Modeling market activity using 1D non-homogeneous Hawkes Processes

Authors:

Eskil Andersson
tfy11ean@student.lu.se

Lunds Tekniska Högskola

Supervisor: Magnus Wiktorsson
Examinator: Erik Lindström
Date: December 3, 2017

Abstract

This paper can be seen as a light introduction to the study of Hawkes processes and their applicability in the realms of finance. In particular, this paper is concerned on the topic of modeling market activity and elaborates on how Hawkes processes are superior to non-homogeneous Poisson processes in this regard. After some rudimentary theory on point processes it goes more in depth into the above mentioned processes and their likelihood estimators. The rest of the paper is dedicated to the actual modeling procedure, its necessary preparations and results, delving into the possible real-world interpretations of what the modeling tells us.

Acknowledgements

This work was done in cooperation with the Department of Mathematical Statistics at Lund University with the intent to probe the topic of Hawkes processes and their applications to finance. I would like to thank my supervisor Magnus Wiktorsson for dedicating his time to tutor me through this process and for not giving up on me. I would also like to thank my family for always encouraging me and being there when their support was needed the most. Lastly, to all the people in my vicinity who make life enjoyable: Thank you.

Contents

Introduction

Identifying patterns in data is in itself very fundamental, and it is an integral part of the scientific machinery that has enabled the technological advancements of modern society. And at the heart of our modern society—at the very core—there is money. If society is a living organism, then our monetary system is its cardiovascular system *and* nervous system in one. It is the intricate plumbing which feeds our society and sustains it, through every malaise and every hiccup. It binds all things; like a force of nature it interacts between entities of all forms. It permeates everything: from the smallest household to the largest corporation and mightiest government. Money is—effectively—the means to all ends.

And where there is money, there is a market. And where there are stocks, there is the Stock Market; whose physical embodiment is the Stock Exchange. Tasked with pumping money and credit around the system, it plays a vital role in our society. It is not so strange then, that people would want to be able to foresee price movement in stocks by identifying patterns in trading-data—the incentives are clearly there. Having said that, let's be clear about what this thesis *is not*. It is not an attempt to foresee price-movements in stocks; or any kind of arbitrage-seeking tool for real-time analysis.

What it is though, is an attempt to see if activity in stock-market transactions can be modeled well using Hawkes processes and if we can get even better models by making the exogenous input time-dependent. Furthermore, we'll investigate if there are any weekly trading-patterns in our data-set as we look for weekday-discrepancies in the parameters.

Starting with some basic definitions for point processes, we then move on to focus on the two types of specific point processes that will be used in this thesis: namely, Poisson- and Hawkes processes. After that, we look at the data-set obtained for this thesis and briefly discuss its limitations. Then, plots galore follows as we take a look at the results from our models. Lastly, some contemplation occurs with respect to the obtained results: potential improvements and reasonable ways forward are discussed. With that said—let's dive into the theory!

Theory

2.1 Point processes

A Point process is basically just a bunch of mathematical points randomly located on an underlying mathematical space. It is a very useful tool for modeling and analyzing spatial data—which in our case is time-events on the positive real line.

Without getting too formal, we'll start off by defining a slightly more general form of a point process than the specific cases we'll use in the actual modeling. So, the point processes used in this paper are built on Definition 1–5 as follows. [5]

Definition 1. *Let* $\big(\Omega, \mathscr{F}, \mathbb{P}\big)$ *be some probability space and* $\{t_i\}_{i \in \mathbb{N}^+}$ *be a sequence of non-negative random variables such that* $\forall i \in \mathbb{N}^+$, $t_i < t_{i+1}$.
Then, $\{t_i\}_{i \in \mathbb{N}^+}$ *is called a simple point process.* $\square$

Definition 2. *Let* $\{t_i\}_{i \in \mathbb{N}^+}$ *be a point process. The right-continuous process*

$$\mathcal{N}(t) = \sum_{i \in \mathbb{N}^+} \mathbb{1}_{t_i \leq t}$$

is called the counting process associated with $\{t_i\}_{i \in \mathbb{N}^+}$. $\square$

Definition 3. *The process* $\{\delta t_i\}_{i \in \mathbb{N}^+}$ *where*

$$\forall i \in \mathbb{N}^+,\, \delta t_i = t_i - t_{i-1}$$

is called the duration process associated with $\{t_i\}_{i \in \mathbb{N}^+}$. $\square$

Definition 4. *Let $\mathcal{N}$ be a point process adapted to a filtration $\mathscr{F}_t$ where $t \in \mathbb{R}^+$. The left-continuous intensity process is defined as*

$$\lambda\big(t|\mathscr{F}_t\big) = \lim_{h\downarrow 0} \mathbb{E}\bigg(\frac{\mathcal{N}(t+h) - \mathcal{N}(t)}{h} \,\bigg|\, \mathscr{F}_t\bigg)$$

which is equivalent to

$$\lambda\big(t|\mathscr{F}_t\big) = \lim_{h\downarrow 0} \frac{1}{h}\, \mathbb{P}\Big(\mathcal{N}(t+h) - \mathcal{N}(t) > 0 \,\Big|\, \mathscr{F}_t\Big).$$

□

More specifically, we will only be looking at a slightly less general form of Definition 4 in this thesis:

$$\left.\begin{aligned} &\lim_{h\downarrow 0} \frac{1}{h}\, \mathbb{P}\big(\mathcal{N}(t+h) - \mathcal{N}(t) = 1 \,\big|\, \mathscr{F}_t\big) = \lambda(t|\mathscr{F}_t) \\ &\lim_{h\downarrow 0} \frac{1}{h}\, \mathbb{P}\big(\mathcal{N}(t+h) - \mathcal{N}(t) \geq 2 \,\big|\, \mathscr{F}_t\big) = 0 \end{aligned}\right\} \tag{2.1}$$

That is, two (or more) events cannot happen at the same time.

Definition 5. *The function Λ is defined as:*

$$\Lambda(t_{i-1}, t_i) = \int_{t_{i-1}}^{t_i} \lambda(s|\mathscr{F}_s)\ \mathrm{d}s\,,\ \forall i \in \mathbb{N}^+$$

and is called the integrated intensity function (also known as the compensator of the point process). □

To round things off, the following theorem will be very useful to us a bit further down the road:

Time-rescaling Theorem. *Let $0 < t_1 < \cdots < t_n < T$ be a realization from a point process with conditional intensity function $\lambda(t|\mathscr{F}_t)$ satisfying $0 < \lambda(t|\mathscr{F}_t)\ \forall t \in (0, T]$. Define the transformation*

$$\Lambda(t_k) = \int_0^{t_k} \lambda(s|\mathscr{F}_s)\ \mathrm{d}s$$

for $k = 1, 2, \ldots, n$ and assume $\mathbb{P}\big(\Lambda(t) < \infty\big) = 1\ \forall t \in (0, T]$. Then the $\Lambda(t_k)$'s are a Poisson process with unit rate (i.e. $\lambda = 1$).[1] □

In this paper, the filtration will always be assumed to be the natural filtration of the process; i.e., $\mathscr{F}_t = \mathscr{F}_t^{\mathcal{N}}$. Furthermore, for notational simplicity, this filtration will always be implied; i.e., $\lambda(t) = \lambda(t|\mathscr{F}_t)$.

With all that out of the way, let's take a closer look at the *actual* processes we will compare throughout this paper.

2.1.1 Poisson processes

Hopefully, the reader will be somewhat familiar with Poisson point processes already; so we will just cover the basics here without getting too deep into proofs and technicalities. For a more comprehensive look at the subject, reading the referenced material is highly recommended.

There are several equivalent ways to define a Poisson process. Here is the one we will use: [2]

Definition 6. *A Poisson process* $\{\mathcal{P}(t)\}_{t\geq 0}$ *is a non-negative, integer-valued stochastic process such that* $\mathcal{P}(0) = 0$ *and*

(a) *the process has independent increments,*

(b) $\mathbb{P}(\mathcal{P}(t+h) - \mathcal{P}(t) = 1) = \lambda(t) + o(h)$ *as* $h \to 0$, $\quad \lambda(t) > 0$

(c) $\mathbb{P}(\mathcal{P}(t+h) - \mathcal{P}(t) \geq 2) = o(h)$ *as* $h \to 0$. □

In fact, Definition 6 is the definition of a non-homogeneous Poisson process—which is the only Poisson model we will use in this thesis. We ignore the regular (homogeneous) Poisson process on the basis of common sense. This will be apparent later when we take a closer look at the data-set. We will also motivate the shape of the intensity-function $\lambda(t)$ after looking at the data-set.

