Statistica Sinica **25** (2015), 1503-1526
doi:http://dx.doi.org/10.5705/ss.2013.217

SEMIPARAMETRIC ESTIMATION OF A SELF-EXCITING REGRESSION MODEL WITH AN APPLICATION IN RECURRENT EVENT DATA ANALYSIS

Fangfang Bai, Feng Chen and Kani Chen

University of International Business and Economics,
University of New South Wales and
Hong Kong University of Science and Technology

Abstract: We consider a semi-parametric self-exciting point process regression model where the excitation function is assumed to be smooth and decreasing but otherwise unspecified, and the baseline intensity is assumed to be a linear function of the regressors. We propose an estimation method for this model based on monotone splines. The estimator for the regression coefficients is shown to be consistent, asymptotically normal, and semi-parametrically efficient. The consistency of the estimator for the nonparametric excitation function is also established. The numerical performance of the estimators was found to be satisfactory through simulation studies. We illustrate the application of the model to a bladder cancer data set.

Key words and phrases: B-spline, efficient estimator, Hawkes process, monotone spline, point process, self-exciting process, semiparametric efficiency, sieve estimator.

1. Introduction

Recurrent event data arise frequently in such areas as seismology and medical statistics. As examples, a specific geographic location can experience earthquakes repeatedly over time, and patients with a certain medical condition might experience the same condition repeatedly over a period of time. Recurrent event data can come in different forms. One is as a single, typically rather long, string of event recurrence times, and possibly also other information of each occurrence of the event. An example of here is a sequence of earthquakes in a certain geographical region that records the time of each earthquake together with coordinates and depth of the epicenter, and a magnitude measure. Recurrent event data can consist of multiple, typically short, strings of event recurrence times with some strings of specific covariates. Thus, to evaluate the efficacy of treatments of a certain medical condition, the recurrence times of the condition might be recorded on a sample of patients together with the treatment administered and

other potentially relevant characteristics of the patient. In these forms of recurrent event data, a commonly encountered feature is temporal clustering of the events. Proper modeling of the event clustering phenomenon is important to prediction of future recurrence times and to the assessment of the influence of external explanatory variables on the recurrence rate.

The self-exciting process (Hawkes (1971)), also known as the Hawkes process, has proved to be a useful model for recurrent event data with the event clustering feature. This model is a point process model whose intensity process depends on previous events of the point process itself. The occurrence of an event is assumed to cause the intensity process to jump upwards by a certain amount and then gradually revert toward a baseline level of event intensity. This simple assumption about the evolution of the intensity process makes sense in many contexts and often agrees well with the data. It has been applied in seismology (Ogata (1988)), neuroscience (Chornoboy, Schramm, and Karr (1988)), social science (Crane and Sornette (2008)), marketing research (Kopperschmidt and Stute (2009)), finance (Embrechts, Liniger, and Lin (2011); Errais, Giesecke, and Goldberg (2010)), and criminology (Mohler et al. (2011)).

In the applications of the self-exciting process, the excitation effect associated with an individual event is typically assumed to decay over time and eventually approach zero. The residual excitation effect of an event as a function of time elapsed since the occurrence of the event is referred to as the excitation function. In applications, popular choices of the excitation function include exponential decay (Embrechts, Liniger, and Lin (2011); Errais, Giesecke, and Goldberg (2010); Kopperschmidt and Stute (2009)) and polynomial decay (Crane and Sornette (2008); Mohler et al. (2011); Ogata (1988)). In some applications the choice of the parametric form the excitation function is supported by empirical evidence, in others the choice is based on ad hoc arguments. For data exploration, it is more desirable to let the data speak for itself. A purpose of this paper is to consider the estimation of the self-exciting process model with an excitation function that is assumed to be decreasing and smooth but otherwise unspecified.

There have been works on inference of the parametric self-exciting process model. Ogata (1978) established the consistency and asymptotic normality of the maximum likelihood estimator of the stationary self-exciting process model. Chornoboy, Schramm, and Karr (1988) established the consistency and asymptotic normality of the maximum likelihood estimator of the multivariate extension of the self-exciting process model, or the mutually exciting model. Rathbun (1996) showed the consistency and asymptotic normality of the maximum likelihood estimator of the spatio-temporal self-exciting process model. In these works the asymptotic inference is developed for long time spans, and stationarity or some similar stability condition is typically required of the model. More recently, Chen and Hall (2013) showed the consistency and asymptotic normality

of the maximum likelihood estimator of a non-stationary self-exciting process in the infill asymptotic sense.

There have also been works devoted to semi- and non-parametric self-exciting process models. Zhuang, Ogata, and Vere-Jones (2002) proposed iterative estimation algorithms for a semi-parametric marked spatial-temporal self-exciting process model in which the background intensity is a nonparametric spatial function and the excitation function is parametric, and applied the estimation algorithms to earthquake data. Marsan and Lengliné (2008) proposed an algorithm to declustering earthquakes based on a semi-parametric marked spatio-temporal self-exciting process model in which the background intensity is a constant but the excitation function is a non-parametric function of space, time and event mark. Mohler et al. (2011) modelled crime data using a non-parametric spatial-temporal self-exciting process where the background intensity is the product of a nonparametric function of time and a nonparametric function of space, and the excitation function is a nonparametric function of time and space, and proposed an iterative kernel-type estimation procedure for the nonparametric functions. Although Marsan and Lengliné and Mohler et al. assessed the finite sample performance of their respective estimators via simulation, theoretical properties of these estimators were not investigated. Works on semi-parametric inference for related models that have an implicit self-exciting feature include those of Cox (1972); Lin and Fine (2009); Oakes and Cui (1994) on the modulated renewal process model and its semi-parametric inference, and those of Engle and Russell (1998); Bauwens and Giot (2000); Zhang, Russell, and Tsay (2001); Hautsch (2002); Engle and Lunde (2003), and many other authors on the ACD (autoregressive conditional duration) type models.

These works mostly focus on modelling a single long string of events recorded over a wide time window. In this paper we consider a semi-parametric Hawkes self-exciting process regression model which is suitable for the modelling of recurrent event data in the multiple-string form that often arises in biostatistics and medical studies. With recurrent event data in this form, an important question is to assess the potential effects of the covariate variables on the event recurrent rate. Therefore we consider an extension of the self-exciting process to include a regression component to account for contributions of the covariate variables on the risk of event recurrence. To maintain interpretability of the model, we assume the regression component is a linear function of the covariates. Therefore the model we consider is semi-parametric. We propose estimators for the parametric and non-parametric parts of the model and study the asymptotic behavior of the estimators. The proposed estimators are based on a monotone B-spline approximation (de Boor (2001); Schumaker (2007)) of the excitation function. The estimator of the parametric part of the model is shown to be consistent,

asymptotically normally distributed, and asymptotically optimal in the sense of achieving semi-parametric efficiency (Bickel et al. (1993); van der Vaart (1998)). The estimator of the nonparametric part of the model is shown to be consistent, with non-parametrically optimal rate of convergence in the sense of Stone (1980).

The rest of the paper is organized as follows. In Section 2 we recall the Hawkes self-exciting process and its semi-parametric extension, and present the proposed estimators for the parametric and the nonparametric parts of the model. The asymptotic properties of the estimators are given in Section 3. The numerical performance of the estimators is assessed using simulations in Section 4, and an application is made to a data set from cancer research in Section 5. Section 6 concludes with discussion. All proofs are contained in the supplementary file.

2. The Model, the Data, and the Estimation Method

2.1. A semi-parametric self-exciting regression model for recurrent event data

The self-exciting process proposed by Hawkes (1971) is a simple point process N with an intensity process λ depending on past events of the point process. The intensity at time t is given by

$$\lambda(t) = \mu + \sum_{t_i<t} g(t-t_i) = \mu + \int_{[0,t)} g(t-s)\,\mathrm{d}\,N(s),$$

where $\mu > 0$ is the baseline intensity, $t_1 < t_2 < \cdots$ denote the points, or event times, of the point process, and $g(\cdot) > 0$ is the excitation function. For stationarity, it is assumed that $\int_0^\infty g(t)\,\mathrm{d}\,t < 1$. With this specification, the occurrence of an event makes the intensity process jump instantly by the amount $g(0)$, giving an increased chance of another event occurring in a short time.

