GRAPHICAL MODELING FOR MULTIVARIATE HAWKES PROCESSES WITH NONPARAMETRIC LINK FUNCTIONS

MICHAEL EICHLER, RAINER DAHLMHAUS, AND JOHANNES DUECK

Maastricht University and University of Heidelberg

October 23, 2015

Abstract. Hawkes (1971a) introduced a powerful multivariate point process model of mutually exciting processes to explain causal structure in data. In this paper it is shown that the Granger causality structure of the process is fully encoded in the corresponding Hawkes kernels. A new nonparametric estimator of the Hawkes kernels based on a time-discretized version of the point process is introduced by using an infinite order autoregression. Consistency and asymptotic normality of the estimator is derived. The estimator is applied to simulated data and to neural spike train data from the spinal dorsal horn of a rat.

Keywords: Hawkes process, Granger causality, graphical model, mutually exciting process, nonparametric estimation

2010 Mathematics Subject Classification. Primary: 60G55; Secondary: 62M10.

1. Introduction

In two seminal papers, Hawkes (1971a,b) Hawkes introduced a multivariate model for point processes with mutually exciting components now referred to as the Hawkes model. In the beginning it was motivated by modeling aftershocks and seismological phenomena (cf. Vere-Jones 1970, Vere-Jones and Ozaki 1982, Ogata 1999). It also served as a first model for neuron firing and stimulated the introduction of more complex nonlinear models to include inhibitory couplings and the refractory period (Okatan et al. 2005, Cardanobile and Rotter 2010). The usage of the Hawkes model has been more and more spread out to different research areas: Brantingham et al. (2011) examines insurgency in Iraq, Mohler et al. (2011) use it for modeling crime, Reynaud-Bouret and Schbath (2010) apply it to genome analysis, and Carstensen et al. (2010) model the occurrence of regulatory elements. Recently, the Hawkes model has become popular in particular in finance for modeling price fluctuations or transactions, cf. Bacry et al. (2013), Bacry et al. (2012) and Embrechts et al. (2011).

In this paper we put the focus on the causal structure of the Hawkes model by applying causality concepts to mutually exciting point processes. Granger Granger (1969) defined the notion of Granger causality. It reflects the belief that a cause should always occur before the effect and that the prediction of a process with the knowledge of a possible cause should improve if there is a causal relation present. However, temporal precedence alone is not a sufficient condition for establishing cause–effect relationships, and it is commonly accepted that empirical evaluation of
Granger causality can lead to false detection of causal links. Nevertheless, the concept of Granger causality together with suitable graphical representations remains a useful tool for causal learning as has been shown in Eichler (2012a, 2013). The first objective of this paper is to set the framework for such causal learning approaches by defining the necessary graphical concepts. In particular, we establish a global Markov property, which relates the Granger causalities observed for part of the variables to the causal structure of the full system. Such global Markov properties play a key role in the graphical approach to causal learning.

The original definition of Granger applies only to processes in discrete time. Extensions to continuous time processes have been developed in the general framework of continuous–time semimartingales by Florens and Fougere (1996) and for mean square continuous processes by Comte and Renault (1996). The notion of Granger causality in continuous time is closely related to the definition of local independence for composable Markov processes (Schweder 1970) and marked point processes (Didelez 2008).

A Hawkes process is a multivariate point process with conditional intensity

$$\lambda(t) = \nu + \int_0^t \phi(u) dN(t-u),$$

where $\nu$ is a vector of positive constants (often referred to as background or Poisson rates) and $\phi(\cdot)$ is a matrix of nonnegative functions that vanish on the negative half axis. It may be viewed as a continuous analogon to classical auto-regression in time series analysis. Whenever a component process jumps it increases the conditional firing rate of the other processes specified by the corresponding Hawkes kernel $\phi(\cdot)$. Hence, a causal structure is encoded in the Hawkes kernels. The problem is to estimate these decay kernels. The most popular approach is an ML-approach as in Ozaki (1979) where $\phi(\cdot)$ is assumed to be of parametric form, e.g. consisting of exponential functions or Laguerre polynomials. In recent literature other estimation procedures evolved. Bacry et al. (2012) use a numerical method for nonparametric estimation based on martingale and Laplace transform techniques and Lewis and Mohler (2011) the EM-algorithm. As these methods involve complex computations, their use for causal learning algorithms, which require fitting of a large number of models, is limited.

In this paper, we present a simple and fast alternative to the existing nonparametric estimation method for the Hawkes kernels: we propose to discretize the point process by considering the the increments over equidistant time points and then to fit a vector autoregressive model by least squares. Part of the derivation will be along the ideas in Lewis and Reinsel (1985). The paper is organized as follows: in Section 2 we define the Hawkes model and summarize some basic properties. Section 3 contains a discussion of the causality structure with respect to Granger causality and from the point of view of graphical models. In Section 4 we introduce the nonparametric estimator and prove consistency when the observation interval tends to infinity and the discretization step size tends to zero. As an illustration, Section 4.1 contains an application to EEG data from the spinal dorsal horn of a rat. Section 5 contains some concluding remarks. Part of the proofs have been put into the appendix.
2. Multivariate Hawkes Processes

We consider multivariate point processes \( N = (N_1, \ldots, N_d)' \) on a probability space \((\Omega, \mathcal{F}, \mathbb{P})\), that is, the components \( N_i, i = 1, \ldots, d \) are a random counting measure on \( \mathbb{R} \). For simplicity, we will write \( N_i(t) = N_i([0, t]) \) for the counting process giving the number of events of the \( i \)-th component process up to time \( t \). Throughout the paper, we make the following basic assumption.

Assumption 2.1. The point process \( N = (N_1, \ldots, N_d)' \) is stationary, 
\[
N_1(A_1), \ldots, N_d(A_d) \sim N_1(t + A_1), \ldots, N_d(t + A_d)
\]
for all measurable sets \( A_1, \ldots, A_d \subseteq \mathbb{R} \), where \( t + A = \{t + a \in \mathbb{R} | a \in A\} \). Furthermore, \( N \) is a simple point process, that is, the counting processes \( N_j, 1 \leq j \leq d \), have almost surely step size 1 and do not jump simultaneously.

