TI 2014-067/III
Tinbergen Institute Discussion Paper

Interpreting Financial Market Crashes as Earthquakes: A New early Warning System for Medium Term Crashes

Francine Gresnigt
Erik Kole
Philip Hans Franses

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

1 Introduction

This paper proposes a modeling framework that draws upon the self-exciting behavior of stock returns around a financial market crash, which is similar to the seismic activity around earthquakes. Incorporating the tendency for shocks to be followed by new shocks, our framework is able to create probability predictions on a medium-term financial market crash. A large literature in finance has focused on predicting the risk of downward price movements one-step ahead with measures like Value-at-Risk and Expected Shortfall. Our approach differs however as we interpret financial crashes as earthquakes in the financial market, which allows us to develop an Early Warning System (EWS) for crash days within a given period. The EWS is tested on S&P 500 data during the recent financial crisis, starting from September 1, 2008. As will become apparent in later sections, our modeling framework differs from Extreme Value models as we allow dependencies across arrival times and magnitudes of shocks. At the same time, our framework differs from the conventional GARCH models by generating highly insightful medium term forecasts, while not having to make stringent assumptions on the tail behavior of error distributions. This makes our approach rather easy to implement and understand in practice.

The identification and prediction of crashes is very important to traders, regulators of financial markets and risk management because a series of large negative price movements during a short time interval can have severe consequences. For example, on Black Monday, that is October 19, 1987, the S&P 500 index registered its worst daily percentage loss of 20.5%. During the recent credit crisis, financial indices declined dramatically for numerous days, thereby suffering its worst yearly percentage loss of 38.5 % in 2008. Unfortunately, crashes are not easy to predict, and there still is a need for tools to accurately forecast the timing of a series of large negative price movements in financial markets.

To initiate the construction of our modeling framework for stock market crashes, we first focus on the potential causes of such crashes. Sornette (2003) summarizes that computer trading, and the increased trading of derivative securities, illiquidity, and trade and budget deficits and also overvaluation can provoke subsequent large negative price movements. More importantly, Sornette (2003) points out that speculative bubbles leading to crashes are likely to result from a positive herding behavior of investors. This positive herding behavior causes crashes to be locally self-enforcing. Hence, while bubbles can be triggered by an exogenous factor, instability grows endogenously. A model for stock market crashes should therefore be able to capture this self-excitation. Notably, such a self-excitation can also be observed in seismic behavior around earthquake sequences, where an earthquake usually generates aftershocks which in turn can generate new aftershocks and so on. For many academics (and perhaps practitioners), earthquakes and stock returns therefore share characteristics typically observable as the clustering of extremes and serial dependence.

Potential similarities across the behavior of stock returns around crashes and the dynamics of earthquake sequences have been noted in the so-called econophysics literature, in which physics models are applied to economics.[1] In contrast to the studies in the econophysics literature and also to related studies like Bowsher (2007) and Clements and Liao (2013), in our framework we do not model the (cumulative) returns but only the extreme returns. As such, we most effectively exploit the information contained in the returns about the crash behavior. As Aït-Sahalia et al. (2013) already show, only taking the jump dynamics of returns into account to approximate the timing of crashes gives more accurate results than using the full distribution of the returns. As is well known, the distribution of stock returns is more heavy-tailed than the Gaussian distribution as extreme returns occur more often than can be expected under normality. Furthermore, the distribution of stock returns is usually negatively skewed. As risk in financial markets is predominantly related to extreme price movements, we propose to model only extreme (negative) returns in order to improve predictions.

To model the extreme (negative) returns we use a particular model that is often used for earthquake sequences, and which is the so-called Epidemic-type Aftershock Sequence model (ETAS). This model has been developed by Ogata (1988) and its use for earthquakes is widely investigated by geophysicists.[2] In the ETAS model a Hawkes process, an inhomogeneous Poisson process, is used to model the occurrence rate of earthquakes above a certain threshold. The jump rate of the Hawkes process increases when a jump (or shock) arrives after which the rate decays as a function of the time passed since the jump. As the probability of jumps increases after a jump has occurred, the Hawkes process is thus called self-exciting. The ETAS model has been exploited for crime rates (Mohler et al., 2011) and for the spread of red banana plants (Balderama et al., 2011). Interestingly, the ETAS model has also been applied to financial data, for example to model arrival data of buy and sell trades (Hewlett, 2006), the duration between trades (Bauwens and Hautsch, 2009) or the returns on multiple indices (Aït-Sahalia et al. 2013, Embrechts et al. 2011, and Grothe et

[1] See amongst others: Sornette, 2003, Weber et al., 2007, Petersen et al., 2010, Baldovin et al., 2011, Baldovin et al., 2012a, Baldovin et al., 2012b, and Bormetti et al., 2013

[2] See amongst others: Ogata, 1998, Helmstetter and Sornette, 2002, Zhuang et al. 2002, Zhuang and Ogata, 2004, Saichev et al., 2005, Hardebeck et al., 2008, and Veen and Schoenberg, 2008

al. 2012).

Our modeling framework entails that we use the ETAS model as a tool to warn for an upcoming crash (read: earthquake) in a financial market. As Herrera and Schipp (2009), Chavez-Demoulin et al. (2005) and Chavez-Demoulin and McGill (2012), already showed when deriving their Value-at-Risk and Expected Shortfall estimates, the ETAS model can contribute to the modeling and prediction of risk in finance. However, in contrast to Herrera and Schipp (2009), Chavez-Demoulin et al. (2005) and Chavez-Demoulin and McGill (2012) who do not provide a practical tool like an Early Warning System or an easily interpretable measure to quantify the risk of crashes, but instead we provide a ready-to-use application of the information from an estimated ETAS model by means of an EWS.

In somewhat more detail, we consider several specifications of the key triggering functions. The parameters of the Hawkes models are estimated by maximum likelihood. And, to judge the fit of the different models, we compare the log-likelihoods and Akaike information criterion (AIC) values. We also develop simulation procedures to graphically assess whether data generated by the models can reproduce features of, for example, the S&P 500 data. The correctness of the ETAS model specification is further evaluated by means of the residual analysis methods as proposed in Ogata (1988). We review the performance of our Early Warning System using the hit rate and the Hanssen-Kuiper Skill Score, and compare it to EWS based on some commonly used and well known volatility models.

