INAUGURAL – DISSERATION

zur

Erlangung der Doktorwürde

der

Gesamtfakultät für Mathematik, Ingenieur- und Naturwissenschaften

der

Ruprecht – Karls – Universität

Heidelberg

vorgelegt von

Marilena Müller, MSc

aus Gehrden

Nonparametric Estimation of Locally Stationary Hawkes Processes

Betreuer: Prof. Dr. Enno Mammen

Zusammenfassung

In dieser Dissertation betrachten wir multivariate Hawkes-Prozesse mit bedingten Intensitäten und Reproduktionsfunktionen, die von der Zeit abhängig sind. In diesem Fall wird das Modell durch die bedingte Intensitätsfunktion $\lambda = (\lambda^1, \ldots, \lambda^d)^T$ charakterisiert, die zu einem Zeitpunkt $t \in [0, T]$ durch

$$\lambda\left(\frac{t}{T}\right) = \nu\left(\frac{t}{T}\right) + \int_{-\infty}^{t-} \mu(t-s, t/T) \, \mathrm{d}N_s$$

für eine in dem Zeitfenster $[0, T]$ zu beobachtende multivariate Realisation des Hawkes-Prozesses $N = (N^{(1)}, \ldots, N^{(m)})^T$.

Dadurch wird eine Klasse von lokal stationären Prozessen definiert. Ein Verfahren zur Schätzung des Vektors von zeitabhängigen Immigrationsfunktionen ν und des Vektors der Reproduktionsfunktionen μ, welches auf einem lokalen Kriterium beruht, wird vorgestellt. Die Theorie über stationäre Hawkes-Prozesse wird um asymptotische Theorie über den Schätzer im lokal stationären Modell erweitert. Simulationsstudien runden die Überlegungen ab.

Im zweiten Teil betrachten wir lokal stationäre Hawkes-Prozesse, die sich in bestimmten Punkten ähneln. Die vorangegangene Arbeit ermöglicht es nun, Testergebnisse zu formulieren. Wir beobachten zwei Hawkes-Prozesse und testen, ob sie zu einem festen Zeitpunkt Realisierungen derselben Immigrations- und Reproduktionsfunktionen sind. Alternativ, aber inhaltlich identisch, könnte man auch das Verhalten eines Hawkes-Prozesses zu zwei verschiedenen Zeitpunkten vergleichen. Dies ermöglicht die Behandlung von Saisonalität, was ein typisches Problem der Anwendung darstellt.

Abstract

In this work we consider multivariate Hawkes processes with baseline conditional intensities and reproduction functions that depend on time. In this case, the model is characterized via the conditional intensity function $\lambda = (\lambda^1, \ldots, \lambda^d)^T$ at a time point $t \in [0, T]$ given by

$$\lambda\left(\frac{t}{T}\right) = \nu\left(\frac{t}{T}\right) + \int_{-\infty}^{t-} \mu(t-s, t/T)\, \mathrm{d}N_s$$

for a multivariate realization of the Hawkes process $N = (N^{(1)}, \ldots, N^{(m)})^T$, which is observed in a time frame $[0, T]$.

Thus, a class of locally stationary processes is defined. The discussed estimation procedure of the vector of time-dependent baseline intensities ν and vector of reproduction functions μ is grounded on a localized criterion. Theory on stationary Hawkes processes is extended to develop asymptotic theory for the estimator in the locally stationary model. Simulation studies round off the considerations.

In the second part we consider the option for local alignment of locally stationary Hawkes processes. The previous work enables the possibility to formulate testing results. We observe two Hawkes processes and test whether they are realizations of the same underlying immigration and reproduction functions at a fixed point in time. From an alternative point of view, but identically contentwise, one could compare the behaviour of one Hawkes process at two distinct time points. This enables reasearch sorrounding seasonarity.

Acknowledgements

First and foremost, I want to thank my supervisor, Enno Mammen, for initially giving me the opportunity to pursue this work. Over the progression of research and development I always highly valued the academic and personal help I was given. I have great appreciation and thankfulness for all the work invested in me.

It is of great importance to me to give special thanks to many (former) members of the Statistics group in Heidelberg who welcomed me to the world of science and inspired me with their own successes and advise, as well as my current colleagues who gave rise to a fruitful and supporting atmosphere. Especially, I would like to thank Xavier Loizeau, Stefan Richter, Joseph Meyer, Ricardo Blum, Sergio Brenner Miguel, Bianca Neubert, Nathawut Phandoidaen, Rainer Dahlhaus and Jan Johannes.

I thank the DFG for making it possible to do the resulting reasearch by financing the Research Training Group 1953 "Statistical Modeling of Complex Systems" in Heidelberg and Mannheim and thus also opening up various opportunities for me to connect socially and academically.

Finally, I want to value the personal support by Stefanie Volz, Emil Volz and Christiane Bös that I have enjoyed over the years, which rounded off these very influencial times. Thank you.

Contents

1 Introduction

As researchers in this field the goal consists of formalizing and modelling certain behaviours and progressions we detect in our lives and nature itself. More specifically, we observe some events in a certain frame of time. In this explicit case we use these observations and try to model this behaviour with a stochastic process on the real line or a subset of it. In our further implications it is notable that we choose a model which is inherintly not reproducable as we only observe it once. This is natural as we would not be able to recreate a lot of phenomenons but would still like to use the data to further our knowledge. On the other hand as statisticians we would like to repeat any experiment as often as possible to get asymptotic results, which enables us to use well known tools.
In this work we develop theory for a specific point processes, meaning the data consists of designated points in time and the goal is to make inference to deduce properties and significant behaviors. Point processes themselves have generated and inspired a vast number of scientific fields in the past and the present. In this specific thesis we devote ourselves to an explicit model, while leaving flexibilty to the model via a nonparametric approach.

1.1 Motivation

An initial idea and inspiration for the upcoming formulation of the problem can be found in an example from epidemiology. We consider the spreading of an infectious disease. Each time a person gets infected then corresponds to an event in the model. Each infectious patient then triggers additional events by infecting hitherto healthy patients. The progression of the infection can then be understood as a whole series of events, and thus might be modelled quiet naturally via stochastic point processes. An underlying strong component of dependency can be detected as the events in the future depend on the events in the past. This results in a significantly harder problem to solve. Whenever

an infectious person enters the community or a member of the observed community gets infected from the exterior, we call this *immigration* of an event. Each immigrating infection triggers new events (*reproduction*) and hence a so-called cluster process is giving the observation a hidden structure which we aim to use to our benefit.
A multivariate setting is obviously intuitive, as the developement of one disease might influence the progression of others. On the other hand, different communities might be considered which affect each other, giving rise to the idea of cross-dependency.
The aim is to find a suitable notion of interdependency using the notion of a so-called (conditional) intensity function as a mean of probabilty for an event to occur in the next upcoming timeframe given the whole history or past of the observed process.
We formalize a variation of the Hawkes process model. This model consists of the merger of two ideas. We consider new events to emerge as either triggered by the immigration of an insect or the reproduction of insects. Thus, we define a baseline conditional intensity, call it ν and a reproduction function μ, also called fertility function or excitement function, which models the influence of recent jumps or events on the imminent conditional intensity. Finally we formalize the model more explicitely, as initially named by and introduced in Hawkes (1971).
Denote by $N(t) = (N^1(t), ..., N^d(t))^\intercal$ for $t \in [0, T]$ the multivariate Hawkes Process, which is characterized by its vector of intensity functions

$$\lambda^l(t) = \nu^l + \sum_{m=1}^{d} \int_{-\infty}^{t-} \mu^{l,m}(t-s)\, \mathrm{d}N_s^{(m)} \tag{1.1}$$

with $l = 1, ..., d$ and for a certain time point $t \in [0, T]$. Here ν^l is a positive constant and $\mu^{l,m}$ for $1 \le l, m \le d$ are some positive functions defined on $\mathbb{R}_0^+$.
One could interpret the model as follows. An infectious member enters the community or gets infected from the outside with probability ν and thus enhences the function λ representing a mean expectation of passing on the infection in the time point t corresponding to the present. For each infection at a past point in time s, i.e. $s < t$, λ grows with a factor generated by μ via the passed by time $t - s$.

1.2 Contribution and Structure of the Thesis

There already is a wide range of literature concerning the case of Hawkes processes as defined in (1.1). In this work the set up differs from the original in the sense that the baseline conditional intensity ν and reproduction function μ are generalized in the following way. Inspired by the example of the spreading of an infection, we propose that the immigration as well as the reproduction should vary over time. In our example, it could be noted that most diseases are prone to spread faster in certain seasons or even only occur in specific times of the year. This effect is called *seasonality* and calls for a more introcate and challenging model, affecting the theoretical manageability as we will see in the following pages.

The original case allows for a notion of stationarity, as will be described in detail later. To make room for this dependence structure we use a locally stationary specification where we assume that the process can be approximated locally around a fixed time point by a stationary Hawkes process.

There are many other similar prominent examples, i.e. lineages of insects. Imagine considering a population of insects. Each birth can be interpreted as an event or a designated point in time. During an insect's lifespan it possibly brings forth additional births of insects, hence new events. Each immigrating insect triggers the forming of a family tree. Again, we find that the development of certain species might have an effect on other species, enabling the wish to formulate cross dependency in a multivariate setting. Concerning seasonality, one might imagine insects immigrating only in certain seasons of the year, which might hold for the reproduction of insects as well.

Making the appropriate changes, in this work we will consider a model for multivariate Hawkes processes $N(t) = (N^1(t), ..., N^d(t))^\intercal$ for $t \in [0, T]$ of the following kind. The multivariate point process is modelled via the conditional intensity function given by

$$\lambda^l(t) = \nu^l\left(\frac{t}{T}\right) + \sum_{m=1}^{d} \int_{-\infty}^{t-} \mu^{l,m}\left(t - s; \frac{t}{T}\right) \mathrm{d}N_s^{(m)} \tag{1.2}$$

with $l = 1, ..., d$ and $t \in [0, T]$ or in vector notation

$$\lambda(t) = \nu\left(\frac{t}{T}\right) + \int_{-\infty}^{t-} \mu\left(t - s, \frac{t}{T}\right) \mathrm{d}N_s.$$

Now, ν^l and $\mu^{l,m}$, where $1 \leq l, m \leq d$, are positive functions defined on $[0, 1]$ and $\mathbb{R}_0^+ \times [0, 1]$, respectively. We will use an asymptotic framework where

$T \to \infty$.

We thus found a model to be investigated. The object of interest λ can give relevant information about when to interfere in the process to take necessary steps to prevent the spreading of said diseases.
In Chapter 2, we formally introduce notions of appropriate point processes and become familiar with fundamental tools, essential features and properties concerning the structures of interest.

While our approach shows a different perspective, many valid strategies are led by other views and great ideas, which are shortly presented in Section 1.3.

The asymptotic framework we will work with will be introduced in the beginning of Chapter 3. For locally stationary Hawkes processes we will propose a nonparametric estimator for the baseline conditional intensity and the reproduction function and develop the asymptotic theory in the following. Consistency of the resulting estimator is shown. The results are demonstrated and displayed in a simulation study. The proofs are found hereafter and we finish the chapter with a discussion of applicability and suitable assumptions.
In the second chapter of theoretical work, (chapter 4) we modify the setting slightly. This additional part of the developed theory is devoted to testing procedures concerning whether or not two Hawkes processes coincide at a certain time point. This is equivalent to testing if one realization of the same underlying process coincides at two different points. This is motivated by e.g. questioning whether or not the reproduction is the same each summer. Again, discussions follow.
The thesis finishes with a short summary and outlook on further research to be made.

1.3 Embedding in Literature

By now there is not only a wide variety of theoretical work, but many fields of application for Hawkes process models as well.
For all the ground work and fundamental work of point processes in general, see Andersen (1993) and Daley and Vere-Jones (1988).
Initially proposed and namegiving of the model is the work of Hawkes (1971). The model is introduced and elemental properties are discovered and despicted. In Hawkes and Oakes (1974) the crossing to cluster processes is made and further investigated.

Considering inference on the stationary model we find a vast amount of literature. For stationary nonparametric multivariate Hawkes processes, Bacry, Dayri and Muzy (2012) are considering the reproduction function of a multivariate nonparametric Hawkes process by estimating the covariance of increments of the Hawkes process and using spectral methods. In Bacry, Dayri and Muzy (2016) estimates of reproduction functions of multivariate Hawkes processes are discussed via solving empirical integral equations with estimated average conditional intensity vectors and estimated conditional laws. This approach is used in Rambaldi, Bacry and Lillo (2016) for the study of order book dynamics. The papers Kirchner (2016) and Kirchner (2017) relate Hawkes processes to integer valued autoregressive processes of infinite order INAR(∞). They approximate these processes by integer valued autoregressive processes of finite order, $p < \infty$, INAR(p), and use methods from statistical inference for INAR(p) processes. In Eichler, Dahlhaus and Dueck (2017) a nonparametric estimator of reproduction functions is proposed for multivariate Hawkes processes based on discretisations of the process. Its consistency is shown and the estimator is used for causal inference. The paper Bai, Chen and Chen (2015) discusses observations of n independent Hawkes processes with constant baseline conditional intensity and nonparametric reproduction function. The paper shows rates of convergence for a nonparametric maximum likelihood sieve estimator and proves asymptotic normality for the parametric estimator of the baseline conditional intensity. The papers Hanssen, Reynaud-Bouret and Rivoirard (2015) and Reynaud-Bouret and Schabt (2010) prove exponential inequalities and tail bounds for statistics of Hawkes processes and apply the results for asymptotic analysis of adaptive and LASSO-estimation of Hawkes processes with nonparametric reproduction function. In the estimation procedure the reproduction function is approximated by a linear combination of functions from a dictionary. For work on nonparametric Bayesian estimation of Hawkes processes, see Donnet, Rivoirard and Rousseau (2018).

There is some work to allow for nonstationary models. The definition of locally stationary Hawkes processes introduced in Roueff, von Sachs and Sansonnet (2016), Roueff and von Sachs (2019) differs from the notion studied here. In particular, they make use of the later discussed cluster representations of one-dimensional Hawkes processes to define spatial locally stationary point processes that generalizes one-dimensional Hawkes processes to higher dimensions. They also relate the definition of their notion of locally stationary processes to the general class of non-stationary Hawkes processes.
The papers Chen and Hall (2013) and Chen and Hall (2015) allow for a

varying baseline conditional intensity. They study estimates in an asymptotic framework where the baseline conditional intensity is multiplied by a factor that converges to infinity. In the asymptotic theory the time horizon is kept fixed. The papers discuss models with parametric reproduction functions and they allow for parametric and nonparametric specifications of the baseline conditional intensity. The work of Omi, Hirata and Aihara (2017) delivers a Bayesian approach for models with time varying baseline conditional intensity. The paper Clinet and Potiron (2018) considers parametric Hawkes processes with the reproduction rate modeled as an exponential function. They allow for the baseline conditional intensity and the parameters of the exponential function to depend on time. In Roueff, von Sachs and Sansonnet (2016), Roueff and von Sachs (2019) a new class of locally stationary multi-dimensional Hawkes processes is proposed. Nonparametric estimation is discussed that is based on local Bartlett spectra. The approach allows one to compute approximations of first and second order moments.

For the application side we see a vast number of applications are found in finance, see Aït-Sahalia, Cacho-Diaz and Laeven (2015) as an example and Bacry, Mastromatteo and Muzy (2015) for an overview. The main idea here lies in the assumption, that limit order books behave like Hawkes processes. In that sense whether you buy or sell stocks influences the buying and selling the nondistant future.
A very interesting usage of the Hawkes model is found in seismology, see Vere-Jones (1970), where the occurance of earthquakes is modeled. The first shocks are innitiated via the immigrationrate and further aftershockes are triggered hence via the reproductionfunction.
We find other applications in the usage of social networks or social media. News and messages get written and shared and influence each other respectively. Examplary, this is despicted in statistical modeling of e-mail networks in Fox, Short, Schoenberg, Coronges, and Bertozzi (2015).
Additionally, application in crime analysis (see Mohler, Short, Brantigam, Schoenberg and Tita (2011)) and genome analysis (see Reynaud-Bouret and Schabt (2010)) are very well worth mentioning.

2 Fundamentals, Basics and Notation

Starting off, a formal introduction of point processes is neccessary. A trail of features and desired properties will lead us to the notion of a Hawkes process, which we will introduce later on. Any fundamental abilities and limits considering the up-coming problems of the following chapers will be discussed and cited.
Definitions and elementary statements about point processes are mainly obtained from work by Daley and Vere-Jones (1988) and Andersen (1993), while some are picked from more applicable work of Hautsch (2012). The more involved segment about Hawkes processes then orignates from the works of Hawkes (1971) and Hawkes and Oakes (1974).

2.1 Point Processes

We first let our intuition lead us to what a point process should be and how we can fomulate it appropriately.

Definition 2.1.1. *A point process on $\mathbb{R}$ is a set of random points $P = \{t_j \in \mathbb{R} : j \in \mathbb{Z}\ , t_j \leq t_{j+1}$ for all $j\}$. Always assume $t_{-1} \leq 0 \leq t_0$.*

An embedding into the notion of measure theory seems intuitive and useful. Hence, the following definition is motivated. This will open up the possibility to use well-known terms and measure theory as well as existing results of probability theory.

Definition 2.1.2. *The (to a point process P corresponding) counting process $N = (N(t))_{t\in\mathbb{R}}$ is given by*

$$N(t) := \sum_{i\in\mathbb{Z}} \mathbb{1}_{\{t_i \leq t\}}(t).$$

N is right-continuous. A point proces is characterized by its counting process. The natural filtration $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{R}}$ is given by

$$\mathcal{F}_t := \sigma(N(s), s \leq t).$$

This very well-known notion of point processes now enables us to express similar expressions and assertions as we would for existing and established theory.
As a side note, the following notion is often used as well.

Definition 2.1.3. *From a different pespective define the counting measure of a point process P by*

$$N(A) := \#(P \cap A)$$

for $A \in \mathcal{B}(\mathbb{R})$.

Note, that the notation $N(t)$ refers to $N((-\infty, t])$ in the setting of observing the counting process N over time up until a point t.

We see that here a point process is charakterized by the number of events one finds in an intervall. This allow for serveral ideas of sliding windows over time as an equivalent to a pointwise shifting. This correspondence will open up the possibility to form a suitable alternative definiton of stationarity.

Remark. *The finite dimension distributions, i.e. the joint distributions of*

$$N(I_1), \ldots, N(I_k),$$

where $k \in \mathbb{N}$ and bounded intervals $I_i \subset \mathbb{R}$, $i = 1, \ldots, k$, characterize P.

For a formal definition of stationarity for counting processes and all many up-coming properties, we make the following assumption.

Definition 2.1.4. *A point process P is called orderly, if $t \neq t'$ for all $t, t' \in P$.*
Always assume $t_{-1} < 0 \leq t_0$.

From now on we will only consider orderly processes, which allow for finiteness, in the sense that multiplicity of deisgnated points implies the possibility for singularities. This would impede theoretical results. One could argue that it is even reasonable. Consider the example from seismology. We observe earth quakes and their aftershocks. In a reasonable experiment two earthquakes, i.e. two simultaneously triggered quakes, would be indistinguishable from the sole observation.

The general assumption, that two events can never really happen at the exact same point in time, might be agreeable as well.

Now, we want to formulate a suitable definition for stationarity in the observed processes.

Definition 2.1.5. *A point process is called stationary, if for any $k \in \mathbb{N}$ and $t_1, \ldots, t_k, \tau \in \mathbb{R}$ it holds that*

$$\mathcal{L}\big((N(t_1+\tau), \ldots, N(t_k+\tau))\big) = \mathcal{L}\big((N(t_1), \ldots, N(t_k))\big).$$

2.2 The Intensity Function

The strucutre we oberve does not directly show the object of interest. The generated data is merely a realization of the function of interest, i.e. the intersity function, which will be introduced below.
As motivated in the last section 2.1, we will use a notion of infinitesimally small intervalls in the imminent future to underlie our point processes and perform inference. This leads us to the structure of interest, i.e. the underlying model specification.

Definition 2.2.1. *Define by $H_t := \{t_j \in P : t_j < t\}$ the history of P at time t. Define the complete intensity function or hazard function of P, denoted by $\Lambda(t, H_t)$ as follows:*

$$\lim_{\delta \to 0} \delta^{-1} \mathbb{P}(N(t+\delta) - N(t) = 1 \mid H_t) = \lambda_{com}(t, H_t),$$
$$\lim_{\delta \to 0} \delta^{-1} \mathbb{P}(N(t+\delta) - N(t) = 0 \mid H_t) = 1 - \lambda_{com}(t, H_t).$$

As mentioned earilier, one shall keep in mind, that the intensity function does not characterize the point process.
The conditional intensity function of P is given by

$$\lambda(t, H_t) := \lim_{\delta \to 0} \delta^{-1} \mathbb{E}[N(t+\delta) - N(t) |\ H_t] = \frac{\mathbb{E}[\mathrm{d}N(t) |\ H_t]}{\mathrm{d}t}$$

This definition suggests that Point processes of interest would have the significant porperty to be highly influenced by its history. This is nort necessarily the case. We will go on by defining a vast general class of point processes not depending on their history. Our final outlook and the center of this thesis focuses on point processes of such a form.

Concerning notation the argument of the history, i.e. the conditioning of the past events is often omitted in writing.

The following notation is often found and is consistent with the one used up until now. We define the (conditional) intesity function as a measure. The *intensity measure* $\lambda^N : \mathcal{B}(\mathbb{R}) \to \mathbb{R}$ of a point process N is given by $\lambda^N(B) = \int_B \lambda(u, H_u)\ \mathrm{d}u$, where λ denotes the intensity function of N.
These notations are interchangeable.

Definition 2.2.2. *A point process is called simple if its conditional intensity function in locally finite and non-atomic. Orderly point processes are simple. A measure μ is called non-atomic if there do not exist atoms, i.e. measurable sets A such that for all $B \subset A$ with $\mu(B) < \mu(A)$ it holds that $\mu(B) = 0$.*

We shall only be concerned with processes for which $\Lambda(\cdot, \cdot)$ is well-defined and finite with probability 1.
The (conditional) intensity function does not necessarily characterize the process.

Lemma 2.2.3. *Given a point process N with well-defined (conditional) intensity function $\lambda(\cdot, \cdot)$ we see that*

$$\left(N((0,t]) - \int_0^t \lambda(s, H_s)\ \mathrm{d}s \right)_{t \geq 0}$$

is a martingale with resect to the canonical filtration of the process $(N((0,t]))_{t\geq 0}$.

2.3 Martingale Theory

Lemma 2.2.3 motivates us to make use of matingale properties in our setting. Thus, we formally introduce martingales and features of relevance.

Definition 2.3.1. *Let $M = (M_t)_{t\in\mathbb{R}}$ be an adapted stochastic process w.r.t. the filtration $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{R}}$ with $\mathbb{E}[M_t] < \infty$ for all $t \in \mathbb{R}$.*
M is called a $\mathcal{F}$-martingale, if for $s < t < \infty$

$$\mathbb{E}[M_t|\mathcal{F}_s] = M_s \quad a.s.$$

and M is called a $\mathcal{F}$-submartingale, if for $s < t < \infty$

$$\mathbb{E}[M_t|\mathcal{F}_s] \geq M_s \quad a.s.$$

In the work of this thesis we wil be confronted with terms of the form $\int g(s)\, dM_s$ for some function g and a martingale M generated via Lemma **??**. From classical martingale theory we have a vast knowledge of asymptoctics of such forms including limiting behavior, which we will use to our advantage.

Remark. *Any counting process is a submartingale w.r.t. to its history.*

We recall a very well-known property, which we are to use considerably often.

Theorem 2.3.2 (Doob-Meyer decomposition)**.** *Any $\mathcal{F}$-submartingale $(N(t))_{t\in\mathbb{R}}$ with can be decomposed into a unique zero-mean $\mathcal{F}$-martingale $(M(t))_{t\in\mathbb{R}}$ and a unique $\mathcal{F}$-predictable cumulative process $(\tilde{\Lambda}(t))_{t\in\mathbb{R}}$, i.e.*

$$N(t) = M(t) + \tilde{\Lambda}(t)$$

for $t \in \mathbb{R}$. $(\tilde{\Lambda}(t))_{t\in\mathbb{R}}$ is called the compensator.

Lemma 2.3.3. *If the considered submartingale is a counting process w.r.t. the history H the compensator is given by*

$$\tilde{\Lambda}(t) = \int_{-\infty}^{t} \lambda(t, H_t)\ \mathrm{d}t.$$

This is again consistent with results and the notation of the previous section.

The well-known limit theory of Rebolledo (1980) for local martingales will be enabling us to formulate our asymptotic results. This will be rigourously discussed in Chapter 4.

2.4 Poisson Processes

Before actually defining Hawkes processes, we want to get into a specific process to then develop and extend the theory to our needs.

Definition 2.4.1. *A point process with constant intensity function, i.e. $\lambda(t, H_t) =: \lambda$ is called a (homogeneous) Poisson process with intensity rate λ.*

Remark. *Let N be a Poisson process with intensity rate λ. It holds that*

$$N((s,t]) \sim \mathrm{Poi}(\lambda(t-s)).$$

This remark gives the intuition to its name. Homogeneous Poisson processes are obviously stationary.

Ultimately we will want to treat processes with an intensity function with random behavior. Towards this direction we establish this other special case.

Definition 2.4.2. *A point process N is called an inhomogeneous Poisson process if the number of events in an intervall $(a, b]$ for $a, b \in \mathbb{R}$ is Poisson distributed as follows*

$$N(a, b] \sim \mathrm{Poi}\left(\int_a^b \lambda(u) \, \mathrm{d}u \right)$$

for a locally integrable positive function λ, which is called the intensity function.

2.5 Hawkes Processes

We define Hawkes processes via their (conditional) intensity function. The realized point process is then considered to be the Hawkes process, while the underlying model is called the Hawkes model, which is not directly observable.

Definition 2.5.1. *Point processes with the conditional intensity function given by*

$$\lambda(t) = \nu + \int_{-\infty}^{t-} \mu(t - u) \, \mathrm{d}N(u).$$

are called Hawkes processes or self-exciting processes where $\nu > 0$ and $\mu : \mathbb{R} \to \mathbb{R}$ with $\mu(u) \geq 0$ for all $u \geq 0$ and $\mu(u) = 0$ for all $u < 0$.

As already motivated in Chapter 2, we find application in ν as the immigration rate and μ as the reproduction rate. Whenevery an event happened as a time point s in the past, the conditional intensity in a time point t for $s < t$ increases depending on the amount of time passed $t - s$.
We see that the formal definition of the intensity function gives rise to the perspective of looking at the immigration and reproduction separately. We note, that the immigration itself is then following a Poisson process model.
Note, that Hawkes processes are thus inhomogenous Poisson processes. The practical problem lies in the additonal randomness of the intensity function, via the conditioning on the history of the process. Even with the knowledge of the reproduction function at a point in time, one knows the mean number

of events in the next instant, but could not know it in the following instant. This behavior makes the model so fascinating and obviously gives rise to a vast amount of applications follow such a pattern.