It is also worth to point out the following:

Definition 7. *If intervals* X *of a point process satisfy*

$$\mathbb{P}(X > s + t | X > t) = \mathbb{P}(X > s)$$

then the processes is memoryless. Poisson processes are memoryless. [10] □

The characteristic of memorylessness comes from the fact that the duration process of a Poisson process is exponentially distributed. [2] [10]

Maximum likelihood function

The maximum likelihood function of a non-homogeneous Poisson process,

$$\mathcal{L}^{\mathcal{P}} = \mathcal{L}\big(\vartheta \bigm| \{\mathcal{P}_t\}_{t\in[0,T]}\big),$$

is given by [10]

$$\mathcal{L}^{\mathcal{P}} = \prod_{i=1}^{n} \lambda(t_i|\vartheta)\, \mathrm{e}^{-\int_0^T \lambda(t|\vartheta)\ \mathrm{d}t}$$

which means that the log-likelihood is given by

$$\ln \mathcal{L}^{\mathcal{P}} = \sum_{i=1}^{n} \ln\big(\lambda(t_i|\vartheta)\big) - \int_0^T \lambda(t|\vartheta)\ \mathrm{d}t \tag{2.2}$$

2.1.2 Hawkes processes

Hawkes processes belong to a class of point processes called self-exciting. Intuitively, one could say that a process is self-exciting when past events make future events more possible. To put it a bit more formally:

Definition 8. *A point process $\mathcal{N}$ is called self-exciting if*

$$\mathbb{C}\big(\mathcal{N}(s,t), \mathcal{N}(t,u)\big) > 0 \quad \textit{for } s < t < u$$

where $\mathbb{C}(\cdot\,,\cdot)$ denotes the covariance function and $\mathcal{N}(s,t) = \mathcal{N}(t) - \mathcal{N}(s)$. □

The thing we are interested in however is the intensity of the self-exciting process. To see that, look no further than Definition 9.

Definition 9. *The intensity of a general self-exciting process $\mathcal{N}$ is defined as:*

$$\begin{aligned}\lambda(t) &= \mu(t) + \int_{-\infty}^{t} \eta(t-s)\ \mathrm{d}\mathcal{N}_s \\ &= \mu(t) + \sum_{t_i<t} \eta(t-t_i)\end{aligned}$$

where $\mu : \mathbb{R} \mapsto \mathbb{R}^+$ is a deterministic base intensity and $\eta : \mathbb{R}^+ \mapsto \mathbb{R}^+$ is the self-exciting kernel which expresses the positive influence of past events t_i on the current value of the intensity process. □

It should be quite obvious from Definitions 8 & 9 that self-exciting processes *do not* have independent increments and they are *not* memoryless like the Poisson process is (since the whole history of past events affects future events). This is very good if we anticipate some kind of clustering-behaviour in the data-set that we want to catch in our models (which we do).

To be more specific—as the title of this section hints at, we will use a certain type of self-exciting process, called a Hawkes process (see Definition 10).

Definition 10. *As suggested by Hawkes*[3]*, a Hawkes process uses the exponential kernel* $\eta(t) = \sum_{\jmath=1}^{\varpi} \alpha_\jmath \mathrm{e}^{-\beta_\jmath t} \mathbb{1}_{\mathbb{R}^+}$ *, which makes the intensity of the model:*

$$\begin{aligned}\lambda(t) &= \mu(t) + \int_0^t \sum_{\jmath=1}^{\varpi} \alpha_\jmath \mathrm{e}^{-\beta_\jmath (t-s)} \, \mathrm{d}\mathcal{N}_s \\ &= \mu(t) + \sum_{t_i<t} \sum_{\jmath=1}^{\varpi} \alpha_\jmath \mathrm{e}^{-\beta_\jmath (t-t_i)} \qquad (2.3)\end{aligned}$$

□

For this thesis, we will only use $\varpi = 1$; so (2.3) reduces to

$$\lambda(t) = \mu(t) + \sum_{t<t_i} \alpha \, \mathrm{e}^{-\beta(t-t_i)} \qquad (2.4)$$

where $\alpha, \beta > 0$ and $t \geq 0$.

If we are to get a more intuitive understanding of (2.4), we can look at $\mu(t)$ as the exogenous part of the intensity—a sort of base intensity representing behavioural patterns *outside* the model itself—and the sum in the end as the endogenous (self-exciting) part. The choice of whether to keep the exogenous part constant or time-dependent will depend on how the data-set looks (more on that later).

Stationarity

Let's have a quick reminder of what stationarity implies:

Definition 11. *For a point process* $\mathcal{N}$*, if*

$$\mathbb{E}(\lambda(t)) = \kappa \ \textit{(const.)}$$

and

$$\mathbb{C}(\mathcal{N}_s, \mathcal{N}_t) < \infty$$

and

$$\mathbb{C}(\mathcal{N}_s, \mathcal{N}_t) = \mathbb{C}(\mathcal{N}_{s-t}, 0) \equiv \mathbb{C}(\mathcal{N}_\tau) \,, \quad \tau = s - t,$$

then the process is called weakly stationary (or covariance stationary). Every strictly stationary process with finite variance is also weakly stationary.[4] □

Let's not dabble in the definition of strict stationarity; assuming weak stationarity will be quite sufficient for our purposes.

If stationarity is assumed for a general self-exciting process and we keep the exogenous parameter μ constant, using Definition 9, we get:

$$\begin{aligned}
\kappa = \mathbb{E}\big(\lambda(t)\big) &= \mathbb{E}\bigg(\mu + \int_{-\infty}^{t} \eta(t-s)\ \mathrm{d}\mathcal{N}_s\bigg) \\
&= \mu + \mathbb{E}\bigg(\int_{-\infty}^{t} \eta(t-s)\lambda(s)\ \mathrm{d}s\bigg) \\
&= \mu + \int_{-\infty}^{t} \eta(t-s)\kappa\ \mathrm{d}s \\
&= \mu + \kappa \int_{0}^{\infty} \eta(r)\ \mathrm{d}r \\
&\Updownarrow \\
\kappa &= \frac{\mu}{1 - \int_0^\infty \eta(r)\ \mathrm{d}r}
\end{aligned} \tag{2.5}$$

where κ is a constant.

For a general 1D-Hawkes process (with constant μ) we have the following stationarity condition:[5]

$$\sum_{\jmath=1}^{\varpi} \frac{\alpha_\jmath}{\beta_\jmath} < 1. \tag{2.6}$$

Combining (2.5) and (2.6), with $\varpi = 1$, gives us the average intensity of our simple 1D-Hawkes process:

$$\mathbb{E}\big(\lambda(t)\big) = \frac{\mu}{1 - {}^{\alpha}\!/_{\beta}}\ , \quad \frac{\alpha}{\beta} < 1. \tag{2.7}$$

Maximum likelihood function

Since we are only interested in the simple ($\varpi = 1$) one-dimensional case, I will omit the calculations for the general 1D-Hawkes process and go straight for the simple version for the sake of brevity and clarity. The calculations for the general case is very straight forward anyway. So, before we get into the log-likelihood function, we need to do some calculations of the integrated intensity function Λ:

$$\begin{aligned}
\Lambda(t_{i-1}, t_i) &= \int_{t_{i-1}}^{t_i} \lambda(s)\ \mathrm{d}s \\
&= \int_{t_{i-1}}^{t_i} \mu(s)\ \mathrm{d}s + \int_{t_{i-1}}^{t_i} \sum_{t_k < s} \alpha\, \mathrm{e}^{-\beta(s - t_k)}\ \mathrm{d}s \\
&= \int_{t_{i-1}}^{t_i} \mu(s)\ \mathrm{d}s + \int_{t_{i-1}}^{t_i} \sum_{t_k \leq t_{i-1}} \alpha\, \mathrm{e}^{-\beta(s - t_k)}\ \mathrm{d}s \\
&= \int_{t_{i-1}}^{t_i} \mu(s)\ \mathrm{d}s + \sum_{t_k \leq t_{i-1}} \frac{\alpha}{\beta} \Big(\mathrm{e}^{-\beta(t_{i-1} - t_k)} - \mathrm{e}^{-\beta(t_i - t_k)} \Big) \qquad (2.8)
\end{aligned}$$