The estimation results confirm that crashes are self-enforcing. Furthermore we find that on average larger events trigger more events than smaller events and that larger extremes are observed after the occurrence of more and/or big events than after a tranquil period. Testing our EWS on S&P 500 data during the recent financial crisis, we find positive Hanssen-Kuiper Skill Scores. Thus as our modeling framework exploits the self-exciting behavior of stock returns around financial market crashes, it is capable of creating crash probability predictions on the medium term. Furthermore our modeling framework seems capable of exploiting information in the returns series not captured by the volatility models.

Our paper is organized as follows. In Section 2 the model specifications are discussed, as well as the estimation method. Estimation results are presented in Section 3. Section 4 contains an assessment of the models by means of simulations and residual analysis. The Early Warning Systems are reviewed in Section 5 and compared to EWS based on volatility models in Section 6. Section 7 concludes also with directions for further research.

2 Models

The Epidemic-Type Aftershock Sequence (ETAS) model is a branching model, in which each event can trigger subsequent events, which in turn can trigger subsequent events of their own. The ETAS model is based on the mutually self-exciting Hawkes point process, which is an inhomogeneous Poisson process. For the Hawkes process, the intensity at which events arrive at time t depends on the history of events prior to time t.

Consider an event process (t_1, m_1),...,(t_n, m_n) where t_i defines the time and m_i the mark of event i. Let $\mathcal{H}_t = \{(t_i, m_i) : t_i < t_g\}$ represent the entire history of events up to time t. The conditional intensity of jump arrivals following a Hawkes process is given by

$$\lambda(t|\theta; \mathcal{H}_t) = \mu_0 + \sum_{i:t_i<t} g(t - t_i, m_i) \tag{1}$$

where $\mu_0 > 0$ and $g(s-t_i, m_i) > 0$ whenever $s > 0$ and 0 elsewhere. The conditional intensity consists of a constant term μ_0 and a self-exciting function $g(s)$, which depends on the time passed since jumps that occurred before t and the size of these jumps. The rate at which events take place is thus separated in a long-term background component and a short-term clustering component describing the temporal distribution of aftershocks. The conditional intensity uniquely determines the distribution of the process.

We consider the following specifications of triggering functions

$$g_{pow}(t - t_i, m_i) = \frac{K_0}{(\gamma(t - t_i) + 1)^{1+\omega}} e^{\alpha(m_i - M_0)} \tag{2}$$

$$g_{exp}(t - t_i, m_i) = K_0 e^{-\beta(t-t_i)+\alpha(m_i-M_0)} \tag{3}$$

where K_0 controls the maximum triggering intensity. Furthermore in (3) K_0 covers the expected number of events directly triggered by an event in (3). In (2) the parameter γ controls the expected number of descendants of an event.

The possibility of an event triggering a subsequent event decays according to a power law distribution for (2), while it decays according an exponential distribution for (3). The parameters ω and β determine how fast the possibility of triggering events decays with respectively the time passed since an event. When ω and β are larger, the possibility that an event triggers another event dies out more quickly.

When $\alpha \neq 0$ the intensity at which subsequent events are triggered by an event is influenced by the size of the event. The minimum magnitude of events that are modeled and marked as an event is represented by M_0. How much the size of an event affects the possibility of triggering other events is determined by α. Assuming that larger events trigger more events than smaller events, so that $\alpha > 0$, the possibility of triggering events increases with the size of the excess magnitude of an event ($x = m_i - M_0$). The larger α, the more pronounced is the influence of the size of events.

The process is stationary when the expected number of off springs of an event, that is the branching ratio n, is smaller than 1. When $n \geq 1$ the number of events arriving will grow to infinity over time. The condition for stationarity of the Hawkes process with triggering function (2) and (3) can be stated as respectively

$$\int_0^\infty g_{pow}(t - t_i, m_i)dt = \frac{K_0}{\gamma\omega} < 1 \tag{4}$$

$$\int_0^\infty g_{exp}(t - t_i, m_i)dt = \frac{K_0}{\beta} < 1 \tag{5}$$

As shown in Herrera and Schipp (2009), Chavez-Demoulin, Davison and McNeill (2005) and Chavez-Demoulin and McGill (2012), the sizes of excess magnitude of the events in our model follow a Generalized Pareto Distribution, that is

$$G_{\xi,\sigma}(x) = \begin{cases} 1 - \left(1 + \xi \frac{x}{\sigma(t)}\right)^{-1/\xi} & \xi \neq 0 \\ 1 - e^{-\frac{x}{\sigma(t)}} & \xi = 0 \end{cases}$$

where $\sigma(t) = \phi + \eta \sum_{i:t_i<t} g(t - t_i, m_i)$. We examine models with a constant scale parameter ($\eta = 0$) and a history dependent scale parameter ($\eta \neq 0$). The hypothesis underlying the first class of models states that the sizes of the events are unpredictable, whereas in the second class of models the times and sizes of previous events affect the probability distribution of the sizes of subsequent events. The larger η, the more pronounced is the influence of the history of events on the size of subsequent events. The mean and variance of the distribution of the sizes of excess magnitudes of the events scale with $\sigma(t)$. Therefore when ϕ or η is larger, the events modeled are on average larger and deviate more in size.

We investigate several specifications of the ETAS model. We consider both the power law triggering function (2) and the exponential triggering function (3) with and without influence of the magnitude of events on the triggering of subsequent events and with and without influence of the history of events on the magnitude of subsequent events. In Table 1 it is set out how the different models are configured.

[Table 1 about here.]

While Bacry et al. (2011) use a non-parametric kernel estimation technique for a symmetric Hawkes process on high frequency data, we prefer parametric kernel estimation to make the model more interpretable. We can advocate this technique as the literature is not consistent in which triggering function to use for financial data. A well known stylized fact of the absolute returns is that they decay roughly according to a power law (Cont, 2001). Selçuk and Genlçay (2006), Weber et al. (2007) and Petersen et al. (2010) conclude that the intraday volatility of stock returns above a certain threshold decays roughly according a power-law, approximating the intraday volatility by the absolute returns. However while for example Hardiman et al. 2013, find power law functions fit the S&P 500 data, others report the superior performance of exponential functions (Filimov and Sornette, 2013, among others). We consider both functions.

