

THE STATISTICAL ANALYSIS OF SELF-EXCITING POINT PROCESSES

Kai Kopperschmidt and Winfried Stute

The Nielsen Company (Germany) GmbH and University of Giessen

Abstract: In this paper we study minimum distance estimation for self-exciting point processes. These processes allow for a flexible semi-parametric modeling of the compensator. As a main result, we show the strong consistency and asymptotic normality of our estimator. We also present an application to a data set from micro-economics.

Key words and phrases: Asymptotic normality, consistency, market research, self-exciting point processes.

1. Introduction

Suppose we are given a collection of increasing random points $T_1 < T_2 < \dots$ observed over time, and let N_t , for $t \in I$, denote the number of T_i that fall below t . Point processes, equivalently counting processes, have become an indispensable tool for analyzing the dynamic behavior of random phenomena over time. For example, in a medical or engineering context, the focus is often on the time T_1 elapsed until the first breakdown (death, recidivism, technical default); further events, if any, are then investigated within a so-called recurrent times-to-event analysis. In applications the statistical analysis of such data should allow for a maximum of flexibility when it comes to modeling the underlying process. A crucial role is played by the cumulative intensity or hazard process $\Lambda = \Lambda_t, t \in I$, the compensator in the Doob-Meyer decomposition $N_t = \Lambda_t + M_t$ of N . M is a (local) martingale w.r.t. a given filtration $\mathcal{F}_t, t \in I$, and therefore trend-free, while Λ compensates the monotonicity of N . Moreover, Λ is predictable and thus may serve as a predictor of N in continuous time. In practice, we observe N while Λ , though being predictable, may depend on unknown parameters. For the homogeneous Poisson process $d\Lambda_t = \lambda dt$, where $\lambda > 0$ is a constant intensity. A simple extension that takes into account seasonal effects is the so-called heterogeneous Poisson process in which $\lambda = \lambda_t$ is (only) a function of t . Another simple counting process is, with $X \equiv T_1$,

$$N_t = 1_{\{X \leq t\}}, t \in I,$$

the so-called Single Event process. Here X is a random variable with distribution function F . The corresponding Λ is

$$\Lambda_t = \int_{(-\infty, t]} \frac{1_{\{X \geq x\}}}{1 - F(x-)} F(dx),$$

with $F(\cdot-)$ denoting the left-continuous version of F . The hazard measure

$$\Lambda(dx) \equiv \frac{1_{\{X \geq x\}}}{1 - F(x-)} F(dx)$$

vanishes on $x > X$. This reflects the fact that for the Single Event process no further events may be expected once X has been observed. If rather than one X we observe an i.i.d. sample X_1, \dots, X_n from the distribution function F , we can take the empirical distribution function

$$F_n(t) = n^{-1} \sum_{i=1}^n 1_{\{X_i \leq t\}}.$$

The martingale is

$$M_n(t) = F_n(t) - \int_{(-\infty, t]} \frac{1 - F_n(x-)}{1 - F(x-)} F(dx).$$

As before the hazard measure vanishes right to the largest order statistic. In these examples the intensity measure is random. When F admits a Lebesgue density f , then

$$\Lambda_n(dx) = [1 - F_n(x-)] \lambda(x) dx,$$

where

$$\lambda(x) = \frac{f(x)}{1 - F(x)}$$

is the hazard function of F . This is a special case of a multiplicative model, in which the density of Λ_n is a product of an unknown deterministic function and an observable predictable process.

In a parametric framework the functions f, F , and hence λ , may depend on an unknown parameter ϑ , so that the Radon-Nikodym derivative of $d\Lambda_n$ w.r.t. Lebesgue measure can be written as a product of a deterministic parametric function and an observable predictable random process not depending on ϑ .

This example has motivated many researchers to model also point processes much more general than Single Event processes in a similar way. Let N_1, \dots, N_n be a sample of i.i.d. point processes with the same distribution as N . As before, we can aggregate all N_i to come up with an extension of F_n that aggregates

the Single Event processes. A popular assumption for the cumulative intensity process of N_i is the parametric multiplicative intensity model

$$d\Lambda_{\vartheta}^i = \alpha_i(x, \vartheta)Y_i(x)dx.$$

In many cases it is assumed that α_i is the same for all i . Fundamental contributions to the statistical inference on the true parameter ϑ_0 may be found in Borgan (1984) and Jacobsen (1982, 1984). The monograph of Andersen et al. (1993), in its chapter VI, is based on Borgan (1984). Another relevant reference is Karr (1986). He remarks (p.175) that multiplicative intensity models provide the broadest setting in which good asymptotic theory based on i.i.d. copies of point processes is available. At the same time he admits (p. 170) that precise results may be obtained when α_i is not very – or not at all – random. The statistical inference about the “true parameter” ϑ_0 is usually performed through maximum likelihood. This presumes that the model under study is dominated. The technical justification may come from Theorem 2.31 in Karr (1986), which describes densities of point processes obtained from exponential martingale transformations of the Poisson process.

An important role in the context of point processes, intensities and martingales is the choice of the filtration \mathcal{F}_t , $t \in I$. To make the process $(N_t)_t$ adapted, \mathcal{F}_t must include $\sigma(N_s : s \leq t)$. Very often, one assumes

$$\mathcal{F}_t = \sigma(N_s : s \leq t) \text{ for all } t \in I, \quad (1.1)$$

the left-hand side sometimes being enriched by some events known at $t = 0$. Andersen et al. (1993) call any point process satisfying (1.1) self-exciting irrespective of whether $(N_t)_t$ has special features or not.

Historically, this notion was, however, first coined by Hawkes (1971a,b). He studied a point process with

$$\Lambda(t) = \nu + \int_{(-\infty, t]} g(t-u)N(du), \quad (1.2)$$

where N is a stationary process. By this, Λ captures the information given by the past values of N as transferred through g . For example, if g decreases exponentially fast, the process with cumulative hazard function Λ features so-called shot noise effects. Hawkes and Oakes (1974) studied the connection of these processes with immigration-birth processes. Ogata and Akaike (1982) modified the right-hand side of (1.2) to also incorporate convolution integrals w.r.t. other processes. Moller and Rasmussen (2005, 2006) studied algorithms to simulate Hawkes processes. To the best of our knowledge, Snyder (1975) was the first monograph discussing point processes featuring such special dynamics. A somewhat different model was proposed by Engle and Russell (1997, 1998). They introduced and

studied so-called autoregressive conditional duration models, which constitute an adaptation of ARCH-time series to the point process context.

The following question motivated the present work and is discussed as an example in Section 3. In market research it is important to understand the purchase behavior of customers. Clearly, each customer gives rise to a point process in which each point denotes the time of a purchase of a certain pre-specified fast-moving consumer good. Very often, just after a purchase, we observe saturation effects leading to a downward jump of λ_t where, for the moment, we assume that the cumulative intensity process Λ admits a Lebesgue-density λ_t :

$$\Lambda(dt) = \lambda_t dt.$$

Hence, in general, λ_t is a stochastic process admitting jumps. This fact does not mean that our model has been designed to incorporate structural changes (change points). In our situation, discontinuities of λ_t do not constitute structural changes but are systemic parts of the model. In addition, λ_t often contains components being in charge of seasonal effects that only depend on time and not on individual customer issues. Finally, we might also include external random components such as promotional activities that are not yet part of the internal purchase history of the customer.

One such component, discussed in detail in Section 3, is the impact of TV-advertising. From the company's point of view, advertising hopefully creates an impulse leading to an upward jump in the intensity process. As is well known, in practice, such effects are followed by certain adstock phenomena, namely, that customers tend to forget about advertising when time passes by. It is then of interest to know how these partial effects enter into the overall Λ and how repeated advertising may overcome the adstock effect. From our experience in market research, any kind of proportionality of λ_t across individuals is too restrictive to explain the behavior of different customers.

In a medical context λ_t is considered the relevant statistical parameter, describing the risk status of a patient at time t . Hence most models and technical approaches in survival analysis are based on hazard terms. It is only understandable that every patient undergoing a treatment expects some relief. In statistical terms this means that the after-treatment effect results in a downward jump of λ_t , possibly delayed.

Since our Theorems 1 and 2 are of a general nature, they can be applied whenever modeling takes place through hazards. Of course, the model-building process described in Section 3 needs adjustment in a new context, after consulting experts in the relevant area.

In such a complicated situation there is little reason to assume that λ_t is multiplicative or that the driving processes are stationary. Worse than that, Λ is

not predictable (or even adapted) w.r.t. the filtration $\sigma(N_s : s \leq t)$, but should respect external effects. Generally, the model is not dominated so that likelihood methods are difficult to apply.

Summarizing, to obtain acceptable models for practical point processes, one often needs to accept intensity processes that combine the internal history of the process with external shocks or impulses to the effect that the model is no longer dominated and straightforward likelihood methods don't exist; allow for jumps which typically are followed by special patterns as, e.g., shot noise effects; have relevant filtration strictly larger than the internal history of the process.

Let $\mathcal{F}_t, t \in I$, denote the filtration containing the information up to time t . The notion of "self-exciting" is used in a general way to describe any filtered adapted point process N . In particular, in most cases of interest, \mathcal{F}_t strictly includes $\sigma(N_s : s \leq t)$. Also, no (local) stationarity, Markov-property, or independence of increments is assumed. Of course, all this comes with a price, since the required information now comes from independent replications of N . In market research there are different households, while in a medical context this requires several patients in the panel. In modern statistical language, our approach constitutes a functional data analysis where each (random) function comes from a point process with a possibly complicated dynamics.

As a final comment, the notation λ_t only expresses the dependence on time. In our approach, λ_t may also depend on previous values $N(s), s < t$, or external measurements taken before t . Typically, this part has a nonparametric flavor when one wants to avoid restrictive distributional assumptions. In addition, there may be unknown parameters connecting the nonparametric input processes. As a result the parameters should be accessible to the practitioner to aid the understanding of the effects of and the interplay between the individual components.

