<table>
<thead>
<tr>
<th><strong>Statistica Sinica Preprint No: SS-13-217R2</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong></td>
</tr>
<tr>
<td><strong>Manuscript ID</strong></td>
</tr>
<tr>
<td><strong>URL</strong></td>
</tr>
<tr>
<td><strong>DOI</strong></td>
</tr>
<tr>
<td><strong>Complete List of Authors</strong></td>
</tr>
<tr>
<td><strong>Corresponding Author</strong></td>
</tr>
<tr>
<td><strong>E-mail</strong></td>
</tr>
</tbody>
</table>

Notice: Accepted version subject to English editing.
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 only assumed to be smooth and decreasing but otherwise unspecified, while the baseline intensity is assumed to be a linear function of the regressors. We propose an estimation method for this model based 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 were found to be satisfactory through simulation studies. We also illustrate the application of the model and the estimators in recurrent event data analysis with 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 arises frequently in many areas of scientific endeavor such as seismology and medical statistics. For example, a specific geographic location can experience earthquakes repeatedly over time. Patients with a certain medical condition might experience the same condition repeatedly over a period of time. Depending on the purpose of the study and the data collection method, the recurrent event data can come in different forms. One is a single, typically rather long, string of event recurrence times, and possibly also other information of each occurrence of the event. An example of this form of recurrent event data is a sequence of earthquakes in a certain geographical region which records the time of each earthquake together with other information such as coordinates and depth of the epicenter and a magnitude measure. Another form of recurrent event data consists of multiple, typically short, strings of event recurrence
times and some string specific covariates. For example, to evaluate the efficacy of treat-
ments of a certain recurrent 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 each patient. In both forms of recurrent event
data, a commonly encountered feature of recurrent event data is the temporal clustering
of the events. Properly modeling of the event clustering phenomenon is important to
predict the future recurrence times of the event and to assess the influence of external
explanatory variables on the event 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 base-
line level of event intensity. This simple assumption about the evolution of the intensity
process makes sense in many contexts and often agrees with the data well. Therefore,
the self-exciting process has been applied in a wide range of areas such as seismology
(Ogata, 1988), neuroscience (Chornoboy et al., 1988), social science (Crane and Sor-
nette, 2008), marketing research (Kopperschmidt and Stute, 2009), finance (Embrechts
et al., 2011; Errais et al., 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 approaches
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 the exponential decay function (Embrechts
et al., 2011; Errais et al., 2010; Kopperschmidt and Stute, 2009) and the polynomial
decay function (Crane and Sornette, 2008; Mohler et al., 2011; Ogata, 1988). While in
some applications the choice of the parametric form the excitation function is supported
by empirical evidence, in many other applications the choice of the parametric forms are
based on essentially ad hoc arguments, if there is any attempt to justify the choice at all.
From the point of view of data exploration, it is more desirable to leave the excitation
function unspecified and estimate it non-parametrically, or let the data speak for itself,
so to speak. Therefore, a purpose of this paper is to consider the estimation of the self-
exciting process model with an excitation function that is only assumed to be decreasing
and smooth but otherwise unspecified.

In the literature, 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 et al. (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 all these works the asymptotic inference is developed in the long time span scenario, 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 an infill asymptotic scenario.

There have also been works devoted to semi- and non-parametric self-exciting process models. For example, Zhuang et al. (2002) proposed iterative estimation algorithms for a semi-parametric marked spatial-temporal self-exciting process model where 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 where 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, the theoretic 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 et al. (2001); Hautsch (2002); Engle and Lunde (2003) and many other authors on the ACD (autoregressive
conditional duration) type models.

