; \mathbf{s}_0)dxdy &- \frac{2}{n_0}\sum_{k=1}^{N}\sum_{\substack{\ell=1\\ \|\mathbf{s}_k-\mathbf{s}_\ell-\mathbf{s}_0\|^2\le\epsilon}}^{N} \hat{f}_{b,h,-(k,\ell)}(Y_k, Y_\ell; \mathbf{s}_0) \\ &\times I_{\{a\le Y_k\le A\}}I_{\{a\le Y_\ell\le A\}},\end{aligned} \qquad (7.10)$$

where n_0 is the number of the pairs (k, ℓ) such that $\|\mathbf{s}_k - \mathbf{s}_\ell - \mathbf{s}_0\|^2 \le \epsilon$ with $a \le Y_k, Y_\ell \le A$. Here $\|\cdot\|$ is the Euclidean norm in R^2, and $\epsilon > 0$ is a small number that needs to be specified (see below). Further, $\hat{f}_{b,h,-(k,\ell)}(Y_k, Y_\ell; \mathbf{s}_0)$ is the leave-two-out estimator of $f(Y_k, Y_\ell; \mathbf{s}_0)$ in the form

$$\hat{f}_{b,h,-(k,\ell)}(Y_k, Y_\ell; \mathbf{s}_0) = \frac{\sum_{i\in S_{k\ell}, j\in S_{k\ell}} L_b(\mathbf{s}_i - \mathbf{s}_j - \mathbf{s}_0)K_h(Y_i - Y_k)K_h(Y_j - Y_\ell)}{\sum_{i\in S_{k\ell}, j\in S_{k\ell}} L_b(\mathbf{s}_i - \mathbf{s}_j - \mathbf{s}_0)}, \qquad (7.11)$$

where $S_{k\ell} = \{i : 1 \le i \le N,\ i \ne k,\ i \ne \ell)\}$.

For the calculation in the first term on the right-hand side of (7.10), although we may compute the double integral $\int_a^A\int_a^A \hat{f}_{b,h}^2(x, y; \mathbf{s}_0)dxdy$ as done in Li and Racine (2007, sec. 1.3), it would lead to four-fold summations from 1 to N, the calculation of which becomes time-demanding when the sample size N is not small, say $N > 100$. To simplify the calculation, we suggest approximating the double integration by applying numerical integration as follows:

$$\int_a^A\int_a^A \hat{f}_{b,h}^2(x, y; \mathbf{s}_0)dxdy \approx \sum_{u=1}^{M}\sum_{v=1}^{M} \hat{f}_{b,h}^2(x_u, y_v; \mathbf{s}_0)\delta^2,$$

where $\delta = (A - a)/M$ and $x_u = y_u = a + (u - 1)\delta$ for $u, v = 1, 2, \ldots, (M + 1)$. In application, we may take, say, $M = 30$ to simplify the calculation. For the second term on the RHS of (7.10), the calculation can be greatly simplified by considering only those pairs of (k, ℓ) satisfying $\|\mathbf{s}_k - \mathbf{s}_\ell - \mathbf{s}_0\|^2 \le \epsilon$, where ϵ may in general be dependent on N.

In the empirical examples in the next section, we simply take $\epsilon = 0.01$, which appears to work quite well.

Thus we have an approximate CV

$$\begin{aligned}\mathrm{CV}(b, h; \mathbf{s}_0) = \sum_{u=1}^{M}\sum_{v=1}^{M} \hat{f}_{b,h}^2(x_u, y_v; \mathbf{s}_0)\delta^2 &- \frac{2}{n_0}\sum_{k=1}^{N}\sum_{\substack{\ell=1\\ \|\mathbf{s}_k-\mathbf{s}_\ell-\mathbf{s}_0\|^2\le\epsilon}}^{N} \hat{f}_{b,h,-(k,\ell)}(Y_k, Y_\ell; \mathbf{s}_0) \\ &\times I_{\{a\le Y_k\le A\}}I_{\{a\le Y_\ell\le A\}},\end{aligned} \qquad (7.12)$$

where n_0 is as before. We then select $(b_{\mathrm{opt}}, h_{\mathrm{opt}})$ minimizing $\mathrm{CV}(b, h; \mathbf{s}_0)$ over $b \in [b_L, b_U]$ and $h \in [h_L, h_U]$, where $0 < b_L < b_U$ and $0 < h_L < h_U$ are appropriately given.

The CV bandwidth selection as described in the above is relatively computationally intensive. To simplify the computation, we adapt an empirical rule for selecting a bandwidth due to Fan, Yao, and Cai (2003, p. 64) by incorporating the asymptotic outcome in Theorem 4 into determining the bandwidth; see also

Lu et al. (2009). Recalling (7.9) and Theorem 4, it follows that the integrated mean squared error

$$\begin{aligned}\mathrm{IMSE}(b, h; \mathbf{s}_0) &= E[\mathrm{ISE}(b, h; \mathbf{s}_0)]\\ &= \int_a^A \int_a^A E[\hat{f}_{b,h}(x, y; \mathbf{s}_0) - f(x, y; \mathbf{s}_0)]^2 dx dy\\ &\approx \bar{c}_0 + (\bar{c}_{11} b^2 + \bar{c}_{12} h^2)^2 + \bar{c}_2 (Nbh)^{-2},\end{aligned}$$

where $\bar{c}_0$, $\bar{c}_{11}$, $\bar{c}_{12}$ and $\bar{c}_2$ are constants. Notice that we cannot simultaneously select optimal b and h to minimize $\mathrm{IMSE}(b, h; \mathbf{s}_0)$. The reasons are as follows: First of all, the equation in the above, which indicates a kind of symmetric involvement of b and h in $\mathrm{IMSE}(b, h; s_0)$, is derived on the basis of the condition that $Nhb^2 \to 0$ (as $N \to \infty$) in Assumption (IV, b, iii), which is critical in the derivation of asymptotic bias (Theorem 5.1) and asymptotic variance (Theorem 5.3). Second, we notice that because of the condition $Nhb^2 \to 0$, b and h are actually not as symmetrically connected as perhaps indicated in $\mathrm{IMSE}(b, h; s_0)$. Mathematically speaking, we could look at the cases of $h/b \to 0$ and $h/b = O(1)$. The problem in these two cases is that the asymptotic bias of $\hat{f}(x, y, \mathbf{s}_0)$ would be larger than the bias order of $O(h^2)$ for usual kernel density estimation (see Theorem 5.1). Statistically speaking, it looks more reasonable to choose b with $b/h \to 0$ for bias reduction (see Corollary 1). Therefore, for a given b, for which we would asymptotically assume $b/h \to 0$ (as $N \to \infty$), up to first-order asymptotics, the optimal bandwidth for h is thus $h^b_{\mathrm{opt}} = \{c_2/[2(Nb)^2 c_1]\}^{1/6}$, minimizing

$$\begin{aligned}\mathrm{CV}_b(h) = \mathrm{CV}(b, h; \mathbf{s}_0) &= c_0 + c_1 h^4 + c_2 (Nbh)^{-2}\\ &\quad + o_P\{h^4 + (Nb)^{-2} h^{-2}\},\end{aligned}$$

where c_0, c_1, and c_2 are constants that may possibly depend on b.

In practice, we may partition $[b_L, b_U]$ and $[h_L, h_U]$ into r points, $b_1, b_2, \ldots, b_r$, and q points, $h_1, h_2, \ldots, h_q$, respectively. For a given $b = b_\ell$, $\ell = 1, \ldots, r$, the coefficients c_0, c_1, and c_2 will be estimated from $\mathrm{CV}_b(h_k)$, $k = 1, 2, \ldots, q$, via least squares regression,

$$\min_{c_0, c_1, c_2} \sum_{k=1}^{q} \left\{\mathrm{CV}_b(h_k) - c_0 - c_1 h_k^4 - c_2/(Nbh_k)^2\right\}^2, \quad (7.13)$$

where $\mathrm{CV}_b(h_k) = \mathrm{CV}(b, h_k; \mathbf{s}_0)$ is obtained from (7.12), and thus we let $h^b_{\mathrm{opt}} = \{\hat{c}_2/(2(Nb)^2 \hat{c}_1)\}^{1/6}$ when both $\hat{c}_1$ and $\hat{c}_2$, the estimators of c_1 and c_2, are positive, and we calculate the corresponding $\mathrm{CV}_b(h^b_{\mathrm{opt}}) = \hat{c}_0 + \hat{c}_1 (h^b_{\mathrm{opt}})^4 + \hat{c}_2 (Nbh^b_{\mathrm{opt}})^{-2}$. In the unlikely event that one of them is nonpositive, we let $h^b_{\mathrm{opt}} = \arg\min_{1 \le k \le q} \mathrm{CV}_b(h_k)$ and the corresponding minimum for $\mathrm{CV}_b(h^b_{\mathrm{opt}})$. Thus we can select

$$b_{\mathrm{opt}} = \arg \min_{1 \le \ell \le r} \mathrm{CV}_{b_\ell}\left(h^{b_\ell}_{\mathrm{opt}}\right), \text{ and } h_{\mathrm{opt}} = h^{b_{\mathrm{opt}}}_{\mathrm{opt}}.$$

This bandwidth selection procedure will be applied below. It is computationally efficient as r and q are moderately small, that is, we only need to compute (rq) CV-values; see remark 2(c) in Fan, Yao, and Cai (2003). Furthermore, the least squares estimation (7.13) also serves as a smoother for the CV bandwidth estimates. Also see Ruppert (1997).

8. NUMERICAL EVIDENCES

8.1 Simulation

In this section, we show the performance of our proposed estimation of the marginal and joint probability density functions. To evaluate the performance of our estimation procedure, we need to use a spatial process $\{Y(\mathbf{s})\}$, where the theoretical joint probability density function of $Y(\mathbf{s}_i)$ and $Y(\mathbf{s}_i + \mathbf{s}_0)$, $\mathbf{s}_0 \neq 0$, can be computed analytically. Therefore, we are considering a special spatial process, $Y(\mathbf{s}_i)$, that is generated through a mixture of Gaussian spatial moving average processes as follows:

Step 1 (Generating locations): Generate the locations irregularly positioned in R^2: Define a lattice (u_i, v_j) with $u_i = u_0 + (i-1)\delta$ and $v_j = v_0 + (j-1)\delta$, for $i = 1, \ldots, N$ and $j = 1, \ldots, N$. We take $\delta = 0.3$, and $u_0 = \delta$, $v_0 = 2\delta$. Then we randomly select N locations from the lattice as the irregular locations at which our observations are made. We denote these locations as $\mathbf{s}_k = (u_{i_k}, v_{j_k})$, $k = 1, \ldots, N$, which are fixed in the repetition of the simulation below, where $1 \le i_k, j_k \le N$.