Here λ_t resp. Λ_t constitutes a flexible semi-parametric model containing an unknown multivariate parameter ϑ . It is our aim to provide a methodology for estimating the parameter of interest when complicated input processes are present. Since dominance cannot be guaranteed, we do not dwell on likelihood but hold to the fact that $N_t - \Lambda_t$ is a (local) martingale.

Remark 1. The main focus is on statistical inference of the parametric part in i.i.d. copies of a point process with complicated dynamics. There is a rich literature on estimating parameters when only one point process is available. If $I = [0, T]$ is the observation period, increasing information comes in from letting T go to infinity. Proposition 3.23 in Karr (1986) is a prototype of such results. It yields asymptotic normality for the intensity in a simple Poisson Process, as $T \rightarrow \infty$. Important extensions are, e.g., due to Ogata (1978), Rathbun (1996), Schoenberg (2005) and Waagepetersen and Guan (2009), who extended

the Poisson case to spatio-temporal processes satisfying some stationarity and mixing conditions. For many applications T is fixed so that i.i.d. copies of N need to be sampled. The advantage here is that complicated dynamics can be handled without requiring strong distributional assumptions.

2. Main Results

Let $N = N(t), t \in I$, be a counting process over a compact interval $I = [t, \bar{t}]$. For each $t \in I$, let \mathcal{F}_t be the σ -field of events observable prior to t . Since we assume that N is observable we have $\sigma(N(s) : s \leq t) \subset \mathcal{F}_t$. In most cases the inclusion is strict.

$$\text{Set } M_t = N_t - \Lambda_t, \quad t \in I.$$

The process Λ_t serves as a predictor process for N_t . In particular, if N_t is integrable, $\mathbb{E}N_t = \mathbb{E}\Lambda_t$. From time to time, we assume that $d\Lambda_t = \lambda_t dt$. Here $\lambda_t, t \in I$, is a predictable stochastic process, the hazard or intensity process, with possible jumps. The λ -process takes its values in the space of left-hand continuous functions with right hand limits. This is the predictable analog of the better known Skorokhod space $D(I)$. See Billingsley (1968).

Let $\vartheta \in \Theta \subset \mathbb{R}^d$ denote the parameter of interest, and let

$$\mathcal{M} = \{\Lambda_\vartheta : \vartheta \in \Theta\}$$

be the associated model for the cumulative hazard process, the dependence on the nonparametric part being suppressed throughout this section. We let ϑ_0 be the true parameter, $\Lambda = \Lambda_{\vartheta_0}$ for some $\vartheta_0 \in \Theta$. Hence the associated innovation martingale is $M = N - \Lambda_{\vartheta_0}$. We provide some methodology on how to estimate ϑ_0 . This estimate takes into account n independent replicates of N , say N_1, \dots, N_n . If each N_i is a simple Single Event process, our results take on a special form for empirical distributions. For general self-exciting processes our approach yields a contribution to the analysis of dynamic functional (point process) data.

For further motivation, let $\Lambda_{\vartheta,i}$, for $1 \leq i \leq n$ and $\vartheta \in \Theta$, be the individual cumulative hazard processes. The unknown parameter $\vartheta_0 \in \Theta$ is the same for all $1 \leq i \leq n$ and hence is intended to describe effects that are equal among the group. The remaining components are individual, usually random, but their distributional properties remain unspecified and perhaps complicated. To exploit the martingale structure, we apply a minimum distance procedure that yields robust consistent estimates of ϑ_0 and does not require additional distributional assumptions.

We formulate our main result. Let N_1, \dots, N_n be i.i.d. copies of N observed over I . For each $1 \leq i \leq n$, let $\mathcal{F}_i(t), t \in I$, be an increasing filtration comprising

the relevant information about N_i . Let $\Lambda_{\vartheta,i}$ with $\vartheta \in \Theta \subset \mathbb{R}^d$ be a given class of semiparametric cumulative intensities such that the true Λ_i of N_i satisfies $\Lambda_i = \Lambda_{\vartheta_0,i}$ for some $\vartheta_0 \in \Theta$. Let μ be a finite measure on I . If f and g are square integrable functions w.r.t. μ , we set

$$\langle f, g \rangle_\mu = \int_I fg d\mu$$

with corresponding semi-norm

$$\|f\|_\mu = \left[\int_I f^2(t) \mu(dt) \right]^{1/2}.$$

Our first f is the difference between the aggregated point process and the aggregated compensator:

$$\bar{N}_n = \frac{1}{n} \sum_{i=1}^n N_i \quad \bar{\Lambda}_{\vartheta,n} = \frac{1}{n} \sum_{i=1}^n \Lambda_{\vartheta,i},$$

with $n\bar{N}_n$ the point process obtained by pooling the points in the individual processes. The associated innovation martingale \bar{M}_n is

$$d\bar{M}_n = d\bar{N}_n - d\bar{\Lambda}_{\vartheta_0,n}.$$

If, for μ , we take $\mu = \bar{N}_n$, the quantity $\|\bar{N}_n - \bar{\Lambda}_{\vartheta,n}\|_{\bar{N}_n}$ represents an overall measure of fit of $\bar{\Lambda}_{\vartheta,n}$ to \bar{N}_n . Our final estimator of ϑ_0 is

$$\vartheta_n = \arg \inf_{\vartheta \in \Theta} \|\bar{N}_n - \bar{\Lambda}_{\vartheta,n}\|_{\bar{N}_n}.$$

Other measures and cumulative functions that appear later are Λ_ϑ and $\mathbb{E}\Lambda_\vartheta$. Throughout, we assume that

$$\mathbb{E}N(\bar{t}) < \infty \text{ and } \mathbb{E}\Lambda_\vartheta(\bar{t}) < \infty \text{ for each } \vartheta \in \Theta.$$

Under weak identifiability and smoothness conditions, ϑ_n is a strongly consistent estimator of ϑ_0 .

Theorem 1. *Let $\Theta \subset \mathbb{R}^d$ be a bounded open set and suppose that, for each $\varepsilon > 0$,*

$$\inf_{\|\vartheta - \vartheta_0\| \geq \varepsilon} \|\mathbb{E}\Lambda_{\vartheta_0} - \mathbb{E}\Lambda_\vartheta\|_{\mathbb{E}\Lambda_{\vartheta_0}} > 0. \tag{2.1}$$

$$\text{The process } (t, \vartheta) \rightarrow \Lambda_\vartheta(t) \text{ is continuous with probability one} \tag{2.2}$$

and admits a continuous extension to $I \times \Theta^c$, where Θ^c is the closure of Θ . Then

$$\lim_{n \rightarrow \infty} \vartheta_n = \vartheta_0 \text{ with probability one.}$$

Remark 2. Condition (2.1) is a weak identifiability condition. Since $\vartheta \rightarrow \mathbb{E}\Lambda_\vartheta$ admits a continuous extension to Θ^c , the infimum in (2.1) is attained, and (2.1) is equivalent to $\mathbb{E}\Lambda_{\vartheta_0}(t) \neq \mathbb{E}\Lambda_\vartheta(t)$ for all $\vartheta \neq \vartheta_0$ and $t \in I_\vartheta$, where $I_\vartheta \subset I$ is a set of t 's with $\mu(I_\vartheta) > 0$. Here μ is the measure with cumulative function $\mathbb{E}\Lambda_{\vartheta_0}$.

As to (2.2), in our applications Λ_ϑ has a (random) Lebesgue density λ_ϑ with values in an appropriate Skorokhod space. This guarantees continuity (but not differentiability) of Λ_ϑ in t and allows for unexpected jumps in the intensity function λ_ϑ as well.

The boundedness of Θ is assumed only for convenience; Theorem 1 also holds if Λ has a continuous extension to the compactification of Θ in the extended Euclidean space.

Remark 3. In the context of Theorem 1, (2.1) is only a technical condition. We can get an extension to the case in which the minimizer ϑ_0 is not unique. In such a situation let Θ_n be the set of estimators. With the Hausdorff distance $d(\Theta_0, \Theta_n)$ of Θ_0 and Θ_n , it can then be shown that with probability one, $d(\Theta_0, \Theta_n) \rightarrow 0$. In the numerical examples of Section 3, ϑ_n was always unique.

For a second result, we assume that $\vartheta \rightarrow \Lambda_\vartheta(t)$ is a twice continuously differentiable function in a neighborhood of ϑ_0 ; first and second order derivatives explicitly appear in the distributional approximation of ϑ_n . Let

$$\Phi_0(\vartheta) = \frac{\partial}{\partial \vartheta} \int_I (\mathbb{E}\Lambda_\vartheta(t) - \mathbb{E}\Lambda_{\vartheta_0}(t)) \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t)^T \mathbb{E}\Lambda_{\vartheta_0}(dt),$$

a matrix-valued function, where T denotes transposition. Under regularity conditions (see (2.4)) we may interchange differentiation and integration to come up with

$$\begin{aligned} \Phi_0(\vartheta) &= \int_I \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t)^T \mathbb{E}\Lambda_{\vartheta_0}(dt) \\ &\quad + \int_I (\mathbb{E}\Lambda_\vartheta(t) - \mathbb{E}\Lambda_{\vartheta_0}(t)) \mathbb{E} \frac{\partial^2}{\partial \vartheta^2} \Lambda_\vartheta(t)^T \mathbb{E}\Lambda_{\vartheta_0}(dt). \end{aligned}$$

At $\vartheta = \vartheta_0$, the second integral vanishes so that

$$\Phi_0(\vartheta_0) = \int_I \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t)^T \mathbb{E}\Lambda_{\vartheta_0}(dt) \Big|_{\vartheta=\vartheta_0}. \quad (2.3)$$

The integrand in (2.3) is nonnegative definite for each t . So, on a set of positive $\mathbb{E}\Lambda_{\vartheta_0}$ measure, $\Phi_0(\vartheta_0)$ is positive definite.

