Bachelor Thesis in Statistics, 15 ECTS

LUND UNIVERSITY

A temporal Hawkes process model for shooting occurrences in Sweden

Sara Kopelman

Lund University
School of Economics and Management
Department of Statistics

January 2024

Supervisor: Jonas Wallin

Abstract

The Hawkes process, also referred to as a self-exciting point process, is a class of point processes where the intensity is conditioned on previous events. More specifically, an event occurrence *excites* the process, temporarily increasing the probability of more events occurring. One application of Hawkes processes is to model occurrences of crimes, such as burglaries or shootings. The present thesis attempts to apply the theory of Hawkes processes to shooting occurrences in Sweden, by exploring whether a temporal Hawkes process model is suitable to model shooting occurrences in the Police regions of Stockholm, Väst (West) and Syd (South). The results indicate that although there is reason to believe that shooting occurrences in the regions exhibit self-exciting tendencies, a temporal Hawkes process model with time invariant parameters is not adequate. The present thesis suggests that the reason for this lie in the parameters not being constant over time. Furthermore, based on previous research it is likely that a spatial component is needed to adequately capture the underlying process.

Keywords: Point processes, stochastic processes, statistics, conditional intensity, crime, shootings

Conventions and abbreviations

$\mathbb{N}_0$ Set of natural numbers with zero, i.e. $\{0, 1, 2, \dots\}$

E Expectation

P Probability

$\mathcal{H}$ History

λ Intensity

Λ Compensator

f Probability density function

F Cumulative distribution function

$\mathcal{O}$ Big order

Uni Uniform distribution

Exp Exponential distribution

CDF Cumulative distribution function

ECDF Empirical distribution function

PDF Probability density function

Q-Q Quantile-quantile

Table of Contents

1 Introduction

When performing statistical analysis on crime data, there are multiple approaches that can be considered. According to Reinhart and Greenhouse (2018), some of the more common methods include: identifying crime 'hot-spots', studying near-repeat effects or performing regression analysis focusing on leading indicators or local risk factors. However, another promising method that has received attention in the last decade is to model crime using a Hawkes process.

The Hawkes process, also known as a self-exciting point process, is defined by Hawkes (1971) as a type of point process wherein occurrences in the past determine the intensity of current occurrences, with the intensity decaying with time in accordance with some function. It can be used to create statistical models for the occurrences of many disparate phenomena, with some examples being earthquakes, trade orders, bank defaults and gang violence (Laub, Lee & Taimre 2021:7). The commonality for all of these occurrences lies in all of them exhibiting self-exciting tendencies; i.e. an event occurring increases the chance of more occurrences during some period after the initial occurrence.

The aim of this bachelor thesis is to determine whether a Hawkes process is suitable to model shooting occurrences in Sweden. This will be achieved by first conducting a simulation study in order to determine a good estimation procedure. Following this, model estimations will be made based on real data of shooting occurrence times from the three police regions of Stockholm, Syd (South) and Väst (West).

As previously stated, the Hawkes process has emerged as a new promising method of statistical analysis on crime data and has thus far been utilized to model, among others, burglaries (Mohler et al. 2011, Reinhart and Greenhouse 2018), mass shootings (Boyd & Molnyeux 2021) and gang-related violent crimes (Park et al. 2021). However, as of yet, it has seen little application on Swedish or even European data. A Hawkes process model of shooting occurrences in Sweden could therefore serve to increase the understanding of the statistical backgrounds of shootings in Sweden.

The organization of this thesis is as follows: after the introduction, a background chapter will follow wherein a description of the Hawkes-process and summary of research of relevance to the present thesis will be given. After the background chapter, a chapter detailing the present study itself and its results will follow. The chapter thereafter will further discuss the results and the potential reasons behind them. The concluding chapter will summarize the results and conclusions of the present thesis, as well as give some suggestions for potential avenues for future research.

2 Background

The aim of this chapter is to provide an overview of the Hawkes process as well as an outline of previous research of relevance to the present thesis. However, in order to properly explain the Hawkes process it is necessary to have an understanding of point process and in particular of the Poisson process. An overview of fundamental concepts relating to point processes and the Poisson process will thus be provided as well.

2.1 Point-processes

Point processes are a class of stochastic processes "whose realizations consist of point events in time or space"(Cox & Isham 1980:0). According to Cox & Isham (1980:1) , point processes have a large variety of applications, with some examples being the study of the time sequence of radioactive emissions, the sequence of dates for disasters, road traffic studies and operational research. Consider a store where customers can arrive at any time the store is open. If we propose that the event of a customer arriving follows some kind of stochastic process we can regard the series of arrival times as series of random variables, giving us the point process which we define as

Definition 1 (Laub, Lee & Taimre 2021:7) If a sequence of random variables $T = \{t_1, t_2, ...\}$, taking values in $[0, \infty)$, has $P(0 \leq t_1 \leq t_2 \leq ...) = 1$, and the number of points in a bounded region is almost surely finite, then T is a *(simple) point process.*

Also of interest is the concept of the time between events, the *inter-arrival time*, which can be defined as

Definition 2 (Bas 2019:9) Consider the point process $T = \{t_0, t_1, ..., t_n\}$, where t_n is the arrival time of the nth event. Then

$$\tau_1 = t_1 - t_0$$

$$\tau_2 = t_2 - t_1$$

$$\cdots$$

$$\tau_n = t_n - t_{n-1}$$

are the 1st, 2nd, $\cdots$, nth inter-arrival time for the point process T and $\tau = \tau_0, \tau_1, ..., \tau_n$ is a stochastic process with the random variables denoting the inter-arrival times.

Note that: $t_n = \tau_1 + \tau_2 + \cdots + \tau_n$

Finally, instead of considering the sequence of variables (arrival times), we may instead focus on the number of arrivals, i.e. the *count*, which brings us to the counting process:

Definition 3 (Laub, Lee & Taimre 2021:7) A *counting process* is a stochastic process $N(t) : t \geq 0$ taking values in $\mathbb{N}_0$ that satisfies $N(0) = 0$, is almost surely finite and is a right-continuous step function with increments of size $+1$. Further, denote by $(\mathcal{H}(t) : t \geq 0)$ the *history* of the arrivals up to time t. (Strictly speaking $\mathcal{H}(\cdot)$ is a filtration, that is, an increasing sequence of σ-algebras.)

As apparent from the above definitions, one might consider the point-process, the set of inter-arrival times, and the counting process to be three different realisations of the same underlying process, where the point process is defined as the total time until arrival n, the counting process as total number of arrivals until time t, and the inter-arrival time as the time between the nth and $(n-1)$th arrival. According to (Laub, Lee & Taimre 2021:11) one may characterise a point process by specifying the distribution function of the next arrival conditional on the past, defining the following conditional CDF (also referred to as the conditional arrival function), and joint PDF.

$$F(t|\mathcal{H}(u))) = \int_u^t P(T_{k+1} \in [s, s + ds]|\mathcal{H}(u)) \, ds = \int_u^t f(s|\mathcal{H}(u)) \, ds \tag{1}$$

$$f(t_1, t_2, ..., t_k) = \prod_{i=1}^{k} f(t_i|\mathcal{H}(t_{i-1})) \tag{2}$$

However, as it is rather cumbersome to have to continually specify that the functions are conditional on $\mathcal{H}(\cdot)$ it is customary to denote this by a superscript asterisk (Laub, Lee & Taimre 2021:11). $F(t|\mathcal{H}(u)))$ will thus be abbreviated as F^* and $f(t|\mathcal{H}(u)))$ as f^* .

As the conditional arrival distribution is difficult to work with, it is generally more common to characterise a point process by its conditional intensity function(Laub, Lee & Taimre 2021:11-12).

Definition 4 (Laub, Lee & Taimre 2021:11-12) Consider a counting process $N(\cdot)$ with associated histories $\mathcal{H}(\cdot)$. If a (non-negative) function $\lambda^*(t)$ exists such that

$$\lambda^*(t) = \lim_{h \downarrow 0} \frac{E[N(t+h) - N(t)|\mathcal{H}(t)]}{h} \tag{3}$$

which only relies on information of $N(\cdot)$ in the past (that is, $\lambda^*(t)$ is $\mathcal{H}(t)$-measurable), then it is called the *conditional intensity function* of $N(\cdot)$.

Also of importance is the integrated conditional intensity function, which is utilized in parameter estimation and evaluating goodness-of fit. It is defined as

Definition 5 (Laub, Lee & Taimre 2021:13) For a counting process $N(\cdot)$ the non-decreasing function

$$\Lambda(t) = \int_0^t \lambda^*(s)\, ds \tag{4}$$

is called the *compensator* of the counting process.

2.2 The Poisson Process

Having defined the point process we will now move to discuss the perhaps defining example of one, the *Poisson process*. As previously stated, the Poisson process is of particular importance among all the point processes. More specifically, it can be said to play a similar role to that of the normal distribution in the study of random variables, while at the same time being perhaps the simplest point process (Cox & Isham 1980:45). There are several ways to define the Poisson process but three common equivalent ways to define it are: via its intensity function $\lambda(t)$:

Definition 6 (Gut 2009:221-222) A Poisson process is a counting process $\{N(t),\ t \geq 0\}$ with independent, stationary, Poisson-distributed increments. Also $N(0) = 0$. In other words,

1. $N(0) = 0$
2. the increments $\{N(t_k) - N(t_{k-1}),\ 1 \leq k \leq n\}$ are independent random variables for all $0 \leq t_0 \leq t_1 \leq \cdots \leq t_n$ and all n;
3. There exists $\lambda > 0$ such that

$$N(t+s) - N(s) \in Po(\lambda(t-s)), \quad \text{for } 0 \leq s \leq t$$

The constant λ i called the intensity (or rate), of the process.

Note that by the law of large numbers $N(t)/t \xrightarrow{p} \lambda$ as $t \to \infty$, meaning that the intensity measures the frequency or density of events (Gut 2009:222).

Another equivalent definition of the Poisson process is through its increments:

Definition 7 (Ross 2014:314) The counting process $\{N(t), t \geq 0\}$ is said to be a *Poisson process having rate* λ, $\lambda > 0$, if

1. $N(0) = 0$
2. The process has stationary and independent increments
3. $P[(N(h) = 1] = \lambda h + o(h)$
4. $P[N(h) \geq 2] = o(h)$

Finally, one may define the Poisson process by the distribution of its inter-arrival times:

Definition 8 (Gut 2009:231) Let $\{N(t),\ t \geq 0\}$ be a counting process with $N(0) = 0$, let t_1 be the time of the first occurrence, and let t_k be the time between the $(k-1)$th and the kth occurrences for $k \geq 2$. If $\{t_k, k \geq 1\}$, are independent Exp[1](θ)-distributed random variables for some $\theta > 0$, then $N(t)$ is a Poisson process with intensity $\lambda = \theta^{-1}$.

The basic Poisson process that has been thus defined is also sometimes referred to as the *homogeneous* Poisson process, due to the fact that its intensity is constant. However, when the we allow the intensity to be a function of t, we get the *inhomogeneous* Poisson process, which is defined as:

Definition 9 (Gut 2009:231) The counting process $\{N(t),\ t \geq 0\}$ is said to be a *inhomogeneous Poisson process* with intensity function $\lambda(t)$, $t \geq 0$, if

1. $N(0) = 0$
2. $\{N(t),\ t \geq 0\}$ has independent increments
3. $P[N(t + h) - N(t) \geq 2] = o(h)$
4. $P[N(t + h) - N(t) = 1] = \lambda(t)h + o(h)$

As can be seen from the above definition, the homogeneous Poisson process differs from the homogeneous in that it does not have stationary increments.