In applications, two popular choices of the excitation function are the exponential decay function $g(t) = ae^{-bt},\ t \geq 0$, with parameters $a, b > 0$, and the polynomial decay function $g(t) = K(t+c)^{-p},\ t \geq 0$, with parameters $K, c > 0$ and $p > 1$. With the corresponding constraints on the parameters, these forms of the excitation function are decreasing; it seems reasonable to assume that the residual excitation effect due to an individual event diminishes toward zero over time. More specific assumptions about the excitation function are not always justified. To reduce the risk of model misspecification, it is desirable to leave the form of the excitation function unspecified and estimate it non-parametrically based on the observed data. Here the excitation function $g(\cdot)$ is assumed to be a smooth and bounded decreasing function, otherwise unspecified.

We are interested in the regression problem which assesses the influence of some exogenous variables on the intensity of the self-exciting process. For ease of

interpretation, we assume the influence of the explanatory variables, or suitable transformations of them, on the intensity is linear. Let $X \in \mathbb{R}^p$ be the vector of covariates. The intensity process of the self-exciting process regression model is given by

$$\lambda(t) = X^\top \beta + \int_0^t g(t-s)\,\mathrm{d}\,N(s), \tag{2.1}$$

where $\beta \in \mathbb{R}^p$ is the vector of regression coefficients. We assume that the self-exciting process N is only observable up to a random censoring time C. Given the covariates X, the censoring variable C is assumed to be independent of the point process N. Our interest in the excitation function $g(\cdot)$ is restricted to an interval $[0,\tau]$, so $P(C \le \tau) = 1$ and $P(C > \tau - \epsilon) > 0$ for all $\epsilon > 0$.

The parameter space $\mathcal{B}$ for the regression coefficient is taken to be a bounded convex set in $\mathbb{R}^p$. If K is the upper bound of $g(0)$ and $g(\cdot)$ is r times continuously differentiable for some r, the parameter space for g is the space $\mathcal{F}_r$ of all r times continuously differentiable and decreasing functions on $[0,\tau]$ with values in $[0,K]$. The full parameter space of the model is then $\Theta = \mathcal{B} \times \mathcal{F}_r$.

The data that we use to identify the model consists of n independently obtained right-censored sample paths of N, together with the associated covariates and the censoring time:

$$\{W_i \equiv (X_i^\top, C_i, N_i(t), 0 \le t \le C_i)^\top;\ i = 1,\dots,n\}. \tag{2.2}$$

We also assume the $(X_i^\top, C_i)^\top$ follow a common design distribution, so that the W_i are independent and identically distributed (i.i.d.). Since the sample path of N_i is a jump function with jump sizes equal to 1, it is completely determined by the jump times or the event times $t_{i1} < t_{i2} < \cdots < t_{in_i}$, where $n_i = N_i(C_i)$. Therefore, the data can be equivalently represented as $\{(X_i^\top, C_i, n_i, t_{i1},\dots,t_{in_i})^\top;$ $i = 1,\dots,n\}$.

If $f_{X,C}$ is the joint design density of the covariate vector and the censoring variable, relative to some reference measure ν on $\mathbb{R}^p \times \mathbb{R}_+$, then the density of a generic data point $W = (X^\top, C, N)^\top$ is

$$\begin{aligned} f_\theta(W) = f_{X,C}(X,C) \exp\Bigg[&\int_0^C \log\Big\{X^\top\beta + \int_0^t g(t-s)\,\mathrm{d}\,N(s)\Big\}\,\mathrm{d}\,N(t) \\ &- \int_0^C \Big\{X^\top\beta + \int_0^t g(t-s)\,\mathrm{d}\,N(s) - 1\Big\}\,\mathrm{d}\,t\Bigg], \end{aligned} \tag{2.3}$$

where the reference measure is $\nu \otimes \sigma$, σ being the distribution of the Poisson process on $[0,\tau]$ with unit rate (see e.g. Daley and Vere-Jones (2003, Chapter

7)). The log likelihood for the parameter (β, g) based on the generic data point W is up to an additive constant,

$$\ell(\theta) \equiv \ell(\theta, W) = \int_0^C \log\Big\{X^\top\beta + \int_0^t g(t-s)\,\mathrm{d}\,N(s)\Big\}\,\mathrm{d}\,N(t) - \int_0^C \Big\{X^\top\beta + \int_0^t g(t-s)\,\mathrm{d}\,N(s)\Big\}\,\mathrm{d}\,t. \tag{2.4}$$

Remark 1. For numerical computation, an alternative form of the log likelihood is useful:

$$\ell(\theta) = \sum_{i=1}^{N(C)} \log\Big\{X^\top\beta + \sum_{j=1}^{i-1} g(t_i - t_j)\Big\} - X^\top\beta C - \sum_{i=1}^{N(C)} \int_0^{C-t_i} g(s)\,\mathrm{d}\,s,$$

where t_i, $i = 1, \ldots, N(C)$ denote the jump times of N up to time C.

2.2. A monotone B-spline based sieve estimator

The estimation of semi-parametric models is generally more difficult than the estimation of parametric or nonparametric models. A number of methods have been proposed to deal with the nonparametric component in semi-parametric models: penalized least squares (Engle et al. (1986); Green (1985)); penalized likelihood (Green (1987)), kernel smoothing (Speckman (1988); Zeger and Diggle (1994)), profile likelihood (Nielsen et al. (1992); Huang (1996)), the local polynomial method (Huggins et al. (2007)), piecewise polynomial approximation (Chen and Jin (2006)), the sieve likelihood method (Huang and Rossini (1997); Xue, Lam, and Li (2004)), the nonparametric maximum likelihood method (Zeng and Lin (2006)), pseudo-likelihood method (Wellner and Zhang (2007)), penalized P-splines (Hazelton and Turlach (2011)), and the B-spline approximation-based method (Chen and Tong (2010); Lu, Zhang, and Huang (2007, 2009)). Due to their flexibility for incorporating shape constraints and their remarkable numerical stability (Mammen et al. (2001); Schumaker (2007, §4.9)), B-splines are a natural method of choice for shape-constrained curve estimation. We use B-splines to deal with the monotonicity constraint placed on the excitation function.

Fix a positive integer $d \geq r + 1$. Let $\kappa_n > d$ be an integer, depending on the sample size n, such that $\kappa_n \to \infty$ as $n \to \infty$. Let ξ^n be a sequence of length $\kappa_n + d$ such that $0 = \xi_1 = \cdots = \xi_d < \xi_{d+1} < \cdots < \xi_{\kappa_n+1} = \cdots = \xi_{\kappa_n+d} = \tau$, with

$$\Delta(\xi^n) = \max\{\xi_{i+1} - \xi_i;\ i = 1, \ldots, \kappa_n + d - 1\} \to 0,\ \text{as } n \to \infty.$$

Let $B(t) = (B_1(t), \ldots B_{\kappa_n}(t))^\top$ denote the order d B-spline basis functions associated with the knot sequence ξ. Set

$$\mathcal{F}_r^n = \{B(t)^\top\gamma;\ \gamma \in \mathbb{R}^{\kappa_n}, K \geq \gamma_1 \geq \cdots \geq \gamma_{\kappa_n} \geq 0\}, \Theta_n = \mathcal{B} \times \mathcal{F}_r^n.$$

By the variation diminishing property of B-splines (Schumaker (2007, §4.9)), $B(t)^\top\gamma$ is a positive and decreasing function of t since $\gamma = (\gamma_1, \ldots, \gamma_{\kappa_n})$ is a positive and decreasing sequence, and therefore $\mathcal{F}_r^n \subset \mathcal{F}_r$ and $\Theta_n \subset \Theta$. By the Jackson-type theorem for B-splines (de Boor (2001, Theorem XII.6)), we have $\bigcup_n \Theta_n = \Theta^o$.

Let P_n denote the empirical probability measure corresponding to a sample of size n, and let

$$\ell_n(\theta) = P_n\ell(\theta) = \frac{1}{n}\sum_{i=1}^{n} \ell(\theta, W_i),$$

Then our estimator for the parameter $\theta = (\beta^\top, g)^\top$ is taken as a maximizer of the likelihood function over Θ_n,

$$\widehat{\theta}_n = (\widehat{\beta}_n^\top, \widehat{g}_n)^\top = \underset{\theta\in\Theta_n}{\operatorname{argmax}}\ \ell_n(\theta).$$

By construction it is clear that the estimator is of the sieve type (Grenander (1981)), with $\{\Theta_n; n = 1, 2, \ldots\}$ being a sieve for the parameter space Θ. This estimator is nonparametric in nature because the dimension of the sieve space depends on the sample size and grows to infinity when the sample size tends to infinity. However, from the computational point of view, the estimator is parametric because the sieve space Θ_n is finite-dimensional. Moreover, the optimal dimension of the sieve space is significantly smaller than the sample size. With the optimal choice of dimension, the optimization problem required to evaluate the estimator can often be done using standard optimization routines available from any modern software package. Optimizations in this paper were done in R R Core Team (2013) using the `optim` routine.