Step 2 (Intermediate variables): Generate two intermediate processes $\tilde{Y}_{i,1}$ and $\tilde{Y}_{i,2}$ from two independent Gaussian spatial moving averages,

$$\tilde{Y}_{i,r} = \mu_r + \sum_{k=-1}^{1} \sum_{\ell=-1}^{1} a_{k,r} Z_{i-k,r}, \ 1 \le i \le m_1, \ r = 1, \ 2, \quad (8.14)$$

where $Z_{i,r}$, $r = 1, 2$, are two independent iid samples from $\mathcal{N}(0, \sigma^2_{Z,r})$, $r = 1, 2$, respectively. We take $\mu_1 = -1$ and $\mu_2 = 0.3$, let $a_{k,1}$ and $a_{k,2}$ be the $(k+2)$th elements of $a_1 = (1/5, 2/5, -4/5)$ and $a_2 = (-3/5, -2/5, -1/5)$, respectively, for $-1 \le k \le 1$, and $\sigma_{Z,1} = 0.67$, $\sigma_{Z,2} = 0.76$. Here the marginal distribution of $\tilde{Y}_{i,1}$ is Gaussian $\mathcal{N}(\mu_1 = -1, \sigma_1^2 = 0.3771)$, and $\tilde{Y}_{i,2}$ is $\mathcal{N}(\mu_2 = 0.3, \sigma_2^2 = 0.3235)$, where $\mathcal{N}(\mu, \sigma^2)$ stands for the univariate Gaussian distribution of mean μ and variance σ^2. We then generate spatial process by first generating independent $R_j \sim \mathrm{Binomial}(1, p = 0.4)$, $1 \le j \le m_2$, and then defining

$$\begin{aligned}\tilde{Y}_{ij} &= \tilde{Y}_{i,1} \times R_j\\ &\quad + \tilde{Y}_{i,2} \times (1 - R_j), \ 1 \le i \le m_1, \ 1 \le j \le m_2, \quad (8.15)\end{aligned}$$

where we take $m_1 = m_2 = N$, and R_j's are independent of $Z_{i,r}$'s and hence of $\tilde{Y}_{i,r}$'s, with $r = 1, 2$.

Step 3 (Final observations): We define our final observations, for $\mathbf{s}_k = (u_{i_k}, v_{j_k})$ in Step 1,

$$Y_k = Y(\mathbf{s}_k) = \tilde{Y}_{i_k j_k}, \ k = 1, \ldots, N. \quad (8.16)$$

Thus Y_k's are our simulated observations on the irregular positions $\mathbf{s}_k$, $k = 1, \ldots, N$. Note that the marginal distribution of Y_k is a mixture of normal distributions, in the form

$$\begin{aligned}&0.4 \times \mathcal{N}(\mu_1 = -1, \sigma_1^2 = 0.3771)\\ &\qquad + 0.6 \times \mathcal{N}(\mu_2 = 0.3, \sigma_2^2 = 0.3235).\end{aligned}$$

We generate the simulated spatial data by using the values of the parameters in the above models. Note that our simulated data satisfy the α-mixing assumption, which clearly follows from the simulating models (8.14), (8.15), and (8.16) with the

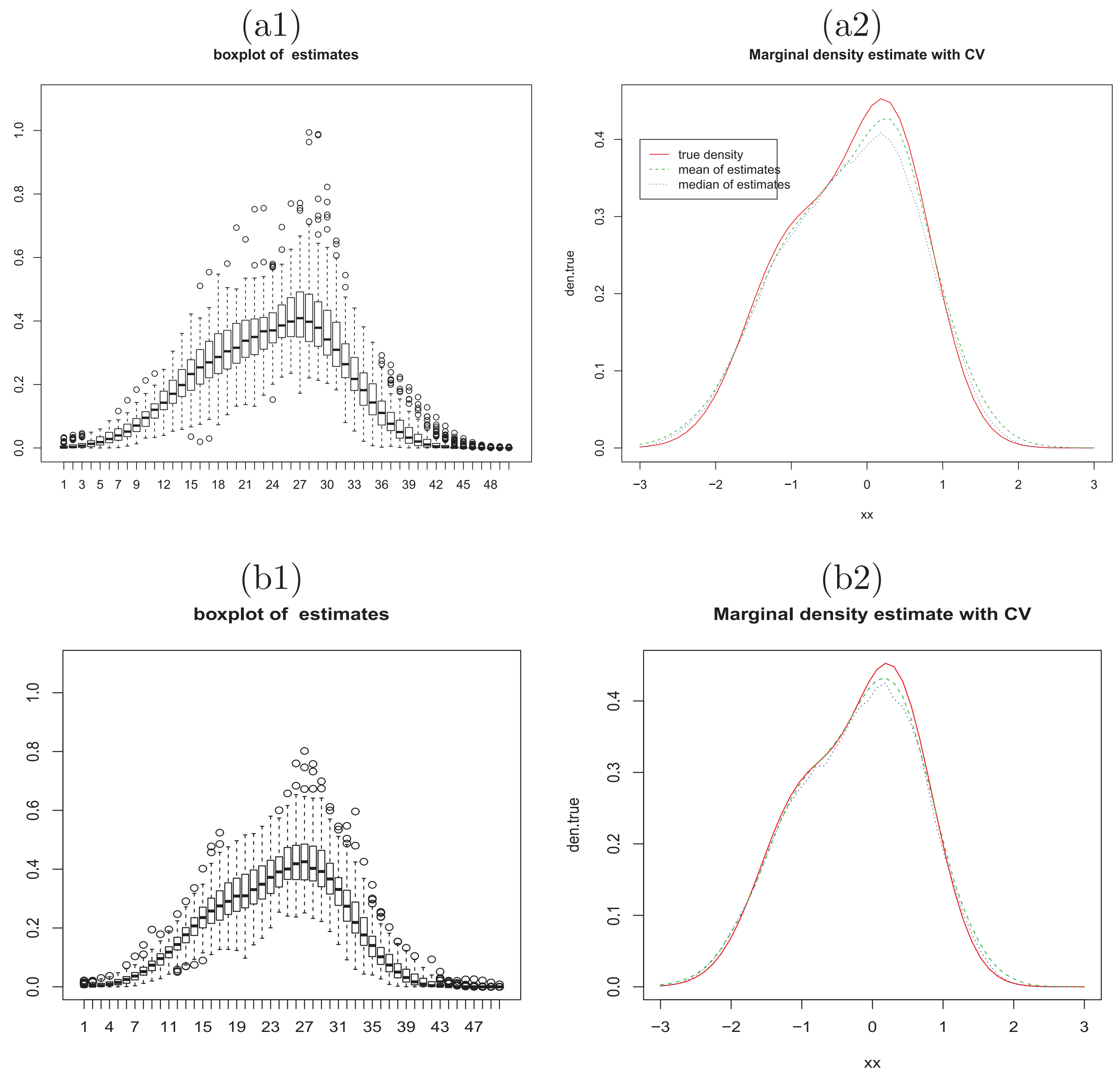

Figure 1. The comparison of the estimated marginal probability density functions with different sample sizes: (a1, a2) $N = 200$, (b1, b2) $N = 500$. Here the left column is for the boxplots of the density estimates, while the right one is for the comparisons of the actual density with the median of the density estimates, of 100 times of simulations.

fact that $Z_{i,1}$, $Z_{i,2}$, and R_j are independent iid processes, which are α-mixing. Then the resultant marginal and the joint density functions considered below are asymmetric, non-Gaussian with one and two modes in marginal and joint density functions, respectively, these features being similar to those of empirical marginal and joint ones, in the example of Tokyo land price below. We are considering the cases of $N = 200$ and $N = 500$, respectively. We repeat the simulation 100 times. The marginal density estimate of $Y(\mathbf{s}_k)$ is depicted in Figure 1, where the left column is for the boxplots of the density estimates (using the *R* function "density" with "bcv" for bandwidth) of the 100 simulated samples while the right column is for the comparisons of the actual density with the median and mean of the density estimates of 100 simulations. Obviously, the simulation result for the marginal distribution is satisfying.

We turn to the joint distribution. Let's look at the joint density $f(x, y; \delta)$ at two sites, $\mathbf{s}$ and $\mathbf{s} + \mathbf{s}_0$, of distance $\mathbf{s}_0 = (\delta, 0) = (0.3, 0)$. By (8.14)–(8.16), we can calculate to obtain the actual density

$$f(x, y; \delta = 0.3) = p\mathcal{N}_2(x, y; (\mu_1, \mu_1), \Sigma_1) + (1 - p)\mathcal{N}_2(x, y; (\mu_2, \mu_2), \Sigma_2),$$

where $\mathcal{N}_2(x, y; (\mu, \mu), \Sigma)$ stands for the two-dimensional joint Gaussian density function of mean (μ, μ) and variance matrix Σ, and Σ_1 is a 2×2 variance-covariance matrix with elements $\sigma_{1,11} = \sigma_{1,22} = 0.3771$ and $\sigma_{1,12} = \sigma_{1,21} = -0.1077$, and Σ_2 is a 2×2 variance-covariance matrix with elements $\sigma_{2,11} = \sigma_{2,22} = 0.3235$ and $\sigma_{2,12} = \sigma_{2,21} = 0.1848$. We also consider the joint density $f(x, y; 3\delta)$ at two sites of distance $\mathbf{s}_0 = (3\delta, 0) = (0.9, 0)$. By (8.15), we can deduce that, owing to the independence of the $Z_{i,r}$'s with $r = 1, 2$, the actual density is

$$f(x, y; 3\delta = 0.9) = p\mathcal{N}_2(x, y; (\mu_1, \mu_1), \Sigma_1) + (1 - p)\mathcal{N}_2(x, y; (\mu_2, \mu_2), \Sigma_2),$$

where Σ_1 is a 2×2 variance-covariance matrix with elements $\sigma_{1,11} = \sigma_{1,22} = 0.3771$ and $\sigma_{1,12} = \sigma_{1,21} = 0$, and Σ_2 with elements $\sigma_{2,11} = \sigma_{2,22} = 0.3235$ and $\sigma_{2,12} = \sigma_{2,21} = 0$. In the simulations, we estimate $f(x, y; 0.3)$ and $f(x, y; 0.9)$ at $x = x_k$ and $y = y_\ell$, for $1 \le k, \ell \le 50$, where $x_k = y_k = a + (k - 1)(A - a)/49$, for $k = 1, 2, \ldots, 50$, is a partition of $[a, A]$ into 49 subintervals in the bandwidth selection as described in Section 7. With some initial experiments, we took $a = -3$ and $A = 3$, and $b_L = 0.1$, $b_U = 1$, $h_L = 0.2$ and $h_U = 0.7$ together with $r = q = 5$ for easy implementation of selection of the bandwidths, which appear to work quite well in the simulation. The contour plots of the theoretical and the average, based on 100 times of simulations, of the estimated density functions for $f(x, y; 0.3)$ and $f(x, y; 0.9)$ are given in Figures 2 and 3, respectively, where the boxplots of the averaged squared estimation errors

$$\text{ASEE} = \frac{1}{2500} \sum_{k=1}^{50} \sum_{\ell=1}^{50} (\hat{f}(x_k, y_\ell) - f(x_k, y_\ell))^2,$$

based on 100 times of simulations, are also provided for comparison. For both experiments the estimated marginal and bivariate distributions are close to the theoretical ones, and the error decreases with sample size. As in the ordinary kernel estimation, due to the smoothing, the peak of the density is slightly underestimated. The positive dependence between the x- and y-variables in experiments 1 and 2 are well taken care of.

8.2 Real Data Examples

We provide two real data examples to demonstrate the proposed joint density estimation.

8.2.1 Baltimore House Sales Price. We are analyzing a spatial dataset in the R package, *spdep*, that gives the locations with longitude and latitude together with house sales price and characteristics for a spatial hedonic analysis in Baltimore, MD 1978 (see Dubin 1992). This dataset consists of 211 observations on 17 variables, among which we are concerned with the following three variables in this article:

u = numeric latitude (in the notation of this article);
v = numeric longitude (in the notation of this article);
PRICE = numeric house sales price.