If we take a closer look at the summation-terms in (2.8), there is a recursion there that can be used to simplify the expression (and, more importantly, the computations):

$$\begin{aligned}
\mho(i-1) &= \sum_{t_k \leq t_{i-1}} \mathrm{e}^{-\beta(t_{i-1} - t_k)} \\
&= 1 + \mathrm{e}^{-\beta(t_{i-1} - t_{i-2})} \sum_{t_k \leq t_{i-2}} \mathrm{e}^{-\beta(t_{i-2} - t_k)} \\
&= 1 + \mathrm{e}^{-\beta(t_{i-1} - t_{i-2})} \mho(i-2). \qquad (2.9)
\end{aligned}$$

Using the expression from (2.9) in (2.8) gives us the final expression for the integrated intensity:

$$\Lambda(t_{i-1}, t_i) = \int_{t_{i-1}}^{t_i} \mu(s)\ \mathrm{d}s + \frac{\alpha}{\beta} \Big(1 - \mathrm{e}^{-\beta(t_i - t_{i-1})} \Big) \mho(i-1)\,, \quad \forall i \in \mathbb{N}^+ \qquad (2.10)$$

where $\mho(0) = 0$.[5]

The log-likelihood of a simple (general) point process $\mathcal{N}$ with intensity λ is given by[9]:

$$\ln \mathcal{L}\big(\vartheta \;\big|\; \{\mathcal{N}_t\}_{t \in [0,T]}\big) = \int_0^T \big(1 - \lambda(s)\big)\ \mathrm{d}s + \int_0^T \ln \lambda(s)\ \mathrm{d}\mathcal{N}_s. \qquad (2.11)$$

By applying (2.11) to our simple 1D-Hawkes process, $\mathcal{H}$, we get[8]:

$$\begin{aligned}
\ln \mathcal{L}\big(\vartheta \;\big|\; \{\mathcal{H}_t\}_{t \in [0,T]}\big) &= T - \Lambda(0, T) + \sum_{i=1}^{n} \ln \lambda(t_i) \\
&= T - \Lambda(0, T) + \sum_{i=1}^{n} \ln \left(\mu(t_i) + \sum_{k=1}^{i-1} \alpha\, \mathrm{e}^{-\beta(t_i - t_k)} \right) \qquad (2.12)
\end{aligned}$$

where n is the number of events.

Once again, if we look at the last term of (2.12) we'll find a recursive pattern that will greatly simplify computations[7]:

$$
\begin{aligned}
\Upsilon(i) &= \sum_{k=1}^{i-1} \mathrm{e}^{-\beta(t_i - t_k)} \\
&= \mathrm{e}^{-\beta(t_i - t_{i-1})} \sum_{k=1}^{i-1} \mathrm{e}^{-\beta(t_{i-1} - t_k)} \\
&= \mathrm{e}^{-\beta(t_i - t_{i-1})} \left(1 + \sum_{k=1}^{i-2} \mathrm{e}^{-\beta(t_{i-1} - t_k)} \right) \\
&= \mathrm{e}^{-\beta(t_i - t_{i-1})} \big(1 + \Upsilon(i-1) \big). \qquad (2.13)
\end{aligned}
$$

So, inserting (2.13) and (2.9) into (2.12) gives us our final (recursively computable) expression of the log-likelihood function:

$$
\ln \mathcal{L}^{\mathcal{H}} = T - \int_0^T \mu(s)\, \mathrm{d}s - \sum_{i=1}^{n} \left(\frac{\alpha}{\beta} \left(1 - \mathrm{e}^{-\beta(t_n - t_i)} \right) + \ln \big(\mu(t_i) + \alpha \Upsilon(i) \big) \right) \quad (2.14)
$$

with $\Upsilon(1) = 0$ and of course $\mathcal{L}^{\mathcal{H}} = \mathcal{L}\big(\vartheta \mid \{\mathcal{H}_t\}_{t \in [0,T]}\big)$.

If the exogenous input parameter μ is constant then the log-likelihood of the process, denoted as $\underline{\mathcal{H}}$, simplifies to

$$
\ln \mathcal{L}^{\underline{\mathcal{H}}} = T(1 - \mu) - \sum_{i=1}^{n} \left(\frac{\alpha}{\beta} \left(1 - \mathrm{e}^{-\beta(t_n - t_i)} \right) + \ln \big(\mu + \alpha \Upsilon(i) \big) \right) \qquad (2.15)
$$

It is worth noting that this estimator has the following properties:

Definition 12. *For a stationary 1D-Hawkes process with constant μ and $\varpi = 1$, the maximum likelihood estimator $\hat{\vartheta}(\hat{\mu}, \hat{\alpha}, \hat{\beta})$ is consistent, asymptotically normal and asymptotically efficient.[6]* □

where the properties of consistency, asymptotic normality and -efficiency are defined as follows (Definition 13–15):

Definition 13. *If it is true that*

$$
\lim_{T \to \infty} \mathbb{P}\Big(\big| \hat{\vartheta} - \vartheta \big| > \varepsilon \Big) = 0, \quad \forall \varepsilon > 0
$$

then $\hat{\vartheta}$ is consistent.[5] □

Definition 14. *If*

$$\sqrt{T}\Big(\hat{\vartheta} - \vartheta\Big) \xrightarrow{\mathrm{d}} \mathbf{N}\big(0, \mathbb{I}^{-1}(\vartheta)\big), \quad \mathbb{I}^{-1}(\vartheta) = \left(\mathbb{E}\left(\frac{1}{\lambda}\frac{\partial\lambda}{\partial\vartheta_i}\frac{\partial\lambda}{\partial\vartheta_\jmath}\right)\right)_{i,\jmath}$$

then, ϑ is asymptotically normal. [5] □

Where $\mathbb{I}^{-1}$ is the inverse Fisher-information matrix.

Definition 15. *If $\hat{\vartheta}$ reaches the Cramér–Rao bound of the variance it is said to be asymptotically efficient.* [5] □

To wrap up the section on models, let's look at a summary of all the models' likelihood-functions:

Model summary:

$$\mathcal{P}: \quad \ln\mathcal{L} = \sum_{i=1}^{n} \ln\big(\lambda(t_i)\big) - \int_0^T \lambda(t)\ \mathrm{d}t$$

$$\underline{\mathcal{H}}: \quad \ln\mathcal{L} = T(1-\mu) - \sum_{i=1}^{n}\left(\frac{\alpha}{\beta}\Big(1 - \mathrm{e}^{-\beta(t_n - t_i)}\Big) + \ln\big(\mu + \alpha\Upsilon(i)\big)\right)$$

$$\mathcal{H}: \quad \ln\mathcal{L} = T - \int_0^T \mu(s)\,\mathrm{d}s - \sum_{i=1}^{n}\left(\tfrac{\alpha}{\beta}\Big(1 - \mathrm{e}^{-\beta(t_n - t_i)}\Big) + \ln\big(\mu(t_i) + \alpha\Upsilon(i)\big)\right)$$

2.2 Model validation

To be able to make judgments on which model out of a selected set of models is to be preferred (i.e. which one fits best to the data-set at hand) some kind of evaluation must take place. This is called testing a model's goodness of fit to the data. The path chosen here is to utilize one quantitative- and one qualitative approach to model validation; namely, the BIC and K-S test respectively.

2.2.1 Bayesian Information Criterion

There are many different ways in statistics to choose a preferred model by using a criterion. In this paper, this is done by comparing their respective BIC-values (Bayesian Information Criterion):

$$\mathrm{BIC} = \ln(n)k - 2\ln\hat{\mathcal{L}},$$

where n is the number of data-points, k is the number of free parameters to be estimated, and $\hat{\mathcal{L}} = \mathcal{L}(\hat{\vartheta})$, that is the maximized value of the likelihood function.

The first term is a penalizing term meant to alleviate the risk of choosing an overfitted model over a more generally correct model while the other term is basically the negative log-likelihood, whose value (relative to another models log-likelihood on the same data-set) is an indication of how good the model is in relation to the data at hand.

2.2.2 Kolmogorov-Smirnov test

On the subject of goodness-of-fit tests, there are two different techniques which boils down to essentially the same plot: Q-Q (quantile-quantile) plots and K-S (Kolmogorov-Smirnov) tests. Out of personal preference, the K-S test was chosen.