[1]PDF: $f(x) = \frac{1}{\theta} e^{-x/\theta}$

2.3 The Hawkes process

Now that sufficient concepts relating to point processes have been defined we will move on to the point process model that is the subject of this thesis, the *Hawkes process*,

2.3.1 Definition

As stated in the introduction, the Hawkes process is named after its originator Alan G. Hawkes (1971:84), who described a self-exciting process wherein "the current intensity of events is determined by events in the past", referring to it as a self-exciting point process with conditional intensity function

$$\lambda^*(t) = \mu + \int_{-\infty}^{t} g(t-u)\, dN(u) \tag{5}$$

Recall the definition of the intensity function, in Definition 4. A self-exciting point process is then a process wherein an arrival causes the conditional intensity function to increase (Laub et al. 2015). Subsequently, the function g, needs to be strictly positive in order for excitation to occur, while also decaying in order to avoid an exploding process. While Hawkes (1971) only considered the case of excitation one may naturally also consider the case of a *self-inhibiting* process, wherein an arrival causes the conditional intensity to decrease (Laub, Lee & Taimre 2021:111). We thus define the Hawkes-process as

Definition 10 (Laub, Lee & Taimre 2021:16) A simple point process with conditional intensity function

$$\lambda^*(t) = \mu + \int_{0}^{t} g(t-u)\, dN(u) \tag{6}$$

with background rate $\mu > 0$ and excitation function (or kernel) $g(\cdot) \neq 0$[2] is a *Hawkes process.*

Furthermore if we let $\{t_1, t_2, ...\}$ represent the observed sequence of past occurrences we get the equivalent form for the conditional intensity function

$$\lambda^*(t) = \mu + \sum_{t_i < t} g(t - t_i) \tag{7}$$

Note that while μ is often assumed to be constant it might also itself be a function of t which turns equation (7) into the following

$$\lambda^*(t) = \mu(t) + \sum_{t_i < t} g(t - t_i) \tag{8}$$

[2]In the case of $g(\cdot) = 0$, the process is a homogeneous Poisson process

2.3.2 The excitation function

Having now defined the Hawkes process we will proceed to describe some common choices for its excitation function. The exponential kernel is the most common and is also the shape that was originally considered by Hawkes (1971). With exponential decay the excitation function is defined as

$$g(t) = \alpha e^{-\beta t} \tag{9}$$

with $\alpha > 0$ representing the increase in intensity after an arrival and $\beta > 0$ representing the subsequent rate of exponential decay. This gives us the following conditional intensity function

$$\lambda^*(t) = \mu + \int_0^t \alpha e^{\beta(t-u)} \, dN(u) = \mu + \sum_{t_i < t} \alpha e^{-\beta(t-t_i)} \tag{10}$$

another common choice for Hawkes processes is the power law kernel defined as

$$g(t) = \frac{k}{(c+t)^P} \tag{11}$$

and is a commonly used for aftershock models for earthquakes (Laub, Lee & Taimre 2021:17)

2.3.3 Stationarity

When describing the stationarity requirements of a Hawkes process it is generally simpler to view it as a branching process, also referred to as the *immigration-birth representation*. Laub, Lee & Taimre (2021:18) explains the representation by likening it to a country where inhabitants can arrive as immigrants or by birth. The arrival of immigrants forms a homogeneous Poisson process with rate λ. Furthermore, each immigrant has the potential of generating births, whose arrivals in turn follow an inhomogeneous Poisson process. By viewing the Hawkes process as a branching process we can utilize the branching ratio n^* which represents the expected number of births generated by an immigrant (Laub, Lee & Taimre 2021:18-19). It is defined as

$$n^* = \int_0^\infty g(t) \, dt \; . \tag{12}$$

The branching ratio is important in that it determines the stability of the Hawkes process. Specifically the stationarity requirement for a Hawkes process is $n^* \in (0, 1)$.[3] In the case of the exponential kernel the branching ratio is

$$n^* = \int_0^\infty \alpha e^{-\beta t} \, dt = \frac{\alpha}{\beta} \tag{13}$$

For an Hawkes process with exponential kernel to be stationary it is thus necessary that $\alpha < \beta$.

[3] For proof, see Laub, Lee & Taimre 2021:18-25

2.3.4 Extensions of Hawkes process models

Having now described some important characteristics of the Hawkes process we will conclude the section with a brief description of some common extensions. While the process as described by Hawkes (1971) only applied to temporal Hawkes processes many phenomena also exhibit self-exciting tendencies in the spatial dimension. In order to correctly capture the self-exciting behaviour of many processes it is thus necessary to extend the model to include a spatial component. The resulting *spatio-temporal Hawkes process* has the following conditional intensity function

$$\lambda^*(s,t) = \mu(s) + \sum_{i:t_i<t} g(s - s_i, t - t_i), \tag{14}$$

where $\{s_1, s_2, \ldots, s_n\}$ denotes the spatial location of the arrivals (Reinhart 2018). Some applications for spatio-temporal Hawkes processes will be described in the next section. Another important extension of the Hawkes process is the *marked Hawkes process*[4], wherein features of the arrivals other than their observation times (or locations in the spatio-temporal case) are included. A typical example is including the observed magnitudes as marks when modelling earthquake arrivals (Reinhart 2018). Methods for parameter inference and evaluating goodness-of-fit are given in the methodology section 3.3.

[4]There are many types of marked point process models and they are thus not a feature specific to Hawkes processes. See Cox & Isham 1980:132-142 for a detailed explanation of marked point processes.

2.4 Previous studies

The following section consists of an overview of previous studies involving Hawkes processes, with focus on those utilizing crime data. In addition, a brief summary of two previous studies on crime data in Sweden is included.

As explained in the introduction, the Hawkes process was first described in Hawkes 1971, then referred to as a self-exciting process, a name that is still used today. It has since then seen a variety of applications. Reinhart (2018) suggests that four major applications of Hawkes processes are: earthquake models, crime forecasting, epidemic infection forecasting, and events on networks. In the present thesis only the first two will be described, and in case of earthquake models only briefly. It is therefore suggested to peruse Reinhart 2018 for a more in-depth view.

Due to the clustering behaviour of earthquake occurrences with main shocks followed by aftershocks, they were from early on (e.g. Adamopoulos 1976) suggested as a potential application for Hawkes processes. However, the most important contribution is the Epidemic Type Aftershock-Sequences (ETAS) model formulated by Ogata (1988,1996), which models the conditional intensity of aftershock activity of magnitude M_0 and larger as:

$$\lambda^*(t) = \mu + \sum_{t_i < t} \frac{K_0}{(t - t_i + c)^p} \cdot e^{\alpha(M_i - M_0)} \tag{15}$$

where the parameters K_0, α, c and p are constants. This initial temporal model was subsequently expanded to also include a spatial component, which gives us the spatio-temporal ETAS model (See Ogata 1996, 1999). The ETAS model for aftershocks has been further developed throughout the decades and has been an important influence on modelling self-exciting processes even outside of seismology (Reinhart 2018).

As with earthquakes, crime exhibits many characteristics that makes modelling it as some form of self-exciting process intuitive. It has been observed that there is a phenomenon of repeat victimization, wherein victims of crimes have a high risk of being victimized again, and that this risk is the greatest in the immediate period after victimization (Farrell 1995). This repeat victimization risk has also been observed to be "contagious", where for example houses in the near vicinity of a burglarized house have an increased risk of also being burglarized. This "contagiousness" is known as the *near-repeat hypothesis* (Townsley, Homel & Chaseling 2003). Similarly, Ratcliffe & Rengert (2008), studied the spatio-temporal patterns of shootings in Philadelphia and found that shootings occurring as part of a lovers triangle and disputes during illegal activities carried an elevated risk of retaliating shootings from the victim or someone close to them occurring nearby in a week or two after the initial shooting, i.e. near repeats. Furthermore, it has also been shown that places that have a high risk of crime, i.e. hot-spots, can be divided into *chronic hot-spots* that always carry an elevated risk, and *temporary hot-spots* which persist for only a shorter period as a result of a flare-up (Gorr & Lee 2015). The near-repeat effects, as well as the possible division of the process into a chronic background part and a temporary flare up part are all characteristics that would suggest the Hawkes process as a potential model for modelling crime occurrences.

Mohler et al. (2011) observed these above mentioned characteristics and noted that the resulting clustering patterns were similar to that of seismological events. They (2011) therefore utilized the extended spatio-temporal ETAS model to construct a self-exciting model of Los Angeles residential burglaries, opting to utilize non-parametric methods to avoid having to specify the underlying parametric structure. Trough this study, Mohler et al. (2011) were able to show that the seismological approach may also be used on crime data. In Mohler 2014, a parametric approach with added leading indicator events to the model is instead utilized to estimate a mixed self-exciting point process model of gun crime in Chicago. With this model Mohler (2014) was able to generate predictive hot-spot maps that were more accurate than traditional methods. Reinhart and Greenhouse (2018) further extended Mohler's (2014) model, opting to also include spatial covariates. They (2018) noted that statistical models of crime focusing purely on either the spatial or temporal aspects have a tendency of confounding, noting that this problem can also be extended to other self-exciting processes.

Finally, Park et al. (2021) estimated a Hawkes process model for gang-related violent crimes in Los Angeles, also including spatial covariates, but opting for a non-parametric method to estimate the background rate. By separating the events into those affected by the Gang Reduction Youth Development (GRYD)-program and those not affected they were able to evaluate the program's effect on the retaliation rate. This shows that the Hawkes process model has the potential to be utilized in many different ways in crime modelling, and is thus not simply a new method for generating predictive hot spot maps.

Moving on to studies focusing on Sweden, the Hawkes process has as of yet not been applied to Swedish crime data. There are, however, some studies focusing on near-repeat effects. Sturup et al. (2018) analysed shootings in the cities of Stockholm, Gothenborg and Malmö during the period of 2011 - 2015 and found that near-repeat patterns could be observed in all three cities, noting that the pattern was stronger in Stockholm and Malmö, and weaker in Gothenborg. Similarly, Sturup, Gerell & Rostami (2020) conducted a near-repeat study of hand grenade detonations and shootings together. They found that hand grenade detonations exhibited similar near-repeat patterns to that of shootings, but that no improvements to the near-repeat analysis of shootings were made by adding the hand grenade detonations to the model. They (2020) note that this suggests that although both crimes are heavily connected to criminal groups, they only partially share spatio-temporal patterns.

3 Study

3.1 Purpose

The purpose of this study is to evaluate whether a temporal Hawkes Process model can be utilized to model shooting occurrences in Sweden. The research question are as follows:

1. Is a temporal Hawkes process a suitable model for shooting occurrences in Sweden
2. Are the parameters constant or do they vary with time?
3. How does the appropriateness of the model and its resulting estimations differ between different Police regions?

3.2 Data

The data for the present study consist of confirmed shootings in Sweden during the period of January 1st 2018[5] until September 2nd 2023. According to the Swedish Police[6] "a confirmed shooting is an incident wherein projectiles have been discharged from a firearm, and tangible evidence thereof is discernible in the form of bullets, casings or damage to materials or individuals arising from the discharge. Alternatively, there must be more than one independent eyewitness to the shooting. Moreover, the shooting must be deemed unlawful and not obviously accidental."

The information provided about each confirmed shooting are as follows: Reference number, date of incident, number of deaths, number of wounded, police region, police district (or local police district), and geographical coordinates in SWEREF 99[7] rounded to approximately 100 meters. The total number of incidents recorded amounted to 3807. However, incidents with several plaintiffs were recorded as multiple incidents with the same reference number. As these incidents should be considered to be the same shooting they were combined, resulting in a total of 2397 incidents. Table 1, on the following page, displays some descriptive statistics for Sweden as a whole, as well as for the three most populous police regions[8] of Stockholm, Väst (West) and Syd (South).