We estimate the parameters $\theta = \{\mu, K_0, \gamma, \omega, \beta, \xi, \phi, \eta\}$ of the models by maximum likelihood. The log-likelihood of the model is given by

$$\begin{aligned}\log L(\theta) = &\sum_{i=1}^{N} \log \lambda(t_i|\theta; \mathcal{H}_t) - \log \sigma(t) + \left(1 + \frac{1}{\xi}\right) \log \left(1 + \xi \frac{m_i - M_0}{\sigma(t)}\right) \\ &- \int_0^T \lambda(t_i|\theta; \mathcal{H}_t) dt\end{aligned} \tag{6}$$

where $\lambda(t_i|\theta; \mathcal{H}_t)$ is the conditional intensity and t_i are the event arrival times in the interval $[0, T]$. We optimize the log-likelihood numerically using the Nelder-Mead simplex direct search algorithm. The difficulty of accurately estimating the parameters of a Hawkes process has well been recognized in the literature on Hawkes processes [3]. Therefore we exploited different estimation methods and optimization algorithms and test our procedure on simulated data series.

The probability of the occurrence of an event following a Hawkes process with conditional intensity $\lambda(t|\theta; \mathcal{H}_t)$ between t_{n-1} and t_n is given by

$$\begin{aligned}\Pr\left(N(t_n) - N(t_{n-1}) > 0\right) &= 1 - \Pr\left(N(t_n) - N(t_{n-1}) = 0\right) \\ &= 1 - F\left(t^* > t_n - t_{n-1}\right) \\ &= \exp\left(- \int_{t_{n-1}}^{t_n} \lambda(t|\theta; \mathcal{H}_t) dt\right)\end{aligned} \tag{7}$$

Thus, using the conditional intensity (1) specified by the estimated parameters of the ETAS models and the history of the stock returns, we are able to predict the probability of the

[3] See amongst others: Rasmussen, 2013, Filimonov and Sornette, 2013, Hardiman et al., 2013, Aït-Sahalia et al., 2013, Chavez-Demoulin et al., 2012, Bacry et al., 2012, Veen and Schoenberg, 2008

occurrence of an event during a given time period. These probability predictions form the basis of our Early Warning system.

3 Application to Financial Data

We consider data of the S&P 500 index over a period from 2 January, 1957, to 1 September, 2008 to calibrate our models and 5 years thereafter for an out-of-sample evaluation of the models. The dataset consists of daily returns $R_t = \frac{p_t - p_{t-1}}{p_{t-1}} \times 100$, where p_t denotes the value of the index at t. Figure 1 shows the evolution of respectively the S&P 500 index and the returns on this index. Severe drops in the price index ans large negative returns corresponding to these drops, are observed around famous crash periods, "Black Monday" (1987) and the stock market downturn of 2002 after the "dot-com bubble" (1997-2000). Furthermore the Figure illustrates the clustering of extreme returns, that is tranquil periods with small price changes alternate with turbulent periods with large price changes. This clustering feature can be related to the positive herding behavior of investors and the endogenous growth of instability in the financial market.

We apply the ETAS models to the 95% quantile of extreme returns and the 95% quantile of extreme negative returns referred to as extremes and crashes, respectively. The quantiles correspond with 687 events over a time period of $13,738$ trading days. The estimation results are presented in Table 2 and Table 3.

Giving interpretation to the parameter μ, returns above the 95% threshold not triggered by previous extremes, occur on average at a daily rate that ranges from 0.0058 (model A) to 0.0078 (model H). Over the time period approximately 80–107 of the total of 687 events arrived spontaneously according to the models. This means that about 84–88% of the events took place by self-excitement. For the crashes, the mean background intensity of events ranges from 0.0075 (model A, B and E) to 0.0111 (model H), so that about 78–85% of the events are triggered by other events according to the models. Also for both sets of returns the branching ratio (n), that is the expected number of descendants of an event, lies between 0.8 and 0.9 in the models where the magnitude of an event has no influence on the triggering of descendants ($\alpha = 0$). When $\alpha \neq 0$, the branching ratio, depends on the magnitude of the event. However the branching ratio of an event with the threshold magnitude (M_0) lies above 0.7 in the models. Therefore, as $\alpha > 0$, the minimum expected number of descendants of an event in these models lies above 0.7.

We can therefore state that many extreme movements in the S&P 500 index are triggered by previous extreme movements in this index. This does not come as a surprise as the clustering and serial dependence of extremes is a well known feature of stock returns. It confirms our expectation that crashes are local self-enforcing and grow endogenously as events provoke the occurrence of new events.

The ETAS models with a power law triggering function (models A to D) have a higher log-likelihood and a lower AIC value, than their counterparts with an exponential triggering function (models E to H) for both sets of returns. The decay of the triggering probability seems slower than exponential for our data. When the estimate for ω is large or not significant, this indicates that other distributions like the exponential or hyperbolic distribution can be more appropriate.

Comparing the ETAS models with the parameter restriction $\alpha = 0$ to the ETAS models without this restriction, the magnitude of an extreme has a significant positive influence on the probability of triggering another extreme for both sets of returns. This means that on average larger events trigger more events than smaller events. The models B, D, F and H have a higher ranking both in terms of log-likelihood as in AIC value than their counterparts with $\alpha = 0$, that is model A, C, E and G respectively. Incorporating the size of the events into an ETAS model for the extreme (negative) returns thus improves the model.

The estimates for η in the models C, D, G and H are positive and significant for both sets of returns. The models score better in both log-likelihood and AIC value than the models A, B, E and F. This suggests a model which incorporates the history of the event process to prospect the sizes of subsequent events, matches the extreme (negative) returns closer than a model which assumes the sizes of events are independent of the past. When $\eta > 0$, the mean and variance of the distribution of the excess magnitudes of the events scale with the value of the cumulative triggering function, and thus the probability of the arrival of an event triggered by another event. This means that on average larger extremes are observed after the occurrence of more and/or big events than after a tranquil period.

A likelihood ratio test shows that all the estimated parameters of the models are significant at a 5% level. All together model D with a power law triggering function, a non-zero influence of the size of the events on the triggering of subsequent events and predictable event sizes, fits best according to the log-likelihoods and AIC values for both the extremes as the crashes.