The K-S test falls into the category of non-parametric tests and can be used to compare a sample with a reference probability distribution. The way this is done in our case is by first taking the durations of the compensator

$$\tau_k = \Lambda(t_k) - \Lambda(t_{k-1}), \quad k = 1, 2, \ldots, n$$

where $\forall \tau_k \overset{\text{i.i.d.}}{\sim} \mathbf{Exp}(1)$ according to the time-rescaling theorem; then, we make the transformation

$$z_k = 1 - \mathrm{e}^{-\tau_k}$$

which makes $\forall z_k \overset{\text{i.i.d.}}{\sim} \mathbf{U}(0,1)$. After that, we sort the transformations

$$z_k \mapsto z_{(i)}, \quad z_{(1)} < z_{(2)} < \cdots < z_{(n)}$$

and plot them against the values of the c.d.f. of a standard uniform distribution denoted as

$$\upsilon_k = \frac{k - \frac{1}{2}}{n}.$$

Thus, we have created the graph $\left(\upsilon_k, z_{(k)}\right)$ which represents the K-S plot. If the distribution of the sample is the same as the reference distribution, the dots should be relatively close to the line $x = y$ on $[0,1] \times [0,1]$. For moderate to large samples, the approximation for the 95 % confidence bounds are given by $\upsilon \pm 1.36/\sqrt{n}$.[1]

Data-set

In this section, we will be looking at the data-set at our disposal—and with the acquired inference, we will then motivate the choice of time-dependent behaviour for our intensities.

3.1 Processing the data

The set contains stock-market data for shares of Volvo B from 2003-01-02 to 2004-11-26 (a total of 477 trading days). All in all, there where 429 132 separate events in the set initially; but, due to mathematical constraints in the modeling, some events had to be scrubbed.

After removing trading activity outside of regular bourse-hours, canceled orders, and OTC (Over The Counter) trades, it ended up with 422 804 events remaining. The result of these actions can be seen in Figure 3.1.

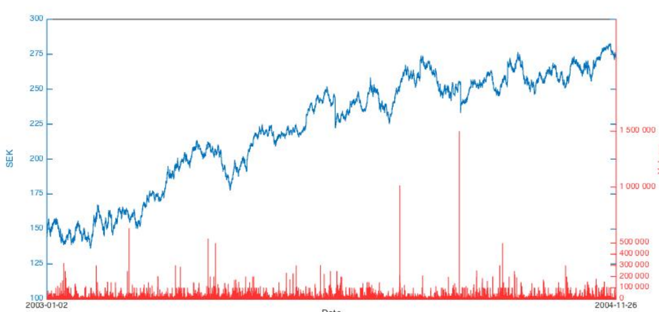

Figure 3.1: This graph shows the price and volume of all the events in the data-set that will be used for modeling. The price is the the amount paid per share of Volvo B-stock at each instance of trading and the volume represents the amount of shares trading hands at each purchase.

Furthermore, some processing of the remaining data has to be done. Because of (2.1), we can only have at most one event happening at every moment in time; however, the resolution of the time-stamps is only [1 s].

This means that all events occurring during the same one-second tick-window will be registered as having happened at the same time. Sure, we are talking about the early 2000's here—HFT (High-Frequency Trading) didn't really exist yet—but there are still *a lot* of events registered at the same time: making modeling impossible.

To get around the problem of incomputable log-likelihood functions, we need to scramble the time-stamps a little. This is achieved by adding a stochastic time $\varepsilon_i \sim \mathbf{U}(0,1)$ to event t_i for all i (where the unit of t_i is [s]) and then sorting each day individually with respect to time.

To put it more clearly:

Algorithm: Shake 'n' Sort. Denote the set of our data as $\mathcal{S}$. Then, $\mathcal{S}$ can be partitioned into a family of disjoint indexed subsets $\mathcal{S}_{\ell\in[1,D]}$ where D is the number of days we have data on; i.e.,

$$\mathcal{S} = \bigcup_{\ell=1}^{D} \mathcal{S}_\ell \text{ and } \bigcap_{\ell=1}^{D} \mathcal{S}_\ell = \varnothing,$$

where $\mathcal{S}_\ell = \{t_i^{(\ell)} : i \in \mathcal{I}_\ell\}$ and $\mathcal{I}_\ell$ is the index-set for $\mathcal{S}_\ell$. The algorithm is then as follows:

1. $\mathcal{S}_\ell \to \mathcal{S}'_\ell : t_i^{(\ell)} \mapsto t_i^{(\ell)} + \varepsilon_i^{(\ell)} \quad \forall i, \ell \ \land \ \varepsilon \sim \mathbf{U}(0,1)$
2. $\mathcal{S}_\ell \leftarrow \underset{t}{\operatorname{sort}}\big(\mathcal{S}'_\ell\big) \quad \forall \ell$

I realize that the notation in Algorithm 1 (i.e., "Shake 'n' Sort") *might* be seen as a little sloppy since, technically, $\mathcal{S}$ contains "event-objects"—let's denote them $\mathcal{E}$—in the form of tuples (whose size could differ depending how true one would want to stay to the raw data-file).

If trade-types and other (for our purposes) non-essential information is ignored, a more proper notation would be:

$$\mathcal{S} = \big\{\mathcal{E}_i : \mathcal{E} = \big(\langle\text{date}\rangle, \langle\text{time}\rangle, \langle\text{volume}\rangle, \langle\text{price}\rangle\big)\,,\, i \in \mathcal{I}\big\}$$

where $\mathcal{I}$ is the index-set for the entire set $\mathcal{S}$, and

$$\mathcal{S}_\ell = \Big\{\mathcal{E}_i^{(\ell)} : \Big(\forall \mathcal{E}_i^{(\ell)}, \mathcal{E}_\jmath^{(\ell)}, i \neq \jmath\Big)\big[\langle\text{date}\rangle_i = \langle\text{date}\rangle_\jmath\big] \,:\, i, \jmath \in \mathcal{I}_\ell \land \ell \in [1, D]\Big\}.$$

However, in this paper we are only interested in the date and time of an event; and the date is implied in the partition. So, the time-value is really the only thing we are interested in. Therefore, the notation was made to reflect that fact and, of course, to keep it as simple as possible.

Anyway—with the data being clean and compatible with our modeling efforts, it is a good time to take a closer look at possible behavioural patterns inherent in the data.

3.2 Choosing the exogenous intensity function

How does one motivate the choice of shape for an exogenous time-dependent intensity? The obvious choice is plotting a histogram of all events in $\mathcal{S}$. This can be seen in Figure 3.2.

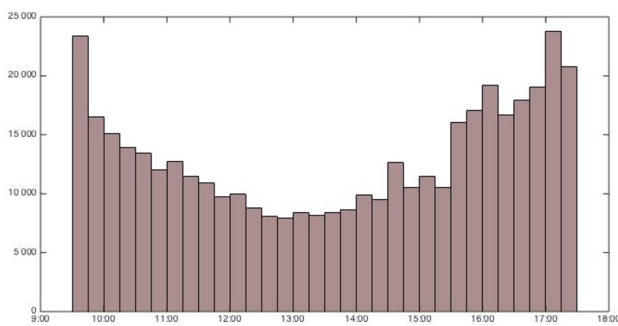

Figure 3.2: The histogram shows the time-distribution of all events in $\mathcal{S}$. Each bin represents the number of events happening inside its 15 min time-frame.

It does not take a lot of imagination to see that the intensity seems to be following some kind of convex second-order curve. To be on the safe side, let's look at each trading-day of the week separately to see if this is ubiquitous behaviour or some cumulative effect.

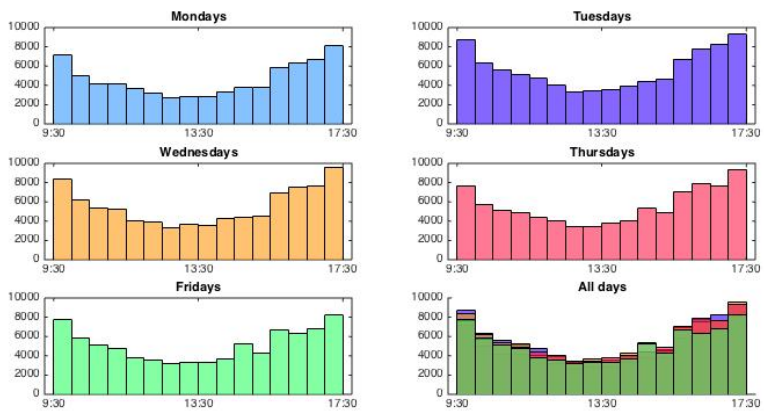

Figure 3.3: The histograms show the time-distribution of all events in $\mathcal{S}$ divided into separate week-days. Each bin represents the number of events happening inside its 30 min time-frame.

