ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis Econometrie en Operationele Research

Modeling market crashes by assuming similarities with earthquakes in occurrence behavior

Glenn Cato Visser (458412)

Supervisor: dr. A.J. Koning

Second assessor: dr. P. Wan

Date final version: July 7, 2019

Abstract

I investigate similarities in the occurrences of earthquakes and financial market crashes, specifically their self-exciting behavior. This paper revolves around the Hawkes process, as it accounts for this behavior. Residual analysis of the maximum likelihood estimated model for an S&P 500 dataset shows the self-enforcing characteristic of financial market crashes, leading to the conclusion that financial market crashes follow a Hawkes process. Furthermore, I confirm that this model is suitable for the investigation of earthquake occurrences by estimating it for a dataset consisting of data about earthquakes in the vicinity of Japan. Evaluation of simulated data series that are created using the estimated Hawkes process, is inconclusive in determining whether the data contains the characteristics of the original datasets because of an error in the simulation procedure.

Contents

1 Introduction

Being able to predict a market crash in the medium-term future is beneficial for a number of different people. Practically every person that has a connection to financial markets has a lot to gain from knowing when large negative price movements will occur, from traders to monetary policymakers at a central bank.

To be able to predict certain market crashes, one first has to understand what causes them. Sornette (2003) lists five main reasons for crashes: computer trading, derivative securities because they increase the risk in the US market even though they did not yet exist during previous market crashes, illiquidity, trade, and budget deficits and overvaluation. The latter does not seem to trigger market crashes every time. The most important reason that Sornette (2003) listed, however, is the herding behavior of investors. This occurs when investors follow others instead of thinking for themselves. The result is that market crashes can be considered self-enforcing. There exist similarities between market crashes and earthquakes, in the sense that once they happened, the chance exists that there will be aftershocks. Therefore, the use of a model that allows for aftershocks can be beneficial in the modeling of market crashes.
Hence, for the modeling of the crashes, I will use the Epidemic-Type Aftershock Sequence model (ETAS), which is developed by Ogata (1988). This model is widely used to investigate earthquakes and their aftershock occurrences. In the ETAS model, the Hawkes process is traditionally used to model the occurrences of earthquakes, and will now be used to model the occurrences of market crashes. After a shock in the Hawkes model, the change of a new shock increases, after which it slowly starts declining over time, thus mimicking the aftershock behavior. This brings me to my research question:
Do the occurrences of financial market crashes follow a Hawkes process?

The ETAS model served many purposes besides modeling earthquakes. Balderama et al. (2012), for example, used it to study the spread of the red banana plant, whose population was increasing drastically in Costa Rica. Mohler et al. (2011) modeled crime using the ETAS model. It has however also been exploited in the financial sector. It is used by Aït-Sahalia et al. (2015), Embrechts et al. (2011) and Grothe et al. (2014) to model the returns on indices, while Bauwens and Hautsch (2009) used the Hawkes model to study the duration between trades.

I investigate the fit of the Hawkes model for two types of data. The first of the two datasets I use is the "Catalog of Large Earthquakes in the Region of Japan From 1885 Through 1980"

published by Utsu (1982). This includes every earthquake in the vicinity of Japan with a magnitude greater or equal to 6. I use this dataset in order to confirm the suitability of the Hawkes model for earthquake occurrences.
The second dataset is the set of historical prices of the S&P 500 index over the period January 2, 1957, to September 1, 2008. Analyzing the modeling performance of the Hawkes process for this data answers the research question.

In order to get the best information with regards to market crashes, I only model the extreme values of the returns, as proposed by Aït-Sahalia et al. (2015).
Following Gresnigt et al. (2015), I perform the test for the goodness-of-fit for the estimated models with the use of the residual analysis technique of Ogata (1988). To test whether the residual process follows a Poisson distribution with intensity equal to 1, I apply the Kolmogorov-Smirnov test. Furthermore, I perform a simulation study using the estimation results in order to research whether simulated data shows the characteristics of the original earthquake and market crash data.

The residual analyses confirm that the Hawkes model is a good fit for the earthquakes. On top of that, I find out that the occurrences of financial market crashes can be described by the Hawkes process. This fact confirms the self-exciting behavior of the crashes. These results pave the way for the prediction of financial market crashes.
Due to an incorrect simulation procedure, I cannot provide any conclusions on whether the data simulated with the use of the estimated Hawkes model possesses the same features as the original data.

My paper is structured as follows. In Section 2, I describe the Hawkes process and the applied techniques. The results are listed and discussed in Section 3, after which Section 4 concludes and proposes directions for further research.

2 Methodology

2.1 Self-excitement

A counting process is a form of point process. It counts the number of events over time. The counting process $N(t)$ denotes the number of events that occurred in the interval $[0, t]$, while $N(s, t)$ does the same for the interval $[s, t]$. A counting process qualifies for the category "self-exciting point process" if for $s < t < u$

$$cov(\ N(s,t), N(t,u)\ ) > 0, \tag{1}$$

where $cov(X,Y)$ is the covariance between X and Y (Stover, n.d.). The intuition behind this definition is that the probability of future events increases with the occurrence of events in the past. The aftershock behavior of earthquakes makes it interesting to look at models that capture this aspect. An earthquake can trigger a subsequent earthquake, which in its turn can cause another earthquake, and so on. As mentioned in the introduction, herding behavior causes market crashes to show similarities with that of earthquakes. Given a certain market crash, the probability of a next crash increases.