Let \( \mathcal{F} = (\mathcal{F}(t))_{t \in \mathbb{R}} \) be a filtration and suppose that \( N \) is adapted to \( \mathcal{F} \). Then, under mild regularity conditions (see Daley and Vere-Jones 2003), the point process \( N \) has a conditional intensity which satisfies 
\[
\mathbb{P}(dN(t) = 1 | \mathcal{F}(t)) = \lambda(t) dt + o_{\mathbb{P}}(dt).
\]

Multivariate Hawkes processes are point processes where the conditional intensity for each component depends linearly on the past events.

Definition 2.2. A multivariate Hawkes process is a stationary and simple multivariate point process \( N = (N_1, \ldots, N_d)' \) such that \( N_i \) has conditional intensity
\[
\lambda_i(t) = \nu_i + \sum_{j=1}^d \int_0^t \phi_{ij}(u) dN_j(t - u),
\]
where \( \nu_i > 0 \) is the baseline intensity of the \( i \)-th component and the link functions \( \phi_{ij} \) satisfy \( \phi(u) = 0 \) for \( u \leq 0 \) and 
\[
\int_0^\infty \|\phi_{ij}(u)\| du < 1.
\]

The integrability condition (2.2) ensures the existence and uniqueness of a stationary point process with conditional intensities given by (2.1) (e.g. Hawkes and Oakes 1974). Furthermore, the condition \( \phi_{ij}(0) = 0 \) ensures that the conditional intensity \( \lambda(t) \) is left-continuous and hence is predictable. As a consequence, the components of a Hawkes process are special semi-martingales with respect to the canonical filtration \( \mathcal{F} \). This means that every component \( N_i \) has an almost surely unique decomposition 
\[
N_i(t) = \Lambda_i(t) + M_i(t), \quad t \in \mathbb{R},
\]
where \( \Lambda = (\Lambda(t))_{t \in \mathbb{R}} \) is a predictable process with \( \Lambda(0) = 0 \) and \( M = (M(t))_{t \in \mathbb{R}} \) is a martingale, both with respect to \( \mathcal{F} \). The process \( \Lambda(t) \) is called the compensator of the point process \( N \) and is related to the conditional intensity \( \lambda(t) \) by 
\[
\Lambda(t) = \int_0^t \lambda(s) ds.
\]
In terms of stochastic differentials, we thus have 
\[
dN(t) = \lambda(t) dt + dM(t).
\]
The decomposition (2.3) is important for the definition of Granger non-causality in section 3.

For a stationary Hawkes process, the mean intensity $p_N = \mathbb{E}(N(1))$ is related to the link function $\phi$ and baseline intensity $\nu$ by

$$p_N = \left( I_d - \int_0^\infty \phi(u) \, du \right)^{-1} \nu$$

(2.4)

where $I_d$ denotes the $d \times d$ identity matrix. The covariance structure of a stationary simple point process $N$ is given by

$$\text{cov}(N(A), N(B)) = \int_A \int_B q_{NN}(t-s) \, dt \, ds + \int_{A \cap B} p_N \, dt,$$

were $q_{NN}$ is the covariance density of $N$. Here, the second integral is due to the fact that for a simple point process we have $\mathbb{E}(dN(t)^2) = \mathbb{E}(dN(t))$. For Hawkes processes, explicit expressions for the covariance density are available, for instance, in the case of exponentially decaying link functions (Hawkes 1971a); Bacry et al. (2012) provides a detailed analysis of the covariance structure in the general case.

Similarly, for integrable functions $f_1, \ldots, f_k$, we have for the cumulants of higher order

$$\int_{\mathbb{R}^k} \prod_{i=1}^k f_i(t_i) \, \text{cum}(dN_{i_1}(t_1), \ldots, dN_{i_k}(t_k))$$

$$= \sum_{P_1, \ldots, P_m} \int_{\mathbb{R}^m} \prod_{j \in P_1} f_j(t_1) \cdots \prod_{j \in P_m} f_j(t_m) q_{P_1,\ldots,P_m}(t_1, \ldots, t_m) \, dt_1 \cdots dt_m,$$

(2.5)

where the first sum extends over all partitions $\{P_1, \ldots, P_m\}$ of $\{1, \ldots, k\}$ with $m = 1, \ldots, k$. Furthermore, $q_{i_1,\ldots,i_m}$ denotes the cumulant density of $N_{i_1,\ldots,i_m}$ and the first sum extends over all partitions $\{P_1, \ldots, P_m\}$, $m = 1, \ldots, k$, of $\{1, \ldots, k\}$. Explicit expressions for the cumulants of a Hawkes Process can be found in Jovanović et al. (2015).

3. **Graphical modelling of multivariate Hawkes processes**

Let $N = (N_1, \ldots, N_d)'$ be a stationary $d$-dimensional point process with canonical filtration $\mathcal{F} = (\mathcal{F}(t))_{t \in \mathbb{R}}$. For any $A \subseteq V = \{1, \ldots, d\}$ let $\mathcal{F}_A(t)$ be the sub-$\sigma$-algebra that corresponds to the sub-process $N_A = (N_a)_{a \in A}$. Following the general definition of Florens and Fougerere (1996), we obtain the following definition of Granger non-causality. We note that Florens and Fougerere (1996) use the term “instantaneous Granger non-causality” to distinguish it from non-causality over longer time horizons.

**Definition 3.1 (Granger non-causality).** Let $N$ be a stationary multivariate point process with canonical filtration $\mathcal{F}$. Then the $i$-th component $N_i$ does not Granger-cause the $j$-th component $N_j$ with respect to $\mathcal{F}$ if the compensator $\Lambda_j(t)$—or equivalently the conditional intensity function $\lambda_j(t)$—is $\mathcal{F}_-(t-)$-measurable for all $t \in \mathbb{R}$.

The above definition could be generalized by including also additional exogenous variables $X$ into the filtration. For our discussion of multivariate Hawkes processes
as given by 2.2 it is sufficient to consider canonical filtrations generated by the point process $N$.

For the $i$–th component $N_i$ of a Hawkes process, the conditional intensity $\lambda_i(t)$ is $\mathcal{F}_j(t–)$–measurable if and only if the $j$–th link function $g_{ij}$ is identical to zero. Thus we have the straightforward result.

**Proposition 3.2.** Let $N$ be a multivariate Hawkes process with intensities as in (2.1). Then $N_i$ does not Granger–cause $N_j$ with respect to $N$ if and only if $\phi_{ji}(u) = 0$ for all $u \in \mathbb{R}$.

