International Journal of Statistics and Probability; Vol. 3, No. 3; 2014
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Modeling Event Clustering Using the *m*-Memory Cox-Type Self-Exciting Intensity Model

Feng Chen[1] & Kani Chen[2]

[1] School of Mathematics and Statistics, University of New South Wales, Sydney, Australia

[2] Department of Mathematics, The Hong Kong University of Science and Technology, Kowloon, Hong Kong

Correspondence: Feng Chen, School of Mathematics and Statistics, The University of New South Wales, Sydney, NSW 2052, Australia. Tel: 61-2-9385-7026. E-mail: feng.chen@unsw.edu.au

Received: June 10, 2014 Accepted: July 14, 2014 Online Published: July 28, 2014

doi:10.5539/ijsp.v3n3p126 URL: http://dx.doi.org/10.5539/ijsp.v3n3p126

Abstract

In the analysis of point processes or recurrent events, the self-exciting component can be an important factor in understanding and predicting event occurrence. A Cox-type self-exciting intensity point process is generally not a proper model because of its explosion in finite time. However, the model with *m*-memory is appropriate to analyze sequences of recurrent events. It assumes the most recent *m* events multiplicatively affect the conditional intensity of event occurrence. Aside from the interpretability, one advantage is the simplicity of the estimation and inference–the Cox partial likelihood can be applied and the resulting estimator is consistent and asymptotically normal. Another advantage is that the model can be applied to the analysis of case-cohort data via the pseudo-likelihood approach. The simulation studies support the asymptotic theory. Application is illustrated with analysis of a bladder cancer dataset and of an Australian stock index dataset, which shows evidence of self-excitation.

Keywords: longitudinal data, proportional intensity model, partial likelihood, recurrent events

1. Introduction

Recurrent event data are encountered frequently in many areas of scientific endeavor, such as the modeling and predictions of earthquakes and other disastrous events, study of the patterns of neural firings in neuroscience, assessing the efficacy of cancer medications in suppressing the recurrence of tumors, and analysis of the risk of default on debt repayments by borrowers. Point processes are natural stochastic process models for the modeling and analysis of recurrent event data. Depending on the form of the data available and the research questions of interest, one of two types of point process models might be appropriate. If the data is in the form of a single long string of event recurrence times, it might be of interest to predict the next event recurrence time by exploiting potential dependence of the waiting times between events on past events or on exogenous covariates. Models of this type include the self-exciting point process (Hawkes, 1971; Ogata, 1978), the modulated renewal process (Cox, 1972; Oakes & Cui, 1994; Lin & Fine, 2009) and the autoregressive conditional duration models (Engle & Russell, 1998; Fernandes & Grammig, 2006). Another form of data, which appears most often in medical statistics, consists of multiple strings of event times and covariates for each string. The number of events in each string is typically small due to censoring, and some individuals might not have experienced a single event by the censoring time. For data in this form, the main interest in practice is to assess the effects of the covariates on the frequency of event recurrence. Examples of the models that suits this purpose include the Cox proportional intensities (CoxPI) model (Andersen & Gill, 1982) and the proportional means model (Lin et al., 2000; Wellner & Zhang, 2007).

In this paper we consider a model that suits the analysis of data in the multiple string form. We are motivated by the temporal clustering of event times observed in individual strings with multiple events. The temporal clustering of event times indicates potential self-exciting effect among the events, which, if not properly accounted for, can lead to erroneous inferences about the effects of the covariate. Although the CoxPI model does not explicitly account for the potential self-exciting effect and therefore is not directly suitable for data with signs of event clustering, its many well-known theoretical and computational advantages motivate us to build our model based on it. The aim is to explicitly incorporate a self-exciting feature in the model, while at the same time retaining as many advantages of the CoxPI model as possible.

The method to model event clustering in this paper is motivated by the aforementioned Hawkes self-exciting point process model, which is a simple point process $N(t)$ with intensity process $\lambda(t)$ in a self-exciting form,

$$\lambda(t) = \nu + \int_0^t g(t-u)\,\mathrm{d}N(u), \tag{1}$$

where $\nu > 0$ is the background event intensity and $g(\cdot) \geq 0$ is the excitation function. The CoxPI model is a simple point process with intensity process given by

$$\lambda(t) = \lambda_0(t)\exp\{z(t)^{\mathrm{T}}\beta\},$$

where $\lambda_0(t)$ is a baseline intensity function, $z(t)$ is a vector valued process of covariates, and β a vector of parameters. A naïve extension of the CoxPI model by including the integral term in (1) to the logarithm of the intensity, i.e.,

$$\log\lambda(t) = \log\lambda_0(t) + z(t)^{\mathrm{T}}\beta + \int_0^t g(t-u)\,\mathrm{d}N(u), \tag{2}$$

does not lead to an appropriate model because such a model can easily be explosive; see Remark 1 below for an explanation. However, if we modify the integral term in (2) by restricting the contribution of past events on the current event intensity to the most recent m ($< \infty$) events, then the resulting model does not suffer from the explosion issue and still posses an explicit self-excitation feature. Such a model, which we call the m-memory Cox-type self-exciting intensity (CoxSEI(m)) model, shall be an appropriate model for recurrent event data with temporal clustering of event times.

The rest of this paper is organized as follows. In Section 2, we present the CoxSEI(m) model and the estimation procedure. In Section 3, we present some asymptotic properties of the estimators. In Section 4, we report the results of some simulation studies and analysis of a bladder cancer data set and an Australian stock index data set. Section 5 concludes with discussion. Technical proofs are relegated to the Appendix. All computation was done in R (R Core Team, 2014) with the aid of the package `coxsei` written by the authors.

