Statistica Sinica **27** (2017), 555-574
doi: https://doi.org/10.5705/ss.202015.0173

SINGLE-INDEX MODEL FOR INHOMOGENEOUS SPATIAL POINT PROCESSES

Yixin Fang and Ji Meng Loh

New Jersey Institute of Technology

Abstract: We introduce a single index model for the intensity of an inhomogeneous spatial point process, relating the intensity function to an unknown function ρ of a linear combination of measurements of a p-dimensional spatial covariate process. Such a model extends and generalizes a commonly used model where ρ is known. We derive an estimating procedure for ρ and the coefficient parameters $\boldsymbol{\beta}$ and show consistency and asymptotic normality of estimates of $\boldsymbol{\beta}$ under some regularity assumptions. We present results of some simulation studies showing the effectiveness of the procedure. Finally, we apply the procedure to a dataset of fast food restaurant locations in New York City.

Key words and phrases: Asymptotic normality, consistency, fast food restaurant data, single-index model, spatial point processes.

1. Introduction

In the analysis of inhomogeneous spatial point patterns, a common interest is the study of the relationship between the intensity function λ and various measured spatial covariates $\mathbf{Z}$. For example, Waagepetersen and Guan (2009) examined the intensity of trees in a plot of land in relation to land and soil characteristics, and Illian et al. (2012) modeled the locations of muskoxen herds with an index of vegetation productivity and other spatial covariates such as altitude.

To be more specific, let X represent an inhomogeneous spatial point process defined on $\mathbb{R}^2$ that is observed within a finite observation window W_n, and let $X(B)$ denote the number of events observed in a Borel set $B \subset \mathbb{R}^2$ and $|B|$ the area of B. The intensity function of X is defined as

$$\lambda(s|\,\mathbf{Z}(s) = \mathbf{Z}) = \frac{\lim_{|ds|\to 0} E\{X(ds)|\,\mathbf{Z}(s) = \mathbf{Z}\}}{|ds|}, \tag{1.1}$$

which incorporates the dependence of the intensity function on a p-dimensional stationary covariate process $\mathbf{Z}(\cdot)$. See, e.g. Diggle (2003).

Recent work on the modeling of λ as a function of $\mathbf{Z}$ has been based on assuming

$$\lambda(s|\,\mathbf{Z}(s)) = \exp(\mathbf{Z}(s)^{\mathrm{T}}\boldsymbol{\beta}_0), \tag{1.2}$$

and estimating the parameter $\boldsymbol{\beta}_0$ using the estimating equation derived from maximizing the Poisson process log-likelihood,

$$\widetilde{l}_n(\boldsymbol{\beta}) = \frac{1}{|W_n|}\left[\sum_{s\in X\cap W_n}\mathbf{Z}(s)^{\mathrm{T}}\boldsymbol{\beta} - \int_{W_n}\exp(\mathbf{Z}(t)^{\mathrm{T}}\boldsymbol{\beta})\,dt\right]. \tag{1.3}$$

It is important to point out that, although the estimating equation is derived from the Poisson process log-likelihood, Schoenberg (2004) showed that estimates obtained this way are still consistent even when the underlying point process is non-Poisson, under some regularity conditions. Schoenberg (2016) showed further that estimates may be consistent even when some covariates are excluded from the model, provided that the effect of the excluded covariates on the intensity is small. Waagepetersen and Guan (2009) further established asymptotic normality for estimates of $\boldsymbol{\beta}$ obtained from the Poisson estimating equations as part of a two-step procedure introduced by Waagepetersen (2007) for fitting Neyman-Scott models to inhomogeneous spatial point processes.

In this paper, we generalize the model (1.2), replacing the exponential function with an unknown function ρ, so that

$$\lambda(s|\,\mathbf{Z}(s)) = \rho(\mathbf{Z}(s)^{\mathrm{T}}\boldsymbol{\beta}_0), \tag{1.4}$$

where $\|\,\boldsymbol{\beta}_0\,\| = 1$. In regular regression settings, this is known as a single-index model (Ichimura (1993)) and has been used in such areas as econometrics and biometrics. As a semi-parametric model, the single-index model (1.4) is more flexible than the parametric model (1.2). By using only one nonparametric dimension, the single-index model avoids computational difficulties that are common with fully nonparametric models.

We expect the single-index model (1.4) to be more useful than the log-linear model (1.2) for predicting the intensity, even if the two models perform similarly for estimating coefficients $\boldsymbol{\beta}_0$. We give an estimating procedure for the single-index model applied to spatial point patterns and show its effectiveness. We hope this will motivate more research on the use of other semi-parametric models, such as generalized additive models (Hastie and Tibshirani (1990)) and generalized additive partial linear models (Wang et al. (2011)) to model the intensity function.

The single-index model (1.4) can be estimated using either iterative or direct

methods (Horowitz (1998, Ch. 2)). We consider the iterative method, and show consistency and asymptotic normality of the estimators. Guan (2008) studied a more general model, the sliced inverse regression (SIR) model (Li (1991)), and considered the direct method of estimation. The iterative method is appealing for investigating theoretical properties, while the direct method is appealing for high-dimensional data analysis as it is more computationally efficient.

The rest of the paper is organized as follows. In Section 2, we describe the estimation procedure for ρ and $\boldsymbol{\beta}_0$. In Section 3, we derive consistency and asymptotic normality of the estimates of $\boldsymbol{\beta}_0$, while the implementation of the procedure is described in Section 4. In Section 5, we illustrate the procedure using a simulation study and an application to data. The Appendix in the online Supplementary Materials contains the lemmas and some of the more detailed proofs.

2. Method

With the single-index model (1.4) for the intensity function, the Poisson process log-likelihood (1.3) is

$$\widetilde{l}_n(\boldsymbol{\beta}) = \frac{1}{|W_n|}\left[\sum_{s\in X\cap W_n} \log \rho(\mathbf{Z}(s)^{\mathrm{T}}\boldsymbol{\beta}) - \int_{W_n} \rho(\mathbf{Z}(t)^{\mathrm{T}}\boldsymbol{\beta})\, dt\right]. \tag{2.1}$$

Since ρ is an unknown function, $\boldsymbol{\beta}_0$ cannot be estimated by maximizing $\widetilde{l}_n$ directly. Instead, following Ichimura (1993), we estimate ρ using kernel regression with kernel function $K(\cdot)$ and bandwidth h_n,

$$\widehat{\rho}^*(u;\boldsymbol{\beta}) = \frac{\sum_{s\in X\cap W_n} K(|\,\mathbf{Z}(s)^{\mathrm{T}}\boldsymbol{\beta} - u|/h_n)}{\int_{W_n} K(|\,\mathbf{Z}(t)^{\mathrm{T}}\boldsymbol{\beta} - u|/h_n)\, dt}, \tag{2.2}$$

and estimate $\boldsymbol{\beta}_0$ using

$$\widehat{\boldsymbol{\beta}}_n = \mathrm{argmax}_{\|\,\boldsymbol{\beta}\,\|=1}\widehat{l}_n(\boldsymbol{\beta}), \tag{2.3}$$

where

$$\widehat{l}_n(\boldsymbol{\beta}) = \frac{1}{|W_n|}\left[\sum_{s\in X\cap W_n} \log \widehat{\rho}^*(\mathbf{Z}(s)^{\mathrm{T}}\boldsymbol{\beta};\boldsymbol{\beta}) - \int_{W_n} \widehat{\rho}^*(\mathbf{Z}(t)^{\mathrm{T}}\boldsymbol{\beta};\boldsymbol{\beta})\, dt\right]. \tag{2.4}$$

Here $\widehat{\rho}^*$ estimates the conditional intensity function given $\mathbf{Z}(s)^{\mathrm{T}}\,\boldsymbol{\beta} = u$,

$$\begin{aligned}\rho^*(u;\boldsymbol{\beta}) &= \frac{\lim_{|ds|\to 0} E\{X(ds)\big|\,\mathbf{Z}(s)^{\mathrm{T}}\,\boldsymbol{\beta} = u\}}{|ds|}\\ &= \frac{\lim_{|ds|\to 0} E\big\{E[X(ds)|\,\mathbf{Z}(s)]\big|\,\mathbf{Z}(s)^{\mathrm{T}}\,\boldsymbol{\beta} = u\big\}}{|ds|}\end{aligned}$$

$$= E\{\rho(\mathbf{Z}(s)^{\mathrm{T}}\boldsymbol{\beta}_0)|\,\mathbf{Z}(s)^{\mathrm{T}}\,\boldsymbol{\beta} = u\}. \tag{2.5}$$