Figure 2 presents the intensity with which extremes occur estimated with model D, over the estimation period, that is from 2 January, 1957, to 1 September, 2008. The estimated intensity shows large spikes around the famous crash periods, "Black Monday" (1987) and the "stock market downturn of 2002" (2002) after the "dot-com bubble" (1997-2000). As expected, the rate at which events arrive is high around crashes, reflecting the increase in the triggering probability after the occurrence of events.

[Figure 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

[Figure 2 about here.]

4 Goodness-of-fit

4.1 Simulation

To check whether our estimated models can reproduce features of the extreme (negative) returns we develop two different simulation procedures and compare their generated data with the observed data. While in the first procedure the probability of occurrence of an event is used to realize a series of events in discrete time, the second procedure is carried out in continuous time employing the branching structure of the ETAS model. In the first procedure events can occur at a daily frequency. In the second procedure event times are not integers and multiple events can occur during one day. As the first procedure seems to resemble the data generating process more closely, we only discuss results from this procedure. Both procedures can be found in the appendix.

We generate 1000 data series from the models using the parameters estimates derived from the extreme negative returns on the S&P 500 index (Table 3). We set the sample period equal to the number of trading days over which we estimated the models for the S&P 500 crashes. Estimation results for these series are shown in Table 4. One thing that stands out is the estimation results of the ETAS models with a power law triggering function (models A to D) are not so satisfactory. The maximum likelihood estimation does not converge in a number of simulations. Furthermore the estimated $\widehat{\omega}$ of the triggering functions deviate much from the ω used to simulate the data and the standard deviations of the $\widehat{\omega}$ are much larger than the standard deviation of $\widehat{\omega}$ derived from the crashes. The estimates for ω derived from data series generated with a continuous time procedure are much closer to values used to simulate the series. Also the standard deviations of these $\widehat{\omega}$ are much smaller.

We have examined several methods to simulate and estimate the ETAS model with the power law triggering function. When estimating the models, the Expectation-Maximization procedure of Veen and Schoenberg (2008), the Bayesian procedure of Rasmussen (2011) and gradient-based optimization algorithms give inferior results in terms of speed and robustness for our kind of data. The estimated ETAS models with the exponential triggering function (models E to H) appear more reliable.

In Figure 3 the S&P 500 crashes are compared to a series simulated with the discrete time procedure from model D (power law triggering function) and H (exponential triggering function), the ETAS models with a non-zero influence of the magnitude on the triggering of subsequent events and a non-zero influence of the history of the event process on the sizes of subsequent events. Models D and H have the highest log-likelihoods and lowest AIC values amongst the models, when applied to the crashes. The simulated series share the major features characteristic to the models and similar to the crashes like the clustering of events, heavy-tailed distributed event sizes, and large events are especially observed after the occurrence of more and/or other big events. Histograms show the data simulated with the discrete time procedure differs from the S&P 500 data as much less event pairs are observed with an inter event time of 1 or 2 days. Examining plots of the logarithm of the cumulative number of events against the logarithm of time, model H seems to match the crashes more closely than model D. However, when looking at the figures that show the magnitudes and times of events, the clustering of the S&P 500 crashes is more similar to the clustering observed when events are simulated from model D.

[Table 4 about here.]

[Figure 3 about here.]

4.2 Residual analysis

We also assess the goodness-of-fit of our models using the residual analysis technique of Ogata (1988). This method states that if the event process $\{t_i\}$ is generated by the conditional intensity $\lambda(t)$, the transformed times

$$\tau_i = \int_0^{t_i} \lambda(t)dt \tag{8}$$

are distributed according a homogeneous Poisson process with intensity 1. Furthermore the transformed interarrival times, that is

$$\tau_i - \tau_{i-1} = \int_{t_{i-1}}^{t_i} \lambda(t)dt \tag{9}$$

are independent exponential random variables with mean 1. If the models are correctly specified, $\lambda(t)$ can be approximated by $\lambda(t|\widehat{\theta}; \mathcal{H}_t)$. The sequence $\{\tau_i\}$ is called the residual process. In order to verify whether the residual process derived from the models is Poisson with unit intensity, we perform the Kolmogorov-Smirnov (KS) test. The null hypothesis of our test is that the distribution of the residual process and the unit Poisson distribution are equal.

The p-values of the KS tests are reported in Table 5. Figure 4 shows the cumulative number of S&P 500 crashes against the transformed times derived from model D and model H together with the 95% and 99% error bounds of the KS statistic. The p-values and the Figure indicate that the extreme (negative) returns do not deviate from an event process specified by an ETAS model at a 5% level. However the p-values belonging to the power law models are higher than the p-values of the exponential models. The models with a non-zero influence of the magnitude of events on the triggering of events and the models with a zero influence of the history of the event process on the magnitude of events have higher p-values than their counterparts. Overall, p-values are higher for the extremes than for the crashes. This suggests that ETAS models are more suitable for the extremes than for the crashes, whereas model B fits both sets of returns the best.

[Table 5 about here.]

[Figure 4 about here.]

5 Forecasting

5.1 Early Warning System

The identification of financial market crashes is of great importance to traders, regulators of financial markets and risk management. They can benefit from an Early Warning System that sets an alarm when the probability of a crash becomes too high, urging the traders, regulators and risk managers to take action. We develop an Early Warning System for extremes and crashes in the financial market within a certain time period using the conditional intensity specified by the estimated parameters of the ETAS models and the history of the stock returns. The probability of an extreme or a crash occurring between t_{n-1} and t_n is given by (7). To evaluate the performance of the EWS, we use measures reported in Candelon et al. (2012). We do not compute the optimal threshold value for giving an alarm. Instead we set the threshold at 0.5. Therefore an alarm is given when the models predict that it is more likely that at least one event occurs than that no event occurs within a certain time period. Here we consider the occurrence of events within a time period of 5 days during the last few years, that is from 1 September, 2008, to 1 January, 2013, and during the recent financial crisis, that is from 1 September, 2008, to 1 January, 2010.

To compare the probability predictions made by the different models, we compute the Quadratic Probability Score (QPS) and the Log Probability Score (LPS) for each model, that is

$$QPS = \frac{2}{T} \sum_{t=1}^{T} (\hat{p}_t - y_t)^2 \tag{10}$$

$$LPS = -\frac{1}{T} \sum_{t=1}^{T} \left[(1 - y_t) \log (1 - \hat{p}_t) + y_t \log (\hat{p}_t)\right] \tag{11}$$

where p_t represents the predicted probability of crash and y_t is an indicator function taking the value one when a crash occurs and the value zero otherwise. The QPS and LPS range respectively from 0 to 1 and from 0 to ∞, with 0 indicating perfect accuracy.