Remark 2. With the knot sequence $\xi_1 = \cdots = \xi_d < \cdots < \xi_{\kappa+1} = \cdots = \xi_{\kappa+d}$, the set of order d (≥ 1) basis functions $B_i(t) \equiv B_i^d(t),\ i = 1, \ldots, \kappa$ is defined recursively via the recurrence,

$$B_i^k(t) = \frac{t - \xi_i}{\xi_{i+k-1} - \xi_i} B_i^{k-1}(t) + \frac{\xi_{i+k} - t}{\xi_{i+k} - \xi_{i+1}} B_{i+1}^{k-1}(t),\ k = d, d-1, \ldots, 2,$$

and the initial condition,

$$B_i^1(t) = \begin{cases} I\left\{t \in [\xi_i, \xi_{i+1})\right\}, & i \neq \kappa,\ i \in \{1, \ldots, \kappa + d - 1\}, \\ I\left\{t \in [\xi_i, \xi_{i+1}]\right\}, & i = \kappa, \end{cases}$$

where $I\{\cdot\} = 1$ when $\{\cdot\}$ is true and $= 0$ otherwise.

Remark 3. To compute the log-likelihood function, one needs to compute integrals of the form $\int_0^t g(s)\,\mathrm{d}\,s$ for $t \in [0,\tau]$ (c.f., Remark 1). Numerical integration might be required for this purpose. The B-spline method has the advantage that numerical integration can be avoided, since when g is approximated by the order d B-spline $B^d(t)^\top\gamma$, the integral $\int_0^t g(s)\,\mathrm{d}\,s$ is simply an order $d+1$ B-spline (de Boor (2001, p. 128)),

$$\int_0^t \sum_{i=1}^{\kappa} \gamma_i B_i^d(s)\,\mathrm{d}\,s = \sum_{i=1}^{\kappa} \Big\{ \sum_{j=1}^{i} \frac{\gamma_j(\xi_{j+d}-\xi_j)}{k} \Big\} B_i^{d+1}(t), \quad t \in [\xi_1, \xi_{\kappa+1}] \equiv [0,\tau],$$

where the extra knot value $\xi_{\kappa+d+1}$ needed in the definition of B_κ^{d+1} can be an arbitrary value $\geq \xi_{\kappa+d}$.

Remark 4. To make sure the estimated excitation function $\hat{g}(t) = B(t)^\top\hat{\gamma}$ is positive and decreasing, one only needs to make sure that the estimates $\hat{\gamma} = (\hat{\gamma}_1, \ldots, \hat{\gamma}_{\kappa_n})$ are a positive and decreasing sequence. To guarantee this, one can reparametrize γ in terms of the logarithms of their successive differences, $\gamma_i' = \log(\gamma_i - \gamma_{i+1})$, with $\gamma_{\kappa_n+1} = 0$. This simple treatment effectively eliminates the need for constrained numerical optimization, and has worked very well in our numerical studies.

3. Asymptotic Properties of the Estimator

In this section, under some regularity conditions on the model and the choice of the knot sequence in the definition of the estimator, we show that the estimator of the regression coefficient is consistent, asymptotically normal, and semi-parametrically efficient, and that the estimator of the excitation function is consistent with optimal rate of convergence. We begin with the regularity conditions. In the sequel, $\theta_0 = (\beta_0^\top, g_0)^\top$ denotes the true value of the parameter, and $||\cdot||$ denotes the Euclidean norm.

C1 β_0 is an interior point of $\mathcal{B}$.

C2 For any $\beta \in \mathcal{B}$, there exists an $\epsilon > 0$ such that $X^\top\beta \geq \epsilon$ almost surely.

C3 There exists a constant $M > 0$ such that $||X|| \leq M$ almost surely.

C4 If for any $\beta^1, \beta^2 \in \mathcal{B}$ and $g^1, g^2 \in \mathcal{F}_r$,

$$X^\top\beta^1 + \int_0^t g^1(t-s)\,\mathrm{d}\,N(s) \equiv X^\top\beta^2 + \int_0^t g^2(t-s)\,\mathrm{d}\,N(s)$$

almost surely, then $\beta^1 = \beta^2$ and $g^1 = g^2$.

C5 With $\Delta(\xi) = \max_{d\le i\le\kappa_n} |\xi_{i+1} - \xi_i|$ and $\delta(\xi) = \min_{d\le i\le\kappa_n} |\xi_{i+1} - \xi_i|$, the sequence of knots ξ^n satisfies $\Delta(\xi^n) = O(n^{-q})$ for some $q \in (0, 1/2)$ and $\Delta(\xi^n)/\delta(\xi^n)$ is bounded.

Remark 5. C1 is commonly assumed in semi-parametric estimation problems. C2 guarantees the positivity of the intensity process. C3 is typically satisfied in practical applications. C4 is an identifiability condition. C5 is used to balance the model bias induced by the finite-dimensional approximation to the infinite-dimensional parameter, and is similar to what is assumed by Lu, Zhang, and Huang (2009) and Zhou, Shen, and Wolfe (1998) in studying the asymptotic properties of B-spline based estimators.

Theorem 1 (Consistency)**.** *Under* C1−C5*, the estimator* $\widehat{\theta}_n$ *is consistent,*

$$||\widehat{\beta}_n - \beta_0|| + \int_0^\tau \left|\widehat{g}_n(s) - g_0(s)\right| \mathrm{d}\, s \xrightarrow{P} 0.$$

Theorem 2 (Rate of convergence)**.** *In addition to* C1−C5 *and the condition that* $\kappa_n \to \infty$ *as* $n \to \infty$*, suppose that* $\lim_{n\to\infty} \kappa_n/n = 0$*. Then*

$$\left|\left|\widehat{\beta}_n - \beta_0\right|\right| + \int_0^\tau \left|\widehat{g}_n(s) - g_0(s)\right| \mathrm{d}\, s = O_p\big((\frac{\kappa_n}{n})^{1/2} + \kappa_n^{-r}\big).$$

Remark 6. If we choose $\kappa_n = n^{1/(2r+1)}$ in Theorem 2 up to a positive constant, then it follows that $\int_0^\tau \left|\widehat{g}_n(s) - g_0(s)\right| \mathrm{d}\, s = O_p(n^{-r/(2r+1)})$. The rate of convergence for the nonparametric component, $n^{-r/(2r+1)}$, is the asymptotically optimal rate (Stone (1980)).

Theorem 3 (Asymptotic normality)**.** *Assume the conditions of Theorem* 2*, and that* $\lim_{n\to\infty} \kappa_n^2/n = 0$ *and* $\lim_{n\to\infty} n\kappa_n^{-4r} = 0$*. Then*

$$\sqrt{n}(\widehat{\beta}_n - \beta_0) \xrightarrow{D} N(0, \Sigma_\beta),$$

where Σ_β *is positive definite, and the estimator* $\widehat{\beta}_n$ *is semi-parametrically efficient. Moreover,* Σ_β *is consistently estimated by* $\hat{\Sigma}_\beta = A\mathcal{I}_n^{-1}A^\top$*, where* $\mathcal{I}_n$ *is the observed information matrix, and* $A = A_{p\times(p+\kappa_n)} = (I_p, 0)$ *is the identity matrix of size* p *padded with zeros.*

The proof of these results can be found in the supplementary file.

4. Simulation Studies

Our simulation model was

$$\lambda_i(t) = X_i^\top \beta + \int_0^t g(t-s)\, \mathrm{d}\, N_i(s),\ t \in [0, \tau],\ i = 1, \ldots, n,$$

where $\beta = (\beta_0, \beta_1, \beta_2)^\top = (0.5, 1, 2)^\top$, $g(t) = ae^{-bt} = 4e^{-8t}$, $X_i = (1, X_{i1}, X_{i2})^\top$, $i = 1, \ldots, n$ were i.i.d. with $X_{i1} \sim$ Uniform$[0, 1]$, $X_{i2} \sim$ Bernoulli(0.5), $C_i \equiv \tau = 1$, and $n = 100$, 200, or 400. We simulated data from this model and subsequently estimated the parameter using the monotone B-spline (MBS) method described in Section 2.2. The fully parametric maximum likelihood (ML) method was also used for the purpose of comparison. The order d of the B-spline was set to 3. The interior knots of the B-spline were evenly placed in $[0, 1]$: $(\xi_d, \ldots, \xi_{\kappa_n+1}) = (0, 1/m, \ldots, m/m)$, where $m = \kappa_n - d + 1$ were chosen, in line with C5 and Remark 6, to be integers near $n^{1/3}$, where n is the sample size. To assess the sensitivity of the estimator to the order and the number of knots of the B-spline, we let d vary in the set $\{2, 3, 4\}$ and m vary in the set $\{1, 2, \ldots, 30\}$. With each combination of sample size and estimation method, the process of data simulation and parameter estimation was repeated 1,000 times. For each sample size, the covariates were simulated once and held constant across the 1,000 simulated data sets.