2.2 Hawkes' point processes

Hawkes (1971) was the first to explicitly define a self-exciting point process model when looking for a model that could capture this aspect of earthquake occurrences. This resulted in the name Hawkes process. Ogata (1988) compared a Hawkes process and several other models with respect to their ability to predict earthquakes. This led to the conclusion that the proposed Hawkes model, defined as Epidemic Type Aftershock-Sequences (ETAS) model, is superior. Over the years, many different extensions of the ETAS model have been used in the field of seismology, and are still of frequent occurrence (Liniger, 2009).

Hawkes models take into account the positive relationship between the arrival of one event and the probability of subsequent arrivals. A Hawkes process is a non-homogeneous Poisson process. This means that its intensity rate is a function of time. The history of events before a certain point in time determines the chance of a new event at that point. The conditional intensity of jump arrivals following a Hawkes process is given by

$$\lambda(t) = \mu + \int_0^t g(t-s)dN(s), \tag{2}$$

where μ is the constant part of the intensity, N(s) is a counting process and $g(t)$ is called the response function (Ozaki, 1979). In order to account for the elastic aftereffect described by Lomnitz (1974), I set

$$g(t) = ae^{-\alpha t}. \tag{3}$$

This effect is based on the theory of elastic rebound. It concerns the situation where an object, in this case a tectonic plate, gets forcibly deformed within its elastic limits. When this happens, a certain amount of time passes before the plate returns to its initial form. This is called the elastic aftereffect. In financial terms, this effect can be interpreted as the time it takes for stock prices to return to their normal level after a market crash.

The efficient market hypothesis (EMH) is a theory that states that stock prices are always fair, or in other words, on their true level. This could only be the case when all available information is reflected in these prices. This would mean that it is impossible for any investor to generate profits by buying undervalued or selling overvalued stocks. The behavior of stock prices after a market crash, however, is an indication that prices can deviate from their fair level. This is in contradiction with the EMH. The fact that Hawkes models account for an effect of market crashes that influences stock prices over a longer period of time, implies that not all information is instantly reflected in stock prices, thus contravening the EMH.
In the case of an event at time t, equation (2) shows that $dN(t) = 1$ and consequently $d\lambda(t) = g(0) = a$. This event has a decreasing influence on $\lambda(t)$, such that at time $u > t$ the increase in $\lambda(t)$ caused by the event at time t is equal to $g(u-t) = ae^{-\alpha(u-t)}$ (Bacry et al., 2012). This shows the self-excited aspect of the process.

2.3 Estimation

In the early applications of Hawkes processes, the parameters of the models were often estimated through spectral analysis techniques, but this grew less popular over time. The maximum likelihood method for point processes that is developed by Rubin (1972) was further investigated and applied to Hawkes models by Ozaki (1979). When performing maximum likelihood estimation, one focuses on the likelihood function

$$L(\boldsymbol{\theta}|y) = f(y|\boldsymbol{\theta}), \tag{4}$$

where $\boldsymbol{\theta}$ is the vector of parameters and y is the observed data. The likelihood function expresses the likelihood of the parameters in $\boldsymbol{\theta}$ given the observed data. By maximizing the likelihood function, one chooses the parameters that make the observed data most likely. This means that these parameters have the largest probability of being equal to the parameters in the data generating process. The functions $L(\boldsymbol{\theta}|y)$ and $log(L(\boldsymbol{\theta}|y))$ are monotonically related. This means that when $L(\hat{\boldsymbol{\theta}}_1|y) < L(\hat{\boldsymbol{\theta}}_2|y)$, then $log(L(\hat{\boldsymbol{\theta}}_1|y)) < log(L(\hat{\boldsymbol{\theta}}_2|y))$. The consequence of this fact is that maximizing either one of these functions will give the same parameter estimates. Since the optimization of the log-likelihood function $log(L(\boldsymbol{\theta}|y))$ requires less computational power, this is the function that is used in maximum likelihood estimation (Myung, 2003). Given the observed events at time $t_1, t_2.., t_n$ in the interval $[0, T]$ the log-likelihood function is given by

$$logL(\boldsymbol{\theta}|t_1, ..., t_n) = \sum_{i=1}^{n} log(\ \mu + \int_0^{t_i} g(t_i - s)dN(s)\ ) - \int_0^T (\ \mu + \int_0^t g(t-s)dN(s)\ )dt, \tag{5}$$

where $\boldsymbol{\theta} = (\mu, a, \alpha)^T$ is the parameter vector (Rubin, 1972). After substituting and rewriting this comes down to

$$\begin{aligned} &logL(\boldsymbol{\theta}|t_1, ..., t_n) = \\ &\sum_{i=1}^{n} log(\ \mu + \int_0^{t_i} ae^{-\alpha(t_i-s)} dN(s)\ ) - \int_0^T (\ \mu + \int_0^t ae^{-\alpha(t-s)} dN(s)\ )dt = \\ &\sum_{i=1}^{n} log(\ \mu + \sum_{t_j<t_i} ae^{-\alpha(t_i-t_j)}\ ) - (\ \mu T + \sum_{t_j<T} \frac{a}{\alpha}(\ 1 - e^{-\alpha(T-t_j)}\ )\ ). \end{aligned} \tag{6}$$