The martingale property of $M$ with respect to $\mathcal{F}$, that is, $\mathbb{E}(dM(t) | \mathcal{F}(t)) = 0$, suggests that the definition of Granger causality considers only dependence in the mean and hence corresponds to what is known as Granger causality in the mean in the context of time series. However, we note that for simple point processes the conditional intensities determine the full conditional distribution. This rules out any higher-order dependences of $M(t)$ on the past. Additionally, the increments of the components of $M(t)$ are mutually independent since simultaneous occurrence of events is almost surely not possible for a simple process. Consequently, the above definition in fact describes a notion of strong Granger (non–)causality formulated in terms of conditional independence.

Schweder (1970) introduced the concept of local independence to describe dynamic dependences in time–continuous Markov processes. We note that the above notion of Granger noncausality and that of local independence are equivalent in the present context of point processes.

With the above definition of Granger noncausality, the definition of Granger causality graphs in Eichler (2007, 2012b) directly extends to the present case of multivariate Hawkes processes.

**Definition 3.3.** For a multivariate Hawkes process $N = (N_1, \ldots, N_d)'$, the Granger causality graph of $N$ is given by a graph $G$ with vertices $V = \{1, \ldots, d\}$ and directed edges $i \rightarrow j$ with $i, j \in V$ satisfying

$$i \rightarrow j \notin G \iff \phi_{ji}(u) = 0 \text{ for all } u \in \mathbb{R}.$$  

The use of Granger causality graphs goes beyond simple visualization of dynamic dependences. The key feature of the graphical approach is that it relates Granger noncausality to pathwise separation in the graph by so-called global Markov properties. This allows to derive Granger noncausality relations for arbitrary subprocesses from the graph. More importantly, the graphical approach yields also criteria for identifying the Granger causal structure of a system which is only partially observed. This is particularly of interest in neurological applications where the activity of only a small number of neurons can be recorded.

The global Granger causal Markov property is a general result that goes beyond the framework of multivariate Hawkes processes. For multivariate simple stationary point processes, it simplifies compared to the time series case as the increments of the martingale $M(t)$ are mutually independent. For its formulation, we need some terminology from graph theory.

Let $G$ be a directed graph with vertex set $V$ and edges $E$. A path in the graph $G$ is a sequence of edges $\pi = (e_1, \ldots, e_n)$ with $e_i \in \{a_{i-1} \rightarrow a_i, a_{i-1} \leftarrow a_i\}$ for vertices $a_0, \ldots, a_n \in V$ with $a_0 = a$ and $a_n = b$. If the last edge on the path is $a_{n-1} \rightarrow b$,
that is, the path ends with an arrowhead at \( b \), we speak of a path \( \pi \) from \( a \) to \( b \) (also referred to as \( B \)-pointing path, cf Eichler 2007). A vertex \( a_i \) on a path \( \pi \) is called a collider if the adjacent edges form the subpath \( a_{i-1} \rightarrow a_i \leftarrow a_{i+1} \); otherwise \( a_i \) is called a non-collider. A path \( \pi \) is blocked by a set \( C \) if and only if there exists one collider on the path that does not lie in \( C \) or the exists one non-collider that lies in \( C \).

**Definition 3.4 (Global Markov properties).** A multivariate stationary simple point process \( N \) satisfies the global Granger causal Markov property with respect to a directed graph \( G \) if the following condition holds: \( N_A \) does not Granger cause \( N_B \) with respect to \( N_S \) if every path from a vertex \( a \in A \) to a vertex \( b \in B \) is blocked by the set \( S \setminus A \).

Furthermore, we say that \( N \) satisfies the global Markov property with respect to an undirected graph \( U \) if the processes \( N_A \) and \( N_B \) are independent conditionally on \( N_C \) whenever the sets \( A \) and \( B \) are separated by \( C \) in \( U \), that is, every path between some vertex \( a \in A \) and some vertex \( b \in B \) contains at least one vertex \( c \in C \).

The following result states that a stationary Hawkes processes \( N \) is Granger–Markov with respect to its Granger–causality graph \( G \). Likewise, \( N \) is also Markov with respect to the moral graph \( G^m \) derived from the Granger–causality graph \( G \). Here, the moral graph of a directed graph \( G \) is defined as the undirected graph \( G^m \) that has the same vertex set as \( G \) and has edges \( i \rightarrow j \in G^m \) if \( i \) and \( j \) are adjacent in \( G \) or there exists a third vertex \( k \) such that \( G \) contains both edges \( i \rightarrow k \) and \( j \rightarrow k \). Furthermore, a vertex \( a \) is an ancestor of another vertex \( b \) if there exists a path \( a \rightarrow \ldots \rightarrow b \) in \( G \); the set of all ancestors of vertices in \( B \subseteq V \) is denoted by \( \text{an}(B) \). Finally, \( G_A \) for some subset \( A \subseteq V \) denotes the subgraph obtained from the graph \( G \) by retaining all edges that connect vertices in \( A \).

**Theorem 3.5.** Let \( N \) be a stationary multivariate Hawkes process and let \( G \) be the Granger causality graph of \( N \). Then

(i) \( N \) satisfies the global Granger causal Markov property with respect to \( G \);

(ii) every subprocess \( N_S, S \subseteq V \), satisfies the global Markov property with respect to \( (G_{\text{an}(S) \cup S})^m \).

We note that the graph \( H = (G_{\text{an}(S) \cup S})^m \) in (ii) can be reduced further to a graph \( H(S) \) with vertex set \( S \) by extending the subgraph \( H_S \) by additional edges \( i \rightarrow j \) whenever \( i \) and \( j \) are not separated by \( S \setminus \{i, j\} \) in \( H \).

The global Granger causal Markov property allows an intuitive interpretation of pathways in Granger causality graphs. Moreover, it is of fundamental importance for graphical approaches to causal discovery. Here, the main problem is to distinguish true cause–effect relationships from so-called spurious causation due to unobserved variables. Under the global Granger–causal Markov property, the causal structure of the system including any relevant unobserved variables implies certain Granger non-causal relations among the observed variables and any subset thereof. Algorithms for causal discovery exploit this link by identifying all causal structures that are consistent with the observed Granger non-causal relations. For more details, we refer to Eichler (2012a, 2013).

The largest problem for the implementation of such algorithms for causal discovery is the extremely large number of models that need to be fitted: \( 2^d - d - 1 \) models for \( d \).
variables. This prohibits the use of iterative methods for parameter estimation such as, for instance, the EM algorithm by Lewis and Mohler (2011). In the next section, we therefore discuss nonparametric estimation of the link function by discretizing the point process and applying standard least squares estimation for autoregressive time series.