[5]The actual data set includes shootings from 2016 but according to the former police commissioner Lars Ojelind the data is considered to be of lower quality before 2018 (SVT, no date) and it was therefore excluded.

[6]Polisen, no date-a

[7]See Lantmäteriet, no date

[8]See Polisen no date-b for a description of the organization of the Swedish police, including its police regions.

Table 1: Shooting occurences in Sweden

		Days between shootings			
Region	N. of shootings	Mean	St.d	Min	Max
Sweden	2088	1.00	1.12	0	11
Stockholm	725	2.89	2.98	0	21
Väst	279	7.45	8.27	0	51
Syd	422	4.95	5.82	0	35

As explained in 2.4, there are intuitive reasons to use Hawkes processes to model shooting occurrences, in that retaliating shootings in the period shortly after the initial occurrence results in the arrivals following a clustering pattern.

Figure 1, below, displays the arrival times of shootings in each of the three police regions, during the period of June to December 2021. As can be seen all regions seem to display a tendency of clustering, an indication that a Hawkes process might be an appropriate model for the arrivals.

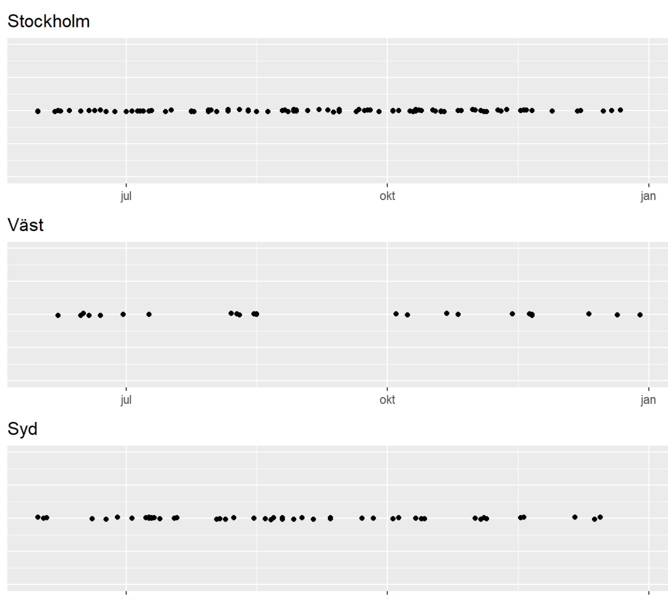

Figure 1: Arrival times of shooting occurrences during June-December 2021, by police region

It is also of relevance to ascertain whether the arrivals seem to follow some kind of underlying trend. Figure 2, 3, and 4, on the following pages, display the observed shooting occurrences aggregated by month, weekday and year for the three police regions. Note that the 2023 data was excluded as it would otherwise skew the monthly data due to it not including the whole year.

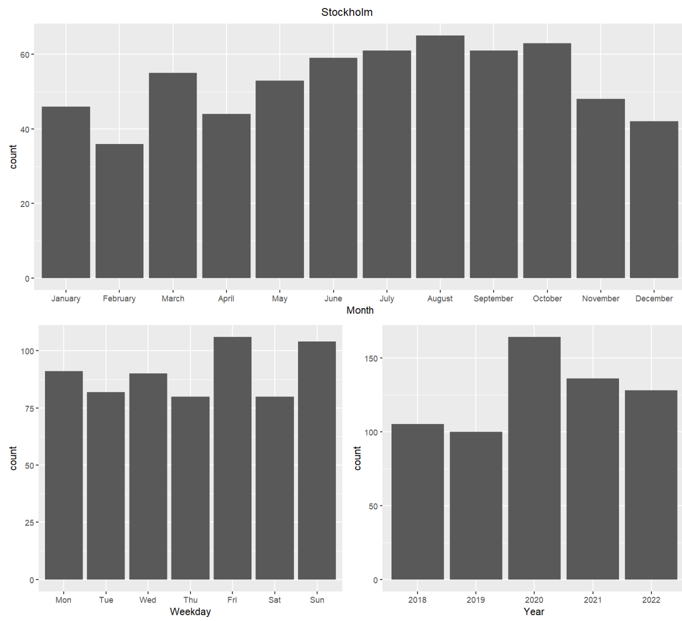

Figure 2: Shooting occurrences in the Stockholm police region 2018-2022, by weekday, month and year

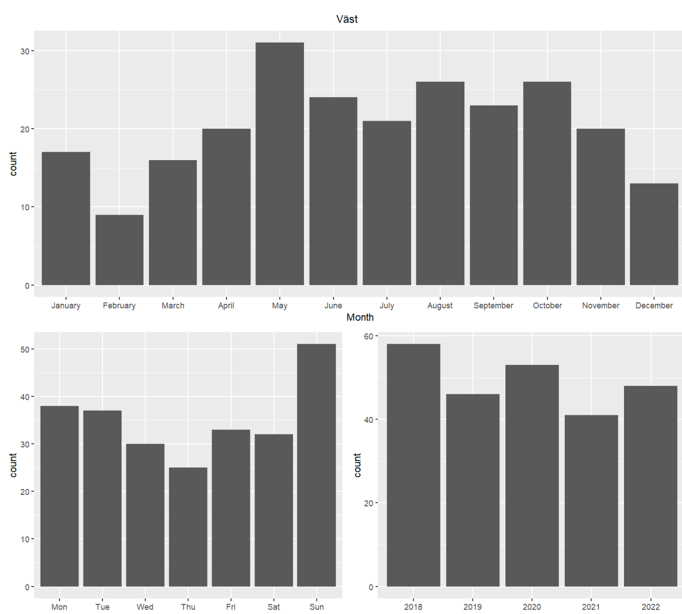

Figure 3: Shooting occurrences in the Väst police region 2018-2022, by weekday, month and year

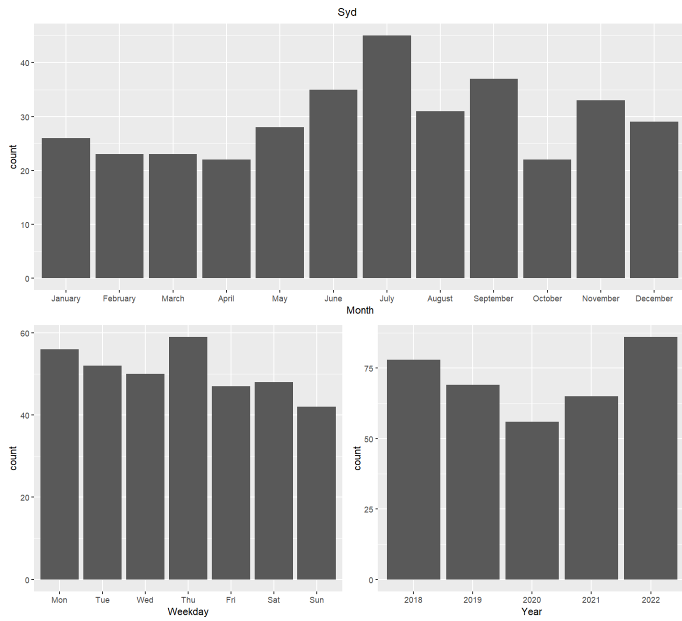

Figure 4: Shooting occurrences in the Syd police region 2018-2022, by weekday, month and year

As can be seen, all of the regions seem to display an underlying seasonal trend, with more shootings occurring during the warmer months. Furthermore it is clear that the numbers of shootings per year is not constant, with Stockholm in particular showing a clear spike in 2020. Finally the regions all have differing patterns when it comes to day of the week.

3.3 Methodology

The following section provides a brief overview of the statistical methods that will be utilized in the study. Statistical analysis was performed in R version 4.3.2. The packages that were utilized are described in the bibliography.

3.3.1 Estimation methods

Maximum Likelihood: Maximum likelihood is one of the most widely used methods for parameter inference in frequentist statistics. It is beyond the scope of the present thesis to provide a rigorous explanation of the theory underpinning maximum likelihood estimation, but generally it relies on numerically maximizing the log-likelihood function[9]. While Bayesian methods are becoming increasingly more common for inference on Hawkes processes (particularly in the multivariate case), the frequentist approach of parameter estimations using

[9]See Millar 2011 for a detailed overview of the theory and practise of inference with maximum likelihood.

maximum likelihood is still the most popular approach. The likelihood function for the temporal Hawkes process is derived from that of a simple point process, which is given by the following theorem

Theorem 1 (Laub, Lee & Taimre 2021:38) Let $N(\cdot)$ be a simple point process with conditional intensity $\lambda^*(\cdot)$, and compensator $\Lambda(\cdot)$. If we observe all arrival times over the time period $[0, T]$, denoted $\{t_1, \ldots, t_{n(T)}\}$, then the likelihood function L for $N(\cdot)$ is

$$L = \left[\prod_{i=1}^{n(T)} \lambda^*(t_i)\right] e^{-\Lambda(t)} \tag{16}$$

and the log-likelihood is of the form

$$\ell = \sum_{i=1}^{n(T)} \log\left(\lambda^*(t_i)\right) - \Lambda(T) \tag{17}$$

In the case of a Hawkes process with exponential kernel, the likelihood function is then given by Equation (18), below

$$\ell = \sum_{i=1}^{n} \log\left[\mu + \alpha \sum_{j=1}^{i-1} e^{-\beta(t_i - t_j)}\right] - \mu T - \frac{\alpha}{\beta} \sum_{i=1}^{n(t)} \left[1 - e^{-\beta(T - t_i)}\right] \tag{18}$$

This equation is problematic as it has a complexity of $\mathcal{O}(n(T)^2)$. It can however be simplified to a complexity of $\mathcal{O}(n(T))$, if a recursive approach is utilized (Laub, Lee & Taimre 2021:41).

As explained in 2.3, the Hawkes process is a counting process model on a continuous measure. However, in the case of real life count data, the actual event times are often not available, with the data instead being collected into bins of fixed sizes (such as hours, days, or months), which is also the case for the data used in the present thesis. As a consequence, estimation procedures of point processes are no longer applicable. One way to handle this is to "scramble" the data, i.e to add some noise, typically by adding a random simulations from the Uniform distribution to each arrival time in order to "scramble" their position in the bins (Cheysson & Lang 2023, Meyer, Elias & Höhle 2021). The resulting arrival times are then sorted from smallest to largest, upon which maximum-likelihood estimation is performed as if the data was originally continuous. However, this naturally assumes that bin-positions of the occurrences prior to binning were uniformly distributed, which may not actually be the case. The ramifications of this will be further discussed in Chapter 4.

Whittle likelihood While the above estimation procedure is one method to deal with the problem of binned arrival times, another method proposed by Cheysson and Lang (2022) is to utilize log-spectral, or Whittle, likelihood. They found that for an Hawkes process with exponential kernel and small bin sizes, Whittle likelihood estimation fared almost as well as the MLE from continuous data.

Sliding Windows: The sliding window technique is an algorithm wherein a ordered sequences of events (such as a count sequence or time series) are categorized into active and expired elements. The *window* of the sliding window consist of the currently active elements. Operations are performed on the active elements after which the window is slid along the line, expanding to include the next expired element, and in the case of a window with fixed size discarding the last active element in the previous window (Braverman 2007). While there are existing estimation methods for estimating time-variant Hawkes parameters they come with the caveat that they require an *a priori* assumption about their underlying function. By utilizing a sliding window approach and treating the parameters as constant we are able to capture the time-variant nature of the parameters without depending on prior assumptions. This approach is utilized in Godoy et al. 2016, where simple local likelihood estimation is used to estimate time variant Hawkes parameters. While their approach utilizes weights to create truly smooth windows, due to having to rely on already existing estimation packages, the present thesis will solely utilize overlapping windows, where each window is estimated as if it were an independent process.