In order to find the parameter estimates, the maximum of the log-likelihood function has to be determined. To find an extremum, i.e. maximum or minimum, of a continuous and differentiable function, the chosen parameters must result in a first derivative that is equal to zero. This is called the likelihood equation. Therefore, the set of likelihood equations is given by

$$\frac{\delta logL(\boldsymbol{\theta}|t_1, ..., t_n)}{\delta \boldsymbol{\theta}} = \mathbf{0}. \tag{7}$$

See the appendix for the three likelihood equations of the specified Hawkes process. When the log-likelihood function is non-linear and contains multiple parameters, it is often not possible to determine the solutions to the likelihood equations analytically, like in this case. The best approach is then to use a nonlinear optimization algorithm.

While Newton-type methods calculate the Hessian matrix $\boldsymbol{H}$ directly, quasi-Newton methods approximate the Hessian in order to be more computational and time-efficient. This Hessian matrix consists of the second-order derivatives. Multiple methods for estimating $\boldsymbol{H}$ have been proposed over the years, of which the one created by Broyden, Fletcher, Goldfarb, and Shanno (BFGS algorithm) is widely regarded as the most effective. In this method, $\boldsymbol{H_0}$ is a random symmetric, positive definite matrix. The updating of the approximated Hessian then happens in the following manner:

$$\boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \frac{\boldsymbol{q}_k \boldsymbol{q}_k^T}{\boldsymbol{q}_k^T \boldsymbol{s}_k} - \frac{\boldsymbol{H}_k \boldsymbol{s}_k \boldsymbol{s_k}^T \boldsymbol{H}_k^T}{\boldsymbol{s}_k^T \boldsymbol{H}_k \boldsymbol{s}_k}, \tag{8}$$

with

$$\boldsymbol{q}_k = \boldsymbol{\nabla f}(\boldsymbol{x}_{k+1}) - \boldsymbol{\nabla f}(\boldsymbol{x}_k), \tag{9}$$

$$\boldsymbol{s}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k \tag{10}$$

(Mathworks, n.d.). As Matlab only provides the BFGS algorithm as part of a minimization algorithm, I minimize $-logL(\boldsymbol{\theta}|t_1, ..., t_n)$, which yields the same parameter estimates as maximizing $logL(\boldsymbol{\theta}|t_1, ..., t_n)$.

2.4 Residual analysis

In order to assess the goodness-of-fit of the proposed model, I perform the residual analysis Ogata (1988) describes. As mentioned before, I assume the data $\{t_i\}$ is generated by a self exciting point process with the conditional intensity given by equation (2). Now consider the transformed times $\{\tau_i\}$, which are determined by

$$\begin{aligned}\tau_i &= \int_0^{t_i} \lambda(t)dt \\ &= \int_0^{t_i} (\ \mu + \int_0^{t} ae^{-\alpha(t-s)}dN(s)\ )dt \\ &= \int_0^{t_i} \mu dt + \int_0^{t_i}\int_s^{t_i} ae^{-\alpha(t-s)}dtdN(s) \\ &= \mu t_i + \sum_{t_j<t_i}(\ \frac{-a}{\alpha}e^{-\alpha(t_i-t_j)} + \frac{a}{\alpha}\ ).\end{aligned} \tag{11}$$

According to Papangelou (1972), the transformed times $\{\tau_i\}$ follow a stationary, homogeneous Poisson(1) process. The transformed data can loosely be interpreted as the residuals or noise of the original point process $\{t_i\}$. That is why $\{\tau_i\}$ is also called the residual process. Figure 2.1 shows a plot of the timeline of the earthquake times, as well as the sequence of transformed times. For the purpose of visibility, I plot only a part of the full sequences. I choose observation 271-308 for both plots, because these observations in $\{t_i\}$ clearly show the clustering characteristic of the earthquake occurrences times. At the same time, it is visible that the residual process $\{\tau_i\}$ is "spread out" much more evenly over time. This can be an indication that they follow a homogeneous Poisson process.