Theorem 2. *Suppose (2.1) and (2.2) hold, Furthermore, assume that*

$$\left\| \frac{\partial}{\partial \vartheta} (\mathbb{E}\Lambda_{\vartheta}(t) - \mathbb{E}\Lambda_{\vartheta_0}(t)) \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t)^T \right\| \leq C(t) \tag{2.4}$$

for all ϑ in a neighborhood of ϑ_0 , where the majorant function C is integrable w.r.t. $\mathbb{E}\Lambda_{\vartheta_0}$, and

$$\varphi(x) = \int_{[x, \bar{t}]} \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \mathbb{E}\Lambda_{\vartheta_0}(dt) \Big|_{\vartheta=\vartheta_0}, \underline{t} \leq x \leq \bar{t} \tag{2.5}$$

is square integrable w.r.t. $\mathbb{E}\Lambda_{\vartheta_0}$. Then

$$\begin{aligned} n^{1/2}\Phi_0(\vartheta_0)(\vartheta_n - \vartheta_0) &= n^{1/2} \int_I \int_{[x, \bar{t}]} \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \mathbb{E}\Lambda_{\vartheta_0}(dt) \bar{M}_n(dx) \Big|_{\vartheta=\vartheta_0} + o_{\mathbb{P}}(1) \\ &\equiv n^{1/2} \int_I \varphi(x) \bar{M}_n(dx) + o_{\mathbb{P}}(1). \end{aligned} \tag{2.6}$$

Under (2.5), the integral in (2.6) is a sum of centered i.i.d. square-integrable random vectors to which the Central Limit Theorem can be applied.

Corollary 1. *Under the assumptions of Theorem 2, as $n \rightarrow \infty$,*

$$n^{1/2}\Phi_0(\vartheta_0)(\vartheta_n - \vartheta_0) \rightarrow \mathcal{N}_d(0, C(\vartheta_0)) \quad \text{in distribution,}$$

where $C(\vartheta_0)$ is a $d \times d$ matrix with entries

$$C_{ij}(\vartheta_0) = \int_I \varphi_i(x) \varphi_j(x) \mathbb{E}\Lambda_{\vartheta_0}(dx)$$

and $\varphi = (\varphi_1, \dots, \varphi_d)^T$.

The matrix $\Phi_0(\vartheta_0)$ is nonnegative definite; if positive definite, as $n \rightarrow \infty$,

$$n^{1/2}(\vartheta_n - \vartheta_0) \rightarrow \mathcal{N}_d(0, \Sigma) \quad \text{in distribution,}$$

where $\Sigma = \Phi_0^{-1}(\vartheta_0)C(\vartheta_0)\Phi_0^{-1}(\vartheta_0)$.

Remark 4. The i.i.d. representation (2.6) is useful when we plan to derive some goodness-of-fit tests of $H_0 : \Lambda_0 \in \mathcal{M}$, where \mathcal{M} is a semiparametric model for the true Λ_0 . Since this is beyond the scope of the paper, we only sketch the arguments. The basic test process is $t \rightarrow n^{1/2}[\bar{N}_t - \hat{\Lambda}_t]$, where $\hat{\Lambda}$ is taken from \mathcal{M} with estimated parameter ϑ_n . $\bar{N} - \hat{\Lambda}$ is often called the martingale residual process though estimation of parameters usually destroys the martingale property. Under H_0 , this process has the expansion

$$n^{1/2}[\bar{N} - \Lambda_0] - n^{1/2}[\hat{\Lambda} - \Lambda_0]. \tag{2.7}$$

The first part of (2.7) is \bar{M} (standardized), under H_0 , while under the smoothness assumptions of Theorem 2 and because of (2.6), Taylor-expansion yields an i.i.d. representation of the second part of (2.7). For aggregated Single Event processes, this decomposition was studied by Durbin (1973), and for regression by Stute (1997). Estimation of unknown parameters typically changes the distributional character of the test process. Stute, Thies, and Zhu (1998) discusses how in the regression case a martingale transformation can be applied to come up with some asymptotically distribution-free tests. For model checks of time series, see Koul and Stute (1999). In principle, this may be also done in the context of this paper; this will be studied in detail in future work.

3. A Data Example

In this section we apply our methodology to analyze the dynamic behavior of customers over time. We are interested in the impact of TV-advertising and so-called “Adstock-effects”. In Kopperschmidt and Stute (2009) we gave a review of the most prominent timing models in market research. These models are unable to provide a satisfactory dynamic framework for advertising effects.

We start with the modeling of λ_t in a seasonal market. The product of interest was a premium brand of packaged ice cream on the German market. Our λ_t acknowledges seasonality effects, the internal purchase history including adstock effects, and advertising effects. The seasonality effect is assumed to be the same for all households. It serves as a baseline and depends on parameters α, β, γ , and δ through

$$\lambda_1(t) = \alpha \sin(\beta t + \gamma) + \delta.$$

Here δ is a basic consumption rate which is the same throughout the year. The rest is a properly shifted sinus curve assumed to have a peak in the months June-September. We expect that, due to seasonal effects, $\delta > \alpha > 0, \beta \sim 2\pi/365$ and $\gamma \sim \pi/2$.

For household i , we denote with $Y_{i1} < \dots < Y_{ik_i}$ the ordered purchase times. For some households $k_i = 0$. The quantity $Y_{iN_i(t-)}$ is the time of the last purchase before t , hence $t - Y_{iN_i(t-)}$ is the “age” of the system. We suppose that the intensity increases as time passes and the consumer’s stock of ice cream diminishes. The “speed” for recovery of the purchase inclination is controlled by a parameter $\varepsilon > 0$ that enters into some $\lambda_2^i(t)$ as

$$\lambda_2^i(t) = \left(1 - e^{-\varepsilon(t - Y_{iN_i(t-)})} \mathbf{1}_{\{t > Y_{i1}\}}\right).$$

Our final component is

$$\lambda_3^i(t) = \xi \sum_{h=1}^{W_i(t)} e^{\eta(t - X_{ih})}.$$

Here $W_i(t)$ is the number of advertising contacts observed by household i up to time t . The advertising times are $X_{i1} < X_{i2} < \dots$ hence $t - X_{ih}$ is the time elapsed between t and the h th advertising contact (before t). The parameter η is the so-called adstock parameter; if $\eta < 0$ is small, the impact of advertising rapidly decreases as time passes by. The parameter ξ measures the mean impulse of an advertising contact at the moment the message is received. At each X_{ih} there is a jump of size ξ followed by a decay. Hence $\lambda_3^i(t)$ is a superposition of $W_i(t)$ of these effects, and

$$\lambda_{\vartheta,i} = \lambda_1 \lambda_2^i + \lambda_3^i.$$

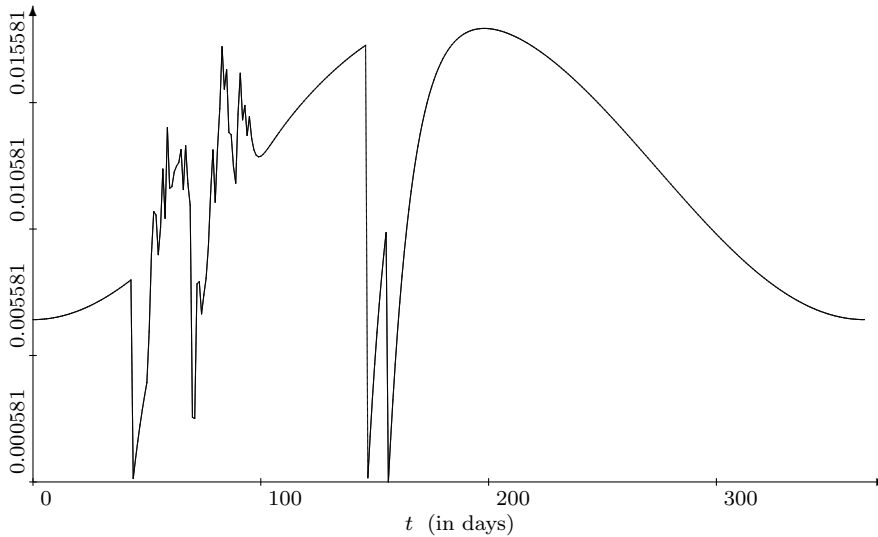
We used a multiplicative-additive version for $\lambda_{\vartheta,i}$ so that the advertising effect could be better isolated from the rest. For the parameter ϑ , we of course have $\vartheta = (\alpha, \beta, \gamma, \delta, \varepsilon, \eta, \xi)$.

The data set stems from a “Single Source Panel” of AC Nielsen, Germany. Until December 2006, AC Nielsen equipped several thousand households of the panel with technical devices that allowed for the capture of their purchase and TV behavior. Each household provided information on purchases of arbitrary fast-moving consumer goods on a daily basis. This was achieved by scanning the barcodes at home. In our analysis the amount of money spent was neglected. Each panel household was also provided with a device by which one knew whether and when an advertising spot was watched. This gave us the information contained in W_i and hence in λ_3^i . The statistical analysis was based on 1,660 households, 1375 of which purchased at least once. The households were observed over a period of 546 days, $I = [0, 546]$. Only 27 households did not watch a single TV advertisement for packaged ice cream throughout the 18 months. The whole sample was divided into 11 subgroups of size 100-200 according to sociodemographic factors “income” and “size of household”. We present the results of our analysis for two of these 11 subgroups that differ in household size and monthly income.