The spatial plot of the variable, logarithmic PRICE over latitude (u) and longitude (v) is provided in Panel (a) of Figure 4. Obviously the price dataset is irregularly positioned in space and it appears we can see a spatial trend of logarithmic price over space in Panel (a), indicating that the spatial trend of the logarithmic price data may need to be removed to obtain the observations approximately from a spatial stationary process. Panels (b) and (c) display the spatial trend of logarithmic price, fitted by sm.regression in the R package *SM*, and the residuals after removing the spatial trend, over (u, v), respectively. Clearly, the residual of the logarithmic house sales price looks more stationary in Panel (c). We are interested in examining the distribution function and spatial dependence of the residuals, that is, the logarithmic house sales price after removing spatial trend, denoted by $Y_i = Y(\mathbf{s}_i)$ below, where $\mathbf{s}_i = (u_i, v_i)$ for $i = 1, 2, \ldots, 211$.

The marginal kernel density estimate of $Y_i = Y(\mathbf{s}_i)$ by the R function "density" is plotted in solid line in Panel (d) of Figure 4 with dashed line for the normal density function of the same mean and variance. Clearly, it is neither normally distributed nor symmetric, with heavier LHS tail than that of the normal distribution on $Y_i = Y(\mathbf{s}_i)$. This implies that the spatial dependence measured by the spatial autocorrelation, the analysis of which was made by Dubin (1992), may not be adequate, and it is useful to estimate the joint distributions of $(Y(\mathbf{s}_i), Y(\mathbf{s}_i + \mathbf{s}_0))$ for $\mathbf{s}_0 \neq (0, 0)$.

We are estimating the joint probability density functions, $f(x, y; \mathbf{s}_0)$, of $(Y(\mathbf{s}_i), Y(\mathbf{s}_i + \mathbf{s}_0))$ for $\mathbf{s}_0 = (1, 1)$, $\mathbf{s}_0 = (5, 5)$, and $\mathbf{s}_0 = (9, 9)$, respectively, by applying the methodology proposed in this article. We use the cross-validation principle detailed in the last section with $(a, A) = (-1.9, 0.9)$, $M = 30$ and $\epsilon = 0.01$ in (7.12), to select the bandwidths (b, h). The estimated joint probability density functions for $(Y(\mathbf{s}_i), Y(\mathbf{s}_i + \mathbf{s}_0))$ are depicted in Figure 5, in which Panel (a) corresponds to $\mathbf{s}_0 = (1, 1)$ with CV-bandwidth $(b, h) = (0.9556, 0.1025)$, (b) to $\mathbf{s}_0 = (5, 5)$ with CV-bandwidth $(b, h) = (2.222, 0.1)$, and (c) to $\mathbf{s}_0 = (9, 9)$ with CV-bandwidth $(b, h) = (1.211, 0.2425)$, where each panel contains the contour plot of the estimated joint density function. Apparently, our estimated joint probability density functions give a good characterization of the nonlinear, non-Gaussian spatial dependence existent in the logarithmic house sales price (after removing spatial trend), which cannot be captured by spatial auto-correlation $\gamma(\|\mathbf{s}_0\|) = \text{corr}(Y(\mathbf{s}_i), Y(\mathbf{s}_i + \mathbf{s}_0))$, where the correlogram of $Y(\mathbf{s}_i)$ is plotted in Figure 6 by using the R package NCF. Whereas the correlogram characterizes (linear) dependence by just furnishing one number for each lag, the plots of Figure 5 give a complete distributional pattern for each lag. This can be used to further characterize dependence in terms of nonlinear dependence measures. Comparing to the simulations, it is clear that there is significant positive dependence at lag $\mathbf{s}_0 = (1, 1)$, while it is seen that we get closer to independence as the lag increases. A more formal and quantitative analysis can be undertaken but is outside the scope of the present article.

8.3 Tokyo Land Prices

Our second real example is a larger dataset from Matsuda and Yajima (2009), who considered the parametric and nonparametric spectral estimation to land price data, collected by the Japanese Ministry of Land, Infrastructure and Transport in 2001. This dataset is the record of land prices (yen per square meter) irregularly spaced in the residential areas around Tokyo. For details, see Figure 1(a) of Matsuda and Yajima (2009), where the coordinates are modified with units of kilometers. They selected 1431 points inside the rectangular region ($20 \le x \le 45$,

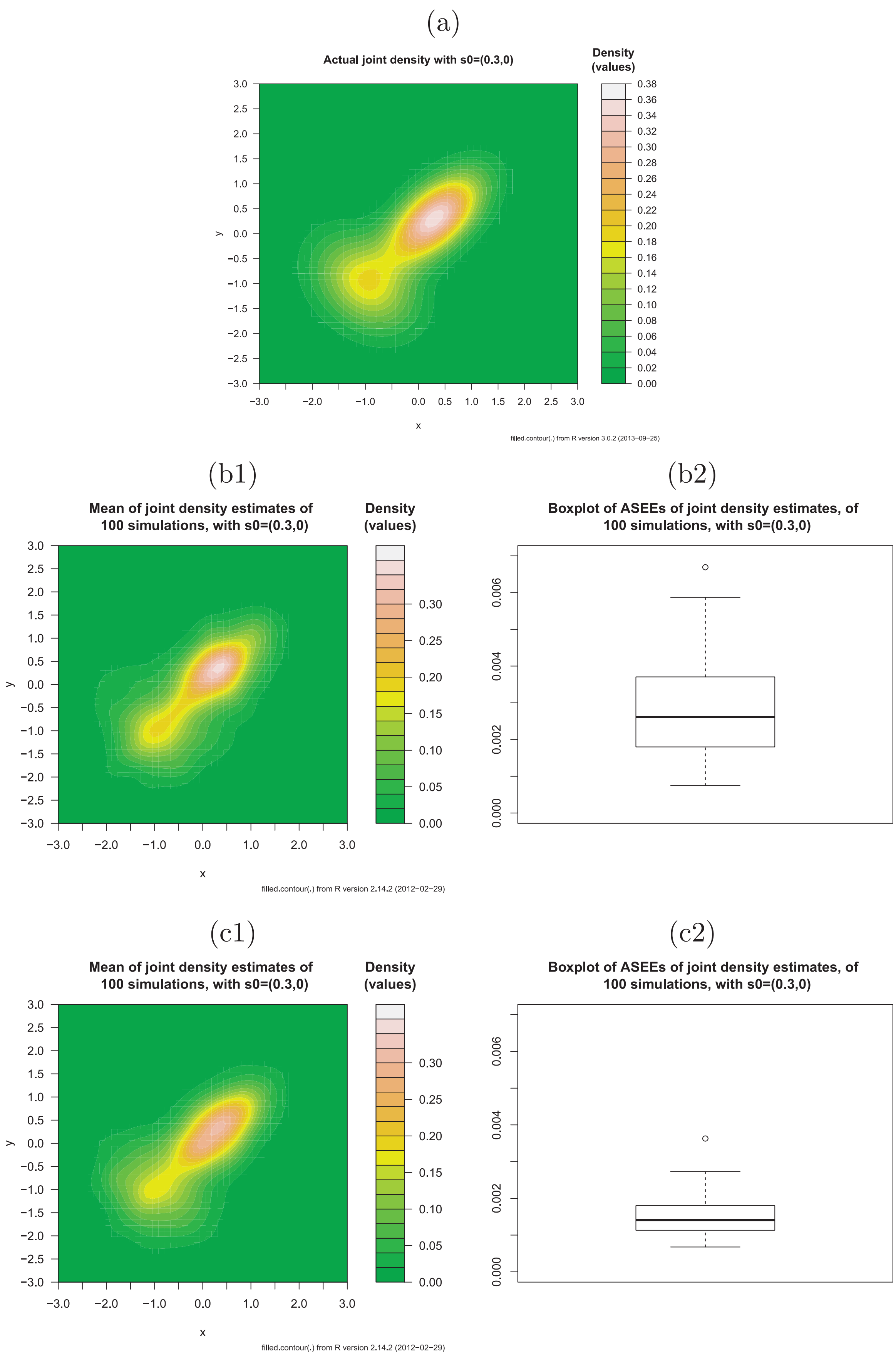

Figure 2. The comparison of the theoretical and the estimated joint probability density functions with distance $\mathbf{s}_0 = (0.3, 0)$, under different sample size N: (a) contour of theoretical density, (b1, b2) estimates with $N = 200$, (c1, c2) estimates with $N = 500$. Here in (b) and (c), the left column is for the contour of the mean of the density estimates, while the right one is for the boxplot of the averaged squared estimation errors of the density estimates, of 100 times of simulations.

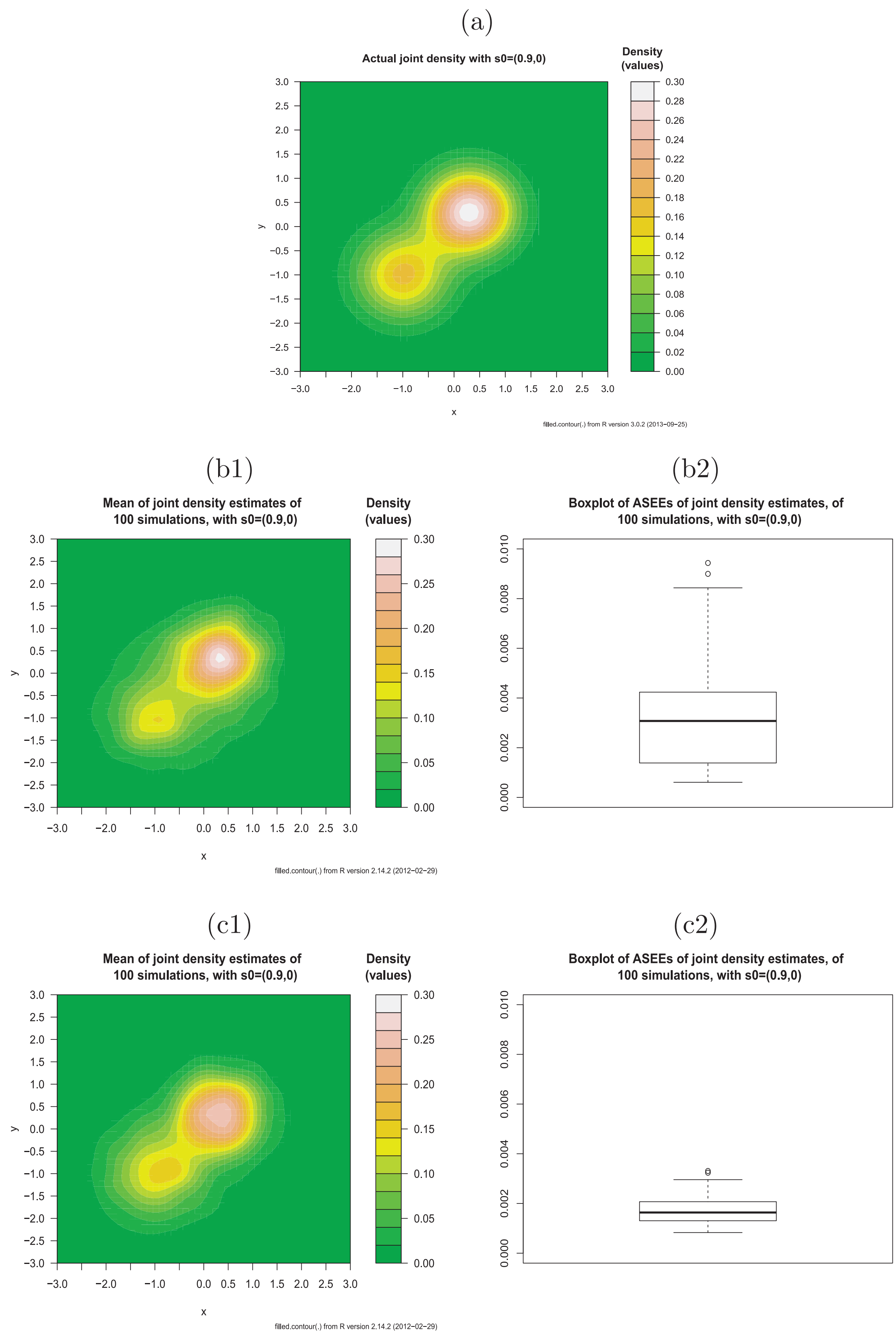

Figure 3. The comparison of the theoretical and the estimated joint probability density functions with distance $\mathbf{s}_0 = (0.9, 0)$, under different sample size N: (a) theoretical density, (b1), b2) $N = 200$, (c1, c2) $N = 500$. Here in (b) and (c), the left column is for the contour of the mean of the density estimates, while the right one is for the boxplot of the averaged squared estimation errors of the density estimates, of 100 times of simulations.