2. The CoxSEI(m) Model and the Estimation Procedure

Consider a point process $N(t) = \sum_{j=1}^{\infty} 1_{\{T_j \leq t\}}$, with $t \in [0,\infty)$ and T_j denoting the j-th event time. As a CoxSEI(m) point process, $N(\cdot)$ has a conditional intensity process given by

$$\lambda(t) = \mu(t)\exp\{Z(t)^{\mathrm{T}}\beta + \phi(t)\} \tag{3}$$

where $\mu(t)$ is an unspecified baseline intensity, $Z(t)$ is a possibly time-varying p-vector of covariates, β is a p-vector of regression coefficients which measures the effects of the covariates to the intensity on the log scale, and $\phi(t)$ is a self-exciting term depending on past events of the process,

$$\phi(t) = \phi(t,\alpha,\gamma) = \sum_{j=1}^{m\wedge N(t-)} \alpha g(t - T_{N(t-)-j+1}, \gamma) = \sum_{j\in\mathcal{M}(t)} \alpha g(t-T_j,\gamma), \tag{4}$$

where $\mathcal{M}(t) = \{j\colon \{N(t-)+1-m\}\vee 1 \leq j \leq N(t-)\}$ denotes the set of indexes of the most recent m events in the past. The *excitation function* g is specified up to a parameter γ. Normally g is a positive decaying function, and the parameter γ regulates the decay rate. The decay of g implies that the more recent events have stronger direct effects on the current event intensity than the events in the more remote past. Typical examples of g include the exponential decay function $g(t,\gamma) = \exp(-\gamma t)$ and the polynomial function $g(t,\gamma) = (1+t)^{-\gamma}$, with $\gamma > 0$ (e.g., Errais et al., 2010; Ogata, 1988). The parameter α measures the initial magnitude of the self-exciting effect. While a positive α implies the self-exciting effect is genuinely excitatory, a negative α would imply that the "self-exciting" effect is in fact *inhibitory* (Kopperschmidt & Stute, 2009).

Remark 1 We assume m to be a positive integer. If $m = 0$, the self-exciting component vanishes and the CoxSEI(m) model (3) reduces to a CoxPI model. If $m = \infty$, the CoxSEI(m) model becomes an infinite-memory Cox-type self-exciting process. In this case, the process will be explosive under fairly general conditions if $\alpha > 0$. To see this, suppose the baseline intensity $\mu(\cdot)$ is bounded away from 0 and ∞, $g(t,\gamma) > 0$ is decreasing in t, the covariate processes $Z(\cdot)$ are bounded, and the regression coefficients β are all finite. Write $c = \inf\{\mu(t)\exp(Z(t)^{\mathrm{T}}\beta)$:

$t \geq 0\} > 0$. Let $\Delta T_1 = T_1$, $\Delta T_j = T_j - T_{j-1}$, $j \geq 2$ denote the durations between events. For any fixed $t > 0$, there exists $\varepsilon > 0$ such that $\varepsilon \sum_{j=1}^{\infty} 1/j^2 < t$. As a result,

$$\Pr(N(t) = \infty) = \Pr(\sum_{j=1}^{\infty} \Delta T_j \leq t) \geq \Pr(\Delta T_j \leq \varepsilon / j^2, j = 1, 2, \dots). \tag{5}$$

Clearly the probability on the right hand side of (5) can be written as

$$\begin{aligned} & \prod_{j=1}^{\infty} \Pr(\Delta T_j \leq \varepsilon/j^2 \,|\, \Delta T_k \leq \varepsilon/k^2, 1 \leq k \leq j-1) \\ \geq & \prod_{j=1}^{\infty} \left(1 - \exp\left[-c \, \exp\{(j-1)\alpha g(t,\gamma)\} \, \varepsilon/j^2\right]\right) > 0. \end{aligned}$$

For CoxSEI(m) processes with finite m, under mild regularity conditions, such as C1-C4 to be presented later, the intensity process $\lambda(\cdot)$ is bounded away from 0 and ∞ with probability one. As a result, it will not be explosive for sure (with probability 1). We shall only consider the CoxSEI(m) model with finite m.

Remark 2 Under the CoxSEI(m) model, certain Markov property can be derived for the process. Set $T_k = 0$ for $k \leq 0$ for notational convenience. Let $\xi_j(t) = T_{N(t-)-j+1}$, $1 \leq j \leq m$, be the times of the most recent m events before time t. Let $\xi(t) = (\xi_1(t), ..., \xi_m(t))^{\mathrm{T}}$, $t \geq 0$, be an m-vector continuous time process. It can be verified that given the covariates and $\xi(t)$, $\xi(s)$ and $\xi(\tau)$ with $s < t < \tau$ are conditionally independent. Therefore $\xi(t)$ is a continuous time Markov process of dimension m, conditioning on the covariates.

Suppose we have n independent observations of the CoxSEI(m) process $N(t)$ and the covariate process $Z(t)$ until a censoring time C which is assumed to be independent of $N(t)$ conditional on $Z(t)$. Denote the observations by

$$\{N_i(t), Z_i(t);\ t \leq C_i,\ i = 1, ..., n\}.$$

Write $\theta = (\beta, \alpha, \gamma)^{\mathrm{T}}$, $\Psi(t, \theta) = Z(t)^{\mathrm{T}}\beta + \phi(t, \alpha, \gamma)$, and $Y(t) = I\{C \geq t\}$. Denote the corresponding i.i.d. copies of $T_j, 1 \leq j \leq N(C)$, $\mathcal{M}(\cdot)$, $\Psi(\cdot)$, and $Y(\cdot)$ respectively by $T_{ij}, 1 \leq j \leq N_i(C_i)$, $\mathcal{M}_i(\cdot)$, $\Psi_i(\cdot)$, and $Y_i(\cdot)$, $i = 1, \dots, n$.