Remark. *Let $\lambda(t, H_t)$ denote the intensity function of a Hawkes process N. Suppose that the mean of the intensity function λ be constant, i.e. $\mathbb{E}[\lambda(t)] = \lambda^*$ for all $t \in \mathbb{R}$. Then it holds that*

$$\lambda^* = \mathbb{E}[\lambda(t)] = \nu + \int_{-\infty}^{t} g(t-s)\mathbb{E}[\mathrm{d}N(s)] = \nu + \lambda \int_{0}^{\infty} \mu(s)\mathrm{d}s,$$

i.e. it holds that

$$\lambda = \frac{\nu}{1 - \int_0^\infty \mu(s)\mathrm{d}s}.$$

In particular, it has to hold that $\int_0^\infty \mu(s)\mathrm{d}s < 1$.

Remark. *Therefore, a Hawkes process has to fulfill $\int_0^\infty \mu(s)\mathrm{d}s < 1$ to be stationary.*

We will see in Theorem 2.6.3, which are conditions for a Hawkes process to be stationary.

Lemma 2.5.2. *There exists at most one stationary orderly point process of finite rate which has the condition intensity of a Hawkes process from Definition 2.5.1 with $0 < \int_0^\infty \mu(s)\mathrm{d}s < 1$.*

2.6 Cluster Processes

Beforehand we have already motivated the following idea. A point in a Hawkes process is always triggered by either the Poisson process share via ν, the immigration rate i.e. events in the past history via the summed up weights of values of μ. Recalling the insect lineage analogy, this means, that the population grows either by reproduction or immigration. This thus implies a family tree-like structure on the point process, as each ancestoral triggers a family tree.
We formally introduce this notion as follows.

Definition 2.6.1. *A (Poisson) cluster process is a point process which fulfills the following properties*

(1) It contains highlighted points, so-called cluster centers or immigrants which are generated by a Poisson process with a rate, the so-called immigration measure ν^c,i.e.

$$N^c(a,b] \sim \mathrm{Poi}\bigg(\int_a^b \nu^c(t)\ \mathrm{d}t\bigg)$$

for $a, b \in \mathbb{R}$.
The process of cluster centers N^c ist also called the first generation.

Denote by K_i the number of individuals in a generation $i \in \mathbb{N}$.

(2) We denote the set of times of birth of individuals of generation i by $\{t_j^{(i)}\}_{j=1,\ldots,K_i}$.
The random variables $P_{i,j}$ have a Poisson distribution. Each individual j in a generation i triggers the birth of $P_{i,j}$-many (direct) descendents whose times of births are characterized by some random variables $\{X_{i,j,k}\}_{k=1,\ldots,P_{i,j}}$, which denote the time passed after the birth of the ancestor such that the set of individuals of $(i+1)$th generation is given by $\{t_j^{(i)} + X_{i,j,k} : j = 1,\ldots,K_i,\ k = 1,\ldots,P_{i,j}\}$.
We denote by $\mu_{i,j}^c(\cdot)$ the reproduction measure, i.e. the intensity of the process of decendents of individual j in generation i, where $\mu_{i,j}^c(t)$ denotes the intensity at time $t_j^{(i)} + t$.

Remark. *A Hawkes process N^H is a (Poisson) cluster process. We notice the cluster representation of Hawkes processes as follows.*

(1) The immigration measure is given by $\nu^c = \nu$. In this case ν is constant, which is not necessary for a cluster process.

(2) The reproduction measure is given by $\mu_{i,j}^c = \mu$, not depending on the time of birth of the ancestor or its generation.
The number of (direct) descendents of the jth ancestor in generation i,$P_{i,j}$, is Poisson distributed with parameter $\int_0^\infty g(u)\ \mathrm{d}u$.

Definition 2.6.2. *We say that a cluster process N exists if with probability one there are only a finite number of points in any finite interval.*

Lemma 2.6.3. *If $\nu > 0$ is constant and $m := \int_0^\infty \mu(u)\ \mathrm{d}u < 1$ there exists a stationary Poisson cluster process with intensity $\lambda := \frac{\nu}{1-m}$ and conditional intensity of a Hawkes process and $m > 0$.*

This enables us to use existing theory of cluster processes on Hawkes processes.
Additionally, it was shown, that the length of a cluster is almost surely finite in the sense that the last birth of a decendant of every immigrant is finite..

Theorem 2.6.4. *Assume that* $0 < \int_0^\infty \mu(s)\,\mathrm{d}s < 1$*. The length of the cluster is integrable if and only if the mean of the birthing time is finite.*

This now enables the idea, that one can impose an independence structure on a cluster process, as each cluster is independently generated via a Poisson process of births of cluster centers. This yields that a Hawkes process fulfills a notion of independence from a time point as long away as the length of a cluster.

Remark. *For Galton-Watson theory we know, that the size of a cluster is almost surely finite if* $m < 1$*. The expected cluster size, i.e. the number of the events triggered by the cluster center, is* $\frac{1}{1-m} = \sum_{k \in \mathbb{N}} m^k$*.*

2.7 Locally Stationary Hawkes Processes

We have by now seen great advantages to using stationary Hawkes processes, but fairly obviously they lack flexibility in certain aspects. As seen in Theorem 2.6.3 for stationarity we can not allow for time varying immigrationrate oder reproduction function. Thus, an extended version of the model will be introduced. Developed inference ist the centerpiece of this work and will be handled in the following chapters.

Definition 2.7.1. *We consider a model for multivariate Hawkes processes* $N(t) = (N^1(t), ..., N^d(t))^\intercal$ *for* $t \in [0, T]$ *which has an underlying conditional intensity function given by defined by*

$$\lambda^l(t) = \nu^l\left(\frac{t}{T}\right) + \sum_{m=1}^{d} \int_{-\infty}^{t-} \mu^{l,m}\left(t - s; \frac{t}{T}\right) \mathrm{d}N_s^{(m)} \tag{2.1}$$

with $l = 1, ..., d$ *and* $t \in [0, T]$ *or in vector notation*

$$\lambda(t) = \nu\left(\frac{t}{T}\right) + \int_{-\infty}^{t-} \mu\left(t - s; \frac{t}{T}\right) \mathrm{d}N_s.$$

ν^l *and* $\mu^{l,m}$ *are positive functions for* $1 \le l, m \le d$ *defined on* $[0, 1]$ *and* $\mathbb{R}_0^+ \times [0, 1]$*, respectively* $\mathbb{R}_0^+ := [0, \infty)$*.*

3 Setup and Estimation Procedure

The goal is to implement inference for the conditional intensity function λ at a fixed time point $x_0 = \frac{t_0}{T}$ for $t_0 \in [0, T]$ and a fixed dimension l in a local neighborhood.
We will use an asymptotic framework where $T \to \infty$. Intuitively, for growing T more events are contained in said local neighborhoods in the time line. This idea is inspired by the work on locally stationary processes in Dahlhaus (1996).
We will make assumptions corresponding to the observed Hawkes process being locally stationary. In this chapter reproduction functions μ are considered, which have compact support. In the testing procedure of Chapter 4 we change this setting to a weaker assumption.

3.1 Estimation Strategy

For a fixed value $x_0 \in (0, 1)$ and for a fixed value of l we will develop theory for estimation of the parameter $\nu^l(x_0)$ and the function $\mu^{l,m}(u; x_0)$ for $1 \leq m \leq d$. Without loss of generality we choose $l = 1$ and we write $\nu^* = \nu^1(x_0)$, $\mu^{*,m}(u) = \mu^{1,m}(u; x_0)$, and $\mu^*(u) = (\mu^{*,m}(u))_{1 \leq m \leq d}$. Correspondingly, we write $t_0 = x_0 T$. Note that this value depends on T. We assume that the Hawkes process is observed on an interval $[0, T]$. To simplify discussions we assume that the observed process has the conditional intensity function (2.5.1) for $(-\infty, T]$. All counting processes considered in this paper are normed to be equal to 0 for $t = 0$.
We will discuss the so-called cluster representation of a Hawkes process as mentioned in Section 2.6. In this representation each jump of the counting process is interpreted as the birth of an individual. There are d different types of individuals. The lth component of the counting process jumps if an individual of type l is born. We distinguish between individuals with exactly

one parent and individuals without parents. Individuals without parents are called immigrants or ancestors. The parent of an individual who is not an immigrant can be of type $1 \leq m \leq d$. We allow for the type of a parent to differ from the type of its child. For the cluster representation of the Hawkes process we have to define with which conditional probability $\lambda^l(t)\mathrm{d}t$ an individual of type l is born in an infinitesimal interval $\mathrm{d}t$, given the past of the process. Let us write for the birth times of individuals of type l: $(t_{l,i} : i \in \mathbb{Z})$. According to (2.5.1) we can write

$$\lambda^l(t)\mathrm{d}t = \nu^l\left(\frac{t}{T}\right)\mathrm{d}t + \sum_{m=1}^{d}\sum_{i\in\mathbb{Z},t_{m,i}<t}\mu^{l,m}\left(t - t_{m,i}; \frac{t}{T}\right)\mathrm{d}t. \tag{3.1}$$

Each summand of this sum can be interpreted as conditional probability of a birth of an individual of type l in the interval $\mathrm{d}t$ where for each summand the birth is caused by another reason, for details see Section 3.3.3.

We now come to the definition of our smoothing estimator. It is based on B-spline fits with accuracy measured by a local criterion function that is localized around x_0. The local criterion makes use of locally polynomial kernel smoothing of fixed order K. Thus, when estimating the function $\mu^*(u) = \mu^{1,m}(u; x_0)$ in the smoothing procedure the function $\mu^{1,m}(u; x)$ is approximated locally for x in a neighborhood of x_0 by the sum of products of a B-spline function with argument u and a power of $x - x_0$. For the B-spline approximation we use a B-spline basis $\psi_1, \psi_2, ..., \psi_J$ for d-dimensional functions on $[0, A]$ with equidistant knot-points and with norm $\|\psi_l\|^2 = \int_0^A \psi_l^\intercal(x)\psi_l(x)\mathrm{d}x = 1$ for $1 \leq l \leq J$. We suppose that the support of the basis elements is a rectangle with side lengths bounded by a constant times $J^{-1/d}$ and that the basis elements are absolutely bounded by $J^{1/2}$. Other theory can be easily extended to other basis choices with these properties. But, we have no idea how to implement other smoothing methods with respect to the first argument of the reproduction functions. In particular, this concerns kernel smoothing. But also for smoothing splines this is not clear to us, because of the additional kernel smoothing done with respect to the second argument of the reproduction functions.

We now give the definition of our estimator for the case $K = 1$. The definiton for $K > 1$ will be introduced in the following Subsection 3.1.1.

The estimator $(\hat{\nu}^*, \hat{\mu}^*(\cdot))$ of $(\nu^*, \mu^*(\cdot))$ is equal to

$$\begin{aligned} \hat{\nu}^* &= \hat{\beta}_0, \\ \hat{\mu}^*(u) &= \sum_{j=1}^{J}\hat{\beta}_j\psi_j(u), \end{aligned}$$

where $\hat{\beta}_0, ..., \hat{\beta}_J$ are defined as follows.

For some bandwidth $h = h_T \to 0$ and basis dimension $J = J_T \to \infty$, both depending on T,

$$\begin{aligned}\rho(\beta) &= -\frac{2}{T}\int \lambda^{\#}(t;\beta)h^{-1}L\left(\frac{t-t_0}{Th}\right)\mathrm{d}N_t^1 \\ &\quad +\frac{1}{T}\int \lambda^{\#}(t;\beta)^2 h^{-1}L\left(\frac{t-t_0}{Th}\right)\mathrm{d}t\end{aligned}$$

is minimized for $\beta = \hat{\beta}$. The function L is a kernel function, i.e. a probability density function. Furthermore,

$$\lambda^{\#}(t;\beta) = \beta_0 + \sum_{j=1}^{J}\beta_j \int_{t-A}^{t-} \psi_j(t-u)^{\intercal}\mathrm{d}N_u$$

is a local approximation of $\lambda^1(t)$ with $\nu^1(x) = \beta_0$ and $(\mu^{1,m}(u,x))_{k=1}^{d} = \sum_{j=1}^{J}\beta_j\psi_j(t-u)^{\intercal}$. The criterion $\rho(\beta)$ =min! can be interpreted as a localized least squares criterion. This becomes clear if one writes in abuse of notation :

$$\rho(\beta) = \int\left[\left(\frac{\mathrm{d}N_t}{\mathrm{d}t}(t) - \lambda^{\#}(t;\beta)\right)^2 - \left(\frac{\mathrm{d}N_t}{\mathrm{d}t}(t)\right)^2\right]^2 h^{-1}L\left(\frac{t-t_0}{Th}\right)\mathrm{d}t.$$

Here the undefined term $\left(\frac{\mathrm{d}N_t}{\mathrm{d}t}(t)\right)^2$ cancels out and $\frac{\mathrm{d}N_t}{\mathrm{d}t}\mathrm{d}t$ is read as $\mathrm{d}N_t$.

It holds that $\rho(\hat{\beta}) = -2\hat{\tau}^{\intercal}\hat{\beta} + \hat{\beta}^{\intercal}\Delta\hat{\beta}$, where $\hat{\tau} = (\hat{\tau}_0, ..., \hat{\tau}_J)^{\intercal}$ is a $(J+1)$-dimensional vector with

$$\hat{\tau}_0 = \frac{1}{Th}\int_{t_0-Th}^{t_0+Th} L\left(\frac{t-t_0}{Th}\right)\mathrm{d}N_t^1, \tag{3.2}$$

$$\hat{\tau}_j = \frac{1}{Th}\int_{t_0-Th}^{t_0+Th}\int_{t-A}^{t-}\psi_j^{\intercal}(t-u)L\left(\frac{t-t_0}{Th}\right)\mathrm{d}N_u\mathrm{d}N_t^1, \tag{3.3}$$

and where Δ is the following random $(J+1)\times(J+1)$-matrix

$$\Delta = \begin{pmatrix} \Delta_{00} & \Delta_{01} & \dots & \Delta_{0J} \\ \Delta_{01}^{\intercal} & \Delta_{11} & \dots & \Delta_{1J} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta_{0J}^{\intercal} & \Delta_{J1} & \dots & \Delta_{JJ} \end{pmatrix} \tag{3.4}$$

with

$$\Delta_{0,0} = \frac{1}{Th}\int_{t_0-Th}^{t_0+Th} L\left(\frac{t-t_0}{Th}\right)\mathrm{d}t,$$

which is equal to 1, for h small enough,

$$\Delta_{0,j} = \frac{1}{Th}\int_{t_0-Th}^{t_0+Th} R_j(t) L\left(\frac{t-t_0}{Th}\right) \mathrm{d}t, \tag{3.5}$$

$$\Delta_{j,j'} = \frac{1}{Th}\int_{t_0-Th}^{t_0+Th} R_j(t) R_{j'}(t)\ L\left(\frac{t-t_0}{Th}\right) \mathrm{d}t, \tag{3.6}$$

$$R_j(t) = \int_{t-A}^{t-} \psi_j^{\intercal}(t-u)\mathrm{d}N_u.$$

Note that $\hat{\beta} = \Delta^{-1}\hat{\tau}$ as long as Δ is invertible. We see that our resulting estimators $\hat{\nu}^*$ and $\hat{\mu}^*(u)$ are based on locally weighted least squares estimation, where the local weights are coming from Nadaraya-Watson type smoothing.

3.1.1 Estimation using Locally Polynomial Smoothing

We now describe our smoothing method for $K > 1$ where the weights are coming from locally polynomial smoothing of order K. As in classical nonparametric regression locally polynomial smoothing leads to a better performance at boundaries and it achieves faster rates in case of higher order smoothness of $\nu^1(x)$ and $\mu^{1,m}(u;x)$, here with respect to x. Now, the estimator $(\hat{\nu}^*, \hat{\mu}^*(\cdot))$ of $(\nu^*, \mu^*(\cdot))$ is equal to

$$\hat{\nu}^* = \hat{\theta}_{0,1},$$

$$\hat{\mu}^*(u) = \sum_{j=1}^{J} \hat{\theta}_{j,1}\psi_j(u),$$

where $\hat{\theta} = (\hat{\theta}_{0,1}, \hat{\theta}_{1,1}, ..., \hat{\theta}_{J,1}) = (\hat{\theta}_0^{\intercal}, ..., \hat{\theta}_J^{\intercal})^{\intercal}$ with $\hat{\theta}_j = (\hat{\theta}_{j1}, ..., \hat{\theta}_{jK})^{\intercal}$ for $0 \le j \le J$ is defined as follows. The criterion

$$\rho(\theta) = -\frac{2}{T}\int \lambda^{\#}(t;\theta) h^{-1} L\left(\frac{t-t_0}{Th}\right) \mathrm{d}N_t^1 + \frac{1}{T}\int \lambda^{\#}(t;\theta)^2 h^{-1} L\left(\frac{t-t_0}{Th}\right) \mathrm{d}t$$

is minimized for $\theta = \hat{\theta}$. Now,

$$\lambda^{\#}(t;\theta) = \sum_{k=1}^{K} \theta_{0,k}\left(\frac{t-t_0}{Th}\right)^{k-1} + \sum_{j=1}^{J}\sum_{k=1}^{K} \theta_{jk}\int_{t-A}^{t-}\left(\frac{t-t_0}{Th}\right)^{k-1} \psi_j(t-u)^{\intercal}\mathrm{d}N_u$$

is a higher order approximation of $\lambda^1(t)$ with $\nu^1(x) = \sum_{k=1}^{K} \theta_{0,k} \left(\frac{x-x_0}{h}\right)^{k-1}$ and $(\mu^{1,m}(u,x))_{k=1}^{d} = \sum_{j=1}^{J} \sum_{k=1}^{K} \theta_{jk} \left(\frac{x-x_0}{h}\right)^{k-1} \psi_j(t-u)^\intercal$. As above, $\rho(\theta)$ =min! can be interpreted as a localized least squares criterion. Furthermore, we have $\hat{\theta} = \Delta^{-1}\hat{\tau}$ as long as Δ is invertible where now $\hat{\tau} = (\hat{\tau}_0^\intercal, ..., \hat{\tau}_J^\intercal)^\intercal$ is a $K(J+1)$-dimensional vector and Δ is a $K(J+1) \times K(J+1)$-matrix. The vector $\hat{\tau}$ is defined by

$$\hat{\tau}_0 = \left(\frac{1}{Th} \int_{t_0-Th}^{t_0+Th} \left(\frac{t-t_0}{Th}\right)^{k-1} L\left(\frac{t-t_0}{Th}\right) \mathrm{d}N_t^1 \right)_{1\le k\le K},$$

$$\hat{\tau}_j = \left(\frac{1}{Th} \int_{t_0-Th}^{t_0+Th} \int_{t-A}^{t-} \psi_j^\intercal(t-u) \left(\frac{t-t_0}{Th}\right)^{k-1} L\left(\frac{t-t_0}{Th}\right) \mathrm{d}N_u \mathrm{d}N_t^1 \right)_{1\le k\le K}.$$

Furthermore, the matrix Δ is given by

$$\Delta_{0,0} = \left(\frac{1}{Th} \int_{t_0-Th}^{t_0+Th} \left(\frac{t-t_0}{Th}\right)^{k+k'-2} L\left(\frac{t-t_0}{Th}\right) \mathrm{d}t \right)_{1\le k,k'\le K},$$

$$\Delta_{0,j} = \left(\frac{1}{Th} \int_{t_0-Th}^{t_0+Th} \int_{t-A}^{t-} \psi_j^\intercal(t-u) \left(\frac{t-t_0}{Th}\right)^{k+k'-2} L\left(\frac{t-t_0}{Th}\right) \mathrm{d}N_u \mathrm{d}t \right)_{1\le k,k'\le K},$$

$$\Delta_{j,j'} = \left(\frac{1}{Th} \int_{t_0-Th}^{t_0+Th} \int_{t-A}^{t-} \int_{t-A}^{t-} \psi_j(t-u)\mathrm{d}N_u \; \psi_{j'}^\intercal(t-v)\mathrm{d}N_v \right.$$
$$\left. \times \left(\frac{t-t_0}{Th}\right)^{k+k'-2} L\left(\frac{t-t_0}{Th}\right) \mathrm{d}t \right)_{1\le k,k'\le K}.$$

We are back in the case $K = 1$ for the particular choice $\beta = (\theta_{0,1}, ..., \theta_{J,1})^\intercal$.

3.2 Asymptotic Analysis

We make the following assumptions:

(A1) The support of the functions $\mu^{l,m}(s,x)$ is contained in $[0,A] \times [0,1]$ for some known $A > 0$ and the functions are bounded and positive for $l, m = 1, ..., d,\ s \in [0,A], x \le 1$. The matrix Γ^+ has spectral radius strictly smaller than 1. Here we define Γ^+ as the matrix with elements

$$\int_0^A \sup_{x\le 1} \mu^{l,m}(s,x) ds.$$

The function $\nu(x)$ is bounded and bounded away from 0 for $x \le 1$.

(A2) The partial derivatives of $\mu^{l,m}(s,\cdot)$ with respect to the second argument exist and they are uniformly absolutely bounded for $l,m = 1,...,d$ and $0 \le s \le A$.

(A3) The smoothing parameters $h = h_T$ and $J = J_T$ depend on T. The order K is fixed. It holds that $J(hT)^{-1/2}T^{\delta} \to 0$ for $\delta > 0$ small enough and that $\log T J^{-1/d} \to 0$.

(A4) There exist $\theta^*_{j,k}$ for $k = 1,..,K$ and $j = 0,...,J$, depending on J such that for $u \in [0,A]$ and $|x - x_0| \le h$

$$\left|\nu^1(x) - \sum_{k=1}^{K} \theta^*_{0,k}(x-x_0)^{k-1}\right| \le \varepsilon_{1,T},$$

$$\left|\mu^{1,m}(u,x) - \sum_{k=1}^{K}\sum_{j=1}^{J} \theta^*_{j,k}\psi_j(u)(x-x_0)^{k-1}\right| \le \varepsilon_{2,T}$$

for some sequences $\varepsilon_{1,T}, \varepsilon_{2,T} \to 0$.

Let us shortly discuss the assumptions. In the common case where $\mu^{l,m}(s,x)$ does not depend on x Assumption (A1) is a standard assumption to get the existence of a stationary solution for (2.5.1). It guarantees that the Hawkes process does not explode. For $d = 1$ it boils down to the integrability condition of the kernel function to be below 1, in order to avoid that the number of events of the Hawkes process converges to infinity which excludes a stationary version of the Hawkes process. In (A1) we assume that the functions $\mu^{l,m}(\cdot,x)$ have bounded support $[0,A]$. From an applied point of view it would be important to generalize our approach to functions $\mu^{l,m}(\cdot,x)$ with unbounded support. To simplify our theoretical discussions we do not allow that $A = \infty$ because nonparametric estimation of a function with unbounded support needs additional considerations, in particular on the tails of the function. In a first step one could generalize our results to an asymptotic setting where $A \to \infty$. In an alternative approach one could use a parametric model for the tails and one could apply the nonparametric estimator only in a finite interval. Both extensions of our approach will not be considered in this paper. Assumption (A2) is used at a technical point in the proof. It could be replaced by weaker smoothness assumptions which is not done here because stronger assumptions are already needed to get (A4) with reasonable rates $\varepsilon_{1,T}, \varepsilon_{2,T} \to 0$. Assumption (A3) demands standard assumptions on the rate of convergence of the smoothing parameters $h = h_T$ and $J = J_T$ for $T \to \infty$. Note that J/h could be interpreted as the order of the number of unknown parameters and that, under our assumptions, the number of events is of order $O_P(T)$. Thus T could be seen as the order of the sample size and in (A3) we assume that the number of unknown parameters is of order $O((T/h)^{1/2}T^{-\delta})$ for some $\delta > 0$ small enough. We conjecture that this assumption can be weakened. For an interpretation of Assumption (A4) suppose that, for $1 \leq m \leq d$, for u in $[0, A]$ and for x in a neighborhood of x_0, $\mu^{1,m}(u, x)$ has absolutely bounded partial derivatives of order M with respect to u and x. Assume further that B-splines of order M with n intervals of $[0, A]$ are used and locally polynomial smoothing with $K = M$ is applied. Then one can show for some constants $C_1, C_2, ...$ that Assumption (A4) holds with $\varepsilon_{1,T} \leq C_1 h^M$ and $\varepsilon_{2,T} \leq C_2 n^{-M} + C_3 h^M$. Because J is of order n^d we get that $\varepsilon_T = \max(\varepsilon_{1,T}, \varepsilon_{2,T}) \leq C_4 J^{-M/d} + C_5 h^M$. In the following theorem we get that the estimator is of order $\varepsilon_T + \sqrt{\frac{J}{hT}}$. This order is minimized by choosing h and $J^{-1/d}$ of order $T^{-1/(2M+d+1)}$. Then one gets an estimator with rate of convergence of order $T^{-M/(2M+d+1)}$. In nonparametric curve estimation this rate of convergence shows up as optimal rate in many settings for the estimation of M-times differentiable functions with $(d+1)$-dimensional argument. We get that this rate can also be attained in our estimation problem after appropriate choice of the smoothing parameters J and h.

Theorem 3.2.1. *Make Assumptions (A1) – (A4). With* $\varepsilon_T = \max(\varepsilon_{1,T}, \varepsilon_{2,T})$ *it holds that*

$$\hat{\nu}^* - \nu^* = O_P\left(\varepsilon_T + \sqrt{\frac{J}{hT}}\right),$$

$$\left(\int [\hat{\mu}^{*,(l)}(u) - \mu^{*,(l)}(u)]^2 \mathrm{d}u\right)^{1/2} = O_P\left(\varepsilon_T + \sqrt{\frac{J}{hT}}\right)$$

for $l = 1, ..., d$.