3.3.2 Goodness-of-fit

Poisson transformation: The most important tool in determining the goodness-of-fit of an estimated Hawkes process is to transform it to an unit rate Poisson process using the compensator defined in Definition 5 and the random time change theorem, which is defined as follows:

> **Theorem 2 (Laub, Lee & Taimre 2021:79)** Let $\{t_1, t_2, \ldots, t_k\}$ be a realisation over time $[0, T]$ from a point process with conditional intensity function $\lambda^*(\cdot)$. If $\lambda^*(\cdot)$ is positive over $[0, T]$ and $\Lambda(T) < \infty$ almost surely, then the transformed points $\{\Lambda(t_1), \Lambda(t_2), \ldots, \Lambda(t_k)\}$ form a Poisson process with unit rate.

Thus, by utilizing the compensator to transform the estimated process into a Poisson process, we can ascertain the goodness of fit using the same method as with a Poisson process.

Kolmogorov-Smirnov test: The Kolmogorov-Smirnov (KS) test is a test commonly used to test whether a sample is derived from a specific parametric distribution F with the following null and alternative hypotheses (Massey 1951):

- H_0 : Sample belongs to F
- H_A : Sample does not belong to F

In addition to the KS-test, graphical comparisons between the ECDF and CDF will also be utilized.

Ljung-Box test: The Ljung-Box (LB) test is a goodness-of-fit test for whether time series observation are autocorrelated at the mth lag. The null and alternative hypotheses are as follows:

- H_0 : The observations are independent
- H_A : The observations exhibit auto-correlation

In the present thesis, the LB-test is mainly used to test for independent inter-arrival times in a suspected Poisson process. See Ljung & Box 1978 for a more detailed explanation of the test.

3.4 Results

3.4.1 Simulation Study

Before estimating parameters from the actual data of shooting occurrences, a simulation study was conducted in order to compare different methods for estimation. The simulation study consists of three parts: In the first part Hawkes processes are generated and then estimated using maximum likelihood estimation. In the next part, the resulting simulations are binned and estimations are made based upon that binned data using the methods discussed in 3.3. Finally, in the last part a sliding window approach is explored in order to determine whether it can be used to estimate time variant parameters.

As mentioned above, the first part of the simulation study consisted of estimating the parameters of simulated Hawkes processes using maximum likelihood estimation. 100 simulations were generated from two different Hawkes process with exponential kernel, H1 and H2, with parameters $\{\mu, \alpha, \beta\} = \{0.36, 0.5, 1.2\}$ and $\{\mu, \alpha, \beta\} = \{0.7, 0.1, 0.8\}$. The maximum time horizon for the simulation was set to 2000. A 95% confidence interval for the resulting parameter estimations for each simulation are plotted in Figure 5 and Figure 6, on the following page.

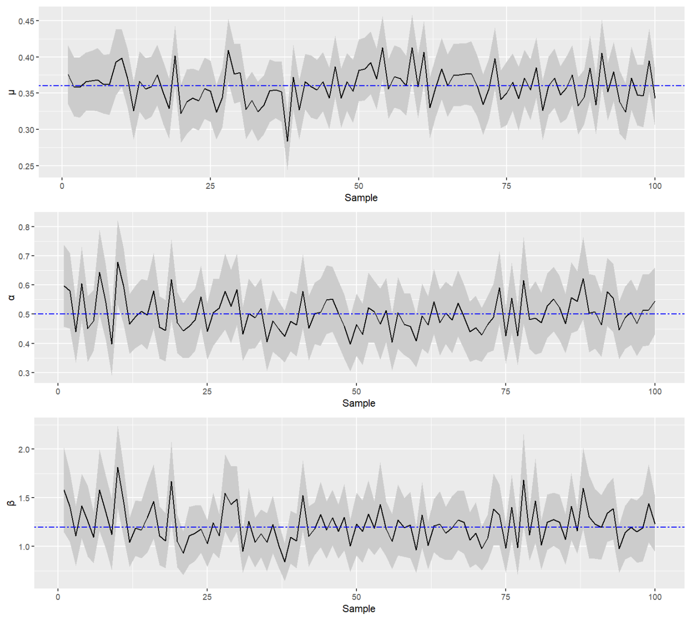

Figure 5: A 95% confidence interval for parameter estimations from 100 simulations of H1

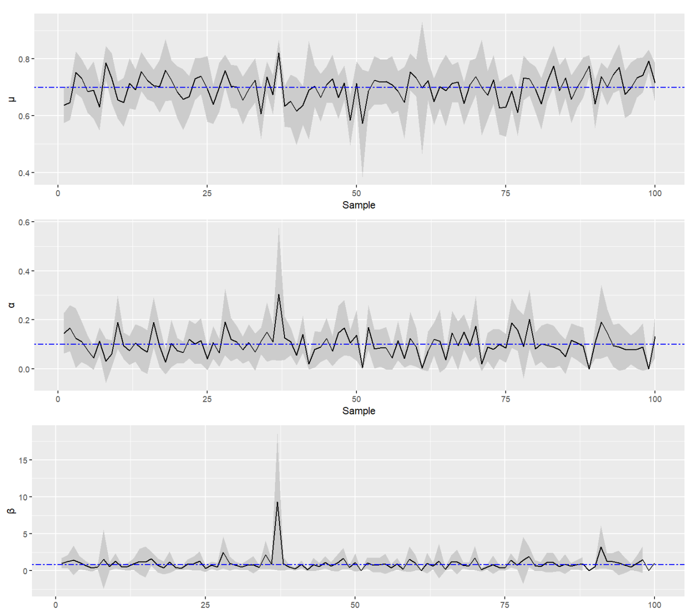

Figure 6: A 95% confidence interval for parameter estimations from 100 simulations of H2

Table 2: MLE Estimations from H1 & H2

	Parameter	Coverage	St.d	Mean Bias
H1				
	μ	94%	0.0231	-0.0004
	α	97%	0.0583	-0.0014
	β	94%	0.1833	-0.0253
H2				
	μ	91%	0.0463	0.0021
	α	87%	0.0489	-0.0005
	β	85%	0.9910	-0.1616

Table 2, above, describes descriptive statistics regarding the parameter estimations. The estimations from H1 generally show low bias and coverage is close to what one would expect. In the case of H2, however, coverage and bias is generally worse, which most likely stems from α being smaller. Since the Hawkes process approaches a Poisson process when α approaches 0, it is natural that the estimation procedure becomes more difficult when α is small. In particular the algorithm were sometimes unable to generate standard errors for β when α was estimated to be very close to 0. This is once again not very surprising as the existence of β depends on α being non-zero. As can be seen in 6, some confidence intervals for α also go below 0, which would make the process self-inhibiting rather than self-exciting. However, as the algorithms used for estimating the parameters do not actually allow for estimations less than 0, we ought to not consider these as actual confidence interval boundaries. Also note that while some confidence intervals for β in the same figure also go below 0, such a process would have an exponentially increasing intensity and would therefore explode.

As discussed in the previous chapter, data of event times are often aggregated to some unit, which causes problems when performing parameter estimation. Consequently, two potential methods to handle this were proposed: "scrambling" the data by adding uniform noise or estimating the parameters by Whittle likelihood. In order to explore the efficiency of these two methods, the resulting simulations of H1 and H2 above were aggregated into bins of size 1, representing "days". Following this, noise in the form of a $Uni(0, 1)$ random variable were added to each observation. The resulting arrival times were then sorted and parameters estimated using the same method of maximum likelihood estimation as above. Whittle likelihood estimation was also performed on the binned event times without added noise. The resulting estimations from the ML with noise procedure is shown in Figure 7 and 8, while the estimations from Whittle likelihood are shown in Figure 9 and 10 on the following two pages. Note, that the package *hawkesbow* used to perform Whittle likelihood estimation does not provide standard errors. Confidence intervals could therefore only be calculated for the parameters estimated with maximum likelihood.

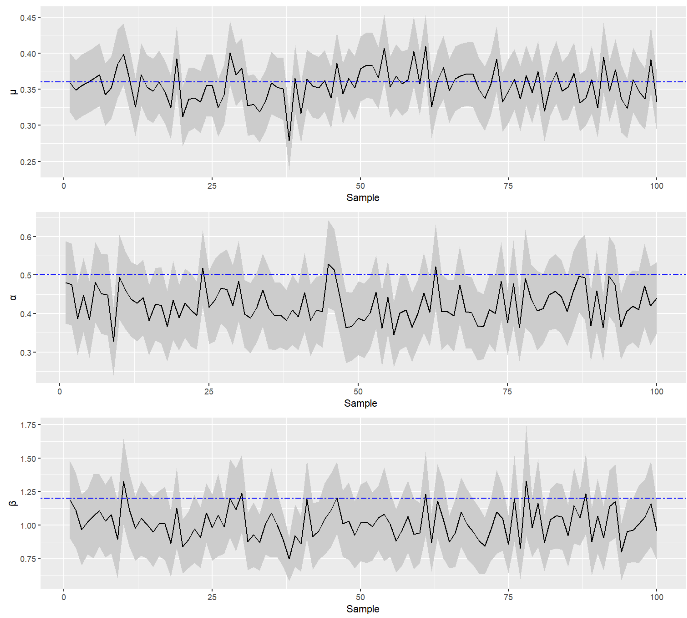

Figure 7: A 95% confidence interval for maximum likelihood parameter estimations from 100 binned simulations of H1

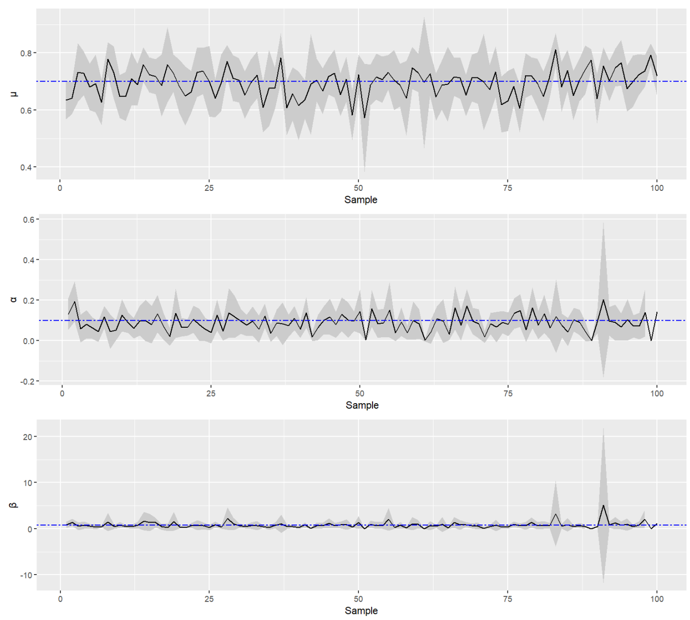

Figure 8: A 95% confidence interval for maximum likelihood parameter estimations from 100 binned simulations of H2

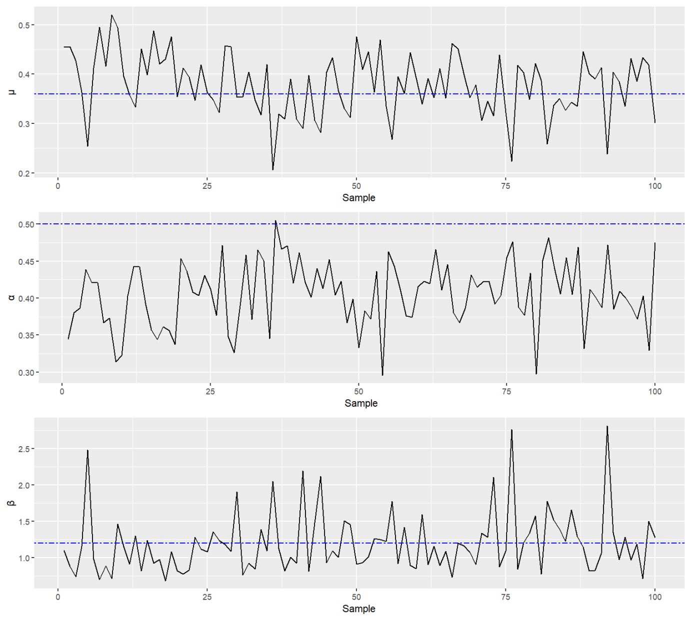

Figure 9: Whittle likelihood parameter estimations from 100 binned simulations of H1