The aforementioned 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 arise often 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 will consider an extension of the self-exciting process to include a regression component which accounts for any potential contribution of the covariate variables on the risk of event recurrence. To maintain interoperability 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) to the excitation function. The estimator of the parametric part of the model will be 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 both 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 of the model and the estimators is illustrated with a data set from cancer research in Section 5. Finally, Section 6 concludes with discussion. All technical 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 its intensity process $\lambda$ depending on past events of the point process. Specifically,
the intensity at time $t$ is given by

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

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 also assumed that $\int_0^\infty g(t) \, dt < 1$. By the above specification, we note in particular that the occurrence of an event will make the intensity process jump instantly by the amount $g(0)$, which implies an increased chance of another event occurring in a short time interval following the event. This makes the self-exciting process an amenable model for recurrent event data with temporal clustering of events.

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

$$g(t) = \frac{K}{(t + c)^p}, \; t \geq 0,$$

with parameters $K, c > 0$ and $p > 1$. With the corresponding constraints on the parameters, these two forms of the excitation function are both decreasing. From a practical point of view, it seems reasonable to assume that the residual excitation effect due to an individual event wears out and diminishes toward zero as time elapses. However, more specific assumptions, such as the exponential and polynomial forms, for 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. In this paper, the excitation function $g(\cdot)$ shall only be assumed to be a bounded smooth decreasing function but otherwise unspecified, which means we are faced with a nonparametric problem.

As mentioned in the introduction, we are also 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 then given by

$$
\lambda(t) = X^\top \beta + \int_0^t g(t - s) \, dN(s),
$$

(2.1)

where $\beta \in \mathbb{R}^p$ is the vector of regression coefficients. Motivated by typical applications of recurrent data analysis in areas such as medical statistics, 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 $C$ will be assumed to be independent of the point process $N$. Suppose our interest about the excitation function $g(\cdot)$ is restricted to an interval $[0, \tau]$. Then we shall also assume that $P(C \leq \tau) = 1$ and $P(C > \tau - \epsilon) > 0$ for all $\epsilon > 0$.

The parameter space $B$ for the regression coefficient is assumed to be a bounded convex set in $\mathbb{R}^p$. Let $K$ be the upper bound of $g(0)$ and assume $g(\cdot)$ to be $r$ times continuously differentiable for some $r$. Then the parameter space for $g$ is the space $F_r$ of all $r$ times continuously differentiable and decreasing functions on $[0, \tau]$ with values in $[0, K]$, and the full parameter space of the model is given by

$$
\Theta = B \times F_r.
$$

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

$$
\{W_i \equiv (X_i^\top, C_i, N_i(t), 0 \leq t \leq C_i)^\top; \ i = 1, \ldots, n\}.
$$

(2.2)

As a part of the model specification, 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 $t_{i1} < t_{i2} < \ldots < t_{in_i}$, where $n_i = N_i(C_i)$. Therefore, the data can be equivalently represented in the follow form,

$$
\{(X_i^\top, C_i, n_i, t_{i1}, \ldots, t_{in_i})^\top; \ i = 1, \ldots, n\}.
$$

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

\[
f_\theta(W) = f_{X,C}(X, C) \exp \left[ \int_0^C \log \left\{ X^\top \beta + \int_0^t g(t - s) \, d \, N(s) \right\} \, d \, N(t) \right. \\
\left. - \int_0^C \left\{ X^\top \beta + \int_0^t g(t - s) \, d \, N(s) - 1 \right\} \, d \, t \right], \tag{2.3}
\]

where the reference measure is \( \nu \otimes \sigma \) with \( \sigma \) 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 \((\beta, g)\) based on the generic data point \( W \) is therefore given, up to an additive constant, by

\[
\ell(\theta) = \ell(\theta, W) = \int_0^C \log \left\{ X^\top \beta + \int_0^t g(t - s) \, d \, N(s) \right\} \, d \, N(t) \\
- \int_0^C \left\{ X^\top \beta + \int_0^t g(t - s) \, d \, N(s) - 1 \right\} \, d \, t. \tag{2.4}
\]