When the QPS or the LPS are higher, the probability predictions deviate more from a binary variable indicating the occurrence of events. The LPS punishes large deviations heavier than small deviations. The QPS and the LPS are displayed in Table 6. LPS of the models with a size-dependent triggering probability are slightly lower. QPS and LPS of the models with an exponential triggering function are slightly lower during the financial crisis of 2008 than the QPS and LPS of the models with a power law triggering function. The probability predictions for crashes seem somewhat less accurate than the probability predictions for the extremes as the QPS and LPS of the models for crashes are higher than the QPS and LPS of the models for extremes.

Furthermore we compare the accuracy of the probability predictions made by the different models. For nested models, that is the models with the same triggering function with and without influence of the size of the events on the triggering probability, we compute the adjusted Mean Squared Prediction Error (adjusted MSPE). According to the adjusted

MSPEs the probability predictions based on models with a size-dependent triggering probability are significantly more accurate from 1 September, 2008, to 1 January, 2013. For the non-nested models, that is the models with a different triggering function, we compute the Diebold Mariano (DM) statistic. According to the DM statistics the models with an exponential triggering function predict the probability on crashes significantly more accurate during the financial crisis of 2008 than the models with a power law triggering function.

Table 9 reports the number of correct predictions, the number of false predictions, the hit rate, the false alarm rate and the Hanssen-Kuiper Skill Score of the EWS. As a reference, no events are predicted by an EWS using a homogeneous Poisson model. Here the intensity of the Poisson process is set equal to the number of events during the forecast period divided by the length of the forecast period. The Hanssen-Kuiper Skill Score (KSS) is computed as the hit rate minus the false alarm rate. The KSS of all EWS are positive, meaning that the rate of correct predictions is higher than the rate of false predictions, whereas KSS of the Poisson-EWS is zero. The KSS of models with a size-dependent triggering probability are slightly higher. The KSS of the models with an exponential triggering function are slightly higher during the financial crisis of 2008 than the KSS of the models with a power law triggering function.

Figure 5 shows the predicted probability of a financial market crash occurring within 5 days based on the models C and G, from 1 September, 2008, to 1 January, 2013. The Figure shows that during the financial crisis of 2008 the crash probability is exceptionally high. The crash probability also becomes high when the stock market falls at the end of July/the beginning of August, 2011. Comparing the crash probabilities based on the different models, the model with the power law triggering function (model C) forecasts a higher probability on a crash than the model with the exponential triggering function (model G). Furthermore during the financial crisis of 2008, the risk of a crash decays more slowly according to model C than to G.

[Table 6 about here.]

[Table 7 about here.]

[Table 8 about here.]

[Table 9 about here.]

[Figure 5 about here.]

6 Comparison volatility models

We are interested in whether our models are capable of getting information out of stock market data on future crashes or extremes, not captured by commonly used and well known volatility models. In order to assess whether this is the case, we compare the performance of the Early Warning Systems based on the ETAS models with EWS based on GARCH-type and ACD-type models. For the EWS we again consider the occurrence of events within a time period of 5 days during the last few years, that is from 1 September, 2008, to 1 January, 2013.

In General AutoRegressive Conditional Heteroskedasticity (GARCH) type models, the variance of the current error term is a function of the error terms and innovations in previous periods. The time-varying conditional variance enables the models to capture the volatility clustering feature of stock market returns. After evaluating the performance of several GARCH-type models and error distributions we consider two GARCH-type models, that is the GARCH(1,1) model and the GJR(1,1) model, in combination with a Student-t distribution for the error terms. The heavy-tailed Student-t distribution accounts for the stylized fact that, even after correcting for volatility clustering, extreme returns occur more often than under normality (Cont, 2001). As shown by Hansen and Lunde (2005), it is difficult to beat the GARCH(1,1) model when forecasting conditional volatility. The GJR(1,1) model of Glosten, Jagannathan and Runkle differs from the GARCH(1,1) model, as the model allows for separate influences of positive and negative innovations on future volatility. This asymmetric response to shocks, or so-called "leverage effect", is in line with the observation that measures of the volatility of assets tend to correlate negatively with the returns on those assets (Cont, 2001). This can be of an advantage when modelling returns (see Hua and Mansan, 2013). The conditional variance σ_t^2 in respectively the GARCH(1,1) and the GJR(1,1) model is specified as follows

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{12}$$

$$\sigma_t^2 = \omega + \left(\alpha\epsilon_{t-1}^2 + \gamma\mathcal{I}\left[\epsilon_{t-1} < 0\right]\right)\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{13}$$

Autoregressive Conditional Duration type models, see Engle and Russell (1998), focus on modeling the expected duration between events. As well as in the ETAS models, the conditional intensity in these models is a function of the time between past events. Futhermore the event process on which the models are based is self-exciting. Analogous to GARCH-type and Hawkes models, ACD-type models are therefore able to pick up the characteristic clustering of extreme stock market returns. After evaluating the performance of several

ACD-type models and error distributions we consider two ACD-type models, that is the ACD(1,1) model and the log-ACD(1,1) model, in combination with a Generalized Gamma distribution. Like Bauwens et al. (2004) and Allen et al. (2009), we find other ACD-type models much more costly to estimate and to evaluate, but not of superior performance, while the use of the Generalized Gamma distribution instead of the epxonential or Weibull distribution does add to the performance of the models. A survey on ACD-type models is provided by Pacurar (2008).

In the regular ACD model the expected duration is a linear function of past durations and conditional durations. The logarithmic version of the ACD model implies a nonlinear relation between the variables, which guarantees positive durations without imposing restrictions on the parameters. The duration in the ACD(1,1) and the log-ACD(1,1) model is given by $\tau_t = \theta_t z_t$, where z_t independent and identically distributed according to the Generalized Gamma distribution such that $\mathbb{E}[z_t] = 1$. In respectively the ACD(1,1) and the log-ACD(1,1) model is θ_t is specified as follows

$$\theta_t = \omega + \alpha\tau_{t-1} + \beta\theta_{t-1} \tag{14}$$

$$\log(\theta_t) = \omega + \alpha\log(\tau_{t-1}) + \beta\log(\theta_{t-1}) \tag{15}$$

We apply the models again for daily return data on the S&P 500 index between 2 January, 1957, and 1 September, 2008. Like the ETAS models, the volatility models are estimated by maximum likelihood in combination with the Nelder-Mead simplex direct search algorithm. Consulting Andersen, Bollerslev et al. (2006) and Lau and McSharry (2010) for forecasting multi-step ahead densities with GARCH-type models, we use Monte Carlo methods to derive the probability of the occurence of no event within a certain period. For the ACD-type models this probability follows easily from the model specification, by noticing it is equal to the probability of the duration exceeding the time period.