The estimating procedure (2.3) can also be obtained by minimizing the Kullback-Leibler (KL) distance between the conditional intensity function and the true intensity function, $\rho^*(u;\boldsymbol{\beta})$ and $\rho(u)$. Since $\lambda(s|\,\mathbf{Z}(s)) = \rho(\mathbf{Z}(s)^{\mathrm{T}}\,\boldsymbol{\beta}_0) = \rho^*(\mathbf{Z}(s)^{\mathrm{T}}\boldsymbol{\beta}_0;\boldsymbol{\beta}_0)$, the KL distance between it and $\rho^*(\mathbf{Z}(s)^{\mathrm{T}}\,\boldsymbol{\beta};\boldsymbol{\beta})$ is given by

$$d(\boldsymbol{\beta},\boldsymbol{\beta}_0) = \frac{1}{|W_n|}E\Bigg\{\Big[\sum_{s\in X\cap W_n}\log\rho^*(\mathbf{Z}(s)^{\mathrm{T}}\boldsymbol{\beta}_0;\boldsymbol{\beta}_0) - \int_{W_n}\rho^*(\mathbf{Z}(t)^{\mathrm{T}}\boldsymbol{\beta}_0;\boldsymbol{\beta}_0)\,dt\Big]$$
$$-\Big[\sum_{s\in X\cap W_n}\log\rho^*(\mathbf{Z}(s)^{\mathrm{T}}\,\boldsymbol{\beta};\boldsymbol{\beta}) - \int_{W_n}\rho^*(\mathbf{Z}(t)^{\mathrm{T}}\,\boldsymbol{\beta};\boldsymbol{\beta})\,dt\Big]\Bigg\}. \tag{2.6}$$

By the property of the KL distance, $d(\boldsymbol{\beta},\boldsymbol{\beta}_0) = 0$ implies that $\rho^*(\mathbf{Z}(s)^{\mathrm{T}}\boldsymbol{\beta};\boldsymbol{\beta}) = \rho(\mathbf{Z}(s)^{\mathrm{T}}\,\boldsymbol{\beta}_0)$. Under some regularity conditions on $\mathbf{Z}(\cdot)$ and $\rho(\cdot)$, such as Assumptions 4.1 and 4.2 in Ichimura (1993), this in turn implies that $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, suggesting that the model is identifiable using (1.4). Next, the KL distance $d(\boldsymbol{\beta},\boldsymbol{\beta}_0)$ between $\rho^*(\mathbf{Z}(s)^{\mathrm{T}}\boldsymbol{\beta};\boldsymbol{\beta})$ and $\rho^*(\mathbf{Z}(s)^{\mathrm{T}}\boldsymbol{\beta}_0;\boldsymbol{\beta}_0)$ can be estimated by

$$\widehat{d}(\boldsymbol{\beta},\boldsymbol{\beta}_0) = \frac{1}{|W_n|}\Bigg\{\Big[\sum_{s\in X\cap W_n}\log\rho^*(\mathbf{Z}(s)^{\mathrm{T}}\boldsymbol{\beta}_0;\boldsymbol{\beta}_0) - \int_{W_n}\rho^*(\mathbf{Z}(t)^{\mathrm{T}}\boldsymbol{\beta}_0;\boldsymbol{\beta}_0)\,dt\Big]$$
$$-\Big[\sum_{s\in X\cap W_n}\log\widehat{\rho}^*(\mathbf{Z}(s)^{\mathrm{T}}\,\boldsymbol{\beta};\boldsymbol{\beta}) - \int_{W_n}\widehat{\rho}^*(\mathbf{Z}(t)^{\mathrm{T}}\,\boldsymbol{\beta};\boldsymbol{\beta})\,dt\Big]\Bigg\}. \tag{2.7}$$

Ignoring the constant terms not depending on $\boldsymbol{\beta}$, the minimizer of (2.7) is equivalent to that of (2.4). Finally, $\rho(u)$ can be estimated by

$$\widehat{\rho}(u) = \widehat{\rho}^*(u;\widehat{\boldsymbol{\beta}}_n). \tag{2.8}$$

3. Theoretical Results

In this section we show the consistency and asymptotic normality of $\widehat{\boldsymbol{\beta}}$. Before stating the assumptions needed for the results, we need to define a mixing coefficient to specify the dependence structure of the point process X. Since we apply a spatial version of Bernstein's inequality developed in Zhu and Lahiri (2007), we adopt the definition of the mixing coefficient used there. For $S_1, S_2 \in \mathcal{B}(\mathbb{R}^2)$, let $\alpha_1(S_1,S_2) = \sup\big\{|Pr(A_1\cap A_2) - Pr(A_1)Pr(A_2)| : A_1 \in \mathcal{F}_X(S_1), A_2 \in \mathcal{F}_X(S_2)\big\}$, where S_1 and S_2 are any two subsets in $\mathbb{R}^2$ and $\mathcal{F}_X(S)$ is the σ-algebra generated by the variables $\{X(s) : s \in S\}$. Then the α-mixing coefficient is

$$\alpha(m;b) = \sup\big\{\alpha_1(S_1,S_2) : |S_1| \le b, |S_2| \le b, d(S_1,S_2) \ge m\big\},$$

where $|S|$ is the area of S, $d(S_1,S_2)$ is the minimal distance between S_1 and S_2.

3.1. Assumptions

We require some assumptions. In them, c and C are generic positive constants.

A1. Covariate $\{\mathbf{Z}(s) : s \in \mathbb{R}^2\}$ is a stationary process and $\| \mathbf{Z}(s)\| \leq C$.

A2. The observable window W_n is of the form $[a_1 n, a_2 n] \times [b_1 n, b_2 n]$, where $a_2 > a_1$ and $b_2 > b_1$. For sufficient large $k > 0$, $E\{X^k(W_1)\} < \infty$, where $X(W)$ is the number of events in window W.

A3. For any $\| \boldsymbol{\beta} \| = 1$, the density of $\mathbf{Z}(s)^{\mathrm{T}} \boldsymbol{\beta}$, denoted as $f(u; \boldsymbol{\beta})$, is three times continuously differentiable with respect to u. Also, $\inf_{\| \boldsymbol{\beta} \|=1} f(\mathbf{Z}(s)^{\mathrm{T}} \boldsymbol{\beta}; \boldsymbol{\beta}) \geq c > 0$.

A4. The function $\rho^*(u; \boldsymbol{\beta})$ is continuously differentiable with respect to u and $\boldsymbol{\beta}$, $0 < c \leq \rho^*(u; \boldsymbol{\beta}) \leq C$, and $\|(\partial \rho^*(u; \boldsymbol{\beta})/\partial u, \partial \rho^*(u; \boldsymbol{\beta})/\partial \boldsymbol{\beta}^{\mathrm{T}})\| \leq C$.

A5. The α-mixing coefficient $\alpha(m; b)$ of $X(\cdot)$ satisfies $\alpha(m; b) \leq C m^{-\tau} b^{\delta}$, for some $\tau > 0$ and $\delta > 0$.

A6. The α-mixing coefficient $\tilde{\alpha}(m; b)$ of $\mathbf{Z}(\cdot)$ satisfies $\tilde{\alpha}(m; b) \leq C m^{-\tau} b^{\delta}$, for some $\tau > 0$ and $\delta > 0$.

A7. The kernel $K(u)$ is symmetric, twice continuously differentiable and the second derivative satisfies a Lipschitz condition; $\int K(u) du = 1$ and $K(u) = 0$ for $|u| > 1$.

Remark 1. In Assumption A1, the stationarity of the covariate process is added for mathematical convenience in stating some of the limits, such as the last term in (3.2).

Remark 2. The boundedness assumption A1 for the covariate process is also for convenience; it was also assumed in Waagepetersen and Guan (2009) and in Ichimura (1993).

Remark 3. The form of the window W_n in Assumption A2 was assumed in Waagepetersen and Guan (2009). The results can be easily generalized to any window $[a_1 \lambda_n, a_2 \lambda_n] \times [b_1 \lambda_n, b_2 \lambda_n]$, where $\lambda_n \to \infty$ as $n \to \infty$. In addition, the moment assumption is satisfied if the distribution of $X(W_1)$ is light-tailed. We determine the value of k in the proof.