4. Nonparametric Estimation and Identification

Our approach for nonparametric estimation of the link function $\phi$ is via discretization and consequently using methods from time series analysis. Again, as in section 2 we observe a multivariate point process $N = (N(t))_{t \in \mathbb{R}}$ with component processes $N_i$, $1 \leq i \leq d$, where $d$ is the dimension of the process $N$. The conditional intensity function is once more given by (2.1), where the component functions of $\phi$ belong to a class of non-negative integrable functions that is specified later on in the section. Our objective is the nonparametric estimation of the link functions of the Hawkes process, that is, we do not assume any parametric form of the link functions such as an exponential form. For the purpose of discretization we define for fixed $h > 0$

$$Y_{i,t}^h = N_i((t-1)h) - N_i((t-1)h)$$

for all $t \in \mathbb{Z}$ and $1 \leq i \leq d$. This is equivalent to dividing the real line into intervals of width $h$. For every fixed $h$, $Y^h$ represents a $d$-dimensional time series displaying the number of jumps in time intervals $((t-1)h, t h]$ for $t \in \mathbb{Z}$. Thus, for $h$ small enough, the random variables $Y^h$ are approximately binary. Notice that the considered point processes are orderly such that the probability that more than one jump takes place in an interval of length $h$ is of order $o(h)$. Additionally, Hawkes processes are stochastically continuous. With these observations we may calculate

$$E[Y^h_{i,t+1} | F_{ht}] = P(N((t+1)h) - N((t)h) = 1 | F_{ht}) + o(h)$$

$$= h \nu_i + h \sum_{j=1}^{d} \int_{0}^{\infty} \phi_{ij}(s) dN_j(t h - s) + o(h)$$

$$= h \nu_i + h \sum_{j=1}^{d} \sum_{u=0}^{h} \phi_{ij}(uh + \alpha) dN_j(t (h - u) - \alpha) + o(h).$$

If the link function $\phi$ is continuous and $h$ is small enough, we can approximate $\phi$ by a piecewise constant function, which yields

$$E(Y^h_t | F_{h(t-1)}) \approx h \nu + h \sum_{u=1}^{\infty} \phi(u h) Y^h_{t-u}.$$  \hspace{1cm} (4.1)

This suggests to estimate the link function by a least squares approach.

We introduce some notation. First let $\Gamma^h(u) = \text{cov}(Y^h_t, Y^h_{t-u})$ for $u \geq 0$ and $\Gamma^h(j) = \Gamma^h(-j)'$ for $j < 0$ be the covariance function of the process $Y^h$, which depends on $h$. Furthermore with $Y^h_{t,k} = \vec{Y}^h_{t-1, \ldots, t-k}$ we set

$$\gamma_{h,k} = \text{cov}(Y^h_t, Y^h_{t-k}) = (\Gamma^h(u))_{u=1, \ldots, k},$$

$$\Gamma_{h,k} = \text{cov}(Y^h_{t,k}, Y^h_{t-k}) = (\Gamma^h(u - v))_{u,v=1, \ldots, k}.$$  

Now suppose that the process $N$ has been observed on the interval $[0, T]$ and set $T_h = T/h$. Then the above linear approximation for the conditional mean of $Y^h_t$...
leads to the least squares problem of minimizing
\[ \sum_{t=k+1}^{T_h} \| Y_t^h - \nu^h - \phi^{h,k} Y_t^{h,k} \|_2^2. \]
over the parameters \( \nu^h = \nu h \) and \( \phi^{h,k} = (h \phi(h), \ldots, h \phi(hk)) \).

The above expression is minimized by \( \hat{\phi}^{h,k} = \hat{\gamma}_{h,k} \hat{\Gamma}^{-1}_{h,k} \)
and \( \hat{\nu}^{h,k} = \bar{Y}^h - \hat{\phi}^{h,k} \bar{Y}^{h,k} \),
where with \( T_{h,k} = T_h - k \)
\[ \hat{\gamma}_{h,k} = \frac{1}{T_{h,k}} \sum_{t=k+1}^{T_h} (Y_t^h - \bar{Y}^h)(Y_t^{h,k} - \bar{Y}^{h,k})' \]
is the sample covariance of \( Y_t^h \) and \( Y_t^{h,k} \) and \( \hat{\Gamma}_{h,k}, \bar{Y}^h, \bar{Y}^{h,k} \) are defined similarly.

With these definitions we are able to derive the desired asymptotic results for \( \phi^{h,k} \). Therefore we denote by \( \| B \|_2^2 = \text{tr}(B'B) \) the Euclidean norm and by \( \| B \| = \sup_{\| x \| \leq 1} \| Bx \|_2 \) the spectral norm of \( B \). We note that \( \| B \|_2^2 \) equals the largest eigenvalue of the matrix \( B'B \). For subsequent proofs recall the inequalities \( \| AB \|_2 \leq \| A \|_2 \| B \|_2 \) and \( \| A \| \leq \| A \|_2 \leq \sqrt{r} \| A \| \), where \( r \) is the rank of \( A \).

**Theorem 4.1.** Let \( N \) be a Hawkes process with baseline intensity \( \nu \) and link function \( \phi \) satisfying Assumption 2.1. Additionally suppose that the following conditions hold:
(i) Let \( k = k_T \) and \( h = h_T \) be functions of \( T \) such that
\[ k_T h_T \to \infty, \quad k_T h_T^2 \to 0, \quad \text{and} \quad \frac{k_T^2}{T} \to 0 \]
as \( T \to \infty \).
(ii) The link function \( \phi \) satisfies \( \| \int_0^\infty \phi(u) \, du \| < \infty \).
(iii) The link function \( \phi \) is Lipschitz continuous and decreases to zero with \( \| \phi(u) \| \leq C u^{-1} \) and \( \int_{h_T k_T}^\infty \| \phi(v) \| \, dv = o(1), \quad T \to \infty \).