6.1 In sample results

In order to evaluate the ability of ETAS models to exploit information in the returns series not captured by GARCH-type models, we apply the ETAS models to the standardized residuals from the GARCH-type models. The results of this exercise are reported in 10 and 11. As estimated parameters are significant, our evidence for ETAS models is not simply driven by volatility clustering. A likelihood ratio test shows that only the α parameter is not significant at a 5% level, such that the size of previous extreme residuals does not seem to influence the probability of the arrival of new extremes. As the η parameter is significantly different from zero at a 5% level, and the models with a non-zero η parameter have a higher log-likelihood and a lower AIC value than their counterparts with $\eta = 0$, the size of future extreme residuals appears to be affected by the history of the event process.

We check the goodness-of-fit of the ETAS models by means of residual analysis, and verify whether the distribution of the transformed times is unit Poisson with the Kolmogorov-Smirnov (KS) test. The p-values of the KS tests are reported in Table 12. The (negative) extremes do not deviate from an event process specified by a ETAS model at a 1% level. At a 5% level the ETAS models are only inappropriate for negative extremes from the GARCH(1,1) model at the 95% and 96% quantile. Not surprisingly the models without influence of the magnitude of events on the triggering of events have slightly higher p-values than their counterparts. Overall, p-values are higher for absolute extremes than for negative extremes. Also p-values are higher for the standardized residuals from the GJR(1,1) model than for the standardized residuals from the GARCH(1,1) model, especially at higher quantiles. This suggests ETAS models are most suitable for absolute extremes of the standardized residuals from the GJR(1,1) model.

[Table 10 about here.]

[Table 11 about here.]

[Table 12 about here.]

6.2 Out-of-sample results

Tables 13 and 14 report the results of the EWS based on the GARCH-type, the ACD-type and two ETAS models for the extreme (negative) returns above the 95–99% in sample quantile. It is immediately apparent from the tables that the ACD-type models do not perform well. While the EWS based on the the ACD(1,1) model predicts the occurrence of an event almost every period, the EWS based on the log-ACD(1,1) model predicts far too few events over the out-of-sample period. This results in a KSS around zero (the KSS is even slightly negative for some model-quantile combinations).

In constrast to the ACD-type models, both GARCH-type models are capable of delivering accurate warning signals. Overall the GJR(1,1) does slightly better than the GARCH(1,1) in terms of their KSS. Compared to the ETAS models the KSS of the EWS for crashes are slightly higher. However the EWS based on the ETAS model outperform the EWS based on the GARCH-type models for extremes, especially at higher quantiles. Examining whether using information on positive extreme returns improves forcasting the occurrence of negative extreme returns with ETAS models, our findings are in line with Embrechts et al. (2011). Negative extremes do help to forecast positive ones, however this does not apply the other way around.

However forecasting with the GARCH-type models is much more time consuming than forecasting with the other models. Using these models, the probability distribution of the occurence of one or more events during an certain time period, has to be derived empirically by means of a Monte Carlo procedure. Roughly about one to two hours are needed to execute the Monte Carlo simulation and deduce alarm signals over the out-of-sample period (when using $10,000$ replications), while with EWS based on ETAS models, it takes no more than half a second to compute forecasts.

[Table 13 about here.]

[Table 14 about here.]

7 Conclusion

This paper explores similarities between stock returns during a financial market crash and earthquakes to make predictions of the probability of a crash in the financial market. We provide a ready-to-use application of this information by means of an Early Warning System.

The basis of the models examined is the self-exciting Hawkes point process. The rate at which events arrive is separated in a long-term background component and a short-term clustering component describing the temporal distribution of triggered events.

The models are applied to the 95% quantile of extreme (negative) returns on the S&P 500 index over a period from 2 January, 1957, to 1 September, 2008. The estimation results confirm that like earthquakes, crashes are self-enforcing. The decay of the probability of triggering events seems better modeled by the power law distribution than by the exponential distribution. The triggering probability is size-dependent, as larger events trigger on average more events than smaller events. The sizes of events are history dependent, as on average larger extremes are observed after the occurrence of more and/or big events than after a tranquil period.

Simulated series have the major features that are characteristic to the models and similar to the extreme (negative) returns like the clustering of events, heavy-tailed distributed event sizes, and that the large events are especially observed after the occurrence of more and/or other big events. Furthermore performing residual analysis, we find that the extreme (negative) returns do not significantly deviate from an event process specified by a Hawkes model.

We develop an Early Warning System for events in the financial market based on the probability of the occurrence of an event within a certain time period predicted by the models. These are reviewed from 1 September, 2008, to 1 January, 2010, and from 1 September, 2008, to 1 January, 2013. Testing the EWS, the rate of correct predictions is higher than the rate of false predictions. Thus as our modeling framework exploits the self-exciting behavior of stock returns around financial market crashes, it is capable to create crash probability predictions on the medium term. The models with an exponential triggering function and a non-zero influence of the size of events on the triggering probability perform best according to the Hanssen-Kuiper Skill Score.

From 1 September, 2008, to 1 January, 2013, we also consider EWS based on some commonly used and well known volatility models. While the ACD models do not perform well, the GARCH models are, like our models, capable of delivering accurate warning signals. However our models outperform the GARCH models for extremes, especially at higher quantiles. Moreover, forecasting with GARCH models is much more time consuming, taking over a hour compared to less than half a second using our models.

In order to further evaluate the ability of our models to exploit information in the returns series not captured by GARCH models, the models are applied to the standardized residuals from the GARCH models. The significance of the parameters indicates that GARCH-models do not completely capture the self-exciting behavior of crashes. Moreover, checking the goodness-of-fit of the models by means of residual analysis, we find that our models are appropriate for modeling the extreme residuals from the GARCH models.