By using some simplifications in our proofs one gets the following result for a stationary Hawkes process N_t with conditional intensity function

$$\lambda^l(t) = \nu^l + \sum_{m=1}^{d} \int_{t-A}^{t-} \mu^{l,m}(t-s)\, \mathrm{d}N_s^{(m)} \tag{3.7}$$

for some constants ν^l and functions $\mu^{l,m}(\cdot)$ $(1 \leq l, m \leq d)$. Now, the estimator $(\hat{\nu}, \hat{\mu}(\cdot))$ of $(\nu, \mu(\cdot))$ is defined by $\hat{\mu}(u) = \sum_{j=1}^{J} \hat{\beta}_j \psi_j(u)$, where $(\hat{\nu}, \hat{\beta}_1, ..., \hat{\beta}_J)$ minimizes

$$-\frac{2}{T}\int \lambda^{\#}(t; \nu, \beta)\mathrm{d}N_t^1 + \frac{1}{T}\int \lambda^{\#}(t; \nu, \beta)^2 \mathrm{d}t$$

with $\lambda^{\#}(t;\nu,\beta) = \nu + \sum_{j=1}^{J} \beta_j \int_{t-A}^{t-} \psi_j(t-u)^\intercal \mathrm{d}N_u$. In the stationary case we get the following result:

Corollary 3.2.2. *For a stationary process with* (3.7) *assume that for* $1 \leq l, m \leq d$ *the support of the functions* $\mu^{l,m}(s)$ *is contained in* $[0, A]$ *for some* $A > 0$*, that the functions are positive for* $l, m = 1, ..., d,\ s \in [0, A]$*, and that the matrix* Γ^+ *has spectral radius strictly smaller than 1. Now, we define* Γ^+ *as the matrix with elements* $\int_0^A \mu^{l,m}(s)ds$*. Furthermore, we assume that* $\nu > 0$*, that* $J(T)^{-1/2+\delta} \to 0$ *for* $\delta > 0$ *small enough, that* $\log T J^{-1/d} \to 0$ *and that there exist* β_j^* *for* $j = 1, ..., J$*, depending on* J*, such that*

$$\left| \mu^{1,m}(u) - \sum_{j=1}^{J} \beta_j^* \psi_j(u) \right| \leq \varepsilon_T$$

for some sequence $\varepsilon_T \to 0$ *and for* $u \in [0, A]$*. Then, it holds that* $\hat{\nu} - \nu = O_P\left(\varepsilon_T + \sqrt{J/T}\right)$ *and*

$$\left(\int [\hat{\mu}^{1,m}(u) - \mu^{1,m}(u)]^2 \mathrm{d}u \right)^{1/2} = O_P\left(\varepsilon_T + \sqrt{J/T}\right)$$

for $m = 1, ..., d$.

Up to an additional log factor this result can also be proved along the lines of arguments used in Hanssen, Reynaud-Bouret and Rivoirard (2015). There an additional log factor appears because adaptive LASSO estimation is considered.

3.3 Proof of Theorem 3.2.1

3.3.1 Outline of the Proof

In our proofs the quantities $C, C^*, C_1, ...$ are positive constants that are chosen large enough and $c, c^*, c_1, ...$ are strictly positve constants that are chosen small enough. The same names will be used for different constants, also in the same formula. For simplicity we assume that $K = 1$. All arguments go through for $K > 1$ at the cost of a more complex notation. For the case $K = 1$ we can make use of a simplified notation as already outlined in Section 3.1. We write $\hat{\beta}_j = \hat{\theta}_{j,1}$ and $\beta_j^* = \theta_{j,1}^*$. Our estimator $(\hat{\nu}^*, \hat{\mu}^*(\cdot))$ of $(\nu^*, \mu^*(\cdot))$ is defined as $\hat{\mu}^*(u) = \sum_{j=1}^{J} \hat{\beta}_j \psi_j(u)$, where for $\hat{\beta} = (\hat{\beta}_0, ..., \hat{\beta}_J)^\intercal$ with some bandwidth $h \to 0$ and basis dimension $J \to \infty$

$\rho(\beta) = -\frac{2}{T}\int \lambda^{\#}(t;\beta)h^{-1}L\left(\frac{t-t_0}{Th}\right)\mathrm{d}N_t^1 + \frac{1}{T}\int \lambda^{\#}(t;\beta)^2 h^{-1}L\left(\frac{t-t_0}{Th}\right)\mathrm{d}t$ is minimized. Here, as above, $\psi_1, ..., \psi_J$ is a B-spline basis for d-dimensional functions on $[0, A]$ and $\lambda^{\#}(t;\beta) = \beta_0 + \sum_{j=1}^{J}\beta_j \int_{t-A}^{t-}\psi_j(t-u)^{\intercal}\mathrm{d}N_u$. Note that now $\rho(\beta) = -2\hat{\tau}^{\intercal}\beta + \beta^{\intercal}\Delta\beta$, where

$$\Delta = \begin{pmatrix} 1 & \delta^{\intercal} \\ \delta & \Delta^* \end{pmatrix}$$

with $\delta = (\Delta_{0,1}, ..., \Delta_{0,J})^{\intercal}$ and $\Delta^* = (\Delta_{j,j'})_{1\le j,j'\le J}$ see (3.5), (3.6). Note that $\hat{\beta} = \Delta^{-1}\hat{\tau}$ as long as Δ is invertible. We now decompose $\hat{\tau} = (\hat{\tau}_0, ..., \hat{\tau}_J)^{\intercal} = (\hat{\tau}_0, \hat{\tau}_{-0}^{\intercal})^{\intercal}$ as follows

$$\begin{aligned} \hat{\tau}_0 &= \hat{\tau}_{0,A} + \hat{\tau}_{0,B} + \hat{\tau}_{0,C}, \\ \hat{\tau}_{-0} &= \hat{\tau}_{-0,A} + \hat{\tau}_{-0,B} + \hat{\tau}_{-0,C}, \end{aligned}$$

where

$$\begin{aligned} \hat{\tau}_{0,A} &= \frac{0}{Th}\int_{t_0-Th}^{t_0+Th} L\left(\frac{t-t_0}{Th}\right)\mathrm{d}M_t^1, \\ \hat{\tau}_{0,B} &= \frac{1}{Th}\int_{t_0-Th}^{t_0+Th} L\left(\frac{t-t_0}{Th}\right)\Bigg[\{\nu(t/T) - \beta_0^*\} \\ &\qquad + \int_{t-A}^{t-}\left\{\mu(t-s, t/T) - \sum_{k=1}^{J}\beta_k^*\psi_k^{\intercal}(t-s)\right\}\mathrm{d}N_s\Bigg]\mathrm{d}t, \\ \hat{\tau}_{0,C} &= \frac{1}{Th}\int_{t_0-Th}^{t_0+Th} L\left(\frac{t-t_0}{Th}\right)\left[\beta_0^* + \int_{t-A}^{t-}\sum_{k=1}^{J}\beta_k^*\psi_k^{\intercal}(t-s)\mathrm{d}N_s\right]\mathrm{d}t, \end{aligned}$$

$$\begin{aligned} \hat{\tau}_{-0,A,j} &= \frac{1}{Th}\int_{t_0-Th}^{t_0+Th}\int_{t-A}^{t-}\psi_j^{\intercal}(t-u)L\left(\frac{t-t_0}{Th}\right)\mathrm{d}N_u\mathrm{d}M_t^1, \\ \hat{\tau}_{-0,B,j} &= \frac{1}{Th}\int_{t_0-Th}^{t_0+Th}\int_{t-A}^{t-}\psi_j^{\intercal}(t-u)L\left(\frac{t-t_0}{Th}\right)\mathrm{d}N_u\Bigg[\{\nu(t/T) - \beta_0^*\} \\ &\qquad + \int_{t-A}^{t-}\left\{\mu(t-s, t/T) - \sum_{k=1}^{J}\beta_k^*\psi_k^{\intercal}(t-s)\right\}\mathrm{d}N_s\Bigg]\mathrm{d}t, \\ \hat{\tau}_{-0,C,j} &= \frac{1}{Th}\int_{t_0-Th}^{t_0+Th}\int_{t-A}^{t-}\psi_j^{\intercal}(t-u)K\left(\frac{t-t_0}{Th}\right)\mathrm{d}N_u\Bigg[\beta_0^* \\ &\qquad + \int_{t-A}^{t-}\sum_{k=1}^{J}\beta_k^*\psi_k^{\intercal}(t-s)\mathrm{d}N_s\Bigg]\mathrm{d}t \end{aligned}$$

with $\beta_j^* = \theta_{j,1}^*$ for $j = 1, ..., J$, see (A4).

Note that $(\beta_0^*, ..., \beta_J^*)^\intercal = \Delta^{-1}(\hat{\tau}_{0,C}, \hat{\tau}_{-0,C,1}, ..., \hat{\tau}_{-0,C,J})^\intercal$. For the proof of the theorem we will show that for $1 \le j \le J$

$$\hat{\tau}_{0,A} = O_P\left(\sqrt{\frac{1}{hT}}\right), \qquad \hat{\tau}_{-0,A,j} = O_P\left(\sqrt{\frac{1}{hT}}\right), \tag{3.8}$$

$$\hat{\tau}_{0,B} = O_P\left(\varepsilon_T\right), \qquad \hat{\tau}_{-0,B,j} = O_P\left(\varepsilon_T\right), \tag{3.9}$$

see Subsection 3.3.6 and Lemma 3.3.9. Furthermore we will show that

the smallest eigenvalue of $\mathbb{E}[\Delta]$ is bounded from below
by some constant $c > 0$, (3.10)

see Subsection 3.3.2 and Lemma 3.3.3. Finally we will show that for all $\epsilon > 0$ with probability tending to one it holds that

$$\max_{0 \le j,j' \le J} |\Delta_{j,j'} - \mathbb{E}[\Delta_{j,j'}]| \le CT^{\epsilon}\frac{1}{\sqrt{hT}}, \tag{3.11}$$

see Subsection 3.3.5. From (3.11) and Assumption (A3) we get that the operator norm of $\Delta - \mathbb{E}[\Delta]$ is of order $o_P(1)$. Thus, from (3.10)– (3.11) one concludes that, with probability tending to 1, the smallest eigenvalue of Δ is bounded from below by some constant $c > 0$. In summary, together with (3.8), (3.9), this shows that the statement of Theorem 3.2.1 follows from (3.8)–(3.11).

For the proof of (3.10) we need formulas for the first two moments of linear statistics of the Hawkes process. We present such results in Subsection 3.3.2. For the proof of (3.8)–(3.9) we need bounds on the number of jumps of the Hawkes process in subintervals of $[0, T]$. For the proof of (3.11) we need that $\Delta_{j,j'} - \mathbb{E}[\Delta_{j,j'}]$ can be approximated by sums of independent random variables. For these two claims we make use of the cluster representations of Hawkes processes, see Hawkes and Oakes (1974) and Reynaud-Bouret and Roy (2007). This representation has also been used in the statistical analysis of nonparametric estimators of other classes of Hawkes processes in Hanssen, Reynaud-Bouret and Rivoirard (2015), Roueff, von Sachs and Sansonnet (2016), and Roueff and von Sachs (2019). In Subsection 3.3.3 we will define a cluster representation for our non-stationary inhomogeneous process N. We will use the cluster representation in Subsection 3.3.4 to compare N with a stationary Hawkes process by a coupling argument. The coupling will allow to carry over results known for stationary Hawkes process to our locally stationary setting. In Subsection 3.3.5 we will apply the cluster representation to show that for $0 \le j, j' \le J$ $\Delta_{j,j'} - \mathbb{E}[\Delta_{j,j'}]$ can be approximated by sums of independent variables. This will be used to show (3.11). Finally in Section 3.3.6 we show (3.8) and (3.9) by using bounds from Subsection 3.3.4 on the moments of the number of jumps of the Hawkes process in an interval.

3.3.2 Bounds on the Moments of Linear Statistics with an Application for the Calculation of $\mathbb{E}[\Delta]$

We will use the following lemma for the calculation of second order moments of linear statistics of Hawkes processes of the form (2.5.1). This lemma generalises results in Bacry, Dayri and Muzy (2012) for stationary homogeneous Hawkes processes.

Lemma 3.3.1. *For a Hawkes process N_t of the form* (2.5.1) *that fulfils Assumption (A1), it holds for $\Lambda(t/T) = \mathbb{E}\left[\lambda(t)\right]$ that:*

$$\Lambda\left(\frac{t}{T}\right) = \nu\left(\frac{t}{T}\right) + \int \chi\left(t-s, \frac{t}{T}\right)\nu\left(\frac{s}{T}\right)\mathrm{d}s, \tag{3.12}$$

$$\lambda(t) = \Lambda\left(\frac{t}{T}\right) + \int \chi\left(t-s, \frac{t}{T}\right)\mathrm{d}M_s, \tag{3.13}$$

where

$$\chi\left(t-s, \frac{t}{T}\right) = \sum_{k=1}^{\infty} \mu^{(*k)}\left(t-s, \frac{t}{T}\right)$$

with $\mu^{(*1)}\left(t-s, \frac{s}{T}\right) = \mu\left(t-s, \frac{s}{T}\right)$ and

$$\mu^{(*k)}\left(t-s, \frac{t}{T}\right) = \int \mu^{(*(k-1))}\left(t-u, \frac{t}{T}\right)\mu\left(u-s, \frac{u}{T}\right)\mathrm{d}u$$

for $k \geq 2$, and where M_t is the martingale defined by $\mathrm{d}M_t = \mathrm{d}N_t - \lambda(t)\mathrm{d}t$. Finally, we have that

$$\mathbb{E}\left[\mathrm{d}N_t \mathrm{d}N_{t'}^{\intercal}\right] = \left(\Lambda\left(\frac{t}{T}\right)\Lambda\left(\frac{t'}{T}\right)^{\intercal} + \Sigma_{t/T}\delta_{t-t'} + \chi\left(t-t', \frac{t}{T}\right)\Sigma_{t'/T} \right. \tag{3.14}$$
$$\left. + \Sigma_{t/T}\chi\left(t'-t, \frac{t'}{T}\right)^{\intercal} + \int \chi\left(t-s, \frac{t}{T}\right)\Sigma_{s/T}\chi\left(t'-s, \frac{t'}{T}\right)^{\intercal}\mathrm{d}s\right)\mathrm{d}t\mathrm{d}t',$$

where $\Sigma_{t/T}$ is a diagonal matrix with diagonal elements $\Lambda_i(t/T)$ and where $\delta_{t-t'}\mathrm{d}t\mathrm{d}t'$ is equal to $\mathrm{d}t$ for $t = t'$ and equal to 0 otherwise.

A proof of this Lemma can be found in the supplement **?**. By application of the lemma we get immediately the following result.

Lemma 3.3.2. *Make the assumptions of Theorem 3.2.1. For functions* $g : [0, A] \to \mathbb{R}^d$ *and* $\gamma \in \mathbb{R}$ *it holds for*

$$R(g,\gamma) = \mathbb{E}\left[\frac{1}{Th}\int_{t_0-Th}^{t_0+Th}\left(\gamma + \int_{t-A}^{t-} g^\intercal(t-u)\mathrm{d}N_u\right)^2 L\left(\frac{t-t_0}{Th}\right)\mathrm{d}t\right].$$

that:

$$R(g,\gamma) = R(g) + R^*(g,\gamma) \tag{3.15}$$

where

$$R(g) = \frac{1}{Th}\int_{t_0-Th}^{t_0+Th}\int\left(g(t-v) + \int \chi\left(u-v,\frac{u}{T}\right)^\intercal g(t-u)\mathrm{d}u\right)^\intercal \Sigma_{v/T}\left(g^\intercal(t-v) + \int \chi\left(u-v,\frac{u}{T}\right)^\intercal g(t-u)\mathrm{d}u\right) L\left(\frac{t-t_0}{Th}\right)\mathrm{d}v\mathrm{d}t,$$

$$R^*(g,\gamma) = \frac{1}{Th}\int_{t_0-Th}^{t_0+Th}\left(\gamma + \int_{t-A}^{t-} g^\intercal(t-u)\Lambda\left(\frac{u}{T}\right)\mathrm{d}u\right)^2 L\left(\frac{t-t_0}{Th}\right)\mathrm{d}t.$$

Finally, this lemma can be used to prove the following result, see Subsection 3.3.7 for details.

Lemma 3.3.3. *Make the assumptions of Theorem 3.2.1. For functions* $g : [0, A] \to \mathbb{R}^d$ *and* $\gamma \in \mathbb{R}$ *it holds for* T *large enough that:*

$$R(g,\gamma) \geq c\left(\gamma^2 + \int_0^A g^\intercal(t)g(t)\mathrm{d}t\right) \tag{3.16}$$

for some constant $c > 0$*. In particular, it holds that the smallest eigenvalue of* $\mathbb{E}[\Delta]$ *is bounded from below and that, thus, the matrix* $\mathbb{E}[\Delta]$ *is invertible.*

3.3.3 Cluster Representations of Locally Stationary Hawkes Processes

We now come back to the cluster representation of a Hawkes process where the jumps $t_{m,i}$ of the counting process are interpreted as the birth dates of immigrants or children, see (3.1). We now interpret each summand of the sum on the right hand side of (3.1) as conditional probability of a birth of an individual of type l in the interval $\mathrm{d}t$ where for each summand the birth is caused by another reason. The first summand $\nu^l\left(\frac{t}{T}\right)\mathrm{d}t$ is the conditional probability that an immigrant of type l is born. And $\mu^{l,m}\left(t - t_{m,i}; \frac{t}{T}\right)\mathrm{d}t$ is the conditional probabilty that the individual of type m who was born at

$t_{m,i}$ gets a child of type l in the interval dt. Clearly, for an observed Hawkes process we do not know if a birth of a child of type l in the interval dt is caused by immigration or by an individual of type m getting a child. But the interpretation of the Hawkes process as developing families simplifies the mathematics essentially. We now give a more formal description of the cluster representation of locally stationary Hawkes processes.

The cluster representation for locally stationary Hawkes process (2.5.1) is given by independent random variables

- the number of children $P_x^{l,m}$ $(x \in [0,1], 1 \leq l, m \leq d)$ of type l of an individual of type m born at Tx. The number $P_x^{l,m}$ is assumed to have a Poisson distribution with parameter $p_x^{l,m} = \int_0^A \mu^{l,m}(s, x + (s/T))\mathrm{d}s$;
- the number of anchestors $P^{(m)}$ $(1 \leq m \leq d)$ born in the interval $[0,T]$. It is assumed that $P^{(m)}$ has a Poisson distribution with parameter $p^{(m)} = T \int_0^1 \nu^{(m)}(s)\mathrm{d}s$;
- Possible birth dates $x + X_{x,i}^{l,m}$ $(x \in [0,1], 1 \leq l, m \leq d,\ i \in \mathbb{N})$ of children of type l of an individual of type m born at Tx. The variables $X_{x,i}^{l,m}$ are distributed according to Lebesgue density $\mu^{l,m}(\cdot, x + (\cdot/T))/p_x^{l,m}$; and
- Possible birth dates $Z_i^{(m)}$ $(1 \leq m \leq d,\ i \in \mathbb{N})$ of anchestors. The variables $Z_i^{(m)}$ have a Lebesgue density $\nu^{(m)}(\cdot/T)/p^{(m)}$.

The variables $P_x^{l,m}$ and $X_{x,i}^{l,m}$ are defined for all $x \in [0,1]$ and $X_{x,i}^{l,m}$ and $Z_i^{(m)}$ are defined for all $i \in \mathbb{N}$. In the construction of the locally stationary Hawkes process we will make use of the variables only for random subsets of $[0,1]$ and $\mathbb{N}$. More precisely, we only use the variables $P_x^{l,m}$ for xT equal to the birth times of individuals of type m and we use $X_{x,i}^{l,m}$ for xT equal to the birth times of individuals of type m and with $1 \leq i \leq P_x^{l,m}$ and finally we need $Z_i^{(m)}$ only for $1 \leq i \leq P_x^{(m)}$.

The construction of the Hawkes process now starts with the definition of birth dates of the ancestors. The birth dates of ancestors of type m are given by a Poisson process with conditional intensity $\nu^{(m)}(\cdot/T)$ or equivalently, as $Z_1^{(m)}, ..., Z_{P(m)}^{(m)}$. Each ancestor sets up a separate family with descendants of type $1 \leq l \leq d$. The descendants are born iteratively in generations $n \geq 1$. A member of a family that was born at time S and that is of type m has $P_{S/T}^{l,m}$ children of type l. They are born at the dates

$$S + X_{S/T,1}^{l,m}, ..., S + X_{S/T,P_{S/T}^{l,m}}^{l,m}.$$

In the construction we made use of the fact that with probability 1 no two individuals are born at the same date. This follows because our conditional intensity measures are assumed to have Lebesgue densities. In particular, for this reason none of the variables $P_x^{l,m}$ $(x \in [0,1], 1 \leq l, m \leq d)$ and $X_{x,i}^{l,m}$ $(x \in [0,1], 1 \leq l, m \leq d,\ i \in \mathbb{N})$ is used twice or more times in the construction, with probability equal to 1.

The cluster construction allows to compare the process N with a homogeneous Hawkes process $\bar{N}$ that is constructed by adding a random number of additional individuals. Thus we can apply bounds on the number of birth times in an interval stated for homogeneous Hawkes processes. By construction the bounds are also valid for the locally stationary process N. This will be described in the next subsection.

3.3.4 Bounds on the Number of Births in an Interval, on the Life Spans of Families and on the Number of Family Members

For homogeneous Hawkes processes bounds have been given for the number of births in an interval, for the life spans of families and for the number of family members. We now construct a homogeneous Hawkes process $\bar{N}$ by adding births of a random number of additional individuals. These allows us to get bounds for these quantities also for our locally stationary process. The homogeneous Hawkes process $\bar{N}$ has conditional intensity function

$$\bar{\lambda}^l(t) = \bar{\nu}^l + \sum_{m=1}^{d} \int_{t-A}^{t-} \bar{\mu}^{l,m}(t-u)\, \mathrm{d}\bar{N}_s^{(m)} \tag{3.17}$$

with $\bar{\nu}^l = \sup_{x \leq 1} \nu^l(x)$ and $\bar{\mu}^{l,m}(t) = \sup_{x \leq 1} \mu^{l,m}(t;x)$ for $1 \leq l, m \leq d$ and $0 \leq t \leq A$. Using cluster representations for both processes, it can be seen that there exists a strong construction of N and $\bar{N}$ such that all birth points of N^l are also birth points of $\bar{N}^l$ for $1 \leq l \leq d$. Under strong construction we mean that this holds for two processes constructed on the same probability space with the distribution of N or $\bar{N}$, respectively. Thus, we can carry over the results of Lemma 1 and of Proposition 2 in Hanssen, Reynaud-Bouret and Rivoirard (2015) that treat homogeneous Hawkes processes to our inhomogeneous Hawkes process and we get the following result.

Lemma 3.3.4. *Make the assumptions of Theorem 3.2.1. There exist $\rho > 0$ and $C > 0$, not depending on T, such that*

$$\mathbb{E}\left[e^{\rho W_l}\right] < C, \tag{3.18}$$

$$\mathbb{E}\left[e^{\rho N_{[t-A,t]}}\right] < C \tag{3.19}$$

for $1 \le l \le d$ and $t \le T$. Here W_l is the number of family members in a family with ancestor of type l and $N_{[t-A,t]}$ is the number of birth points of all types of N in the interval $[t-A,t]$.

At this point we would also like to add another result that follows from the theory of homogeneous Hawkes processes and that will be used in the local mathematical analysis of our estimator $\hat{\beta}$. Denote by $T^*_{e,t}$ the last birth date of all types inside all families whose ancestor was born before t and put $T_{e,t} = T^*_{e,t} - t$. This is also called extinction time. For the homogeneous process $\bar{N}$ the distribution of $T_{e,t}$ does not depend on t and it holds that $\mathbb{P}[T_{e,t} \ge s] \le \sum_{l=1}^d \frac{\bar{\nu}^l}{\rho_l} \mathbb{E}\left[e^{\rho_l W_l}\right] e^{-\rho_l s}$, see the proof of Proposition 3 in Hanssen, Reynaud-Bouret and Rivoirard (2015). Again using the above strong approximation, we can carry over this result to our inhomogeneous Hawkes process and we get the following lemma.

Lemma 3.3.5. *Make the assumptions of Theorem 3.2.1. We have for some constants $C, \rho^* > 0$ that do not depend on T*

$$\mathbb{P}[T_{e,t} \ge s] \le C e^{-\rho^* s} \tag{3.20}$$

for $t \le T$ and $s \ge 0$.

For the study of the terms δ and Δ we will use that these quantities can be approximated by sums of independent random variables. For this aim we will use a construction that also has been used in Hanssen, Reynaud-Bouret and Rivoirard (2015) and Reynaud-Bouret and Roy (2007) for the study of homogeneous Hawkes processes. For x^*, x fixed, suppose that $N_{q,n}$ ($q \in \mathbb{N}, n \in \mathbb{Z}$) are independent Hawkes processes with conditional intensity function $\lambda^l_{q,n}(t) = \lambda^l_n(t) = \nu^{(m)}\left(\frac{t}{T}\right) \mathbb{I}_{[x^*+2nx-x \le t < x^*+2nx+x]}$ $+ \sum_{m=1}^d \int_{t-A}^{t-} \mu^{l,m}\left(t-s; \frac{t}{T}\right) \mathrm{d}N^{(m)}_{q,n;s}$ for $1 \le l \le d$. These are Hawkes processes that only contain families with ancestors born in the interval $[x^*+2nx-x, x^*+2nx+x)$. Put $N^+ = \sum_{-\infty<n<\infty} N_{0,n}$ and $N_q = \sum_{-\infty<n<q-1} N_{q,n} + N_{0,q}$. It holds: The Hawkes process N^+ has the same distribution as N. The ancestors of the Hawkes process N_q are all born before $x^* + 2qx + x$ and the process N_q has conditional intensity function

$$\nu^l\left(\frac{t}{T}\right) \mathbb{I}_{[-\infty<t<x^*+2qx+x]} + \sum_{m=1}^d \int_{t-A}^{t-} \mu^{l,m}\left(t-s; \frac{t}{T}\right) \mathrm{d}N^{(m)}_{q;s} \tag{3.21}$$

for $1 \le l \le d$. Furthermore, it holds that the processes N_q ($q \ge 1$) are independent. We now put for a, x fixed with $0 < a < x$

$$M^{x^*,x}_q = N_q\big|_{[x^*+2qx-a, x^*+2qx+x)}. \tag{3.22}$$

These are independent processes for $q \in \mathbb{N}$. With the same arguments as in Section 3.1 of Reynaud-Bouret and Roy (2007) for one-dimensional homogeneous Hawkes processes one gets that

$$\begin{aligned} \left| \mathbb{P}[M_q^{x^*,x} \neq N^+\big|_{[x^*+2qx-a,x^*+2qx+x)}] \right| &\leq 2\mathbb{P}[T_{e,x^*+2qx-x} \geq x-a] \\ &\leq 2Ce^{-\rho^*(x-a)}, \end{aligned}$$

where (3.20) has been used. In particular, we have that for measurable sets $\mathcal{A}$

$$\left| \mathbb{P}[M_q^{x^*,x} \in \mathcal{A}] - \mathbb{P}[N\big|_{[x^*+2qx-a,x^*+2qx+x)} \in \mathcal{A}] \right| \leq 2Ce^{-\rho^*(x-a)}.$$

By a small extension of the arguments one gets the following lemma.