Then the least squares estimators \( \hat{\phi}^{h,k} \) and \( \hat{\nu}^{h,k} \) are consistent,
\[ \| \hat{\phi}^{h,k} - \phi^{h,k} \|_2^2 \to 0 \quad \text{and} \quad \| \hat{\nu}^{h,k} - \nu^h \|_2^2 \to 0. \]

The first assumption of the above theorem requires usual rate conditions on the sequences \( k_T \) and \( h_T \). The first condition ensures that the support of the estimated link function increases with \( T \) while the other two conditions restrict the growth of the number of parameters. The second assumption \( \| \int_0^\infty \phi(u) \, du \| < 1 \) ensures that the Hawkes process is stationary with absolutely integrable autocovariances density. Assumption (iii) restricts the tail of the link function; it is satisfied, for instance, for exponentially decreasing link functions and hence is not restrictive for practical applications.
The next theorem generalizes Theorem 4.1 to a functional convergence. Therefore we set \( \hat{\phi}_T \) to be the step function defined by
\[
\hat{\phi}_T(u) = \frac{1}{h} \hat{\phi}^h_{[u/h]}, \quad 0 \leq u \leq k h
\]
and zero otherwise.

**Theorem 4.2.** Under the assumptions of Theorem 4.1 it holds
\[
\int_0^\infty \| \hat{\phi}_T(u) - \phi(u) \|_2 \, du \overset{P}{\to} 0 \quad \text{as } T \to \infty.
\]

**Proof.** Decomposing the integral into the approximation error and the estimation error we find
\[
\int_0^\infty \| \hat{\phi}_T(u) - \phi(u) \|_2 \, du \leq \int_0^{hk} \| \hat{\phi}_T(u) - \phi(u) \|_2 \, du + \int_{hk}^\infty \| \phi(u) \|_2 \, du.
\]
Here the second term is of order \( O(h \sqrt{k}) \) while the first term can be bounded by
\[
\| \hat{\phi}^h_{k} - \phi^h_{k} \|_2 \leq h \int_0^{hk} \| \phi(uh) - \phi([u+1]h) \|_2 \, du.
\]
Using Lipschitz continuity of the link function, the second term is of order \( O(h^2 k) = o(1) \) while the first term converges to zero in probability by Theorem 4.1. \( \square \)

Figure 4.1 illustrates our estimation procedure based on simulated data. For the simulation of a three dimensional Hawkes process we used the method by Ogata (1981) based on a thinning algorithm. The true link functions, given by the dashed lines, were taken to have the form \( \phi_{ij}(u) = \alpha_{ij} \exp(-\beta_{ij} u) \) for \( 1 \leq i, j \leq 3 \) with different coefficients. In total, the simulated data contained approximately 5000 events. For the estimation of the link function, \( h = 0.05 \) and \( k = 20 \) were chosen as discretization parameters. The solid lines in Figure 4.1 are the obtained estimates of the link functions.
4.1. Application

As an application we analyzed spike train data from the lumbar spinal dorsal horn of a pentobarbital-anaesthetised rat during noxious stimulation. The firing times of ten neurons were recorded simultaneously by a single electrode with an observation time of 100s. The data have been measured and analyzed by Sandkühler and Eblen-Zajjur (1994) who studied discharge patterns of spinal dorsal horn neurons under various conditions.

As an underlying model it is reasonable to choose Hawkes model, since it reflects the mutually exciting structure of neurons dependent on time. Hence, we estimated the Hawkes kernels with the presented estimation procedure (Figure 4.2). As discretization parameters we set $h = 0.5$ and $k = 20$. We can clearly see that most of the 100 link functions vanish and only few show significant peaks. To statistically significantly decide whether the link functions vanish or not a test has to be constructed in future work. As we have seen a Granger causality graph is induced by the estimated link functions, which is displayed in Figure 4.3. It is remarkable, that neuron nine is completely isolated from the other neurons.

Furthermore Figure 4.2 shows, that if there is a peak, then shape, time and intensity behave very similarly. The time instants of the peaks happen roughly around 17 ms with an intensity of approximately 0.38 spikes per millisecond.

Hawkes model has two drawbacks in this application. First it is only capable of modeling excitement, not inhibition, which is well known to play a major role in neuronal data. Secondly neurons possess a refractory period during which the neuron cannot fire again. This can be observed in Figure 4.2 where the link functions on the diagonal become negative leading to a conditional intensity that is below the level of the baseline intensity. Although this behaviour contradicts the conditions of the Hawkes model, the nonparametric estimates still yield meaningful results.

![Figure 4.2. Estimated link functions of Hawkes model based on neural spike train data.](image-url)
whereas incorporating refractory periods into the model would destroy its linear structure. We note that in many different applications such as modeling aftershock effects (Ogata 1999, Vere-Jones 1970, Vere-Jones and Ozaki 1982), insurgency in Iraq (Brantingham et al. 2011), crime (Mohler et al. 2011) or genome analysis (Reynaud-Bouret and Schbath 2010) the Hawkes model does not encounter these problems.

5. Concluding Remarks

In this paper we have investigated the structure of Hawkes models and proved that the Granger causality structure of the process is fully encoded in the corresponding kernels. Moreover we have defined a new nonparametric estimator of the Hawkes kernels based on a time-discretized version of the point process and an infinite order autoregression. The estimator is easy and fast to compute even for higher dimensions, which is of particular importance for the implementation of causal search algorithms that require fitting not only of the full model but also of many submodels.

Given the form of the conditional intensity, and in particular (4.1), the estimator is quite intuitive. However, on a second glance, it is surprising that the method really is consistent since in the limit the discretized time series consists mainly of zeros and some 1s. We have succeeded to establish consistency rigorously but failed up to now to prove asymptotic normality - although we are still convinced that asymptotic normality holds with a reasonably good rate. A closer inspection of the problems reveals that the structure of the discretized time series is quite different from usual infinite order AR-processes in that the innovation are (approximately) a heteroscedastic martingale difference sequence leading to severe technical problems. Furthermore, some terms in the calculations are of higher order as in the AR-case and do not disappear. For this reason we have postponed the proof of asymptotic normality to future work.

Appendix A. Proofs

Proof of Theorem 3.5. We consider stationary multivariate point processes on $\mathbb{R}$ while the proof of Didelez only covers processes on $[0, T]$ (or any compact interval). Therefore, strictly speaking, the result must be extended.
The log–likelihood of the process $N$ on $[t_0, t]$ is given by
\[ \sum_{i=1}^{d} \left[ \int_{t_0}^{t} \log \lambda_i(t) \, dN_i(t) - \int_{t_0}^{t} \lambda_i(t) \, dt \right] \]
where $\lambda_i(t)$ is given by (2.1) and only depends on $N_{cl(i)}(s), s \leq t$, where $cl(A) = A \cup \text{pa}(A)$. It follows that the likelihood can be factorized into factors that are $\mathcal{F}_C(t)$–measurable for sets $C \in \mathcal{C} = \{ cl(i) | i \in V \}$. The sets in $\mathcal{C}$ are complete in the moral graph $G^m$; combining factors with sets in the same clique we obtain a factorization over the cliques of the moral graph. The factorization prevails if we let $t_0$ tend to $-\infty$. This implies the global Markov property with respect to the moral graph.