Looking at Figure 3.3, it looks like the pattern stays consistent no matter what week-day it is. This should give us enough confidence to make an educated guess as to how the modeled exogenous intensity should look like:

Conjecture 1. *Let's set the exogenous time-dependent intensities of our non-homogeneous Poisson process and non-homogeneous Hawkes process, $\mathcal{P}$ and $\mathcal{H}$ respectively, so that they follow a convex second-order polynomial whose minimum is located in the upper positive half-plane of its graph; i.e.,*

$$f : t \mapsto \zeta t^2 + \varphi t + \xi \quad \textit{such that} \begin{cases} \zeta & > 0 \\ \varphi & < 0 \\ \xi & \geq \frac{\varphi^2}{4\zeta} \end{cases}, \quad t \in [0, T],$$

where $t = 0$ and $t = T$ represents the beginning and end of the trading-day respectively. □

Perhaps the reasoning in Conjecture 1 should be a bit more explicit:

$$\begin{aligned} f(t) &= \zeta t^2 + \varphi t + \xi \\ f'(t) &= 2\zeta t + \varphi \qquad \implies f'(t) = 0 \iff t = -\frac{\varphi}{2\zeta}, \\ & \qquad\qquad\qquad t, \zeta > 0 \implies \varphi < 0 \\ f\Big(-\frac{\varphi}{2\zeta}\Big) \geq 0 &\iff \frac{\varphi^2}{4\zeta} - \frac{\varphi^2}{2\zeta} + \xi \iff \xi \geq \frac{\varphi}{4\zeta}. \end{aligned}$$

So—with that being said, now would be a good time to do some actual modeling. This will be done in the following chapter.

Results

4.1 Daily parameter estimates

So, there are several ways this can be presented. The approach taken here is to first look at each day in $\mathcal{S}$ (i.e., $\mathcal{S}_\ell \; \forall \ell \in [1, D]$) as *separate* processes to get a qualitative feeling for how the parameters of the different models fluctuate from day to day. These results can be seen in Figure 4.4–4.6.

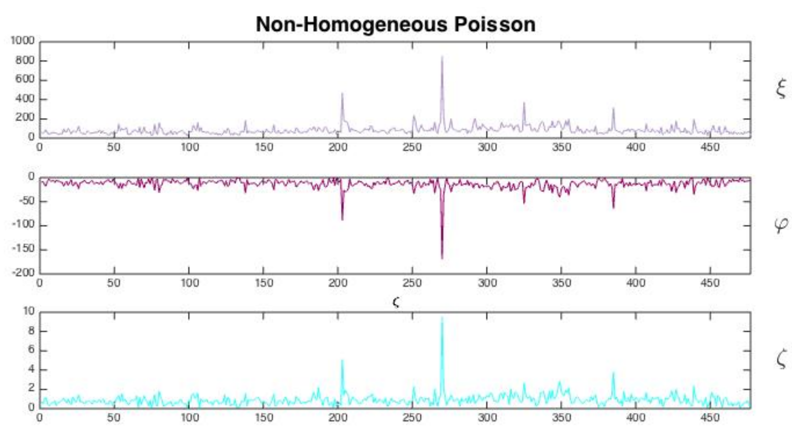

Figure 4.4: These are the results of fitting the data to non-homogeneous Poisson processes, $\mathcal{P}$, by using the maximum likelihood estimator $\mathcal{L}^{\mathcal{P}}$. (See (2.2))

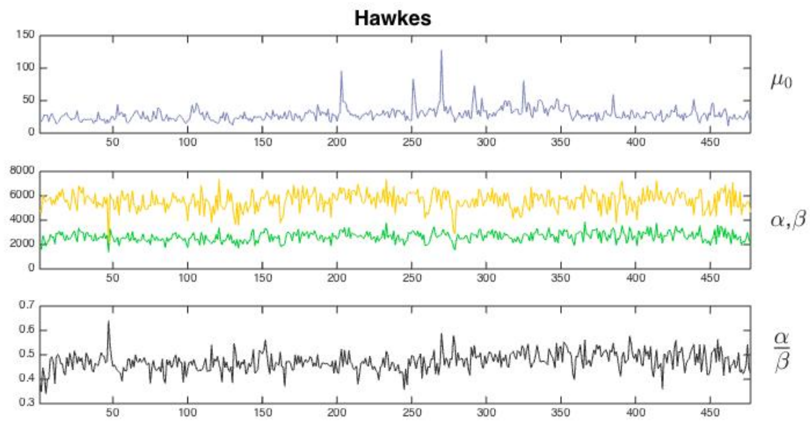

Figure 4.5: These are the results of fitting the data to Hawkes processes, $\underline{\mathcal{H}}$, by using the maximum likelihood estimator $\mathcal{L}^{\underline{\mathcal{H}}}$. (See (2.15))

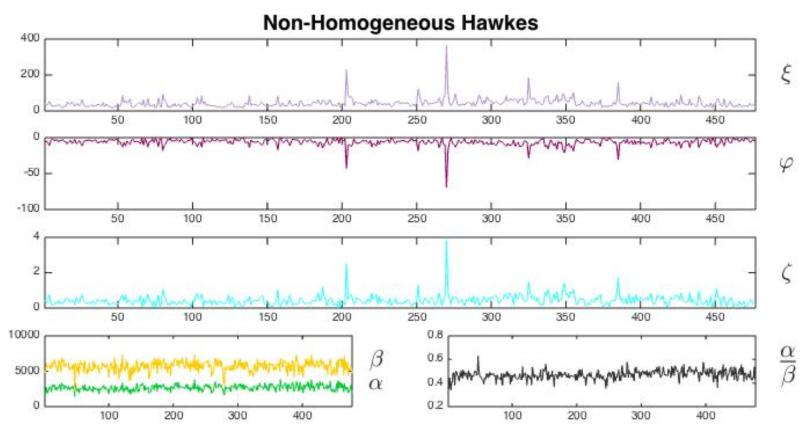

Figure 4.6: These are the results of fitting the data to non-homogeneous Hawkes processes, $\mathcal{H}$, by using the maximum likelihood estimator $\mathcal{L}^{\mathcal{H}}$. (See (2.14))

Simply, what has been done in Figure 4.4–4.6 is finding the set of parameters $\hat{\vartheta}$ that maximizes the likelihood function (or, what *really* has been done is minimizing the negative log-likelihood function; but that equates to the same thing); i.e.,

$$\hat{\vartheta} = \arg\max_{\vartheta} \mathcal{L} = \arg\max_{\vartheta} \ln \mathcal{L} = \arg\min_{\vartheta} -\ln \mathcal{L}$$

There is really no point in trying to get any meaning out of the values of the parameters by just looking at them; the values themselves are not interesting. It should be noted also that, due to numerical reasons, all event times had to be re-scaled for the modeling. The chosen scaling was:

$$t : [0, 28800] \mapsto [0, 16] \qquad \Longrightarrow \qquad \lambda : [(1\,\mathrm{s})^{-1}] \mapsto [(30\,\mathrm{min})^{-1}],$$

i.e., the unit of the intensities went from 'per second' to 'per 30 min'. This scaling is kept throughout the rest of the paper.

If you are interested in how the intensity for a Hawkes process looks like, take a look at Figure 4.7.

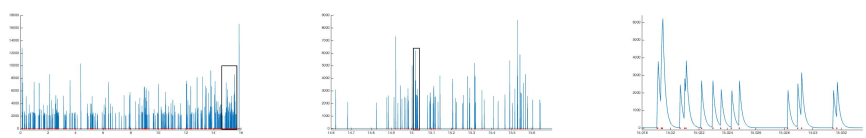

Figure 4.7: These plots show the intensity function $\lambda(t|\hat{\vartheta})$ of a Hawkes process $\underline{\mathcal{H}}$ where the left picture shows how the intensity varies over an entire day (chosen at random) and every step to the right is a zoomed in picture of the box in the previous plot.