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

\[
\ell(\theta) = \sum_{i=1}^{N(C)} \log \left\{ X^\top \beta + \sum_{j=1}^{i-1} g(t_i - t_j) \right\} - X^\top \beta C - \sum_{i=1}^{N(C)} \int_{C-t_i}^{C} g(s) \, d \, s,
\]

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, because not only the estimation of the nonparametric component is often challenging, but the presence of the nonparametric component can also make the estimation of the parametric component more difficult. In the literature a number of methods have been proposed to deal with the nonparametric component in specific semi-parametric models, such as 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 et al., 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.
Due to their flexibility of incorporating shape constraints and their remarkable numerical stability (Mammen et al., 2001; Schu- maker, 2007, §4.9), B-splines are the natural method of choice for shape constrained curve estimation. Therefore in this work we shall 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 \( \xi^n \) be a sequence of length \( \kappa_n + d \) such that
\[
0 = \xi_1 = \ldots = \xi_d < \xi_{d+1} < \ldots < \xi_{\kappa_n+1} = \ldots = \xi_{\kappa_n+d} = \tau
\]
and that
\[
\Delta(\xi^n) = \max\{\xi_{i+1} - \xi_i; \ i = 1, \ldots, \kappa_n + d - 1\} \to 0, \ \text{as} \ n \to \infty.
\]
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 \( \xi \). Define
\[
\mathcal{F}_r^n = \{B(t)^\top \gamma; \ \gamma \in \mathbb{R}^{\kappa_n}, K \geq \gamma_1 \geq \ldots \geq \gamma_{\kappa_n} \geq 0\}, \ \text{and} \ \Theta_n = \mathcal{B} \times \mathcal{F}_r^n.
\]
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 defined as a maximizer of the likelihood function over \( \Theta_n \). That is,
\[
\hat{\theta}_n = (\hat{\beta}_n^\top, \hat{g}_n)^\top = \arg\max_{\theta \in \Theta_n} \ell_n(\theta).
\]
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 \( \Theta \). From a theoretical point of view, the 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 virtually parametric because the sieve space \( \Theta_n \) is finite-dimensional. Moreover, the optimal dimension of the sieve space is significantly smaller than the sample size. With the optimal
choice of the dimension, the optimization problem required to evaluate the estimator can often be done using standard optimization routines available from any modern mathematical and statistical software package. The numerical optimizations in this paper were done in R (R Core Team, 2013) using the \texttt{optim} routine.

**Remark 2.** Specifically with the knot sequence $\xi_1 = \ldots = \xi_d < \ldots < \xi_{\kappa+1} = \ldots = \xi_{\kappa+d}$, the set of order $d \geq 1$ basis functions $B_i(t) \equiv B^d_i(t), \ i = 1, \ldots, \kappa$ are defined recursively via the recurrence,

$$B^k_i(t) = \frac{t - \xi_i}{\xi_{i+k-1} - \xi_i} B^{k-1}_i(t) + \frac{\xi_{i+k} - t}{\xi_{i+k} - \xi_{i+1}} B^{k-1}_{i+1}(t), \ k = d, d-1, \ldots, 2,$$

and the initial condition,

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

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 $\int_0^t g(s) \, ds$ for $t \in [0, \tau]$ (c.f. Remark 1). In general, numerical integration might be required for this purpose. However, our 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 $\int_0^t g(s) \, ds$ is simply an order $d + 1$ B-spline (de Boor, 2001, p. 128),

$$\int_0^t \sum_{i=1}^\kappa \gamma_i B^d_i(t) \, ds = \sum_{i=1}^\kappa \left( \sum_{j=1}^i \gamma_j (\xi_{j+d} - \xi_j)/k \right) B^{d+1}_i(t), \ t \in [\xi_1, \xi_{\kappa+1}] \equiv [0, \tau],$$

where the extra knot value $\xi_{\kappa+d+1}$ needed in the definition of $B^{d+1}_\kappa$ 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 $\gamma$ 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, we study the large sample properties of the proposed estimator. 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 a list of 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. \( \beta_0 \) is an interior point of \( B \).