Remark 4. The Assumptions A3, A4 and A7 are standard for single-index models. Similar assumptions were made in Ichimura (1993).

REMARK 5. The mixing coefficient of X is needed only for using the spatial version of Bernstein's inequality given in Lemma 4.7 of Zhu and Lahiri (2007). Assumptions A5 and A6 are needed in order to apply this inequality. We determine the values of m and b in the proof.

3.2. Consistency

Theorem 1. *Under Assumptions A1-A7, if $n \to \infty$, $h_n \to 0$ and $nh_n^2 \to \infty$, then $\widehat{\boldsymbol{\beta}}_n \to \boldsymbol{\beta}_0$ in probability.*

Proof. Take

$$l_n(\boldsymbol{\beta}) = \frac{1}{|W_n|}[\sum_{s\in X\cap W_n} \log \rho^*(\mathbf{Z}(s)^{\mathrm{T}}\boldsymbol{\beta};\boldsymbol{\beta}) - \int_{W_n} \rho^*(\mathbf{Z}(t)^{\mathrm{T}}\boldsymbol{\beta};\boldsymbol{\beta})\,dt]. \tag{3.1}$$

Then, under Assumption A1,

$$\begin{aligned} E\{l_n(\boldsymbol{\beta})\} &= \frac{1}{|W_n|}E\left\{\int_{W_n} \log \rho^*(\mathbf{Z}(t)^{\mathrm{T}}\boldsymbol{\beta};\boldsymbol{\beta})\rho(\mathbf{Z}(t)^{\mathrm{T}}\boldsymbol{\beta}_0)dt - \int_{W_n} \rho^*(\mathbf{Z}(t)^{\mathrm{T}}\boldsymbol{\beta};\boldsymbol{\beta})\,dt\right\} \\ &= \frac{1}{|W_1|}E\left\{\int_{W_1} \log \rho^*(\mathbf{Z}(t)^{\mathrm{T}}\boldsymbol{\beta};\boldsymbol{\beta})\rho(\mathbf{Z}(t)^{\mathrm{T}}\boldsymbol{\beta}_0)dt - \int_{W_1} \rho^*(\mathbf{Z}(t)^{\mathrm{T}}\boldsymbol{\beta};\boldsymbol{\beta})\,dt\right\}. \end{aligned} \tag{3.2}$$

By definition, $Pr\{\widehat{l}_n(\widehat{\boldsymbol{\beta}}_n) \geq \widehat{l}_n(\boldsymbol{\beta}_0)\} = 1$. So, since we have

$$\begin{aligned} 1 = Pr\{\widehat{l}_n(\widehat{\boldsymbol{\beta}}_n) \geq \widehat{l}_n(\boldsymbol{\beta}_0) \text{ and } \widehat{\boldsymbol{\beta}}_n \in B(\boldsymbol{\beta}_0)\} + Pr\{\widehat{l}_n(\widehat{\boldsymbol{\beta}}_n) \geq \widehat{l}_n(\boldsymbol{\beta}_0) \\ \text{and } \widehat{\boldsymbol{\beta}}_n \notin B(\boldsymbol{\beta}_0)\} \leq Pr\{\widehat{\boldsymbol{\beta}}_n \in B(\boldsymbol{\beta}_0)\} + Pr\{\sup_{\boldsymbol{\beta}\notin B(\boldsymbol{\beta}_0)} \widehat{l}_n(\boldsymbol{\beta}) \geq \widehat{l}_n(\boldsymbol{\beta}_0)\}, \end{aligned}$$

for any small open ball $B(\boldsymbol{\beta}_0)$ that includes $\boldsymbol{\beta}_0$, Theorem 1 follows if

$$Pr\{\sup_{\boldsymbol{\beta}\notin B(\boldsymbol{\beta}_0)} \widehat{l}_n(\boldsymbol{\beta}) \geq \widehat{l}_n(\boldsymbol{\beta}_0)\} \to 0. \tag{3.3}$$

With the identifiability, $E\{l_n(\boldsymbol{\beta}_0)\} - \sup_{\boldsymbol{\beta}\notin B(\boldsymbol{\beta}_0)} E\{l_n(\boldsymbol{\beta})\} > \varepsilon$, for some $\varepsilon > 0$ and for any n. Now, because

$$\begin{aligned} &Pr\{\sup_{\boldsymbol{\beta}\notin B(\boldsymbol{\beta}_0)} \widehat{l}_n(\boldsymbol{\beta}) \geq \widehat{l}_n(\boldsymbol{\beta}_0)\} \\ &= Pr\{\sup_{\boldsymbol{\beta}\notin B(\boldsymbol{\beta}_0)} [\widehat{l}_n(\boldsymbol{\beta}) - l_n(\boldsymbol{\beta}) + l_n(\boldsymbol{\beta}) - E\{l_n(\boldsymbol{\beta})\} + E\{l_n(\boldsymbol{\beta})\}] \geq \widehat{l}_n(\boldsymbol{\beta}_0)\} \\ &\leq Pr\{\sup_{\boldsymbol{\beta}\notin B(\boldsymbol{\beta}_0)} [\widehat{l}_n(\boldsymbol{\beta}) - l_n(\boldsymbol{\beta})] + \sup_{\boldsymbol{\beta}\notin B(\boldsymbol{\beta}_0)} [l_n(\boldsymbol{\beta}) - E\{l_n(\boldsymbol{\beta})\}] \\ &\qquad + [E\{l_n(\boldsymbol{\beta}_0)\} - \widehat{l}_n(\boldsymbol{\beta}_0)] \geq E\{l_n(\boldsymbol{\beta}_0)\} - \sup_{\boldsymbol{\beta}\notin B(\boldsymbol{\beta}_0)} E\{l_n(\boldsymbol{\beta})\}\}, \end{aligned}$$

(3.3) follows if we can show that

$$Pr\Big\{ \sup_{\|\boldsymbol{\beta}\|=1} \Big|\widehat{l}_n(\boldsymbol{\beta}) - l_n(\boldsymbol{\beta})\Big| \geq \frac{\varepsilon}{3}\Big\} \to 0; \tag{3.4}$$

$$Pr\Big\{ \sup_{\|\boldsymbol{\beta}\|=1} |l_n(\boldsymbol{\beta}) - E\{l_n(\boldsymbol{\beta})\}| \geq \frac{\varepsilon}{3}\Big\} \to 0; \tag{3.5}$$

$$Pr\Big\{ \Big|E\{l_n(\boldsymbol{\beta}_0)\} - \widehat{l}_n(\boldsymbol{\beta}_0)\Big| \geq \frac{\varepsilon}{3}\Big\} \to 0. \tag{3.6}$$

Since (3.6) is implied by (3.4) and (3.5), it suffices to prove (3.4) and (3.5), which are respectively implied by Lemmas A.7 and A.8 in the Appendix available in the online Supplementary Materials.

3.3. Asymptotic normality

Let $\boldsymbol{\beta}_0 = (\beta_{01}, \ldots, \beta_{0p})^{\mathrm{T}}$, with $\|\boldsymbol{\beta}_0\| = 1$. Without loss of generality, assume that $\beta_{01} \neq 0$. We reparameterize the parameters as $\boldsymbol{\beta}_0^{\mathrm{T}} = \beta_{01}(1, \check{\boldsymbol{\beta}}_0^{\mathrm{T}})$ and $\rho(\mathbf{Z}(s)^{\mathrm{T}}\boldsymbol{\beta}_0) = \check{\rho}(Z_1(s) + \check{\mathbf{Z}}(s)^{\mathrm{T}}\check{\boldsymbol{\beta}}_0)$, where $\mathbf{Z}^{\mathrm{T}} = (Z_1, \check{\mathbf{Z}}^{\mathrm{T}})$. In this section, for notational simplicity, we use $\boldsymbol{\beta}$, ρ and $\mathbf{Z}$ to refer to $\check{\boldsymbol{\beta}}$, $\check{\rho}$ and $\check{\mathbf{Z}}$ respectively.