Lemma 3.3.6. *Make the assumptions of Theorem 3.2.1. There exists a Hawkes process N^+ that has the same distribution as N with the following property. For a finite subset I of $\mathbb{N}$, $x^* \in [0,1)$, and $x > a > 0$ there exist independent Hawkes processes N_q for $q \in I$ with conditional intensity function* (3.21) *such that for the processes $M_q^{x^*,x}$ defined in* (3.22) *it holds that*

$$\mathbb{P}[M_q^{x^*,x} \neq N^+\big|_{[x^*+2qx-a,x^*+2qx+x)} \text{ for some } q \in I] \leq 2C|I|e^{-\rho^*(x-a)} \tag{3.23}$$

3.3.5 Treatment of $\Delta - \mathbb{E}[\Delta]$

We now consider the variables δ and Δ^*. With the help of Lemma 3.3.6 with $a = A$ we can approximate these variables by the sum of two terms, where each term is the sum of independent variables. Such a splitting device has also been used in Reynaud-Bouret and Roy (2007) to prove Hoeffding and Bernstein inequalities for averages of flows induced by stationary Hawkes processes. We start by discussing Δ^*. With $G(t) = (G_{jk}(t))_{1\leq j,k\leq J}$ where

$$G_{jk}(t) = \int_{t-A}^{t-} \int_{t-A}^{t-} L\left(\frac{t-t_0}{Th}\right) \psi_j^\intercal(t-u) \mathrm{d}N_u \psi_k^\intercal(t-v) \mathrm{d}N_v \mathbb{I}_{(t_0-Th\leq t\leq t_0+Th)}$$

we get that $\Delta^* - \mathbb{E}[\Delta^*] = \Delta_1^* + \Delta_2^*$ with

$$\begin{aligned} \Delta_1^* &= \frac{1}{Q}\sum_{q=0}^{Q} \int_{x^*+2qx}^{x^*+(2q+1)x} \frac{Q}{Th} \left\{G(t) - \mathbb{E}[G(t)]\right\} \mathrm{d}t, \\ \Delta_2^* &= \frac{1}{Q}\sum_{q=0}^{Q} \int_{x^*+(2q+1)x}^{x^*+(2q+2)x} \frac{Q}{Th} \left\{G(t) - \mathbb{E}[G(t)]\right\} \mathrm{d}t, \end{aligned}$$

where $x^* = t_0 - hT$ and where x is a value that depends on T and that we will choose below as T^ϵ with $\epsilon > 0$ small enough. Furthermore, Q is the smallest integer larger than $hT/x - 1$. Note that Δ_1^* has the same distribution as Δ_1^+ which is defined as Δ_1^* but with $G(t)$ replaced by $G^+(t)$. The function $G^+(t)$ is defined as $G(t)$ but with the counting process N replaced by N^+. We also define $G_q^+(t)$ as $G^+(t)$ but with N^+ replaced by $M_q^{x^*,x}$. From Lemma 3.3.6 with $a = A$ we get that with probability $\geq 1 - CQ\exp(-\rho^*(x - A))$

$$\Delta_1^+ = \frac{1}{Q}\sum_{q=0}^{Q}\int_{x^*+2qx}^{x^*+(2q+1)x} \frac{Q}{Th}\left\{G^+(t) - \mathbb{E}[G(t)]\right\} \mathrm{d}t = \frac{1}{Q}\sum_{q=0}^{Q}\eta_q$$

with

$$\eta_q = \int_{x^*+(2q+1)x}^{x^*+(2q+2)x} \frac{Q}{Th}\left\{G_q^+(t) - \mathbb{E}[G(t)]\right\} \mathrm{d}t.$$

The variables η_q are mean zero independent random $d \times d$ matrices. Furthermore, they are bounded as follows. Denote the jump points of the components of $M_q^{x^*,x}$ by $t_1^q, t_2^q, \ldots.$ First we have that for $x^* + 2qx \leq t < x^* + (2q+1)x$, $1 \leq j, j' \leq J$

$$\begin{aligned} |G^+_{q,j,j'}(t)| &\leq C\sum_{k,l\geq 1}\|\psi_j(t - t_k^q)\|\|\psi_{j'}(t - t_l^q)\| \qquad (3.24)\\ &\leq CJ\sum_{k,l\geq 1}\sum_{d\geq r,r'\geq 1}\mathbb{I}_{(|t-t_k^q-\tau_j^r|\leq CJ^{-1/d})}\mathbb{I}_{(|t-t_l^q-\tau_{j'}^{r'}|\leq CJ^{-1/d})}, \end{aligned}$$

where $\tau_j^l \in (\tau_j^{l,-}, \tau_j^{l,+})$ and $\tau_{j'}^l \in (\tau_{j'}^{l,-}, \tau_{j'}^{l,+})$ with rectangles $\prod_{l=1}^d [\tau_j^{l,-}, \tau_j^{l,+}]$ and $\prod_{l=1}^d [\tau_{j'}^{l,-}, \tau_{j'}^{l,+}]$ equal to the support of ψ_j or $\psi_{j'}$, respectively. We now construct upper bounds for the number of jump points of the components of $M_q^{x^*,x}$. For this purpose first note that Lemma 3.3.4 implies that $\mathbb{E}[\exp(\rho N_{[t-2A,t]}/2)] < 2C$. This implies that $\sup_{0\leq t\leq T} N_{[t-A,t]} < C\log T$ with probability tending to 1. Because ν_0^l and $\mu^{l,m}$ are bounded by Assumption (A1) for $1 \leq l, m \leq d$ we have that $\lambda^l(t) \leq C\log T$ with probability tending to 1 for C chosen large enough. With similar argumennts as above we now argue that a homogeneous Poisson proces $\tilde{N}_t^q$ can be constructed with conditional intensity $\tilde{\lambda}^{(m)}(t) \equiv C\log T$ such that all jumps of a component of $M_q^{x^*,x}$ are also jumps of the corresponding component of $\tilde{N}^q$, with probability tending to 1. Denote the jump points of the components of $\tilde{N}^q$ in the interval $[x^* + 2qx - A, x^* + 2qx + x)$ by $\tilde{t}_1^q, \tilde{t}_2^q, \ldots.$

Using this argument and (3.24) we get that for

$$\eta^+_{q;j,j'} = \frac{Q}{Th}\int_{x^*+2qx}^{x^*+(2q+1)x} G^*_{q,j,j'}(t)\mathrm{d}t$$

$$
\begin{aligned}
|\eta^+_{q;j,j'}| &\leq \frac{C}{x}\int_{x^*+2qx}^{x^*+(2q+1)x} J|G^+_{q,j,j'}(t)|\mathrm{d}t \qquad (3.25)\\
&\leq C\frac{J}{x}\int_{x^*+2qx}^{x^*+(2q+1)x} \sum_{k,l\geq 1}\sum_{r,r'\geq 1} \mathbb{I}_{(|t-\tilde{t}^q_k-\tau^r_j|\leq CJ^{-1/d})}\mathbb{I}_{(|t-\tilde{t}^q_l-\tau^{r'}_{j'}|\leq CJ^{-1/d})}\mathrm{d}t\\
&\leq C\frac{J}{x}\int_{x^*+2qx}^{x^*+(2q+1)x} \sum_{k,l\geq 1}\sum_{r,r'\geq 1} \mathbb{I}_{(|t-t^q_k-\tau^r_j|\leq CJ^{-1/d})}\mathbb{I}_{(|t-t^q_l-\tau^{r'}_{j'}|\leq CJ^{-1/d})}\mathrm{d}t\\
&\leq C\frac{J}{x}\int_{x^*+2qx}^{x^*+(2q+1)x} \left(\sum_{k\geq 1}\sum_{r\geq 1} \mathbb{I}_{(|t-\tilde{t}^q_k-\tau^r_j|\leq CJ^{-1/d})}\right)^2\mathrm{d}t\\
&\qquad +C\frac{J}{x}\int_{x^*+2qx}^{x^*+(2q+1)x} \left(\sum_{l\geq 1}\sum_{r'\geq 1} \mathbb{I}_{(|t-\tilde{t}^q_l-\tau^{r'}_{j'}|\leq CJ^{-1/d})}\right)^2\mathrm{d}t.
\end{aligned}
$$

The first term on the right hand side of (3.25) can be bounded by

$$
\begin{aligned}
&\leq C\frac{J}{x}\sum_{i=1}^{I}\int_{x^*+2qx+C(i-1)J^{-1/d}}^{x^*+2qx+CiJ^{-1/d}} \left(\sum_{k\geq 1}\sum_{r\geq 1} \mathbb{I}_{(|\tilde{t}^q_k+\tau^r_j-x^*-2qx-CiJ^{-1/d}|\leq 2CJ^{-1/d})}\right)^2\mathrm{d}t\\
&\leq C\frac{1}{x}\sum_{i=1}^{I} Z^2_{q,i}
\end{aligned}
$$

with $Z_{q,i}=\sum_{k\geq 1}\sum_{r\geq 1}\mathbb{I}_{(|\tilde{t}^q_k+\tau^r_j-x^*-2qx-CiJ^{-1/d}|\leq 2CJ^{-1/d})}$ and where I is of order $xJ^{1/d}$. Note that for $1\leq q\leq Q$, $1\leq l\leq 4$ the variables $Z_{q,l}, Z_{q,l+4}, Z_{q,l+8}, ...$ are independent Poisson random variables with parameter of order $\log TJ^{-1/d}$. We now use that with $X_j=(Z^2_{q,l+4(j-1)}-\mathbb{E}[Z^2_{q,l+4(j-1)}])$ and with constants C_k depending on k we have that $\mathbb{E}[X_j^{2k}]\leq C_k\log TJ^{-1/d}$. Note that $\log TJ^{-1/d}$ converges to 0 according to Assumption (A3). By application of Rosenthals inequality, see Rosenthal (1970), we have that $\mathbb{E}[(X_1+...+X_{I/4})^{2k}]\leq C_k\{I\mathbb{E}[X_1^{2k}]+(I\mathbb{E}[X_1^2])^k\}\leq C_k(I\log TJ^{-1/d})^k$. Here we assume without loss of generality that I is a multiple of 4. This gives that $\mathbb{P}(x^{-1}(X_1+...+X_{I/2})>v)\leq C_k(\log T/x)^k v^{-2k}$. Because of $|\frac{1}{x}\sum_{i=1}^{I}\mathbb{E}[Z^2_{q,i}]|\leq C\log T$ for $C>0$ large enough, we get that

$$
\begin{aligned}
&\mathbb{P}\left(|\eta^+_{q;j,j'}|\geq C\log T+v \text{ for some } 1\leq q\leq Q, 1\leq j,j'\leq d\right) \qquad (3.26)\\
&\leq C\left(\frac{\log T}{x}\right)^k QJ^2v^{-2k}
\end{aligned}
$$

for $v>0$. We will use these considerations for the proof of the following lemma.

Lemma 3.3.7. *Make the assumptions of Theorem 3.2.1. Choose $\epsilon > 0$ small enough such that Q converges to infinity where Q is chosen as the smallest integer larger than $hT^{1-\epsilon} - 1$. Then for $1 \leq j, j' \leq J, 1 \leq q \leq Q$ there exist independent mean zero variables $\tilde{\eta}_{q,j,j'}$ and $\tilde{\eta}_{q,j}$ with $|\tilde{\eta}_{q,j,j'}| \leq CT^{\epsilon}$, $|\tilde{\eta}_{q,j}| \leq CT^{\epsilon}$ for some $C > 0$ such that for all $\kappa > 0$ there exists $C^* > 0$ with*

$$\left| \Delta^+_{j,j'} - \mathbb{E}[\Delta^+_{j,j'}] - \frac{1}{Q} \sum_{q=1}^{Q} \tilde{\eta}_{q,j,j'} \right| \leq C^* T^{-\kappa},$$

$$\left| \delta^+_{j} - \mathbb{E}[\delta^+_{j}] - \frac{1}{Q} \sum_{q=1}^{Q} \tilde{\eta}_{q,j} \right| \leq C^* T^{-\kappa}$$

with probability $\geq 1 - C^* T^{-\kappa}$. Here, Δ^+ is a $d \times d$ random matrix and δ^+ is a $\mathbb{R}^d$-valued random variable that have the same distribution as Δ or δ, respectively.

Proof. Choose Δ^+ as Δ but with $G(t)$ replaced by $G^+(t)$. With this choice the lemma holds. This can be seen by the considerations made above with the choice

$$\tilde{\eta}_{q,j,j'} = \eta^*_{q,j,j'} \mathbb{I}_{[|\eta^*_{q,j,j'}| \leq CT^{\epsilon}]} - \mathbb{E}\left[\eta^*_{q,j,j'} \mathbb{I}_{[|\eta^*_{q,j,j'}| \leq CT^{\epsilon}]} \right].$$

This follows by application of (3.26) with $x = T^{\epsilon}$ and with $k > 0$ large enough. In particular, one gets from (3.26) with $k > 0$ large enough that

$$\mathbb{E}\left[\eta^*_{q,j,j'} \mathbb{I}_{[|\eta^*_{q,j,j'}| \leq CT^{\epsilon}]} \right] \leq C^* T^{-\kappa}$$

for all $\kappa > 0$ with $C, C^* > 0$ large enough. This shows the statements on Δ^+ in the lemma. The statements for δ^+ can be shown by similar arguments for an appropriately chosen δ^+. □

Lemma 3.3.8. *Make the assumptions of Theorem 3.2.1. Then it holds uniformly for $\nu \in \mathbb{R}^d$ and $b_1, ..., b_J \in \mathbb{R}$ with $g = \sum_{j=1}^{J} b_j \psi_j$*

$$\begin{aligned} &(\nu, b_1, ...b_J)\, \Delta\, (\nu, b_1, ...b_J)^{\intercal} \\ &= \frac{1}{Th} \int_{t_0 - Th}^{t_0 + Th} \left(\nu + \int_{t-A}^{t-} g^{\intercal}(t-u) \mathrm{d}N_u \right)^2 L\left(\frac{t - t_0}{Th} \right) \mathrm{d}t \\ &\geq c \left(\|\nu\|^2 + \int_0^A g^{\intercal}(t) g(t) \mathrm{d}t \right) \end{aligned}$$

with probability tending to one for $c > 0$ small enough. In particular, we have that $\|\Delta^{-1} x\| \leq c^{-1} \|x\|$ for all $x \in \mathbb{R}^{J+1}$ with probability tending to one.

Proof. By application of Bernstein's inequality we get that for all $\epsilon, \kappa^* > 0$ with $C, C^* > 0$ large enough that for $1 \leq j, j' \leq J$

$$\mathbb{P}\left(\left|\frac{1}{Q}\sum_{q=1}^{Q}\tilde{\eta}_{q,j,j'}\right| > CT^{\epsilon}\frac{1}{\sqrt{hT}}\right) \quad \leq \quad C^*T^{-\kappa^*},$$

$$\mathbb{P}\left(\left|\frac{1}{Q}\sum_{q=1}^{Q}\tilde{\eta}_{q,j}\right| > CT^{\epsilon}\frac{1}{\sqrt{hT}}\right) \quad \leq \quad C^*T^{-\kappa^*}.$$

By application of Lemma 3.3.7 this gives with probability $\geq 1 - J^2C^*T^{-\kappa^*}$ that

$$\max_{1\leq j,j'\leq J} |\Delta_{j,j'} - \mathbb{E}[\Delta_{j,j'}]| \leq CT^{\epsilon}\frac{1}{\sqrt{hT}}$$

and thus $\|\Delta - \mathbb{E}[\Delta]\|_2 \leq CJT^{\epsilon}\frac{1}{\sqrt{hT}}$. By Assumption (A4) and Lemma 3.3.3 we get the result of the lemma. □

3.3.6 Treatment of the $\hat{\tau}$

According to (A4), we have that $|\beta_0^* - \nu_0^*| \leq \varepsilon_T$ and and $|\mu^{*,(l)}(u) - \sum_{j=1}^{J}\beta_j^*\psi_j(u)| \leq \varepsilon_T$. This remark shows that for the proof of Theorem 3.2.1 it remains to show the following lemma.

Lemma 3.3.9. *Make the assumptions of Theorem 3.2.1. Then* (3.8) – (3.9) *hold for* $j = 1, ..., J$.

Proof. For the proof of the second part of claim (3.8) note that for $1 \leq j \leq J$

$$\begin{aligned} \mathbb{E}[\hat{\tau}^2_{-0,A,j}] &= \frac{1}{(Th)^2}\int_{t_0-Th}^{t_0+Th}\mathbb{E}[(\int_{t-A}^{t-}\psi_j^\intercal(t-u)\mathrm{d}N_u)^2L^2\left(\frac{t-t_0}{Th}\right)\lambda(t)\mathrm{d}t \\ &\leq \frac{C}{(Th)^2}\int_{t_0-Th}^{t_0+Th}\mathbb{E}[N^2_{[t-A,t]} + N^3_{[t-A,t]}]\mathrm{d}t = O(1/(Th)) \end{aligned}$$

because of Lemma 3.3.4 and $\lambda(t) \leq C(1 + N_{[t-A,t]})$. The first part of claim (3.8) follows similarly.

For the proof of (3.9) one can use the inequality $|ab| \leq ((a^2/\varepsilon_T) +$

$(b^2\varepsilon_T))/2$. Note that

$$\mathbb{E}\Bigg[\frac{1}{Th}\int_{t_0-Th}^{t_0+Th} L\left(\frac{t-t_0}{Th}\right)\Bigg[\{\nu(t/T)-\beta_0^*\}$$
$$+\int_{t-A}^{t-}\Big\{\mu(t-s,t/T)-\sum_{j=1}^{J}\beta_j^*\psi_j^\intercal(t-s)\Big\}\mathrm{d}N_s\Bigg]^2\mathrm{d}t\Bigg]=O(\varepsilon_T^2),$$
$$\mathbb{E}\Bigg[\frac{1}{Th}\int_{t_0-Th}^{t_0+Th} L\left(\frac{t-t_0}{Th}\right)\Bigg[\int_{t-A}^{t-}\psi_j^\intercal(t-u)\mathrm{d}N_u\Bigg]^2\mathrm{d}t\Bigg]=O(1),$$

by Assumption (A4) and Lemma 3.3.4. The first part of Claims (3.9) can be shown by similar arguments. □

3.3.7 Additional Proofs

Proof of Lemma 3.3.1. Using the conditional intensity defined in (3.17) we can construct a stationary homogeneous process $\bar{N}_t$ with the property that all jumps of N_t are also jumps of $\bar{N}_t$. In particular, this implies the existence of N_t. Note that

$$\begin{aligned}
\lambda(t) &= \nu\left(\frac{t}{T}\right)+\int\mu\left(t-u,\frac{t}{T}\right)\mathrm{d}N_u\\
&= \nu\left(\frac{t}{T}\right)+\int\mu\left(t-u,\frac{t}{T}\right)\lambda(u)\mathrm{d}u+\int\mu\left(t-u,\frac{t}{T}\right)\mathrm{d}M_u\\
&= \nu\left(\frac{t}{T}\right)+\int\mu\left(t-u,\frac{t}{T}\right)\nu\left(\frac{u}{T}\right)\mathrm{d}u+\int\mu^{(*2)}\left(t-u,\frac{t}{T}\right)\lambda(u)\mathrm{d}u\\
&\quad +\int\sum_{l=1}^{2}\mu^{(*l)}\left(t-u,\frac{t}{T}\right)\mathrm{d}M_u\\
&= \nu\left(\frac{t}{T}\right)+\int\sum_{l=1}^{k-1}\mu^{(*l)}\left(t-u,\frac{t}{T}\right)\nu\left(\frac{u}{T}\right)\mathrm{d}u\\
&\quad +\int\mu^{(*k)}\left(t-u,\frac{t}{T}\right)\lambda(u)\mathrm{d}u+\int\sum_{l=1}^{k}\mu^{(*l)}\left(t-u,\frac{t}{T}\right)\mathrm{d}M_u
\end{aligned}$$

for $k\geq 1$. Because λ and $\mu^{(*l)}$ are positive and can be bounded by $\bar{\lambda}$ and $\bar{\mu}^{(*l)}$, where $\bar{\mu}^{(*l)}$ is defined as $\mu^{(*l)}$ but with N_t replaced by $\bar{N}_t$, we can conclude that the right hand side of the last equation converges to

$$\nu\left(\frac{t}{T}\right)+\int\chi\left(t-u,\frac{t}{T}\right)\nu\left(\frac{u}{T}\right)\mathrm{d}u+\int\chi\left(t-u,\frac{t}{T}\right)\mathrm{d}M_u.$$

This convergence holds almost shurely and also in the sense that the second moment of the difference converges to 0. Thus, $\lambda(t)$ is equal to this expression and we conclude (3.12) by taking the expectation of this expression. This also implies (3.13).

For the proof of (3.14) one proceeds similarly as in the proof of Proposition 2 in Bacry, Dayri and Muzy (2012). Note that:

$$\mathbb{E}\left[\mathrm{d}N_t \mathrm{d}N_{t'}^\intercal\right] = I_1 + ... + I_4$$

with

$$I_1 = \mathbb{E}\left[\mathrm{d}M_t \mathrm{d}M_{t'}^\intercal\right] = \Sigma_{t/T}\delta_{t-t'}\mathrm{d}t\mathrm{d}t',$$

$$\begin{aligned}
I_2 &= \mathbb{E}\left[\lambda(t)\mathrm{d}M_{t'}^\intercal\right]\mathrm{d}t \\
&= \mathbb{E}\left[\left\{\Lambda\left(\frac{t}{T}\right) + \int \chi\left(t-s, \frac{t}{T}\right)\mathrm{d}M_s\right\}\mathrm{d}M_{t'}^\intercal\right]\mathrm{d}t \\
&= \int \chi\left(t-s, \frac{t}{T}\right)\Sigma_{s/T}\delta_{s-t'}\mathrm{d}s\mathrm{d}t\mathrm{d}t' \\
&= \chi\left(t-t', \frac{t}{T}\right)\Sigma_{t'/T}\mathrm{d}t\mathrm{d}t', \\
I_3 &= \Sigma_{t/T}\chi\left(t'-t, \frac{t'}{T}\right)^\intercal \mathrm{d}t\mathrm{d}t', \\
I_4 &= \mathbb{E}\left[\lambda(t)\lambda(t')^\intercal\right]\mathrm{d}t\mathrm{d}t' \\
&= \mathbb{E}\left[\left\{\Lambda\left(\frac{t}{T}\right) + \int \chi\left(t-s, \frac{t}{T}\right)\mathrm{d}M_s\right\}\lambda(t')^\intercal\right]\mathrm{d}t\mathrm{d}t' \\
&= \Lambda\left(\frac{t}{T}\right)\Lambda\left(\frac{t'}{T}\right)^\intercal \mathrm{d}t\mathrm{d}t' \\
&\qquad + \mathbb{E}\left[\left\{\int \chi\left(t-s, \frac{t}{T}\right)\mathrm{d}M_s\right\}\left\{\int \chi\left(t'-u, \frac{t'}{T}\right)^\intercal \mathrm{d}M_u\right\}\right]\mathrm{d}t\mathrm{d}t' \\
&= \Lambda\left(\frac{t}{T}\right)\Lambda\left(\frac{t'}{T}\right)^\intercal \mathrm{d}t\mathrm{d}t' + \int \chi\left(t-s, \frac{t}{T}\right)\Sigma_{s/T}\chi\left(t'-s, \frac{t'}{T}\right)^\intercal \mathrm{d}s\mathrm{d}t\mathrm{d}t'.
\end{aligned}$$

This shows equation (3.14). □

Proof of Lemma 3.3.3. For the second statement of the lemma note that for $\beta = (\beta_0, ..., \beta_J)^\intercal$ it holds that $\beta^\intercal \mathbb{E}[\Delta]\beta = R(g, \beta_0)$ with $g(u) = \sum_{j=1}^J \beta_j \psi_j(u)$. For the proof of (3.16) we will show that

$$R(g) \geq c^* \int_0^A g^\intercal(t)g(t)\mathrm{d}t \tag{3.27}$$

for some constant $c^* > 0$. Note that for all $0 < D \leq 1$ with a constant $C > 0$

$$\begin{aligned} R^*(g,\gamma) &\geq DR^*(g,\gamma) \\ &\geq \frac{D}{2}\gamma^2 - DC\int_0^A g^\intercal(t)g(t)\mathrm{d}t, \end{aligned}$$

where the inequality $(a+b)^2 \geq a^2/2 - b^2$ has been used. Choosing $D > 0$ small enough we get inequality (3.16) from (3.28) and from the last lemma if one chooses $c > 0$ small enough.