To get a better view of how the intensity behaves when a lot of events are happening in a very small window of time, see Figure 4.8

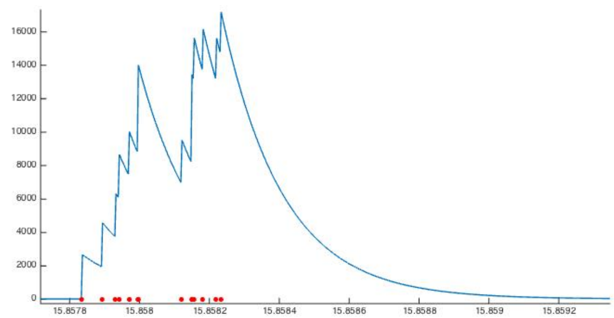

Figure 4.8: A close-up look at the intensity $\lambda(t|\hat{\vartheta})$ of a Hawkes process $\underline{\mathcal{H}}$ during extreme clustering.

As fun as it is to look at intensity functions, the interesting part is comparing the models. These comparisons, done by BIC, can be seen in Figure 4.9.

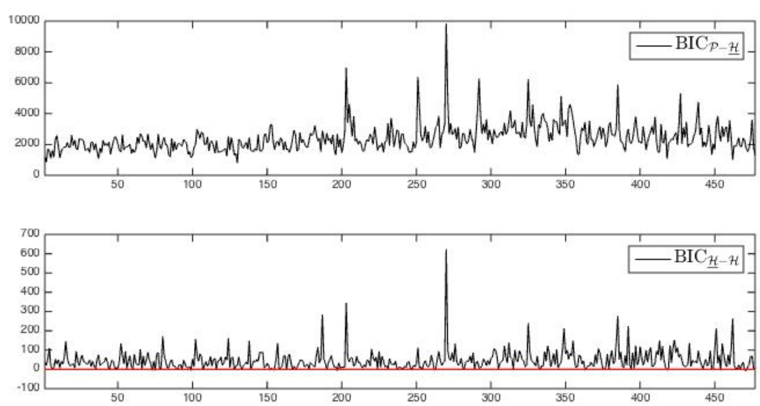

Figure 4.9: These figures are comparing the BIC-values for individual days for $\mathcal{P}$ to $\underline{\mathcal{H}}$ at the top and for $\underline{\mathcal{H}}$ to $\mathcal{H}$ at the bottom.

To clarify any possible confusion that might arise from Figure 4.9,

$$\left.\begin{array}{l} \text{BIC}_{\mathcal{P}-\underline{\mathcal{H}}}(\ell) = \text{BIC}_{\mathcal{P}}(\ell) - \text{BIC}_{\underline{\mathcal{H}}}(\ell) \\ \text{BIC}_{\underline{\mathcal{H}}-\mathcal{H}}(\ell) = \text{BIC}_{\underline{\mathcal{H}}}(\ell) - \text{BIC}_{\mathcal{H}}(\ell) \end{array}\right\} \quad \forall \ell \in [1, D],$$

and the smallest BIC is always preferred; meaning that when $\text{BIC}_{\mathcal{P}-\underline{\mathcal{H}}}$ is positive, $\underline{\mathcal{H}}$ is preferred over $\mathcal{P}$, etc.

For obvious reasons, the plot of $\text{BIC}_{\mathcal{P}-\mathcal{H}}$ was considered superfluous and were therefore omitted; but all tallies can be seen in Table 4.1.

	$\mathcal{P}$	$\underline{\mathcal{H}}$	$\mathcal{H}$
$\mathcal{P}$	—	0	0
$\underline{\mathcal{H}}$	477	—	27
$\mathcal{H}$	477	450	—

Table 4.1: This is the final tallies of the BIC-comparisons, were each row is compared to each column.

Before we move on, it is easily noted that some relatively extreme spikes exist in Figure 4.9. To address this, Figure 4.10 will be used as a source of discussion in the following chapter.

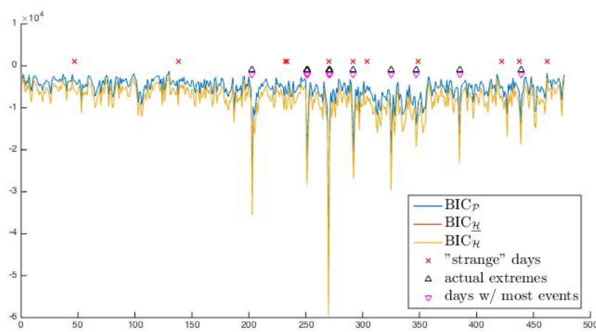

Figure 4.10: This is the actual BIC-values for all individual days plotted in the same plot along with some markers. See discussion in the next chapter for further elaboration.

4.2 Estimating one set of parameters

Seeing every new day as a new and completely fresh process is a nice way to get an intuitive feeling for the volatility of the parameters between consecutive days. However, we really want to view each new day as the *same* process repeating over and over. For this purpose, the approach has to be slightly changed.

If we look at the likelihood under the assumption that all days are independent, we want to find the parameters that maximizes the sum off all individual days' likelihoods. This can (somewhat informally) be expressed as:

$$\breve{\vartheta} = \arg\max_{\vartheta} \ \mathcal{L}_{\Sigma} = \sum_{\ell \in [1,D]} \mathcal{L}(\vartheta \,|\, \mathcal{N}_t \in \mathcal{S}_\ell),$$

where $\mathcal{N}_t \in \mathcal{S}_\ell$ is the counting process for day ℓ and will vary depending on which model is used. The results of these calculations can be seen (compared with the mean of its corresponding individual likelihoods) in Table 4.2.

	ξ	φ	ζ
$\mathbb{E}(\hat{\vartheta}_{\mathcal{P}})$	83.4517	-13.1638	0.9054
$\breve{\vartheta}_{\mathcal{P}}$	83.5449	-13.1818	0.9060

	μ	α	β
$\mathbb{E}(\hat{\vartheta}_{\underline{\mathcal{H}}})$	29.0847	2645.8	5599.6
$\breve{\vartheta}_{\underline{\mathcal{H}}}$	28.9788	2639.5	5484.4

	ξ	φ	ζ	α	β
$\mathbb{E}(\hat{\vartheta}_{\mathcal{H}})$	43.1671	-6.3732	0.4397	2678.3	5705.3
$\breve{\vartheta}_{\mathcal{H}}$	42.6151	-6.2627	0.4336	2663.9	5561.3

Table 4.2: These tables show the mean parameter values for all individual days versus the parameters that maximizes the 'sum-likelihood' for all models: $\mathcal{P}$, $\underline{\mathcal{H}}$, and $\mathcal{H}$ respectively.

Once again, the values themselves are not that interesting; it's the BICs we want to compare to see which model fits best to the data. To see that, look at Table 4.3

$\mathbb{BIC}_{\mathcal{P}-\underline{\mathcal{H}}}$	$\mathbb{BIC}_{\mathcal{P}-\mathcal{H}}$	$\mathbb{BIC}_{\underline{\mathcal{H}}-\mathcal{H}}$
1 203 773	1 224 259	20 486

Table 4.3: This table shows the differences between the BICs using the 'sum-likelihoods' of the different models. As before, a positive value indicates that the latter subscript is preferred over the first subscript.

Once again, the notation in Table 4.3 was chosen with the aim of reducing as much clutter as possible; hopefully, it is clear enough.

As good as the BIC is at comparing different models against each other, it doesn't really tell us if our preferred model is an *actual* good fit to the data; it just tells us it is the best one out of the models we have chosen to investigate. To address this apparent flaw, the K-S plots of the different models are shown in Figure 4.11.

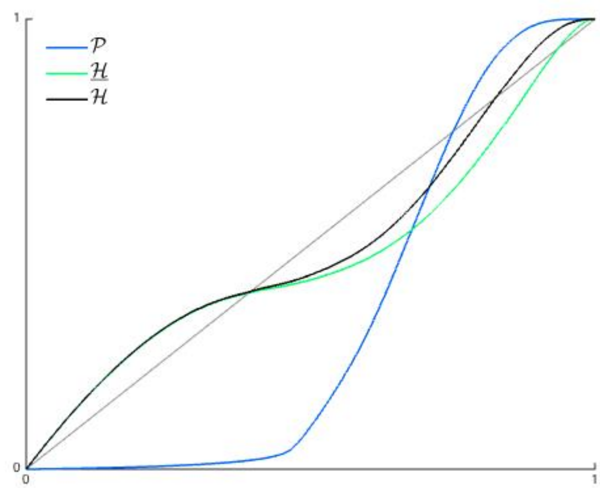

Figure 4.11: This plot shows the results of the K-S test on the different models for the entire data-set $\mathcal{S}$. Because of the amount of data, the 95 % confidence bounds are too small to be visible on this scale.