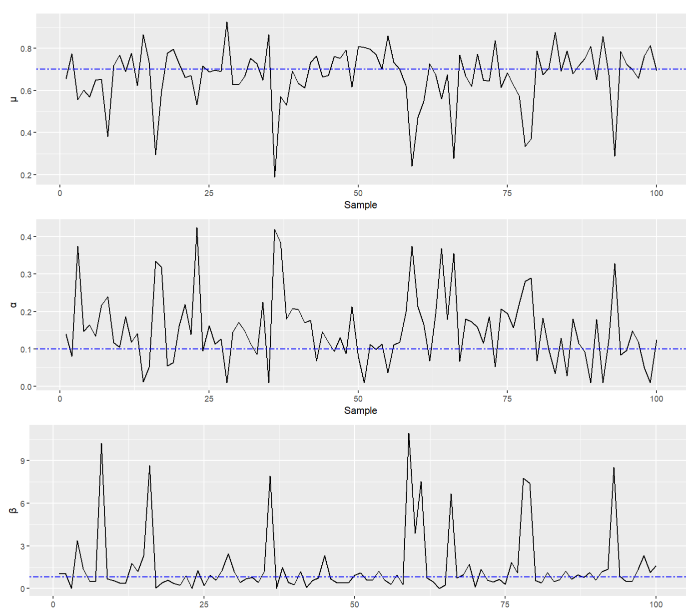

Figure 10: Whittle likelihood parameter estimations from 100 binned simulations of H1

Table 3: ML Estimations from binned simulations of H1 & H2

	Parameter	Coverage	St.d	Mean Bias
H1				
	μ	95%	0.0229	0.0044
	α	59%	0.0435	0.0756
	β	65%	0.1173	0.1848
H2				
	μ	91%	0.0467	0.0041
	α	84%	0.0418	0.0119
	β	83%	0.6593	-0.0047

Table 4: Whittle likelihood estimations from binned simulations of H1 & H2

	Parameter	St.d	Mean Bias
H1			
	μ	0.0628	-0.0188
	α	0.0447	0.0942
	β	0.4212	0.0047
H2			
	μ	0.1397	0.0309
	α	0.0951	-0.0524
	β	2.2918	-0.7481

Descriptive statistics regarding the estimations are provided in Table 3 and 4, above. Unsurprisingly, the mean bias is generally higher for the parameters estimated from the binned data compared to those in Table 2. As for the two different methods utilized, with the exception of β in H1, the mean bias is lower for the estimations performed with maximum likelihood. Similarly the variance is also lower for the maximum likelihood estimated parameters. Furthermore returning to the figures, both methods seem to systematically underestimate α for H1, with the maximum likelihood method also systematically underestimating β in the same model. However, while the bias in the Whittle likelihood looks less systematic for some parameters, many of the biased estimations are very large, making the estimations from the Whittle method ultimately more biased.

Finally, a sliding window approach was attempted in order to explore whether it could be used to estimate time variant parameters, first from a continuous process with maximum likelihood followed by a binned process, once again comparing maximum likelihood with added noise and Whittle likelihood. 100 simulations were run from which parameter estimations were made on sliding windows. The simulated Hawkes process had the same values for $\{\alpha, \beta\} = \{0.5, 1.2\}$ as H1. However, in contrast to H1 a time dependent intensity function was chosen where:

$$\mu(t) = 0.15 + 0.1 \left(1 - \cos\left(\frac{2\pi}{0.5T}t\right)\right) \tag{19}$$

T refers to the maximum time horizon for the simulation which is as before 2000. As with the previous simulation procedures, 100 simulations were generated. However, instead of estimating the parameters from all data points in a sample the estimation procedure was applied to a symmetric sliding window of size 71. Rolling quantiles were then computed for the resulting estimates from all samples.

Figure 11: Rolling quantiles of sliding window estimates from 100 simulations of a Hawkes process with time variant μ

Figure 11 above show the resulting estimations (in gray), and respective rolling quantiles of maximum likelihood estimations from 100 continuous simulations. The sliding window approach is clearly able to capture the time variant pattern of $\mu(t)$ while also keeping α and β approximately constant. The estimations do however seem biased with the 2nd quartile being consistently situated above the true value.

The resulting estimations from the binned simulations are shown in Figure 12 and 13, on the following page.

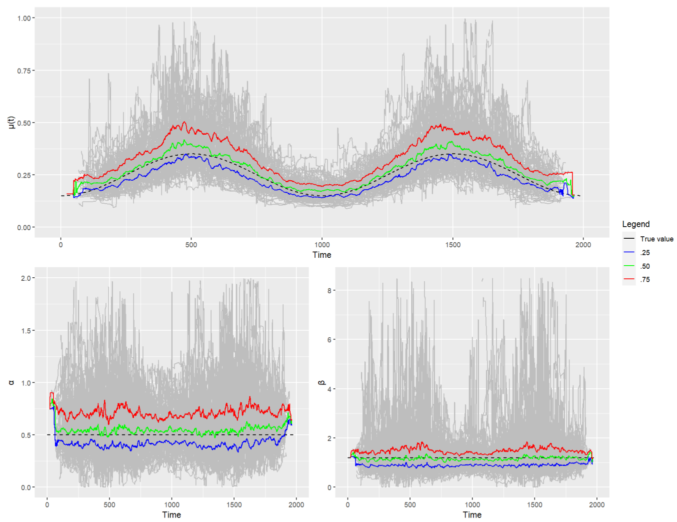

Figure 12: Rolling quantiles of sliding window estimates with maximum likelihood with added noise from 100 binned simulations of a Hawkes process with time variant μ

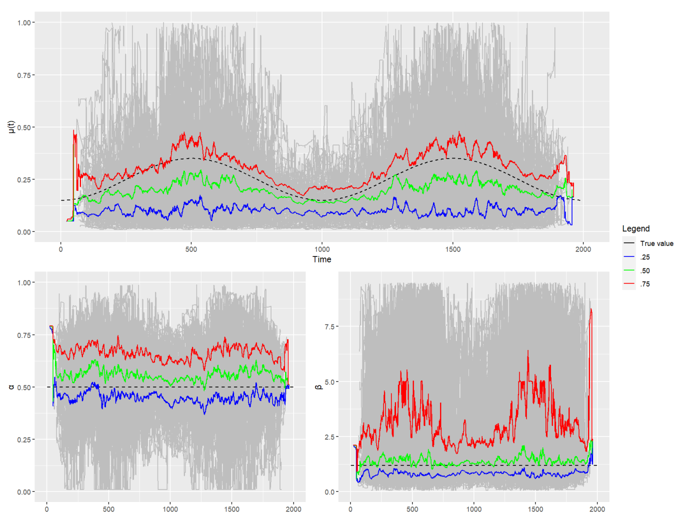

Figure 13: Rolling quantiles of sliding window estimates with Whittle likelihood from 100 binned simulations of a Hawkes process with time variant μ

As with the previous part of the study, parameter estimations with maximum likelihood with added noise seem to perform better than those using Whittle likelihood. In particular, the latter has difficulty differentiating the parameters from one another, resulting in much of the variation in $\mu(t)$ being miss-attributed to β. Furthermore, although the maximum likelihood estimations seem more consistent, individual estimations may be very far from the true value. While the sliding window estimations may on average be able to capture the time variant effects reasonably well, it is likely that the estimation in any individual window may be very far from the true value. Overall, the results of the simulation study indicate that maximum likelihood with added noise generates more consistent estimations than those generated with Whittle likelihood. As a consequence, the estimations from real data in the next section will utilize the former method.

3.4.2 Shooting occurrences

In the following section, parameter estimation of Hawkes processses with an exponential kernel will be made from data of shooting occurrences in the police regions of Stockholm, Väst and Syd, as explained in 3.2. The estimation procedure is as follows: first we control for the case that the process might also be described by a simple Poisson process. This is accomplished by three different methods: By comparing the empirical distribution function (henceforth, ECDF) of the inter-arrival times to that of an exponential distribution function with the mean inter-arrival time, $\bar{\tau}$, as parameter, by performing a Kolmogorov-Smirnov test with the same distribution as the null hypothesis and by testing the assumption of independent inter-arrival times (see 2.2) with a Ljung-Box-test. After controlling for the simple Poisson process, time invariant parameter estimations will be performed using the whole data set for each region, with the method of choice being maximum likelihood estimation with added noise, described in 3.3. Goodness-of- fit for the model is then evaluated using methods discussed in the same section. Finally, a sliding window approach, as performed in 3.4.1, is utilized in order to estimate time variant parameters. The size of the window is 71, as in the simulation study.

Stockholm As stated, the first region to be explored was the Stockholm region. Figure 14, on the following page, displays the ECDF of the inter-arrival times in blue and the theoretical CDF of the exponential function previously discussed, in red. As can be seen the two functions overlap quite heavily, making it not unthinkable that the arrival times of the shootings might be following a simple Poisson process.

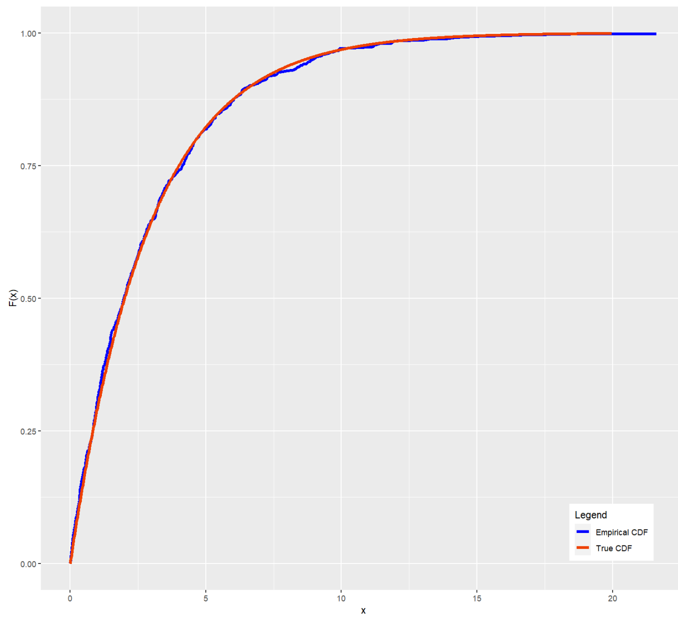

Figure 14: ECDF vs Hypothetical exponential CDF for Stockholm inter-arrival times

While the graphical analysis gave indications that the inter-arrival times follow an exponential distribution, thus giving credence to the potential of the arrival times following a simple Poisson process the same was not true for the KS and LB tests. The tests resulted in p-values amounting to < 0.0001 and 0.025, respectively, meaning that the null hypotheses for both tests may be rejected. There are thus indications that evaluating another model than the simple Poisson process for the inter-arrival times may be worthwhile.

Table 5 below displays the parameter estimations for the time-invariant Hawkes model. As can be seen the standard errors for both α and β are almost the same (or in the case of β larger) than the estimations themselves. As a consequence confidence intervals for α will include 0, meaning that the case of a simple Poisson process cannot be discounted.