For the proof of (3.28) note that because of (A2) with $c > 0$ small enough for functions g with $\int g(w)^\intercal g(w)\mathrm{d}w = 1$ the left hand side of (3.28) can be bounded from below by

$$\begin{aligned} & c\frac{1}{Th}\int_{t_0-Th}^{t_0+Th}\int\left(g(w)+\int\chi\left(w-s,\frac{t-s}{T}\right)^\intercal g(s)\mathrm{d}s\right)^\intercal \\ & \qquad \left(g^\intercal(w)+\int\chi\left(w-s,\frac{t-s}{T}\right)^\intercal g(s)\mathrm{d}s\right)L\left(\frac{t-t_0}{Th}\right)\mathrm{d}w\mathrm{d}t \\ & \geq c\frac{1}{Th}\int_{t_0-Th}^{t_0+Th}\int\left(g(w)+\int\chi^{\#}\left(w-s\right)^\intercal g(s)\mathrm{d}s\right)^\intercal \\ & \qquad \left(g^\intercal(w)+\int\chi^{\#}\left(w-s\right)^\intercal g(s)\mathrm{d}s\right)L\left(\frac{t-t_0}{Th}\right)\mathrm{d}w\mathrm{d}t - Ch \\ & \geq c\int\left(g(w)+\int\chi^{\#}\left(w-s\right)^\intercal g(s)\mathrm{d}s\right)^\intercal \\ & \qquad \left(g(w)+\int\chi^{\#}\left(w-s\right)^\intercal g(s)\mathrm{d}s\right)\mathrm{d}w - Ch \\ & \\ & \geq c+\int\int g(w)^\intercal\sum_{k=1}^{\infty}\mu^{\#,(*k)}\left(w-s\right)g(s)\mathrm{d}s\mathrm{d}w \\ & \qquad +\int\int g(w)^\intercal\sum_{k=1}^{\infty}\mu^{\#,(*2k)}\left(w-s\right)g(s)\mathrm{d}s\mathrm{d}w - Ch \\ & \geq c - Ch \geq c \end{aligned}$$

for $c > 0$ small enough and T large enough. Here we use our convention that a constant $c > 0$ may denote different constants also in the same formula. Furthermore, we used $\chi^{\#}\left(t-s\right) = \sum_{k=1}^{\infty}\mu^{\#,(*k)}\left(t-s\right)$ with $\mu^{\#,(*1)}\left(t-s\right) = \mu^*\left(t-s,t_0/T\right)$ and

$$\mu^{\#,(*k)}\left(t-s\right) = \int\mu^{\#,(*(k-1))}\left(t-u\right)\mu^{\#,(*1)}\left(u-s\right)\mathrm{d}u$$

for $k \geq 2$. Note that the operator norm of $\mu^{\#,(*k)}(w-s) - \mu^{(*k)}\left(w-s, \frac{t-s}{T}\right)$ is bounded by $C(\gamma^+)^{k-1}h$, uniformly for $k \geq 1$, $|w-s| \leq A$ and $|t-t_0| \leq CTh$. Here, we used (A1) to get the last inequality. Note that this implies that the operator norm of $\chi^{\#}(t-s) - \chi\left(t-s, \frac{t-s}{T}\right)$ is bounded by Ch, again uniformly $k \geq 1$, $|w-s| \leq A$ and $|t-t_0| \leq CTh$. This shows (3.28) and concludes the proof of the lemma. □

Proof of Lemma 4.1. Using the conditional intensity defined in equation (18) of the main text we can construct a stationary homogeneous process $\bar{N}_t$ with the property that all jumps of N_t are also jumps of $\bar{N}_t$. In particular, this implies the existence of N_t. Note that

$$\begin{aligned}
\lambda(t) &= \nu\left(\frac{t}{T}\right) + \int \mu\left(t-u, \frac{t}{T}\right) \mathrm{d}N_u \\
&= \nu\left(\frac{t}{T}\right) + \int \mu\left(t-u, \frac{t}{T}\right) \lambda(u)\mathrm{d}u + \int \mu\left(t-u, \frac{t}{T}\right) \mathrm{d}M_u \\
&= \nu\left(\frac{t}{T}\right) + \int \mu\left(t-u, \frac{t}{T}\right) \nu\left(\frac{u}{T}\right) \mathrm{d}u + \int \mu^{(*2)}\left(t-u, \frac{t}{T}\right) \lambda(u)\mathrm{d}u \\
&\quad + \int \sum_{l=1}^{2} \mu^{(*l)}\left(t-u, \frac{t}{T}\right) \mathrm{d}M_u \\
&= \nu\left(\frac{t}{T}\right) + \int \sum_{l=1}^{k-1} \mu^{(*l)}\left(t-u, \frac{t}{T}\right) \nu\left(\frac{u}{T}\right) \mathrm{d}u \\
&\quad + \int \mu^{(*k)}\left(t-u, \frac{t}{T}\right) \lambda(u)\mathrm{d}u + \int \sum_{l=1}^{k} \mu^{(*l)}\left(t-u, \frac{t}{T}\right) \mathrm{d}M_u
\end{aligned}$$

for $k \geq 1$. Because λ and $\mu^{(*l)}$ are positive and can be bounded by $\bar{\lambda}$ and $\bar{\mu}^{(*l)}$, where $\bar{\mu}^{(*l)}$ is defined as $\mu^{(*l)}$ but with N_t replaced by $\bar{N}_t$, we can conclude that the right hand side of the last equation converges to

$$\nu\left(\frac{t}{T}\right) + \int \chi\left(t-u, \frac{t}{T}\right) \nu\left(\frac{u}{T}\right) \mathrm{d}u + \int \chi\left(t-u, \frac{t}{T}\right) \mathrm{d}M_u.$$

This convergence holds almost shurely and also in the sense that the second moment of the difference converges to 0. Thus, $\lambda(t)$ is equal to this expression and we conclude equation (13) of the main text by taking the expectation of this expression. This also implies equation (14) of the main text.

For the proof of equation (15) of the main text one proceeds similarly as in the proof of Proposition 2 in Bacry, Dayri and Muzy (2012). Note that:

$$\mathbb{E}\left[\mathrm{d}N_t \mathrm{d}N_{t'}^{\intercal}\right] = I_1 + \ldots + I_4$$

with

$$I_1 = \mathbb{E}\left[\mathrm{d}M_t \mathrm{d}M_{t'}^{\intercal}\right] = \Sigma_{t/T}\delta_{t-t'}\mathrm{d}t\mathrm{d}t',$$

$$\begin{aligned}
I_2 &= \mathbb{E}\left[\lambda(t)\mathrm{d}M_{t'}^{\intercal}\right]\mathrm{d}t \\
&= \mathbb{E}\left[\left\{\Lambda\left(\frac{t}{T}\right) + \int \chi\left(t-s, \frac{t}{T}\right)\mathrm{d}M_s\right\}\mathrm{d}M_{t'}^{\intercal}\right]\mathrm{d}t \\
&= \int \chi\left(t-s, \frac{t}{T}\right)\Sigma_{s/T}\delta_{s-t'}\mathrm{d}s\mathrm{d}t\mathrm{d}t' \\
&= \chi\left(t-t', \frac{t}{T}\right)\Sigma_{t'/T}\mathrm{d}t\mathrm{d}t', \\
I_3 &= \Sigma_{t/T}\chi\left(t'-t, \frac{t'}{T}\right)^{\intercal}\mathrm{d}t\mathrm{d}t', \\
I_4 &= \mathbb{E}\left[\lambda(t)\lambda(t')^{\intercal}\right]\mathrm{d}t\mathrm{d}t' \\
&= \mathbb{E}\left[\left\{\Lambda\left(\frac{t}{T}\right) + \int \chi\left(t-s, \frac{t}{T}\right)\mathrm{d}M_s\right\}\lambda(t')^{\intercal}\right]\mathrm{d}t\mathrm{d}t' \\
&= \Lambda\left(\frac{t}{T}\right)\Lambda\left(\frac{t'}{T}\right)^{\intercal}\mathrm{d}t\mathrm{d}t' \\
&\qquad + \mathbb{E}\left[\left\{\int \chi\left(t-s, \frac{t}{T}\right)\mathrm{d}M_s\right\}\left\{\int \chi\left(t'-u, \frac{t'}{T}\right)^{\intercal}\mathrm{d}M_u\right\}\right]\mathrm{d}t\mathrm{d}t' \\
&= \Lambda\left(\frac{t}{T}\right)\Lambda\left(\frac{t'}{T}\right)^{\intercal}\mathrm{d}t\mathrm{d}t' + \int \chi\left(t-s, \frac{t}{T}\right)\Sigma_{s/T}\chi\left(t'-s, \frac{t'}{T}\right)^{\intercal}\mathrm{d}s\mathrm{d}t\mathrm{d}t'.
\end{aligned}$$

This shows equation equation (15) of the main text. □

Proof of Lemma 4.3. For the second statement of the lemma note that for $\beta = (\beta_0, ..., \beta_J)^{\intercal}$ it holds that $\beta^{\intercal}\mathbb{E}[\Delta]\beta = R(g, \beta_0)$ with $g(u) = \sum_{j=1}^{J}\beta_j\psi_j(u)$. For the proof of equation (17) of the main text we will show that

$$R(g) \quad \geq \quad c^* \int_0^A g^{\intercal}(t)g(t)\mathrm{d}t \tag{3.28}$$

for some constant $c^* > 0$. Note that for all $0 < D \leq 1$ with a constant $C > 0$

$$\begin{aligned}
R^*(g,\gamma) &\geq DR^*(g,\gamma) \\
&\geq \frac{D}{2}\gamma^2 - DC\int_0^A g^{\intercal}(t)g(t)\mathrm{d}t,
\end{aligned}$$

where the inequality $(a+b)^2 \geq a^2/2 - b^2$ has been used. Choosing $D > 0$ small enough we get inequality (17) of the main text from (3.28) and from the last lemma if one chooses $c > 0$ small enough.

For the proof of (3.28) note that because of (A2) with $c > 0$ small enough for functions g with $\int g(w)^\intercal g(w)\mathrm{d}w = 1$ the left hand side of (3.28) can be bounded from below by

$$\begin{aligned}
&c\frac{1}{Th}\int_{t_0-Th}^{t_0+Th}\int\left(g(w)+\int\chi\left(w-s,\frac{t-s}{T}\right)^\intercal g(s)\mathrm{d}s\right)^\intercal \\
&\quad\left(g^\intercal(w)+\int\chi\left(w-s,\frac{t-s}{T}\right)^\intercal g(s)\mathrm{d}s\right)L\left(\frac{t-t_0}{Th}\right)\mathrm{d}w\mathrm{d}t \\
&\geq c\frac{1}{Th}\int_{t_0-Th}^{t_0+Th}\int\left(g(w)+\int\chi^{\#}(w-s)^\intercal g(s)\mathrm{d}s\right)^\intercal \\
&\quad\left(g^\intercal(w)+\int\chi^{\#}(w-s)^\intercal g(s)\mathrm{d}s\right)L\left(\frac{t-t_0}{Th}\right)\mathrm{d}w\mathrm{d}t - Ch \\
&\geq c\int\left(g(w)+\int\chi^{\#}(w-s)^\intercal g(s)\mathrm{d}s\right)^\intercal \\
&\quad\left(g(w)+\int\chi^{\#}(w-s)^\intercal g(s)\mathrm{d}s\right)\mathrm{d}w - Ch
\end{aligned}$$

$$\begin{aligned}
&\geq c + \int\int g(w)^\intercal\sum_{k=1}^{\infty}\mu^{\#,(*k)}(w-s)\,g(s)\mathrm{d}s\mathrm{d}w \\
&\quad + \int\int g(w)^\intercal\sum_{k=1}^{\infty}\mu^{\#,(*2k)}(w-s)\,g(s)\mathrm{d}s\mathrm{d}w - Ch \\
&\geq c - Ch \geq c
\end{aligned}$$

for $c > 0$ small enough and T large enough. Here we use our convention that a constant $c > 0$ may denote different constants also in the same formula. Furthermore, we used $\chi^{\#}(t-s) = \sum_{k=1}^{\infty}\mu^{\#,(*k)}(t-s)$ with $\mu^{\#,(*1)}(t-s) = \mu^*(t-s, t_0/T)$ and

$$\mu^{\#,(*k)}(t-s) = \int\mu^{\#,(*(k-1))}(t-u)\,\mu^{\#,(*1)}(u-s)\,\mathrm{d}u$$

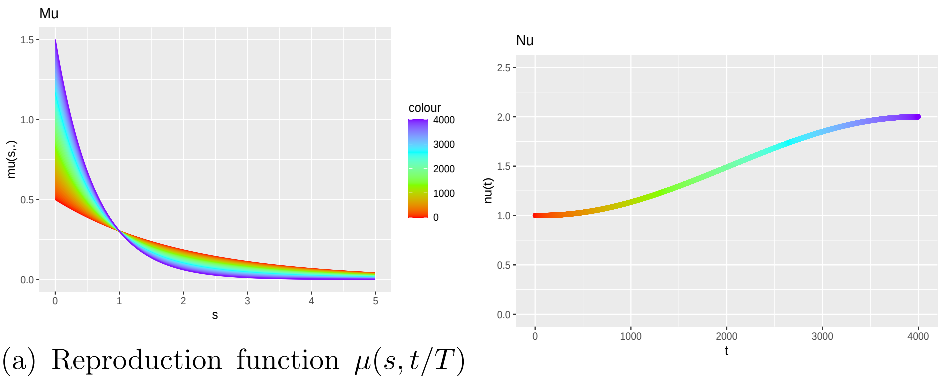
for $k \geq 2$. Note that the operator norm of $\mu^{\#,(*k)}(w-s) - \mu^{(*k)}\left(w-s,\frac{t-s}{T}\right)$ is bounded by $C(\gamma^+)^{k-1}h$, uniformly for $k \geq 1$, $|w-s| \leq A$ and $|t-t_0| \leq CTh$. Here, we used (A1) to get the last inequality. Note that this implies that the operator norm of $\chi^{\#}(t-s) - \chi\left(t-s,\frac{t-s}{T}\right)$ is bounded by Ch, again uniformly $k \geq 1$, $|w-s| \leq A$ and $|t-t_0| \leq CTh$. This shows (3.28) and concludes the proof of the lemma. □

3.4 Simulation Study

We now illustrate our approach by a small simulation study. We generated one-dimensional locally stationary Hawkes processes on a time frame $[0, T]$ with $T = 4000$. This was done for the following choice of the reproduction function μ and the conditional baseline intensity function ν with support $[0, A] = [0, 5]$:

$$\mu\Big(s, \frac{t}{T}\Big) := \left(\frac{3}{2}\exp\left(\frac{-8s}{5}\right) \cdot \frac{t}{T} + \frac{1}{2}\exp\left(\frac{-s}{2}\right)\Big(1 - \frac{t}{T}\Big)\right) \cdot 1_{[0,A]}(s),$$
$$\nu\Big(\frac{t}{T}\Big) := \frac{3}{2} + \frac{1}{2}\cos\left(\Big(1 - \frac{t}{T}\Big)\pi\right).$$

For a plot of the functions see Figure 3.1 where the second argument of μ is despicted in the colour scheme as explained in the legend and again illustrated in the graph of ν.



(a) Reproduction function $\mu(s, t/T)$ with argument t indicated by colour.

(b) Immigration function $\nu(t/T)$.

Figure 3.1: Choice of reproduction functions and conditional baseline intensity functions in the simulation.

We calculated our locally linear estimator at $t_0 = 2000$ for the choice $K = 1$ and $J = 7$ using kernel smoothing with biweight kernel $\text{Kernel}_{biw} : [-1, 1] \to \mathbb{R}, t \mapsto \frac{15}{16}(1 - t^2)^2$. The seven B-Spline basis elements of degree two which are used are depicted in Figure 3.2. The knots $(0, 0.5, 1, 2, 3, 5)$ are not equidistant.

The bandwidth h is chosen such that 20000 events are contained in the support of the smoothing kernel. The simulation used 200 replications. The estimated conditional baseline intensity function ν is despicted via a boxplot in Figure 3.3 (a). The green graph represents the true value of ν at time t_0.

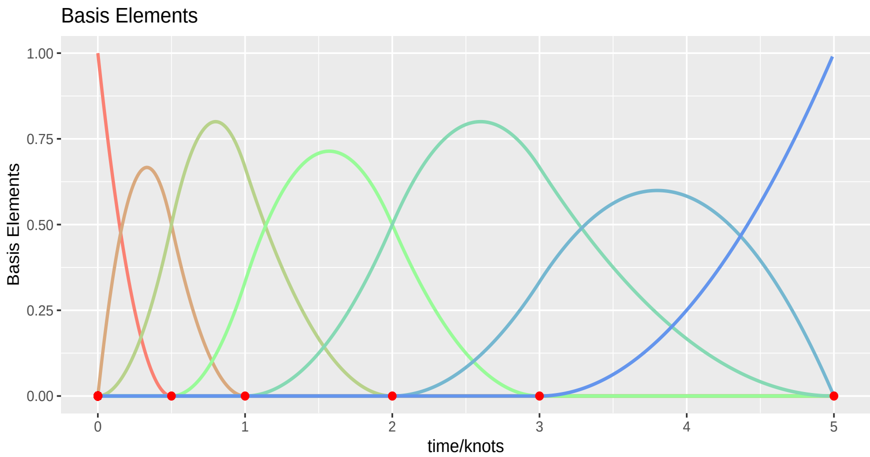


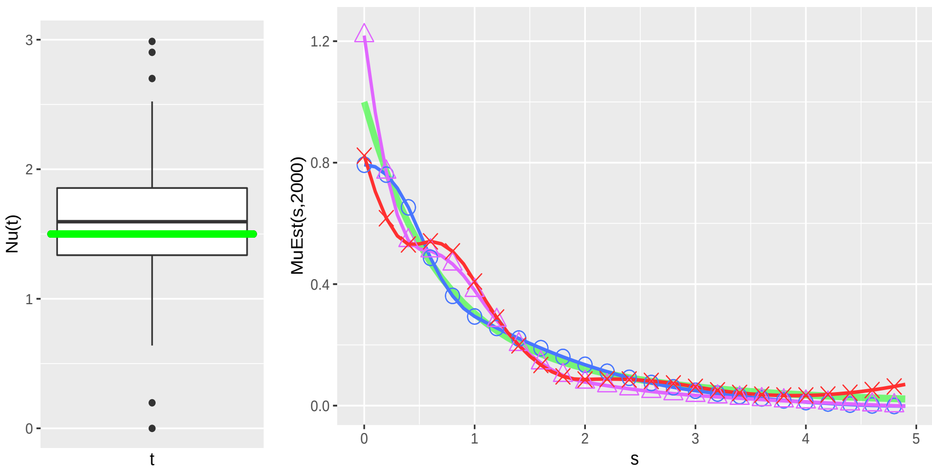
Figure 3.2: Basis elements in B-spline estimation.

In Figure 3.3 (b) the 0.25-, 0.50-, 0.75- best estimated reproduction functions with respect to the integrated squared error are shown (in blue, purple and red respectively). Recall that

$$ISE(\tilde{\mu}) := \int_0^A \left(\mu\left(u, \frac{t_0}{T}\right) - \tilde{\mu}(u)\right)^2 \, \mathrm{d}u$$

for an estimator $\tilde{\mu}$ of $\mu(\cdot, t_0/T)$. The true reproduction function is displayed in transparent green for comparision.

We see that the conditional baseline intensity function was not always accurately estimated. In fact, there are some outliers. On the other side μ was estimated quite accurately. In Figure 3.4 the tuples of ISE for the estimated reproduction function μ and the squared error of the estimated conditional baseline intensity ν are shown.



(a) Boxplot of simulation results for the estimation of $\nu(1/2)$. True value is $1/2$.

(b) Plots of simulated estimators of μ with 0.25-, 0.50-, 0.75 best ISE performance, in blue, purple and red respectively. Simulated reproduction function in transparent green.

Figure 3.3: Performance of simulated estimators of ν and μ at $t_0 = 2000$.

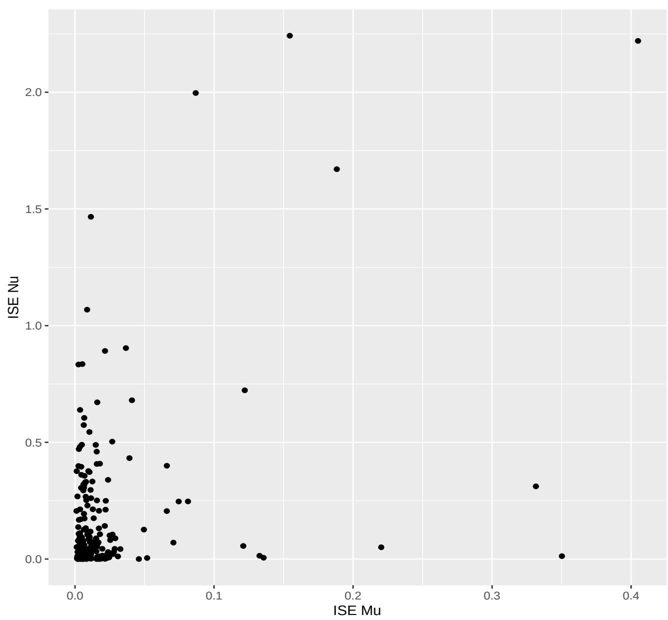


Figure 3.4: Plot of ISE of estimated μ versus SE of estimated ν.

4 | Testing for Local Alignment

The developed estimation procedure from Chapter 3 is inspiring the formulation of testing results. Suppose, we observe two Hawkes processes and want to test whether their reproduction function and immgration coincide at a designated point in time. Alternatively one could compare the behaviour of one Hawkes process at two distinct time points.

We will not make all the same assumptions on the model as in Chapter 3, as there are slightly different requirements to be fulfilled for the necessary tools as well as technical calculations. We do still consider the same model for Hawkes processes N which are defined via their conditional intensity function

$$\lambda(t) = \nu\left(\frac{t}{T}\right) + \int_{\infty}^{t-} \mu\left(t - s, \frac{t}{T}\right) \, \mathrm{d}N_s$$

for $t \in [0, T]$ in an asymptotic framework where $T \to \infty$. Note, that this is the vector notation and $\mu : \mathbb{R}_0^+ \times [0,1] \to (\mathbb{R}_0^+)^{d \times d}$ and $\nu : \times[0,1] \to (\mathbb{R}_0^+)^d$.

The underlying true parameter functions are denoted by μ_0 and ν_0 and the corresponding intensity function is called λ_0 as before. We denote by L a kernel function in the sense that $L : [-1,1] \to \mathbb{R}$ with $\int_{\mathbb{R}} L(x) \, \mathrm{d}x = 1$ and $\int_{\mathbb{R}} xL(x) \, \mathrm{d}x = 0$.

According to our intuition and as it is handled in chapter 3, it is not necessarily restrictive to assume, that μ has support in the compact interval $[0, A]$. In most examples, we would not expect that events in the past carry much relevance in the far future. The test statictic, that we will be considering allows us to let μ have unbounded support over $\mathbb{R}^+$, while we make additional restrictions. We will see, that a reproduction function with compact support fulfills the newly formulated requirements.

4.1 Motivation

In application we would observe two Hawkes processes $N^{(1)}, N^{(2)}$ and want to test whether the underlying models of the realizations $N^{(1)}$ and $N^{(2)}$ coincide in the following sense.

$$\mu_0^{(1)}(s, t_0/T) = \mu_0^{(2)}(s, t_0/T), \tag{4.1}$$

$$\nu_0^{(1)}(t_0/T) = \nu_0^{(2)}(t_0/T) \tag{4.2}$$

at a fixed point of interest $t_0/T \in [0,1]$ and all $s \in \mathbb{R}^+ = (0, \infty)$. Respectively we write $\cdot^{(1)}$ and $\cdot^{(2)}$ whenever we consider two different realizations.
Alternatively, one could compare the behaviour of one Hawkes process at two different time points $t_{(1)} < t_{(2)}$. Formally we would then consider $N^{(1)} := N$ and $N^{(2)} = N_{-(t_{(2)}-t_{(1)})}$ given by $N_t^{(2)} := N_{t-(t_{(2)}-t_{(1)})}$ and use the same local procedure at timepoint $t_{(1)}$.
As simulations have shown, estimation in such manner is very much unstable, which is why we remain in a finite dimensional setting as will be introduced below.

4.2 Assumptions and Setup

The following set of assumptions are made throughout this chapter. These will be cited explicitely, whenever they are used. Keep in mind, that they partly vary from the assumptions in Chapter 3.

Assumption 4.2.1. *(A1) The functions $\mu_0^{l,m}$ are positive and bounded for all $1 \le l, m \le d$ and the support of μ_0 is contained in $\mathbb{R}_0^+ \times [0,1]$.*
Define Γ^+ as the matrix

$$\left(\int_{\mathbb{R}_0^+} \sup_{x \le 1} \mu_0^{l,m}(s, x) \, \mathrm{d}s \right)_{1 \le l,m \le d}.$$

Γ^+ has spectral radius smaller than 1.
The function ν_0 is bounded and bounded away from 0 for any values smaller oder equal to 1.
The partial derivatives of μ_0 with respect to the second argument exist and are uniformly absolutely bounded.
Suppose that the basis functions ϕ_j are absolutely bounded and have compact support for all $j = 1, \dots, J$.

(A2) Concerning the asymptotic behaviour of T and h suppose, that

$$Th \to \infty,$$
$$\sqrt{Th^5} \to 0$$

for $T \to \infty$ and $h \to 0$ and there exists a $\delta > 0$, such that

$$h > T^{-\delta}.$$

Recall the following definition from chapter 3.

$$\chi(t-s, t/T) = \sum_{i=1}^{\infty} \mu_0^{(*i)}(t-s, t/T),$$
$$\mu_0^{(*i)}(t-s, t/T) = \int_{\mathbb{R}} \mu_0^{(*(i-1))}(t-u, t/T)\mu_0(u-s, u/T)\ \mathrm{d}u$$
$$\mu_0^{(*1)} = \mu_0.$$

In contrast to the last chapter we endow χ with an additional assumption.

(A3) For any $s \in \mathbb{R}_0^+$ and $x \in [0,1]$ there exist constants $c, c' > 0$ such that

$$\chi(s, x) \leq c \exp(-c's).$$

(A4) For the designated point t_0 from (4.1) and (4.2) and $u \in [0, T]$ there exist constants $c_1, c_2 > 0$ such that

$$|\chi(t-u, t/T) - \chi(t-t_0, t/T)| \leq c_1 |u - t_0|$$
$$|\chi(t-u, t/T) - \chi(t-u, t_0/T)| \leq c_2 \left| \frac{t-t_0}{T} \right|.$$

Remark. *The assumption (A3) holds for μ with compact support. In this case (A4) holds with a weak additional assumption on μ. For a disscussion on that topic, see the Appendix 4.5 of chapter 4.*

Consistently, we write for functions $f : \mathbb{R}_0^+ \to \mathbb{R}, h : [0,1] \to \mathbb{R}$ with only one argument and functions $g : \mathbb{R}_0^+ \times [0,1]$

$$(f * g)(t-s) := \int_{\mathbb{R}} f(t-r)g(r-s, r/T)\ \mathrm{d}r,$$
$$(g * f)(t-s, t/T) := \int_{\mathbb{R}} g(t-r, t/T)f(r-s)\ \mathrm{d}r,$$
$$(f * h)(t) := \int_{\mathbb{R}} f(t-r)h(r/T)\ \mathrm{d}r.$$

Note, that this notation is associative but not necessarily symmetric, e.g. it might hold that $(f * \chi) \neq (\chi * f)$.

We find motivation in the estimator investigated in chapter 3. Recall the estimators from chapter 3 as $\hat{\theta} = \Delta^{-1}\hat{\tau}$ for invertible Δ for a bandwidth $h \to 0$, the Kernel function L, the basis of functions on $\mathbb{R}_0^+$, namely $\{\phi_j\}_{j=1,\dots,J}$ and Δ and $\hat{\tau}$ as in 3.1.

4.3 Testing Procedure

We proceed on the idea, that the two underlying Hawkes models coincide whenever the estimators or related terms have a small square distance. Hence, we define a suitable test statistic for each component $j = 0, \dots, J$.

$$\mathcal{T}_j = \hat{\tau}_j^{(1)} - \hat{\tau}_j^{(2)}.$$

Recall, that

$$\hat{\tau} = (\tau_0^T, \tau_1^T, \dots, \tau_J^T)^T$$

with

$$\hat{\tau}_0 := \left(\frac{1}{Th} \int_{t_0-Th}^{t_0+Th} \left(\frac{t-t_0}{Th} \right)^{k-1} L\left(\frac{t-t_0}{Th} \right) \mathrm{d}N_t \right)_{k=1,\dots,K},$$

$$\hat{\tau}_j := \left(\frac{1}{Th} \int_{t_0-Th}^{t_0+Th} \left(\frac{t-t_0}{Th} \right)^{k-1} L\left(\frac{t-t_0}{Th} \right) \int_{-\infty}^{t_-} \phi_j(t-u)\ \mathrm{d}N_u \mathrm{d}N_t \right)_{k=1,\dots,K},$$

for $j = 1, \dots, J$. Note, that we will proof assertions for $\mathcal{T}_j$, i.e. in a componentwise sense.