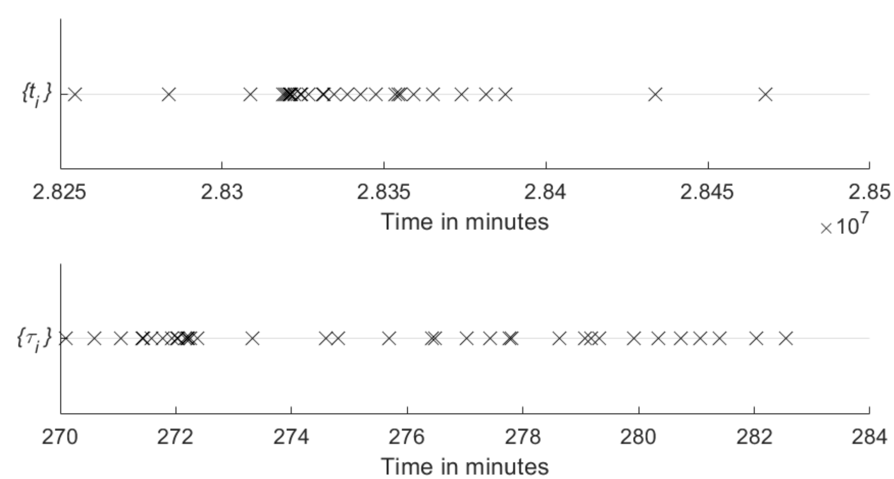

Figure 2.1: A timeline of the same 38 events for both the earthquake sequence and the residual process of this sequence

When the estimated conditional intensity denoted by $\lambda(t, \hat{\boldsymbol{\theta}})$ is a good approximation of $\lambda(t)$, then $\tau_i(t, \hat{\boldsymbol{\theta}})$ should have the characteristics of a Poisson(1) process. If this is the case, then the transformed interarrival times, $\zeta_i = \tau_{i+1} - \tau_i$, are independent exponential random variables with mean equal to 1 (Gresnigt et al., 2015). In order to verify this, I perform a Kolmogorov-Smirnov test to obtain a test decision for the null hypothesis that the transformed interarrival times follow an exponential distribution with intensity 1. For the Kolmogorov-Smirnov test, I use Matlab to determine the empirical cumulative distribution function, which I will denote by $\hat{F}$. The one-sample Kolmogorov-Smirnov test statistic is the absolute value of the maximized difference between $\hat{F}$ and the hypothetical cumulative distribution function, denoted by G (Massey Jr., 1951). In this situation, G is the cumulative distribution function of the *exponential*(1) distribution, i.e.

$$KS = max_{\boldsymbol{x}}(\,|\hat{F}(\boldsymbol{x}) - G(\boldsymbol{x})|\,). \tag{12}$$

3 Results

3.1 Application to earthquake data

Following Ogata (1988), I use the earthquake dataset constructed by Utsu (1982). It contains information about every earthquake within the "off Tohoku district" in the Pacific Ocean near Japan for the period of 1885 until 1980, provided the earthquake had a magnitude equal to or larger than 6.0. The district is located within the Circum-Pacific seismic belt, where the Pacific Plate subducts underneath the Eurasian plate. This causes the district to suffer from earthquakes frequently.

Performing maximum likelihood estimation on the dataset using the previously specified Hawkes model provides the following estimates:

Table 3.1: Maximum likelihood estimation results of the Hawkes model applied to the earthquake data

	$\hat{\mu}$	$\hat{a}$	$\hat{\alpha}$	*-LogL*	*AIC*
Value	6.71E-06	0.000127	0.000426	5796.33	11598.7

These are the parameters estimated for equation (2). As can be seen in this equation, $\hat{\mu}$ is the constant part of the estimated conditional intensity rate $\lambda(t, \hat{\boldsymbol{\theta}})$, and $\hat{a}$ and $\hat{\alpha}$ the estimates for the parameters in the response function.

I derive the standard errors for these estimates from the Hessian matrix of the log-likelihood, evaluated at the optimal point. The negative of the Hessian at this point is called the observed Fisher information matrix $\boldsymbol{I}(\hat{\boldsymbol{\theta}})$. Assuming asymptotic normality for the maximum likelihood estimates with as expected value the true parameter values and variance equal to the inverse of the observed Fisher information criterion is denoted as

$$\hat{\boldsymbol{\theta}} \overset{a}{\sim} \mathcal{N}(\,\boldsymbol{\theta},\ [\boldsymbol{I}(\hat{\boldsymbol{\theta}})]^{-1}\,). \tag{13}$$

This shows that the standard errors of the estimates are equal to the square roots of the diagonal elements of this asymptotic covariance matrix, i.e.

$$SE(\hat{\boldsymbol{\theta}}) = \frac{1}{\sqrt{\boldsymbol{I}(\hat{\boldsymbol{\theta}})}}, \tag{14}$$

(Pawitan, 2013). In order for a certain point to qualify as a local minimum, it is a necessary condition that the Hessian at that point is positive semidefinite. Intuitively, this means that the log-likelihood will not decrease in any direction as seen from that point. The approximated Hessian for this model, however, does not meet this condition. This results in standard errors for the maximum likelihood estimates that are imaginary numbers. I state with fair certainty that the estimated parameters do not provide a local minimum, even though the given point might be close to a minimum. There are several possible causes for this problem. The first focuses on the fact that the BFGS algorithm is of the quasi-Newton type, and thus does not use the actual Hessian, but merely an approximation. It is possible that the actual Hessian would positive semidefinite at point $\hat{\boldsymbol{\theta}}$.

A second, more likely explanation is that the minimization algorithm that I use met one of its stopping conditions before finding an actual minimum. This can occur when for instance the function is very flat, but still decreasing around the possible minimum. If an iteration of the minimization does not result in a sufficient decrease in the objective function, the algorithm is terminated.