To see how $\breve{\vartheta}$ matches up on single days (so we have visible confidence bounds), a small selection is shown in Figure 4.12.

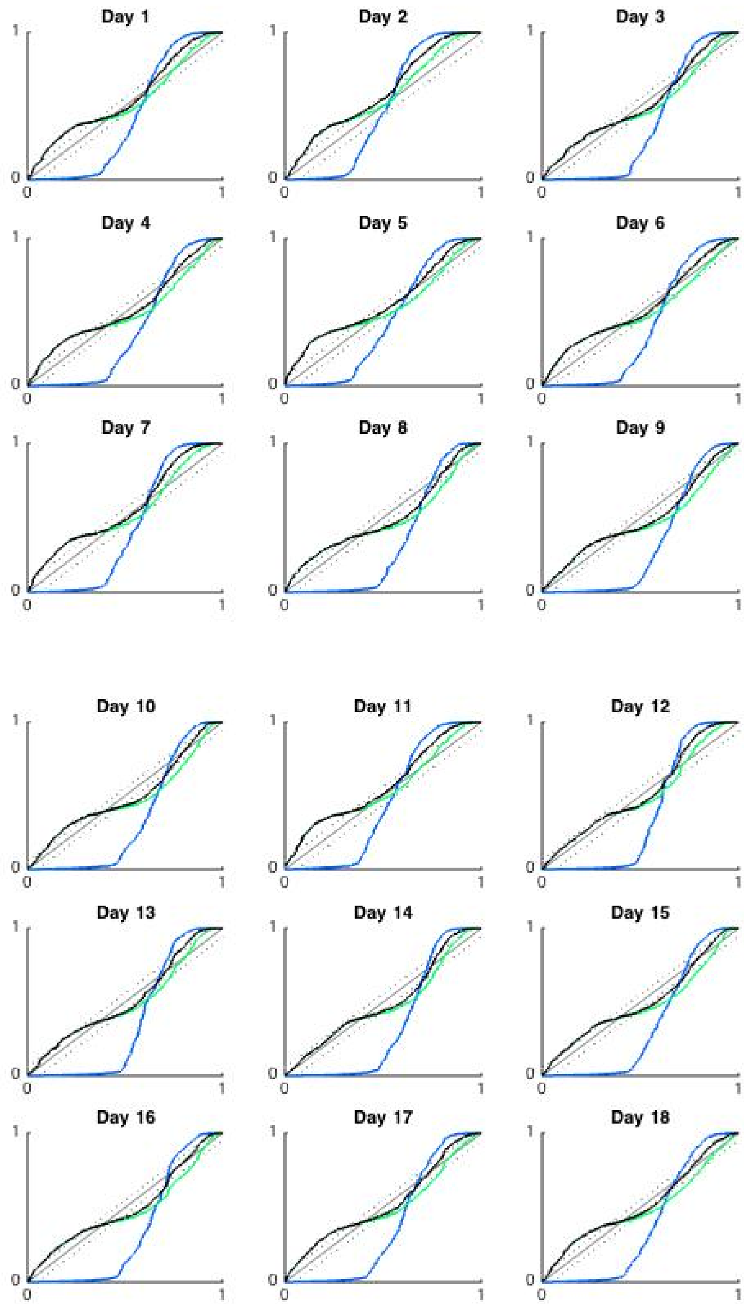

Figure 4.12: These plots are a selection of the eighteen first daily K-S plots for the different models with parameters $\breve{\vartheta}$. They are accompanied by the 95 % confidence interval.

4.3 Estimating parameters on a week-day basis

Finally, we will be looking at how the parameters $\breve{\vartheta}$ differ for different week-days; i.e., we make another partition of $\mathcal{S}$ into sets $\mathcal{S}_{w \in \mathcal{W}}$ where $\mathcal{S}_\ell \subseteq \mathcal{S}_w$ if and only if $\forall \mathcal{E}_i^{(\ell)} : \langle \text{week-day} \rangle_i = w$. As it is quite obvious from Table 4.3 that $\mathcal{H}$ is the best model for our data, we will only be doing these calculations for that model. The results can be seen in Table 4.4.

$\breve{\vartheta}$	ξ	φ	ζ	α	β
Mon	39.6825	-6.2597	0.4317	2694.6	5705.8
Tue	44.6943	-6.6212	0.4522	2630.0	5462.3
Wed	44.8655	-6.7265	0.4620	2710.4	5619.9
Thu	41.8077	-5.7824	0.4144	2678.8	5576.6
Fri	41.8895	-5.8884	0.4053	2611.4	5474.2
$\mathcal{H}$	42.6151	-6.2627	0.4336	2663.9	5561.3

Table 4.4: This table shows each parameter value for $\breve{\vartheta}_w$ where $w \in \mathcal{W}$ and $\mathcal{W}$ is the set of week-days. At the bottom, the parameter values that maximizes $\mathcal{L}_\Sigma^\mathcal{H}$ over the *entire* data-set $\mathcal{S}$ is included for comparison.

Actually, looking at Table 4.4, one starts to wonder if the approach of assigning different parameters to different week-days is a significant improvement of our previous modeling. This begs to be investigated.

To do this, a new model has to be defined—let's call it $\mathring{\mathcal{H}}$. Let the parameter-set be

$$\mathring{\vartheta} = \{\forall \breve{\vartheta}_w : w \in \mathcal{W}\}$$

where, of course, $|\mathring{\vartheta}| = 25$.

Since we assume that all days are independent, then so are the set of all different week-days; meaning that the log-likelihood $\ln \mathcal{L}^{\mathring{\mathcal{H}}}$ is just the sum of the log-likelihoods of the separate week-days. As the underlying data-set is the same for both models, the BIC can be used for evaluation. The result of these calculations is:

$\mathbb{BIC}_{\mathcal{H}-\mathring{\mathcal{H}}}$
555.44

i.e., $\mathring{\mathcal{H}}$ is indeed preferred over $\mathcal{H}$.

Lastly—to aid in the visualization of the information contained in Table 4.4, Figure 4.13 was produced.

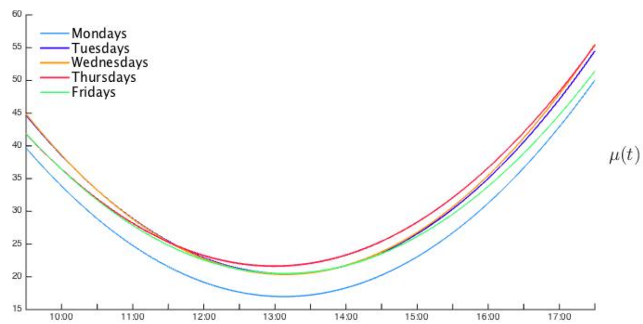

Figure 4.13: This plot shows how the estimated exogenous intensity $\breve{\mu}(t)$ varies over $[0, T]$ for each week-day.

Conclusion

At last, it is time to conclude our findings and discuss the results. First and foremost, it is very clear that the non-homogeneous Hawkes process $\mathcal{H}$ is vastly superior to the alternatives provided in this paper—looking at Table 4.3 confirms that. This shouldn't really come as a surprise to anyone. From the start, $\mathcal{P}$ was already virtually rejected as a viable model. If one thinks about what it implies, it basically says that all events on the stock-market (within one stock) are independent from one another; no herd- or reactionary behaviour whatsoever! Just the mere suggestion that that is a reasonable assumption is ludicrous. However, it does serve its purpose as a kind of benchmark—no matter how easy it is to beat. As it is non-homogeneous though, it does take into consideration that some external forces outside the model are somewhat responsible for the observed intensity of events on the underlying space. We see for example from Figure 3.2 that, generally, orders are arriving at a higher rate near the opening- and closing hours of the day; while being at its lowest around what most people consider reasonable lunch-hours. This is not a coincidence: this is expected from simple common sense. Without any endogenic behaviour at all present in the model however, it does fall flat like previously mentioned. It honestly never stood a chance.

The interesting part of comparing the models lies in looking at how the different Hawkes-models perform vis-à-vis each other. What we are really interested in is knowing whether the time-dependence of its exogenous input is significant enough to warrant two extra parameters in the model. As Table 4.3 loudly proclaims, this is indeed the case. It should however be noticed that Table 4.1 indicates that the non-homogeneity of the base intensity could potentially be seen as a tad superfluous on a few days in our data-set. But—the modeling approach used in that example was really just to get a qualitative feel of the data, and it is quite different from our preferred approach with the 'sum-likelihood'.