The estimation of the CoxSEI(m) model is along the same lines as that of the CoxPI model. The estimation of the parametric part relies on the Cox partial likelihood, and the estimation of the cumulative baseline intensity is motivated by the Breslow estimator (Breslow, 1972) as in the CoxPI model. Specifically, we note that given the history of the n subjects prior to time t and the observation that an event occurs at time t, the conditional probability that the event pertains to the i-th subject is

$$\frac{\exp\{\Psi_i(t, \theta)\}}{\sum_{j \in \mathcal{R}_t} \exp\{\Psi_j(t, \theta)\}},$$

where $\mathcal{R}_t = \{k : C_k \geq t, 1 \leq k \leq n\}$. Therefore, the Cox partial likelihood is

$$L(\theta) = \prod_{i=1}^{n} \prod_{0 \leq t \leq C_i} \Big[\frac{\exp\{\Psi_i(t, \theta)\}}{\sum_{j \in \mathcal{R}_t} \exp\{\Psi_j(t, \theta)\}} \Big]^{\mathrm{d}N_i(t)}.$$

The maximum partial likelihood estimator $\hat{\theta}$ is defined as the maximizer of $L(\theta)$ over the parameter space $\Theta \subset \mathbb{R}^{p+2}$. The estimator of the cumulative baseline intensity function $U(\cdot) = \int_0^{\cdot} \mu(t)\, \mathrm{d}t$ is similar to the Breslow estimator (Breslow, 1972) and is given by

$$\hat{U}(t) = \int_0^t \frac{\mathrm{d}N_{\cdot}(s)}{\sum_{j \in \mathcal{R}_s} \exp\{\Psi_j(s, \hat{\theta})\}} \tag{6}$$

where $N_{\cdot}(t) = \sum_{i=1}^{n} N_i(t \wedge C_i)$.

3. Large Sample Properties of the Estimators

The following conditions are needed to prove the large sample properties of $\hat{\theta}$. Let θ_0 be the true value of θ in Θ. We use the symbols ∂_θ and $\partial^2_{\theta\theta^{\mathrm{T}}}$ to denote the operators of finding first and second order partial derivatives with respective to θ.

C1. The covariate process $Z(\cdot)$ is bounded.

C2. The parameter space Θ is closed, bounded and connected, and contains θ_0 as an interior point. Moreover, $\Pr(\Psi(\cdot,\theta_1) = \Psi(\cdot,\theta_2)) < 1$ for any $\theta_1 \neq \theta_2 \in \Theta$.

C3. The excitation function $g(t,\gamma)$ is positive, bounded, decreasing in t, and twice continuously differentiable in γ. The baseline intensity $\mu(\cdot)$ is bounded and continuous.

C4. The matrix $\Sigma(\theta)$ is finite and positive definite and continuous at θ_0 where

$$\Sigma(\theta) = E[\int_0^\infty \{\partial_\theta \Psi(t,\theta) - \overline{\partial_\theta \Psi}(t,\theta)\}^{\otimes 2} Y(t)\lambda(t)\,\mathrm{d}t],$$

and

$$\overline{\partial_\theta \Psi}(t,\theta) = \frac{E[\{\partial_\theta \Psi(t,\theta)\}Y(t)\exp\{\Psi(t,\theta)\}]}{E[Y(t)\exp\{\Psi(t,\theta)\}]}.$$

Remark 3 C1 is commonly assumed in the literature and C2 is an identifiability condition. In C3, the monotonicity of g is practical but can be relaxed. C4 is essential to the asymptotic normality of $\hat{\theta}$. Unlike the classical Cox model, the global concavity of the log-partial likelihood does not automatically hold in the CoxSEI(m) process.

The large sample properties of $\hat{\theta}$ and $\hat{U}(\cdot)$ are given in the following two propositions.

Proposition 4 *Assume C1-C3 hold. Then, $\hat{\theta}$ is strongly consistent. If, moreover C4 holds,*

$$\sqrt{n}(\hat{\theta} - \theta_0) \to N(0, \Sigma^{-1}).$$

where $\Sigma = \Sigma(\theta_0)$, which can be consistently estimated by $-\frac{1}{n}\partial^2_{\theta\theta^{\mathrm{T}}} \log L(\theta)\big|_{\theta=\hat{\theta}}$.

Remark 5 Similar to the efficiency of the Cox partial likelihood estimator in the proportional hazards model, it can be verified that Σ is the Fisher information matrix for θ and that, as a result, $\hat{\theta}$ is a semiparametric efficient estimator of θ.

Proposition 6 *Let $\mu_0(\cdot)$ and $U_0(\cdot)$ be the true baseline intensity and baseline cumulative intensity functions respectively. Let $r^{(m)}(t,\theta)$, $R^{(m)}(t,\theta)$, $m = 0,1,2$, and $\mathscr{I}(\theta)$ be as those defined by (A.1)-(A.7) in the Appendix. Assume C1-C4 hold. Then the process $\sqrt{n}\{\hat{U}(\cdot) - U_0(\cdot)\}$ converges weakly to a Gaussian process with mean zero and covariance function*

$$\int_0^{t_1\wedge t_2} \frac{\mu_0(s)\,\mathrm{d}s}{r^{(0)}(s,\theta_0)} + \int_0^{t_1} \frac{r^{(1)}(s,\theta_0)^{\mathrm{T}}}{r^{(0)}(s,\theta_0)}\mu_0(s)\,\mathrm{d}s\,\Sigma^{-1} \int_0^{t_2} \frac{r^{(1)}(s,\theta_0)}{r^{(0)}(s,\theta_0)}\mu_0(s)\,\mathrm{d}s,$$

which can be estimated uniformly consistently by

$$n\Big\{\int_0^{t_1\wedge t_2} \frac{\mathrm{d}N_\cdot(s)}{R^{(0)}(s,\hat{\theta})^2} + \int_0^{t_1} \frac{R^{(1)}(s,\hat{\theta})^{\mathrm{T}}}{R^{(0)}(s,\hat{\theta})^2}\,\mathrm{d}N_\cdot(s)\mathscr{I}(\hat{\theta})^{-1} \int_0^{t_2} \frac{R^{(1)}(s,\hat{\theta})}{R^{(0)}(s,\hat{\theta})^2}\,\mathrm{d}N_\cdot(s)\Big\}.$$

Remark 7 The large sample distribution of $\hat{\theta}$ is approximately normal with mean θ_0 and variance $\mathscr{I}(\hat{\theta})^{-1}$, and the distribution of $\hat{U}(t)$ is approximately normal with mean $U_0(t)$ and variance

$$\int_0^{t} \frac{\mathrm{d}N_\cdot(s)}{R^{(0)}(s,\hat{\theta})^2} + \int_0^{t} \frac{R^{(1)}(s,\hat{\theta})^{\mathrm{T}}}{R^{(0)}(s,\hat{\theta})^2}\,\mathrm{d}N_\cdot(s)\mathscr{I}(\hat{\theta})^{-1} \int_0^{t} \frac{R^{(1)}(s,\hat{\theta})}{R^{(0)}(s,\hat{\theta})^2}\,\mathrm{d}N_\cdot(s). \tag{7}$$