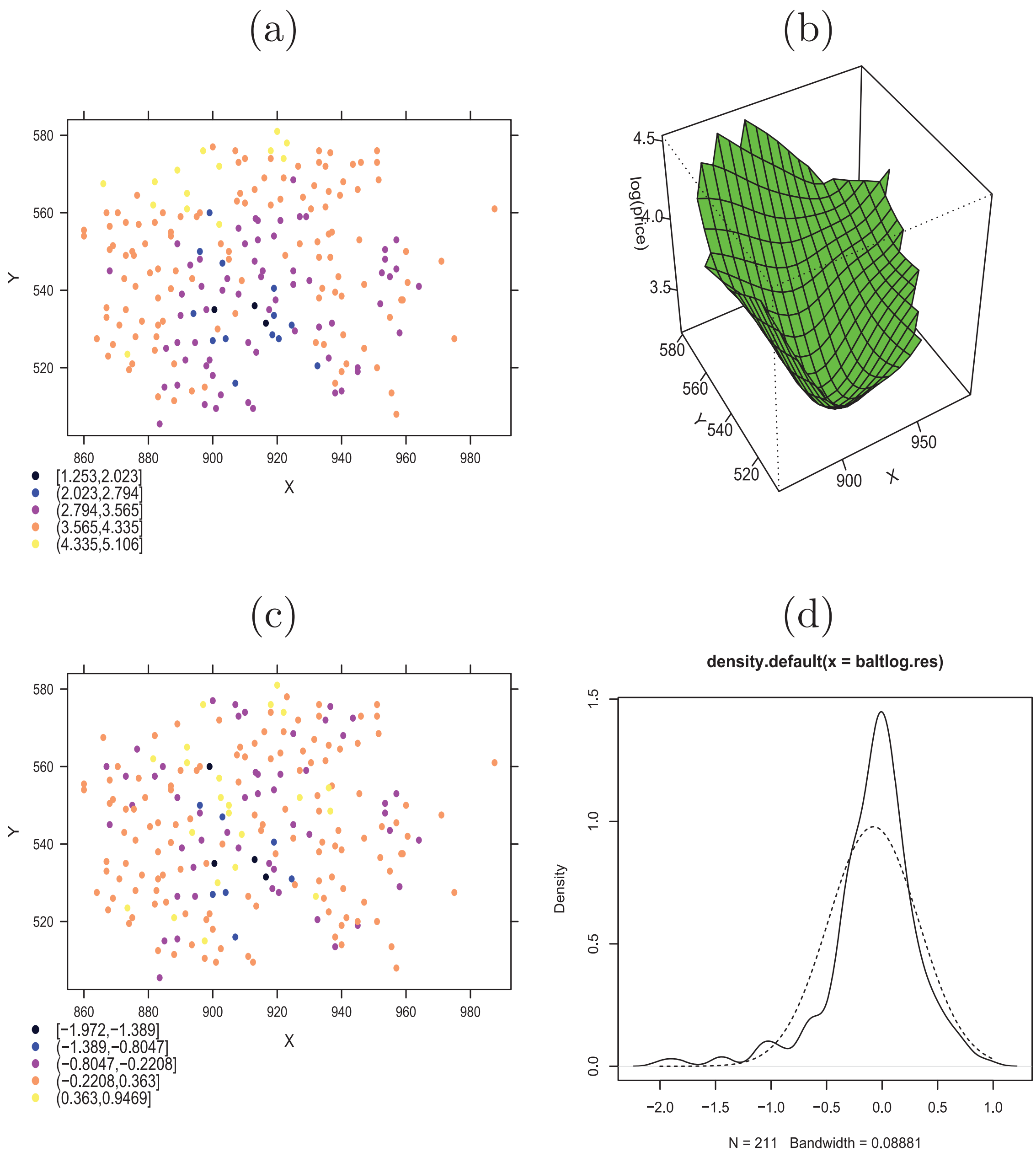

Figure 4. Spatial plot of house sales price in Baltimore. Panel (a): Logarithmic house sales price; Panel (b): Spatial trend of logarithmic house sales price, fitted by sm.regression; Panel (c): Residual of logarithmic house sales price by removing spatial trend, denoted by $Y_i = Y(\mathbf{s}_i)$; Panel (d): Marginal kernel density (solid line) and Gaussian density (dashed line) of $Y_i = Y(\mathbf{s}_i)$.

$35 \le y \le 65$) that is shown in Figure 1(b) of Matsuda and Yajima (2009), to which they applied nonparametric and parametric spectral estimation. To account for the mean component, Matsuda and Yajima (2009) considered the regression model for land price data $X_{\mathbf{t}_j}$, $j = 1, \ldots, 1431$,

$$X_{\mathbf{t}_j} = \mu_0 + \mu_1 u_{\mathbf{t}_j} + \mu_2 v_{\mathbf{t}_j} + Z_{\mathbf{t}_j},\ j = 1, \ldots, 1431, \quad (8.17)$$

where $u_{\mathbf{t}_j}$ and $v_{\mathbf{t}_j}$ are the covariates that stand for the distances from the nearest train station and the terminal station in central Tokyo to the location $\mathbf{t}_j \in R^2$, respectively. They then analyzed the residual process $Z_{\mathbf{t}}$, $\mathbf{t} \in R^2$, by the spectral density function that was identified by the nonisotropic Matern class; see Section 6.2 of of Matsuda and Yajima (2009) for details.

In this section, we are interested in the distributions of the residual process $Z_{\mathbf{t}}$. Is this process a Gaussian process, and if not, what is the nonlinear spatial interdependence pattern? Here we consider the residuals $Z_{\mathbf{t}_j}$, $j = 1, \ldots, 1431$, after the least squares regression of (8.17), the spatial plot of which is depicted in Figure 7(a), where $Z_{\mathbf{t}_j}$ has been scaled by 10^{-4}. The kernel estimate of the marginal density distribution of $Z_{\mathbf{t}_j}$ is depicted in Figure 7(b), which appears to deviate somewhat from a Gaussian density of the same mean and variance. We, therefore, further examine the joint density function, $f(x, y; \mathbf{s}_0)$, of $(Z_{\mathbf{t}_j}, Z_{\mathbf{t}_j+\mathbf{s}_0})$ for $\mathbf{s}_0 \neq (0, 0)$. The contour plots of the estimated joint probability density functions for the residuals $Z_{\mathbf{t}_j}$, $j = 1, \ldots, 1431$, are given Figure 8 with (a) $\mathbf{s}_0 = (1, 1)$ and CV-bandwidth $(b, h) = (2, 0.53)$, (b) $\mathbf{s}_0 = (5, 5)$ with CV-bandwidth $(b, h) = (1.5, 1.0)$, and (c) $\mathbf{s}_0 = (9, 9)$ with CV-bandwidth $(b, h) = (2.5, 1.0)$, respectively. Here the cross-validation principle detailed in the last section with $(a, A) = (-8.16, 10.65)$, $M = 30$, and $\epsilon = 0.01$ in

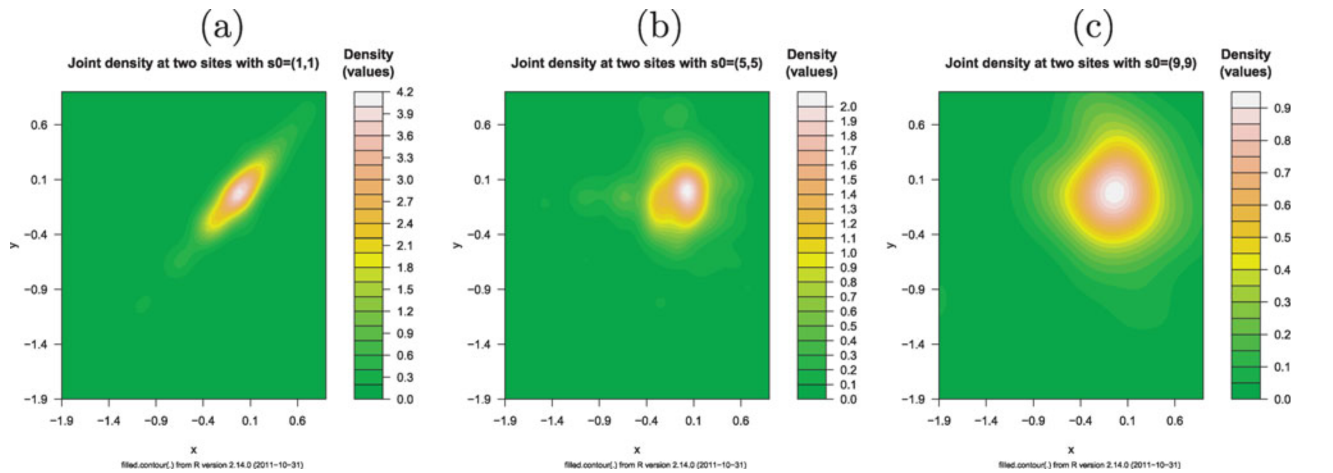

Figure 5. The contour plots of the estimated joint probability density functions for $(Y(\mathbf{s}_i), Y(\mathbf{s}_i+\mathbf{s}_0))$: (a) $\mathbf{s}_0=(1,1)$ with CV-bandwidth $(b,h)=(0.9556, 0.1025)$, (b) $\mathbf{s}_0=(5,5)$ with CV-bandwidth $(b,h)=(2.222, 0.1)$, (c) $\mathbf{s}_0=(9,9)$ with CV-bandwidth $(b,h)=(1.211, 0.2425)$.

(7.12), is applied to select the bandwidths (b,h), and the plots are highly non-Gaussian.

The Tokyo land price data are quite different in structure from the Baltimore house price data. It is seen in Figure 7(a) that large negative residuals are clustered in several geographical areas. This is not captured by the bivariate density at lag $\mathbf{s}_0=(1,1)$, which shows positive dependence similar to the Baltimore data. However, at higher lags the structure of Figure 7(a) has the effect of creating two modes. Such effects can easily be produced by simulation or computation as in Section 8.1, by mixture models. Depending on the distance between the two components of the mixture, the corresponding bivariate density simulated (or computed) as in Section 8.1, may have one or two modes. (It is also possible to construct spatial moving average models with more than two modes.) In the present case, there are two: one composed by pairing the peak value of Figure 7(b) with itself, and one by pairing this peak value with the point of change of the slope at about -5 of Figure 7(b). Clearly the bivariate density plots of Figure 8 give much more information about spatial dependence than a linear covariance analysis would. For both Figures 8(c) and (b) such an analysis would probably give a correlation close to zero, ignoring the nonlinear and heteroscedastic structure of these plots.