We indicate four directions for further research. The first is the application of the models to high-frequency stock market data. The second is a multivariate extension of the models as events tend to occur simultaneously in financial markets. Furthermore, the models could benefit from the addition of a time-varying exogenous component to the conditional intensity. This allows the models to incorporate information of precursors of financial market crashes. Lastly, models with a regime-switching conditional intensity could match the data more closely.

8 References

[1] Aït-Sahalia, Y., Cacho-Diaz, J., and Laeven, R. (2013). Modeling financial contagion using mutually exciting jump processes. NBER Working paper.

[2] Allen, D., Lazarov, Z., McAleer, M., and Peiris, S. (2009). Comparison of alternative ACD models via density and interval forecasts: Evidence from the Australian stock market. *Mathematics and Computers in Simulation*, 79(8): 2535–2555.

[3] Andersen, T. G., Bollerslev, T., Christoffersen, P. F., and Diebold, F. X., (2006). Volatility and correlation forecasting. In *Handbook of Economic Forecasting*, 777–878. North-Holland Amsterdam.

[4] Bacry, E., Dayri, K., and Muzy, J. F. (2012). Non-parametric kernel estimation for symmetric Hawkes processes. application to high frequency financial data. *The European Physical Journal B*, 85(5): 157.

[5] Balderama, E., Schoenberg, F., Murray, E., and Rundel, P. (2012). Application of branching models in the study of invasive species. *Journal of the American Statistical Association*, 107(498): 467–476.

[6] Baldovin, F., Bovina, D., Camana, F., and Attilio, A. (2011). Modeling the non-Markovian, non-stationary scaling dynamics of financial markets. *Econophysics of order-driven markets*, 1: 239–252.

[7] Baldovin, F., Camana, F., Caporin, M., and Attilio, A. (2012a). Ensemble properties of high frequency data and intraday trading rules. *Available at SSRN 2001981.*

[8] Baldovin, F., Camana, F., Caraglio, M., Stella, A., and Zamparo, M. (2012b). Aftershock prediction for high-frequency financial markets dynamics. In *Econophysics of Systemic Risk and Network Dynamics*, 49–58. Springer Milan.

[9] Bauwens, L., Giot, P., Grammig, J., and Veredas, D. (2004). A comparison of financial duration models via density forecasts. *International Journal of Forecasting*, 20(4): 589–609.

[10] Bauwens, L., and Hautsch, N. (2009). Modeling financial high frequency data using point processes. In *Handbook of Financial Time Series*, 953–979. Springer Berlin Heidelberg.

[11] Bormetti, G., Calcagnile, L., Treccani, M., Corsi, F., Marmi, S., and Lillo, F. (2013). Modeling systemic cojumps with Hawkes factor models. *arXiv preprint arXiv:1301.6141*.

[12] Bowsher, C. (2007). Modeling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141: 876–912.

[13] Candelon, B., Dumitrescu, E., and Hurlin, C. (2012). How to evaluate an Early Warning System? *IMF Economic Review*, 60(1): 75–113.

[14] Chavez-Demoulin, V., Davison, A., and McNeill, A. (2005). Estimating value-at-risk: a point process approach. *Quantitative Finance* 5(2): 227–234.

[15] Chavez-Demoulin, V., and McGill, J. (2012). High-frequency financial data modeling using Hawkes processes. *Journal of Banking & Finance*, 36: 3415–3426.

[16] Clark, T., West, K., (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics* 138: 291–311.

[17] Clements, A., and Liao, Y. (2013). The dynamics of co-jumps, volatility and correlation. NCER Working paper.

[18] Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1: 223–236.

[19] Diebold, F., and Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13: 253–265.

[20] Embrechts, P., Liniger, T., and Lin, L. (2011). Multivariate Hawkes processes: an application to financial data. *Journal of Applied Probability*, 48: 367-378.

[21] Engle, R. F., and Russell, J. R. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, 1127–1162.

[22] Filimonov, V., and Sornette, D. (2013). Apparent criticality and calibration issues in the Hawkes self-excited point process model: application to high-frequency financial data. *arXiv preprint arXiv:1308.6756*.

[23] Grothe, O., Korniichuk, V., and Manner, H., (2014). Modeling multivariate extreme events using self-exciting point processes. Forthcoming in *Journal of Econometrics*

[24] Lau, A., and McSharry, P., (2010). Approaches for multi-step density forecasts with application to aggregated wind power. *The Annals of Applied Statistics* 4(3): 1311–1341.

[25] Hansen, P. R., and Lunde, A., (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, 20, 873–889.

[26] Hardiman, S. J., Bercot, N., and Bouchaud, J. P. (2013). Critical reflexivity in financial markets: a Hawkes process analysis. *Physical Review E*, 85(5): 056108.

[27] Hardebeck, J., Felzer, K., and Michael, A. (2008). Improved tests reveal that the accelerating moment release hypothesis is statistically insignificant. *Journal of Geophysical Research*, 113: B08310.

[28] Herrera, R. and Schipp, B. (2009). Self-exciting extreme value models for stock market crashes. In *Statistical Inference, Econometric Analysis and Matrix Algebra*, 209–231. Physica-Verlag HD.

[29] Hewlett, P., (2006). Clustering of order arrivals, price impact and trade path optimization. In *Workshop on Financial Modeling with Jump processes, Ecole Polytechnique*.

[30] Helmstetter, A., and Sornette, D. (2002). Sub-critical and super-critical regimes in epidemic models of earthquake aftershocks, *Journal of Geophysical Research*, 107(B10): 2237.

[31] Hua, J., and Mansan, S. (2013). Forecasting the return distribution using high-frequency volatility measures, *Journal of Banking & Finance*, 37(11): 4381–4403.

[32] Mohler, G., Short, M., Brantingham, P., Schoenberg, F., and Tita, G. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493): 100–108.

[33] Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401): 9–27.

[34] Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2): 379–402.

[35] Petersen, A., Wang, F., Havlin, S., and Stanley, H. (2010). Market dynamics immediately before and after financial shocks: quantifying the Omori, productivity and Bath laws. *Physical Review*, 82: 036114.

[36] Pacurar, M. (2008). Autoregressive conditional duration models in finance: a survey of the theoretical and empirical literature. *Journal of Economic Surveys*, 22(4): 711–751.