If the baseline intensity function $\mu(\cdot)$ rather than its integral is of interest, then we can estimate it using one of the many nonparametric methods available, such as kernel smoothing (Ramlau-Hansen, 1983) and the local polynomial method (Chen et al., 2011). To this end, we first note the intensity process of the aggregate process $N_\cdot(t)$ has a multiplicative form $\{\sum_{i\in\mathcal{R}_t}\Psi_i(t,\theta)\}\mu(t)$. Since the nonparametric estimator of $\mu(t)$ with the exposure process $\sum_{i\in\mathcal{R}_t}\Psi_i(t,\theta)$ fully known typically has a rate of convergence slower than $\sqrt{n}$, while the plug-in estimator of the exposure process $\sum_{i\in\mathcal{R}_t}\Psi_i(t,\hat{\theta})$ has a $\sqrt{n}$ rate, we can simply estimate $\mu(t)$ by assuming the estimated exposure process is the unknown true exposure process.

The proof of Proposition 4 is given in the Appendix. The proof of Proposition 6 is essentially the same as that of Theorem 3.4 and Corollary 3.5 in Andersen and Gill (1982), and is omitted.

4. Numerical Studies

4.1 Simulation

This section reports the results of a simulation study. The simulation model is CoxSEI(2) with baseline intensity $\mu(t) = 1 + 0.5\cos(2\pi t)$ and excitation function $g(t) = \alpha \exp(-\gamma t) = 0.7e^{-10t}$. The covariate process has three static components Z_i, $i = 1, 2, 3$. Their design distributions are Uniform[0.5,1.5], Uniform[1.5,2.5] and Bernoulli(0.5), respectively. The regression coefficients associated with the Z_i are $\beta_1 = 0.2$, $\beta_2 = 0.4$, $\beta_3 = 0.6$. The censoring variable is independently generated, following lognormal$(0, 0.1)$. The sample size is $n = 100$. The simulation was repeated 100 times. The results are summarized in Table 1. It is seen that the estimates of the parameters β_i, α and γ seem unbiased, and the estimates of the standard errors are close enough to the empirical ones. The empirical distributions of all estimates are very close to normal distributions, with the two-sided Kolmogorov-Smirnov tests of normality all having p-values much greater than 0.05.

Table 1. Results of the simulation–fitting the correct model

Parameter	β_1	β_2	β_3	α	γ
True	0.2	0.4	0.6	0.7	10
Mean Est.	0.200	0.403	0.616	0.690	10.832
Mean SE Est.	0.147	0.147	0.096	0.084	3.382
Empirical SE	0.166	0.133	0.095	0.086	3.326
P-value for the K-S test of normality	0.976	0.937	0.833	0.761	0.622

The estimates of the cumulative baseline intensity function are shown in the left panel of Figure 1, which are close to the true cumulative intensity function. The standard error estimates calculated from the variance estimator (7) are shown in the right panel of Figure 1 together with the empirical standard errors, from which we note the variance estimator (7) for the cumulative intensity estimator is slightly biased upward, but not by much. We therefore conclude that the proposed estimation procedure works well and conforms with the theory.

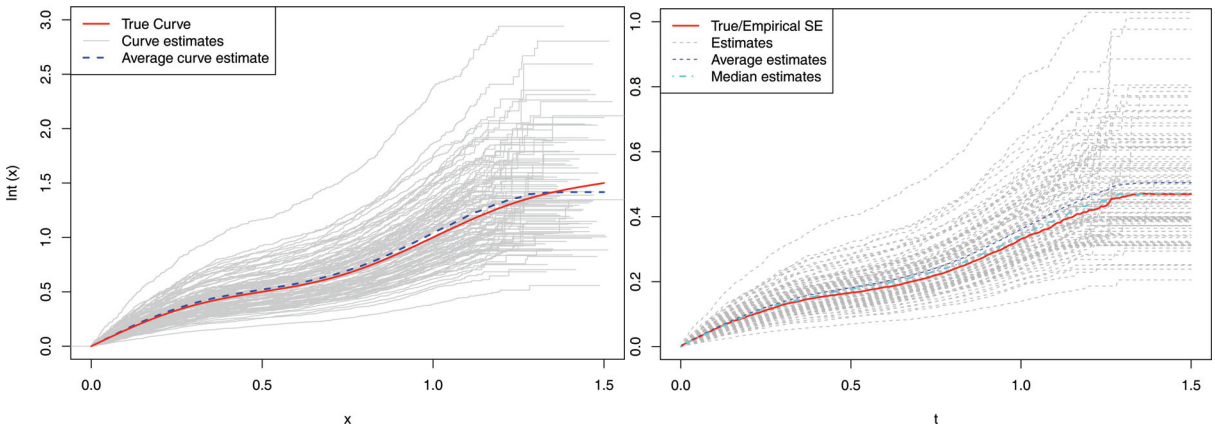

Figure 1. The estimates of the cumulative baseline intensity function (left) and of the standard errors (right) based on the simulated data

Table 2. Results of the simulation–fitting the CoxPI model to data generated by CoxSEI(2) models

	β_1	β_2	β_3
True value	0.2	0.4	0.6
Average estimate	0.267	0.543	0.813
Empirical SE	0.225	0.183	0.121
Average SE estimate	0.146	0.146	0.092

To evaluate the effects of neglecting self-excitation on the estimation of the covariate effect, we fit the ordinary CoxPI model to the data generated from the CoxSEI(2) model. The results are shown in Table 2. The estimated

covariate effects are clearly inflated and the standard errors are generally underestimated. This implies that application of the CoxPI model to recurrent event data without accounting for the potential self-exciting effect may lead to erroneous inference about the covariate effects.