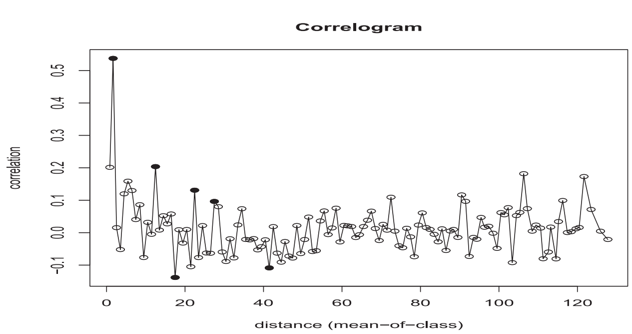

Figure 6. The correlogram of $Y_i = Y(\mathbf{s}_i)$, the residual of the logarithmic house sales price after removing spatial trend.

9. THE PROOF FOR THE MAIN THEOREMS

We collect the proofs of the main theorems in this section, with the necessary lemmas involved to be proved in the Appendix (online supplementary material).

The following lemma is often applied, borrowed from Ibragimov and Linnik (1971) or Deo (1973).

Lemma 1. Let $\mathcal{L}_r(\mathcal{F})$ denote the class of $\mathcal{F}$-measurable random variables ξ satisfying $\|\xi\|_r := (\mathrm{E}|\xi|^r)^{1/r} < \infty$. Let $U \in \mathcal{L}_r(\mathcal{B}(\mathcal{S}))$ and $V \in \mathcal{L}_s(\mathcal{B}(\mathcal{S}'))$, where $\mathcal{B}(\mathcal{S})$ and $\mathcal{B}(\mathcal{S}')$ denote the σ-fields generated by $\{Y(\mathbf{s}) : \mathbf{s} \in \mathcal{S}\}$ and $\{Y(\mathbf{s}) : \mathbf{s} \in \mathcal{S}'\}$, respectively. Then, for any $1 \le r, s, t < \infty$ such that $r^{-1}+s^{-1}+t^{-1}=1$,

$$|\mathrm{E}[UV]-\mathrm{E}[U]\mathrm{E}[V]| \le C\|U\|_r\|V\|_s[\alpha(\mathcal{S},\mathcal{S}')]^{1/t}, \tag{9.18}$$

where $\alpha(\mathcal{S},\mathcal{S}') = \sup\{|P(AB)-P(A)P(B)| : A \in \mathcal{B}(\mathcal{S}), B \in \mathcal{B}(\mathcal{S}')\}$.

9.1 Proofs for Section 4

In this section, we only provide the full proof for the asymptotic variance (4.7). The proof of Theorem 1 is only sketched. It is similar to that of Theorem 4 (which is more involved) given in Section 9.4 by applying some of the details of this section.

Proof of (4.7). Note that

$$\begin{aligned}\mathrm{var}(\hat{f}(x)) &= \frac{1}{(Nh)^2}\Bigg[\sum_{i=1}^{N}\mathrm{var}(K((Y_i-x)/h)) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \\ &\qquad \times \mathrm{cov}(K((Y_i-x)/h), K((Y_j-x)/h))\Bigg] \\ &= \frac{1}{Nh}[V_{N1}+V_{N2}], \end{aligned}\tag{9.19}$$

where

$$\begin{aligned} V_{N1} &= h^{-1}\mathrm{var}(K((Y_i-x)/h)) \\ &\to f(x)\int K^2(u)du \equiv V_1(x), \end{aligned}\tag{9.20}$$

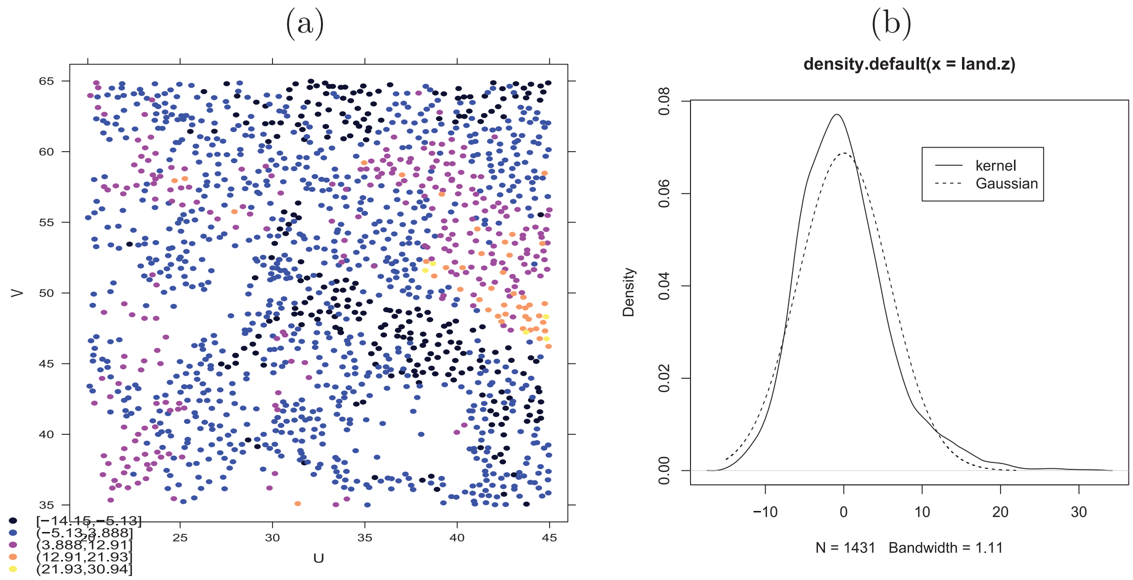

Figure 7. The residuals $Z_{\mathbf{t}_j}$, , $j = 1, \ldots, 1431$, after the least squares regression of (8.17) for Tokyo land prices: (a) Spatial plot; (b) Marginal kernel density (solid line) and Gaussian density (dashed line).

and

$$V_{N2} = (Nh)^{-1} 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \times \mathrm{cov}(K((Y_i - x)/h), K((Y_j - x)/h)). \quad (9.21)$$

We are showing $V_{N2} \to 0$ as $N \to \infty$: First of all, for $i \neq j$,

$$\begin{aligned} &|\mathrm{cov}(K((Y_i - x)/h), K((Y_j - x)/h))| \\ &\leq \int\int K((u-x)/h), K((v-x)/h)|f_{ij}(u, v) \\ &\quad - f(u)f(v)|dudv \\ &\leq Ch^2 \int\int K(u)K(v)dudv = O(h^2), \end{aligned}$$

where $O(\cdot)$ is uniform with respect to i and j owing to the first part of Assumption (I)(iii). Further, applying Lemma 1 given in the above, we can easily derive that, for the $\kappa > 0$ specified in Assumption (I)(i),

$$\begin{aligned} &|\mathrm{cov}(K((Y_i - x)/h), K((Y_j - x)/h))| \\ &\quad \leq C\|K((Y_i - x)/h)\|_{2+\kappa} \|K((Y_j - x)/h)\|_{2+\kappa} \\ &\quad \times [\alpha(d(\mathbf{s}_i, \mathbf{s}_j))]^{\kappa/(2+\kappa)} \\ &\quad = O(h^{2/(2+\kappa)})[\alpha(d(\mathbf{s}_i, \mathbf{s}_j))]^{\kappa/(2+\kappa)}, \end{aligned}$$

where C is a generic finite positive constant that may differ at different places throughout the remainder of this article. Thus it follows from (9.21) that

$$\begin{aligned} &|V_{N2}| \leq C(Nh)^{-1} \\ &\times \sum_{0<d(\mathbf{s}_i, \mathbf{s}_j) \leq c_N} |\mathrm{Cov}(K((Y_i - x)/h), K((Y_j - x)/h))| \\ &+ C(Nh)^{-1} \sum_{d(\mathbf{s}_i, \mathbf{s}_j) > c_N} |\mathrm{Cov}(K((Y_i - x)/h), K((Y_j - x)/h))| \end{aligned}$$

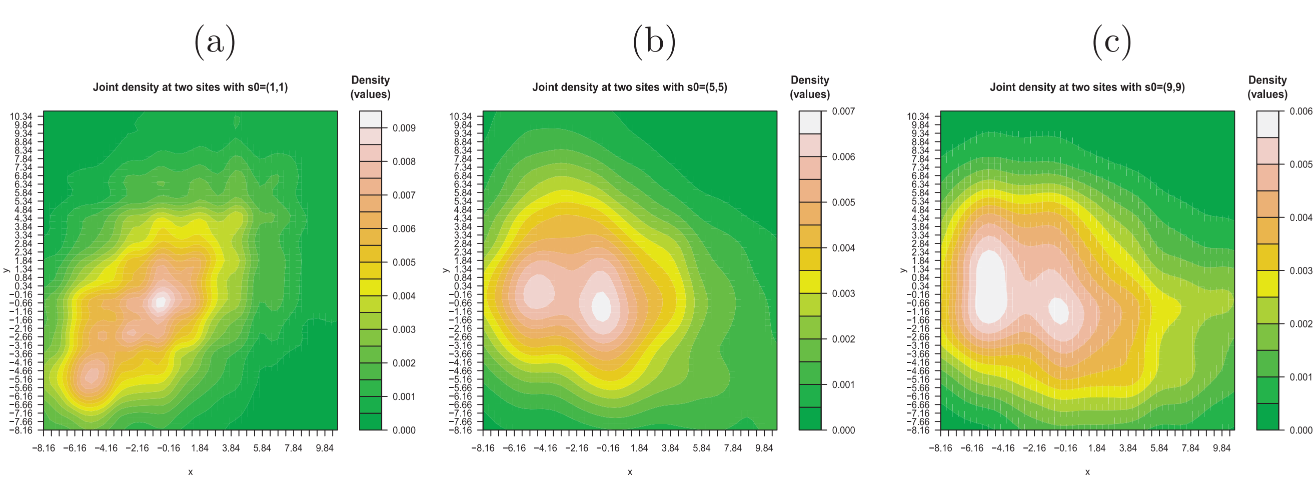

Figure 8. The contour plots of the estimated joint probability density functions for the residuals $Z_{\mathbf{t}_j}$, , $j = 1, \ldots, 1431$, after the least squares regression of (8.17) for Tokyo land prices: (a) $\mathbf{s}_0 = (1, 1)$ with CV-bandwidth $(b, h) = (2, 0.53)$, (b) $\mathbf{s}_0 = (5, 5)$ with CV-bandwidth $(b, h) = (1.5, 1.0)$, (c) $\mathbf{s}_0 = (9, 9)$ with CV-bandwidth $(b, h) = (2.5, 1.0)$.

$$\le (Nh)^{-1} \sum_{0<d(\mathbf{s}_i,\mathbf{s}_j)\le c_N} O(h^2) + (Nh)^{-1} \times \sum_{d(\mathbf{s}_i,\mathbf{s}_j)>c_N} O(h^{2/(2+\kappa)})[\alpha(d(\mathbf{s}_i,\mathbf{s}_j))]^{\kappa/(2+\kappa)},$$

where $\sum_{0<d(\mathbf{s}_i,\mathbf{s}_j)\le c_N}$ stands for the summation over $\{(i,j): 1 \le i, j \le N, 0 < d(\mathbf{s}_i,\mathbf{s}_j) \le c_N\}$, the cardinality of which is controlled by $CN(c_N/\delta_N)^2$, and $\sum_{d(\mathbf{s}_i,\mathbf{s}_j)>c_N}$ stands for the summation over $\{(i,j): 1 \le i, j \le N, d(\mathbf{s}_i,\mathbf{s}_j) > c_N\}$, which is a subset of $\bigcup_{m=c_N}^{\infty}\{(i,j): 1 \le i, j \le N, m < d(\mathbf{s}_i,\mathbf{s}_j) \le (m+1)\}$. Thus

$$\begin{aligned} |V_{N2}| &\le (Nh)^{-1} CN(c_N/\delta_N)^2 O(h^2) + (Nh)^{-1} \\ &\quad \times \sum_{m=c_N}^{\infty} O(h^{2/(2+\kappa)}) N((m+1)/\delta_N)^2 [\alpha(m)]^{\kappa/(2+\kappa)} \\ &\le O(1)\{(c_N/\delta_N)^2 h\} + O(1)\left\{\delta_N^2 h^{\kappa/(2+\kappa)}\right\}^{-1} \\ &\quad \times \sum_{m=c_N}^{\infty} m^2 [\alpha(m)]^{\kappa/(2+\kappa)}. \end{aligned} \tag{9.22}$$

Let c_N be the integer part of $\{\delta_N^2 h^{\kappa/(2+\kappa)}\}^{-1/\gamma}$ for the $\gamma > 0$ specified in Assumption I(i), by which the second part of (9.22) tends to zero as $N \to \infty$. Now note that

$$(c_N/\delta_N)^2 h = \left\{\delta_N^2 h^{\kappa/(2+\kappa)}\right\}^{-2/\gamma} \delta_N^{-2} h = \delta_N^{-2(1+2/\gamma)} h^{1-2\kappa/\{(2+\kappa)\gamma\}},$$

which tends to zero, following from Assumption IV(a). It hence follows from (9.19)–(9.22) that

$$(Nh)\mathrm{var}(\hat f(x)) \to V_1(x) = f(x)\int K^2(u)du \tag{9.23}$$

as $N \to \infty$.