C2. For any \( \beta \in B \), there exists an \( \varepsilon > 0 \) such that \( X^\top \beta \geq \varepsilon \) almost surely.

C3. The covariate \( X \) is bounded, that is, there exists a constant \( M > 0 \) such that \( \|X\| \leq M \) almost surely.

C4. If for any \( \beta_1, \beta_2 \in B \) and \( g_1, g_2 \in F_r \),

\[
X^\top \beta_1 + \int_0^t g_1(t - s) \, dN(s) \equiv X^\top \beta_2 + \int_0^t g_2(t - s) \, dN(s)
\]

almost surely, then \( \beta_1 = \beta_2 \) and \( g_1 = g_2 \).

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

Remark 5. C1 is commonly assumed in semi-parametric estimation problems. C2 is to guarantee 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 those assumed by Lu et al. (2009) and Zhou et al. (1998) in studying the asymptotic properties of B-spline based estimators.

We can now state the main results concerning the asymptotic behavior of the estimators.

Theorem 1 (Consistency). Under C1-C5, the estimator \( \hat{\theta}_n \) is consistent, that is

\[
\|\hat{\beta}_n - \beta_0\| + \int_0^T |\hat{g}_n(s) - g_0(s)| \, ds \overset{P}{\longrightarrow} 0.
\]
Theorem 2 (The rate of convergence). In addition to C1-C5 and the condition that \(\kappa_n \to \infty\) as \(n \to \infty\), also assume that \(\lim_{n \to \infty} \kappa_n/n = 0\). Then

\[
|\hat{\beta}_n - \beta_0| + \int_0^\tau |\hat{g}_n(s) - g_0(s)| \, ds = O_P\left(\left(\frac{\kappa_n}{n}\right)^{1/2} + \kappa_n^{-r}\right).
\]

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

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

\[
\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{D} N(0, \Sigma_{\beta}),
\]

where \(\Sigma_{\beta}\) is positive definite, and the estimator \(\hat{\beta}_n\) is semi-parametrically efficient. Moreover, \(\Sigma_{\beta}\) is consistently estimated by \(\hat{\Sigma}_{\beta} = A\hat{\mathcal{I}}_{n}^{-1}A^\top\), where \(\mathcal{I}_n\) is the observed information matrix, i.e. the negative Hessian matrix of the sieve log-likelihood, 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

The simulation model was

\[
\lambda_i(t) = X_i^\top \beta + \int_0^t g(t - s) \, dN_i(s), \quad t \in [0, \tau], \quad i = 1, \ldots, n.
\]

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 \(X_{i1} \sim \text{Uniform}[0, 1]\), \(X_{i2} \sim \text{Bernoulli}(0.5)\), \(C_i \equiv \tau = 1\), and \(n = 100, 200, \text{ or } 400\). 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 \(\{1, 2, \ldots, 30\}\).
the process of data simulation and parameter estimation was repeated 1000 times. For each sample size, the covariates were simulated once and held constant across the 1000 simulated data sets.

Table 4.1 shows a summary of the estimates of the regression parameters using respectively 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 $d = 4$, were not shown. From these results we note that both 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 respectively 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 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 4.1 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, the MSE decreases roughly at rate $n$, as expected 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 4.2 shows the point-wise 95% confidence limits and the median of the estimates of the excitation function, using the parametric method and the MBS method 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 4.1, from which 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 equals 2 when $n = 100$ or 200, and equals 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 approximation to the excitation function. For practical applications, it is desirable to have a data driven approach to select these two tuning parameters. To this end, we propose to select the values of $d$ and $m$ by minimizing the Akaike Information Criterion (AIC, Akaike, 1974),

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

where $\ell_{\text{max}}$ 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 $\hat{m}_\text{AIC}$, for different values of $d$ and sample size $n$ is shown in Figure 4.3. From this figure we note that the $\hat{m}_\text{AIC}$ values seem to be distributed around $m_{\text{MIAE}}$, 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 4.2, 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 the excitation function based on the MBS estimator with varying \( d \in \{2, 3, 4\} \) and \( m = \hat{m}_{\text{MIAE}} \) are shown in Figure 4.4 together with those based on the parametric MLE. We note that the empirical 95% confidence bands, though point-wise and not simultaneous ones, 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 practical applications.