Table 1 shows a summary of the estimates of the regression parameters using the fully parametric method and the MBS method with several different order and knot sequence choices, for sample sizes $n = 100$, $n = 200$, and $n = 400$, respectively. To save space, the results with $d = 2$, which were similar to those with $d = 3$ and 4, are not shown. From the results we note that the parametric method and MBS method gave essentially unbiased estimates of the regression coefficients. The standard error (SE) of the estimators of the regression coefficients decreases with increasing sample size n roughly at the root-n rate as expected. The SE of the MBS method is also very close to that of the parametric method. These observations support the theory on the rate of convergence and semiparametric efficiency of the MBS estimator established earlier. The average of the standard error estimates (SEE) is also close to the true/empirical standard error, with the discrepancy between them narrowing as the sample size increases, both in the parametric method and in the MBS method. This provides empirical evidence of the consistency of the variance estimator for the estimator of the regression coefficients. The coverage probabilities (CPs) of the 95% confidence intervals based on different sample sizes are all above 91%, and approach the nominal rate of 95% as the sample size increases. The CPs based on the parametric and the MBS methods are close to each other, giving further evidence of the semiparametric efficiency of the MBS method. It is also clear that the bias, SE, and SEE of the MBS estimator and the CP of the corresponding confidence interval are nearly identical with different combinations of d and m, which suggests that the MBS estimator of the parametric part is, to some extent, robust to the choice of the order and knot sequence in the MBS approximation to the excitation function. Further evidence of the robustness of the MBS estimator

Table 1. Estimates of the regression coefficients with simulated data (sample size $n = 100$, 200, or 400), using the fully parametric method (PAR) and the monotone B-spline method [MBS(d, m)] with order d and equally spaced interior knots $\{\xi_d, \ldots, \xi_{\kappa_n+1}\} = \{i/m;\ i = 0, \ldots, m\}$. Note: Bias and SE stand for the bias and standard error of the estimator respectively, SEE for the average of the standard error estimates, and CP for the coverage probability of the 95% confidence interval.

		$\beta_0 = 0.5$			$\beta_1 = 1.0$			$\beta_2 = 2.0$		
	n	100	200	400	100	200	400	100	200	400
PAR	Bias	0.002	0.002	0.006	0.022	0.007	-0.004	0.004	0.012	-0.005
	SE	0.265	0.177	0.132	0.563	0.330	0.240	0.354	0.266	0.177
	SEE	0.247	0.170	0.130	0.537	0.327	0.236	0.345	0.248	0.170
	CP	0.914	0.930	0.950	0.937	0.942	0.949	0.938	0.937	0.947
MBS(3,5)	Bias	-0.002	-0.002	0.004	0.014	0.001	-0.008	-0.010	0.001	-0.012
	SE	0.264	0.176	0.132	0.560	0.328	0.239	0.352	0.265	0.176
	SEE	0.247	0.170	0.130	0.536	0.327	0.236	0.345	0.248	0.171
	CP	0.914	0.932	0.951	0.929	0.944	0.950	0.939	0.939	0.947
MBS(3,6)	Bias	-0.002	-0.002	0.004	0.013	0.001	-0.008	-0.011	0.000	-0.012
	SE	0.264	0.176	0.132	0.559	0.328	0.239	0.352	0.265	0.176
	SEE	0.247	0.169	0.130	0.536	0.327	0.236	0.345	0.248	0.171
	CP	0.920	0.930	0.950	0.941	0.942	0.950	0.940	0.939	0.950
MBS(3,7)	Bias	-0.002	-0.002	0.004	0.013	0.001	-0.008	-0.011	0.000	-0.012
	SE	0.264	0.176	0.132	0.560	0.328	0.239	0.352	0.265	0.176
	SEE	0.248	0.169	0.130	0.538	0.327	0.236	0.347	0.248	0.171
	CP	0.915	0.930	0.949	0.934	0.940	0.948	0.944	0.937	0.945
MBS(4,4)	Bias	-0.002	-0.001	0.004	0.014	0.002	-0.008	-0.010	0.001	-0.011
	SE	0.264	0.176	0.132	0.559	0.328	0.239	0.352	0.265	0.176
	SEE	0.247	0.169	0.130	0.536	0.327	0.236	0.345	0.248	0.171
	CP	0.916	0.932	0.951	0.933	0.941	0.949	0.940	0.938	0.947
MBS(4,5)	Bias	-0.002	-0.002	0.004	0.013	0.002	-0.008	-0.010	0.001	-0.012
	SE	0.264	0.176	0.132	0.559	0.329	0.239	0.352	0.265	0.176
	SEE	0.247	0.169	0.130	0.536	0.327	0.236	0.345	0.248	0.171
	CP	0.912	0.930	0.952	0.936	0.943	0.952	0.942	0.942	0.948
MBS(4,6)	Bias	-0.002	-0.002	0.004	0.014	0.001	-0.008	-0.010	0.000	-0.012
	SE	0.264	0.176	0.132	0.560	0.328	0.239	0.352	0.265	0.176
	SEE	0.247	0.169	0.130	0.536	0.326	0.236	0.345	0.248	0.171
	CP	0.922	0.930	0.951	0.934	0.937	0.952	0.945	0.939	0.949

of the regression coefficients to the choice of d and m is revealed by comparing the average mean square errors (MSE) of the MBS estimators of the regression coefficients with different values of d, m. The left panel of Figure 1 is a graph of the average MSE as a function of $m \in \{1, 2, \ldots, 30\}$, for each $d \in \{2, 3, 4\}$, and $n \in \{100, 200, 400\}$, from which we note that when $m \geq 3$ the MSE is virtually constant in m. It also suggests that for $m \geq 2$, when the sample size n increases,

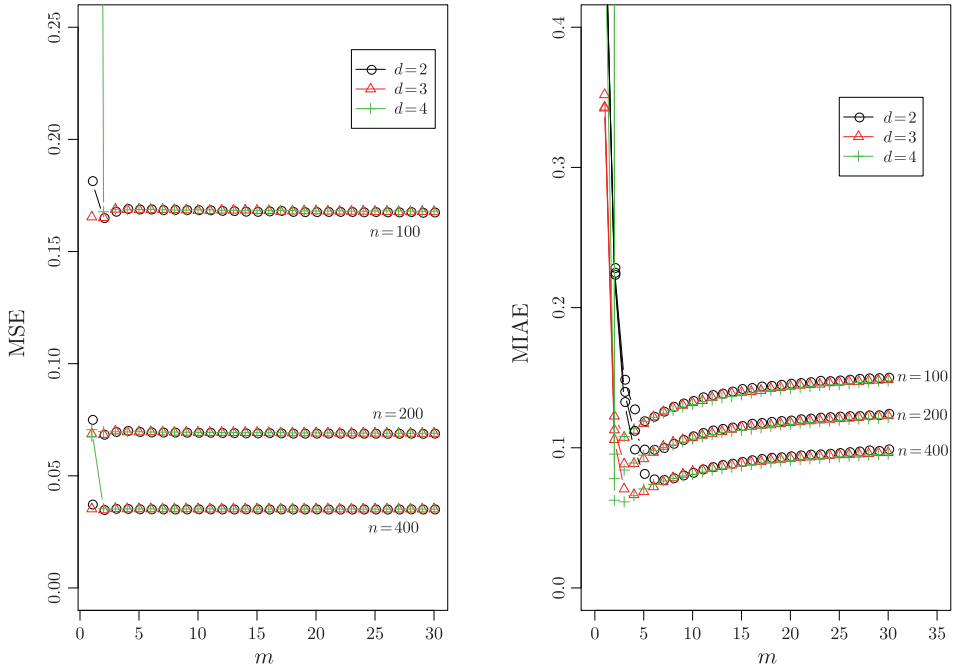

Figure 1. The average MSE of the MBS regression coefficient estimator(left panel), and the MIAE of the excitation function estimator(right panel), with values of the order d of the MBS in $\{2, 3, 4\}$ and the number m of equal sized interior knot intervals of the MBS in $\{1, 2, \ldots, 30\}$, for sample sizes $n = 100$, 200 and 400.

the MSE decreases roughly at rate n, as predicted by the asymptotic theory. The order d of the MBS has little influence on the MSE when $m \geq 2$, although the value $d = 3$ seems to have a slightly better overall performance than $d = 2$ and $d = 4$.