The consequence for this estimation is the absence of standard errors for the estimates, which are most likely not an optimal solution for the maximum likelihood estimation.

Since I use the same dataset as Ogata (1988) and the estimating approach is similar, it is interesting to compare the results to the ones listed in this paper. Unfortunately, Ogata (1988) does not list the parameter estimates, but only the log-likelihood evaluated at the minimum. For the same response function I assume in the Hawkes process, equation (3), Ogata (1988) lists the negative log-likelihood as ($-logL = 2288.4$). This differs greatly from the resulting log-likelihood ($-logL = 5796.33$) in this paper. Comparing the Akaike information criterion (AIC) of the two papers provides an insight into which model has the better fit for the data. Since the same model is used in both papers, a higher log-likelihood results in a higher AIC. However, since not only the model is identical but also the used dataset, the reasonable expectation would be that the AICs of the models would be close to each other. That is why an AIC of 4582.8 for Ogata (1988) against 11598.7 for my model is another surprising result.

Assuming that the same equation (6) is used for the log-likelihood, this is an indication of substantially different parameter estimates. On top of that, the AICs suggest that the model Ogata (1988) estimates, is superior in terms of fitting the dataset. This is remarkable as the models are supposed to be each other's equivalent.

A possible explanation for this observation lies in the usage of the dataset, which lists the earthquake occurrence times down to the minute. If Ogata (1988) chose to set the interval lengths to daily or hourly instead of by the minute as I do in this paper, this causes different results.

Another possible cause of these deviant results I consider more plausible is the fact that the found solution for the minimization of the negative log-likelihood function does not actually provide a minimum, as discussed before. This results in an estimated model that is inferior to a model that has estimated parameters that do provide a local minimum.

3.2 Application to S&P 500 financial data

Besides the earthquake data, I use the described methodology on the financial dataset, consisting of the price levels of the S&P 500 index for the period January 2, 1957, to September 1, 2008. For my research, I consider the returns of the index, given by

$$R_t = \frac{p_t - p_{t-1}}{p_{t-1}} \times 100, \tag{15}$$

where p_t is the closing price of the S&P 500 index on day t, adjusted for dividends and splits. The corresponding set of returns consists of 13005 data points. It is necessary to decide what classifies as a market crash in terms of returns. Following Gresnigt et al. (2015), I set the 95% quantile of negative returns as market crashes. This results in a total of 650 crashes. Performing maximum likelihood estimation on these observations provided the following estimates for the Hawkes model specified by equation (2):

Table 3.2: Maximum likelihood estimation results of the Hawkes model applied on the S&P 500 market crash data

	$\hat{\mu}$	$\hat{a}$	$\hat{\alpha}$	*-LogL*	*AIC*
Value	0.0120 (0.0017)	0.0304 (0.0044)	0.0397 (0.0061)	2355.69	4717.38

Unfortunately, it is not possible to compare these results directly with those listed by Gresnigt et al. (2015). While several models are estimated in the aforementioned paper, the one specified by equation (2) is absent. It is possible however to compare the performance of the model I estimate with that of the other models of Gresnigt et al. (2015) by means of their AIC. These indicate that my model is of a higher quality for the financial dataset, as it has a lower AIC than any one of the models estimated by Gresnigt et al. (2015).
Even though the dataset I use consists of returns from the same index, over the same period, it consists of around 700 fewer returns than the one Gresnigt et al. (2015) uses. This is possibly due to the fact that Gresnigt et al. (2015) add stock price levels to the dataset for the national holidays, which are not trading days, for example the prices of the day before. This might contribute to part of the difference in the results.

3.3 Residual Analysis

When the Kolmogorov-Smirnov test statistic, given by equation (12), becomes larger than the approximated critical value, it means that the maximum difference between the empirical and hypothetical cumulative distribution function becomes too large for it to be likely that the data is generated by the process that I assume under the null hypothesis. Depending on the chosen significance level and the calculated p-value, the decision is made whether or not the null hypothesis is rejected. I set the significance level equal to 0.05, which is often considered to be the default level.

3.3.1 Earthquake data

In order to be able to judge the fit of the Hawkes model for the earthquake data, I perform a residual analysis on this dataset. The results for the Kolmogorov-Smirnov test with the null hypothesis that the transformed times follow a Poisson(1) process are listed below:

Table 3.3: Results of the Kolmogorov-Smirnov (K-S) test with the null hypothesis that the transformed interarrival times of the earthquakes are exponential(1) random variables

	Approximated critical value	*K-S test statistic*	*p-value*
Value	0.0615	0.0586	0.0700

These results indicate that even though the estimates are suboptimal and the model has a larger AIC than the ones Ogata (1988) tests, it does seem to fit the data reasonably well. The results lead to the test decision of not rejecting the null hypothesis that the transformed interarrival times are independent *exponential*(1) random variables, or equivalently, that the transformed times follow a homogeneous Poisson process with intensity 1. This means that at a 5% level, the earthquake occurrences do not deviate significantly from the Hawkes point process model specified by equation (2) and the estimated parameters in table 3.1.