Finally we note that every path from $A$ to $B$ is blocked by $B \cup C$ if and only if
\[ \text{pa}(B) \setminus (B \cup C) \not\propto A | B \cup C \]
in the moral graph $(G_{an(A \cup B \cup C)})^m$. This implies by the global Markov property that
\[ \mathcal{F}_{\text{pa}(B)}(t) \perp \mathcal{F}_A(t) | \mathcal{F}_{B \cup C}(t). \]

Finally we have for every $b \in B$
\[ \lambda_{A \cup B \cup C}^b(t) = \mathbb{E}(\lambda^V(t) | \mathcal{F}_{A \cup B \cup C}(t)) = \mathbb{E}(\lambda_{pa}^b(t) | \mathcal{F}_{A \cup B \cup C}(t)) 
= \mathbb{E}(\lambda_{pa}^b(t) | \mathcal{F}_{B \cup C}(t)) = \mathbb{E}(\lambda^V(t) | \mathcal{F}_{B \cup C}(t)) = \lambda_{B \cup C}^b(t). \]

Hence $N_A$ does not Granger cause $N_B$ with respect to the subprocess $N_{A \cup B \cup C}$. \hfill \Box

For the proof of Theorem 4.1, we need the following two technical lemmas.

**Lemma A.1.** Under the assumptions of Theorem 4.1 we have

(i) \[ \left\| \frac{1}{T_{h,k}} \sum_{t=k+1}^{T_h} (Y^h_{t-u} - p_N h) \right\|_2^2 = o_P \left( \frac{h^2}{T} \right); \]

(ii) \[ \left\| \hat{\Gamma}_{h,k} - \tilde{\Gamma}_{h,k} \right\|_2^2 = O_P \left( \frac{k^2 h^4}{T} \right) \quad \text{and} \quad \left\| \hat{\gamma}_{h,k} - \tilde{\gamma}_{h,k} \right\|_2^2 = O_P \left( \frac{k^4 h^4}{T} \right) \]

where $\hat{\Gamma}_{h,k}$ and $\hat{\gamma}_{h,k}$ are defined with $\hat{Y}^h$ and $\hat{Y}^{h,k}$ substituted by their mean $p_N h$.

**Proof.** For (i), we note that
\[ \sum_{t=k+1}^{T_h} (Y^h_{t-u} - p_N h) = \int_{kh}^{(k+1)h} d\hat{N}(t - hu). \]

This implies that
\[ \mathbb{E} \left\| \frac{1}{T_{h,k}} \sum_{t=k+1}^{T_h} (Y^h_{t-u} - p_N h) \right\|_2^2 \]
\[ = \frac{1}{T_{h,k}^2} \int_{kh}^{(k+1)h} \int_{kh}^{(k+2)h} \left( q_{NN}(t - s) + p_N \delta(t - s) \right) \, dt \, ds = O \left( \frac{h}{T_{h,k}} \right). \]

For the first part of (ii), we note that
\[ \left\| \hat{\Gamma}_{h,k} - \tilde{\Gamma}_{h,k} \right\|_2 = (\hat{Y}^{h,k} - (p_N h) \otimes I_d)'(\hat{Y}^{h,k} - (p_N h) \otimes I_d) \]
and hence
\[ \mathbb{E} \left\| \hat{\Gamma}_{h,k} - \tilde{\Gamma}_{h,k} \right\|_2 = \sum_{u=1}^{k} \mathbb{E} \left\| \frac{1}{T_{h,k}} \sum_{t=k+1}^{T_h} (Y^h_{t-u} - p_N h) \right\|_2^2 = O \left( \frac{k^2}{T_{h,k}} \right). \]
Proof. For the first two assertions, we note that $Y^h$ is a discretized version of a linear transform of the stationary point process $N$. Thus we obtain for the covariance function of $Y^h$

$$\text{cov}(Y^h_{u+1}, Y^h_t) = \int_0^h \int_0^h (q_{NN}(uh + t - s) + p_N \delta(uh + t - s)) \, dt \, ds$$

$$= \frac{1}{2\pi} \int_R \int_0^h \int_0^h \hat{q}_{NN}(\omega) e^{i\omega(t-s+uh)} \, d\omega \, dt \, ds + p_N h \delta(u)$$

and further with $H_h(\omega) = \int_0^h I_d e^{-i\omega t} \, dt$

$$= \frac{1}{2\pi} \int_R H_h(-\omega) \hat{q}_{NN}(\omega) H_h(\omega) e^{i\omega hu} \, d\omega + p_N h \delta(u)$$

$$= \frac{1}{2\pi} \frac{h}{\pi} \int_R H_h\left(\frac{-\omega}{h}\right) \hat{q}_{NN}\left(\frac{\omega}{h}\right) H_h\left(\frac{\omega}{h}\right) e^{i\omega u} \, d\omega + p_N h \delta(u)$$

$$= \frac{h}{2\pi} \int_{-\pi}^{\pi} \left[ \sum_{u \in \mathbb{Z}} H_1(-\omega - u) \hat{q}_{NN}\left(\frac{\omega + u}{h}\right) H_1(\omega + u) + \frac{1}{2\pi} p_N \right] e^{i\omega u} \, d\omega,$$

where we have used that $H_h\left(\frac{\omega}{h}\right) = h H_1(\omega)$. which implies

$$f_{Y^hY^h}(\omega) = h \sum_{u \in \mathbb{Z}} H_1(-\omega - u) \hat{q}_{NN}\left(\frac{\omega + u}{h}\right) H_1(\omega + u) + \frac{h}{2\pi} p_N.$$