4.2 Analysis of a Bladder Cancer Dataset

We illustrate the CoxSEI(m) model with two real-life examples. The first is a bladder cancer study reported by Byar (1980) and frequently used to illustrate event history data analysis methods (e.g. Wei et al., 1989; Therneau & Hamilton, 1997; Wellner & Zhang, 2007). A total of 118 patients with superficial bladder tumors were admitted to the study between November, 1971 and August, 1976. The tumors were removed transurethrally and patients were randomly assigned to one of three treatment groups: placebo, pyridoxine, and thiotepa. For patients who experienced tumor recurrence, the new tumors were removed at each visit. The initial number of tumors and the size of the largest initial tumor were recorded for each patient. The censoring time was the earlier of death due to bladder cancer or other causes and end of study. The follow-up time of all patients varies from 0 to 64 months, the number of recurrences experienced by the patients varies between 0 and 9 with mean 1.6 and variance 5.3; see Figure 2. The data is available from the `R` package `survival` (Therneau and original Splus→R port by Thomas Lumley, 2011). We have made slight modifications to the data by adding 0.5 to the two 0 censoring times and to the censoring times that equal the corresponding patient's last recurrence time. These modifications caused no appreciable difference to the numerical result of fitting the CoxPI model using the `coxph` function from the `R` package `survival`.

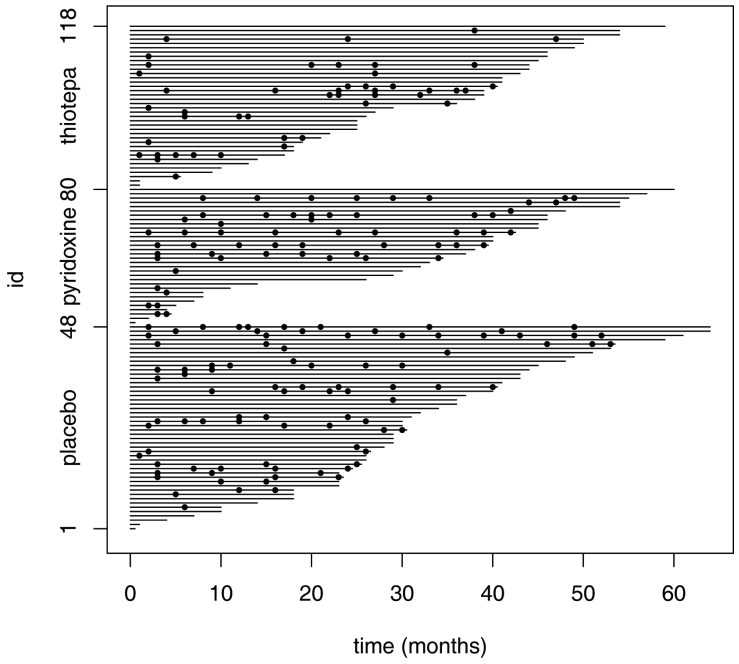

Figure 2. Bladder tumor recurrence (solid point) and censoring (end of line) times of the 118 bladder cancer patients

We fitted the CoxSEI(m) model to the modified data with $m = 0, 1, \ldots, 9$ and calculated the corresponding values of the Akaike information criterion (AIC), which is defined as minus twice the maximized log-partial likelihood value plus twice the number of parameters involved in the partial likelihood. With $m = 0$ the AIC value was 1626.5, while with $m \geq 1$ the AIC values was in the range $[1552.9, 1571.6]$ with the minimal value achieved by $m = 2$, which suggests the CoxSEI(2) model gives the best fit to the data. The results of fitting the CoxPI and the CoxSEI(2) models are shown in Table 3. It is noted that by the CoxPI model the treatment thiotepa has a statistically significant suppressing effect on tumor recurrence intensity in the presence of other covariates. However, in the CoxSEI(2) model, while thiotepa still seems to have a beneficial effect in the presence of other covariates and the self-exciting effect, the beneficial effect is much less conclusive with a p-value substantially greater than 0.05, even if a single-sided alternative is assumed. Since the estimated α parameter of the self-exciting term is positive and statistically highly significant, and the AIC suggests CoxSEI(m) with $m > 0$ fits much better to the data than the CoxPI model, it seems plausible to conclude the self-exciting effect among bladder tumor recurrences is genuine.

From a biological point of niew, it also seems natural to suspect the occurrence of a tumor and the ensuing surgery to remove it could damage the bladder tissue, rendering further tumor recurrences more likely to happen. The neglect of the self-exciting effect could have been the cause of inflated beneficial effect of thiotepa in the ordinary CoxPI model, which is similar to the false positives caused by fitting generalized linear models to overdispersed data without properly accounting for overdispersion.

Table 3. Results of fitting the CoxPI and CoxSEI(2) models to the bladder cancer data

	CoxPI model			CoxSEI(2) model		
	Estimate	StdErr	P-value	Estimate	StdErr	P-value
`pyridoxine`	0.019	0.171	0.91	0.118	0.173	0.50
`thiotepa`	-0.518	0.186	0.0054**	-0.253	0.191	0.19
`number`	0.187	0.036	2.12e-07***	0.115	0.040	0.004**
`size`	-0.007	0.044	0.87	-0.004	0.046	0.94
α	NA	NA	NA	0.924	0.122	3.09e-14***
γ	NA	NA	NA	0.005	0.010	0.31

Signif. codes: 0 *** 0.001 ** 0.01 * 0.5. 0.1 1

4.3 Analysis of an Australian Stock Index Data Set

As an example where the baseline event intensity might also be of interest, we consider data on intra-day times of exceedance of a threshold value by the tick-by-tick return of an Australian stock index, the All Ordinaries Index. Our consideration of the index return exceedance process is motivated by Embrechts et al. (2011). During the period from 1 January 1996 to 3 June 2011 GMT, there were roughly 4,000,000 price moves of the All Ordinaries Index. The corresponding tick-by-tick log-returns varied in the range $[-1.114, 1.103] \times 10^{-1}$, with the maximum and minimum returns attained at 10:13:33.614 and 10:14:01.295 respectively on 28 Jun 2010. The 99th percentile of the returns was $q_{(0.99)} = 4.39 \times 10^{-4}$. For the purpose of illustrating the CoxSEI(m) model, we only considered the intra-day times in year 2010 at which the returns exceeded $q_{(0.99)}$. There were 3,131 such exceedances on 254 trading dates in 2010. We filtered out the data on the 24th and 31st December 2010 as the stock exchange closed early at 14:10 on these two days and the baseline event intensity near 14:10 on these days would be substantially higher than on regular trading days when the market closes at 16:10. Since the market dynamics of after hours trading is expected to be different from that of normal hours trading, we also excluded the data outside the normal trading hours, 10:00-16:10. This left us with 3,030 exceedances on 252 trading days. The daily number of exceedances varied between 0 and 66, with mean 12.02 and variance 111.14.