Group	1	2
HH Size	4	2
Income (Euro)	< 2,000	≥ 3,000
n	118	180
α_n	0.0017	0.0021
β_n	0.0210	0.0146
γ_n	1.5700	1.5700
δ_n	0.0113	0.0062
ε_n	0.1000	0.6580
ξ_n	0.0039	0.0846
η_n	-0.5000	-0.4999

We see that the plausibility check $\beta \sim 2\pi/365 \sim 0.017$ was fulfilled for both groups. The model also “recognized” the annual seasonality with a seasonal peak in the summer. The δ_n 's being in charge of an off-seasonal demand are small. The most important parameters for the producer of the product were ξ and η . Interestingly enough, the adstock parameter η_n was always close to -0.5 . The parameter ξ_n describing the strength of the advertising act differed among the groups. Typically households with children had a smaller reaction parameter ξ ($\xi_n = 0.0039$) compared with small but richer households ($\xi_n = 0.0846$). This may be due to the fact that the product of interest was a premium brand so that families with children, though attracted by the advertisement, showed a tendency to buy cheaper products from a discounter.

We show a typical graph of the purchase-intensity of a household who is subject to manipulation. We see four larger downward jumps indicating purchases which result, on the short run, in saturation effects. Smaller downward jumps are adstock effects appearing after several upward jumps occurring after advertising.



4. Proofs

For $\varepsilon > 0$ we let

$$\mathcal{B}_\varepsilon(\vartheta_0) = \{\vartheta : \|\vartheta - \vartheta_0\| < \varepsilon\}$$

be the open ε -ball with center ϑ_0 ; for $\varepsilon > 0$ and $r > 0$, let

$$\mathcal{B}_r^\varepsilon(\vartheta) = \{\vartheta' \in \Theta^c \setminus \mathcal{B}_\varepsilon(\vartheta_0) : \|\vartheta' - \vartheta\| < r\}$$

be the part of the r -ball with center ϑ , that does not belong to $\mathcal{B}_\varepsilon(\vartheta_0)$. Let $\vartheta \in \Theta$ be a given, fixed parameter.

Lemma 1. *Under the assumptions of Theorem 1, with probability one,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sup_{\vartheta' \in \mathcal{B}_r^\varepsilon(\vartheta)} \|N_i - \Lambda_{\vartheta',i}\|_{N_i}^2 = \mathbb{E} \left\{ \sup_{\vartheta' \in \mathcal{B}_r^\varepsilon(\vartheta)} \|N - \Lambda_{\vartheta'}\|_N^2 \right\}, \quad (4.1)$$

$$\lim_{n \rightarrow \infty} \frac{1}{n(n-1)} \sum_{i \neq j} \sup_{\vartheta' \in \mathcal{B}_r^\varepsilon(\vartheta)} \|N_i - \Lambda_{\vartheta',i}\|_{N_j}^2 = \mathbb{E} \left\{ \sup_{\vartheta' \in \mathcal{B}_r^\varepsilon(\vartheta)} \|N_1 - \Lambda_{\vartheta',1}\|_{N_2}^2 \right\}, \quad (4.2)$$

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n(n-1)} \sum_{i \neq j} \sup_{\vartheta' \in \mathcal{B}_r^\varepsilon(\vartheta)} \langle N_i - \Lambda_{\vartheta',i}, N_j - \Lambda_{\vartheta',j} \rangle_{N_i} \\ &= \mathbb{E} \left\{ \sup_{\vartheta' \in \mathcal{B}_r^\varepsilon(\vartheta)} \langle N_1 - \Lambda_{\vartheta',1}, N_2 - \Lambda_{\vartheta',2} \rangle_{N_1} \right\}, \end{aligned} \quad (4.3)$$

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n(n-1)(n-2)} \sum_{i \neq j \neq k} \sup_{\vartheta' \in \mathcal{B}_r^\varepsilon(\vartheta)} \langle N_i - \Lambda_{\vartheta',i}, N_j - \Lambda_{\vartheta',j} \rangle_{N_k} \\ &= \mathbb{E} \left\{ \sup_{\vartheta' \in \mathcal{B}_r^\varepsilon(\vartheta)} \langle N_1 - \Lambda_{\vartheta',1}, N_2 - \Lambda_{\vartheta',2} \rangle_{N_3} \right\}. \end{aligned} \quad (4.4)$$

Proof. The four statements are immediate consequences of (5.1), our strong law for U -statistics of point processes.

If, in Lemma 1, we replace the sup over $\vartheta' \in \mathcal{B}_r^\varepsilon(\vartheta)$ by ϑ , we obtain the limit results as follows. Proofs again are based on (5.1).

Lemma 2. *With probability one,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \|N_i - \Lambda_{\vartheta,i}\|_{N_i}^2 = \mathbb{E} \left\{ \|N - \Lambda_{\vartheta}\|_N^2 \right\}, \quad (4.5)$$

$$\lim_{n \rightarrow \infty} \frac{1}{n(n-1)} \sum_{i \neq j} \|N_i - \Lambda_{\vartheta,i}\|_{N_j}^2 = \mathbb{E} \left\{ \|N_1 - \Lambda_{\vartheta,1}\|_{N_2}^2 \right\}, \quad (4.6)$$

$$\lim_{n \rightarrow \infty} \frac{1}{n(n-1)} \sum_{i \neq j} \langle N_i - \Lambda_{\vartheta,i}, N_j - \Lambda_{\vartheta,j} \rangle_{N_i} = \mathbb{E} \left\{ \langle N_1 - \Lambda_{\vartheta,1}, N_2 - \Lambda_{\vartheta,2} \rangle_{N_1} \right\}, \quad (4.7)$$

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n(n-1)(n-2)} \sum_{i \neq j \neq k} \langle N_i - \Lambda_{\vartheta,i}, N_j - \Lambda_{\vartheta,j} \rangle_{N_k} \\ &= \mathbb{E} \left\{ \langle N_1 - \Lambda_{\vartheta,1}, N_2 - \Lambda_{\vartheta,2} \rangle_{N_3} \right\}. \end{aligned} \quad (4.8)$$

Lemma 3. *We have*

$$\mathbb{E} \left\{ \langle N_1 - \Lambda_{\vartheta,1}, N_2 - \Lambda_{\vartheta,2} \rangle_{N_3} \right\} = \|\mathbb{E}\Lambda_{\vartheta_0} - \mathbb{E}\Lambda_{\vartheta}\|_{\mathbb{E}\Lambda_{\vartheta_0}}^2. \quad (4.9)$$

Proof. The expectation of the left-hand side is

$$\mathbb{E} \left[\int_I (N_1(t) - \Lambda_{\vartheta,1}(t)) (N_2(t) - \Lambda_{\vartheta,2}(t)) N_3(dt) \right].$$

Conditionally on N_3 , the expectation is, by independence of $N_1 - \Lambda_{\vartheta,1}$ and $N_2 - \Lambda_{\vartheta,2}$, and by Fubini's Theorem,

$$\int_I \mathbb{E}^2 [N_1(t) - \Lambda_{\vartheta,1}(t)] N_3(dt) = \int_I [\mathbb{E}\Lambda_{\vartheta_0}(t) - \mathbb{E}\Lambda_{\vartheta}(t)]^2 N_3(dt).$$

Integrate out to get the result.

We next study local fluctuations of the process

$$\vartheta \rightarrow \int_I (N_1(t) - \Lambda_{\vartheta,1}(t)) (N_2(t) - \Lambda_{\vartheta,2}(t)) N_3(dt). \tag{4.10}$$

Lemma 4. *For given $\varepsilon > 0$ and $\delta > 0$, and for each $\vartheta \in \Theta^c$, there is a small but positive $r > 0$ such that*

$$\left| \mathbb{E} \left\{ \sup_{\vartheta' \in \mathcal{B}_r^\varepsilon(\vartheta)} \langle N_1 - \Lambda_{\vartheta',1}, N_2 - \Lambda_{\vartheta',2} \rangle_{N_3} \right\} - \inf_{\vartheta' \in \mathcal{B}_r^\varepsilon(\vartheta)} \|\mathbb{E}\Lambda_{\vartheta_0} - \mathbb{E}\Lambda_{\vartheta'}\|_{\mathbb{E}\Lambda_{\vartheta_0}}^2 \right| \leq \delta, \tag{4.11}$$

$$\left| \mathbb{E} \left\{ \inf_{\vartheta' \in \mathcal{B}_r^\varepsilon(\vartheta)} \langle N_1 - \Lambda_{\vartheta',1}, N_2 - \Lambda_{\vartheta',2} \rangle_{N_3} \right\} - \|\mathbb{E}\Lambda_{\vartheta_0} - \mathbb{E}\Lambda_{\vartheta}\|_{\mathbb{E}\Lambda_{\vartheta_0}}^2 \right| \leq \delta. \tag{4.12}$$

Proof. By assumption, the process (4.10) is continuous in ϑ . Hence

$$\sup_{\vartheta' \in \mathcal{B}_r^\varepsilon(\vartheta)} \langle N_1 - \Lambda_{\vartheta',1}, N_2 - \Lambda_{\vartheta',2} \rangle_{N_3} \downarrow \langle N_1 - \Lambda_{\vartheta,1}, N_2 - \Lambda_{\vartheta,2} \rangle_{N_3}$$

as $r \downarrow 0$. By monotone convergence, the expectation of the left-hand side converges to

$$\mathbb{E} \{ \langle N_1 - \Lambda_{\vartheta,1}, N_2 - \Lambda_{\vartheta,2} \rangle_{N_3} \} = \|\mathbb{E}\Lambda_{\vartheta_0} - \mathbb{E}\Lambda_{\vartheta}\|_{\mathbb{E}\Lambda_{\vartheta_0}}^2,$$

by Lemma 3. Similarly,

$$\inf_{\vartheta' \in \mathcal{B}_r^\varepsilon(\vartheta)} \|\mathbb{E}\Lambda_{\vartheta_0} - \mathbb{E}\Lambda_{\vartheta'}\|_{\mathbb{E}\Lambda_{\vartheta_0}}^2 \uparrow \|\mathbb{E}\Lambda_{\vartheta_0} - \mathbb{E}\Lambda_{\vartheta}\|_{\mathbb{E}\Lambda_{\vartheta_0}}^2$$

as $r \downarrow 0$. This proves the first part of the lemma. The second part is shown in a similar way.