Since each summand in the first term is positive definite, we find that

$$f_{Y^hY^h}(\omega) \geq \frac{h}{2\pi} \min\{p_{N,1}, \ldots, p_{N,d}\} I_d$$

for all $\omega \in [-\pi, \pi]$ and all $h > 0$. This proves that

$$\|\Gamma_{h,k}^{-1}\| = O(h^{-1})$$

for $h \to 0$ (the bound does not depend on $k$). Furthermore, under the assumptions on the link function the spectrum of $N$,

$$f_{NN}(\omega) = (I_d - \Phi(\omega))^{-1} D_N (I_d - \Phi(-\omega))^{-1}$$

and hence $\hat{q}_{NN}(\omega) = f_{NN}(\omega) - \frac{1}{2\pi} D_N$ is uniformly bounded for all $\omega \in \mathbb{R}$. Since $|H_1(\omega + u)|^2$ satisfies

$$\sum_{u \in \mathbb{Z}} |H_1(\omega + u)|^2 \leq C$$

for all $\omega \in [-\pi, \pi]$, we get $\|\Gamma_{h,k}\| \leq \|f_{Y^hY^h}(\omega)\| \leq C h$ for some constant $C > 0$. □
For the third assertion, we first use the triangle inequality to get
\[ \| \tilde{\Gamma}_{h,k} - \Gamma_{h,k} \|_2 \leq \| \hat{\Gamma}_{h,k} - \tilde{\Gamma}_{h,k} \|_2 + \| \hat{\Gamma}_{h,k} - \Gamma_{h,k} \|_2. \]
Since by Lemma A.1 the first term has the required order, it suffices to prove the assertion for the second term. We note that \( \mathbb{E}(\tilde{\Gamma}_{h,k}) = \Gamma_{h,k} \) and \( \mathbb{E}(\tilde{\Gamma}_t) = 0 \). Thus we obtain by the product formula for cumulants
\[
\mathbb{E}(\| \hat{\Gamma}_{h,k} - \Gamma_{h,k} \|^2) = \frac{1}{\hat{\Gamma}_{h,k}} \sum_{i_1,\ldots,i_4=1}^d \sum_{t,u_1,\ldots,u_4=1} \sum_{t,s=k+1}^T \text{cum}((\tilde{\Gamma}_{i_1,t-u_1}^{i_2,t-u_2}^{i_3,s-u_3}^{i_4,s-u_4}) + \text{cum}((\tilde{\Gamma}_{i_1,t-u_1}^{i_2,t-u_2}^{i_3,s-u_3}^{i_4,s-u_4}) + \text{cum}((\tilde{\Gamma}_{i_1,t-u_1}^{i_2,t-u_2}^{i_3,s-u_3}^{i_4,s-u_4}))].
\]
By (2.5) we find that all cumulants summed over \( t \) are at most of order \( O(h) \) which yields the required order.

The fourth result can be derived from (iii) similarly as in the proof of Lemma 3 of Berk (1974) by noting that
\[ \| \hat{\Gamma}_{h,k}^{-1} - \Gamma_{h,k}^{-1} \|_2 = \| \hat{\Gamma}_{h,k}^{-1}(\hat{\Gamma}_{h,k} - \Gamma_{h,k})\Gamma_{h,k}^{-1} \|_2 \leq \left( \| \hat{\Gamma}_{h,k}^{-1} - \Gamma_{h,k}^{-1} \| + \| \Gamma_{h,k}^{-1} \| \right) \| \hat{\Gamma}_{h,k} - \Gamma_{h,k} \|_2. \]
Rewriting this as
\[ (1 - \| \Gamma_{h,k}^{-1} \| \| \hat{\Gamma}_{h,k} - \Gamma_{h,k} \| ) \| \hat{\Gamma}_{h,k}^{-1} - \Gamma_{h,k}^{-1} \|_2 \leq \| \Gamma_{h,k}^{-1} \|^2 \| \hat{\Gamma}_{h,k} - \Gamma_{h,k} \|_2, \]
we obtain the required convergence since by (ii) and (iii) \( \| \Gamma_{h,k}^{-1} \| \| \hat{\Gamma}_{h,k} - \Gamma_{h,k} \| = O_P(\sqrt{k^2 \frac{T}{T}}) = O_p(1) \) and \( \| \Gamma_{h,k}^{-1} \|^2 \| \hat{\Gamma}_{h,k} - \Gamma_{h,k} \| = O_P(\sqrt{k^2 \frac{T}{T^2}}) \).

For (v), we finally note that
\[ \| \hat{\Gamma}_{h,k}^{-1} \| \leq \| \Gamma_{h,k}^{-1} \| + \| \hat{\Gamma}_{h,k} - \Gamma_{h,k} \| = O_P(h^{-1}) \]
by (ii) and (iv).

In the following, \( A \otimes B \) will denote the Kronecker product of matrices \( A \) and \( B \) with suitable dimensions.

Proof of Theorem 4.1. First of all we note that assumption 2 assures the stationarity of the process \( N \) and hence of the processes \( Y^h \) for all \( h > 0 \). For notational convenience, let \( dN(t) = dN(t) - p_N dt \). We start by rewriting
\[ \tilde{\phi}^{h,k} - \phi^{h,k} = (\tilde{\phi}^{h,k} - \phi^{h,k}) \hat{\Gamma}_{h,k}^{-1} = (\tilde{\Gamma}_{h,k} - \phi^{h,k} \hat{\Gamma}_{h,k}) \hat{\Gamma}_{h,k}^{-1} + (\tilde{\phi}^{h,k} - \phi^{h,k}) \hat{\Gamma}_{h,k}^{-1} + \phi^{h,k} (\hat{\Gamma}_{h,k} - \Gamma_{h,k}) \hat{\Gamma}_{h,k}^{-1}. \]
By Lemmas A.1 and A.2 the last two terms are in Euclidean norm at most of order \( O_P(kh/T) \) and thus converge to zero in probability. For the first term, we get
\[ (\tilde{\phi}^{h,k} - \phi^{h,k}) \hat{\Gamma}_{h,k}^{-1} = \frac{1}{\hat{\Gamma}_{h,k}} \sum_{t=k+1}^T \epsilon_t^{h} (\tilde{\Gamma}_t^{h,k}) \hat{\Gamma}_{h,k}^{-1}.
\]
where \( \epsilon^{h,k}_t = \tilde{Y}^{h}_t - \phi^{h,k}_t Y^{h,k}_t \). This yields the upper bound
\[
\|\hat{\phi}^{h,k} - \phi^{h,k}\|_2 \leq \|\tilde{Y}^{h}_{t-h} - \mathbb{E}(\tilde{Y}^{h}_t | \mathcal{F}_{t-h})\|_2 \leq \big(\|U_{1,T}\| + \|U_{2,T}\| + \|U_{3,T}\|\big) + o_P(1) \tag{A.1}
\]
with
\[
U_{1,T} = \frac{1}{T_{h,k}} \sum_{t=k+1}^{T_h} \big[ \tilde{Y}^{h}_t - \mathbb{E}(\tilde{Y}^{h}_t | \mathcal{F}_{t-h}) \big] (\hat{Y}^{h,k}_t)' ,
\]
\[
U_{2,T} = \frac{h}{T_{h,k}} \sum_{t=k+1}^{T_h} \bigg( \int_0^h \phi(u) d\tilde{N}(t h - u) - \sum_{u=1}^k \phi(h u) Y^{h}_t \bigg) (\hat{Y}^{h,k}_t)' ,
\]
\[
U_{3,T} = \frac{h}{T_{h,k}} \sum_{t=k+1}^{T_h} \bigg( \int_{h k}^\infty \phi(u) d\tilde{N}(t h - u) \bigg) (\hat{Y}^{h,k}_t)' ,
\]
where we have used that \( p_N = \nu + \int \phi(u) du N \). We proceed in showing that the three terms \( U_{1,T}, U_{2,T} \) and \( U_{3,T} \) are of order \( o_P(h) \). Together with Lemma A.2 (ii) this proves \( \|\hat{\phi}^{h,k} - \phi^{h,k}\|_2 = o_P(1) \).