To apply the CoxSEI(m) model, we need the assumption that the return exceedance processes on different days are conditionally independent. A times series plot of the daily number of return exceedances showed quite strong serial correlation even after weekday and month of year were accounted for using a Poisson regression. However, if we fit an order 1 autoregressive time series model with weekday and month as external categorical covariate variables, then the Ljung-Box tests revealed no significant serial correlation among the residuals, with p-values > 0.05 up to lag 14 and > 0.01 up to lag 20. Therefore we assumed that the daily exceedance processes were conditionally independent given weekday, month and the number of exceedances on the previous trading day. We fitted CoxSEI(m) models with exponential excitation function $g(t) = \alpha \exp(-\gamma t)$ and different m values to the data. We then selected the value of m using the AIC. The unit used in measuring time is the hour.

The AIC value was 31766.2 when $m = 0$, and in the range $[31435.5, 31634.7]$ when $m \geq 1$, with the minimal value 31435.5 attained by $m = 1$. The parametric part of the results of fitting the CoxSEI(1) model are shown in Table 4, from which we note that the number of exceedances on the previous day (`yesterday`) has a highly significant positive effect on the current day exceedance intensity. This could be interpreted as an inter-day exciting effect among the return exceedances on the All Ordinaries Index. The month effect is significant with February, June, July and August seeing more and April seeing less exceedances than January. In the presence of other variables, the differences between March, May, September, October, November, December and January were not significant. The weekday effects do not seem to be individually significant. The parameter α is highly significant with a positive value, suggesting the existence of intra-day exciting effect among the return exceedances. The parameter γ is also highly significantly different from 0, indicating the self-exciting effect is decaying over time. The month effect we have observed on the return exceedance intensity is reminiscent of the January effect in financial returns observed in the US financial market. In view of the common theory which relates the January effect to the end of the fiscal year in US, we might also speculate that Australia's end of the fiscal year in June have contributed to the

increased market volatility which is reflected by the increased intensity of the return exceedance process.

Table 4. Parametric part of the results of fitting the CoxSEI(1) model to the all ordinaries index data

	Estimate	StdErr	Z-value	P-value	
Tuesday	0.0710	0.0586	1.2119	0.2255	
Wednesday	0.0439	0.0581	0.7555	0.4500	
Thursday	-0.0316	0.0581	-0.5436	0.5867	
Friday	-0.1127	0.0592	-1.9054	0.0567	.
February	0.2482	0.1117	2.2217	0.0263	*
March	-0.0233	0.1216	-0.1913	0.8483	
April	-0.3778	0.1506	-2.5091	0.0121	*
May	0.0637	0.1181	0.5397	0.5894	
June	0.2458	0.1091	2.2522	0.0243	*
July	0.2764	0.1081	2.5570	0.0106	*
August	0.2279	0.1139	2.0003	0.0455	*
September	0.1345	0.1136	1.1839	0.2365	
October	0.1894	0.1134	1.6698	0.0950	.
November	0.1598	0.1167	1.3694	0.1709	
December	-0.0476	0.1275	-0.3730	0.7091	
yesterday	0.0378	0.0019	20.3590	< 2.2e-16	***
α	1.1441	0.0670	17.0764	< 2.2e-16	***
γ	1.3515	0.1830	7.3859	7.573e-14	***

Signif. codes: 0 *** 0.001 ** 0.01 * 0.5. 0.1 1

In Figure 3 we show the estimated cumulative baseline intensity function and a local linear estimate of the baseline intensity function using the method discussed in Remark 7. From the figure we note the baseline intensity of return exceedance at market open is substantially much higher than in the rest of the trading hours and the intensity during the morning hours are generally higher than in the afternoon hours. The very high return exceedance intensity at market open is to be expected considering that the occurrence of large and sudden price changes of the constituents of the index are likely to be due to the availability of price impacting information accumulated overnight when the local exchange is closed but many overseas exchanges are still running. The relatively high intensity during the rest of morning hours could be linked with the opening of Asian stock exchanges, such as the Malaysian and Singapore stock exchanges at 11:00 AEST (Australian Eastern Standard Time), the Hong Kong and Mainland China exchanges at 11:30 AEST. The opening times are to be postponed by an hour when Australia observes the Daylight Saving Time from the first Sunday of October to the first Saturday of April.

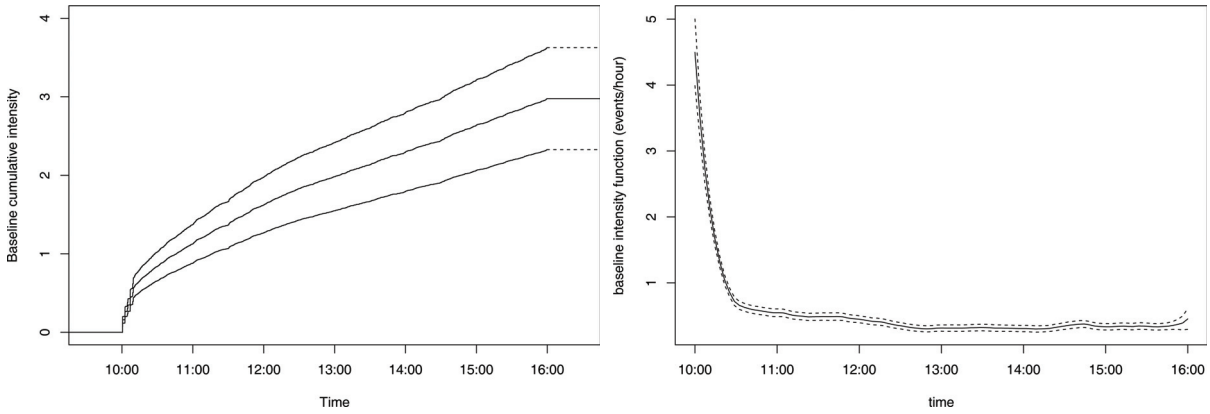

Figure 3. Estimated cumulative baseline intensity (left) and baseline intensity (right) of the All Ordinaries Index return exceedance process with point-wise 95% confidence limits

5. Discussion

In this paper we have considered an extension of the CoxPI model called the CoxSEI(m) model for the analysis of recurrent event data that has the feature of temporal clustering of events experienced by the same individual.