5. A Real Data Example

In this section we illustrate our model and estimation method with a data set arising in bladder cancer study. The data was reported by Byar (1980) and has since been frequently used to illustrate statistical methods for recurrent event data analysis, e.g. by Wei et al. (1989), Zeng and Lin (2006), Wellner and Zhang (2007), and Lu et al. (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 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 and an exponential function respectively, and the other semi-parametric with the excitation function only assumed to be positive and monotone decreasing. Note that 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 5.1 and Figure 5.1 respectively. The standard errors estimates in Table 5.1 were obtained by inverting the observed information matrix discussed earlier. They were nearly identical to the standard errors obtained using a bootstrap method (not reported) to be discussed below.

From Table 5.1 we note that 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 by other authors, e.g. Zeng and Lin (2006), Wellner and Zhang (2007) and Lu et al. (2009). However, 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 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 $\chi^2$-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 $\chi^2$-test, suggesting that the excitation effect decays over time.

Figure 5.1 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 5.1 and omitted to save space. Figure 5.1 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 also inspected the estimates of the regression coefficients with the number \( m \) of interior knot intervals varying in the range 1 to 20, and 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 \( \hat{\Lambda}_i(t) = \int_0^t \hat{\lambda}_i(s) \, ds = \int_0^t \{ \mathbf{X}_i^\top \hat{\beta} + \int_0^s \hat{g}(s-u) \, d N_i(u) \} \, ds, \ t_{i,0} \equiv 0, \ t_{i,n_i+1} \equiv C_i, \) 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 5.2, which shows that the Nelson-Aalen plot of 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 as well. This is further evidence that a parametric model with parametric excitation function produces 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.
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 the monotone B-spline was proposed, and its large sample properties and asymptotic optimality 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 remaining. 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.

Acknowledgment The comments from the Co-Editor, the 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 Hong Kong Research Grant Council grants 601011 and 601612 (K. Chen).
References


CoFE Discussion Paper 02-05, Center of Finance and Econometrics, University of Konstanz.


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
Table 4.1: 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 \([\text{MBS}(d, m)]\) with order \( d \) and equally spaced interior knots \( \{\xi_0, \ldots, \xi_{n+1}\} = \{i/m; \ i = 0, \ldots, m\} \). 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.