Starting with the first term \( U_{1,T} \), we find
\[
\mathbb{E}\|U_{1,T}\|^2 = \frac{1}{T_{h,k}^2} \mathbb{E}\bigg( \bigg[ \sum_{t=k+1}^{T_h} \big[ \tilde{Y}^{h}_t - \mathbb{E}(\tilde{Y}^{h}_t | \mathcal{F}_{t-h}) \big] (\hat{Y}^{h,k}_t)' \bigg]^2 \bigg) = \frac{1}{T_{h,k}} \sum_{t=k+1}^{T_h} \mathbb{E}\bigg( \bigg[ \tilde{Y}^{h}_t - \mathbb{E}(\tilde{Y}^{h}_t | \mathcal{F}_{t-h}) \bigg]' (\tilde{Y}^{h,k}_t)' \bigg) = \frac{1}{T_{h,k}} \sum_{t,k=1}^{T_h} \mathbb{E}\bigg[ \|\tilde{Y}^{h}_t - \mathbb{E}(\tilde{Y}^{h}_t | \mathcal{F}_{t-h})\|_2^2 \|\hat{Y}^{h,k}_t\|_2^2 \bigg] \leq \frac{C h}{T_{h,k}}.
\]

Thus \( h^{-1} \|U_{1,T}\| \) converges to zero in probability according to assumption (i).

For the second term in (A.1) we obtain
\[
\mathbb{E}\|U_{2,T}\| \leq \frac{h}{T_{h,k}} \sum_{t=k+1}^{T_h} \bigg( \mathbb{E}\bigg[ \bigg[ \sum_{u=1}^k \int_0^h (\phi(h u + \alpha) - \phi(h u)) d\tilde{N}(t h - u - \alpha) \bigg]^2 \bigg] \bigg)^{1/2} \|\tilde{Y}^{h}_t\|_2^2 \bigg)^{1/2}.
\]

By Lemma A.2 we obtain for the second mean \( \mathbb{E}\|\tilde{Y}^{h,k}_t\|^2 = \text{tr} \Gamma_{h,k} = O(h k) \). For the first mean, we have
\[
\mathbb{E}\bigg[ \sum_{u=1}^k \int_0^h (\phi(h u + \alpha) - \phi(h u)) d\tilde{N}(t h - u - \alpha) \bigg]^2 \leq \frac{k^2}{h} \int_0^h \int_0^\infty (\phi(h u + \alpha) - \phi(h u))^2 d\tilde{N}(t h - u - \alpha) d\tilde{N}(t h - u - \alpha)
\]
\[
\times \bigg( \phi(h u + \beta) - \phi(h u) \bigg)^2 d\alpha + \frac{k}{h} \int_0^h \int_0^\infty (\phi(h u + \alpha) - \phi(h u)) D_N(\phi(h u + \alpha) - \phi(h u)) d\alpha d\beta \leq C h^2 \int_0^h \int_0^\infty \|q_N(\alpha - \beta)\|_1 d\alpha d\beta + C h^3 k \|D_N\|_1 \leq C h^3 k.
\]

Combining the results, we find \( \mathbb{E}\|U_{2,T}\| = O(h^3 k) = o(h) \).
For the last term in (A.1) we obtain similarly as for $U_{2,T}$

\[
\mathbb{E} \left\| \int_{h_k}^{\infty} \phi(u) \left( dN(th - u) - p_N du \right) \right\|^2_2 \\
\leq \int_{h_k}^{\infty} \int_{h_k}^{\infty} \| \phi(u) \| \| \phi(v) \| \| q_{N}(u - v) \|_2 du dv + \int_{h_k}^{\infty} \| \phi(u) \|^2 \| D_N \|_2 du = O(1/hk),
\]

where we have used assumption (iii). Together with $\mathbb{E} \| \tilde{Y}_h \|_2^2 = O(hk)$ this yields $\mathbb{E} \| U_{3,T} \|_2 = o(h)$.

Finally, we note for the consistency of $\hat{\nu}^{h,k}$ that

\[
\hat{\nu}^{h,k} - \nu^h = \frac{1}{T_{h,k}} \sum_{t=k+1}^{T_h} (Y_t^h - \nu^h - \hat{\phi}^{h,k} Y_t^h) \\
= \frac{1}{T_{h,k}} \sum_{t=k+1}^{T_h} (\tilde{Y}_t^h - \mathbb{E}(\tilde{Y}_t^h \mid \mathcal{F}_{h(t-1)})) + \frac{1}{T_{h,k}} \sum_{t=k+1}^{T_h} (\hat{\phi}^{h,k} - \hat{\phi}^{h,k}) \tilde{Y}_t^h \\
+ \frac{h}{T_{h,k}} \sum_{t=k+1}^{T_h} \sum_{k=1}^{k} \int_0^h \left( \phi(hu + \alpha) - \phi(hu) \right) dN(t(h - u) - \alpha) \\
+ \frac{h}{T_{h,k}} \sum_{t=k+1}^{T_h} \int_{h_k}^{\infty} \phi(u) dN(th - u) + o_P(h)
\]

Convergence to zero of all four terms follows by similar arguments as above. \( \square \)

**References**