With that in mind, the non-homogeneous model almost didn't happen. The reason for that is because, initially, I had the underlying one-dimensional space for the models, $[0, T]$, scaled in seconds (i.e., $T = 28800$). For a domain of that size, mapping to a codomain in the form of a second-order polynomial with a relatively limited range (due to the fact that the intensity per second is very low for our data) puts a lot off pressure on the second-order term of the polynomial to get very small. Computers does not like to handle floats that are too small. Or at least, MATLAB's program for finding the minimum of constrained non-linear multivariable-functions was not able to find what it was designed to find. This led to *weeks* of scouring the internet in hopeless despair, trying to find an explanation for the optimization-debacle (at the time, the cause of the failure was not obvious). At first (after I had confirmed it was not due to a typo in the code, of course), I was completely convinced it was purely mathematical and that I was missing something. However—in time, it started to dawn on me that the size of ζ was indeed really small and that the nature of the problem was probably more practical than theoretical. Re-scaling the domain confirmed this. After that, things got more straight forward.

Going back to the Hawkes processes again: the premise of there being some self-exciting element to the data is a much more reasonable one than that of no dependency at all on recent events. People (and remember: this is the early 2000's we are talking about—algorithmic trading was not a thing yet) are reactionary by nature. I can't scientifically back that up; but I'm claiming it anyway. When people see stock-prices go up: they want to buy. If people see stock-prices fall: they want to sell (I am generalizing here, but bear with me). These sort of herd-like tendencies tend to exaggerate price-movements in stocks that would most likely not be as prominent had all participants acted oblivious of the activity taking place on the exchange. This is the very essence of self-exciting behaviour, and our intuition tells us that—for a model of this kind to be any good—it has to capture this behaviour somehow. This is exactly what the Hawkes process does, and if you throw in an exogenous part of the model taking in common daily human- and/or societal patterns, a reasonably realistic model should be a fait accompli! At least, that's the reasoning behind the choice of models; and as mentioned earlier, the exogenous part seems to be significant enough to warrant the extra complexity without loosing generality. In fact, even the choice of week-day seems to demand extra complexity—but more on that later.

How about the assumptions then, regarding the modeling process? Is it reasonable to assume that each day is independent from all others? Maybe. Probably not. The K-S plots in Figure 4.11 indicates that there are still some information left in the data that we have not unravelled yet in this paper. One can only imagine what that might be.

As far as, somewhat alarmingly, a major proportion of the populace—and perhaps stock-brokers in particular—are very short-term minded in nature; people do like to ride long-term trends. No effort has been made in this paper to catch such a phenomenon; even though we all know it's there. To also assume the absence of medium-term trends and interday dependencies could also be frowned upon. The important thing to realize is that taking all those things into considerations adds *a lot* of complexity to the models; which might not always be a bad thing, but as a rule of thumb: it often is. Anyway, it might be a possible avenue for further study to add more complex features to the models and see what sticks. *Something* (or perhaps some things) is (are) obviously missing here.

With regard to Figure 4.10, I did want to touch upon the somewhat extreme spikes in the BIC-values (and implicitly: the likelihood functions). Although we are not really interested in this approach to modeling, I was personally a bit intrigued by what the source of these occasions could be—at first, I suspected it had something to do with extreme clustering behaviour—so I decided to go to the bottom of it.

I started out by manually going through the counting-process graphs of *every single day* in the data-set. I was on the look-out for 'strange' behaviour manifesting in the plots; meaning some signs of abnormal clustering. My long and arduous search resulted in the red crosses you can see in the figure. To my disappointment, of the eleven days in which I assertively identified out-of-the-ordinary clustering, only one or two coincided with actual extremes. Back to the drawing board. Not long thereafter, I managed to deduce that the size of the data-set under consideration when calculating both the likelihood and the BIC have a profound effect on the magnitude of the acquired value. A quick check confirmed that the ten days with the most events coincided with the ten most extreme BICs. Perfect correlation. It might be a trivial observation, but I decided (obviously) to include it.

Moving on to Table 4.2, it can be seen that the parameters $\breve{\vartheta}$ are pretty close to the mean of $\{\hat{\vartheta}_\ell\}$. This is not strange considering the fact that we have assumed all days to be independent of each other and the linear nature of the mean function. There is not much to add to that.

Turning our attention then towards the estimations on specific week-days as a group we are back in behavioural science territory again. If we look at Figure 4.13, the first thing that strikes us is that Mondays are slow. You would think that brokers are eager to get on top of things at the start of the week, considering a lot of things could potentially have happened over the weekend (when the market is closed). That should mean that $\mu(t)$ on Mondays is higher than the other days in the opening hours. This does not seem to be the case. In fact, Mondays have the lowest starting-intensity (and ending-intensity for that matter) of all days of the week.

By just looking at the exogenous intensities, it seems like the most intense days are Thursdays—with Wednesdays and Tuesdays not far behind though. By following the same line of reasoning as one did when assuming that Mondays would be the most intense day during opening hours, one would probably also come to the conclusion that closing hours on Fridays should be the most hectic closing hours of the week. This line of thought is also refuted by the data.

Since many prudent traders probably don't want to hold too large positions (or any at all) over the weekend, it would be natural to assume that they get liquid (i.e., sell their assets) in the latter part of the trading-day to avoid potential macroeconomic (or other) disasters. However, the graph suggests that the activity in the market tapers off after lunch-time relative to all other days; meaning that a lot of people check out of the game early when they feel that the weekend is just around the corner. It also seems like people go for lunch a little earlier than usual on Thursdays; I have no idea why that would be the case. All in all, the only thing that is clear from Figure 4.13 is that Mondays are considerably more sluggish than the other days. I rest my case.

Whether one can discern lunch-break patterns or not out of the parameters, it is clear that the week-day specific behaviour on the stock exchange is significant enough to call for substantial extra complexity to the model. We're talking about a quintupled amount of parameters here! I was personally quite surprised by the result myself. Perhaps there is something about lunch-times on Thursdays that are much more significant than I thought. . .

That was about everything in this paper that I intend to comment on. As for ideas of how one could possibly take this further, I already mentioned the more-complex-models bit. It is also worth pointing out that what I have done here is the absolute minimal, completely bare bones, most trivial thing one could possibly do with Hawkes processes. The more interesting cases involve multidimensional Hawkes processes where log-likelihoods and other related equations can get very long and very involved very fast. One could also incorporate marked point processes to give events different weights based on, say volume, and so forth.

This kind of process is not merely suitable for financial modeling though: Hawkes himself came up with it to study earthquakes, for example. Other interesting fields—besides economics—could probably be epidemiology, neural networks, seismology, queueing theory, population theory, etc.

Bibliography

[1] Brown et al. *The Time-Rescaling Theorem and its Applications to Neural Spike Train Data Analysis*,
Neural Computation, **14**: 325–346, 2001.

[2] Gut, A. *An Intermediary Course in Probability*, 2nd ed.,
Springer Science+Business Media, pp. 221–268, 2009.

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

[4] Lindgren, G., Rootzén, H., Sandgren, M. *Stationary Stochastic Processes for Scientists and Engineers*,
CRC Press, Taylor & Francis Group, p. 32, 2014.

[5] Muni Toke, I. *An Introduction to Hawkes Processes with Applications to Finance*,
`http://lamp.ecp.fr/MAS/fiQuant/ioane_files/HawkesCourseSlides.pdf`

[6] Ogata, Y. *The asymptotic behaviour of maximum likelihood estimators for stationary point processes*,
Annals of the Institute of Statistical Mathematics, **30**(1): 243–261, 1978.

[7] Ogata, Y. *On Lewis' simulation method for point processes*,
IEEE Transactions on Information Theory, **27**(1): 23–31, 1981.

[8] Ozaki, T. *Maximum likelihood estimation of Hawkes' self-exciting process*,
Annals of the Institute of Statistical Mathematics, **31**(1): 145–155, 1979.

[9] Rubin, I. *Regular point processes and their detection*,
IEEE Transactions on Information Theory, **18**(5): 547–557, 1972.

[10] Shimazaki, H. *Poisson process*,
`http://www.neuralengine.org/res/lecture/ipm/1_Poisson_process.pdf`

[11] Stover, C. *Self-exciting point process*,
`http://mathworld.wolfram.com/Self-ExcitingPointProcess.html`