Let $\rho_{\boldsymbol{\beta}_0,2}(s,t)$ denote the second-order product density of X. The second-order product density describes the behavior of point pairs in X. See, e.g. Diggle (2003). The pair correlation function is related to the second-order product density and is defined as

$$g(s,t) = \frac{\rho_{\boldsymbol{\beta}_0,2}(s,t)}{\rho(\mathbf{Z}(s)^{\mathrm{T}}\boldsymbol{\beta}_0)\rho(\mathbf{Z}(t)^{\mathrm{T}}\boldsymbol{\beta}_0)}. \tag{3.7}$$

Theorem 2. *Under Assumptions A1-A7, if $n \to \infty$, $h_n \to 0$ and $nh_n^4 \to \infty$, then*

$$|W_n|^{1/2}\Sigma_n^{-1/2}V_n(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{D} \mathrm{MVN}(\mathbf{0}, I);$$

where

$$V_n = \frac{1}{|W_n|}\int_{W_n} \frac{[\rho'(\mathbf{Z}(s)^{\mathrm{T}}\boldsymbol{\beta}_0)]^2\widetilde{\mathbf{Z}}(s)\widetilde{\mathbf{Z}}(s)^{\mathrm{T}}}{\rho(\mathbf{Z}(s)^{\mathrm{T}}\boldsymbol{\beta}_0)}\,ds,$$

$$\Sigma_n = V_n + \frac{1}{|W_n|}\int_{W_n\times W_n} \rho'(\mathbf{Z}(s)^{\mathrm{T}}\boldsymbol{\beta}_0)\rho'(\mathbf{Z}(t)^{\mathrm{T}}\boldsymbol{\beta}_0)\widetilde{\mathbf{Z}}(s)\widetilde{\mathbf{Z}}(t)^{\mathrm{T}}[g(s,t)-1]\,ds\,dt,$$

where $\widetilde{\mathbf{Z}}(s) = \mathbf{Z}(s) - E\{\mathbf{Z}(s)|\,\mathbf{Z}(s)^{\mathrm{T}}\boldsymbol{\beta}_0\}$.

Proof. By Theorem 1, with probability close to one, for $\overline{\boldsymbol{\beta}}$ between $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$,

$$\widehat{l}_n(\widehat{\boldsymbol{\beta}}_n) = \widehat{l}_n(\boldsymbol{\beta}_0) + \frac{d\widehat{l}_n(\boldsymbol{\beta}_0)}{d\,\boldsymbol{\beta}^{\mathrm{T}}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^{\mathrm{T}}\frac{d^2\widehat{l}_n(\overline{\boldsymbol{\beta}})}{d\,\boldsymbol{\beta}\,d\,\boldsymbol{\beta}^{\mathrm{T}}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0).$$

Then, by the same arguments used in the proof of Theorem 5.2 in Ichimura (1993), Theorem 2 follows from Lemmas A.14-A.16.

3.4. Efficiency

Here we discuss without rigorous proof the relative efficiency of the estimator (2.3). If X is a Poisson process, then $\widehat{\boldsymbol{\beta}}_n$ is an efficient estimate of $\boldsymbol{\beta}_0$ in the sense of asymptotic efficiency in Newey (1994). To see this, assume that the true unknown nonparametric component $\rho(\cdot)$ is in some compact set. If X is a Poisson process, the log-likelihood for ρ and $\boldsymbol{\beta}_0$ is

$$l_n(\varrho, \boldsymbol{\beta}) = \frac{1}{|W_n|}[\sum_{s \in X \cap W_n} \log \varrho(\mathbf{Z}(s)^{\mathrm{T}} \boldsymbol{\beta}) - \int_{W_n} \varrho(\mathbf{Z}(t)^{\mathrm{T}} \boldsymbol{\beta})\, dt]. \tag{3.8}$$

For fixed $\boldsymbol{\beta}$, let $\varrho(\cdot; \boldsymbol{\beta})$ denote the one that minimizes

$$E\{l_n(\varrho, \boldsymbol{\beta})\} = \frac{1}{|W_n|} E[\int_{W_n} \log \varrho(\mathbf{Z}(t)^{\mathrm{T}} \boldsymbol{\beta}) \rho(\mathbf{Z}(s)^{\mathrm{T}} \boldsymbol{\beta}_0)\, dt - \int_{W_n} \varrho(\mathbf{Z}(t)^{\mathrm{T}} \boldsymbol{\beta})\, dt]. \tag{3.9}$$

The curve $\{\varrho(\cdot; \boldsymbol{\beta}) : \| \boldsymbol{\beta} \| = 1\}$ is a least favorable curve in the sense of Stein (1956), and it can be shown that $\varrho(u; \boldsymbol{\beta}) = \rho^*(u; \boldsymbol{\beta})$. By Severini and Wong (1992), since $\rho^*(\cdot; \boldsymbol{\beta})$ is a least favorable curve, the profile estimator $\widehat{\boldsymbol{\beta}}_n$ is an efficient estimator of $\boldsymbol{\beta}$.

If X is not a Poisson process, the efficiency lost resulting from using the objective function (2.1) instead of its true log-likelihood can be measured using $D_n = \Sigma_n - V_n$. Following the proof in Appendix A.2, we can show that

$$D_n = \frac{1}{|W_n|} \int_{W_n \times W_n} \rho'(\mathbf{Z}(s)^{\mathrm{T}} \boldsymbol{\beta}_0) \rho'(\mathbf{Z}(t)^{\mathrm{T}} \boldsymbol{\beta}_0) \widetilde{\mathbf{Z}}(s) \widetilde{\mathbf{Z}}(t)^{\mathrm{T}} [g(s,t) - 1]\, ds\, dt \geq 0.$$

4. Implementation

The integrals in both (2.2) and (2.4) can be approximated using a method proposed in Berman and Turner (1992) and Baddeley and Turner (2000). The same approximation method was also used in Guan (2008). Let $w_1, \cdots, w_N$ be a set of non-overlapping small windows with $\cup_{i=1}^N w_i = W_n$, such that any point in X is in $\{w_1, \cdots, w_N\}$, and each window contains at most one point in X. Let $t_i \in w_i, i = 1, \cdots, N$ be representative locations of w_i, e.g. the centroids of w_i. Then, the integral $\int_{W_n} \zeta(t) dt$, for function $\zeta : \mathbb{R}^2 \mapsto \mathbb{R}$, can be approximated by

$$\int_{W_n} \zeta(t) dt = \sum_{i=1}^{N} \zeta(t_i) |w_i|. \tag{4.1}$$

As pointed out by Guan (2008), the use of (4.1) presents little constraint in practice since the covariate process $\mathbf{Z}(\cdot)$ is often observed only at some discrete locations. In this case, each t_i corresponds to a location where $\mathbf{Z}(\cdot)$ is observed

and w_i is chosen to be the set associated with t_i. Specifically,

$$\int_{W_n} K(\frac{|\,\mathbf{Z}(t)^{\mathrm{T}}\boldsymbol{\beta}-u|}{h_n})dt \doteq \sum_{i=1}^{N} K(\frac{|\,\mathbf{Z}(t_i)^{\mathrm{T}}\boldsymbol{\beta}-u|}{h_n})|w_i|, \tag{4.2}$$

$$\int_{W_n} \widehat{\rho}^{*}(\mathbf{Z}(t)^{\mathrm{T}}\boldsymbol{\beta};\boldsymbol{\beta})dt \doteq \sum_{i=1}^{N} \widehat{\rho}^{*}(\mathbf{Z}(t_i)^{\mathrm{T}}\boldsymbol{\beta};\boldsymbol{\beta})|w_i|. \tag{4.3}$$

Based on (4.2) and (4.3), the estimator at (2.3) can be calculated numerically. The R code implementing the proposed method is available as an online supplementary file.

5. Numerical Results

5.1. Simulation

We performed a simulation study to examine the empirical properties of our estimators (2.3) and (2.8) for the single-index model (1.4). The point patterns and covariates are simulated using functions available from the R packages `spatstat` and `RandomFields`. R code to estimate $\boldsymbol{\beta}_0$ and ρ is available from the authors upon request.