3.3.2 S&P 500 financial data

I perform the same residual analysis on the S&P 500 financial data, to evaluate the goodness-of-fit of the Hawkes model. The results for the Kolmogorov-Smirnov test with the null hypothesis that the transformed times follow an unit Poisson process are listed below:

Table 3.4: Results of the Kolmogorov-Smirnov (K-S) test with the null hypothesis that the transformed interarrival times of the market crashes are exponential(1) random variables

	Approximated critical value	*K-S test statistic*	*p-value*
Value	0.0530	0.0514	0.0630

These results lead to the conclusion that the Hawkes point process model specified by equation (2) and the estimates in table 3.2, fits the S&P 500 data well enough for me not to dismiss it based on this test. I do not reject the null hypothesis that the transformed times follow a homogeneous Poisson process with intensity 1. That is, the negative returns above the 95% quantile, that I consider as crashes, do not deviate significantly from the model at a 5% level.

3.4 Simulation study

In order to simulate the times at which an event happens according to the estimated Hawkes model, I follow the procedure Gresnigt et al. (2015) describes:

1. Since the time of the first event is not influenced by any other event occurrences, the expected time until the first event t_1 is exponentially distributed with intensity $\hat{\mu}$.
2. For every t_n after t_1, I calculate the probability P_{no_event} that there is no event in the interval (t_{n-1}, t_n).
3. Draw a number u from a uniform distribution on the interval $(0, 1)$.
4. If $u > P_{no\_event}$, an event occurs at t_n. If $u < P_{no\_event}$, do nothing.
5. Repeat until $t_n = T$.

P_{no_event} for the interval (t_{n-1}, t_n) is the probability of no event occurring in said interval. This probability is defined as

$$P_{no\_event} = P(\ N(t_n) - N(t_{n-1}) = 0\ )$$

$$\begin{aligned} &= 1 - exp(\ -\int_{t_{n-1}}^{t_n} \lambda(t)\ ) \\ &= 1 - exp(\ (\mu t_{n-1} - \mu t_n) - \sum_{t_j < t_{n-1}} (\ -\frac{a}{\alpha} e^{-\alpha(t_n - t_j)} + \frac{a}{\alpha} e^{-\alpha(t_{n-1} - t_j)}\ ) - \frac{a}{\alpha} e^{-\alpha(t_n - t_{n-1})} + \frac{a}{\alpha}\ ). \end{aligned} \tag{16}$$

For each of the two datasets used, I simulate 10,000 new datasets using the estimated parameter values as listed in 3.1 and 3.2. Over these new datasets I determine the new estimates for the parameters. For each of the three parameters, I take the average of the 10,000 values and compare them to the real values used in the data generating process. These results are listed below.

3.4.1 Earthquake data

Table 3.5: Results of maximum likelihood estimation of the Hawkes model for the simulated earthquake datasets and p-values for the t-tests with null hypothesis that the estimated parameters are equal to the data generating parameters

	μ	a	α
true value	6.71e-06	0.000127	0.000426
estimated value	5.41e-08	0.775	0.546
	(2.77e-08)	(0.420)	(0.264)
p-value	< 0.00001	< 0.00001	< 0.00001

It can be seen that the estimated values deviate from the true data generating parameters. I test whether these differences are significant, by means of performing two-tailed t-tests. The null hypothesis in these tests is that the estimated parameters are equal to their "true" counterparts in the data generating process. The p-values for these tests are listed in table 3.5 as well, showing that all three estimated parameters differ significantly from their true value.

3.4.2 S&P 500 financial data

Table 3.6: Results of maximum likelihood estimation of the Hawkes model for the simulated S&P crash datasets and p-values for the t-tests with null hypothesis that the estimated parameters are equal to the data generating parameters

	μ	a	α
true value	0.0120	0.0304	0.0397
estimated value	1.11e-04	0.272	0.308
	(4.64e-05)	(0.125)	(0.161)
p-value	< 0.00001	< 0.00001	< 0.00001

For this dataset, I observe from table 3.6 that the estimated values are different from the true values as well. Furthermore, the p-values of the t-tests show that again, these differences are significant.

Neither the results of the simulation of the earthquake occurrences, nor those of the S&P 500 crashes are in accordance with expectations. This is most likely due to an error I make in calculating the probabilities of no event happening given by equation (16). When evaluating the probability, it becomes clear that it starts out small, but then increases monotonically in $N(s)$. It does not provide the intended characteristic of self-exciting behavior, since the probability of a subsequent event decreases every time an event occurs. This results in simulated datasets with events clustered at the beginning of the sample period. After a certain point the probability of an event happening becomes so small, that no more events are generated. This point is reached in such a fast manner, that for the 10,000 S&P 500 crash simulations, the latest simulated crash is on average at day 127. This is a clear indication of a faulty simulation, considering that the sample period is $[0, 13005]$. The average number of generated crashes is 28, against 650 in the estimation sample. For the earthquake data, the average last simulated earthquake is at minute 149534, which is at about 0.3% of the total sample period $[0, 50490720]$. The average number of simulated earthquakes for the 10,000 earthquake simulations is 127, against 483 for the estimation sample. To give a visual representation of the problem with the simulated datasets, I plot the first simulation for the S&P data. This is shown in figure 3.1.