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \beta_0 = 0.5 )</th>
<th>( \beta_1 = 1.0 )</th>
<th>( \beta_2 = 2.0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAR</td>
<td>Bias</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.265</td>
<td>0.177</td>
</tr>
<tr>
<td></td>
<td>SEE</td>
<td>0.247</td>
<td>0.170</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.914</td>
<td>0.930</td>
</tr>
<tr>
<td>( \text{MBS}(3,5) )</td>
<td>Bias</td>
<td>-0.002</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.264</td>
<td>0.176</td>
</tr>
<tr>
<td></td>
<td>SEE</td>
<td>0.247</td>
<td>0.170</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.914</td>
<td>0.932</td>
</tr>
<tr>
<td>( \text{MBS}(3,6) )</td>
<td>Bias</td>
<td>-0.002</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.264</td>
<td>0.176</td>
</tr>
<tr>
<td></td>
<td>SEE</td>
<td>0.247</td>
<td>0.169</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.920</td>
<td>0.930</td>
</tr>
<tr>
<td>( \text{MBS}(3,7) )</td>
<td>Bias</td>
<td>-0.002</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.264</td>
<td>0.176</td>
</tr>
<tr>
<td></td>
<td>SEE</td>
<td>0.248</td>
<td>0.169</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.915</td>
<td>0.930</td>
</tr>
<tr>
<td>( \text{MBS}(4,4) )</td>
<td>Bias</td>
<td>-0.002</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.264</td>
<td>0.176</td>
</tr>
<tr>
<td></td>
<td>SEE</td>
<td>0.247</td>
<td>0.169</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.916</td>
<td>0.932</td>
</tr>
<tr>
<td>( \text{MBS}(4,5) )</td>
<td>Bias</td>
<td>-0.002</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.264</td>
<td>0.176</td>
</tr>
<tr>
<td></td>
<td>SEE</td>
<td>0.247</td>
<td>0.169</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.912</td>
<td>0.930</td>
</tr>
<tr>
<td>( \text{MBS}(4,6) )</td>
<td>Bias</td>
<td>-0.002</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.264</td>
<td>0.176</td>
</tr>
<tr>
<td></td>
<td>SEE</td>
<td>0.247</td>
<td>0.169</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.922</td>
<td>0.930</td>
</tr>
</tbody>
</table>
Figure 4.1: 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.
Figure 4.2: The excitation function (central solid curve) and the point-wise 2.5th, 50th, and 97.5th percentiles of the 1000 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 $\{i/m; \ i = 0, \ldots, m\}$. 
Figure 4.3: Distribution of the optimal $m$ values selected using the AIC.
Table 4.2: Summary of the regression coefficient estimates using the MBS\((d,m)\) estimator with \(m = \hat{m}_{\text{AIC}}\).

\[
\begin{array}{ccccccccccc}
\text{d} = 2 & \text{Bias} & -0.002 & 0.000 & 0.006 & 0.015 & 0.001 & -0.009 & -0.011 & 0.000 & -0.013 \\
& \text{SE} & 0.264 & 0.176 & 0.133 & 0.560 & 0.329 & 0.240 & 0.352 & 0.265 & 0.176 \\
& \text{SEE} & 0.247 & 0.169 & 0.130 & 0.536 & 0.327 & 0.236 & 0.344 & 0.248 & 0.171 \\
& \text{CP} & 0.912 & 0.932 & 0.949 & 0.936 & 0.943 & 0.949 & 0.937 & 0.938 & 0.943 \\
\text{d} = 3 & \text{Bias} & -0.001 & -0.002 & 0.005 & 0.013 & -0.002 & -0.012 & -0.012 & -0.006 & -0.02 \\
& \text{SE} & 0.264 & 0.176 & 0.132 & 0.558 & 0.327 & 0.238 & 0.352 & 0.264 & 0.175 \\
& \text{SEE} & 0.247 & 0.170 & 0.130 & 0.535 & 0.326 & 0.235 & 0.344 & 0.247 & 0.170 \\
& \text{CP} & 0.915 & 0.933 & 0.951 & 0.933 & 0.943 & 0.949 & 0.938 & 0.941 & 0.942 \\
\text{d} = 4 & \text{Bias} & -0.001 & -0.001 & 0.005 & 0.015 & 0.002 & -0.008 & -0.010 & 0.000 & -0.012 \\
& \text{SE} & 0.264 & 0.176 & 0.133 & 0.560 & 0.329 & 0.240 & 0.351 & 0.265 & 0.176 \\
& \text{SEE} & 0.247 & 0.169 & 0.130 & 0.536 & 0.327 & 0.236 & 0.344 & 0.248 & 0.171 \\
& \text{CP} & 0.918 & 0.931 & 0.949 & 0.932 & 0.943 & 0.948 & 0.938 & 0.939 & 0.944 \\
\end{array}
\]
Figure 4.4: 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.
Table 5.1: 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      | -3.3e-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       |
Figure 5.1: 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% (point-wise) confidence bands based on the MBS estimator were obtained by parametric bootstrap.
Figure 5.2: 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 semiparametric self-exciting process model with a monotone excitation function estimated using the MBS method (SEPP-MBS).