We consider a framework with J basis functions. This is due to the unstable nature of the procedure which became clear in the simulation studies in 3.4. The integer K is chosen such that the estimation procedure captures the slope of the true parameters well enough.

Theorem 4.3.1. *Make the Assumptions 4.2.1. It holds that*

$$\sqrt{Th}(\hat{\tau}^{(1)} - \hat{\tau}^{(2)}) \to \mathcal{N}(0, \Sigma^{(1)} + \Sigma^{(2)})$$

in probability for $Th \to \infty$ and a covariance matrix $\Sigma^{(i)}$ as stated in Theorem 4.4.1.

4.4 Proofs

Theorem 4.3.1 can be shown by using the Cramér-Wold device. One shows asymptotic normality of

$$a_{01}\hat{\tau}_{01} + \ldots + a_{0K}\hat{\tau}_{0K} + a_{11}\hat{\tau}_{11} + \ldots + a_{JK}\hat{\tau}_{JK}$$

for a fixed matrix $(a_0, \ldots, a_J) \in \mathbb{R}^{K\times(J+1)}$.
For simplicity we show the statement for $\hat{\tau}_j$ for $j = 0, \ldots, J$. Furthermore, without loss of generality we consider the case of a one-dimensional Hawkes process ($d = 1$) and $K = 1$. Whenever we omit the index i, i.e. write λ instead of $\lambda^{(i)}$, as well as μ_0 instead of $\mu_0^{(i)}$, we refer to the terms corresponding to the same $i \in \{1, 2\}$.

4.4.1 Proof of Theorem 4.3.1

We will show the following theorem.

Theorem 4.4.1. *Make the Assumptions 4.2.1. It holds that for* $i = 1, 2$

$$\sqrt{Th}(\hat{\tau}_0^{(i)} - \mathbb{E}[\hat{\tau}_0^{(i)}]) \to \mathcal{N}(0, \Sigma_0^{(i)})$$

and for $j = 1, \ldots, J$

$$\sqrt{Th}(\hat{\tau}_j^{(i)} - \mathbb{E}[\hat{\tau}_j^{(i)}]) \to \mathcal{N}(0, \Sigma_j^{(i)})$$

in probability with matrices $\Sigma_j^{(i)}$ *stated in the proof.*

Then, Theorem 4.3.1 immediately follows from Theorem 4.4.1 as a corollary if we can show the following lemma concering the expectation of the $\hat{\tau}_j$.

Lemma 4.4.2. *For any* $j = 0, \ldots, J$ *there exists a constant* $c > 0$*, such that*

$$\left|\mathbb{E}[\hat{\tau}_j^{(1)}] - \mathbb{E}[\hat{\tau}_j^{(2)}]\right| = c \cdot h^2.$$

We give a brief overview of the proof which is split into many lemmas and then prove the remaining assertions in section 4.4.2.

We start by writing the terms of interest in a different notion to be used afterwards.

Proof of Theorem 4.4.1: First of all, we show the following lemma, giving an alternative form of the $\hat{\tau}_j$.

Lemma 4.4.3. *Under assumption (A1) from Assumptions 4.2.1 it holds that*

$$\begin{aligned}
\hat{\tau}_0 &= \frac{1}{Th}\int_{t_0-Th}^{t_0+Th} L\Big(\frac{t-t_0}{Th}\Big)\mathrm{d}N_t \\
&= \tau_0 + \frac{1}{Th}\int_{\mathbb{R}} L\Big(\frac{t-t_0}{Th}\Big)g_0(t)\ \mathrm{d}M_t + O_{\mathbb{P}}\Big(\frac{1}{Th}\Big), \\
\hat{\tau}_j &= \frac{1}{Th}\int_{\mathbb{R}} L\Big(\frac{t-t_0}{Th}\Big)\Big(\int_{-\infty}^{t-}\phi_j(t-u)\ \mathrm{d}N_u\Big)\mathrm{d}N_t \\
&= \tau_j + O_{\mathbb{P}}\Big(\frac{1}{Th}\Big) \\
&\quad + \frac{1}{Th}\int_{\mathbb{R}} L\Big(\frac{t-t_0}{Th}\Big)\Big(g_1(t) + g_2(0)\cdot g_0(t) + 2\cdot\int_{-\infty}^{t-}(g_2+\overline{g_2})(t-s)\ \mathrm{d}M_s\Big)\ \mathrm{d}M_t,
\end{aligned}$$

where M_t is the martingale defined by

$$\begin{aligned}
\mathrm{d}M_t &= \mathrm{d}N_t - \lambda(t)\ \mathrm{d}t, \\
\tau_0 &= \frac{1}{Th}\int_{\mathbb{R}} L\Big(\frac{t-t_0}{Th}\Big)\Lambda(t/T)\ \mathrm{d}t, \\
g_0(t) &= \Big(1 + \int_{\mathbb{R}}\chi(s-t, s/T)\ \mathrm{d}s\Big), \\
\tau_j &= \frac{1}{Th}\int_{\mathbb{R}} L\Big(\frac{t-t_0}{Th}\Big)\Big((\phi_{j,0}*\Lambda)(t) + g_2(0)\Big)\Lambda(t/T)\ \mathrm{d}t, \\
g_1(t) &= ((\phi_{j,0}+\overline{\phi_{j,0}})*\Lambda)(t) + (\overline{\chi}*((\phi_{j,0}+\overline{\phi_{j,0}})*\Lambda))(t), \\
g_2(t-u) &= \phi_{j,0}(t-u) + ((\phi_{j,0}+\overline{\phi_{j,0}})*\chi)(t-u) + (\overline{(\phi_{j,0}*\chi)}*\chi)(t-u)
\end{aligned}$$

with $\Lambda(t/T) = \mathbb{E}[\lambda(t)]$ and $\phi_{j,0} := \phi_j \cdot \mathbb{1}_{(0,\infty)}$ and $\overline{\phi_{j,0}}(\cdot) = \phi_{j,0}(-\cdot)$ and $\overline{\chi}(t-u, t/T) := \chi(u-t, t/T)$. Note, that g_0, g_1, g_2 are deterministic.

The remaining goal is to control and investigate the these complex terms.
We will make use a version of Rebolledo's Theorem, namely Proposition 4.2.1 in Ramlau-Hansen (1983):

Proposition 4.4.4. *Let H_n be a sequence of predictable processes where $\mathbb{E}[\int_0^1 H_n^2(s)\lambda_n(s)\ \mathrm{d}s] < \infty$ and introduce $\tilde{M}_n(t) := \int_0^t H_n(s)\mathrm{d}M_n(s)$. Then the following proposition is valid.*
Suppose that

(i) $\forall\ \epsilon > 0 : \int_0^1 H_n(s)^2 \mathbb{1}_{|H_n(s)|>\epsilon}\lambda_n(s)\ \mathrm{d}s \to 0$ *in probability,*

(ii) $\int_0^1 H_n^2(s)\lambda_n(s)\ \mathrm{d}s \to 1$ *in probability when* $n \to \infty$.

Then $\tilde{M}_n(1) \to \mathcal{N}(0,1)$ *in distribution, where* $\mathcal{N}(0,1)$ *is the standard normal distribution.*

Define

$$H(t) := \frac{1}{Th} L\left(\frac{t-t_0}{Th}\right)\left(g_1(t) + g_2(0)\cdot g_0(t) + 2\cdot \int_{-\infty}^{t-} (g_2 + \overline{g_2})(t-s)\ \mathrm{d}M_s\right).$$

Note, that our setting corresponds to the setup of Ramlau-Hansen. We choose a sequence of intervals with growing length and $H_n(\cdot) := H(\cdot\ T_n)$, where T_n denotes the upper bound of the intervals. This strech does not affect the analysis used. Correspondingly we proceed with λ and M.
Accordingly, for the rest of the proof it remains to show that

(i) It holds that

$$Th \int_{\mathbb{R}} (H(t))^2 \lambda(t)\ \mathrm{d}t \to \sigma^2$$

in probability.

(a) First, we show that

$$\mathbb{E}\left[Th \int_{\mathbb{R}} (H(t))^2 \lambda(t)\ \mathrm{d}t\right] \to \sigma^2,$$

(b) Furthermore, we will argue that

$$\mathrm{Var}\left(Th \int_{\mathbb{R}} (H(t))^2 \lambda(t)\ \mathrm{d}t\right) \to 0$$

in probabilty.

(ii) It remains to show the Lindeberg conditon. For any $\varepsilon > 0$ it holds that

$$\int_0^T (\sqrt{Th} H(t))^2 \mathbb{1}_{\{t:|\sqrt{Th}H(t)|>\varepsilon\}} \lambda(t)\ \mathrm{d}t \to 0$$

in probability.

We start with the investigation of the term in (a) and define $G_1(t) := g_1(t) + g_2(0)\cdot g_0(t)$ and $G_2(t) := \Lambda(t/T)$.

We need to articulate a lemma to use the supposed alignment of μ_0 and ν_0 to our advantage.

Lemma 4.4.5. *For $u, t \in [t_0 - Th, t_0 + Th]$ and $s \in [0, T]$ it holds under Assumptions 4.2.1 (A1) and (A2) that there exist distinct constants $c > 0$ such that*

(i) $|\nu_0(t/T) - \nu_0(t_0/T)| = c|\frac{t-t_0}{T}|$,
$|\mu_0(t-u, t/T) - \mu_0(t-t_0, t/T)| = c|u - t_0|$,
$|\mu_0(s, t/T) - \mu(s, t_0/T)| = c|\frac{t-t_0}{T}|$.

(ii) $|(\chi * \nu_0)(t/T) - (\chi * \nu_0)(t_0/T)| = c|t - t_0|$.

It follows from (i) and (ii) that $|\Lambda(t/T) - \Lambda(t_0/T)| = c|t - t_0|$ as $\Lambda(t/T) = \nu_0(t/T) + (\chi * \nu_0)(t/T)$. Analogously to (ii) it holds that

(iii) $|(\phi_{j,0} * \Lambda)(t) - (\phi_{j,0} * \Lambda)(t_0)| = c|t - t_0|$,

(iv) $|(\phi_{j,0} * \chi)(t-u) - (\phi_{j,0} * \chi)(t_0 - u)| = c|t - t_0|$,
$|(\phi_{j,0} * \chi)(t-u) - (\phi_{j,0} * \chi)(t - t_0)| = c|u - t_0|$.

Lemma 4.4.5 together with assumption (A4) from Assumptions 4.2.1 gives information about how in the case of alignment of μ and ν at the designated points, the integrands indeed are similar.

The developed assumptions allow for additional bounds of the terms and factors of interest.

Lemma 4.4.6. *It holds that g_0, g_1 and g_2 are (asymptotically) bounded.*

We will see that these are the only requirements for g_0, g_1 and g_2. We come back to the proof of theorem 4.4.1:

We now know, that $|G_2(t) - \Lambda(t_0/T)| = c|t - t_0|$ for a constant $c > 0$ and thus asymptotically bounded in the consideres intervall around t_0. With Lemma 4.4.6 we know that G_1 is bounded.

For the explicit calculations we need the following lemma.

Lemma 4.4.7. *Let $k \in \mathbb{N}$ and $\{s_1, \ldots, s_l\} = \{t_1, \ldots, t_k\} \subset [0, T]$ such that $\#\{s_1, \ldots, s_l\} = l \leq k$. For a function $f : \mathbb{R}^k \to \mathbb{R}$ with*

$$\int |f(s_{\kappa(1)}, \ldots, s_{\kappa(k)})| \, \mathrm{d}s_1 \ldots \mathrm{d}s_l < \infty$$

it holds that

$$\mathbb{E}\left[\int f(t_1, \ldots, t_k) \, \mathrm{d}M_{t_1} \ldots \mathrm{d}M_{t_k}\right] < \infty$$

with M. martingales as in Lemma 4.4.3 and $\kappa : \{0, \dots, k\} \to \{0, \dots l\}$ *such that* $s_{\kappa(j)} = t_j$ *for* $j = 0, \dots, k$.
In particular, for the case $k = 3$ *it holds that*

$$\mathbb{E}[\mathrm{d}M_{t_1}\mathrm{d}M_{t_2}\mathrm{d}M_{t_3}] = \Lambda(t_1)\mathrm{d}t_1 + \chi(t_3 - t_1, t_3/T)\Lambda(t_1)\mathbb{1}_{t_1<t_3}\mathrm{d}t_1\mathrm{d}t_3.$$

We thus calculate as follows. Plugging in the definition of H, taking he square, we obtain

$$\begin{aligned}
&\mathbb{E}\left[Th\int_{\mathbb{R}}(H(t))^2\lambda(t)\ \mathrm{d}t\right]\\
&= \mathbb{E}\left[\int_{\mathbb{R}}\frac{1}{Th}L^2\left(\frac{t-t_0}{Th}\right)\left(G_1(t) + 2\int_{-\infty}^{t-}(g_2+\overline{g_2})(t-s)\ \mathrm{d}M_s\right)^2\right.\\
&\quad \left.\cdot\left(G_2(t) + \int_{\mathbb{R}}\chi(t-v,t/T)\ \mathrm{d}M_v\right)\ \mathrm{d}t\right]\\
&= \mathbb{E}\left[\int_{\mathbb{R}}\frac{1}{Th}L^2\left(\frac{t-t_0}{Th}\right)\left((G_1(t))^2 + 4G_1(t)\int_{-\infty}^{t-}(g_2+\overline{g_2})(t-s)\ \mathrm{d}M_s\right.\right.\\
&\quad \left. + 4\int_{-\infty}^{t-}\int_{-\infty}^{t-}(g_2+\overline{g_2})(t-s)(g_2+\overline{g_2})(t-s')\ \mathrm{d}M_s\mathrm{d}M_{s'}\right)\\
&\quad \left.\cdot\left(G_2(t) + \int_{\mathbb{R}}\chi(t-v,t/T)\ \mathrm{d}M_v\right)\ \mathrm{d}t\right].
\end{aligned}$$

We use the linearity of the mean, to see, that

$$\begin{aligned}
&\mathbb{E}\left[Th\int_{\mathbb{R}}(H(t))^2\lambda(t)\ \mathrm{d}t\right]\\
&= \int_{\mathbb{R}}\frac{1}{Th}L^2\left(\frac{t-t_0}{Th}\right)G_1(t)^2G_2(t)\ \mathrm{d}t\\
&\quad + \int_{\mathbb{R}}\frac{1}{Th}L^2\left(\frac{t-t_0}{Th}\right)4G_1(t)\int_{-\infty}^{t-}\int_{\mathbb{R}}\chi(t-v,t/T)(g_2+\overline{g_2})(t-s)\mathbb{E}[\mathrm{d}M_v\mathrm{d}M_s]\ \mathrm{d}t\\
&\quad + \int_{\mathbb{R}}\frac{1}{Th}L^2\left(\frac{t-t_0}{Th}\right)4G_2(t)\int_{-\infty}^{t-}\int_{-\infty}^{t-}(g_2+\overline{g_2})(t-s)\\
&\quad \cdot(g_2+\overline{g_2})(t-s')\mathbb{E}[\mathrm{d}M_s\mathrm{d}M_{s'}]\ \mathrm{d}t\\
&\quad + \int_{\mathbb{R}}\frac{1}{Th}L^2\left(\frac{t-t_0}{Th}\right)4G_1(t)\int_{-\infty}^{t-}\int_{-\infty}^{t-}\int_{\mathbb{R}}\chi(t-v,t/T)(g_2+\overline{g_2})(t-s)\\
&\quad \cdot(g_2+\overline{g_2})(t-s')\mathbb{E}[\mathrm{d}M_v\mathrm{d}M_s\mathrm{d}M_{s'}]\ \mathrm{d}t.
\end{aligned}$$

This is where we need the calculation made in lemma 4.4.7. We can actually see an explicit form of the expectations.

$$
\begin{aligned}
&\mathbb{E}\Big[Th \int_{\mathbb{R}} (H(t))^2 \lambda(t) \,\mathrm{d}t\Big] \\
&= \mathbb{E}\Big[\int_{\mathbb{R}} \frac{1}{Th} L^2\Big(\frac{t-t_0}{Th}\Big) G_1(t)^2 G_2(t) \,\mathrm{d}t \\
&\quad + \int_{\mathbb{R}} \frac{1}{Th} L^2\Big(\frac{t-t_0}{Th}\Big)\Big(4G_1(t) \int_{-\infty}^{t-} \chi(t-s, t/T)(g_2 + \overline{g_2})(t-s)\mathbb{E}[\lambda(s)] \,\mathrm{d}s \\
&\quad + 4G_2(t) \int_{-\infty}^{t-} (g_2 + \overline{g_2})^2 (t-s)\mathbb{E}[\lambda(s)] \,\mathrm{d}s \\
&\quad + \int_{\mathbb{R}} \frac{1}{Th} L^2\Big(\frac{t-t_0}{Th}\Big)\Big(4 \int_{-\infty}^{t-} \int_{-\infty}^{s-} \chi(t-v, t/T)(g_2 + \overline{g_2})^2 (t-s) \\
&\quad \cdot \chi(s-v, s/T)\mathbb{E}[\lambda(v)] \,\mathrm{d}v\mathrm{d}s + 4 \int_{-\infty}^{t-} \chi(t-s, t/T)(g_2 + \overline{g_2})^2 (t-s) \\
&\quad \cdot \mathbb{E}[\lambda(s)] \,\mathrm{d}s\Big)\mathrm{d}t \\
&=: \sigma^2.
\end{aligned}
$$

We have to further investigate the denoted term.
σ^2 is obviously deterministic. We have to show, that σ^2 is bounded for which we use the Assumption 4.2.1 (A3) and 4.4.6. Analogously to the case of g_2 it holds that $\overline{g_2}$ is bounded.
We can now easily see that we can bound σ as follows by finding bounds for each term individually.

$$
\begin{aligned}
\sigma^2 \leq c \cdot &\int_{\mathbb{R}} \frac{1}{Th} L^2\Big(\frac{t-t_0}{Th}\Big) \,\mathrm{d}t \\
&+ \int_{\mathbb{R}} \frac{1}{Th} L^2\Big(\frac{t-t_0}{Th}\Big) c \int_{-\infty}^{t-} \exp(-c(t-s)) \,\mathrm{d}s + O(1/T) \\
&+ \int_{\mathbb{R}} \frac{1}{Th} L^2\Big(\frac{t-t_0}{Th}\Big) c(F+1) \int_{-\infty}^{t-} \exp(-c(t-s))) \,\mathrm{d}s \\
&+ \int_{\mathbb{R}} \frac{1}{Th} L^2\Big(\frac{t-t_0}{Th}\Big) c \int_{-\infty}^{t-} \int_{-\infty}^{s-} \exp(-c(t-v))\mathrm{d}v \,\mathrm{d}s \\
&+ \int_{\mathbb{R}} \frac{1}{Th} L^2\Big(\frac{t-t_0}{Th}\Big) c \int_{-\infty}^{t-} \exp(-c(t-s))) \,\mathrm{d}s,
\end{aligned}
$$

which is obviously bounded with distinct constants $c > 0$.

We start with the proof of (b). Note that for a bounded function $\tilde{g}_0$ and functions $\tilde{g}_h, \tilde{h}_l : \mathbb{R} \to \mathbb{R}, l \in \mathbb{N}$ with

$$|\tilde{g}_l(u)|, |\tilde{h}_l| \leq c \exp(-c'u)$$

for $u \in \mathbb{R}^+$ and contants $c, c' > 0$ that for

$$\tilde{H}(t) = \frac{1}{Th} L\left(\frac{t-t_0}{Th}\right)\left(\tilde{g}_0(t) + \int_{-\infty}^{t-} \tilde{g}_1(t-s)\ \mathrm{d}M_s\right)$$

it holds that

$$\begin{aligned}
&Th \int_{\mathbb{R}} \tilde{H}(t)^2 \lambda(t)\ \mathrm{d}t \\
&= Th \int_{\mathbb{R}} \tilde{H}(t)^2 \Lambda(t/T)\ \mathrm{d}t + Th \int_{\mathbb{R}} \tilde{H}(t)^2 \int_{\mathbb{R}} \chi(t-s, t/T)\ \mathrm{d}M_s\ \mathrm{d}t,
\end{aligned}$$

where we used a lemma from the last chapter, i.e. Lemma 3.3.1:
It holds that

$$\lambda(t) = \mathbb{E}[\lambda(t)] + \int_{\mathbb{R}} \chi(t-s, t/T)\ \mathrm{d}M_s.$$

We continue the calculation by plugging in the newly develoed notation.

$$\begin{aligned}
&Th \int_{\mathbb{R}} \tilde{H}(t)^2 \lambda(t)\ \mathrm{d}t \\
&= \int_{\mathbb{R}} \frac{1}{Th} L^2\left(\frac{t-t_0}{Th}\right)\Bigg(\tilde{g}_0(t)^2 \Lambda(t/T) \\
&\quad + 2\tilde{g}_0(t) \int_{-\infty}^{t-} \tilde{g}_1(t-s)\ \mathrm{d}M_s \Lambda(t/T) \\
&\quad + \left(\int_{-\infty}^{t-} \tilde{g}_1(t-s)\ \mathrm{d}M_s\right)^2 \Lambda(t/T) \\
&\quad + \tilde{g}_0(t)^2 \int_{-\infty}^{t-} \chi(t-s, t/T)\ \mathrm{d}M_s \\
&\quad + 2\tilde{g}_0 \int_{-\infty}^{t-} \tilde{g}_1(t-s)\ \mathrm{d}M_s \int_{\mathbb{R}} \chi(t-s, t/T)\ \mathrm{d}M_s \\
&\quad + \left(\int_{-\infty}^{t-} \tilde{g}_1(t-s)\ \mathrm{d}M_s\right)^2 \int_{\mathbb{R}} \chi(t-s, t/T)\ \mathrm{d}M_s\Bigg)\ \mathrm{d}t.
\end{aligned}$$

Note, that the summands are of the form

$$\frac{1}{Th} \int_{\mathbb{R}} L^2\left(\frac{t-t_0}{Th}\right) \tilde{g}_0(t) \int_{\mathbb{R}} \tilde{h}_1(t-s_1) \mathrm{d}M_{s_1} \ldots \int_{\mathbb{R}} \tilde{h}_k(t-s_k)\ \mathrm{d}M_{s_k}$$

for $k \in \{1, \ldots, 3\}$. The exact definitions of the $\tilde{h}_l$ are to be recalled from Lemma 4.4.3. They are linear combinations of convolutions and are exponentially bounded functions as we see in Lemma 4.4.8.
We now argue, that the variance of all such terms converge to zero. This is immediately clear for $k = 1$ and the terms without a stochastic share. For $k = 2$ we decompose the term into

$$\frac{1}{Th} \int_{\mathbb{R}} L^2\Big(\frac{t - t_0}{Th}\Big) \tilde{g}_0(t) \int_{s_1 \neq s_2} \tilde{h}_1(t - s_1) \tilde{h}_2(t - s_2) \, \mathrm{d}M_{s_1} \, \mathrm{d}M_{s_2}$$

and

$$\frac{1}{Th} \int_{\mathbb{R}} L^2\Big(\frac{t - t_0}{Th}\Big) \tilde{g}_0(t) \int_{s} \tilde{h}_1(t - s) \tilde{h}_2(t - s) \, (\mathrm{d}M_s)^2.$$

The first term has mean zero and its variance is equal to its second moment which can easily be shown to converge to zero. For the second term we use that

$$(\mathrm{d}M_s)^2 = \mathrm{d}N_s = \Lambda(s/T) \, \mathrm{d}s + \int_{\mathbb{R}} \chi(s - u, s/T) \, \mathrm{d}M_u.$$

Thus, by plugging this into the formula we ge a constant term and a mean zero term for which the second moment can be easily shown to converge to zero. For the case of $k = 3$ we repeat the steps for $k = 2$ and recall the Lemma 4.4.7 giving us a tool to calculate the moments of the martingale integrators.

For the verification of the Lindeberg condition (ii) we use the bound

$$\begin{aligned}
&\int_0^T (\sqrt{Th} H(t))^2 \mathbb{1}_{\sqrt{Th} H(t) > \varepsilon} \lambda(t) \, \mathrm{d}t \\
&\leq \int_0^T (\sqrt{Th} H(t))^4 \frac{1}{\varepsilon^2} \lambda(t) \, \mathrm{d}t \\
&= \frac{1}{(Th)^2} \int_0^T \Big(\tilde{g}_0(t) + \int_{-\infty}^{t-} \tilde{g}_1(t - s) \, \mathrm{d}M_s \Big)^4 L^4\Big(\frac{t - t_0}{Th}\Big) \, \mathrm{d}t \\
&= \frac{1}{(Th)^2} \int_{\mathbb{R}} L^4\Big(\frac{t - t_0}{Th}\Big) \Bigg(\sum_{k=0}^{4} \int_{\mathbb{R}} \tilde{h}_1(t - s_1) \, \mathrm{d}M_{s_1} \ldots \int_{\mathbb{R}} \tilde{h}_k(t - s_k) \, \mathrm{d}M_{s_k} \Bigg).
\end{aligned}$$

It can be easily seen that this expression is of order $O_{\mathbb{P}}\big(\frac{1}{Th}\big) = o_{\mathbb{P}}(1)$. Again it is mainly essential to investigate the behaviour of the martingale integrators for which we developed a tool in Lemma 4.4.7. This shows (ii).

In particular, we saw that $\Sigma_j := \sigma^2$.