We first obtained three independent covariates Z_1, Z_2, and Z_3 on the 4×4 square region using realizations from a Gaussian random field with covariance function given by $\sigma^2 \exp(-d/d_0)$. We used $(\sigma^2, d_0) = (0.05, 0.4), (0.08, 0.8)$ and $(0.1, 1)$ respectively for Z_1, Z_2 and Z_3. This was done using the `RFsimulate` function in the `RandomFields` R package. With the covariate vector $\mathbf{Z}(s) = (1, Z_1(s), Z_2(s), Z_3(s))^{\mathrm{T}}$, we then defined intensity functions corresponding to three forms for ρ. Specifically,

$$\begin{aligned}
\lambda_1(s) &= \exp(\mathbf{Z}(s)^{\mathrm{T}}\boldsymbol{\beta}_0),\\
\lambda_2(s) &= 10 + \frac{200}{1 + b_2 \exp(-\,\mathbf{Z}(s)^{\mathrm{T}}\boldsymbol{\beta}_0)},\\
\lambda_3(s) &= 25\left[\mathbf{Z}(s)^{\mathrm{T}}\boldsymbol{\beta}_0 + \sin\left(\frac{\pi}{2}\,\mathbf{Z}(s)^{\mathrm{T}}\boldsymbol{\beta}_0\right) + b_3\right].
\end{aligned}$$

We considered three different sets of values for $\boldsymbol{\beta}_0$, specifically $\boldsymbol{\beta}_0^{\mathrm{T}} = (3.5, 3.5, 3.5, 3.5)$, $(1, 2, 4, 8)$ and $(2.5, 8, 4, 2)$, to form the indices $\mathbf{Z}(s)^{\mathrm{T}}\boldsymbol{\beta}_0$. For each specific $\boldsymbol{\beta}_0$, the values of b_2 and b_3 were chosen so that inhomogeneous Poisson realizations generated in the 4×4 window from the intensity functions had between 2,500 and 3,000 points. Specifically, we used $(b_2, b_3) = (8, 2.5), (2, 4), (2, 2.5)$ for $\boldsymbol{\beta}_0^{\mathrm{T}} = (3.5, 3.5, 3.5, 3.5), (1, 2, 4, 8), (2.5, 8, 4, 2)$, respectively.

For each combination of $\boldsymbol{\beta}_0$ value and intensity function $\lambda_i, i = 1, 2, 3$, 100 inhomogeneous Poisson realizations were generated on the 4×4 observation win-

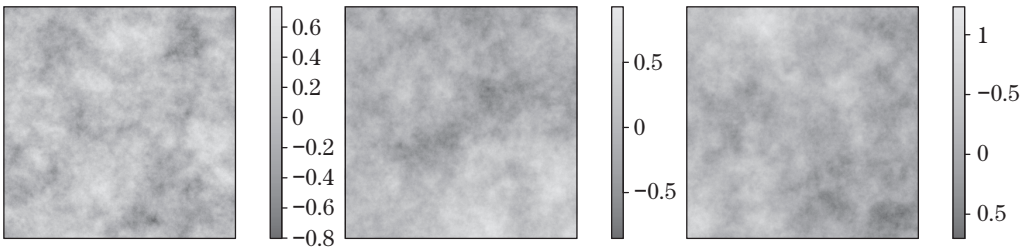

Figure 1. Plots of the three covariates Z_1, Z_2 and Z_3 used in the simulation study.

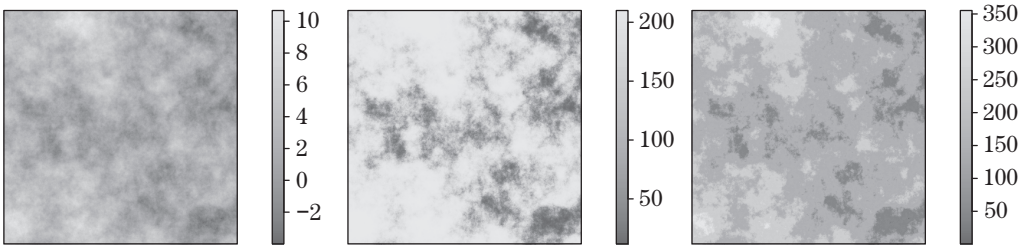

Figure 2. The intensity functions λ_1, λ_2 and λ_3 when $\boldsymbol{\beta}_0 = (1, 2, 4, 8)$. Note that λ_1 is plotted on the log scale.

dow. We also extracted from these realizations smaller point patterns corresponding to the points that fell within the lower left 2×2 and 1×1 squares. With each point pattern, we used (2.3) and (2.8), respectively, to obtain $\widehat{\boldsymbol{\beta}}_0$ and $\widehat{\rho}$. The bandwidths needed for the kernel regression estimation procedure for ρ were selected using a dynamic application of the rule of thumb provided in Silverman (1986) as follows. For every new interim value of $\widetilde{\boldsymbol{\beta}}$ in the optimization procedure, the standard deviation $\widetilde{\sigma}$ of the index $\mathbf{Z}(s)^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}$ for $s \in X$ was computed and a bandwidth $h = 1.06\widetilde{\sigma}N^{-1/5}$ was used with the Gaussian kernel, where N was the number of points in the point pattern.

Since ρ is non-parametric, our estimate $\widehat{\rho}$ can adapt to any scale of the estimates $\widehat{\boldsymbol{\beta}}$. Hence in our fitting procedure, we excluded the intercept and set the coefficient of Z_1 to 1. The resulting estimates of the coefficients of Z_2 and Z_3 were thus β_2/β_1 and β_3/β_1 rather than the original β_2 and β_3. The initial values of $\boldsymbol{\beta}$ used for the procedure were obtained by rescaling the estimates (ignoring the intercept) obtained from applying the `ppm` function in the R package `spatstat`, which fits the log-linear model to the data.

Figure 1 shows plots of the covariates Z_1, Z_2, and Z_3 and Figure 2 shows the resulting intensity functions for $\boldsymbol{\beta}_0 = (1, 2, 4, 8)^{\mathrm{T}}$. Plots of the estimates $\widehat{\rho}$ are shown in Figure 3 for simulated data on the 4×4 observation window. The left,

middle and right columns correspond to λ_1, λ_2, and λ_3, respectively, while each row of plots corresponds to a different value of $\boldsymbol{\beta}_0$ used. The point-wise mean and two standard error limits are shown in thin dashed dark gray lines. The solid black line shows the true function ρ, while the dashed black line shows what the corresponding exponential function would look like if it replaced ρ directly.

We find that ρ is well estimated for all the three intensity functions considered in the simulation. There is some, but not very large, variability in performance between the different true values of $\boldsymbol{\beta}_0$. Performance is best at the mid-range values of the index, where there is more information provided by the data. At the extreme ends of the range of the index, especially at the higher end, there is much more variability in $\widehat{\rho}$.

The intensity function λ_2 is rather flat for a significant range of the index values, as highlighted in the middle plot of Figure 2, which shows large portions of the region with almost constant intensity. This may make it challenging for any estimation procedure to capture the relationship to the covariates. The intensity function λ_3 mostly increases with the index, but is not monotone, due to the sine function in its expression. However, estimates $\widehat{\rho}$ manage to approximately capture the functional form of λ_2 and λ_3.

The estimates $\widehat{\rho}$ obtained from point data on the smaller window sizes (not shown) also follow the true ρ pretty well. For these smaller data sizes, performance is noticeably better for the exponential function λ_1 than for λ_2 or λ_3, in the sense that estimates for the latter intensity functions show significantly more variability. Judging from Figure 2, part of the underperformance may be due to insufficient structure in the intensity function on the smaller windows, especially for λ_2.

In order to assess the performance of the procedure for estimating the intensity function, we define the integrated square error (ISE) for estimating λ as

$$\text{ISE} = \int_W \left(\widehat{\lambda}(s) - \lambda(s)\right)^2 ds, \tag{5.1}$$

where $\lambda(s)$ is the true value of the intensity function at location $s \in W$ and $\widehat{\lambda}(s)$ the corresponding estimate, obtained either from the fitted single-index model or the fitted log-linear model. Since we only have the intensity values on a grid in W, we approximate the ISE by replacing the integral in (5.1) with a sum of the corresponding values over the grid cells. The mean integrated square error (MISE) is then obtained by averaging the estimated ISE over the 100 realizations for each set of parameter values.