Figure 2 shows the point-wise 95% confidence limits and the median of the estimates of the excitation function, using the parametric and MBS methods with several combinations of d and m. The figure suggests that the parametric method estimated the excitation function unbiasedly, and the MBS estimator of the excitation function is biased in general. However, the bias of the MBS estimator tends to be negligible when the order of the B-spline is 3 or 4. The point-wise variance of the estimator seems to increase slightly as d or m increases. In all cases, the 95% point-wise confidence bands seem to cover the true excitation function entirely, suggesting satisfactory performance of the MBS estimator of the excitation function. To assess the influence of d and m on the performance of the MBS excitation function estimator, we calculated the mean integrated absolute error (MIAE) of the estimator based on values of $d \in \{2, 3, 4\}$ and $m \in \{1, 2, \ldots, 30\}$. The results were shown in the right panel of Figure 1, from which

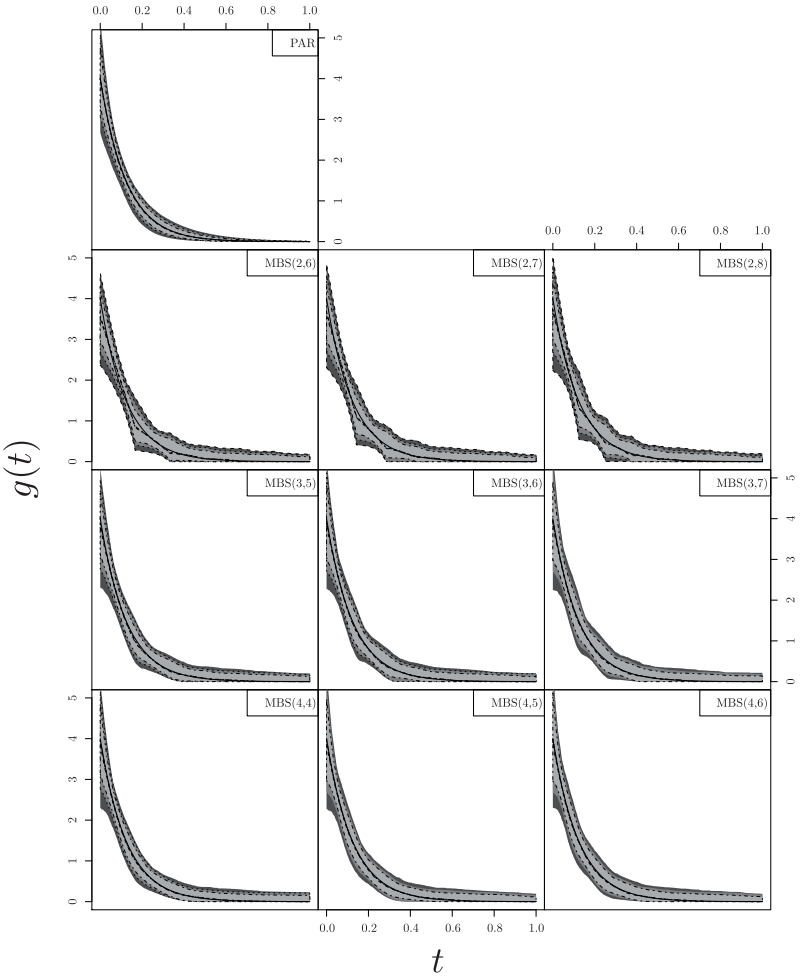

Figure 2. The excitation function (central solid curve) and the point-wise 2.5th, 50th, and 97.5th percentiles of the 1,000 estimates based on simulated data sets of size $n = 100$ (dashed), $n = 200$ (dotted), and $n = 400$ (dot-dashed) respectively, using the fully parametric method (PAR) and the monotone B-spline method [MBS(d,m)] with order d and equally spaced interior knots $\{i/m;\ i = 0, \ldots, m\}$.

we note that, for $m \geq 6$, the influence of d on the MIAE is hardly appreciable; for smaller values of m, $d = 2$ leads to bigger MIAE than $d = 3$ or $d = 4$, or both. In all but the case of $m = 1$, the minimum MIAE is achieved by $d = 4$. With $d = 4$, the m value minimizing the MIAE is 2 when $n = 100$ or 200, and is 3 when $n = 400$.

We have seen that the performance of the MBS estimator depends on the choice of the order d and the number m of interior knot intervals used in the MBS

Table 2. Summary of the regression coefficient estimates using the MBS(d, m) estimator with $m = \hat{m}_{\mathrm{AIC}}$.

$d = 2$	Bias	-0.002	0.000	0.006	0.015	0.001	-0.009	-0.011	0.000	-0.013
	SE	0.264	0.176	0.133	0.560	0.329	0.240	0.352	0.265	0.176
	SEE	0.247	0.169	0.130	0.536	0.327	0.236	0.344	0.248	0.171
	CP	0.912	0.932	0.949	0.936	0.943	0.949	0.937	0.938	0.943
$d = 3$	Bias	-0.001	-0.002	0.005	0.013	-0.002	-0.012	-0.012	-0.006	-0.02
	SE	0.264	0.176	0.132	0.558	0.327	0.238	0.352	0.264	0.175
	SEE	0.247	0.170	0.130	0.535	0.326	0.235	0.344	0.247	0.170
	CP	0.915	0.933	0.951	0.933	0.943	0.949	0.938	0.941	0.942
$d = 4$	Bias	-0.001	-0.001	0.005	0.015	0.002	-0.008	-0.010	0.000	-0.012
	SE	0.264	0.176	0.133	0.560	0.329	0.240	0.351	0.265	0.176
	SEE	0.247	0.169	0.130	0.536	0.327	0.236	0.344	0.248	0.171
	CP	0.918	0.931	0.949	0.932	0.943	0.948	0.938	0.939	0.944

approximation to the excitation function. For applications, it is desirable to have a data-driven approach to selection of these tuning parameters. We propose to select the values of d and m by minimizing the Akaike Information Criterion (Akaike (1974, AIC)),

$$-2\ell_{\max} + 2(m + d - 1),$$

where $\ell_{\max}$ is the maximized log-likelihood value when the excitation function is restricted to the space of monotone B-splines of order d with interior knots $\{0, \tau/m, \ldots, \tau\}$. Since the influence of d on the performance of the MBS estimators tends to be limited compared with the influence of m, it is reasonable to fix the value of d according to the preferred smoothness of the estimated excitation function, and select the value of m by minimizing the AIC. A summary of the optimal m values selected by minimizing the AIC, denoted by $\hat{m}_{\mathrm{AIC}}$, for different values of d and sample size n is shown in Figure 3. From this figure we note that the $\hat{m}_{\mathrm{AIC}}$ values seem to be distributed around m_{MIAE}, the m value minimizing the MIAE for given d and sample size n, shown in the figure as well. This suggests the AIC based selector of the smoothing parameter m has satisfactory numerical performance.

The summary of the MBS regression coefficient estimates with m selected by the AIC is shown in Table 2, from which we note the bias and SE of the estimator, the SEE, and the CP of the 95% confidence intervals are all close to those based on the MBS estimator with fixed m value, and close to those based on the parametric ML estimator (MLE) as well. This suggests that the effect of using the m value selected by the AIC instead of a fixed m value on the inference about the regression coefficients using the MBS estimator is largely negligible. The point-wise 2.5th, 50th and 97.5th percentiles of the estimates of