We thus have shown the statement for $\hat{\tau}_j$ with $j = 1, \dots, J$. The case of $\hat{\tau}_0$ is obviously a simplification, and thus analogous with

$$\Sigma_0 := \mathbb{E}\left[\int_{\mathbb{R}} \frac{1}{(Th)^2} L^2\left(\frac{t-t_0}{Th}\right) g_0^2(t)\lambda(t)\ \mathrm{d}t \right].$$

□

4.4.2 Additional Proofs

Proof of Lemma 4.4.2: We recall the representation of the bias from Lemma 4.4.3 and calculate for the respective Hawkes processes $i = 1, 2$ while using the result from Lemma 4.4.8. The main argument lies in the fact that for each i it holds that $\Lambda(t/T)^{(1)} = \Lambda(t/T)^{(2)} + O(t - t_0)$ for any $t \in [0, T]$ and similar results, which we prove in Lemma 4.4.5.

$$\mathbb{E}[\hat{\tau}_j^{(1)}] = \tau_j^{(1)} = \tau_j^{(2)} + O(h^2) = \mathbb{E}[\hat{\tau}_j^{(2)}] + O(h^2)$$

for all $j = 0, \dots, J$. □

Proof of Lemma 4.4.3: We recall Lemma 3.3.1, which we will be using vigorously.
It holds that

$$\begin{aligned}\lambda(t) &= \mathbb{E}[\lambda(t)] + \int_{\mathbb{R}} \chi(t-s, t/T)\ \mathrm{d}M_s,\\ \mathbb{E}[\lambda(t)] &= \nu\left(\frac{t}{T}\right) + \int_{\mathbb{R}} \chi\left(t-u, \frac{t}{T}\right)\nu\left(\frac{u}{T}\right)\ \mathrm{d}u\\ &= \nu\left(\frac{t}{T}\right) + (\chi * \nu)\left(\frac{t}{T}\right).\end{aligned}$$

Note, that it holds for $s, t \in \mathbb{R}$ and a constant $c > 0$ that

$$\left| L\left(\frac{t-t_0}{Th}\right) - L\left(\frac{s-t_0}{Th}\right) \right| = c\left|\frac{t-s}{Th}\right| \tag{4.3}$$

and

$$\begin{aligned}\int_{t_0-Th}^{t_0+Th} L\left(\frac{t-t_0}{Th}\right) f(t)\ \mathrm{d}M_t &= \int_{\mathbb{R}} L\left(\frac{t-t_0}{Th}\right) f(t)\ \mathrm{d}M_t,\\ \int_{t_0-Th}^{t_0+Th} L\left(\frac{t-t_0}{Th}\right) f(t)\ \mathrm{d}t &= \int_{\mathbb{R}} L\left(\frac{t-t_0}{Th}\right) f(t)\ \mathrm{d}t\end{aligned}$$

for any deterministic or predictable functions f with appropriate assumptions since the support of L lies in $[-1,1]$.

We start our calculation by remembering simply decomposing N into the martingale share and the conditional intensity.

$$\begin{aligned}\hat{\tau}_0 &= \frac{1}{Th}\int_{t_0-Th}^{t_0+Th} L\left(\frac{t-t_0}{Th}\right)\mathrm{d}N_t\\ &= \frac{1}{Th}\int_{t_0-Th}^{t_0+Th} L\left(\frac{t-t_0}{Th}\right)\mathrm{d}M_t + \frac{1}{Th}\int_{t_0-Th}^{t_0+Th} L\left(\frac{t-t_0}{Th}\right)\lambda(t)\ \mathrm{d}t\end{aligned}$$

With the definition of τ_0 and using the alternative form of λ from Lemma 3.3.1

$$\lambda(t) = \Lambda(t/T) + \int_{\mathbb{R}} \chi(t-s, t/T)\ \mathrm{d}M_s,$$

we can calculate

$$\begin{aligned}\hat{\tau}_0 = \tau_0 &+ \frac{1}{Th}\int_{t_0-Th}^{t_0+Th} L\left(\frac{t-t_0}{Th}\right)\mathrm{d}M_t\\ &+ \frac{1}{Th}\int_{t_0-Th}^{t_0+Th} L\left(\frac{t-t_0}{Th}\right)\int_{\mathbb{R}} \chi(t-s, t/T)\ \mathrm{d}M_s\ \mathrm{d}t\end{aligned}$$

Finally, we rephrase the terms and switch the integrals to obtain the desired result. We will use equation (4.3) in the second step, such that linearity of the integral allows for this simplification in exchange of the $O(1/Th)$ term.

$$\begin{aligned}\hat{\tau}_0 &= \tau_0 + \frac{1}{Th}\int_{t_0-Th}^{t_0+Th} L\left(\frac{t-t_0}{Th}\right)\mathrm{d}M_t\\ &\quad + \frac{1}{Th}\int_{\mathbb{R}}\int_{t_0-Th}^{t_0+Th} L\left(\frac{t-t_0}{Th}\right)\chi(t-s, t/T)\ \mathrm{d}t\ \mathrm{d}M_s\\ &= \tau_0 + \frac{1}{Th}\int_{\mathbb{R}} L\left(\frac{s-t_0}{Th}\right)\left(1 + \int_{t_0-Th}^{t_0+Th} \chi(t-s, t/T)\ \mathrm{d}t\right)\ \mathrm{d}M_s\\ &\quad + O\left(\frac{1}{Th}\right) + T^*\\ &= \tau_0 + \frac{1}{Th}\int_{\mathbb{R}} L\left(\frac{s-t_0}{Th}\right)\left(1 + \int_{\mathbb{R}} \chi(t-s, t/T)\ \mathrm{d}t\right)\ \mathrm{d}M_s + O_{\mathbb{P}}\left(\frac{1}{Th}\right),\end{aligned}$$

where

$$T^* = \frac{1}{Th}\int_{\mathbb{R}}\int_{t_0-Th}^{t_0+Th}\left(L\left(\frac{s-t_0}{Th}\right) - L\left(\frac{t-t_0}{Th}\right)\right)\chi(t-s, t/T)\ \mathrm{d}t\ \mathrm{d}M_s.$$

The last equality is implied by the following treatment of T^*.
We start by splitting the term of interest into two and do further calculations individually.

$$\frac{1}{Th}\int_{\mathbb{R}}\int_{t_0-Th}^{t_0+Th}\left(L\left(\frac{s-t_0}{Th}\right)-L\left(\frac{t-t_0}{Th}\right)\right)\chi(t-s,t/T)\;\mathrm{d}t\;\mathrm{d}M_s$$
$$=\frac{1}{Th}\int_{\mathbb{R}}\int_{t_0-Th}^{t_0+Th}\left(L\left(\frac{s-t_0}{Th}\right)-L\left(\frac{t-t_0}{Th}\right)\right)\mathbb{1}_{\{|t-s|\geq r_T\}}\chi(t-s,t/T)\;\mathrm{d}t\;\mathrm{d}M_s$$
$$+\frac{1}{Th}\int_{\mathbb{R}}\int_{t_0-Th}^{t_0+Th}\left(L\left(\frac{s-t_0}{Th}\right)-L\left(\frac{t-t_0}{Th}\right)\right)\mathbb{1}_{\{|t-s|<r_T\}}\chi(t-s,t/T)\;\mathrm{d}t\;\mathrm{d}M_s$$
$$=: \mathrm{I}+\mathrm{II}.$$

We start with the first summand. With the bound of χ we know, that there exist constants $c^*, c' > 0$ such that

$$\begin{aligned}
|\mathrm{I}| &\leq \frac{c^*}{Th}\int_{\mathbb{R}}\int_0^T \exp(-c'r_T)\left|L\left(\frac{s-t_0}{Th}\right)-L\left(\frac{t-t_0}{Th}\right)\right|\;\mathrm{d}t\;\mathrm{d}M_s\\
&\leq \frac{c^*}{h}\left|\int_{\mathbb{R}}\exp(-c'r_T)\;\mathrm{d}M_s\right|\\
&\leq \frac{c^*}{h}\sqrt{T}\exp(-c'r_T)\leq c^*T^{1/2-c'c^{**}+\delta}
\end{aligned}$$

for the choice of $r_T := c^{**}\log(T)$ and where we use the last of the assumptions (A2) of Assumptions 4.2.1. We see that we can choose $c^{**} > 0$ such that the term vanishes asymptotically.
Going on with the second summand, we see for constants $c^*, c' > 0$ which are again given by the bound of χ, that

$$\begin{aligned}
\mathrm{II} &\leq \frac{c^*}{Th}\int_{\mathbb{R}}\left(\mathbb{1}_{\{|\frac{s-t_0}{Th}|\leq 1\}}+\mathbb{1}_{\{|\frac{t-t_0}{Th}|\leq 1\}}\right)\mathbb{1}_{\{|t-s|<r_T\}}\;\frac{r_T}{Th}\;\mathrm{d}t\;\mathrm{d}M_s\\
&\leq \frac{c^*}{Th}\frac{r_T}{Th}(Th+r_T)\int_{\mathbb{R}}\mathbb{1}_{\{|s-t_0|\leq Th+r_T\}}\;\mathrm{d}M_s\\
&\leq \frac{c^*}{Th}\frac{r_T}{Th}(Th+r_T)\;o_{\mathbb{P}}(\sqrt{Th+r_T}).
\end{aligned}$$

We now come to the corresponding term for $j = 1, \ldots, J$. We again start by using the cited decomposition in both used counting processes in the term. This will leave us with multiple terms we will have to investigate individually based on the appropriate integrators. We will repeatedly use the argument made in the last step to show negligability of certain terms.

$$
\begin{aligned}
\hat{\tau}_j &= \frac{1}{Th}\int_{t_0-Th}^{t_0+Th} L\Big(\frac{t-t_0}{Th}\Big)\Bigg(\int_{-\infty}^{t-} \phi_j(t-u)\ \mathrm{d}N_u\Bigg)\mathrm{d}N_t \\
&= \frac{1}{Th}\int_{t_0-Th}^{t_0+Th} L\Big(\frac{t-t_0}{Th}\Big)\int_{-\infty}^{t-} \phi_j(t-u)\ \mathrm{d}M_u\ \mathrm{d}M_t \\
&\quad + \frac{1}{Th}\int_{t_0-Th}^{t_0+Th} L\Big(\frac{t-t_0}{Th}\Big)\int_{-\infty}^{t-} \phi_j(t-u)\Lambda(u/T)\ \mathrm{d}u\ \mathrm{d}M_t \\
&\quad + \frac{1}{Th}\int_{t_0-Th}^{t_0+Th} L\Big(\frac{t-t_0}{Th}\Big)\int_{-\infty}^{t-} \phi_j(t-u)\int_{\mathbb{R}} \chi(u-s,u/T)\mathrm{d}M_s\ \mathrm{d}u\ \mathrm{d}M_t \\
&\quad + \frac{1}{Th}\int_{t_0-Th}^{t_0+Th} L\Big(\frac{t-t_0}{Th}\Big)\int_{-\infty}^{t-} \phi_j(t-u)\ \mathrm{d}M_u\ \Lambda(t/T)\mathrm{d}t \\
&\quad + \frac{1}{Th}\int_{t_0-Th}^{t_0+Th} L\Big(\frac{t-t_0}{Th}\Big)\int_{-\infty}^{t-} \phi_j(t-u)\ \mathrm{d}M_u\ \int_{\mathbb{R}} \chi(t-v,t/T)\ \mathrm{d}M_v\ \mathrm{d}t \\
&\quad + \frac{1}{Th}\int_{t_0-Th}^{t_0+Th} L\Big(\frac{t-t_0}{Th}\Big)\int_{-\infty}^{t-} \phi_j(t-u)\Lambda(u/T)\ \mathrm{d}u\ \Lambda(t/T)\ \mathrm{d}t \\
&\quad + \frac{1}{Th}\int_{t_0-Th}^{t_0+Th} L\Big(\frac{t-t_0}{Th}\Big)\int_{-\infty}^{t-} \phi_j(t-u)\Lambda(u/T)\ \mathrm{d}u\ \int_{\mathbb{R}} \chi(t-v,t/T)\ \mathrm{d}M_v\ \mathrm{d}t \\
&\quad + \frac{1}{Th}\int_{t_0-Th}^{t_0+Th} L\Big(\frac{t-t_0}{Th}\Big)\int_{-\infty}^{t-} \phi_j(t-u)\int_{\mathbb{R}} \chi(u-s,u/T)\mathrm{d}M_s\ \mathrm{d}u\ \Lambda(t/T)\mathrm{d}t \\
&\quad + \frac{1}{Th}\int_{t_0-Th}^{t_0+Th} L\Big(\frac{t-t_0}{Th}\Big)\int_{-\infty}^{t-} \phi_j(t-u)\int_{\mathbb{R}} \chi(u-s,u/T)\mathrm{d}M_s\ \mathrm{d}u \\
&\qquad \cdot \int_{\mathbb{R}} \chi(t-v,t/T)\ \mathrm{d}M_v\ \mathrm{d}t \\
&=: T_1 + \ldots + T_9.
\end{aligned}
$$

We look at the summands seperately and divided in three cases. Those, which are deterministic at first, then those containing one martingale, and thos which contain many martingales for integrators.
Immediately we find that only the sixth summand is deterministic and calculate

$$
\begin{aligned}
T_6 &= \frac{1}{Th}\int_{t_0-Th}^{t_0+Th} L\Big(\frac{t-t_0}{Th}\Big)\int_{-\infty}^{t-} \phi_j(t-u)\Lambda(u/T)\ \mathrm{d}u\ \Lambda(t/T)\ \mathrm{d}t \\
&= \frac{1}{Th}\int_{\mathbb{R}} L\Big(\frac{t-t_0}{Th}\Big)(\phi_{j,0} * \Lambda)(t)\Lambda(t/T)\ \mathrm{d}t.
\end{aligned}
$$

It was easily seen that we can write the integrands as convolutions to obtain the desired form.

We move on to the next section of terms, namely those with one martingales integrator. Examplary, we begin with the second summand.

$$\begin{aligned} T_2 &= \frac{1}{Th} \int_{t_0-Th}^{t_0+Th} L\bigg(\frac{t-t_0}{Th}\bigg) \int_{-\infty}^{t-} \phi_j(t-u)\Lambda(u/T) \; \mathrm{d}u \; \mathrm{d}M_t \\ &= \frac{1}{Th} \int_{\mathbb{R}} L\bigg(\frac{t-t_0}{Th}\bigg) (\phi_{j,0} * \Lambda)(t) \; \mathrm{d}M_t. \end{aligned}$$

The same calculations were used. We are thus able to obtain a simplified form of the integral. In this way, we aim to write all summand such that we obtain the form stated in Lemma 4.4.3. We commence with the fourth and seventh summand.

$$\begin{aligned} T_4 &= \frac{1}{Th} \int_{t_0-Th}^{t_0+Th} L\bigg(\frac{t-t_0}{Th}\bigg) \int_{-\infty}^{t-} \phi_j(t-u) \; \mathrm{d}M_u \; \Lambda(t/T)\mathrm{d}t \\ &= \frac{1}{Th} \int_{\mathbb{R}} L\bigg(\frac{t-t_0}{Th}\bigg) (\overline{\phi_{j,0}} * \Lambda)(t) \; \mathrm{d}M_t + O_{\mathbb{P}}\bigg(\frac{1}{Th}\bigg), \\ T_7 &= \frac{1}{Th} \int_{t_0-Th}^{t_0+Th} L\bigg(\frac{t-t_0}{Th}\bigg) \int_{-\infty}^{t-} \phi_j(t-u)\Lambda(u/T) \; \mathrm{d}u \; \int_{\mathbb{R}} \chi(t-v, t/T) \; \mathrm{d}M_v \; \mathrm{d}t \\ &= \frac{1}{Th} \int_{\mathbb{R}} L\bigg(\frac{t-t_0}{Th}\bigg) (\overline{\chi} * (\phi_{j,0} * \Lambda))(t) \; \mathrm{d}M_t + O_{\mathbb{P}}\bigg(\frac{1}{Th}\bigg). \end{aligned}$$

In both cases, we had to use Fubini's Theorem to exchange the considered integrals. The term of order O was added to exchange the argument in the kernel function L. To be able to use the notation of the extended definition of the convolution we have to interoduce the following notation.
For functions $f : \mathbb{R} \to \mathbb{R}$ and funcions $g : \mathbb{R} \times [0, T] \to \mathbb{R}$ we write

$$\begin{aligned} \overline{f} &: u \mapsto f(-u), \\ \overline{g} &: (u, t) \mapsto g(-u, t). \end{aligned}$$

This way, it is possible to write the integrands such that we can use a convolution to simplify the terms. We come to he last term of that group and see that with our tools for calculation

$$\begin{aligned} T_8 &= \frac{1}{Th} \int_{t_0-Th}^{t_0+Th} L\bigg(\frac{t-t_0}{Th}\bigg) \int_{-\infty}^{t-} \phi_j(t-u) \int_{\mathbb{R}} \chi(u-s, u/T)\mathrm{d}M_s \; \mathrm{d}u \; \Lambda(t/T)\mathrm{d}t \\ &= \frac{1}{Th} \int_{\mathbb{R}} L\bigg(\frac{t-t_0}{Th}\bigg) (\overline{\chi} * (\overline{\phi_{j,0}} * \Lambda))(t) \; \mathrm{d}M_t + O_{\mathbb{P}}\bigg(\frac{1}{Th}\bigg). \end{aligned}$$

We now take a look at the last four summands to investigate. Again, one of them is easiest.

$$\begin{aligned} T_1 &= \frac{1}{Th} \int_{t_0 - Th}^{t_0 + Th} L\left(\frac{t - t_0}{Th}\right) \int_{-\infty}^{t-} \phi_j(t-u) \,\mathrm{d}M_u \,\mathrm{d}M_t \\ &= \frac{1}{Th} \int L\left(\frac{t - t_0}{Th}\right) \int_{\mathbb{R}} \phi_{j,0}(t-u) \,\mathrm{d}M_u \,\mathrm{d}M_t. \end{aligned}$$

It was in that case not necessary to use Fubini's Theorem, as the order of the integrals was already in the desired position. We go on to the next summand.

$$\begin{aligned} T_3 &= \frac{1}{Th} \int_{t_0 - Th}^{t_0 + Th} L\left(\frac{t - t_0}{Th}\right) \int_{-\infty}^{t-} \phi_j(t-u) \int_{\mathbb{R}} \chi(u-s, u/T) \mathrm{d}M_s \,\mathrm{d}u \,\mathrm{d}M_t \\ &= \frac{1}{Th} \int_{\mathbb{R}} L\left(\frac{t - t_0}{Th}\right) \int_{\mathbb{R}} (\phi_{j,0} * \chi)(t-u) \,\mathrm{d}M_u \,\mathrm{d}M_t. \end{aligned}$$

Since we again see that the correct integral is already sorted in the desired postion, we only need elemental calculations and the extention of the integral.

$$\begin{aligned} T_5 &= \frac{1}{Th} \int_{t_0 - Th}^{t_0 + Th} L\left(\frac{t - t_0}{Th}\right) \int_{-\infty}^{t-} \phi_j(t-u) \,\mathrm{d}M_u \int_{\mathbb{R}} \chi(t-v, t/T) \,\mathrm{d}M_v \,\mathrm{d}t \\ &= \frac{1}{Th} \int_{\mathbb{R}} L\left(\frac{t - t_0}{Th}\right) \int_{\mathbb{R}} (\overline{\phi_{j,0}} * \chi)(t-u) \,\mathrm{d}M_u \,\mathrm{d}M_t + O_{\mathbb{P}}\left(\frac{1}{Th}\right). \end{aligned}$$

It is immediately clear, that we have to change to argument in the kernel, as well as using Fubini's Theorem, explaining the order of the lasting term. The last term now is easy to intevestigate and we calculate

$$\begin{aligned} T_9 &= \frac{1}{Th} \int_{t_0 - Th}^{t_0 + Th} L\left(\frac{t - t_0}{Th}\right) \int_{-\infty}^{t-} \phi_j(t-u) \int_{\mathbb{R}} \chi(u-s, u/T) \mathrm{d}M_s \,\mathrm{d}u \\ &\quad \cdot \int_{\mathbb{R}} \chi(t-v, t/T) \,\mathrm{d}M_v \,\mathrm{d}t \\ &= \frac{1}{Th} \int_{\mathbb{R}} L\left(\frac{t - t_0}{Th}\right) (\overline{(\overline{\phi_{j,0}} * \chi)} * \chi)(t-u) \,\mathrm{d}M_u \,\mathrm{d}M_t + O_{\mathbb{P}}\left(\frac{1}{Th}\right), \\ &= \frac{1}{Th} \int_{\mathbb{R}} L\left(\frac{t - t_0}{Th}\right) (\overline{\phi_{j,0} * \chi} * \chi)(t-u) \,\mathrm{d}M_u \,\mathrm{d}M_t + O_{\mathbb{P}}\left(\frac{1}{Th}\right). \end{aligned}$$

All in all, we can now use linearity to write the integrals as one term and define the appropriate integrands. Kepp in mind, that we repeatedly used the argument from the investigation of $\hat{\tau}_0$ on how to show that residual terms vanish asymptotically.

$$\begin{aligned}\hat{\tau}_j &= \frac{1}{Th}\int_{\mathbb{R}} L\left(\frac{t-t_0}{Th}\right)(\phi_{j,0} * \Lambda)(t)(t/T)\ \mathrm{d}t \\ &+ \frac{1}{Th}\int_{\mathbb{R}} L\left(\frac{t-t_0}{Th}\right) g_1(t)\ \mathrm{d}M_t \\ &+ \frac{1}{Th}\int_{\mathbb{R}} L\left(\frac{t-t_0}{Th}\right)\int_{\mathbb{R}} g_2(t-s)\ \mathrm{d}M_s\ \mathrm{d}M_t + O_{\mathbb{P}}\left(\frac{1}{Th}\right).\end{aligned}$$

Being familiar with the assumptions of limit theorems comparable to Rebolledo's Theorem we have to formulate the terms such that the integrands are deterministic or predictable. In practice, this means, that we have to rewrite the stochastic integrals in the integrands. We see that it holds that

$$\begin{aligned}&\frac{1}{Th}\int_{\mathbb{R}} L\left(\frac{t-t_0}{Th}\right)\int_{\mathbb{R}} g_2(t-s)\ \mathrm{d}M_s\ \mathrm{d}M_t \\ &= \frac{1}{Th}\int_{\mathbb{R}} L\left(\frac{t-t_0}{Th}\right)\int_{s<t} g_2(t-s)\ \mathrm{d}M_s\ \mathrm{d}M_t \\ &\quad + \frac{1}{Th}\int_{\mathbb{R}} L\left(\frac{t-t_0}{Th}\right)\int_{s=t} g_2(t-s)\ \mathrm{d}M_s\ \mathrm{d}M_t \\ &\quad + \frac{1}{Th}\int_{\mathbb{R}} L\left(\frac{t-t_0}{Th}\right)\int_{s>t} g_2(t-s)\ \mathrm{d}M_s\ \mathrm{d}M_t.\end{aligned}$$

The split of the integrals now allows us to write

$$\begin{aligned}&\frac{1}{Th}\int_{\mathbb{R}} L\left(\frac{t-t_0}{Th}\right)\int_{\mathbb{R}} g_2(t-s)\ \mathrm{d}M_s\ \mathrm{d}M_t \\ &= \frac{2}{Th}\int_{\mathbb{R}} L\left(\frac{t-t_0}{Th}\right)\int_{-\infty}^{t-} (g_2+\overline{g_2})(t-s)\ \mathrm{d}M_s\ \mathrm{d}M_t \\ &\quad + \frac{1}{Th}\int_{\mathbb{R}} L\left(\frac{t-t_0}{Th}\right)\ g_2(0)\ (\mathrm{d}M_t)^2.\end{aligned}$$

As $(\mathrm{d}M_t)^2 = \mathrm{d}N_t = \mathrm{d}M_t + \lambda(t)\ \mathrm{d}t$ we furthermore see, that it holds that

$$\begin{aligned}&\frac{1}{Th}\int_{\mathbb{R}} L\left(\frac{t-t_0}{Th}\right)\ g_2(0)\ (\mathrm{d}M_t)^2 \\ &= \frac{1}{Th}\int_{\mathbb{R}} L\left(\frac{t-t_0}{Th}\right)\ g_2(0)\ \mathrm{d}M_t + \frac{1}{Th}\int_{\mathbb{R}} L\left(\frac{t-t_0}{Th}\right)\ g_2(0)\Lambda(t/T)\ \mathrm{d}t \\ &\quad + \frac{1}{Th}\int_{\mathbb{R}} L\left(\frac{t-t_0}{Th}\right)\ g_2(0)\int_{\mathbb{R}} \chi(t-s, t/T)\ \mathrm{d}M_s\ \mathrm{d}t\end{aligned}$$

Now, using additonal splitting tools from Lemma 3.3.1, we calculate

$$\begin{aligned}
&\frac{1}{Th}\int_{\mathbb{R}} L\left(\frac{t-t_0}{Th}\right)\, g_2(0)\, (\mathrm{d}M_t)^2 \\
&= \frac{1}{Th}\int_{\mathbb{R}} L\left(\frac{t-t_0}{Th}\right)\, g_2(0)\Lambda(t/T)\, \mathrm{d}t + O_{\mathbb{P}}\left(\frac{1}{Th}\right) \\
&\quad + \frac{1}{Th}\int_{\mathbb{R}} L\left(\frac{t-t_0}{Th}\right)\, g_2(0)\left(1+\int_{\mathbb{R}} \chi(s-t, s/T)\, \mathrm{d}s\right)\, \mathrm{d}M_t \\
&= \frac{1}{Th}\int_{\mathbb{R}} L\left(\frac{t-t_0}{Th}\right)\, g_2(0)\Lambda(t/T)\, \mathrm{d}t \\
&\quad + \frac{1}{Th}\int_{\mathbb{R}} L\left(\frac{t-t_0}{Th}\right)\, g_2(0)\cdot g_0(t)\, \mathrm{d}M_t + O_{\mathbb{P}}\left(\frac{1}{Th}\right),
\end{aligned}$$

where we already denoted g_2 and g_0 appropriately.
And thus, we find that

$$\begin{aligned}
\hat{\tau}_j &= \frac{1}{Th}\int_{\mathbb{R}} L\left(\frac{t-t_0}{Th}\right)\left(g_1(t)+g_2(0)\cdot g_0(t)+2\cdot\int_{-\infty}^{t-}(g_2+\overline{g_2})(t-s)\, \mathrm{d}M_s\right)\, \mathrm{d}M_t \\
&\tau_j + O_{\mathbb{P}}\left(\frac{1}{Th}\right),
\end{aligned}$$

which gives us the statement of the lemma. □

Proof of Lemma 4.4.5: (i) As ν are smooth enough functions by the Assumptions 4.2.1, such that we know

$$\begin{aligned}
\nu(t/T) &= \nu(t_0/T) + \partial\nu(t_0/T)\left(\frac{t-t_0}{T}\right) + \frac{1}{2}\partial^2\nu(\xi)\left(\frac{t-t_0}{T}\right)^2 \\
&= \nu(t_0/T) + \partial\nu(t_0/T)\left(\frac{t-t_0}{T}\right) + \frac{1}{2}(\partial^2\nu(t_0/T)+2C)\left(\frac{t-t_0}{T}\right)^2
\end{aligned}$$

for ξ between t/T and t_0/T and with

$$\sup_{\xi\in[t_0-t,t_0+t]} |(\nu(\xi/T)-\nu(t_0/T))| \leq 2 \sup_{\xi\in[t_0-t,t_0+t]} |\nu(\xi/T)|,$$

where C is the bound of ν from the Assumptions.
Analogously it follows that there exist constants $c_1, c_2 > 0$ such that

$$\begin{aligned}
|\mu(t-u, t/T) - \mu(t-t_0, t/T)| &= c_1|u-t_0|, \\
|\mu(s, t/T) - \mu(s, t_0/T)| &= c_2\left|\frac{t-t_0}{T}\right|.
\end{aligned}$$

That same approximation obviously holds for any smooth function, such as the basis elements ϕ_j.