Table 5: Parameter estimations for the Stockholm region

	Estimate	Std. Error
μ	0.3282722	0.01822392
α	0.0745741	0.05927210
β	1.4317628	1.70787092

Having established that the parameter estimations are not the most indicative of a Hawkes process, we move on to evaluate their goodness-of-fit. Plots of the transformed inter-arrival times are displayed in Figure 15, on the following page. As can be seen there is a prominent gap in the bottom left between the occurrence time of actual events and compensator. However the Q-Q-plot adhere rather well to the theoretical quantiles and no trend of auto-correlation is visible.

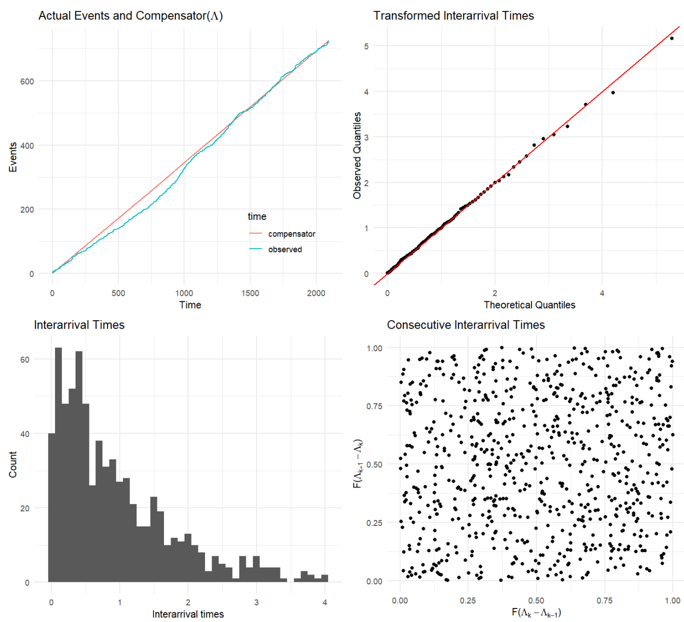

Figure 15: Transformed inter-arrival plots of the time invariant Stockholm model

KS and LB tests of the transformed inter-arrival times resulted in p-values of 0.5628 and 0.04455, respectively. Consequently, while the null hypothesis of the transformed inter-arrival times being exponentially distributed can not be rejected, there is a possibility of auto-correlation. The LB test is thus another indication that a time invariant Hawkes process is not an appropriate model for the Stockholm data.

Figure 16, 17, 18, on the following page, show sliding window estimations of the parameters with rolling quantiles. Note that as the difference between the minimum and maximum estimations were very large for some of the parameters, parts of the estimations are not visible.

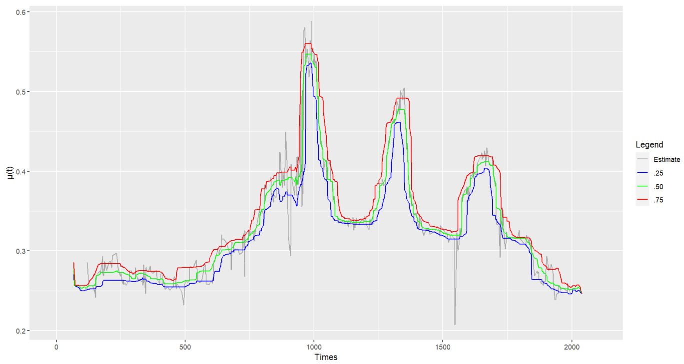

Figure 16: Rolling quantiles of sliding window estimates of μ from the Stockholm data

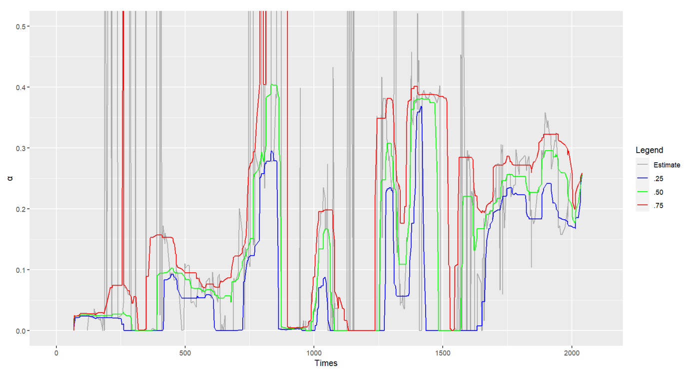

Figure 17: Rolling quantiles of sliding window estimates of α from the Stockholm data

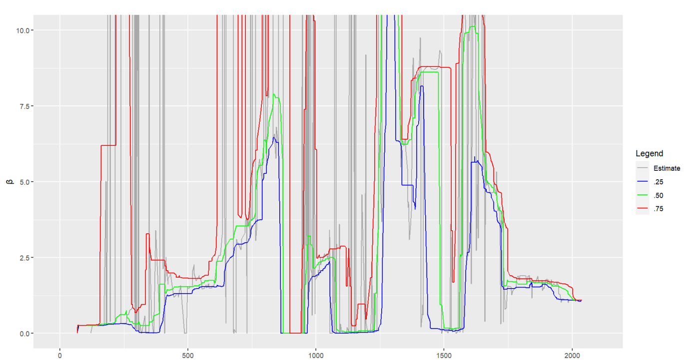

Figure 18: Rolling quantiles of sliding window estimates of β from the Stockholm data

Regarding the sliding window estimates, μ clearly shows signs of some sort of seasonal or cyclical trend. Estimations for both α and β are more erratic in comparison but also seem to possibly display some kind of cyclical pattern, particularly in the case of α

Overall, these results indicate that the temporal Hawkes process might not be an appropriate model for the Stockholm data.

Väst Turning now to Väst, we once again begin by evaluating whether the arrival times might follow a simple Poisson process. Figure 19 show that, as with the Stockholm data, it is not clear that the inter-arrival times are not exponentially distributed, meaning that it is difficult to reject the hypothesis of a simple Poisson process from just graphical analysis alone. However, unlike the Stockholm data there is a prominent gap in the bottom left of the figure.

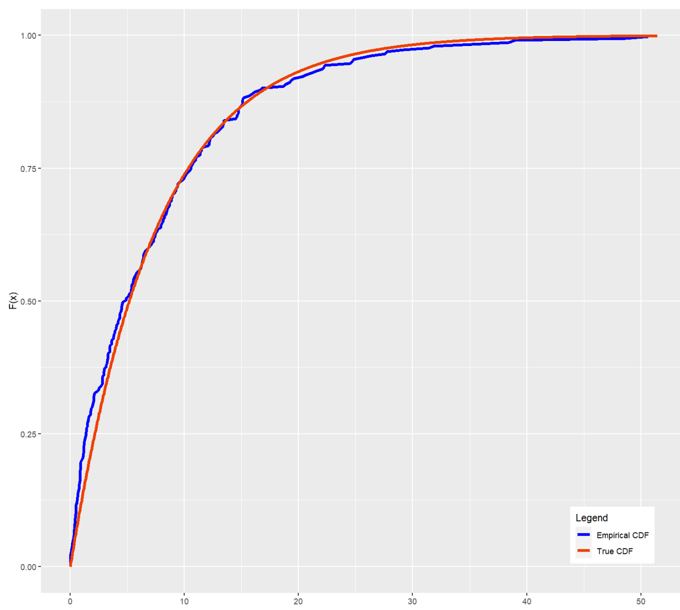

Figure 19: ECDF vs Hypothetical exponential CDF for Väst inter-arrival times

KS and LB tests resulted in p-values of < 0.0001 and 0.459, respectively. While the KS test indicates that one should reject the hypothesis of exponentially distributed inter-arrival times, it is contrary to Stockholm not possible to reject the null hypothesis of them being independent.

Parameter estimations for a time invariant Hawkes model are described in Table 6, on the next page.

Table 6: Parameter estimations for the Väst region

	Estimate	Std. Error
μ	0.132014	0.008041673
α	0.582396	0.439390797
β	29.679715	20.111051228

As with Stockholm, the standard errors for α and β are once again large enough to result in insignificant parameter estimations at the 5% level.

Plots for goodness-of-fit are shown in Figure 20, below. The actual events seems to generally be above what one would expect from the compensator. Furthermore, unlike Stockholm, the upper observed quantiles have a tendency to drift away from the expected. However, there is still no sign of auto-correlation visible.

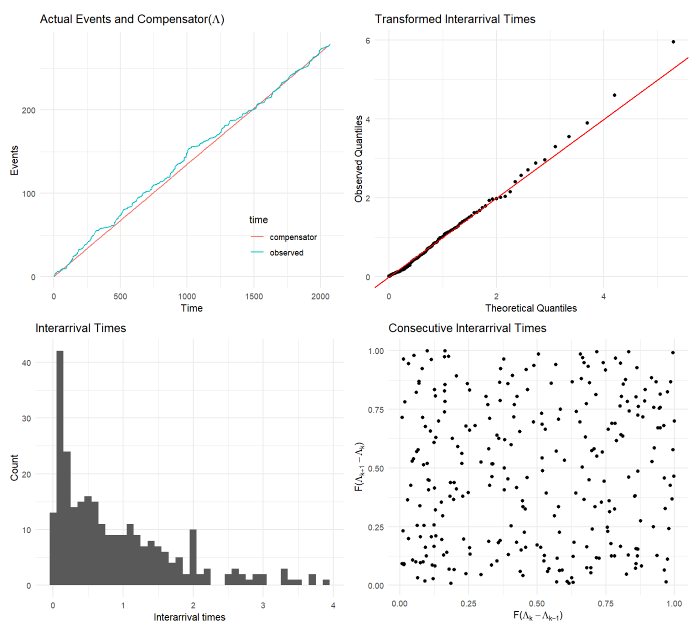

Figure 20: Transformed inter-arrival plots of the time invariant Väst model

KS and LB tests on the transformed inter-arrival times resulted in p-values of 0.087 and 0.462, respectively. In contrast to Stockholm, neither null hypothesis can thus be rejected. It is however notable that the test results are the opposite of those of Stockholm where the larger p-value was that of the KS test.

Sliding window estimates for the parameters are shown in Figure 21, 22 and 23, on the following page.