We turn to the proof of asymptotic normality. Note that by letting $Z_{i,N} := [K_h(Y_i - x) - E\{K_h(Y_i - x)\}]/N$, we have

$$S_N := \sum_{i=1}^{N} Z_{i,N} = \hat f(x) - E[\hat f(x)]. \tag{9.24}$$

We define c_N as in the above, and

$$S_{i,N} := \sum_{\substack{\ell = 1 \\ d(\mathbf{s}_i,\mathbf{s}_\ell) \le c_N}}^{N} Z_{\ell,N} = \sum_{\ell\in J_{3,i}} Z_{\ell,N}, \tag{9.25}$$

where $J_{3,i} := \{1 \le \ell \le N : d(\mathbf{s}_i,\mathbf{s}_\ell) \le c_N\}$. Then, as shown above, we notice from (9.23) that

$$\begin{aligned} a_N : &= \sum_{i=1}^{N} E(Z_{i,N} S_{i,N}) = \mathrm{var}(S_N)(1+o(1)) \\ &= (Nh)^{-1} V_1(x)(1+o(1)), \end{aligned} \tag{9.26}$$

$$\sqrt{Nh}(\hat f(x) - \mathrm{E}\hat f(x))/\sqrt{V_1(x)} = a_N^{-1/2} S_N(1+o(1)). \tag{9.27}$$

Now we need to show that $\bar S_N := a_N^{-1/2} S_N \Rightarrow N(0,1)$ by showing the following, due to Bolthausen (1982, Lemma 2, p. 1048):

a. $\sup_N E(\bar S_N^2) < \infty$;
b. $\lim_{N\to\infty} E[(\iota\lambda - \bar S_N)e^{\iota\lambda \bar S_N}] = 0$ for any $\lambda \in \mathbb{R}$, where $\iota = \sqrt{-1}$.

Here (a) obviously follows from (9.26) in the above. The proof of (b) for the case of marginal density estimator can be done, as in the steps of (9.40)–(9.43) in Section 9.4 together with the details given in Section A.1.3 for the more involved case of joint density estimator, by using the following facts (here bearing in mind $Z_{i,N} := [K_h(Y_i - x) - E\{K_h(Y_i - x)\}]/N$):

$$E Z_{i,N}^2 Z_{\ell,N}^2 = \begin{cases} O(1)(N^4h^3)^{-1} & \text{for } i = \ell, \\ O(1)(N^4h^2)^{-1} & \text{for } i \ne \ell; \end{cases}$$

$$EZ_{i,N}^2 Z_{\ell,N} Z_{\ell',N} = \begin{cases} O(1)(N^4h^2)^{-1} & \text{for i} = \ell \text{ or } i = \ell', \\ O(1)(N^4h)^{-1} & \text{for i} \ne \ell \text{ and } i \ne \ell', \end{cases}$$

where $\ell \ne \ell'$;

and $E Z_{i,N} Z_{i',N} Z_{\ell,N} Z_{\ell',N} = O(1)N^{-4}$ for all i, i', ℓ, ℓ' different from each other, as well as

$$\begin{aligned} &|\mathrm{cov}(Z_{i,N}Z_{\ell,N}, Z_{i',N}Z_{\ell',N})| \\ &\le \|Z_{i,N}Z_{\ell,N}\|_{2+\kappa}\, \|Z_{i',N}Z_{\ell',N}\|_{2+\kappa} \{\alpha(d(\{\mathbf{s}_i,\mathbf{s}_\ell\},\{\mathbf{s}_{i'},\mathbf{s}_{\ell'}\}))\}^{\kappa/(2+\kappa)} \\ &= \begin{cases} O(1)(Nh)^{-2} h^{2/(2+\kappa)} \{\alpha(d(\{\mathbf{s}_i\},\{\mathbf{s}_{i'}\}))\}^{\kappa/(2+\kappa)} \\ \quad \text{for } i = \ell \text{ and } i' = \ell', \\ O(1)(Nh)^{-2} h^{3/(2+\kappa)} \{\alpha(d(\{\mathbf{s}_i\},\{\mathbf{s}_{i'},\mathbf{s}_{\ell'}\}))\}^{\kappa/(2+\kappa)} \\ \quad \text{for } i = \ell \text{ and } i' \ne \ell', \\ O(1)(Nh)^{-2} h^{3/(2+\kappa)} \{\alpha(d(\{\mathbf{s}_i,\mathbf{s}_\ell\},\{\mathbf{s}_{i'}\}))\}^{\kappa/(2+\kappa)} \\ \quad \text{for } i \ne \ell \text{ and } i' = \ell', \\ O(1)(Nh)^{-2} h^{4/(2+\kappa)} \{\alpha(d(\{\mathbf{s}_i,\mathbf{s}_\ell\},\{\mathbf{s}_{i'},\mathbf{s}_{\ell'}\}))\}^{\kappa/(2+\kappa)} \\ \quad \text{for } i \ne \ell \text{ and } i' \ne \ell', \end{cases} \end{aligned}$$

where $\ell \in J_{3,i}$ and $\ell' \in J_{3,i'}$ with $d(\mathbf{s}_i,\mathbf{s}_{i'}) > 3c_N$. The details are, therefore, omitted here to save space. □

9.2 Proofs for Section 5.1

Proof. Denote by $\mathbf{s}_0 = \mathbf{s}_i - \mathbf{s}_k$. For $\hat f(x, y, \mathbf{s}_0)$ defined in (2.2), the bias

$$B(\hat f) = E[\hat f(x, y, \mathbf{s}_0)] - f(x, y, \mathbf{s}_0),$$

where

$$E[\hat f(x, y, \mathbf{s}_0)] = \frac{\sum_{j,\ell=1}^{N} L_b(\mathbf{s}_j - \mathbf{s}_\ell - \mathbf{s}_0) E[K_h(Y_j - x)K_h(Y_\ell - y)]}{\sum_{j,\ell=1}^{N} L_b(\mathbf{s}_j - \mathbf{s}_\ell - \mathbf{s}_0)},$$

and applying Taylor's expansion together with Assumption (II),

$$\begin{aligned} &E[K_h(Y_j - x)K_h(Y_\ell - y)] \\ &= h^{-2}\int\int K((u-x)/h)K((v-y)/h) f(u, v, \mathbf{s}_j - \mathbf{s}_\ell)\, dudv \\ &= \int\int K(u)K(v) f(x + uh, y + vh, \mathbf{s}_j - \mathbf{s}_\ell)\, dudv \\ &= \int\int K(u)K(v)[f(x, y, \mathbf{s}_j - \mathbf{s}_\ell) + f_x(x, y, \mathbf{s}_j - \mathbf{s}_\ell)uh \\ &\quad + f_y(x, y, \mathbf{s}_j - \mathbf{s}_\ell)vh]\, dudv \\ &\quad + \int\int K(u)K(v)\frac{1}{2}[f_{xx}(x_h, y_h, \mathbf{s}_j - \mathbf{s}_\ell)(uh)^2 \\ &\quad + 2f_{xy}(x_h, y_h, \mathbf{s}_j - \mathbf{s}_\ell)(uvh^2) \\ &\quad + f_{yy}(x_h, y_h, \mathbf{s}_j - \mathbf{s}_\ell)(vh)^2]dudv \\ &= f(x, y, \mathbf{s}_j - \mathbf{s}_\ell) + \frac{1}{2}[f_{xx}(x, y, \mathbf{s}_j - \mathbf{s}_\ell) + f_{yy}(x, y, \mathbf{s}_j - \mathbf{s}_\ell)] \\ &\quad \times \mu_{K,2} h^2 (1 + o(1)), \end{aligned} \tag{9.28}$$

where $x_h = x + \xi_1 uh$ and $y_h = x + \xi_2 vh$ with $|\xi_k| \leq 1$ for $k = 1, 2$. Therefore

$$B(\hat{f}) = \frac{\sum_{j,\ell=1}^{N} L_b(\mathbf{s}_j - \mathbf{s}_\ell - \mathbf{s}_0)[f(x, y, \mathbf{s}_j - \mathbf{s}_\ell) - f(x, y, \mathbf{s}_0)]}{\sum_{j,\ell=1}^{N} L_b(\mathbf{s}_j - \mathbf{s}_\ell - \mathbf{s}_0)} + \frac{1}{2}\frac{\sum_{j,\ell=1}^{N} L_b(\mathbf{s}_j - \mathbf{s}_\ell - \mathbf{s}_0)[f_{xx}(x, y, \mathbf{s}_j - \mathbf{s}_\ell) + f_{yy}(x, y, \mathbf{s}_j - \mathbf{s}_\ell)]}{\sum_{j,\ell=1}^{N} L_b(\mathbf{s}_j - \mathbf{s}_\ell - \mathbf{s}_0)} \times \mu_{K,2} h^2(1 + o(1)). \tag{9.29}$$

We need the following lemma. □

Lemma 2. Under Assumptions (I), (II), (III), and (IV), as $N \to \infty$,

$$\frac{1}{N^2}\sum_{j,\ell=1}^{N} L_b(\mathbf{s}_j - \mathbf{s}_\ell - \mathbf{s}_0) H(\mathbf{s}_j - \mathbf{s}_\ell) \to H(\mathbf{s}_0) A_0(\mathbf{s}_0),$$

and

$$\frac{1}{N^2}\sum_{j,\ell=1}^{N} b^{-2} L^2((\mathbf{s}_j - \mathbf{s}_\ell - \mathbf{s}_0)/b) H(\mathbf{s}_j - \mathbf{s}_\ell) \to H(\mathbf{s}_0)\int_{\mathcal{R}^2} L^2(\mathbf{s}_1) d\mathbf{s}_1 A_0(\mathbf{s}_0)$$

for any function $H(\mathbf{s})$ which is continuous at $\mathbf{s}_0$, and

$$\frac{1}{N^2}\sum_{j,\ell=1}^{N} L_b(\mathbf{s}_j - \mathbf{s}_\ell - \mathbf{s}_0)[f(x, y, \mathbf{s}_j - \mathbf{s}_\ell) - f(x, y, \mathbf{s}_0)] = \frac{1}{2} b^2 \left[2\frac{\partial f(x, y, \mathbf{s}_0)}{\partial \mathbf{s}^\tau} \pm b\mu_{L,2} \mathbf{A}_1(\mathbf{s}_0) + \mathrm{tr}\left(\frac{\partial^2 f(x, y, \mathbf{s}_0)}{\partial \mathbf{s}\mathbf{s}^\tau} \pm b\mu_{L,2} \right) A_0(\mathbf{s}_0) \right] (1 + o(1)) \tag{9.30}$$

The proof of Lemma 2 will be given in the Appendix. Now applying (9.29) together with Lemma 2 leads to

$$B(\hat{f}) = \frac{\frac{1}{2} b^2 \left[2\frac{\partial f(x,y,\mathbf{s}_0)}{\partial \mathbf{s}^\tau} \boldsymbol{\mu}_{L,2} \mathbf{A}_1(\mathbf{s}_0) + \mathrm{tr}\left(\frac{\partial^2 f(x,y,\mathbf{s}_0)}{\partial \mathbf{s}\mathbf{s}^\tau} \boldsymbol{\mu}_{L,2} \right) A_0(\mathbf{s}_0) \right] (1 + o(1))}{A_0(\mathbf{s}_0)} + \frac{1}{2}\frac{A_0(\mathbf{s}_0)[f_{xx}(x, y, \mathbf{s}_0) + f_{yy}(x, y, \mathbf{s}_0)]}{A_0(\mathbf{s}_0)} \mu_{K,2} h^2 (1 + o(1)),$$

from which Theorem 2 follows obviously.