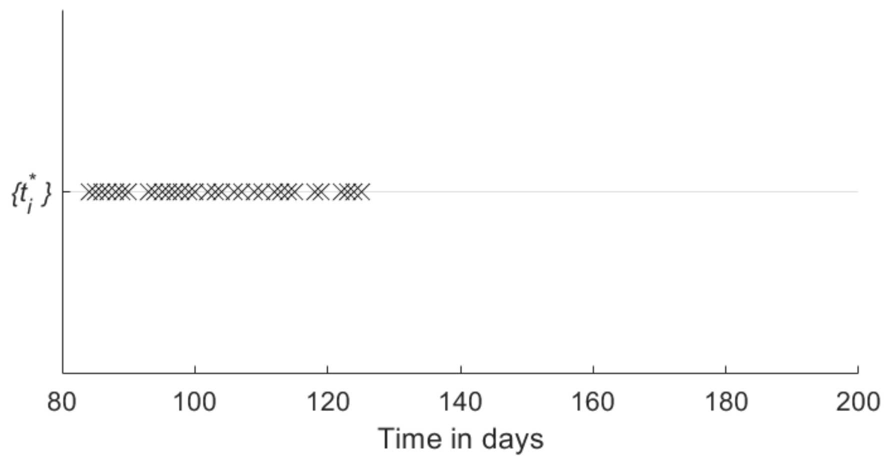

Figure 3.1: A timeline of the 32 market crashes in the first simulated S&P crashes dataset

As the expected time until the first event is exponentially distributed with intensity $\hat{\mu}$, $t_1 = 84$ for every financial simulation. As the last simulated crash in this dataset is at $t = 125$, the horizontal axis is bounded at $t = 200$ in this plot. I am not able to make conclusions from these results, as they are erroneous. The most conspicuous improvement for this procedure is to correctly calculate the probability given by equation (16).

4 Conclusion

This paper investigates the similarities in the behavior of stock prices around market crashes and that of earthquakes. At the center of this research are the Hawkes point process and the question of whether this model can be used to describe financial market crashes. In previous papers, this process has been used to describe earthquake occurrences. Its conditional intensity includes a constant part and a part called the response function, that is affected by previous occurrences. Events can cause subsequent events, while this probability decreases as time passes after the original event, causing a clustering characteristic.
The model is first applied to an earthquake dataset, consisting of every earthquake with a magnitude of 6 or higher, that occurred in the vicinity of Japan. The maximum likelihood parameter estimates are used to construct the residual process, which shows that the model possesses a good fit for the data. This confirms the self-exciting behavior of earthquakes.
The same is done for an S&P 500 index dataset, containing daily returns from a period of approximately 52 years. The model is applied to the 95% quantile of extreme negative returns. Again, the residual analysis confirms that the maximum likelihood estimated Hawkes model that accounts for self-enforcing behavior is a good fit for the data.

For both datasets, the parameter estimates are used to generate 10,000 new datasets, 20,000 in total, according to the algorithm Gresnigt et al. (2015) describes. Over these datasets, the parameters are once more estimated using maximum likelihood, and their average is compared with the true values. For both datasets, these estimates differ significantly from their counterparts in the data generating process. This is the result of a mistake in the simulation procedure, in particular at the calculation of the probability of no event happening in a certain interval. This causes the simulated datasets to consist of fewer events than the estimation samples, all clustered at the beginning of the sample period. For the simulation results to be useful, the probability of no event happening in a certain interval needs to be determined accurately. Unfortunately, it is not possible to achieve this within the given time span for this thesis.

I conclude that the occurrences of financial market crashes follow a Hawkes process. This finding can be used in further research to make predictions about the probabilities of crashes in the future.

More research on this topic can prove to be interesting in multiple directions. The first option is to incorporate different response functions and compare these with the model used in this paper by means of their Akaike information criterion. Besides that, further investigation of the simulated series is interesting, in order to analyze whether they possess the major features that characterize the Hawkes process. Lastly, the models could be applied to high-frequency financial data.

References

Aït-Sahalia, Y., Cacho-Diaz, J., & Laeven, R. (2015). Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, *117*(3), 585-606. doi: https://doi.org/10.1016/j.jfineco.2015.03.002

Bacry, E., K.Dayri, & Muzy, J. F. (2012). Non-parametric kernel estimation for symmetric hawkes processes. application to high frequency financial data. *The European Physical Journal B*, *85*(157). doi: 10.1140/epjb/e2012-21005-8

Balderama, E., Schoenber, F., Murray, E., & Rundel, P. (2012). Application of branching models in the study of invasive species. *Journal of the American Statistical Association*, *107*(498), 467-476. doi: https://doi.org/10.1080/01621459.2011.641402

Bauwens, L., & Hautsch, N. (2009). *Modeling financial high frequency data using point processes*. Springer.