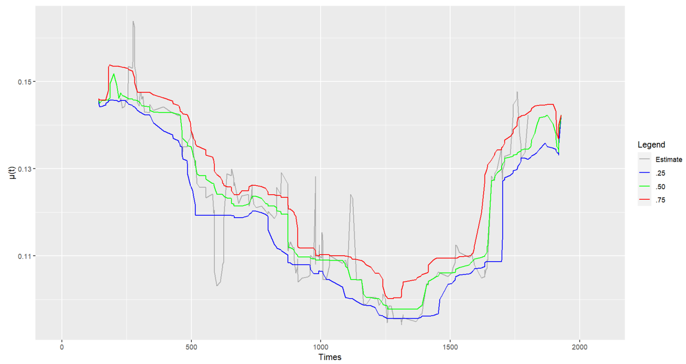

Figure 21: Rolling quantiles of sliding window estimates of μ from the Väst data

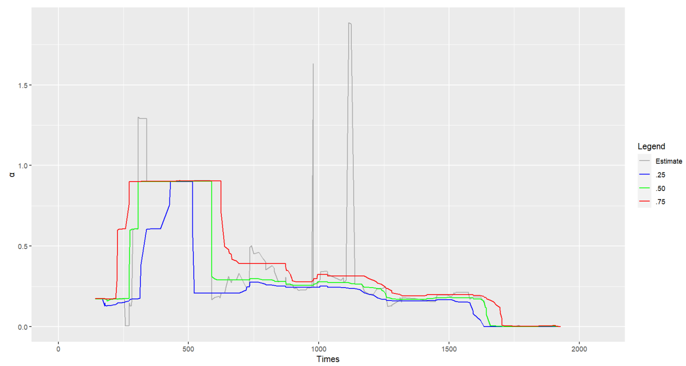

Figure 22: Rolling quantiles of sliding window estimates of α from the Väst data

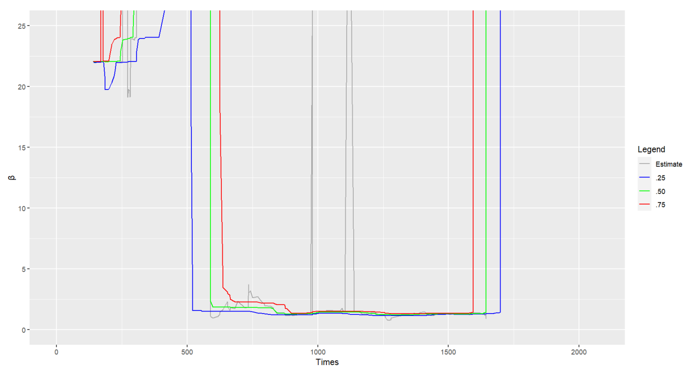

Figure 23: Rolling quantiles of sliding window estimates of β from the Väst data

In contrast to Stockholm, μ does not show signs of following a cyclical trend during the observed time period. There is, however, clearly some kind of trend, with the intensity following a concave shape. Estimates for α and β are not as erratic as for Stockholm, with α in particular looking rather constant after day 500. β, however, shows large spikes in the beginning and end of the observation period.

Overall, although the model has less issues than that of Stockholm, the results still indicate that a temporal Hawkes model might not be suitable to model the arrival times for Vast.

Syd Finally, Syd will be evaluated in the same way as Stockholm and Väst, starting by evaluating whether the inter-arrival times can be described by a simple Poisson process. Figure 24 below illustrates the ECDF and exponential CDF for the inter-arrival times. In contrast to Stockholm and Väst, the ECDF is clearly misaligned with the CDF providing a clear indication that the inter-arrival times are not exponentially distributed.

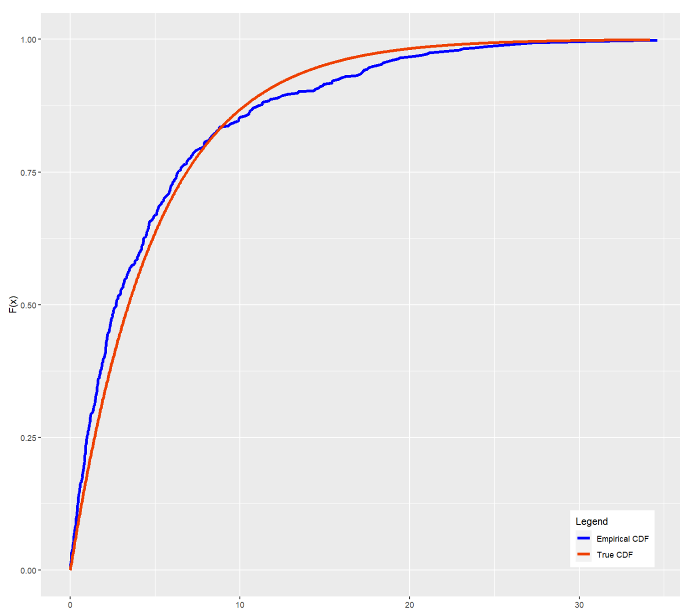

Figure 24: ECDF vs Hypothetical exponential CDF for Syd inter-arrival times

KS and LB tests resulted in p-values of < 0.0001 and 0.3316. We can therefore not reject the null hypothesis for the latter test, meaning that the possibility of independent inter-arrival times cannot be rejected.

Parameter estimations for a time-invariant Hawkes model are described in Table 7, on the next page.

Unlike Stockholm and Väst, the resulting standard errors are not as large relative to the coefficients themselves, resulting in all parameters being significant on a 5%-level.

Table 7: Parameter estimations for the Syd region

	Estimate	Std. Error
μ	0.1568974	0.01354471
α	0.1033079	0.03058870
β	0.4575279	0.15969960

Goodness-of-fit plots are shown in Figure 25, below. Actual Events and Compensator show a gap in the middle but seem to otherwise follow each other rather well. Similarly, theoretical and observed quantiles seem to mostly align, although there is a visible bump around the third quantile.

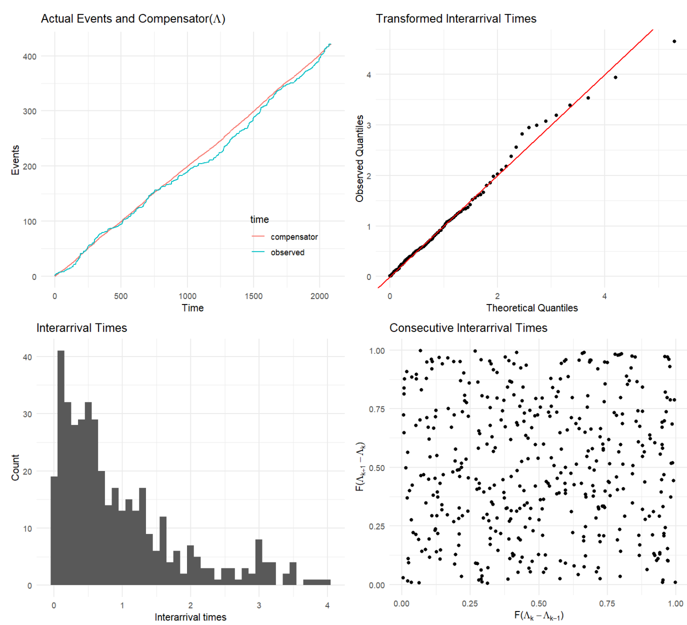

Figure 25: Transformed inter-arrival plots of the time invariant Syd model

KS and LB tests on the transformed inter-arrival times resulted in p-values of 0.896 and 0.909, respectively. The null hypothesis can thus not be rejected for either test, meaning that the inter-arrival times may very well be independent and exponentially distributed. We thus have indications that the Hawkes process might be appropriate.

Sliding window estimates are displayed in Figure 26, 27 and 28, on the following page.

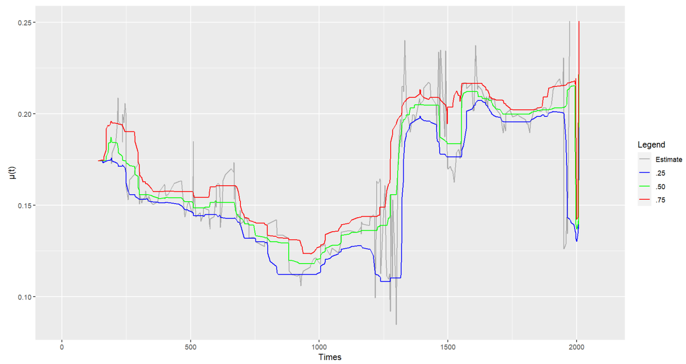

Figure 26: Rolling quantiles of sliding window estimates of μ from the Syd data

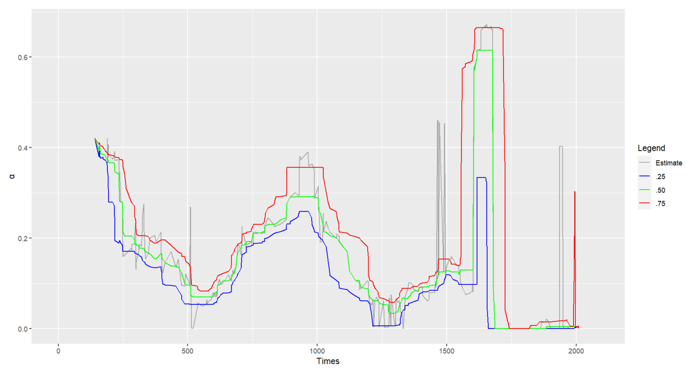

Figure 27: Rolling quantiles of sliding window estimates of α from the Syd data

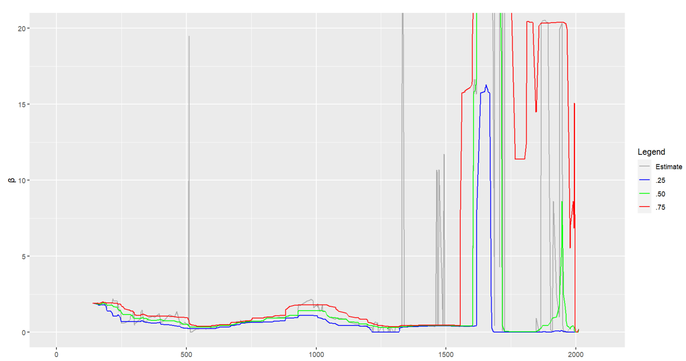

Figure 28: Rolling quantiles of sliding window estimates of β from the Syd data

What particularly stands out is the large jump in the estimate for μ around day 1250, given that such a jump was not visible in the yearly histogram in

Figure 4. Furthermore, there are also indications that μ has some cyclical tendencies, but it is hard to discern. Interestingly, however, a cyclical trend is clearly visible in α, and to a lesser extent, β.

Overall, in comparison to Stockholm and Väst the temporal Hawkes model seems to be a more suitable model for Syd. However, the sliding window estimates showed clear indications that the parameters varied with time.

Taken together, the results for the three regions show that although all regions showed some indications that their arrival times did not follow a simple Poisson process, it is not necessarily the case that a temporal Hawkes model is the most suitable model. Both the Stockholm and Väst region had insignificant parameter estimates for α and β at the 5%-level , which would add credence to the simple Poisson process. There were also problems with goodness-of-fit for both regions. Syd, however, showed more promise, with significant parameters and generally good results on goodness-of-fit plots and tests. However, as sliding window estimates showed indications of time variant parameters for all regions, one would ideally want to evaluate a inhomogeneous model as well before making a verdict on the temporal Hawkes as a potential model.

4 Discussion

As shown in the previous chapter, there were notable differences in the appropriateness of the temporal Hawkes process between the regions. In this chapter the potential factors behind this, as well limitations of the model will be discussed. Suggestions on how the model could be improved will also be given.

The present study showed that there were large differences in the parameter estimations between the different police regions. This in itself is not necessarily notable, given that the regions vary significantly in their populations. However, the differences were not limited to the magnitude of the parameters themselves. Where Stockholm and Syd had ECDF's that did not notably differ from the CDF of an exponential distribution, Syd showed clear abbreviations. Similarly, while both Stockholm and Väst suffered from insignificant parameter estimations for α and β, Syd did not. Given that Väst had the fewest shootings overall, it is possible that the excitation factor α is very small and it could therefore be theorized that the problems with the model for Väst stem from a small population size. However, it is more difficult to argue the same for Stockholm. While it is in theory possible that the shootings in Stockholm simply do not exhibit self-exciting tendencies, given that near-repeat effects are well documented (including in Stockholm), it does not seem very realistic. One problematic factor could be the potential of edge effects affecting the model. Specifically, the fact that the processes have been estimated as if they were regionally independent. Given that it is not necessarily the case that the processes can be neatly delimited into the same geographical area as the police regions, there is a possibility that there are shootings missing from the estimation procedure that ought to have been included. This risk would appear to be particularly high in the Stockholm region, given that it solely includes the city of Stockholm, excluding other cities within commuter distance such as Södertälje and Uppsala. If there is some covariance with Copenhagen, one could also surmise that Syd would have similar problems. Furthermore, one cannot exclude the case that occurrences could affect each other even in non-nearby Police regions. Consequently, to ensure correct specification of the processes one would ideally incorporate detailed criminological knowledge of the networks and agents involved.