9.3 Proofs for Section 5.2

Proof. Denote

$$w_{j\ell}(\mathbf{s}_0) = \frac{L_b(\mathbf{s}_j - \mathbf{s}_\ell - \mathbf{s}_0)}{\sum_{i,k=1}^{N} L_b(\mathbf{s}_i - \mathbf{s}_k - \mathbf{s}_0)}. \tag{9.31}$$

Then

$$\hat{f}(x, y, \mathbf{s}_0) = \sum_{j,\ell=1}^{N} w_{j\ell}(\mathbf{s}_0) K_h(Y_j - x) K_h(Y_\ell - y).$$

Therefore,

$$\mathrm{var}(\hat{f}) = \sum_{j,\ell=1}^{N} w_{j\ell}^2(\mathbf{s}_0) \mathrm{var}(K_h(Y_j - x) K_h(Y_\ell - y)) + \sum_{j,\ell=1}^{N} \sum_{\substack{j',\ell'=1 \\ (j',\ell') \neq (j,\ell)}}^{N} w_{j\ell}(\mathbf{s}_0) w_{j'\ell'}(\mathbf{s}_0) \times \mathrm{cov}(K_h(Y_j - x) K_h(Y_\ell - y), K_h(Y_{j'} - x) K_h(Y_{\ell'} - y)) \equiv V_{1N} + V_{2N}. \tag{9.32}$$

Note that if $j = \ell$, then as N is sufficiently large, $L(-\mathbf{s}_0/b) = 0$ in view of the fact that the kernel function $L(\cdot)$ has a bounded support (Assumption (III, ii)) and $b = b_N \to 0$ (Assumption (IV, ii)) with $\mathbf{s}_0 \neq (0, 0)$ fixed, and therefore as N is sufficiently large,

$$w_{jj}(\mathbf{s}_0) = \frac{L_b(-\mathbf{s}_0)}{\sum_{i,k=1}^{N} L_b(\mathbf{s}_i - \mathbf{s}_k - \mathbf{s}_0)} = \frac{b^{-2} L(-\mathbf{s}_0/b)}{\sum_{i,k=1}^{N} L_b(\mathbf{s}_i - \mathbf{s}_k - \mathbf{s}_0)} = 0. \tag{9.33}$$

Thus, for sufficiently large N, we only need to examine the case of $j \neq \ell$ in (9.32) below:

$$\begin{aligned} &\mathrm{var}(K_h(Y_j - x) K_h(Y_\ell - y)) \\ &= E(K_h(Y_j - x) K_h(Y_\ell - y))^2 - (E K_h(Y_j - x) K_h(Y_\ell - y))^2 \\ &= h^{-4} \int K^2((u - x)/h) K^2((v - y)/h) \times f(u, v, \mathbf{s}_j - \mathbf{s}_\ell) du dv + O(1) \\ &= h^{-2} f(x, y, \mathbf{s}_j - \mathbf{s}_\ell) [\int K^2(u) du]^2 (1 + o(1)). \end{aligned}$$

Then applying Lemma 2,

$$\begin{aligned} V_{1N} &= \sum_{j,\ell=1}^{N} w_{j\ell}^2(\mathbf{s}_0) \mathrm{var}(K_h(Y_j - x) K_h(Y_\ell - y)) \\ &= \frac{N^{-4} \sum_{j,\ell=1}^{N} (L_b(\mathbf{s}_j - \mathbf{s}_\ell - \mathbf{s}_0))^2 h^{-2} f(x, y, \mathbf{s}_j - \mathbf{s}_\ell) [\int K^2(u) du]^2 (1 + o(1))}{\{N^{-2} \sum_{i,k=1}^{N} L_b(\mathbf{s}_i - \mathbf{s}_k - \mathbf{s}_0)\}^2} \\ &= \frac{(Nbh)^{-2} \int_{\mathcal{R}^2} L^2(\mathbf{s}_1) d\mathbf{s}_1 f(x, y, \mathbf{s}_0) [\int K^2(u) du]^2 A_0(\mathbf{s}_0) (1 + o(1))}{\{A_0(\mathbf{s}_0)\}^2} \\ &= (Nbh)^{-2} \frac{\int_{\mathcal{R}^2} L^2(\mathbf{s}_1) d\mathbf{s}_1 f(x, y, \mathbf{s}_0) [\int K^2(u) du]^2}{A_0(\mathbf{s}_0)} (1 + o(1)). \end{aligned} \tag{9.34}$$

To deal with V_{2N}, we need the following □

Lemma 3. (cross-term lemma): Under Assumptions (I), (II), (III), and (IV), as $N \to \infty$,

$$\sum_{j,\ell=1}^{N} \sum_{\substack{j',\ell'=1 \\ (j',\ell') \neq (j,\ell)}}^{N} w_{j\ell}(\mathbf{s}_0) w_{j'\ell'}(\mathbf{s}_0) \times \mathrm{cov}(K_h(Y_j - x) K_h(Y_\ell - y), K_h(Y_{j'} - x) K_h(Y_{\ell'} - y)) = o\left((Nbh)^{-2}\right).$$

The proof of Lemma 3 is quite involved, so we leave it in Appendix. Theorem 3 then follows from (9.32), (9.34), and Lemma 3.

9.4 Proofs for Section 5.3

Proof. First recall $w_{ik}(\mathbf{s}_0)$ defined in (9.31) and

$$\hat{f}(x, y; \mathbf{s}_0) = \sum_{i,k=1}^{N} w_{ik}(\mathbf{s}_0) K_h(Y_i - x) K_h(Y_k - y). \tag{9.35}$$

Put

$$\eta_{ik}(x, y; \mathbf{s}_0) := K_h(Y_i - x)K_h(Y_k - y)$$

and

$$\Delta_{ik}(x, y; \mathbf{s}_0) := w_{ik}(\mathbf{s}_0)(\eta_{ik}(x, y; \mathbf{s}_0) - \mathrm{E}\eta_{ik}(x, y; \mathbf{s}_0)),$$

and define $Z_{i,N} := \sum_{k=1}^{N} \Delta_{ik}$. Then

$$S_N := \sum_{i=1}^{N} Z_{i,N} = \hat{f}(x, y; \mathbf{s}_0) - E[\hat{f}(x, y; \mathbf{s}_0)]. \quad (9.36)$$

Now we define c_N as defined in Lemma 3, and

$$S_{i,N} := \sum_{\substack{\ell = 1 \\ d(\mathbf{s}_i, \mathbf{s}_\ell) \le c_N}}^{N} Z_{\ell,N} = \sum_{\ell \in J_{3,i}} Z_{\ell,N}, \quad (9.37)$$

where $J_{3,i} := \{1 \le \ell \le N : d(\mathbf{s}_i, \mathbf{s}_\ell) \le c_N\}$. Then, as shown in the proof of Theorem 3, we notice that

$$\begin{aligned} a_N : &= \sum_{i=1}^{N} E(Z_{i,N} S_{i,N}) = \mathrm{var}(S_N)(1 + o(1)) \\ &= (Nhb)^{-2} V_1(x, y, \mathbf{s}_0)(1 + o(1)), \end{aligned} \quad (9.38)$$

and hence

$$\begin{aligned} &(Nbh)(\hat{f}(x, y; \mathbf{s}_0) - \mathrm{E}\hat{f}(x, y; \mathbf{s}_0))/\sqrt{V_1(x, y, \mathbf{s}_0)} \\ &\quad = a_N^{-1/2} S_N(1 + o(1)). \end{aligned} \quad (9.39)$$

Now we need to show that $\bar{S}_N := a_N^{-1/2} S_N \Rightarrow N(0, 1)$ by showing the following, due to Bolthausen (1982, Lemma 2, p. 1048):

(a) $\sup_N E(\bar{S}_N^2) < \infty$;
(b) $\lim_{N\to\infty} E[(\imath\lambda - \bar{S}_N)e^{\imath\lambda\bar{S}_N}] = 0$ for any $\lambda \in \mathbb{R}$, where $\imath = \sqrt{-1}$.

Here it follows from Theorem 3 that (a) is obviously true. We focus on the proof for (b) by noticing, with $\bar{S}_{i,N} := a_N^{-1/2} S_{i,N}$,

$$E\left[(\imath\lambda - \bar{S}_N)e^{\imath\lambda\bar{S}_N}\right] = A_{1N} - A_{2N} - A_{3N}, \quad (9.40)$$

where we have to prove

$$A_{1N} = E\left[\imath\lambda e^{\imath\lambda\bar{S}_N}\left(1 - a_N^{-1}\sum_{i=1}^{N} Z_{i,N} S_{i,N}\right)\right] \to 0, \quad (9.41)$$

$$A_{2N} = E\left[a_N^{-1/2} e^{\imath\lambda\bar{S}_N}\sum_{i=1}^{N} Z_{i,N}\left(1 - e^{-\imath\lambda\bar{S}_{i,N}} - \imath\lambda\bar{S}_{i,N}\right)\right] \to 0, \quad (9.42)$$

$$A_{3N} = E\left[a_N^{-1/2}\sum_{i=1}^{N} Z_{i,N} e^{\imath\lambda(\bar{S}_N - \bar{S}_{i,N})}\right] \to 0 \quad (9.43)$$

as $N \to \infty$. Because of the double summations in (9.35), the proof of (9.41)–(9.43) is quite involved. We give the proof of (9.41)–(9.43) one by one in detail in the Appendix (Section A.1.3). Thus it follows from (9.40) and (9.41)–(9.43) that (b) holds true. By (a) and (b) together with (9.39), we complete the proof of this proposition. □

9.5 Proofs for Section 6

We need the following lemma to facilitate the proof of the theorems in this section.