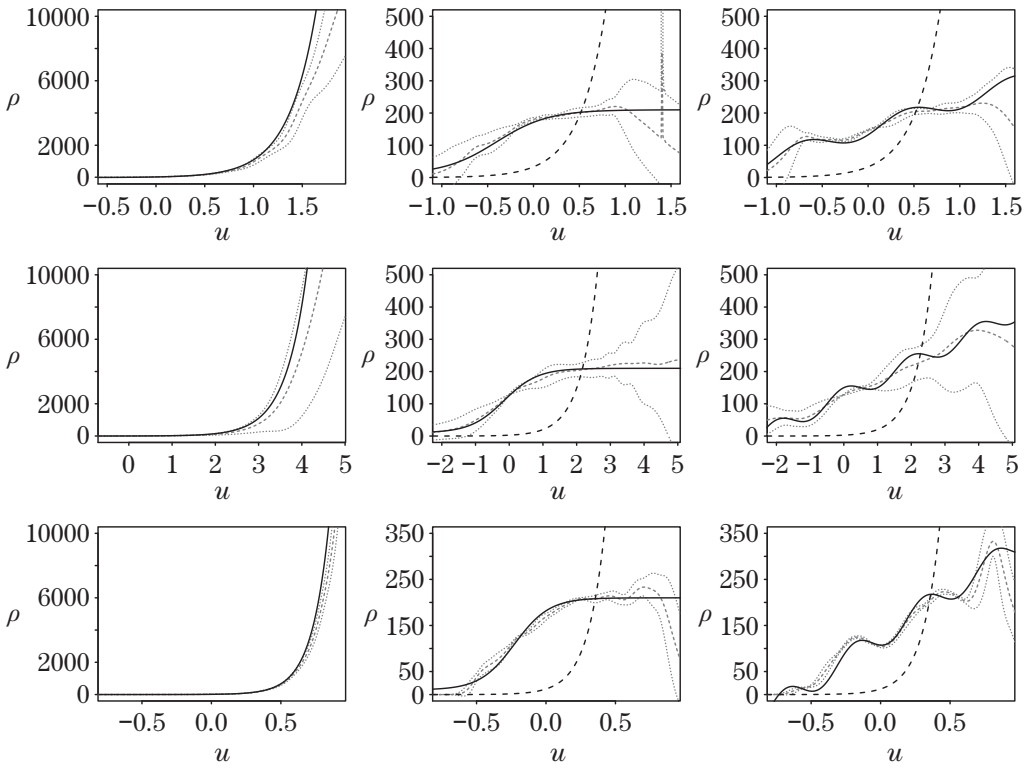

Figure 3. Estimates $\widehat{\rho}$ for λ_1, λ_2 and λ_3 (left, middle and right columns respectively) for $\boldsymbol{\beta}_0 = (3.5, 3.5, 3.5, 3.5), (1, 2, 4, 8), (2.5, 8, 4, 2)$ (top, middle and bottom rows respectively), obtained from inhomogeneous Poisson point data on the 4×4 observation window, with point-wise mean and two standard error limits shown in dashed gray lines. The solid and dashed black curves show the true function and the exponential function respectively.

Table 1 shows the ratio of MISE values obtained from the single-index model to the MISE values obtained from the log-linear model, for the three forms of the intensity function, λ_1 to λ_3, the three sets of $\boldsymbol{\beta}_0$ values, and the three observation window sizes. Values smaller (larger) than 1 correspond to out-performance (under-performance) of the single-index model compared with the log-linear model.

Not surprisingly, we find that for λ_1, where the log-linear model is the correct model, using the single-index model yielded a higher MISE. With the other two intensity functions, however, while there is no clear better performing model for the 1×1 window, the single-index model performs better for all three $\boldsymbol{\beta}_0$ values with the 2×2 window, with even greater out-performance with the 4×4 window.

Table 1. Ratio of mean integrated square errors (MISE) obtained from the single-index model to those from the log-linear model.

$\boldsymbol{\beta}_0$		Window size 4×4	2×2	1×1
$(3.5, 3.5, 3.5, 3.5)$	λ_1	4.26	4.03	2.93
	λ_2	0.10	0.40	23.36
	λ_3	0.04	0.27	2.14
$(1, 2, 4, 8)$	λ_1	1.75	1.77	2.97
	λ_2	0.02	0.13	3.06
	λ_3	0.02	0.20	0.75
$(2.5, 8, 4, 2)$	λ_1	2.08	4.86	6.28
	λ_2	0.04	0.27	1.73
	λ_3	0.02	0.14	0.29

Table 2. Ratio of median absolute errors for estimates of β_2/β_1 and β_3/β_1 in the intensity functions λ_1, λ_2 and λ_3 obtained using the single-index model estimation procedure to the corresponding errors obtained from the log-linear model.

True values	Window size	λ_1		λ_2		λ_3	
		β_2/β_1	β_3/β_1	β_2/β_1	β_3/β_1	β_2/β_1	β_3/β_1
$(3.5, 3.5, 3.5)$	4×4	1.20	1.29	0.93	1.23	0.97	1.12
	2×2	1.20	1.41	1.10	1.02	0.89	1.29
	1×1	1.04	1.24	1.13	1.02	1.29	0.97
$(2, 4, 8)$	4×4	2.05	1.41	0.51	0.52	0.59	0.67
	2×2	1.07	1.38	0.91	1.84	1.04	1.20
	1×1	1.46	1.43	1.10	0.98	1.30	1.09
$(8, 4, 2)$	4×4	1.47	0.93	0.97	0.79	0.68	0.80
	2×2	1.32	1.20	0.81	1.57	0.73	1.05
	1×1	1.07	1.06	0.95	1.06	1.00	1.08

We also examined the individual grid cell mean square errors, found without integrating over the window. Figure 4 shows the distributions of these mean square errors for the different intensity functions and observation windows when $\boldsymbol{\beta}_0 = (3.5, 3.5, 3.5, 3.5)^{\mathrm{T}}$. The dark gray histograms correspond to the single-index model, while the light, translucent gray histograms overlaid on top correspond to the log-linear model. For the exponential function (top row), the histograms for the log-linear model, the correct model, are located slightly to the left of the histograms for the single-index model.

The histograms are more interesting for the other two intensity functions. Focusing on the centers of the histograms, the log-linear model seems to attain smaller MSEs with the 1×1 window, but the difference decreases with the larger window sizes. However, the histograms for the log-linear model exhibit a lot

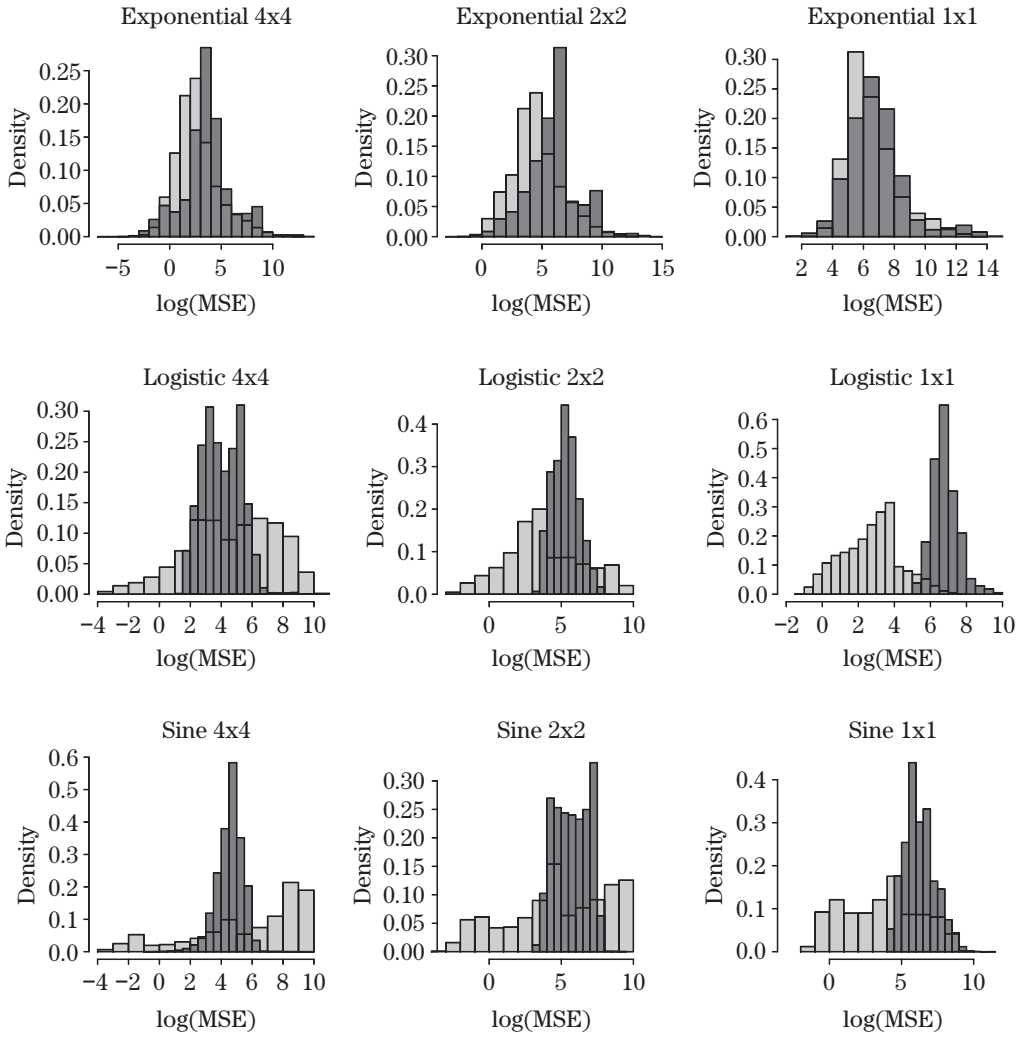

Figure 4. Distributions of grid cell mean square errors (MSEs) over the spatial locations of the observation windows. The dark histograms represent MSEs from the single-index model and the light translucent histograms those from the log-linear model. Overlapping portions appear as medium gray.