Lemma 5. *For each fixed $\vartheta \in \Theta$, we have*

$$\|\bar{N}_n - \bar{\Lambda}_{\vartheta,n}\|_{\bar{N}_n}^2 \rightarrow \|\mathbb{E}\Lambda_{\vartheta_0} - \mathbb{E}\Lambda_{\vartheta}\|_{\mathbb{E}\Lambda_{\vartheta_0}}^2 \text{ with probability one.}$$

In particular, for $\vartheta = \vartheta_0$, the limit equals zero.

Proof. We have

$$\begin{aligned} \|\bar{N}_n - \bar{\Lambda}_{\vartheta,n}\|_{\bar{N}_n}^2 &= \int_I [\bar{N}_n(t) - \bar{\Lambda}_{\vartheta,n}(t)]^2 \bar{N}_n(dt) \\ &= n^{-3} \sum_{i,j,k=1}^n \langle N_i - \Lambda_{\vartheta,i}, N_j - \Lambda_{\vartheta,j} \rangle_{N_k}. \end{aligned}$$

By (4.5) – (4.7), the last triple sum is dominated by the sum restricted to $i \neq j \neq k \neq i$. The limit follows from (5.1), (4.8), and Lemma 3.

Lemma 6. *For each $\varepsilon > 0$ with probability one,*

$$\inf_{\vartheta: \|\vartheta - \vartheta_0\| \geq \varepsilon} \|\bar{N}_n - \bar{\Lambda}_{\vartheta,n}\|_{\bar{N}_n} \rightarrow \inf_{\vartheta: \|\vartheta - \vartheta_0\| \geq \varepsilon} \|\mathbb{E}\Lambda_{\vartheta_0} - \mathbb{E}\Lambda_{\vartheta}\|_{\mathbb{E}\Lambda_{\vartheta_0}}.$$

Proof. For a given $\varepsilon > 0$, let $\delta > 0$ be an arbitrary positive number. According to Lemma 4, there exists, for each $\vartheta \in \Theta^c$, a positive radius $r = r(\vartheta) > 0$ such that (4.11) and (4.12) are satisfied. Then

$$\bigcup_{\vartheta \in \Theta^c \setminus \mathcal{B}_{\varepsilon}(\vartheta_0)} \mathcal{B}_{r(\vartheta)}^{\varepsilon}(\vartheta) = \Theta^c \setminus \mathcal{B}_{\varepsilon}(\vartheta_0)$$

is an open covering of the compact set $\Theta^c \setminus \mathcal{B}_{\varepsilon}(\vartheta_0)$. We can select finitely many $\vartheta_1, \dots, \vartheta_q$ so that $\Theta^c \setminus \mathcal{B}_{\varepsilon}(\vartheta_0)$ is already covered by the $\mathcal{B}_{r_l}^{\varepsilon}(\vartheta_l)$, where $r_l = r(\vartheta_l)$. Conclude that

$$\begin{aligned} & \left| \inf_{\vartheta: \|\vartheta - \vartheta_0\| \geq \varepsilon} \|\bar{N}_n - \bar{\Lambda}_{\vartheta,n}\|_{\bar{N}_n}^2 - \inf_{\vartheta: \|\vartheta - \vartheta_0\| \geq \varepsilon} \|\mathbb{E}\Lambda_{\vartheta_0} - \mathbb{E}\Lambda_{\vartheta}\|_{\mathbb{E}\Lambda_{\vartheta_0}}^2 \right| \\ &= \left| \min_{1 \leq l \leq q} \inf_{\vartheta \in \mathcal{B}_{r_l}^{\varepsilon}(\vartheta_l)} \|\bar{N}_n - \bar{\Lambda}_{\vartheta,n}\|_{\bar{N}_n}^2 - \min_{1 \leq l \leq q} \inf_{\vartheta \in \mathcal{B}_{r_l}^{\varepsilon}(\vartheta_l)} \|\mathbb{E}\Lambda_{\vartheta_0} - \mathbb{E}\Lambda_{\vartheta}\|_{\mathbb{E}\Lambda_{\vartheta_0}}^2 \right| \\ &\leq \max_{1 \leq l \leq q} \left| \inf_{\vartheta \in \mathcal{B}_{r_l}^{\varepsilon}(\vartheta_l)} \|\bar{N}_n - \bar{\Lambda}_{\vartheta,n}\|_{\bar{N}_n}^2 - \inf_{\vartheta \in \mathcal{B}_{r_l}^{\varepsilon}(\vartheta_l)} \|\mathbb{E}\Lambda_{\vartheta_0} - \mathbb{E}\Lambda_{\vartheta}\|_{\mathbb{E}\Lambda_{\vartheta_0}}^2 \right|. \end{aligned}$$

It suffices to prove that for each $1 \leq l \leq q$, with probability one, the limsup of the term in brackets is less than or equal to δ in absolute value. But

$$\inf_{\vartheta \in \mathcal{B}_{r_l}^{\varepsilon}(\vartheta_l)} \|\bar{N}_n - \bar{\Lambda}_{\vartheta,n}\|_{\bar{N}_n}^2 \leq n^{-3} \sum_{i,j,k} \sup_{\vartheta \in \mathcal{B}_{r_l}^{\varepsilon}(\vartheta_l)} \langle N_i - \Lambda_{\vartheta,i}, N_j - \Lambda_{\vartheta,j} \rangle_{N_k},$$

which again is dominated by the sub-sum pertaining to $i \neq j \neq k \neq i$. With Lemma 1 this converges to

$$\mathbb{E} \left\{ \sup_{\vartheta \in \mathcal{B}_{r_l}^{\varepsilon}(\vartheta_l)} \langle N_1 - \Lambda_{\vartheta,1}, N_2 - \Lambda_{\vartheta,2} \rangle_{N_3} \right\}.$$

By Lemma 4 and the choice of r_l , this term is within δ -distance of $\inf_{\vartheta \in \mathcal{B}_{r_l}^{\varepsilon}(\vartheta_l)} \|\mathbb{E}\Lambda_{\vartheta_0} - \mathbb{E}\Lambda_{\vartheta}\|_{\mathbb{E}\Lambda_{\vartheta_0}}^2$. The corresponding lower bound may be obtained in a similar way. This concludes the proof of the lemma.

Proof of Theorem 1. Let $\varepsilon > 0$ be given. We need to show that

$$\mathbb{P} \left(\limsup_{n \rightarrow \infty} \{\|\vartheta_n - \vartheta_0\| \geq \varepsilon\} \right) = 0.$$

But

$$\begin{aligned} \{\|\vartheta_n - \vartheta_0\| \geq \varepsilon\} &\subset \left\{ \inf_{\vartheta: \|\vartheta - \vartheta_0\| \geq \varepsilon} \|\bar{N}_n - \bar{\Lambda}_{\vartheta,n}\|_{\bar{N}_n} < \inf_{\vartheta: \|\vartheta - \vartheta_0\| < \varepsilon} \|\bar{N}_n - \bar{\Lambda}_{\vartheta,n}\|_{\bar{N}_n} \right\} \\ &\subset \left\{ \inf_{\vartheta: \|\vartheta - \vartheta_0\| \geq \varepsilon} \|\bar{N}_n - \bar{\Lambda}_{\vartheta,n}\|_{\bar{N}_n} < \|\bar{N}_n - \bar{\Lambda}_{\vartheta_0,n}\|_{\bar{N}_n} \right\}. \end{aligned}$$

From Lemma 6, the left-hand side goes to

$$\inf_{\vartheta: \|\vartheta - \vartheta_0\| \geq \varepsilon} \|\mathbb{E}\Lambda_{\vartheta_0} - \mathbb{E}\Lambda_{\vartheta}\|_{\mathbb{E}\Lambda_{\vartheta_0}},$$

which is positive by (2.1). At the same time, the right-hand side tends to $\|\mathbb{E}\Lambda_{\vartheta_0} - \mathbb{E}\Lambda_{\vartheta_0}\|_{\mathbb{E}\Lambda_{\vartheta_0}} = 0$. This proves Theorem 1.

The following lemma expresses certain point process integrals in terms of the associated innovation martingale \bar{M}_n .

Lemma 7. *We have*

$$\begin{aligned} &\int_I [\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_0,n}(t)] \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{N}_n(dt) \Big|_{\vartheta=\vartheta_n} \\ &= \int_I \bar{M}_n(t) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{M}_n(dt) \Big|_{\vartheta=\vartheta_n} + \int_I \bar{M}_n(t) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{\Lambda}_{\vartheta_0,n}(dt) \Big|_{\vartheta=\vartheta_n}. \end{aligned} \tag{4.13}$$

Proof. This is an immediate consequence of the identity

$$d\bar{N}_n = d\bar{M}_n + d\bar{\Lambda}_{\vartheta_0,n}, \tag{4.14}$$

and the fact that ϑ_n minimizes $\|\bar{N}_n - \bar{\Lambda}_{\vartheta,n}\|_{\bar{N}_n}^2$ in ϑ and is in the inner open set of Θ , whence

$$\int_I [\bar{N}_n(t) - \bar{\Lambda}_{\vartheta,n}(t)] \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{N}_n(dt) = 0 \text{ at } \vartheta = \vartheta_n.$$

If we multiply (4.13) by $n^{1/2}$ and replace \bar{N}_n by \bar{M}_n , we have with the parametric process

$$\alpha_n(\vartheta) = n^{1/2} \int_I [\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_0,n}(t)] \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{M}_n(dt),$$

where $\alpha_n(\vartheta_0) = 0$. With

$$\beta_n(\vartheta) = n^{1/2} \int_I \bar{M}_n(t) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{M}_n(dt),$$

$$\gamma_n(\vartheta) = n^{1/2} \int_I \bar{M}_n(t) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{\Lambda}_{\vartheta_0,n}(dt),$$

then, after multiplication with $n^{1/2}$, the assertion of Lemma 7 is

$$\begin{aligned} \alpha_n(\vartheta_n) + n^{1/2} \int_I [\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_0,n}(t)] \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{\Lambda}_{\vartheta_0,n}(dt) \Big|_{\vartheta=\vartheta_n} \\ = \beta_n(\vartheta_n) + \gamma_n(\vartheta_n). \end{aligned} \tag{4.15}$$

Since, by Theorem 1, $\vartheta_n \rightarrow \vartheta_0$ with probability one, we have $\alpha_n(\vartheta_n) = \alpha_n(\vartheta_0) + o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1)$, once it has been proved that $\{\alpha_n\}$ is uniformly tight in the space $C(\Theta)$, say uniformly C -tight, of all bounded continuous functions on Θ .