Lemma 4. Under Assumptions (I$'$), (II$'$), (III$'$), and (IV$'$), as $N \to \infty$,

$$\begin{aligned} &\frac{1}{N^2}\sum_{j,\ell=1}^{N} L_b(\|\mathbf{s}_j - \mathbf{s}_\ell\| - \|\mathbf{s}_0\|)H(\|\mathbf{s}_j - \mathbf{s}_\ell\|) \\ &\quad \to H(\|\mathbf{s}_0\|)A_0^*(\|\mathbf{s}_0\|), \end{aligned}$$

and

$$\begin{aligned} &\frac{1}{N^2}\sum_{j,\ell=1}^{N} b^{-1}L^2((\|\mathbf{s}_j - \mathbf{s}_\ell\| - \|\mathbf{s}_0\|)/b)H(\|\mathbf{s}_j - \mathbf{s}_\ell\|) \\ &\quad \to H(\|\mathbf{s}_0\|)\int L^2(s_1)ds_1 A_0^*(\|\mathbf{s}_0\|) \end{aligned}$$

for any function $H(\|\mathbf{s}\|)$ which is continuous at $\|\mathbf{s}_0\|$, and

$$\begin{aligned} &\frac{1}{N^2}\sum_{j,\ell=1}^{N} L_b(\|\mathbf{s}_j - \mathbf{s}_\ell\| - \|\mathbf{s}_0\|) \\ &\quad \times [f(x, y, \|\mathbf{s}_j - \mathbf{s}_\ell\|) - f(x, y, \|\mathbf{s}_0\|)] \\ &= \frac{1}{2}b^2\left[2\frac{\partial f(x, y, \|\mathbf{s}_0\|)}{\partial s}\mu_{L,2}\{A_1^*(\|\mathbf{s}_0\|) + A_0^*(\|\mathbf{s}_0\|)/\|\mathbf{s}_0\|\}\right. \\ &\quad \left. + \left(\frac{\partial^2 f(x, y, \|\mathbf{s}_0\|)}{\partial s^2}\mu_{L,2}\right)A_0^*(\|\mathbf{s}_0\|)\right](1 + o(1)). \end{aligned} \quad (9.44)$$

The proof of Lemma 4 will be given in Appendix.

Proof. The proof of Theorems 5–7 is completely similar to that of Theorems 2–4 with Lemma 2 used there replaced by Lemma 4. The details are omitted. □

SUPPLEMENTARY MATERIALS

The proofs for Lemmas 24 and (9-41)–(9-43) in Section 9 are available in the online supplementary materials.

[Received August 2012. Revised May 2014.]

REFERENCES

Bolthausen, E. (1982), "On the Central Limit Theorem for Stationary Mixing Random Fields," *The Annals of Probability*, 4, 1047–1050. [1551,1561,1563]

Chiles, J-P, and Delfiner, P. (1999), *Geostatistics: Modeling Spatial Uncertainty*, New York: Wiley. [1546]

Cressie, N. (1993), *Statistics for Spatial Data* (revised ed.), New York: Wiley. [1546,1548]

Dalenius, T., Hájek, J., and Zubrzycki, S. (1961), "On Plane Sampling and Related Geometrical Problems," *Proceedings of the Fourth Berkeley Symposium on Mathematics, Statistics and Probilities*, Oakland, CA: University of California Press, Vol. 1, pp. 125–150. [1547]

Deo, C. M. (1973), "A Note on Empirical Processes of Strong Mixing Sequences," *Annals of Probability*, 1, 870–875. [1559]

Du, J., Zhang, H., and Mandrekar, V. S. (2009), "Fixed-Domain Asymptotic Properties of Tapered Maximum Likelihood Estimators," *The Annals of Statistics*, 37, 3330–3361. [1548]

Dubin, R. A. (1992), "Spatial Autocorrelation and Neighborhood Quality," *Regional Science and Urban Economics*, 22, 433–452. [1555]

Fan, J., Huang, T., and Li, R. (2007), "Analysis of Longitudinal Data With Semiparametric Estimation of Covariance Function," *Journal of the American Statistical Association*, 102, 632–641. [1547]

Fan, J., and Wu, Y. (2008), "Semiparametric Estimation of Covariance Matrices for Longitudinal Data," *Journal of the American Statistical Association*, 103, 1520–1533. [1547]

Fan, J., and Yao, Q. (2003), *Nonlinear Time Series: Nonparametric and Parametric Methods*, New York: Springer. [1546]

Fan, J., Yao, Q., and Cai, Z. (2003), "Adaptive Varying-Coefficient Linear Models," *Journal of the Royal Statistical Society,* Series B, 65, 57–80. [1552,1553]

Gao, J., Lu, Z., and Tjøstheim, D. (2006), "Estimation in Semiparametric Spatial Regression," *The Annals of Statistics*, 34, 1395–1435. [1546,1548]

Guyon, X. (1987), "Estimation dun champ par pseudo-vraisemblance conditionnelle: Étude asymptotique et application au cas Markovien," in *Spatial Processes and Spatial Time Series Analysis*, Proceedings of the 6th Franco-Belgian Meeting of Statisticians, eds. J.-J. Droesbeke et al., Brussels: FUSL, pp. 15–62. [1549]

——— (1995), *Random Fields on a Network: Modeling, Statistics, and Application*, New York: Springer-Verlag. [1549]

Hall, P., and Patil, P. (1994), "Properties of Nonparametric Estimators of Autocovariance for Stationary Random Fields," *Probability Theory and Related Fields*, 99, 399–424. [1546,1547,1548,1551]

Hallin, M., Lu, Z., and Tran, L. T. (2001), "Density Estimation for Spatial Linear Processes," *Bernoulli*, 7, 657–668. [1546,1548,1549]

——— (2004a), "Kernel Density Estimation for Spatial Processes: The L_1 Theory," *Journal of Multivariate Analysis*, 88, 61–75. [1546,1548,1549]

——— (2004b), "Local Linear Spatial Regression," *The Annals of Statistics*, 32, 2469–2500. [1546,1548,1549,1551]

Ibragimov, I. A., and Linnik, Yu. V. (1971), *Independent and Stationary Sequences of Random Variables*, Groningen: Wolters-Noordhoff. [1559]

Jenish, N., and Prucha, I. R. (2009), "Central Limit Theorems and Uniform Laws of Large Numbers for Arrays of Random Fields," *Journal of Econometrics*, 150, 86–98. [1546,1551]

Lahiri, S. N. (1996), "On Inconsistency of Estimators Based on Spatial Data Under Infill Asymptotics," *Sankhya,* Series A, 58, 403–417. [1548]

Li, Q., and Racine, J. (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton: Princeton University Press. [1552]

Lu, Z. (1998), "On Geometric Ergodicity of a Non-Linear Autoregressive Model With an Autoregressive Conditional Heteroscedastic Term," *Statistica Sinica*, 8, 1205–1217. [1549]

——— (2001), "Asymptotic Normality of Kernel Density Estimators Under Dependence," *Annals of Institute of Statistical Mathematics*, 53, 447–468. [1549]

Lu, Z., and Linton, O. (2007), "Local Linear Fitting Under Near Epoch Dependence," *Econometric Theory*, 23, 37–70. [1549]

Lu, Z., Lundervold, A., Tjøstheim, D., and Yao, Q. (2007), "Exploring Spatial Nonlinearity Using Additive Approximation," *Bernoulli*, 13, 447–472. [1546,1547]

Lu, Z., Steinskog, D. J., Tjøstheim, D., and Yao, Q. (2009), "Adaptively Varying-Coefficient Spatiotemporal Models," *Journal of Royal Statistical Society,* Series B, 71, 859–880. [1548,1549,1553]

Lu, Z., Tjøstheim, D., and Yao, Q. (2008), "Spatial Smoothing, Nugget Effect and Infill Asymptotics," *Statistics and Probability Letters*, 78, 3145–3151. [1548,1549]

Masry, E. (1986), "Recursive Probability Density Estimation for Weakly Dependent Processes," *IEEE Transactions on Information Theory*, 32, 254–267. [1549]

Matérn, B. (1960), *Spatial Variation*, Meddelanden fran Statens Skogsforsknings Institut, Stockholm. Band 49, No. 5. [1547]

Matsuda, Y., and Yajima, Y. (2009), "Fourier Analysis of Irregularly Spaced Data on R^d," *Journal of the Royal Statistical Society,* Series B, 71, 191–217. [1546,1548,1555,1558]

Nakhapetyan, B. S. (1980), "The Central Limit Theorem for Random Fields With Mixing Conditions," in *Advances in Probability 6, Multicomponent Systems*, eds. R. L. Dobrushin and Ya. G. Sinai, New York: Marcel Dekker, pp. 531–548. [1549]

Neaderhouser, C. C. (1980), "Convergence of Blocks Spins Defined on Random Fields," *Journal of Statistical Physics*, 22, 673–684. [1549,1550]

Novak, E. (1988), *Deterministic and Stochastic Error Bounds in Numerical Analysis*, Lecture Notes in Mathematics, No. 1349, Berlin: Springer-Verlag. [1547]

Pham, T. D. (1986), "The Mixing Properties of Bilinear and Generalized Random Coefficient Autoregressive Models," *Stochastic Processes and Their Applications*, 23, 291–300. [1549]

Pham, T. D., and Tran, L. T. (1985), "Some Mixing Properties of Time Series Models," *Stochastic Processes and Their Applications*, 19, 297–303. [1549]

Quenouille, M. H. (1949), "Approximate Tests of Correlation in Time Series," *Journal of The Royal Statistical Society,* Series B, 11, 68–84. [1547]

Robinson, P. M. (2008), "Developments in the Analysis of Spatial Data," *Journal of the Japanese Statistical Society*, 38, 87–96. [1548]

——— (2009), "Large-Sample Inference on Spatial Dependence," *Econometrics Journal*, 12, s1, S68–S82. [1548]

——— (2011), "Asymptotic Theory for Nonparametric Regression With Spatial Data," *Journal of Econometrics*, 165, 5–19. [1548]

Rosenblatt, M. (1985), *Stationary Sequences and Random Fields*, Boston, MA: Birkhauser. [1549]

Ruppert, D. (1997), "Empirical-Bias Bandwidths for Local Polynomial Regression and Density Estimation," *Journal of the American Statistical Association*, 92, 1049–1062. [1553]

Stein, C. (1972), "A Bound for the Error in the Normal Approximation of a Sum of Dependent Random Variables," in *Proceedings of the Berkeley Symposium*, pp. 583–603. [1551]

Stein, M. L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, New York: Springer. [1546,1548]

Stone, M. (1974), "Cross-Validation and Multinomial Prediction," *Biometrika*, 61, 509–515. [1552]

Takahata, H. (1983), "On the Rates in the Central Limit Theorem for Weakly Dependent Random Fields," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 64, 445–456. [1549,1550]

Teräsvirta, T., Tjøstheim, D., and Granger, C. W. J. (2010), *Modelling Nonlinear Economic Time Series*, Oxford: Oxford University Press. [1546]

Tjøstheim, D. (1990), "Nonlinear Time Series and Markov Chains," *Advances in Applied Probability*, 22, 587–611. [1549]

Tran, L. T. (1990), "Kernel Density Estimation on Random Fields," *Journal of Multivariate Analysis*, 34, 37–53. [1546,1549,1551]

Traub, J. F., Wozniakowski, H., and Wasilkowski, G. W. (1988), *Information-Based Complexity*, New York: Academic Press. [1547]

Withers, C. S. (1981), "Conditions for Linear Processes to be Strong Mixing," *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57, 477–480. [1549]

Zhang, H. (2004), "Inconsistent Estimation and Asymptotically Equal Interpolations in Model-Based Geostatistics," *Journal of the American Statistical Association*, 99, 250–261. [1548]

Zhang, H., and Zimmerman, D. L. (2005), "Towards Reconciling Two Asymptotic Frameworks in Spatial Statistics," *Biometrika*, 92, 921–936. [1548]