more variability in MSEs across the observation window than those for the single-index model. We believe this indicates the mis-fit of the log-linear model - the `ppm` fitting procedure manages to fit the model to some parts of the intensity function at the expense of other portions.

Finally, we also computed the errors of β_2/β_1 and β_3/β_1 for all three intensity functions λ_1 to λ_3 and values of $\boldsymbol{\beta}_0$ we considered. We note that there is quite a lot of variability in the coefficient estimates for both the single-index and log-

linear models, so that some of the absolute errors are very large. Hence we report the median absolute errors rather than the mean square errors.

Table 2 shows the ratio of the median absolute errors for the single-index model to the log-linear model. For the intensity function λ_1, where the log-linear model is the correct model, we find, not surprisingly, that the ratios of median absolute errors are larger than 1 for all observation window sizes and $\boldsymbol{\beta}_0$ values. With the other two intensity functions, however, the relative performance of the single-index model to the log-linear model improves as the observation window gets larger. There is some variability, however, in the relative performance with the true $\boldsymbol{\beta}_0$ values.

In summary, the single-index model is useful when there is reason to believe that the exponential function in the log-linear model is not the correct function. In such cases, using the single-index model not only allows estimation of the unknown function ρ but can improve estimation of the intensity function as well as of the coefficients, especially when the data size is large. If the resulting estimate of ρ suggests that an exponential function is a reasonable choice, the usual estimation procedure using `ppm` for the log-linear model can then be employed.

5.2. Application to fast-food restaurant dataset

We applied the single-index model and its estimation procedure to a dataset consisting of the locations of fast food restaurants (FFR) in New York City. The 817 FFRs were identified using an on-line directory of restaurant inspections from the New York City Department of Health and Mental Hygiene. FFRs were identified as national chains or local establishments that fit a number of criteria, such as not providing table service, but serving customers at a cash register or drive-thru window, requiring payment before eating, and with primary menu items consisting of hamburgers, hot dogs and fried chicken. The identified FFRs include national chains such as McDonald's, Burger King, and Wendy's as well as local chains such as Crown Fried Chicken and Kennedy Fried Chicken.

Studies have found links between fast food consumption and obesity-related measures, e.g. Bowman and Vinyard (2004), Alter and Eny (2005), Jeffrey et al. (2006), and Currie et al. (2009). Kwate et al. (2009) analyzed a FFR dataset similar to ours, and examined how the prevalence of FFRs varies with NYC neighborhood demographic variables. They found, in agreement with studies such as Zenk et al. (2005) and Inagami et al. (2006), that neighborhoods with higher poverty and/or higher percentage of African Americans tend to have more FFRs. In addition, studies such as Austin et al. (2005), Simon et al. (2008),

Table 3. Coefficient estimates for the FFR data, estimated using the log-linear and single-index models, rescaled so that the coefficient of age is 1.

	Age	Income	Pop Density	% Blk	Elementary	High School
Log−linear	1	0.24	−10.00	−3.96	−6.84	−4.52
Single−index	1	−0.64	−19.94	−2.39	−8.37	−12.78

Sturm (2008), Zenk and Powell (2008), Davis and Carpenter (2009), Kwate and Loh (2010), and Neckerman et al. (2010) suggest that FFRs may cluster around schools.

Besides FFR location data, we used 2,000 Census data to provide information, at the block group level, of the following: percent Black, median household income, median age, and population density. Furthermore, we have locations of public elementary and high schools, obtained from the NYC Department of Education. As a simple way to incorporate the school location data as covariates, we obtained kernel intensity estimates of the point processes of elementary schools and high schools, using a value of 250m for the standard deviation of the isotropic Gaussian smoothing kernel. The values of these kernel estimates at the FFR locations were then used as covariates in the single-index and log-linear models. We also set the demographic covariates for each FFR to be the Census variable values of the block group that the FFR resides in.

We modeled intensity as a function of these covariates using the log-linear model (with the `ppm` function in the `spatstat` R package), and using the single-index model. Table 3 shows the coefficient estimates obtained with the two models; we have rescaled them so that the estimate for "Age" is 1 in both cases. The sizes of these estimates cannot be directly compared since the effect of each covariate is modulated by the link function. However, with the exception of "Income", the estimates from the two models have the same signs. This suggests that both models find the same directional effects of the covariates on FFR intensity.

The plot on the left of Figure 5 shows the estimate of the link function ρ obtained from the single-index model estimation procedure, along with the exponential function of the log-linear model. Since the estimate for Age is negative for the log-linear model, the exponential function is reversed. The two functions are similar for positive values of the index down to about -20. With lower values of the index, the two functions behave differently, with $\widehat{\rho}$ showing a moderate increase before tapering off while the exponential function has a much steeper increase.

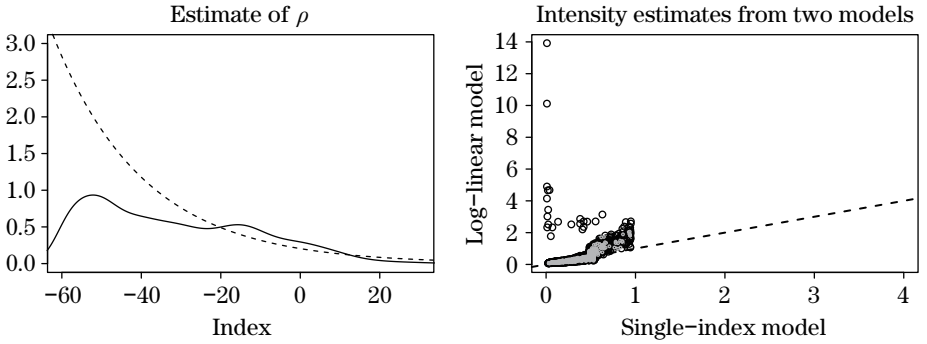

Figure 5. Left: estimate of ρ obtained from the FFR data set using the single-index model. The dashed line is the corresponding exponential function. Right: scatter-plot of intensity estimates obtained from the log-linear and single-index models.

Using the information provided in Table 3 and Figure 5, we find that the FFR intensity tends to be higher in block groups with higher population density, higher percent Black and/or more public elementary and high schools. With the single-index model, this increase in FFR intensity is limited, as indicated by the form of $\widehat{\rho}$ at the smaller index values and, in our opinion, could be more reasonable than the log-linear model, where it is possible to have extremely large FFR intensities.

The plot on the right of Figure 5 compares the intensity estimates obtained from the two models. The black and gray dots are estimates at the block group centroids and the FFR locations, respectively, with the dashed line representing equality of the two estimates. Many of the block groups and FFR locations have roughly equal estimates from the two models. As we might expect from the left plot of Figure 5, for the higher intensities, the log-linear model estimates tend to be larger than the single-index model estimates. More importantly, in several block groups which do not have FFRs, the single-index model gives low intensity estimates while the log-linear model yields very large estimates. Of course, these block groups may have FFRs nearby, in their neighboring block groups.

Finally, Figure 6 shows maps of the intensity estimates obtained from the single-index and log-linear models, with darker colors indicating higher estimated FFR intensity. Also included are the locations of the FFRs. The two maps are very similar with only subtle differences. For example, in the East, near the middle, the intensity estimates from the log-linear model is noticeably higher. There are also slightly more dark spots in Staten Island in the Southwest for the log-linear model. The strip of land on the Southeast part of the map is another location where differences are noticeable.