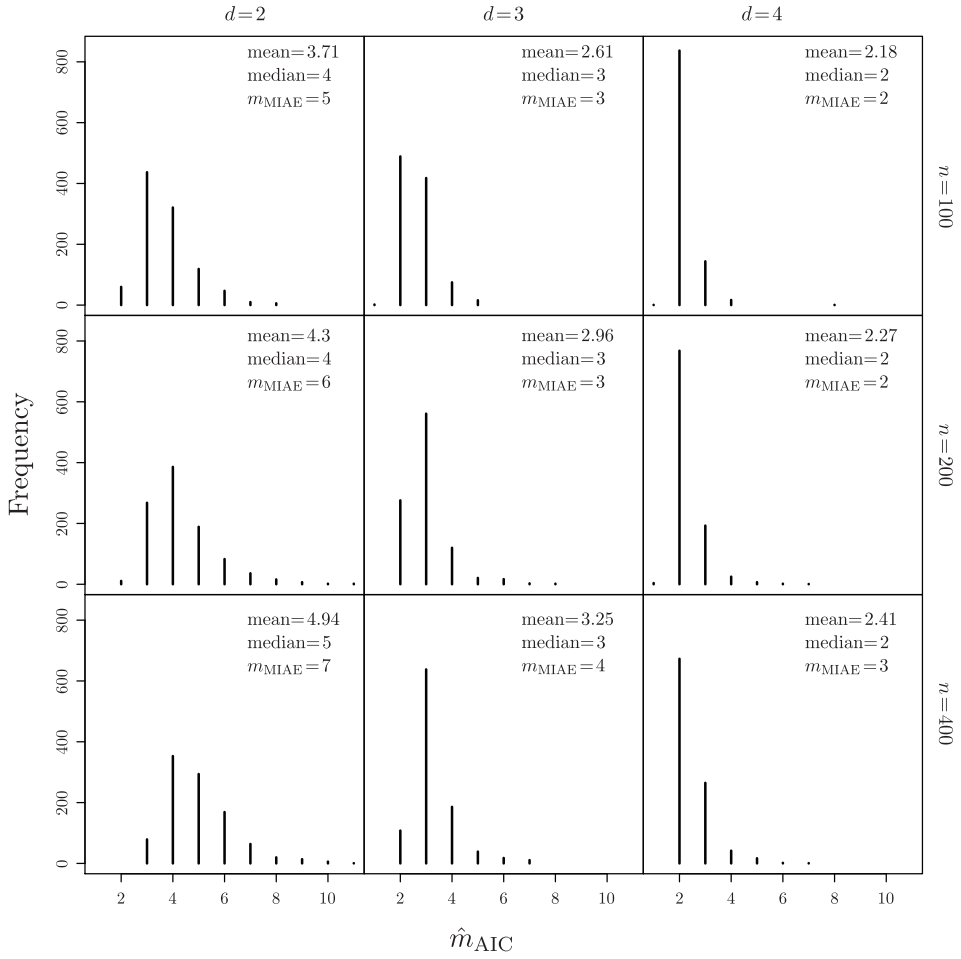

Figure 3. Distribution of the optimal m values selected using the AIC.

the excitation function based on the MBS estimator with varying $d \in \{2, 3, 4\}$ and $m = \hat{m}_{\mathrm{MIAE}}$ are shown in Figure 4, together with those based on the parametric MLE. We note that the empirical 95% confidence bands, though point-wise and not simultaneous, all contain the true curve $g(t)$ entirely. It seems clear that the MBS estimator with $d = 2$ has the worst overall performance, while that with $d = 4$ has the best, which is also confirmed by calculating the MIAE of the MBS estimators with different values of d and n. Therefore, the choice of $d = 4$, which amounts to cubic spline approximation to the excitation function, is recommended as a rule of thumb in applications.

5. A Data Example

We illustrate our model and estimation method with a data set arising in

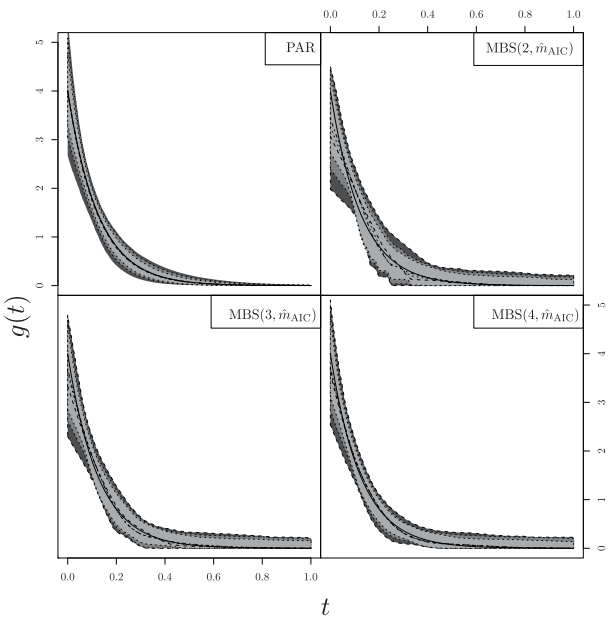

Figure 4. The true excitation function $g(t)$ (central solid curve) in the simulation model and the 2.5th, 50th, and 97.5th percentiles of the MBS(d, m) estimates of the excitation function with different order d and the number m of equal-sized interior knot intervals selected by the AIC, and with simulated data sets of sizes $n = 100$ (dashed), $n = 200$ (dotted) and $n = 400$ (dot-dashed). The results of using the parametric maximum likelihood estimator is also included for ease of comparison.

bladder cancer study. The data was reported by Byar (1980) and has been frequently used to illustrate statistical methods for recurrent event data analysis, e.g. Wei, Lin, and Weissfeld (1989), Zeng and Lin (2006), Wellner and Zhang (2007), and Lu, Zhang, and Huang (2009). The data consists of the bladder tumor recurrence times of 118 Stage-I bladder cancer patients. On entry of the study, all bladder cancer tumors were surgically removed through transurethral resection, and the patients were randomly assigned to one of three groups – placebo, pyridoxine, and thiotepa. Any new tumors discovered at subsequent recurrence times were surgically removed. The recurrence times were recorded as months since entry of the study. For each patient, the number of initial tumors and the size of the largest tumor were also available. The patients were right censored at the earlier of the time of death due to bladder cancer or other causes and the end of study period. The maximum follow-up time was 64 months. The observed number of tumor recurrences ranges from 0 to 9.

We fitted three self-exciting point process regression models to the data – the first two fully parametric with the excitation function forced to be a constant

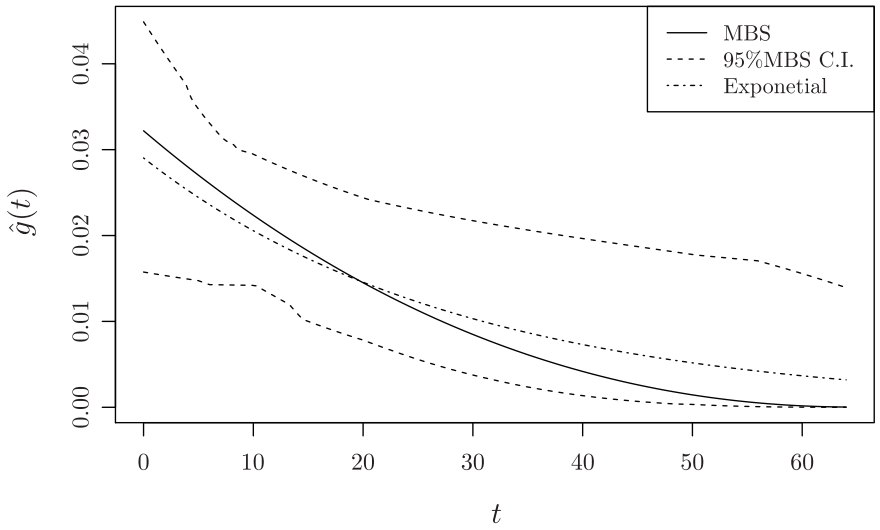

Figure 5. Estimates of the excitation function g for the bladder cancer data by the parametric method with an exponential form for g and by the MBS(d, m) method with $d = 4$ and m selected by the AIC. The 95% (pointwise) confidence bands based on the MBS estimator were obtained by parametric bootstrap.

and an exponential function, respectively, and the other semi-parametric with the excitation function assumed to be positive and monotone decreasing. With the constant excitation function, the model becomes a parametric additive risk model (Aalen (1980, 1989); Lin and Ying (1994)) with cumulative number of previous events as a time-varying covariate. When estimating the semi-parametric model with the proposed MBS estimator, we used $d = 4$ and equally spaced interior knots $i/m \times 64$, $i = 0, \ldots, m$, with $m = 1$ selected by the AIC. For comparison, we also fitted a point process regression model without the excitation effect, which is equivalent to Poisson regression. The estimated regression coefficients and the excitations function are shown in Table 3 and Figure 5, respectively. The standard errors estimates in Table 3 were obtained by inverting the observed information matrix. They were nearly identical to the standard errors obtained using a bootstrap method (not reported) to be discussed below.

From Table 3, by all four models the suppressing effect of the thiotepa on the tumor recurrence intensity is statistically significant at level 0.05, and the number of tumors present at entry of the study is a highly significant risk factor for tumor recurrence. These results are consistent with the analysis of others, e.g. Zeng and Lin (2006), Wellner and Zhang (2007), and Lu, Zhang, and Huang (2009). By the self-exciting process models the effect is less conclusive, with p-values equal to 0.021, 0.047, and 0.025 respectively. This suggests that the self-exciting

Table 3. Estimated regression coefficients under different models, where Poisson Reg. is short for the Poisson regression model, and SEP(con), SEP(exp), and SEP(nonpar) for the self-exciting process regression models with the constant, exponential, and nonparametric excitation functions respectively.