Considering the potentially erroneous inference about the covariate effects that could have been caused by neglecting the self-exciting effects, it seems warranted to develop formal statistical tests to detect the existence of the self-exciting/inhibitory effect. While the likelihood ratio test seems a natural candidate test, the asymptotic null distribution of the likelihood ratio statistic is non-standard. The reason is that under the null hypothesis of no self-exciting effect or equivalently, $\alpha = 0$ in the examples considered in this paper, the parameter γ is unidentifiable and the asymptotic normality of $\hat{\gamma}$ fails, and therefore, the asymptotic distribution of the likelihood ratio statistic under the null fails to be χ^2. In an unreported simulation study, we have found that the empirical distribution of the likelihood ratio statistic deviates substantially from the χ^2_1 and χ^2_2 distributions. Continuing work concerning the asymptotic null distribution of the likelihood ratio test or concerning other tests is desirable.

In constructing the self-excitation term (4) in the CoxSEI(m) model, we have parametrized the effects of the recent m events on the current event intensity in the form of $\alpha g(t - T_{N(t-)+1-j}, \gamma)$ rather than using m unstructured coefficients corresponding to the m events respectively. The consideration behind this choice is interpretability. With a decreasing function g, the individual excitation effects on the current event intensity associated with recent events are monotone with more recent events having more significant effects, which tends to agree with our intuition. In contrast, the unstructured coefficients approach could give rise to estimated coefficients with erratic patterns which are hard to interpret.

The CoxSEI(m) model considered in this work may appear to be a special case of the CoxPI model with a time-dependent covariate $\sum_{j\in\mathcal{M}(t)} g(t-T_j, \gamma)$. However this is generally not the case because of the nonlinear dependence of g on the unknown parameters γ.

In the real data examples, our choice of the parametric form of the excitation function is essentially arbitrary and we have not considered how to select the excitation function using any data-driven procedures. The main reason is that for the correct estimation of the covariate effects and the baseline intensity function, the specific choice of the excitation function is much less important than the inclusion of the self-excitation term in the model. However, further work concerning formal specification tests for the excitation function is clearly desirable.

From the viewpoint of explicitly accounting for potential self-exciting effects in intensity based regression analysis of recurrent event data, one can also consider the combination of the Hawkes self-exciting point processmodel with the Aalen additive intensity regression model (Aalen, 1980). Although care is needed in fitting such a model to guarantee the positivity of the intensity process and the accommodation of self-inhibitory effects might not be as easy, this additive model is arguably more intuitive and easier to interpret in specific contexts. Therefore such a model also deserves investigation, and shall be considered elsewhere.

Another advantage of the CoxSEI(m) model is that it can be applied to the analysis of data collected by the cost effective case-cohort design (Prentice, 1986), with inference based a pseudo-likelihood approach; for details, see the companion paper F. Chen and K. Chen (2014).

Acknowledgements

The stock index data used in this work was supplied by Securities Industry Research Centre of Asia-Pacific (SIRCA) on behalf of Reuters. F. C. was supported by University of New South Wales ECR and SFRG grants. K. C. was supported by Hong Kong Research Grants Coucil grants 601011 and 601612.

References

Aalen, O. O. (1980). A model for nonparametric regression analysis of counting processes. In W. Klonecki, A. Kozek, & J. Rosinski (Eds.), *Mathematical Statistics and Probability Theory*, Volume 2 of *Lecture Notes in Statistics* (pp. 1-25). New York: Springer.

Andersen, P. K., & Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics, 10*(4), 1100-1120.

Breslow, N. (1972). Discussion of paper by D. R. Cox. *Journal of the Royal Statistical Society Series B (Methodological), 34*, 216-217.

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

Chen, F., & Chen, K. (2014). Case-cohort analysis of clusters of recurrent events. *Lifetime Data Analysis, 20*, 1-15.

Chen, F., Yip, P. S. F., & Lam, K. F. (2011). On the local polynomial estimators of the counting process intensity function and its derivatives. *Scandinavian Journal of Statistics, 38*(4), 631-649.

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

Embrechts, P., Liniger, T., & Lin, L. (2011). Multivariate Hawkes processes: an application to financial data. *Journal of Applied Probability, 48*, 367-378.

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

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

Fernandes, M., & Grammig, J. (2006). A family of autoregressive conditional duration models. *Journal of Econometrics, 130*(1), 1-23.

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika, 58*(1), 83-90.

Kopperschmidt, K., & Stute, W. (2009). Purchase timing models in marketing: A review. *AStA Advances in Statistical Analysis, 93*, 123-149.

Lin, D. Y., Wei, L. J., Yang, I., & Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 62*(4), 711-730.

Lin, F., & Fine, J. P. (2009). Pseudomartingale estimating equations for modulated renewal process models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 71*(1), 3-23.

Oakes, D., & Cui, L. (1994). On semiparametric inference for modulated renewal processes. *Biometrika, 81*(1), 83-90.

Ogata, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics, 30*, 243-261.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association, 83*(401), 9-27.

Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika, 73*, 1-11.

R Core Team. (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel function. *Annals of Statistics, 11*(2), 453-466.

Therneau, T. and original Splus−>R port by Thomas Lumley. (2011). *Survival: Survival analysis, including penalised likelihood.* R package version 2.36-9.