Embrechts, P., Liniger, T., & Lin, L. (2011). Multivariate hawkes processes: an application to financial data. *Journal of Applied Probability*, *48A*, 367-378. doi: https://doi.org/10.1239/jap/1318940477

Gresnigt, F., Kole, E., & Franses, P. (2015). Interpreting financial market crashes as earthquakes: A new early warning system for medium term crashes. *Journal of Banking Finance*, *56*, 123-139. doi: https://doi.org/10.1016/j.jbankfin.2015.03.003

Grothe, O., Korniichuk, V., & Manner, H. (2014). Modeling multivariate extreme events using self-exciting point processes. *Journal of Econometrics*, *182*(2), 269-289. doi: https://doi.org/10.1016/j.jeconom.2014.03.011

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, *58*(1), 83-90. doi: 10.2307/2334319

Liniger, T. J. (2009). *Multivariate hawkes processes* (Unpublished doctoral dissertation). Eidgenössische Technische Hochschule Zürich, Rämistrasse 101, 8092 Zurich, Switzerland.

Lomnitz, C. (1974). *Global tectonics and earthquake risk, volume 5* (1st ed.). American Elsevier Publishing Co.

Massey Jr., F. J. (1951). The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical Association*, *46*(253), 68-78. doi: 10.2307/2280095

Mathworks. (n.d.). *Unconstrained nonlinear optimization algorithms.* Retrieved from `https://nl.mathworks.com/help/optim/ug/unconstrained-nonlinear-optimization-algorithms.html#brnpcye`

Mohler, G., Short, M., Brantingham, P., Schoenberg, F., & Tita, G. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, *106*(493), 100-108. doi: https://doi.org/10.1198/jasa.2011.ap09546

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47*(1), 90-100. doi: 10.1016/S0022-2496(02)00028-7

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, *83*(401), 9-27. doi: 10.2307/2288914

Ozaki, T. (1979). Maximum likelihood estimation of hawkes' self-exciting point processes. *Annals of the Institute of Statistical Mathematics*, *31*(1), 145-155. doi: 10.1007/BF02480272

Papangelou, F. (1972). Integrability of expected increments of point processes and a related random change of scale. *Transactions of the American Mathematical Society*, *165*, 483-506. doi: 10.2307/1995899

Pawitan, Y. (2013). *In all likelihood: Statistical modelling and inference using likelihood* (1st ed.). OUP Oxford.

Rubin, I. (1972). Regular point processes and their detection. *IEEE Transactions on Information Theory*, *18*(5), 547-557. doi: 10.1109/TIT.1972.1054897

Sornette, D. (2003). Critical market crashes. *Physics Reports*, *378*(1), 1-98. doi: https://doi.org/10.1016/S0370-1573(02)00634-8

Stover, C. (n.d.). *Self-exciting point process.* Retrieved from `http://mathworld.wolfram.com/Self-ExcitingPointProcess.html`

Utsu, T. (1982). Catalog of large earthquakes in the region of Japan from 1885 through 1980. *Bulletin of the Earthquake Research Institute*, *57*, 401-463.

5 Appendix

5.1 Likelihood equations

$$\frac{\delta LogL}{\delta\mu} = \sum_{i=1}^{n} \frac{1}{\mu + \sum_{t_j<t_i} ae^{-\alpha(t_i-t_j)}} - T \\ = 0. \tag{17}$$

$$\frac{\delta LogL}{\delta a} = \sum_{i=1}^{n} \frac{\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}}{\mu + \sum_{t_j<t_i} ae^{-\alpha(t_i-t_j)}} - \sum_{tj<T} \frac{1}{\alpha}(\ 1 - e^{-\alpha(T-t_j)}\ ) \\ = 0. \tag{18}$$

$$\frac{\delta LogL}{\delta\alpha} = -\sum_{i=1}^{n} \frac{\sum_{t_j<t_i} ae^{-\alpha(t_i-t_j)} \times (t_i - t_j)}{\mu + \sum_{t_j<t_i} ae^{-\alpha(t_i-t_j)}} - \sum_{t_j<T} \frac{a}{\alpha}(\ (\frac{1}{\alpha} + (T - t_j))e^{-\alpha(T-t_j)} - \frac{1}{\alpha}\ ) \\ = 0. \tag{19}$$

5.2 Matlab codes

The matlab codes that I used to perform my research are included in a zip-file. Every file is briefly explained below.

- *createData.m* transforms the dataset given by (Ogata, 1988) into a vector containing the event occurrence times.
- *createTransformedTimes.m* calculates the transformed times τ_i of the sequence t_i, the interarrival times ζ_i and performs the Kolmogorv-Smirnov test specified in the text.
- *dummy.m* is used in *simulation.m* to return a boolean that is true if an event occurs at a certain time, and false otherwise.
- *fixSPData.m* transforms the pricelevels of the SP 500 index into returns.
- *getParametersSimulatedSeries.m* performs MLE on the 10,000 simulated datasets.
- *plotting.m* plots an example of the earthquake times against their transformed times.
- *plottingSimulaties.m* plots the first simulated S&P crash dataset.
- *simulation.m* simulates 10,000 datasets using the parameters estimated for the Hawkes model.
- *thesis.m* performs MLE for the Hawkes model.