	Parameter	Estimate	Std. Error	z value	Pr(> |z|)	
Poisson Reg.	(Intercept)	0.035	0.0093	3.75	1.8e-4	***
	pyridoxine	7.9e-4	0.0094	0.084	0.93	
	thiotepa	-0.027	0.0077	-3.45	5.6e-4	***
	number	0.014	0.0029	4.84	1.3e-6	***
	size	-0.0018	0.0021	-0.86	0.34	
SEI(con)	(Intercept)	0.021	0.0084	2.45	0.014	*
	pyridoxine	-0.0065	0.0084	-0.77	0.44	
	thiotepa	-0.017	0.0072	-2.31	0.021	*
	number	0.010	0.0027	3.62	3.0e-4	***
	size	-9.6e-5	0.0018	-0.05	0.96	
SEI(exp)	(Intercept)	0.019	0.0081	2.4	0.016	*
	pyridoxine	-0.002	0.0085	-0.24	0.81	
	thiotepa	-0.014	0.0069	-1.98	0.047	*
	number	0.0088	0.0026	3.4	8.0e-4	***
	size	-0.3e-4	0.0017	-0.18	0.86	
SEI(nonpar)	(Intercept)	0.0190	0.0083	2.29	0.022	*
	pyridoxine	-0.0070	0.0082	-0.85	0.39	
	thiotepa	-0.0157	0.0070	-2.24	0.025	*
	number	0.0096	0.0027	3.56	4.0e-4	***
	size	0.0001	0.0018	0.055	0.96	

models are less likely to produce falsely significant results. The minimum minus log likelihood value of the Poisson regression model is 732.43, and those of the self-exciting model with constant and exponential excitation functions are 713.08 and 710.33 respectively. The change of the log-likelihood value from the model without self-excitation term to the models with self-excitation term is highly significant by the χ^2-test, which supports the existence of the self-excitation effect among bladder tumor occurrences. The change of the log-likelihood from the self-excitation model with constant excitation function to that with exponential excitation function at the cost of one extra parameter is also significant with a P-value of 0.019 by the χ^2-test, suggesting that the excitation effect decays over time.

Figure 5 shows the ML and the MBS estimates of the excitation function g in the self-exciting intensity models with g assumed to be an exponential decay function and an unspecified decay function, respectively. Point-wise confidence bands based on the MBS estimator were obtained using a parametric bootstrap method. In the bootstrap, 200 bootstrap samples were generated from the self-

exciting process regression model with the regression coefficients and the excitation function fixed at the MBS estimates obtained earlier, and the covariates and censoring times the same as those in the original data set. The MBS estimator with m selected by the AIC and $d = 4$ was then applied on each bootstrap sample to obtain the bootstrap versions of the MBS estimator of g. The point-wise 2.5th and 97.5th percentiles were taken as the lower and upper limits of the point-wise confidence bands for g. We also obtained the bootstrap standard errors for the estimators of the regression coefficients, which were very close to the standard errors reported in Table 3; they were omitted to save space. Figure 5 reveals that the MBS estimator-based confidence band contains the exponential form parametric estimate of g, suggesting the decay of the excitation effect could be modelled by an exponential function.

To assess the sensitivity of the analysis to the choice of the number of knots in the monotone B-spline, we inspected the estimates of the regression coefficients with the number m of interior knot intervals varying in the range 1 to 20; they were nearly identical to the results with $m = 1$. In particular, the treatment effect of thiotepa remained significant for all these m values. This agrees with the numerical evidence of the robustness of the method observed in our simulation studies.

To assess the goodness of fit of the considered models to the data, we also calculated the point process residuals $\hat{r}_{ij} = \hat{\Lambda}_i(t_{ij}) - \hat{\Lambda}_i(t_{i,j-1})$, $j = 1, \ldots, n_i + 1$, $i = 1, \ldots, n$, where $\hat{\Lambda}_i(t) = \int_0^t \hat{\lambda}_i(s)\,\mathrm{d}\,s = \int_0^t \{X_i^\top \hat{\beta} + \int_0^s \hat{g}(s-u)\,\mathrm{d}\,N_i(u)\}\,\mathrm{d}\,s$, $t_{i,0} \equiv 0$, $t_{i,n_i+1} \equiv C_i$, and $n_i \equiv N_i(C_i)$. If a self-exciting intensity model fits the data well, then the corresponding residuals should be close to a sample of independently right-censored unit rate exponential variables, with the $\hat{r}_{i,n_i+1}$ corresponding to the censored observations. One can then graphically check the goodness of fit by comparing the Nelson-Aalen cumulative hazard estimator of the residual with that of the unit rate exponential variable, $H(t) = t$ (Andersen et al. (1993, p. 182)). The Nelson-Aalen residual plots of the four models considered are shown in Figure 6. It shows that the Nelson-Aalen plot of the semiparametric model is the closest to the diagonal line, while those of the Poisson regression model and of the parametric self-exciting process model with a constant excitation function respective deviate from the diagonal rather obviously. Meanwhile, that of the parametric self-exciting process model with an exponential excitation function is close to the diagonal line and fairly similar to that of the semiparametric model. This is further evidence that a parametric model with parametric excitation function produces an acceptable fit to the bladder cancer data, suggesting that the excitation effect of the occurrence (and the ensuing removal) of a bladder tumor on future tumor occurrence decays roughly exponentially.

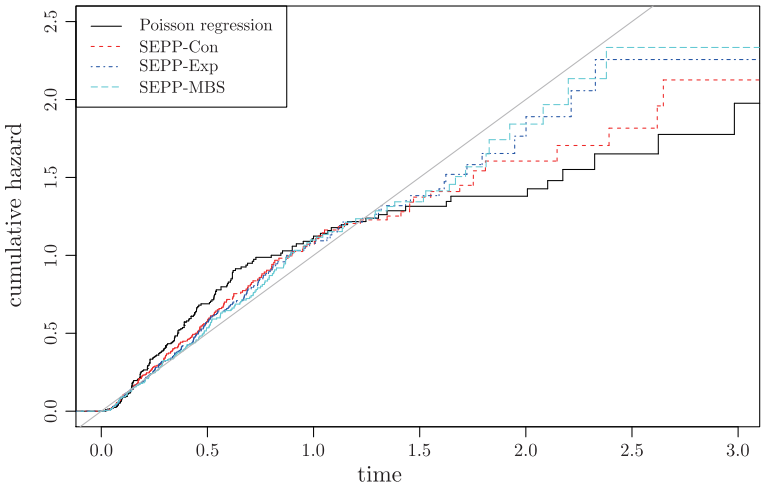

Figure 6. Nelson-Aalen plots of the residuals of the four models — Poisson regression model, the parametric self-exciting intensity point process models with constant excitation function (SEPP-Con), and with exponential excitation function (SEPP-Exp), and the semiparmetric self-exciting process model with a montone excitation function estimated using the MBS method (SEPP-MBS).

6. Discussion

In this paper we considered a semi-parametric extension of the Hawkes self-exciting point process with the excitation function only assumed to be decreasing. An estimator for the model based on monotone B-splines was proposed, and its large sample properties and asymptotic optimality were established. The numerical performance of the estimator was shown to be satisfactory on some simulated data and on a data set arising from bladder cancer studies.

Because of the explicit modeling of the serial correlation among the recurrence times of the events, the proposed model can alleviate the risk of false positives associated with Poisson regression. The B-spline-based estimator for the nonparametric excitation function can help the data analyst to find a suitable parametric form for the excitation function.

There are interesting questions about the model that remain. For instance, a formal nonparametric statistical test for the existence of the self-excitation effect is clearly a question of practical and theoretical interest. A test of specific parametric forms for the excitation function against nonparametric alternatives would also be desirable.

Our choice of a linear function of covariates for the baseline intensity model is motivated more by parsimony and interpretability than by flexibility of the model. If more delicate features of the data, such as time-varying covariate effects, are suspected, then our model can be extended to allow for time-varying

regression coefficients, as considered by Zucker and Karr (1990) and Murphy and Sen (1991), among many others, in the context of the Cox proportional hazards model. Another approach to model the time-varying effect is to use the transformation model, where a transformation such as the Box-Cox transformation of the intensity process, rather than the intensity itself, is modelled by the right-hand side of (2.1). This approach has been adopted by Zeng and Lin (2006) and Zeng and Lin (2007) in the context of the Cox proportional intensities model. Inferences for the extensions of our model along these directions are also interesting questions.