For processes depending on univariate parameters, Billingsley (1968) presents a detailed analysis of tightness tools and techniques. Bickel and Wichura (1971) is an important reference for multiparameter processes. Their work is mainly focussed on stochastic processes with paths in Skorokhod spaces, requiring the incorporation of multidimensional increments. Here the processes α_n (and similarly β_n and γ_n) are continuous processes in the parameter ϑ , and simple increments suffice to guarantee tightness. We first study the local behavior of the processes α_n .

Lemma 8. *The process α_n admits the representation*

$$\alpha_n(\vartheta) = n^{1/2} \int_I [\mathbb{E}\Lambda_{\vartheta}(t) - \mathbb{E}\Lambda_{\vartheta_0}(t)] \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \bar{M}_n(dt) + o_{\mathbb{P}}(1),$$

where the remainder converges to zero uniformly on compact subsets of Θ . Furthermore, the leading term is C -tight.

Proof. To prove the lemma, note that

$$\begin{aligned} & [\bar{\Lambda}_{\vartheta,n} - \bar{\Lambda}_{\vartheta_0,n}] \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n} \\ &= [\bar{\Lambda}_{\vartheta,n} - \bar{\Lambda}_{\vartheta_0,n}] \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta} + [\bar{\Lambda}_{\vartheta,n} - \bar{\Lambda}_{\vartheta_0,n}] \left[\frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n} - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta} \right] \\ &= [\mathbb{E}\Lambda_{\vartheta} - \mathbb{E}\Lambda_{\vartheta_0}] \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta} - [\bar{\Lambda}_{\vartheta_0,n} - \mathbb{E}\Lambda_{\vartheta_0}] \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta} \\ &\quad + [\bar{\Lambda}_{\vartheta,n} - \mathbb{E}\Lambda_{\vartheta}] \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta} + [\bar{\Lambda}_{\vartheta,n} - \bar{\Lambda}_{\vartheta_0,n}] \left[\frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n} - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta} \right] \\ &\equiv [\mathbb{E}\Lambda_{\vartheta} - \mathbb{E}\Lambda_{\vartheta_0}] \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta} - r_n^1 + r_n^2 + r_n^3. \end{aligned}$$

Integration w.r.t. \bar{M}_n therefore leads to

$$\alpha_n(\vartheta) = \tilde{\alpha}_n(\vartheta) - \alpha_n^1(\vartheta) + \alpha_n^2(\vartheta) + \alpha_n^3(\vartheta),$$

where

$$\alpha_n^i(\vartheta) = n^{1/2} \int_I r_n^i(t, \vartheta) \bar{M}_n(dt), \quad i = 1, 2, 3,$$

$$\tilde{\alpha}_n(\vartheta) = n^{1/2} \int_I [\mathbb{E}\Lambda_\vartheta - \mathbb{E}\Lambda_{\vartheta_0}] \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta \bar{M}_n(dt).$$

Therefore, the lemma can be proved by showing that $\tilde{\alpha}_n, \alpha_n^1, \alpha_n^2$ and α_n^3 are uniformly C -tight, and that α_n^1, α_n^2 and α_n^3 go to zero for each fixed ϑ .

Lemma 9. *The processes $\{\tilde{\alpha}_n\}$ are uniformly C -tight on compact subsets of Θ containing ϑ_0 .*

Proof. Obviously, $\tilde{\alpha}_n(\vartheta_0) = 0$. In particular, there exists at least one point where $\tilde{\alpha}_n$ is bounded. See Billingsley (1968), Theorem 12.3, for a discussion of processes on the unit interval. Next we show that

$$\mathbb{E}\|\tilde{\alpha}_n(\vartheta) - \tilde{\alpha}_n(\vartheta')\|^2 \leq \text{const}\|\vartheta - \vartheta'\|^2$$

to complete the proof of the lemma. Now,

$$\tilde{\alpha}_n(\vartheta) - \tilde{\alpha}_n(\vartheta') = n^{1/2} \int_I [\mathbb{E}\Lambda_\vartheta - \mathbb{E}\Lambda_{\vartheta'}] \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta d\bar{M}_n \tag{4.16}$$

$$+ n^{1/2} \int_I [\mathbb{E}\Lambda_{\vartheta'} - \mathbb{E}\Lambda_{\vartheta_0}] \left[\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'} \right] d\bar{M}_n. \tag{4.17}$$

Since for any two vectors a and b we have $\|a + b\|^2 \leq 2(\|a\|^2 + \|b\|^2)$, it suffices to study the terms separately. In each case the integrand is a deterministic function in t . Since for functions f the integrals $\int f dM_i$ are independent and centered, we have

$$\mathbb{E}\|n^{1/2} \int_I f d\bar{M}_n\|^2 = \mathbb{E}\| \int_I f(dN - d\Lambda_{\vartheta_0}) \|^2.$$

Furthermore, Λ_{ϑ_0} is the compensator of N so the last expectation becomes

$$\mathbb{E}\left\{ \left\| \int_I f dN \right\|^2 - \left\| \int_I f d\Lambda_{\vartheta_0} \right\|^2 \right\} \leq \mathbb{E}\left\{ \left\| \int_I f dN \right\|^2 \right\} = \int_I f^T f d\mathbb{E}\Lambda_{\vartheta_0}.$$

When we apply this to (4.16) and (4.17), the conclusion follows from assumptions (2.4) - (2.5). This concludes the proof of the lemma.

Lemma 10. *For all $i = 1, 2, 3$ and each $\vartheta \in \Theta$ we have $\alpha_n^i(\vartheta) = o_{\mathbb{P}}(1)$.*

Proof. We have to show that, in probability,

$$n^{1/2} \int_I [\bar{\Lambda}_{\vartheta_0, n} - \mathbb{E}\Lambda_{\vartheta_0}] \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta d\bar{M}_n \rightarrow 0,$$

$$n^{1/2} \int_I [\bar{\Lambda}_{\vartheta, n} - \mathbb{E}\Lambda_\vartheta] \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta d\bar{M}_n \rightarrow 0,$$

$$n^{1/2} \int_I [\bar{\Lambda}_{\vartheta, n} - \bar{\Lambda}_{\vartheta_0, n}] \left[\frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta, n} - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta \right] d\bar{M}_n \rightarrow 0.$$

The two first assertions follow from

$$n^{1/2} \int_I [\bar{\Lambda}_{\vartheta,n} - \mathbb{E}\Lambda_{\vartheta}] \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'} d\bar{M}_n \rightarrow 0 \tag{4.18}$$

for all $\vartheta, \vartheta' \in \Theta$. The integral in (4.18), however, is

$$n^{-3/2} \sum_{i,k=1}^n \int_I [\Lambda_{\vartheta,k} - \mathbb{E}\Lambda_{\vartheta}] \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'} dM_i \equiv n^{-3/2} \sum_{i,k=1}^n U_{ki}.$$

The sum over $i = k$ consists of i.i.d. centered random variables with expectation zero, since each M_i is a martingale and every $\Lambda_{\vartheta,i}$ is continuous in t and hence predictable. The SLLN guarantees convergence to zero. As to the sum over $i \neq k$, we can apply Lemma 16 to get

$$n^{-3} \mathbb{E} \left\| \sum_{k \neq i} U_{ki} \right\|^2 \leq 2n^{-3} \sum_{k \neq i} \mathbb{E} \|U_{ki}\|^2.$$

To verify condition (5.3) there, again the martingale property and the fact that the integrands are centered is needed. The number of (identical) summands is $O(n^2)$, proving (4.18)

To bound α_n^3 , note that

$$\alpha_n^3(\vartheta) = n^{-5/2} \sum_{p,k,i=1}^n \int_I [\Lambda_{\vartheta,p} - \Lambda_{\vartheta_0,p}] \left[\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k} - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta} \right] dM_i \equiv n^{-5/2} \sum_{p,k,i} U_{pki}.$$

Again it is easy to see that the subsum over $p \neq k \neq i$ dominates the sum. For these index combinations the assumptions of Lemma 17 are satisfied. Conclude that

$$n^{-5} \mathbb{E} \left[\left\| \sum_{p \neq k \neq i} U_{pki} \right\|^2 \right] = o(1).$$

This completes the proof.

Lemma 11. *On compact subsets Θ_0 of Θ we have for all $\vartheta, \vartheta' \in \Theta_0$*

$$\mathbb{E} [\|\alpha_n^i(\vartheta) - \alpha_n^i(\vartheta')\|^2] \leq \text{const} \|\vartheta - \vartheta'\|^2.$$

Proof. The proof is similar to that of Lemma 10. Again the inequalities of Lemmas 15–17 are used. Details are omitted.

Next we study the process $\beta_n(\vartheta)$.