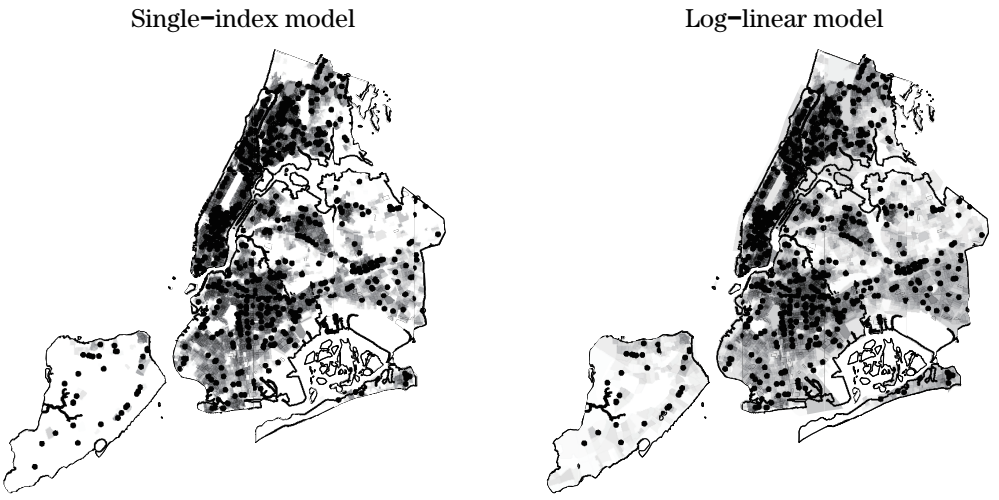

Figure 6. Maps showing FFR intensity estimates from the single-index and log-linear models, with FFR locations.

6. Discussion

One of our future directions is to extend the log-linear model through the use of the generalized additive model, which assumes a known link function but considers more flexible combinations of covariates, say $\lambda(s|\,\mathbf{Z}(s))=\exp\{\sum_{j=1}^{p} f_j(Z_j(s))\}$, where $f_j(\cdot)$ are unknown smooth functions. The generalized additive model and single-index model are two different way to generalize the log-linear model: the former model relaxes the linear predictor but keeps the known link, while the latter relaxes the link function but keeps the linear predictor. Another future direction is to develop a way to select informative covariates from a large list of available covariates. A possible way to proceed is to add a regularization term to the profile likelihood (2.4).

Supplementary Materials

The online supplementary materials contain statements and proofs of lemmas needed for Theorems 1 and 2, and a proof that D_n, the loss of efficiency if the process X is not Poisson, is non-negative.

Acknowledgment

The authors gratefully acknowledge Dr. Jinfeng Xu for his thoughtful comments on early drafts of this manuscript.

References

Alter, D. A. and Eny, K. (2005). The relationship between the supply of fast-food chains and cardivascular outcomes. *Canadian Journal of Public Health* **96**, 173–177.

Austin, S. B., Melly, S. J., Sanchez, B. N., Patel, A., Buka, S. and Gortmaker, S. L. (2005). Clustering of fast-food restaurants around schools: a novel application of spatial statistics to the study of food environments. *American Journal of Public Health* **95**, 1575–1581.

Baddeley, A. and Turner, T. R. (2000). Practical maximum pseudolikelihood for spatial patterns (with discussion). *Australian and New Zealand Journal of Statistics* **42**, 238–322.

Berman, M. and Turner, T. R. (1992). Approximating point process likelihoods with glim. *Applied Statistics* **41**, 31–38.

Bowman, S. A. and Vinyard, B. T. (2004). Fast food consumption of US adults: impact on energy and nutrient intakes and overweight status. *Journal of the American College of Nutrition* **23**, 163–168.

Currie, J., DellaVigna, S., Moretti, E. and Pathania, V. (2009). The effect of fast food restaurants on obesity. *National Bureau of Economic Research Working Paper* 14721.

Davis, B. and Carpenter, C. (2009). Proximity of fast-food restaurants to schools and adolescent obesity. *American Journal of Public Health* **99**, 505–510.

Diggle, P. J. (2003). *Statistical Analysis of Spatial Point Patterns*. Oxford University Press Inc., New York.

Guan, Y. (2008). On consistent nonparametric intensity estimation for inhomogeneous spatial point processes. *Journal of the American Statistical Association* **103**, 1238–1247.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Chapman and Hall, London; New York.

Horowitz, J. (1998). *Semiparametric Methods in Econometrics*. Springer, New York.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics* **58**, 71–120.

Illian, J. B., Sørbye, S. H., Rue, H. and Hendrichsen, D. (2012). Using INLA to fit a complex point process model with temporally varying effects. *Journal of Environmental Statistics* **3**.

Inagami, S., Cohen, D. A., Finch, B. K. and Asch, S. M. (2006). You are where you shop: grocery store locations, weight, and neighborhoods. *American Journal of Preventive Medicine* **31**, 10–17.

Jeffrey, R. W., Baxter, J., McGuire, M. and Linde, J. (2006). Are fast food restaurants an environmental risk factor for obesity? *International Journal of Behavioral Nutrition and Physical Activity* **3**.

Kwate, N. O., Yau, C.-Y., Loh, J. M. and Williams, D. (2009). Inequality in obesigenic environments: Fast food density in New York City. *Health and Place* **15**, 364–373.

Kwate, N. O. A. and Loh, J. M. (2010). Separate and unequal: the influence of neighborhood and school characteristics on spatial proximity between fast food and schools. *Preventive Medicine* **51**, 153–156.

Li, K. C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association* **86**, 316–327.

Neckerman, K. M., Bader, M. D. M., Richards, C. A., Purciel, M., Quinn, J. W., Thomas, J. S.,

Warbelow, C., Weiss, C. C., Lovasi, G. S. and Rundle, A. (2010). Disparities in the food environments of New York City public schools. *American Journal of Preventive Medicine* **39**, 195–202.

Newey, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica* **62**, 1349–1382.

Schoenberg, F. P. (2004). Consistent parametric estimation of the intensity of a spatial-temporal point process. *Journal of Statistical Planning and Inference* **128**, 79–93.

Schoenberg, F. P. (2016). A note on the consistent estimation of spatial-temporal point process parameters. *Statistica Sinica (to appear)*.

Severini, T. A. and Wong, W. H. (1992). Profile likelihood and conditionally parametric models. *The Annals of Statistics* **20**, 1768–1802.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

Simon, P. A., Kwan, D., Angelescu, A., Shih, M. and Fielding, J. E. (2008). Proximity of fast food resturants to schools: do neighborhood income and type of school matter? *Preventive Medicine* **47**, 284–288.

Stein, C. (1956). Efficient nonparametric testing and estimation. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, vol. I*, 187–195. University of California Press.

Sturm, R. (2008). Disparities in the food environment surrounding US middle and high schools. *Public Health* **122**, 681–690.

Waagepetersen, R. (2007). An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics* **63**, 252–258.

Waagepetersen, R. and Guan, Y. (2009). Two-step estimation for inhomogeneous spatial point process. *Journal of the Royal Statistical Society, B* **71**, 685–702.

Wang, L., Liu, X., Liang, H. and Carroll, R. (2011). Estimation and variable selection for generalized additive partial linear models. *The Annals of Statistics* **39**, 1827–1851.

Zenk, S. N. and Powell, L. M. (2008). US secondary schools and food outlets. *Health and Place* **14**, 336–246.

Zenk, S. N., Schulz, A. J., Hollis-Neely, T., Campbell, R. T., Holmes, N., Watkins, G., and et al. (2005). Fruit and vegetable intake in African Americans: income and store characteristics. *American Journal of Preventive Medicine* **29**, 1–9.

Zhu, J. and Lahiri, S. N. (2007). Bootstrapping the empirical distribution function of a spatial process. *Statistical Inference for Stochastic Processes* **10**, 107-145.

Dept of Mathematical Sciences, New Jersey Institute of Technology, University Heights, Newark, NJ 07102, USA.

E-mail: yixin.fang@njit.edu

Dept of Mathematical Sciences, New Jersey Institute of Technology, University Heights, Newark, NJ 07102, USA.

E-mail: loh@njit.edu

(Received May 2015; accepted February 2016)