Acknowledgement

The comments from the Co-Editor, an associate editor, and two reviewers have led to improved presentation, for which we are grateful. This work was partially supported by the University of New South Wales ECR Grants PS27252 and PS30701 (F. Chen) and the Hong Kong Research Grant Council grants 601011 and 601612 (K. Chen).

References

Aalen, O. O. (1980). A model for nonparametric regression analysis of counting processes. In *Mathematical Statistics and Probability Theory* (Edited by Klonecki, W., Kozek, A. and Rosinski, J.), 1–25. Springer, New York.

Aalen, O. O. (1989). A linear regression model for the analysis of life times. *Statist. Medicine* **8**, 907-925.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Control* **19**, 716-723.

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.

Bauwens, L. and Giot, P. (2000). The logarithmic ACD model: An application to the bid-ask quote process of three NYSE stocks. *Annales déconomie et de Statistique* **60**, 117-149.

Bickel, P., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. The John Hopkins University Press, New York.

Byar, D. (1980). The veterans administration study of chemoprophylaxis for recurrent stage I baldder tumors: Comparisons of placebo, pyridoxine, and topical thiotepa. In *Bladder Tumors and Other Topics in Urological Oncology* (Edited by Pavone-Macaluso, M., Smith, P. H., and Edsmyr, F.), 363-370. Plenum Press.

Chen, F. and Hall, P. (2013). Inference for a non-stationary self-exciting point process with an application in ultra-high frequency financial data modeling. *J. Appl. Probab.* **50**, 1006-1024.

Chen, K. and Jin, Z. (2006). Partial linear regression models for clustered data. *J. Amer. Statist. Assoc.* **101**, 195-204.

Chen, K. and Tong, X. (2010). Varying coefficient transformation models with censored data. *Biometrika*, **97**, 969-976.

Chornoboy, E., Schramm, L. and Karr, A. (1988). Maximum likelihood identification of neural point process systems. *Biol. Cybernetics*, **59**, 265-275.

Cox, D. R. (1972). The statistical analysis of dependencies in point processes. In *Stochastic Point Processes* (Edited by Lewis, P. A. W.), 55-66. John Wiley, New York.

Crane, R. and Sornette, D. (2008). Robust dynamic classes revealed by measuring the response function of a social system. *Proc. Nat. Acad. Sci.* **105**, 15649-15653.

Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes Volume I: Elementary Theory and Methods*. Springer-Verlag, New York.

de Boor, C. (2001). *A Practical Guide to Splines*. Springer, New York.

Embrechts, P., Liniger, T. and Lin, L. (2011). Multivariate Hawkes processes: an application to financial data. *J. Appl. Probab.* **48**, 367-378.

Engle, R. F., Granger, C. W. J., Rice, J. and Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *J. Amer. Statist. Assoc.* **81**, 310-320.

Engle, R. F. and Lunde, A. (2003). Trades and quotes: A bivariate point process. *J. Financial Econom.* **1**, 159-188.

Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica* **66**, 1127-1162.

Errais, E., Giesecke, K. and Goldberg, L. R. (2010). Affine point processes and portfolio credit risk. *SIAM J. Financial Math.* **1**, 642-665.

Green, P. J. (1985). Linear models for field trials, smoothing and cross-validation. *Biometrika* **72**, 527-537.

Green, P. J. (1987). Penalized likelihood for general semi-parametric regression models. *Internat. Statist. Rev.* **55**, 245-259.

Grenander, U. (1981). *Abstract Inference*. John Wiley, New York.

Hautsch, N. (2002). Modelling intraday trading activity using Box-Cox-ACD models. CoFE Discussion Paper 02-05, Center of Finance and Econometrics, University of Konstanz.

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58**, 83-90.

Hazelton, M. L. and Turlach, B. A. (2011). Semiparametric regression with shape-constrained penalized splines. *Comput. Statist. Data Anal.* **55**, 2871- 2879.

Huang, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Ann. Statist.* **24**, 540-568.

Huang, J. and Rossini, A. J. (1997). Sieve estimation for the proportional-odds failure-time regression model with interval censoring. *J. Amer. Statist. Assoc.* **92**, 960-967.

Huggins, R. M., Hall, P., Yip, P. S. F. and Bui, Q. M. (2007). Applications of additive semivarying coefficient models: Monthly suicide data from Hong Kong. *Biometrics* **63**, 708-713.

Kopperschmidt, K. and Stute, W. (2009). Purchase timing models in marketing: a review. *AStA Adv. Statist. Anal.* **93**, 123-149.

Lin, D. Y. and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61-71.

Lin, F. and Fine, J. P. (2009). Pseudomartingale estimating equations for modulated renewal process models. *J. Roy. Statist. Soc. Ser. B* **71**, 3-23.

Lu, M., Zhang, Y. and Huang, J. (2007). Estimation of the mean function with panel count data using monotone polynomial splines. *Biometrika* **94**, 705-718.

Lu, M., Zhang, Y. and Huang, J. (2009). Semiparametric estimation methods for panel count data using monotone B-splines. *J. Amer. Statist. Assoc.* **104**, 1060-1070.

Mammen, E., Marron, J. S., Turlach, B. A. and Wand, M. P. (2001). A general projection framework for constrained smoothing. *Statist. Sci.* **16**, 232-248.

Marsan, D. and Lengliné, O. (2008). Extending earthquakes' reach through cascading. *Science* **319**, 1076-1079.

Mohler, G. O., Short, M. B., Brantingham, P. J., Tita, G. E. and Schoenberg, F. P. (2011). Self-exciting point process modeling of crime. *J. Amer. Statist. Assoc.* **106**, 100-108.

Murphy, S. and Sen, P. (1991). Time-dependent coefficients in a Cox-type regression model. *Stochastic Process. Appl.* **39**, 153-180.

Nielsen, G. G., Gill, R. D., Andersen, P. K. and Srensen, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Statist.* **19**, 25-43.

Oakes, D. and Cui, L. (1994). On semiparametric inference for modulated renewal processes. *Biometrika* **81**, 83-90.

Ogata, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Ann. Inst. Statist. Math.* **30**, 243-261.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Assoc.* **83**, 9-27.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rathbun, S. L. (1996). Asymptotic properties of the maximum likelihood estimator for spatio-temporal point processes. *J. Statist. Plann. Inference* **51**, 55-74.

Schumaker, L. L. (2007). *Spline Functions: Basic Theory*. Cambridge University Press, Cambridge.

Speckman, P. (1988). Kernel smoothing in partial linear models. *J. Roy. Statist. Soc. Ser. B* **50**, 413-436.

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8**, 1348-1360.

van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge University Press, New York.

Wei, L. J., Lin, D. Y. and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Amer. Statist. Assoc.* **84**, 1065-1073.

Wellner, J. A. and Zhang, Y. (2007). Two likelihood-based semiparametric estimation methods for panel count data with covariates. *Ann. Statist.* **35**, 2106-2142.

Xue, H., Lam, K. F. and Li, G. (2004). Sieve maximum likelihood estimator for semiparametric regression models with current status data. *J. Amer. Statist. Assoc.* **99**, 346-356.

Zeger, S. L. and Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to cd4 cell numbers in hiv seroconverters. *Biometrics* **50**, 689-699.

Zeng, D. and Lin, D. Y. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika* **93**, 627-640.

Zeng, D. and Lin, D. Y. (2007). Semiparametric transformation models with random effects for recurrent events. *J. Amer. Statist. Assoc.* **102**, 167-180.

Zhang, M. Y., Russell, J. R. and Tsay, R. S. (2001). A nonlinear autoregressive conditional duration model with applications to financial transaction data. *J. Econom.* **104**, 179-207.

Zhou, S., Shen, X. and Wolfe, D. A. (1998). Local asymptotics for regression splines and confidence regions. *Ann. Statist.* **26**, 1760-1782.

Zhuang, J., Ogata, Y. and Vere-Jones, D. (2002). Stochastic declustering of space-time earthquake occurrences. *J. Amer. Statist. Assoc.* **97**, 369-380.

Zucker, D. M. and Karr, A. F. (1990). Nonparametric survival analysis with time-dependent covariate effects: A penalized partial likelihood approach. *Ann. Statist.* **18**, 329-353.

School of Statistics, University of International Business and Economics, 10 Huixin Dongjie, Chaoyang District, Beijing, China.

E-mail: bss03021@gmail.com

School of Mathematics and Statistics, University of New South Wales, Sydney, NSW 2052, Australia.

E-mail: feng.chen@unsw.edu.au

Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong.

E-mail: makchen@ust.hk

(Received July 2013; accepted September 2014)