Lemma 12. *The processes β_n are uniformly tight on compacta, and*

$$\beta_n(\vartheta) = o_{\mathbb{P}}(1) \text{ for each } \vartheta. \tag{4.19}$$

Hence $\beta_n \rightarrow 0$ uniformly on compacta.

Proof. For pointwise convergence, note that

$$\beta_n(\vartheta) = n^{-5/2} \sum_{p,k,i=1}^n \int_I M_k(t) \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,p}(t) M_i(dt).$$

Again we split the sum into terms where all indices are equal up to terms where all are distinct. Since within each group the second moments of the stochastic integrals are finite it suffices to consider the group for which $p \neq k \neq i$. Set

$$U_{pki} = \int_I M_k \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,p} dM_i.$$

According to Lemma 17,

$$\begin{aligned} & n^{-5} \mathbb{E} \left\{ \left\| \sum_{p \neq k \neq i} U_{pki} \right\|^2 \right\} \\ & \leq 64n^{-5} \sum \mathbb{E} \int_I \int_I M_k(t) M_k(s) \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,p}^T(t) \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,q}(s) M_i(dt) M_i(ds), \end{aligned}$$

where the number of summands is of the order n^4 . This proves (4.19).

For the increment we get

$$\beta_n(\vartheta) - \beta_n(\vartheta') = n^{-5/2} \sum_{p,k,i=1}^n \int_I M_k(t) \left[\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,p}(t) - \frac{\partial}{\partial \vartheta'} \Lambda_{\vartheta',p}(t) \right] M_i(dt).$$

The moment bounds are obtained as before. The required factor $\|\vartheta - \vartheta'\|^2$ is obtained from the differentiability of $\mathbb{E} \left\| \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,1} - \frac{\partial}{\partial \vartheta'} \Lambda_{\vartheta',1} \right\|^2$.

Finally, we study the process $\gamma_n(\vartheta)$.

Lemma 13. *We have*

$$\gamma_n(\vartheta) = n^{1/2} \int_I \bar{M}_n(t) \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \mathbb{E} \Lambda_{\vartheta_0}(dt) + o_{\mathbb{P}}(1)$$

uniformly on compacta. Moreover, the leading term is uniformly C-tight in ϑ .

Proof. Set

$$\begin{aligned} \tilde{\gamma}_n(\vartheta) &= n^{1/2} \int_I \bar{M}_n(t) \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \mathbb{E} \Lambda_{\vartheta_0}(dt), \\ \gamma_n^1(\vartheta) &= n^{1/2} \int_I \bar{M}_n(t) \left(\frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right) \mathbb{E} \Lambda_{\vartheta_0}(dt), \\ \gamma_n^2(\vartheta) &= n^{1/2} \int_I \bar{M}_n(t) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) [\bar{\Lambda}_{\vartheta_0,n}(dt) - \mathbb{E} \Lambda_{\vartheta_0}(dt)]. \end{aligned}$$

Then $\gamma_n(\vartheta) = \tilde{\gamma}_n(\vartheta) + \gamma_n^1(\vartheta) + \gamma_n^2(\vartheta)$. Since

$$\tilde{\gamma}_n(\vartheta) = n^{-1/2} \sum_{k=1}^n \int_I M_k(t) \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) \mathbb{E} \Lambda_{\vartheta_0}(dt)$$

is a sum of centered i.i.d. random variables, the CLT can be applied directly. Also the oscillation bound $\mathbb{E} \|\tilde{\gamma}_n(\vartheta) - \tilde{\gamma}_n(\vartheta')\| \leq \text{const} \|\vartheta - \vartheta'\|^2$ is immediate from the smoothness of $\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta$. The uniform convergence of γ_n^1 and γ_n^2 to zero follows, as in previous cases, from the pointwise convergence to zero and the oscillation bound, by applying inequalities in the Appendix.

For the next lemma recall Φ_0 and define

$$\Phi_n(\vartheta) = \frac{\partial}{\partial \vartheta} \int_I [\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_0,n}(t)] \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t)^T \bar{\Lambda}_{\vartheta_0,n}(dt).$$

Lemma 14. *The matrix $\Phi_n(\vartheta)$ admits the expansion $\Phi_n(\vartheta) = \Phi_0(\vartheta) + o_{\mathbb{P}}(1)$. The representation is uniform on compacta.*

Proof. We have

$$\Phi_n(\vartheta) = \int_I [\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_0,n}(t)] \frac{\partial^2}{\partial \vartheta^2} \bar{\Lambda}_{\vartheta,n}(t)^T \bar{\Lambda}_{\vartheta_0,n}(dt) \tag{4.20}$$

$$+ \int_I \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t)^T \bar{\Lambda}_{\vartheta_0,n}(dt). \tag{4.21}$$

Unlike α_n, β_n , and γ_n , the processes here are not standardized. Hence almost sure convergence results related to the Strong Law of Large Numbers apply. Actually, all $\bar{\Lambda}_{\vartheta,n}$ are sample means of i.i.d. nondecreasing processes. A Glivenko-Cantelli argument yields, with probability one, uniform convergence of $\bar{\Lambda}_{\vartheta,n}(t)$ to $\mathbb{E} \Lambda_\vartheta(t)$ uniformly in t on compact subsets of Θ . Similarly, for averages of derivative processes.

Remark 5. Lemma 14 guarantees that in a finite sample situation, we can replace the unknown standardizing matrix $\Phi_0(\vartheta_0)$ by $\Phi_n(\vartheta_n)$ without destroying the distributional approximation through $\mathcal{N}_d(0, C(\vartheta_0))$. Similarly, $C(\vartheta_0)$ can be replaced by $C^n(\vartheta_n)$, where C^n is the sample analog of C .

Proofs of Theorem 2 and Corollary 1. It follows from (4.15) and the preceding lemmas that, at $\vartheta = \vartheta_n$,

$$\begin{aligned} & n^{1/2} \int_I [\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_0,n}(t)] \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(dt) \bar{\Lambda}_{\vartheta_0,n}(dt) \\ &= n^{1/2} \int_I \bar{M}_n(t) \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) \mathbb{E} \Lambda_{\vartheta_0}(dt) \Big|_{\vartheta=\vartheta_0} + o_{\mathbb{P}}(1). \end{aligned} \tag{4.22}$$

Recall $I = [t, \bar{t}]$. By Fubini's theorem, the last integral is

$$n^{1/2} \int_I \int_{[x, \bar{t}]} \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) \mathbb{E} \Lambda_{\vartheta_0}(dt) \bar{M}_n(dx) \Big|_{\vartheta=\vartheta_0}.$$

The left-hand side of (4.22) is $n^{1/2} \Phi_n(\tilde{\vartheta}_n)(\vartheta_n - \vartheta_0)$ for some appropriate $\tilde{\vartheta}_n$ between ϑ_n and ϑ_0 . From Lemma 14 and Theorem 1 it follows that

$$\begin{aligned} & n^{1/2} \Phi_n(\vartheta_0)(\vartheta_n - \vartheta_0) \\ &= n^{1/2} \int_I \int_{[x, \bar{t}]} \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) \mathbb{E} \Lambda_{\vartheta_0}(dt) \bar{M}_n(dx) \Big|_{\vartheta=\vartheta_0} + o_{\mathbb{P}}(1). \end{aligned}$$

Since \bar{M}_n is a sum of centered independent martingales and the integrand is a deterministic function, the CLT can be applied to achieve asymptotic normality of the right-hand side. The covariance matrix of the leading term is easily computed to be $(C_{ij})_{1 \leq i, j \leq d}$. As to the left-hand side, apply Lemma 14 and a Slutsky-type argument. This completes the proofs of Theorem 2 and Corollary 1.

5. Appendix

In the context of point processes or, more generally, functional data analysis, it is important to have access to a methodology that has proved successful for real-valued and multivariate data. One particular class of such statistics are U -statistics, see Serfling (1980) for details. There, if Z_1, \dots , denotes a sequence of i.i.d. random vectors and, for a fixed $m \in \mathbb{N}$, h is an integrable function of m variates (the “kernel”), the associated U -statistic for sample size n is

$$R_n = \frac{(n-m)!}{n!} \sum h(Z_{i_1}, \dots, Z_{i_m}),$$

where the summation takes place over all distinct $i_1 \neq \dots \neq i_m$ from $1, \dots, n$. A crucial result in this context is the strong consistency of R_n as $n \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} R_n = \mathbb{E} h(Z_1, \dots, Z_m) \text{ almost surely and in the mean.} \quad (5.1)$$

The proof shows that, for a proper decreasing filtration $\mathcal{G}_1 \supset \mathcal{G}_2 \supset \dots$, (R_n, \mathcal{G}_n) is a reverse martingale. Then, by the Hewitt-Savage 0-1 law, the limit is constant and has the required expectation. From Hewitt and Savage (1955), it is clear that this also holds for random elements in more general sample spaces. In our case, when we use U -statistics for point processes on a compact interval I , Z_i take values in the Skorokhod space $D(I)$, which is known to be a Polish space. See Billingsley (1968). In such situations, the arguments known for real-valued quantities may be applied as well. With this, we may apply (5.1) when Z_1, \dots are point processes on I .