Therneau, T. M., & Hamilton, S. A. (1997). rhDNase as an example of recurrent event analysis. *Statistics in Medicine, 16*(18), 2029-2047.

Wei, L. J., Lin, D. Y., & Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association, 84*(408), 1065-1073.

Wellner, J. A., & Zhang, Y. (2007). Two likelihood-based semiparametric estimation methods for panel count data with covariates. *The Annals of Statistics, 35*(5), 2106-2142.

Appendix

Technical Details

The following are the quantities referred to in Proposition 6.

$$r^{(0)}(t,\theta) = E[Y(t)\exp\{\Psi(t,\theta)\}] \tag{A.1}$$

$$r^{(1)}(t,\theta) = E[\{\partial_\theta \Psi(t,\theta)\}Y(t)\exp\{\Psi(t,\theta)\}] \tag{A.2}$$

$$r^{(2)}(t,\theta) = E[\{\partial^2_{\theta\theta^{\mathrm{T}}}\Psi(t,\theta) + \partial_\theta\Psi(t,\theta)^{\otimes 2}\}Y(t)\exp\{\Psi(t,\theta)\}], \tag{A.3}$$

$$R^{(0)}(t,\theta) = \sum_{j\in\mathcal{R}_t}\exp\{\Psi_j(t,\theta)\}, \tag{A.4}$$

$$R^{(1)}(\theta,t) = \sum_{j\in\mathcal{R}_t}\{\partial_\theta\Psi_j(t,\theta)\}\exp\{\Psi_j(t,\theta)\} \tag{A.5}$$

$$R^{(2)}(\theta,t) = \sum_{j\in\mathcal{R}_t}\{\partial^2_{\theta\theta^{\mathrm{T}}}\Psi(t,\theta) + \partial_\theta\Psi(t,\theta)^{\otimes 2}\}\exp\{\Psi_j(t,\theta)\} \tag{A.6}$$

$$\mathscr{I}(\theta) = \sum_{i=1}^{n}\int_0^{C_i}\left\{\frac{R^{(2)}(t,\theta)}{R^{(0)}(t,\theta)} - \frac{R^{(1)}(t,\theta)^{\otimes 2}}{R^{(0)}(t,\theta)^2} - \sum_{k\in\mathcal{M}_i(t)}\partial^2_{\theta\theta^{\mathrm{T}}}\Psi_i(t,\theta)\right\}\mathrm{d}N_i(t) \tag{A.7}$$

Proof of Proposition 4. We first show consistency. Write

$$\frac{1}{n}\log\{L(\theta)\} = \frac{1}{n}\sum_{i=1}^{n}\int_0^{C_i}\Big\{\Psi_i(t,\theta) - \log\Big[\sum_{j=1}^{n}Y_j(t)\exp\{\Psi_j(t,\theta)\}\Big]\Big\}\mathrm{d}N_i(t). \tag{A.8}$$

The conditions C1-C3 ensure that $\Psi(t,\theta)Y(t)$ is P-Glivenko-Cantelli over $[0,t_0]\times\Theta$ for any fixed $t_0 > 0$. As a result,

$$\frac{1}{n}\sum_{j=1}^{n}Y_j(t)\exp\{\Psi_j(t,\theta)\} \to E[Y(t)\exp\{\Psi(t,\theta)\}]$$

uniformly over $[0,t_0]\times\Theta$ with probability one. Separate the integration over $[0,\infty)$ into $[0,t_0)$ and $[t_0,\infty)$ in (A.8) and notice that $\Psi(t,\theta)$ is bounded. We have

$$\frac{1}{n}[\log\{L(\theta)\} - \log\{L(\theta_0)\}] \to l(\theta) - l(\theta_0)$$

where

$$\begin{aligned} l(\theta) &= E\Big(\int_0^C [\Psi(t,\theta) - \log E\{\exp\{\Psi(t,\theta)\}\mu_0(t)Y(t)\}]\,\mathrm{d}N(t)\Big) \\ &= \int_0^\infty E\Big([\Psi(t,\theta) - \log E\{\exp\{\Psi(t,\theta)\}\mu_0(t)Y(t)\}]\exp\{\Psi(t,\theta_0)\}Y(t)\mu_0(t)\Big)\,\mathrm{d}t. \end{aligned}$$

Observe that, for any positive random variable ξ and nonnegative η with positive mean, Jensen's inequality implies

$$E[\eta\log\xi]/E[\eta] \le \log E[\xi\eta] - \log E[\eta].$$

Set $\xi = \exp\{\Psi(t,\theta) - \Psi(t,\theta_0)\}$ and $\eta = \exp\{\Psi(t,\theta_0)\}Y(t)\mu(t)$. It is seen that the integrand in the second expression of $l(\theta)$ achieves maximum when $\theta = \theta_0$. By C2, $l(\theta)$ achieves maximum only at θ_0. The uniform convergence over Θ implies that $\hat\theta$ is strongly consistent.

Under C4, in addition to C1-C3, one can apply the empirical approximation to show

$$\frac{1}{n}\partial^2_{\theta\theta^{\mathrm{T}}}\log\{L(\theta)\} \to -\Sigma$$

in probability, uniformly over B_n, which is a ball centered at θ_0 with radius $O(n^{-1/2})$. By Taylor's expansion,

$$\begin{aligned} &\frac{1}{n}\{\log L(\theta) - \log L(\theta_0)\} \\ &= (\theta-\theta_0)^{\mathrm{T}}\frac{1}{n}\sum_{i=1}^{n}\int_0^{C_i}\{\partial_\theta\Psi_i(t,\theta_0) - \overline{\partial_\theta\Psi}(t,\theta_0)\}\,\mathrm{d}N_i(t) - \frac{1}{2}(\theta-\theta_0)^{\mathrm{T}}\Sigma\,(\theta-\theta_0) + o_P(n^{-1/2}) \end{aligned}$$

uniformly over $\theta \in B_n$. Then, the asymptotic normality of $\hat{\theta}$ holds. In addition, Σ can be consistently estimated by $-\frac{1}{n}\partial^2_{\theta\theta^{\mathrm{T}}} \log L(\theta)$ at $\theta = \hat{\theta}$. The proof is complete. □