Another potential cause of errors is the estimation procedure, both due to the estimation procedure itself and due to the added noise. According to Laub, Lee & Taimre (2021:43), the maximum-likelihood estimation procedure is rather sensitive for smaller samples, with many local optima. This was visible in the simulation study in 3.4.1 even for the continuous process, where although the mean bias was quite low, individual estimates could greatly differ from the true value. Furthermore, by utilizing noise with a $Uni(0, 1)$ distribution, we operate under an assumption that the probability of a shooting occurring is unaffected by the time of day. While I was unable to locate studies of the average daily distribution of shootings in Sweden, studies from the US show that shootings generally have a higher probability of occurring from late night until early mornings, and are rarer during the middle of the day (Klerman et al. 2023).

Moving on to potential model improvements, large improvements could likely be made by including spatial information. As discussed in 2.4, the occurrences of crime tend to display not just temporal, but spatial patterns as well, such as the aforementioned near-repeat effects and hot-spots. Indeed, all the studies on crime data discussed in 2.4 utilized some type of spatio-temporal model. By neglecting the spatial component and modelling the shooting occurrences as purely a temporal process, it is thus very likely that important factors have been ignored. An improved model would therefore estimate the parameters of a spatio-temporal point process model while possibly also include spatial covariates. In addition, it is reasonable to suspect that shootings resulting in injuries and deaths could have a higher probability of causing retaliating shootings. Given this, one could also potentially improve the model by including any resulting causalities as marks. Furthermore, given that trends and cyclical patterns were clearly present for all regions one would ideally take this into account into the estimation procedure

Another potential limitation that is overall more difficult to handle is the difficulty of estimating the process when the total amount of occurrences are small. Foreign studies such as Mohler 2014 and Reinhart 2018 discussed in 2.4 were conducted on data sets with tens of thousands of occurrences. In contrast, Stockholm, the region with the largest amount of occurrences, had just 725 occurrences in total. Furthermore it is questionable whether it is proper to view the whole observation period as a single process. In Figure 2, 3 and 4 it was clear that the total number of shootings varied between different years. . Moreover, as shootings are not a natural stochastic process but rather the results of deliberate actions from criminal agents, their motivations and strategies changing could have important ramifications to the process as a whole. Overall, it is clear that only evaluating the events by purely statistical methods without including relevant criminological factors means that there is a possibility of the process not being completely understood.

5 Conclusion

The result of the study showed that although there are indications that shooting occurrences in Sweden could be modeled using a Hawkes process, a purely temporal homogeneous process is not adequate for any of the three regions included in the present study. Given that occurrences of crime follow many spatial patterns, it is not feasible to attempt to model the process without factoring that in. Furthermore, ideally one would also attempt to incorporate criminological factors in order to avoid miss-specifying the process.

Hawkes processes are one important toolbox in statistical analysis of crime and has already shown to be able to generate improved hot-spot maps and evaluate intervention efforts. It is thus a promising new tool in crime prevention and intervention. However, in the case of Sweden there is a dearth of research into not just the application of Hawkes processes on crime data but quantitative statistical analysis of crime overall. Given that gun violence is a salient issue in Sweden it would be pertinent to ensure that the phenomena is better understood from a statistical standpoint. Furthermore, given the great abundance of registry data in Sweden, potential Swedish studies have the potential to not just repeat results from the US but to also improve the model in novel ways. For this reason, more research into Hawkes process models of shootings (and possibly other crimes as well) in Sweden could have the potential to not just improve the understanding of the statistical underpinnings of crime in Sweden, but to generate new knowledge of statistical modelling of crime as a whole.

Bibliography

References:

Adamopoulos, L. (1976). ‘Cluster models for earthquakes: Regional comparisons’, *Journal of the International Association for Mathematical Geology*, 8, pp. 463-475.

Bas, E. (2019). *Basics of probability and stochastic processes*. Cham: Springer.

Boyd, P. & Molyneux, J. (2021). ‘Assessing the contagiousness of mass shootings with nonparametric Hawkes processes´, *PLoS One*, 16(3).

Braverman, V. (2016). ‘Sliding Window Algorithms´, in Ming-Yang, K. (ed.) *Encyclopedia of Algorithms*. 2nd edn. New York: Springer, pp. 2006–2011.

Cheysson, F & Lang, G. (2022). ‘Spectral estimation of Hawkes processes from count data´, *Annals of Statistics*, 50 (3), pp. 1722-1746.

Cox, D. R., & Isham, V. (1980). *Point processes*. London: Chapman and Hall.

Farrell, G. (1995). ‘Preventing Repeat Victimization´, *Crime and Justice*, *19*, pp. 469–534.

Godoy, B. I., Solo, V., Min, J., & Pasha, S. A. (2016). ‘Local likelihood estimation of time-variant hawkes models´, *2016 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, Shanghai, 20-25 March, IEEE, pp. 4199-4203.

Gorr, W.L. & Lee, Y. (2015). ‘Early Warning System for Temporary Crime Hot Spots´, *J Quant Criminol* 31, pp. 25–47.

Gut, A. (2009). *An Intermediate Course in Probability*. New York: Springer, pp. 221-275.

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

Klerman, E. B., Affouf, M., Robbins, R., Iyer, J. M., Griggs, C., Masiakos, P. T., & Sacks, C. A. (2023). ‘Characterizing gun violence by time, day of the week, holidays, and month in 6 US cities, 2015-2021´, *Journal of Biological Rhythms*.

Lantmäteriet. (No date). *SWEREF 99*. Available at: https://www.lantmateriet.se/en/geodata/gps-geodesi-och-swepos/reference-systems/three-dimensional-systems/SWEREF-99/ (Accessed: 1 January 2024).

Laub, P. J., Taimre, T., & Pollett, P. K. (2015). ‘Hawkes processes´. [Preprint]. Available at: https://arxiv.org/pdf/1507.02822.pdf (Accessed: 1 January 2024).

Laub, P. J., Lee, Y., & Taimre, T. (2021). *The elements of Hawkes processes*. Cham: Springer.

Ljung, G. M., & Box, G. E. P. (1978). ‘On a Measure of Lack of Fit in Time Series Models´, *Biometrika*, *65*(2), pp. 297–303.

Massey, F. J. (1951). ‘The Kolmogorov-Smirnov Test for Goodness of Fit´, *Journal of the American Statistical Association*, *46*(253), pp. 68–78.

Meyer, S., Elias, J., & Höhle, M. (2012). ‘A space–time conditional intensity model for invasive meningococcal disease occurrence´, *Biometrics*, *68*(2), pp. 607-616.

Millar, R. B. (2011). *Maximum likelihood estimation and inference: with examples in R, SAS and ADMB*. John Wiley & Sons.

Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., & Tita, G. E. (2011). 'Self-exciting point process modeling of crime', *Journal of the American Statistical Association*, 106(493), pp. 100-108.

Mohler, G. (2014). 'Marked point process hotspot maps for homicide and gun crime prediction in Chicago´. *International Journal of Forecasting*, 30(3), pp. 491-497.

Ogata, Y. (1988). 'Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes´, *Journal of the American Statistical Association*, 83(401), pp. 9–27.

Ogata, Y. (1998). 'Space-time point-process models for earthquake occurrences´, *Annals of the Institute of Statistical Mathematics*, 50, pp. 379-402.

Ogata, Y. (1999). 'Seismicity analysis through point-process modeling: A review´, *Seismicity patterns, their statistical significance and physical meaning*, pp. 471-507.

Park, J., Schoenberg, F. P., Bertozzi, A. L., & Brantingham, P. J. (2021). 'Investigating clustering and violence interruption in gang-related violent crime data using spatial–temporal point processes with covariates´, *Journal of the American Statistical Association*, 116(536), pp. 1674-1687.

Polisen. (No date-a). *Sprängningar och skjutningar - polisens arbete* [Shootings in Sweden - the work of the police]. Available at: https://polisen.se/om-polisen/polisens-arbete/sprangningar-och-skjutningar/ (Accessed: 1 January 2024).

Polisen. (No date-b). *Polismyndighetens organisation* [The organization of the police department]. Available at: https://polisen.se/om-polisen/organisation/ (Accessed: 17 January 2024).

Ratcliffe, J. & Rengert, G. (2008). 'Near-Repeat Patterns in Philadelphia Shootings´, *Secur J*, 21, pp. 58–76.

Reinhart, A. (2018). 'A Review of Self-Exciting Spatio-Temporal Point Processes and Their Applications´, *Statistical Science*, 33(3), pp. 299–318.

Reinhart, A. & Greenhouse, J. (2018). 'Self-exciting point processes with spatial covariates: modelling the dynamics of crime', *Journal of the Royal Statistical Society Series C: Applied Statistics*, 67(5), pp. 1305-1329.

Ross, S. M. (2014). *Introduction to probability models*. 10th edn. Academic press.

Sturup, J., Rostami, A., Gerell, M. & Sandholm, A. (2017). 'Near-repeat shootings in contemporary Sweden 2011 to 2015´, *Secur J*, 31, pp. 73–92.

Sturup, J., Gerell, M., & Rostami, A. (2020). 'Explosive violence: A near-repeat study of hand grenade detonations and shootings in urban Sweden´, *European Journal of Criminology*, 17(5), pp. 661-677.

Sveriges Television [SVT]. (No date). *Skjutningar i Sverige år för år* [Shootings in Sweden by year]. Available at: https://www.svt.se/datajournalistik/skjutningar-i-sverige-ar-for-ar/ (Accessed: 1 January 2024).

Townsley, M., Homel, R., & Chaseling, J. (2003). ‘INFECTIOUS BURGLARIES: A Test of the Near Repeat Hypothesis´. *The British Journal of Criminology*, 43(3), pp. 615–633.

R:

Auguie, B. (2017). *gridExtra: Miscellaneous Functions for "Grid" Graphics*, R package version 2.3.
Available at: https://CRAN.R-project.org/package=gridExtra.

Cheysson, F. (2021). *hawkesbow: Estimation of Hawkes Processes from Binned Observation*, R package version 1.0.2. Available at: https://CRAN.R-project.org/package=hawkesbow.

Grolemund, G. & Wickham, H. (2011). 'Dates and Times Made Easy with lubridate´, *Journal of Statistical Software*, 40(3), pp. 1–25.

Jones-Tod, C.M. & van Helsdingen, A. (2023). *stelfi: Hawkes and Log-Gaussian Cox Point Processes Using Template Model Builder*, R package version 1.0.1. Available at: https://CRAN.R-project.org/package=stelfi.

Millard, S.P. (2013). *EnvStats: An R Package for Environmental Statistics*. New York: Springer.

Pedersen, T.L. (2023). *patchwork: The Composer of Plots*, R package version 1.1.3. Available at: https://CRAN.R-project.org/package=patchwork.

R Core Team. (2016). *R: A Language and Environment for Statistical Computing*, Vienna, Austria. Available at: https://www.R-project.org/.

Tuszynski, J. (2021). *caTools: Tools: Moving Window Statistics, GIF, Base64, ROC AUC, etc*, R package version 1.18.2. Available at: https://CRAN.R-project.org/package=caTools.

Vaughan, D. (2023). *slider: Sliding Window Functions*, R package version 0.3.1. Available at: https://CRAN.R-project.org/package=slider.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*, New York: Springer.

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4. Available at: https://github.com/tidyverse/dplyr.

Wickham, H., Hester, J. & Bryan, J. (2023).*readr: Read Rectangular Text Data*, R package version 2.1.4.
Available at: https://CRAN.R-project.org/package=readr.

Zeileis, A. & Grothendieck, G. (2005). 'zoo: S3 Infrastructure for Regular and Irregular Time Series´, *Journal of Statistical Software*, 14(6), pp. 1–27.