Lemma 15. *Let, for $p, k = 1, \dots, n$, U_{pk} be d -variate random vectors with $\mathbb{E}\|U_{pk}\|^2 < \infty$. If*

$$\mathbb{E}U_{pk}^T U_{ql} = 0 \text{ for } k \neq p, q, l \text{ or } l \neq p, q, k, \tag{5.2}$$

then
$$\mathbb{E}\left\|\sum_{p \neq k} U_{pk}\right\|^2 \leq 2 \sum_{k \leq p, q} \mathbb{E}(U_{pk}^T U_{qk}).$$

Proof. We have

$$\begin{aligned} \mathbb{E}\left\|\sum_{\substack{k, p=1 \\ k \neq p}}^n U_{pk}\right\|^2 &= \mathbb{E}\left\|\sum_{k < p} U_{pk} + \sum_{p < k} U_{pk}\right\|^2 \\ &\leq \mathbb{E}\left\{\left\|\sum_{k < p} U_{pk}\right\| + \left\|\sum_{p < k} U_{pk}\right\|\right\}^2 \leq 2\mathbb{E}\left\|\sum_{k < p} U_{pk}\right\|^2 + 2\mathbb{E}\left\|\sum_{p < k} U_{pk}\right\|^2. \end{aligned}$$

As to the first expectation we obtain

$$\begin{aligned} \mathbb{E}\left\|\sum_{k < p} U_{pk}\right\|^2 &= \mathbb{E}\left\{\left(\sum_{k < p} U_{pk}\right)^T \left(\sum_{l < q} U_{ql}\right)\right\} \\ &= \sum_{k < p} \sum_{l < q} \mathbb{E}(U_{pk}^T U_{ql}) = \sum_{k < p, q} \mathbb{E}(U_{pk}^T U_{qk}), \end{aligned}$$

in view of (5.2). Similarly,

$$\mathbb{E}\left\|\sum_{k > p} U_{pk}\right\|^2 = \sum_{k > p, q} \mathbb{E}(U_{pk}^T U_{qk}).$$

This completes the proof of the lemma.

Lemma 16. *Let, for $k, i = 1, \dots, n$, U_{ki} be d -variate random vectors such that $\mathbb{E}\|U_{ki}\|^2 < \infty$, and*

$$\mathbb{E}(U_{ki}^T U_{lj}) = 0 \text{ whenever one index differs from the rest.} \tag{5.3}$$

Then
$$\mathbb{E}\left\|\sum_{k \neq i} U_{ki}\right\|^2 \leq 2 \sum_{k \neq i} \mathbb{E}\|U_{ki}\|^2.$$

Proof. As in the previous proof,

$$\mathbb{E}\left\|\sum_{k \neq i} U_{ki}\right\|^2 = 2\mathbb{E}\left\|\sum_{k < i} U_{ki}\right\|^2 + 2\mathbb{E}\left\|\sum_{k > i} U_{ki}\right\|^2. \tag{5.4}$$

The first expectation on the right-hand side can be expanded as

$$\begin{aligned} \mathbb{E}\left\|\sum_{k < i} U_{ki}\right\|^2 &= \sum_{k < i, l < j} \mathbb{E}(U_{ki}^T U_{lj}) = \sum_{k < i, j} \mathbb{E}(U_{ki}^T U_{kj}) \\ &= \sum_{k < i} \mathbb{E}\|U_{ki}\|^2. \end{aligned}$$

The second expectation in (5.4) is dealt with similarly.

Our final inequality deals with triple-indexed vectors.

Lemma 17. *Let U_{pki} , for $1 \leq p, k, i \leq n$, be d -variate random vectors satisfying $\mathbb{E} [U_{pki}^T U_{qlj}] = 0$ whenever k, i, l or j differ from the rest. Then*

$$\mathbb{E} \left\| \sum_{p \neq k \neq i} U_{pki} \right\|^2 \leq 64 \sum \mathbb{E} [U_{pki}^T U_{qki}],$$

where the summation takes place over all $p, q < k < i; p, q < i < k; k < p, q < i; i < p, q < k; k < i < p, q;$ or $i < k < p, q$.

Proof. We need the inequality

$$\left\| \sum_{j=1}^l a_j \right\|^2 \leq 2^{l-1} \sum_{j=1}^l \|a_j\|^2,$$

which is valid for all vectors a_1, \dots, a_l in \mathbb{R}^d . It follows that

$$\begin{aligned} & \mathbb{E} \left\| \sum_{p \neq k \neq i} U_{pki} \right\|^2 \\ &= \mathbb{E} \left\| \sum_{p < k < i} U_{pki} + \sum_{p < i < k} U_{pki} + \sum_{k < p < i} U_{pki} + \sum_{i < p < k} U_{pki} + \sum_{k < i < p} U_{pki} + \sum_{i < k < p} U_{pki} \right\|^2 \\ &\leq 64 \left\{ \mathbb{E} \left\| \sum_{p < k < i} U_{pki} \right\|^2 + \dots + \mathbb{E} \left\| \sum_{i < k < p} U_{pki} \right\|^2 \right\}. \end{aligned}$$

We only bound the first expectation, the others being dealt with similarly. Now,

$$\mathbb{E} \left\| \sum_{p < k < i} U_{pki} \right\|^2 = \mathbb{E} \left\{ \left(\sum_{p < k < i} U_{pki} \right)^T \left(\sum_{p < k < i} U_{pki} \right) \right\} = \sum_{\substack{p < k < i \\ q < l < j}} \mathbb{E} U_{pki}^T U_{qlj}.$$

Whenever $i \neq j$, the expectation vanishes. Hence we can restrict summation to $i = j$. In this case the expectations also vanish for $k \neq l$. Conclude that

$$\mathbb{E} \left\| \sum_{p < k < i} U_{pki} \right\|^2 = \sum_{p, q < k < i} \mathbb{E} [U_{pki}^T U_{qki}].$$

Similarly

$$\begin{aligned} \mathbb{E} \left\| \sum_{p < i < k} U_{pki} \right\|^2 &= \sum_{p, q < i < k} \mathbb{E} [U_{pki}^T U_{qki}] \\ \mathbb{E} \left\| \sum_{k < p < i} U_{pki} \right\|^2 &= \sum_{k < p, q < i} \mathbb{E} [U_{pki}^T U_{qki}] \\ &\vdots \\ \mathbb{E} \left\| \sum_{i < k < p} U_{pki} \right\|^2 &= \sum_{i < k < p, q} \mathbb{E} [U_{pki}^T U_{qki}]. \end{aligned}$$

This completes the proof.

6. Conclusion

The paper provides a methodology to statistically study self-exciting point processes in the context of a functional data analysis. Processes of this type appear in such areas as market research, survival analysis, reliability, and credit risk. The method is applied to a “Single Source Panel” of AC Nielsen, Germany aimed at estimating the impact of TV-advertising and “Adstock-effects” on an individual level, taking account of socio-demographic parameters. We show how professional expertise in marketing can lead to a dynamic model, and present results.

Acknowledgement

We are grateful to Prof. Robert Engle for drawing our attention to his work and to a referee for valuable hints on the literature.

References

- Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.
- Bickel, P. J. and Wichura, J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* **42**, 1656-1670.
- Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley.
- Borgan, O. (1984). Maximum likelihood estimation in a parametric counting process model, with applications to censored failure time data and multiplicative models. *Scand. J. Statist.* **11**, 1-16.
- Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1**, 279-290.
- Engle, R. F. and Russell, J. R. (1997). Forecasting the frequency of changes in quoted foreign exchange prices with the ACD model. *J. Empirical Finance* **4**, 187-212.

- Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica* **66**, 1127-1162.
- Hawkes, A. G. (1971a). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58**, 83-90.
- Hawkes, A. G. (1971b). Point spectra of some mutually exciting point processes. *J. Roy. Statist. Soc. Ser. B* **33**, 438-443.
- Hawkes, A. G. and Oakes, D. (1974). A cluster process representation of a self-exciting process. *J. Appl. Probab.* **11**, 493-503.
- Hewitt, E. and Savage, L. J. (1955). Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* **80**, 470-501.
- Jacobsen, M. (1982). *Statistical Analysis of Counting Processes*. Springer, New York.
- Jacobsen, M. (1984). Maximum likelihood estimation in the multiplicative intensity model: a survey. *Internat. Statist. Rev.* **52**, 193-207.
- Karr, A. F. (1986). *Point Processes and Their Statistical Inference*. Marcel Dekker, New York.
- Kopperschmidt, K. and Stute, W. (2009). Purchase timing models in marketing: a review. *Adv. Stat. Anal.* **93**, 123-149.
- Koul, H. and Stute, W. (1999). Nonparametric model checks for time series. *Ann. Statist.* **27**, 204-236.
- Moller, J. and Rasmussen, J. G. (2005). Perfect simulation of Hawkes processes. *Adv. Appl. Probab.* **37**, 629-646.
- Moller, J. and Rasmussen, J. G. (2006). Approximate simulation of Hawkes processes. *Methodol. Comput. Appl. Probab.* **8**, 53-64.
- Ogata, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Ann. Inst. Statist. Math.* **30**, 243-261.
- Ogata, Y. and Akaike, H. (1982). On linear intensity models for mixed doubly stochastic Poisson and self-exciting point processes. *J. Roy. Statist. Soc. Ser. B* **44**, 102-107.
- Rathbun, S. L. (1996). Asymptotic properties of the maximum likelihood estimator for spatio-temporal point processes. *J. Statist. Plann. Inference* **51**, 55-74.
- Schoenberg, F. P. (2005). Consistent parametric estimation of the intensity of a spatial-temporal point process. *J. Statist. Plann. Inference* **128**, 79-93.
- Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.
- Snyder, D. L. (1975). *Random Point Processes*. Wiley, New York.
- Stute, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25**, 613-641.
- Stute, W., Thies, S. and Zhu, L. X. (1998). Model checks for regression: an innovation process approach. *Ann. Statist.* **26**, 1916-1934.
- Waagepetersen, R. and Guan, Y. (2009). Two-step estimation for inhomogeneous spatial point processes. *J. Roy. Statist. Soc. Ser. B* **71**, 685-702.

The Nielsen Company (Germany) GmbH, Insterburger Str. 16, D-60487 Frankfurt, Germany.

E-mail: Kai.Kopperschmidt@nielsen.com

Mathematical Institute, University of Giessen, Arndtstr. 2, D-35392 Giessen, Germany.

E-mail: Winfried.Stute@math.uni-giessen.de

(Received June 2011; accepted October 2012)