(ii) Let us again calculate with the knowledge that ν_0 is bounded

$$\begin{aligned}&(\chi * \nu_0)(t/T)\\&= \int_{\mathbb{R}} \chi(t-s, t/T)\nu_0(s/T)\,\mathrm{d}s\\&= (\chi * \nu_0)(t_0/T) + O(t-t_0)\end{aligned}$$

□

Corollary 4.4.8. *For $u, t \in [t_0 - Th, t_0 + Th]$ it holds under Assumptions 4.2.1 (A1) and (A2) that*

$$\begin{aligned}g_0^{(1)}(t) &= g_0^{(1)}(t_0) + O(t-t_0) = g_0^{(2)}(t) + O(t-t_0),\\g_1^{(1)}(t) &= g_1^{(1)}(t_0) + O(t-t_0) = g_1^{(2)}(t) + O(t-t_0),\\g_2^{(1)}(t-u) &= g_2^{(1)}(0) + O(t-t_0) + O(u-t_0)\\&= g_2^{(2)}(t-u) + O(t-t_0) + O(u-t_0),\\\tau_0^{(1)} &= \tau_0^{(2)} + O(h^2),\\\tau_j^{(1)} &= \tau_j^{(2)} + O(h^2)\end{aligned}$$

for $j = 1, \ldots, J$.

Proof of Corollary 4.4.8: We assumed that $\mu_0^{(1)}(t_0) = \mu_0^{(2)}(t_0)$ as well as $\nu_0^{(1)}(t_0) = \nu_0^{(2)}(t_0)$. With assumption (A4) from 4.2.1 and Lemma 4.4.5 we saw that

$$\begin{aligned}\chi^{(1)}(s, t/T) &= \chi^{(1)}(s, t_0/T) + O\left(\frac{t-t_0}{T}\right)\\&= \chi^{(2)}(s, t_0/T) + O(t-t_0)\\&= \chi^{(2)}(s, t/T) + O(t-t_0)\end{aligned}$$

and analogously that

$$(\chi^{(1)} * \nu_0^{(1)})(t/T) = (\chi^{(2)} * \nu_0^{(2)})(t/T) + O(t-t_0).$$

The remaining cases for the other functions and convolution of functions of Lemma 4.4.5 following analogously. □

Proof of Lemma 4.4.6: Recall, that

$$g_2(t-u) = \phi_{j,0}(t-u) + ((\phi_{j,0} + \overline{\phi_{j,0}}) * \chi)(t-u) + (\overline{(\phi_{j,0} * \chi)} * \chi)(t-u).$$

Note, that $0 < \sup_{x\in[0,A]} \phi_j(x) := F < \infty$. Thus, we look at the summands of g_2 separately. For $t-u \geq 0$ we calculate with elementary knowledge of integration

$$\begin{aligned}
(\phi_{j,0} * \chi)(t-u) &= \int_{\mathbb{R}} \phi_{j,0}(t-s)\chi(s-u, s/T)\ \mathrm{d}s \\
&\leq F \cdot \int_u^{t-} c' e^{-c''(s-u)}\ \mathrm{d}s \\
&= Fc'e^{c''u}(-1)\frac{1}{c''}\left(e^{-c''t} - e^{-c''u}\right) \\
&= F(-1)\frac{c'}{c''}\left(e^{-c''(t-u)} - 1\right).
\end{aligned}$$

Analogously we calculate

$$\begin{aligned}
(\overline{\phi_{j,0}} * \chi)(t-u) &\leq F \int_{t+}^{\infty} c' e^{-c''(s-u)}\ \mathrm{d}s \\
&= Fc'\frac{1}{c''}\left(-e^{-c''(t-u)}\right).
\end{aligned}$$

It remains to take a look at the last summand. Note, that $\mathrm{supp}(\phi_{j,0} * \chi) \subset (0,\infty)$ and $\mathrm{supp}(\chi) \subset [0,\infty)$.

$$\begin{aligned}
(\overline{(\phi_{j,0} * \chi)} * \chi)(t-u) &= \int_{\mathbb{R}} (\phi_{j,0} * \chi)(s-t)\chi(s-u, s/T)\ \mathrm{d}s \\
&\leq \int_{\max(t,u)}^{\infty} c_1''' e^{-c''(s-t)} c' e^{-c''(s-u)}\ \mathrm{d}s \\
&= c_1''' c' e^{t+u} \int_{\max(t,u)}^{\infty} e^{-c''s}\ \mathrm{d}s \\
&= \frac{c_1''' c'}{c''} e^{t+u-c''\max(t,u)} \\
&\leq \frac{c_1''' c'}{c''} e^{-c'''(t-u)}
\end{aligned}$$

as it holds that

$$\begin{aligned}
&\max(t,u) \geq t \\
&\Rightarrow -c''\max(t,u) \leq -c''t \\
&\Rightarrow t+u-c''\max(t,u) \leq (1-c')t + u \leq -((c'')t - u) \leq -c''''(t-u)
\end{aligned}$$

with $c'''' := \max(1, c'' - 1)$.
Thus, the bound for g_2 is proven.

For the rest of the lemma we look at g_0 first. It holds for any $t \in \mathbb{R}$ that

$$\begin{aligned} 0 \leq g_0(t) &= 1 + \int_{\mathbb{R}} \chi(s-t, s/T) \, \mathrm{d}s \\ &\leq 1 + c' \exp(c''t) \int_t^{\infty} \exp(-c''s) \, \mathrm{d}s \\ &= 1 + \frac{c'}{c''}. \end{aligned}$$

For the bound of g_1 we look at both summands individually. It holds that

$$\int_{\mathbb{R}} (\phi_{j,0} + \overline{\phi_{j,0}})(t-s)\mathbb{E}[\lambda(s)] \, \mathrm{d}s = \Lambda(t_0)2M + O\left(\frac{1}{T}\right),$$

where M is the positive constant with $\int_{\mathbb{R}} \phi_j(x) \, \mathrm{d}x \leq M$ and

$$\begin{aligned} (\overline{\chi} * (c + O(1/T)))(t) &= (c + O(1/T)) \int_{\mathbb{R}} \chi(s-t, t/T) \, \mathrm{d}s \\ &\leq (c + O(1/T))c' \exp(c''t) \int_t^{\infty} \exp(-c''s) \, \mathrm{d}s \\ &= (c + O(1/T)) \frac{c'}{(c'')^2} \exp(-c''t). \end{aligned}$$

□

Proof of Lemma 4.4.7: W.l.o.g., assume that $t_1 \leq \ldots \leq t_k$. Note, that for $t_{k-1} \neq t_k$ it holds that

$$\mathbb{E}[\mathrm{d}M_{t_1} \ldots \mathrm{d}M_{t_k}] = 0.$$

Also $(\mathrm{d}M_t)^k = \mathrm{d}N_t = \mathrm{d}M_t + \lambda(t) \, \mathrm{d}t$ for all $k \geq 2$ and $t \in [0, T]$. We recall that with Lemma 3.3.1 under Assumption 4.2.1 (A1) it holds that

$$\lambda(t) = \Lambda(t) + \int_{\mathbb{R}} \chi(t-s, t/T) \, \mathrm{d}M_s.$$

Therefore,

$$\begin{aligned} &\mathbb{E}\left[\int f(t_1, \ldots, t_k) \, \mathrm{d}M_{t_1} \ldots \mathrm{d}M_{t_k}\right] \\ &= \int f(t_1, \ldots, t_k) \, \mathbb{E}\left[\mathrm{d}M_{t_1} \ldots \mathrm{d}M_{t_k}\right] \end{aligned}$$

Now, choose the maximal $N(l) \in \mathbb{N}$, such that

$$\mathbb{E}\left[\mathrm{d}M_{t_1} \ldots \mathrm{d}M_{t_k}\right] = \mathbb{E}\left[\mathrm{d}M_{s_1} \ldots \mathrm{d}M_{t_{k-N(l)}} (\mathrm{d}M_{s_l})^{N(l)}\right].$$

In particular, $N(l) \geq 2$. Also note, that $t_{k-N(l)} = s_{l-1} \neq s_l$.
We proof the assertion by induction.
For a fixed $k \geq 3$ and the corresponding minimal l such that $\{t_1, \ldots, t_k\} = \{s_1, \ldots, s_l\}$ it holds that

$$\begin{aligned}
&\mathbb{E}\left[\int f(t_1, \ldots, t_k)\ \mathrm{d}M_{t_1} \ldots \mathrm{d}M_{t_k}\right] \\
&= \int f(t_1, \ldots t_{k-N(l)}, s_l, \ldots, s_l)\ \mathbb{E}\left[\mathrm{d}M_{t_1} \ldots \mathrm{d}M_{t-N(l)} (\mathrm{d}M_{s_l})^{N(l)}\right] \\
&= \int f(t_1, \ldots t_{k-N(l)}, s_l, \ldots, s_l)\ \mathbb{E}\left[\mathrm{d}M_{t_1} \ldots \mathrm{d}M_{t_{k-N(l)}} \mathrm{d}M_{s_l}\right] \\
&\quad + \int f(t_1, \ldots t_{k-N(l)}, s_l, \ldots, s_l)\ \mathbb{E}\left[\mathrm{d}M_{t_1} \ldots \mathrm{d}M_{t_{k-N(l)}} \lambda(s_l)\right] \mathrm{d}s_l,
\end{aligned}$$

where we used the well-known decomposition of a counting process into a martingale share and the conditional intensity. We go on by using the actual induction step and finally, again, the despiction of Λ from Lemma 3.3.1.

$$\begin{aligned}
&\mathbb{E}\left[\int f(t_1, \ldots, t_k)\ \mathrm{d}M_{t_1} \ldots \mathrm{d}M_{t_k}\right] \\
&= \int f(t_1, \ldots t_{k-N(l)}, s_l, \ldots, s_l)\ \mathbb{E}\left[\mathrm{d}M_{t_1} \ldots \mathrm{d}M_{t_{k-N(l)}} \mathrm{d}M_{s_l}\right] \\
&\quad + \int f(t_1, \ldots t_{k-N(l)}, s_l, \ldots, s_l) \Lambda(s_l)\ \mathbb{E}\left[\mathrm{d}M_{t_1} \ldots \mathrm{d}M_{t_{k-N(l)}}\right] \mathrm{d}s_l \\
&\quad + \int f(t_1, \ldots t_{k-N(l)}, s_l, \ldots, s_l) \int_{\mathbb{R}} \chi(s_l - u, s_l/T)\ \mathbb{E}\left[\mathrm{d}M_{t_1} \ldots \mathrm{d}M_{t_{k-N(l)}} \mathrm{d}M_u\right] \mathrm{d}s_l.
\end{aligned}$$

We note, that χ is integrable via assumption (A3) and that

$$\begin{aligned}
\#\{t_1, \ldots, t_{k-N(l)}, s_l\} &\leq k-1, \\
\#\{t_1, \ldots, t_{k-N(l)}\} &\leq k-2, \\
\#\{t_1, \ldots, t_{k-N(l)}, u\} &\leq k-1.
\end{aligned}$$

Thus the proof is complete, if we can show the assertion for $k = 1$ and $k = 2$.
In the case, that $k = 1$ the mean is zero.
For $k = 2$ we note that

$$\mathbb{E}[\mathrm{d}M_{t_1} \mathrm{d}M_{t_2}] = \mathbb{E}[(\mathrm{d}M_{t_1})^2] = \mathbb{E}[\mathrm{d}N_{t_1}] = \Lambda(t_1)\ \mathrm{d}t_1.$$

Easy calculations hence yield that in the case $k = 3$

$$\mathbb{E}[\mathrm{d}M_{t_1}\mathrm{d}M_{t_2}\mathrm{d}M_{t_3}] = \Lambda(t_1)\mathrm{d}t_1 + \chi(t_3 - t_1, t_3/T)\Lambda(t_1)\mathbb{1}_{t_1<t_3}\mathrm{d}t_1\mathrm{d}t_3.$$

□

4.5 Appendix

Differently than in Chapter 3, we do not require μ to have bounded support. On the other hand our newly formulated assumptions are fulfilled by a reproduction functions with compact support, as we will see now.

Lemma 4.5.1. *If μ has compact support, i.e.*

$$\mathrm{supp}(\mu) \subset [0, A] \times [0, 1]$$

for a constant $A > 0$ Assumption 4.2.1 (A3) is fulfilled, i.e. for any $s \in \mathbb{R}_0^+$ and $x \in [0, 1]$ there exist constants $c, c' > 0$ such that

$$\chi(s, x) \leq c \exp(-c's).$$

Proof of Lemma 4.5.1: Recall that $\mathrm{supp}(\mu_0) \subset [0, A] \times [0, 1]$ as well as $\mu_0(s-u, t/T) \leq C$ for any $s, u \in \mathbb{R}, t \in [0, T]$ and a constant $C > 0$. There exists a $\rho > 0$ with

$$\int_{\mathbb{R}} \mu_0(s - u, t/T) \leq \rho < 1$$

for all $(s, u, t) \in \mathbb{R} \times \mathbb{R} \times [0, T]$.
Note, that

$$\mathrm{supp}(\mu_0^{(*k)})) \subset [0, kA] \times [0, 1]$$

for $k \in \mathbb{N}$ and

$$\mu_0^{(*k)}(s - u, t/T) \leq C\rho^{k-1}$$

for all $k \in \mathbb{N}$, as per induction:

$$\begin{aligned}
\mu_0^{(*1)}(s - u, t/T) &= \mu_0(s - u, t/T) \leq C = C\rho^{1-1}, \\
\mu_0^{(*(k+1))}(s - u, t/T) &= \int_{\mathbb{R}} \mu_0^{(*k)}(s - v, t/T)\mu_0(v - u, v/T)\ \mathrm{d}v \\
&\overset{i.s.}{\leq} C\rho^{k-1} \int_{\mathbb{R}} \mu_0(v - u, v/T)\ \mathrm{d}v \\
&\leq \rho^{k-1}\rho = C\rho^k.
\end{aligned}$$

Now, back to the proof. Consider $l \in \mathbb{Z}$ with $s-u \in (lA, (l+1)A]$, i.e. $l < \frac{s-u}{A}$ and $\mu_0^{(*l)}(s-u, t/T) \equiv 0$ for all $l \leq 0$. Then $\mu_0^{(*k)}(s-u, t/T) = 0$ for all $k \leq l$ and

$$\begin{aligned}
\chi(s-u, t/T) &= \sum_{k=1}^{\infty} \mu_0^{(*k)}(s-u, t/T) = \sum_{k=l+1}^{\infty} \mu_0^{(*k)}(s-u, t/T) \\
&\leq \sum_{k=l+1}^{\infty} C\rho^{k-1} = C\rho^l \sum_{k=0}^{\infty} \rho^k = C\rho^l \frac{1}{1-\rho} \\
&\leq \frac{C}{1-\rho} \cdot \rho^{\frac{s-u}{A}} = \frac{C}{1-\rho} \exp\left(\frac{\log(\rho)}{A}(s-u)\right).
\end{aligned}$$

We thus define

$$\begin{aligned}
c' &:= \frac{C}{1-\rho} > 0, \\
c'' &:= \log(1/\rho) \cdot \frac{1}{A} > 0
\end{aligned}$$

and finish by seeing, that $\text{supp}(\chi) \subset [0, \infty)$. □

Lemma 4.5.2. *Let μ have compact support, i.e.*

$$\text{supp}(\mu) \subset [0, A] \times [0, 1]$$

for a constant $A > 0$. If additionally there exits an absolutely bounded function $g : [0, A] \to \mathbb{R}^+$ such that for all $s \in [0, A]$ and $|x - x_0| \leq h$ it holds that

$$|\mu_0(s, x) - \mu_0(s, x_0)| \leq |x - x_0| g(s).$$

then the Assumption 4.2.1 (A4) is fulfilled.

Proof of Lemma 4.5.2: Note that

$$\begin{aligned}
&\left| \int_{\mathbb{R}} \mu_0^{(*(k-1))}(t_0 - r, t_0/T) \mu_0(r - u, t_0/T) \, \mathrm{d}r \right| \\
&\leq \sup_{r \in \mathbb{R}} |\mu_0^{(*(k-1))}(t_0 - r, t_0/T)| \cdot \gamma \leq \gamma^k
\end{aligned}$$

with

$$\gamma := \sup_{t \in [0,T]} \int_{\mathbb{R}} \mu_0(s, t/T) \, \mathrm{d}s < 1.$$

We now obtain that

$$\begin{aligned}
&{\mu_0}^{(*k)}(t_0-u,t_0/T)\\
&=\int_{\mathbb{R}}\mu_0^{(*(k-1))}(t_0-r,t_0/T)\mu_0(r-u,r/T)\ \mathrm{d}r\\
&=\int_{\mathbb{R}}\mu_0^{(*(k-1))}(t_0-r,t_0/T)\mu_0(r-u,t_0/T)\ \mathrm{d}r\\
&\quad+\int_{\mathbb{R}}\mu_0^{(*(k-1))}(t_0-r,t_0/T)\Big(\mu_0(r-u,r/T)-\mu_0(r-u,t_0/T)\Big)\ \mathrm{d}r.
\end{aligned}$$

For the second summand now holds with the inequality above that

$$\begin{aligned}
&\left|\int_{\mathbb{R}}\mu_0^{(*(k-1))}(t_0-r,t_0/T)\Big(\mu_0(r-u,r/T)-\mu_0(r-u,t_0/T)\Big)\ \mathrm{d}r\right|\\
&\le\int_{\mathbb{R}}\mu_0^{(*(k-1))}(t_0-r,t_0/T)\ \Big|\mu_0(r-u,r/T)-\mu_0(r-u,t_0/T)\Big|\ \mathrm{d}r\\
&\le\int_{\mathbb{R}}\mu_0^{(*(k-1))}(t_0-r,t_0/T)\ \left|\frac{r-t_0}{T}\right||g(r-u)|\ \mathrm{d}r\\
&\le\gamma^{(k-1)}\frac{(k-1)A}{T}C
\end{aligned}$$

as $\mathrm{supp}\mu_0^{(*k)}\subseteq[0,kA]$. Since $\sum_{k=1}^{\infty}\gamma^k k<\infty$ we can write

$$\chi(t_0-u,t_0/T)=\chi_{cl.}(t_0-u,t_0/T)+c\frac{1}{T}$$

for

$$\begin{aligned}
\chi_{cl.}(t-s,t/T)&=\sum_{i=1}^{\infty}\mu_0^{(*_{cl.}i)}(t-s,t/T),\\
\mu_0^{(*_{cl.}i)}(t-s,t/T)&=\int\mu_0^{(*(i-1))}(t-u,t/T)\mu_0(u-s,t/T)\ \mathrm{d}u\\
\mu_0^{(*_{cl.}1)}&=\mu_0.
\end{aligned}$$

We can easily use the well-known properties of $*_{cl.}$ such that we can use a Taylor expansion argument also in the second argument of χ. Therefore, first note, that

$$\partial_2\mu_0^{(*k)}(t_s-s,t_0/T)=\int_{\mathbb{R}}\partial_2\mu_0^{(*(k-1))}(t_0-p,t_0/T)\mu_0(p-s,p/T)\ \mathrm{d}p.$$

From now on $\partial_2\chi$ denotes the derivative with respect to the second argument of $\tilde{\chi_T}:=\chi(t-s,t/T)$. Together with the assumption that $\partial_2\mu_0^{(*1)}=\partial_2\mu_0$

exists, we learn that the derivatives of $\tilde{\chi_T}$ exist and are (uniformly absolutely) bounded. We finally see that for distinct constants c

$$\begin{aligned}\chi(s,t/T) &= \chi_{cl.}(s,t/T) + c\cdot\frac{t-t_0}{T}\\ &= \chi_{cl.}(s,t_0/T) + c\cdot\frac{t-t_0}{T}\\ &= \chi(s,t_0/T) + c\cdot\frac{t-t_0}{T}\end{aligned}$$

This gives us the result with the substitution of $s \mapsto u+t$ The result obviously holds analogously for $i=2$. Note, that the Taylor expansion is immediately applicable in the first argument, such that for a constant $c>0$

$$|\chi(t-u,t/T) - \chi(t-t_0,t/T)| = c|u-t_0|.$$

In particular, we see that with the alternative notation of the convolution for $\mu_0^{(1)}(s,t_0/T) = \mu_0^{(2)}(s,t_0/T)$ it is easy to see that for constants $c_1, c_2 > 0$

$$|\chi_{cl.}^{(1)}(s,t_0/T) - \chi_{cl.}^{(2)}(s,t_0/T)| = \frac{c_1}{T}$$

and thus

$$|\chi^{(1)}(t-u,t/T) - \chi^{(2)}(t-u,t/T)| = c_2\left|\frac{t-t_0}{T}\right|.$$

□

5 | Conclusion and Outlook

After developing an estimator and the presentation of a possible testing procedure there are still a lot of missing possiblities for interesting research to be filled with additional conclusions.
In the first part of this work of research there were many results obtained, that were restricted to reproduction functions, that have compact support. The expansion of the estimator to a non-compact setting seem possible and desirable. For many settings and examples, the compact support were at the very least sufficient if not reasonable. The common argument would be seen e.g. in the example of reproducing species of insects. Insects or any mammel grow and are not always fertile or at some point die themselves. Comparably, if a patient is contagious with a disease, it is often the case that the incubation time is limited to the (not so distant) future.
There is a vast amount of reasearch done in the past and other models for Hawkes processes with the assertion have been considered. For example, that the reproduction function does not have to be of compact support. For lack of better knowledge, I have not seen work on that case in a nonparametric and especially the locally stationary case, which is obviously favourable in the application. Arguments for that have been discussed in the introduction, see seasonarity.
In the second part of this work, we already made the testing procedure possible for the non-compact case.
Either way, additional simulations for the testing procedure would be an obvious next step. Corollaries in the second part might even enable Bootstrap methods. Because of the iterative character of Hawkes processes it has been seen to be very much time consuming to do any calculations. These tries included in particular attempts to find an optimal or even a good choice of bandwidth h. There might be different tools to use or approaches to try to discuss those choices further.

Bibliography

Y. Aït-Sahalia, J. Cacho-Diaz and R.J.A. Leaven (2015). Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, **117**: 585–606.

P. K. Andersen, O. Borgan, R.D. Gill and N. Keiding (1993). Statistical Models Based on Counting Processes. *Springer.*

E. Bacry, K. Dayri and J.F. Muzy (2012). Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data. *Eur. Phys. J. B*, **85**: 157.

E. Bacry, I. Mastromatteo and J.-F. Muzy (2015). Hawkes Processes in Finance, *Mark. Microstructure Liq.* **01** 1550005 [59 pages]

E. Bacry and J.F. Muzy (2016). First- and second order statistics characterization of Hawkes processes and non-parametric estimation, *IEEE Trans. Inform. Theory* **62** 2184–2202

F. Bai, F. Chen and K. Chen (2015), Semiparametric estimation of a self-exciting regression model with an application in recurrent event analysis, *Statistica Sinica* **25**, 1503–1526.

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

F. Chen and P. Hall (2015), Nonparametric estimation for self-exciting point processes – a parsimonious approach, *J. Comp. Graph. Statist.* **17**, 571–595.

S. Clinet and Y. Potiron (2018), Statistical inference for the doubly stochastic self-exciting process, *Bernoulli* **24** 3469–3493.

D. J. Daley and D. Vere-Jones (1988), *An introduction to the theory of point processes*, Springer, New York.

R. Dahlhaus (1996), On the Kullback-Leibler information divergence of locally stationary processes. *Stochastic Process. Appl.* **62**, 139–168.

S. Donnet, V. Rivoirard and J. Rousseau (2020), Nonparametric Bayesian estimation for multivariate Hawkes processes, *Annals of Statistics* **48**, 2698–2727

M. Eichler, R. Dahlhaus and J. Dueck (2017), Graphical Modeling for Multivariate Hawkes Processes with Nonparametric Link Functions, *J. Time Ser. Analysis* **38**, 225–242

E.W. Fox, M.B. Short, F.P. Schoenberg, K.D. Coronges and A.L. Bertozzi (2016), Modeling e-mail networks and inferring leadership using self-exciting point processes, *J. Amer. Statist. Soc.* **111** 564–584

N.R. Hanssen, P. Reynaud-Bouret and V. Rivoirard (2015), Lasso and probabilistic inequalities for multivariate point processes, *Bernoulli* **21**, 83–143.

N. Hautsch (2012). Econometrics of Financial High-Frequency Data *Springer.*

A.G. Hawkes (1971). Spectra of Some Self-Exciting and Mutually Exciting Point Processes *Biometrika.*

A.G. Hawkes and D. Oakes (1974). A cluster process representation of a self-exciting process. *J. Appl. Prob.* **11**, 493-503.

M. Kirchner (2016), Hawkes and INAR(∞) processes, *Stoch. Proc. Appl.* **76**, 1514–1521.

M. Kirchner (2017), An estimation procedure for the Hawkes process, *Quantitative Finance* **17**, 571–595.

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

T. Omi, Y. Hirata and K. Aihara (2017), Hawkes process model with a time-dependent background rate and its application to high-frequency financial data, to appear in *Phys. Rev. E* **96** 012303

M. Rambaldi, E. Bacry and F. Lillo (2016) The role of volume in order book dynamics: a multivariate Hawkes process analysis, *Quantitative Finance* **17**, 999-1020

H. Ramlau-Hansen (1983) Counting process intensities by means of kernel functions, *The Annals of Statistics* **11(2)**, 453–466.

R. Rebolledo (19830) Central Limit Theorems for Local Martingales, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, Springer* **51**, 269–286.

P. Reynaud-Bouret (2006), Compensator and exponential inequalities for some supreme of counting processes, *Statistics and Probability Letters* **76**, 1514–1521.

P. Reynaud-Bouret and E. Roy (2007), Some non-asymptotic tail estimates for Hawkes processes, *Bull. Belg. Math. Soc. Simon Stevin* **13**, 883–896.

P. Reynaud-Bouret and S. Schabt (2010), Adaptive estimation for Hawkes processes; application to genome analysis, *Annals of Statistics* **38**, 2781–2822.

H. P. Rosenthal (1970). On the subspaces of $\mathrm{L}_p (p > 2)$ spanned by sequences of independent random variables, *Israel J. Math.* **8**, 273-303.

F. Roueff, R. von Sachs, and L. Sansonnet (2016). Locally stationary Hawkes processes. *Stochastic Processes and their Applications* **126**, 1710–1743.

F. Roueff, R. von Sachs (2019), Time-frequency analysis of locally stationary Hawkes processes, *Bernoulli* **25** 1355–1385.

D. Vere-Jones (1970). Stochastic models for earthquake occurence. *J. Roy. Statist. Soc., Ser. B* **32**, 1–62.

M. Westcott (1971). On Existence and Mixing Results for Cluster Point Processes. *Journal of the Royal Statistical Society. Series B* **32:2**, 290–300.