[37] Rasmussen, J. G. (2013). Bayesian inference for Hawkes' processes. *Methodology and Computing in Applied Probability*, 15(3): 623–642.

[38] Saichev, A., Helmstetter, A., and Sornette, D. (2005). Power-law distributions of offspring and generation numbers in branching models of earthquake triggering. *Pure and Applied Geophysics*, 162: 1113–1134.

[39] Sornette, D. (2003). Critical market crashes. *Physics Reports*, 378: 1–98.

[40] Selçcuk, F., and Gençcay, R. (2006). Intraday dynamics of stock market returns and volatility. *Physica A*, 367: 375–387.

[41] Veen, A., and Schoenberg, F., (2008). Estimation of space-time branching process models in seismology using an EM-type algorithm. *Journal of the American Statistical Association*, 103(482): 614–624.

[42] Weber, P., Wang, F., Vodenska-Chitkushev, I., Havlin, S., and Stanley, H. (2007). Relation between volatility correlations in financial markets and Omori processes occurring on all scales. *Physical Review*, 76: 016109.

[43] Zhuang, J., Ogata, Y., and Vere-Jones, D. (2002). Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, 97(458): 369–380.

[44] Zhuang, J. and Ogata, Y. (2004). Analyzing earthquake clustering features by using stochastic reconstruction. *Journal of Geophysical Research*, 109: B05301.

A Simulation procedures

Discrete time procedure

1. Simulate the time till the first event from an exponential distribution with parameter μ. As no other events are present yet the occurrence of the first event is Poisson distributed with μ, the constant background rate. The time of the event t_1 recorded is the end of the interval in which the event occurs. Simulate the magnitude of the event from an independent General Pareto Distribution.

2. For t_n after t_1 calculate the probability of the occurrence of no event in the interval $[t_{n-1}, t_n]$, that is (7). Simulate a random number u from a uniform distribution on the interval $[0, 1]$. When $u > P\left(N(t_n) - N(t_{n-1}) = 0\right)$ record the time point t_n as the time of an event and simulate the magnitude of the event from a General Pareto Distribution. When $u < P\left(N(t_n) - N(t_{n-1}) = 0\right)$ do nothing.

3. Repeat for all time points after t_1 till $t_n = T$.

Continuous time procedure

1. Sample the background events

   (a) Simulate the number of background events, N_{back} as $N_{back} = \mu \times T$.

   (b) Simulate the times of the background events $\mathbf{t}_{back}$ as random numbers between zero and T. That is $\mathbf{t}_{back} = \mathbf{u} \times T$, where $\mathbf{u}$ is a $(N_{back} \times 1)$-vector containing random numbers from a uniform distribution on the interval $[0, 1]$.

   (c) Simulate the magnitudes of background events $\mathbf{m}_{back}$ from an independent General Pareto Distribution.

2. Sample the triggered events

   (a) Simulate the number of triggered events N_{off} from a Poisson distribution with an intensity given by the mean number of children of a parent event following that event in the given time window. The mean number of children of a parent is equal to the integrated triggering function, that is

$$G(T - t_{parent}, m_{parent}) = \frac{K_0}{\gamma\omega}\left(1 - \left(\gamma(T - t_{parent}) + 1\right)^{-\omega}\right) e^{\alpha(m_{parent} - M_0)} \quad (16)$$

$$G(T - t_{parent}, m_{parent}) = \frac{K_0}{\beta}\left(1 - e^{-\beta(T - t_{parent})}\right) e^{\alpha(m_{parent} - M_0)} \quad (17)$$

for the power law triggering function and the exponential triggering function respectively.

(b) Simulate the times of these events $\mathbf{t}_{off}$ from the following distribution

$$P\left(t^* \leq t_0 + t | t^* > t_0, t^* < T\right) = \frac{F(t_0 + t) - F(t_0)}{F(T) - F(t_0)} = \frac{S(t_0) - S(t_0 + t)}{S(t_0) - S(T)} \tag{18}$$

where S_t is the survival function of the hazard model, that is $S(t) = 1 - F(t)$. The probability of no event in the interval $[0, t]$ for events distributed according a homogeneous Poisson process is given by $P(N(t) = 0) = F(t^* > t) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t}$. The Hawkes process is an inhomogeneous Poisson process, where the intensity of the process between 0 and t is not constant, so that λt has to replaced with $G(t - t_i, m_i) = -\int_0^t \sum_{i:t_i<t^*} g(t^* - t_i, m_i) dt^*$.

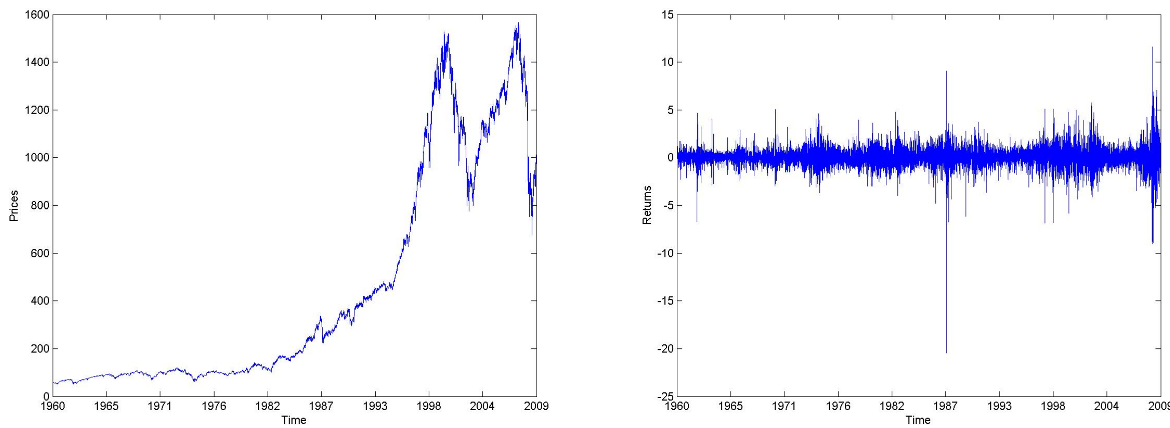
(c) Simulate the magnitudes of these events $\mathbf{m}_{off}$ from an independent General Pareto Distribution.

(d) Repeat the simulation of triggered events till $N_{off} = 0$.

List of Figures

Figure 1: S&P 500 index

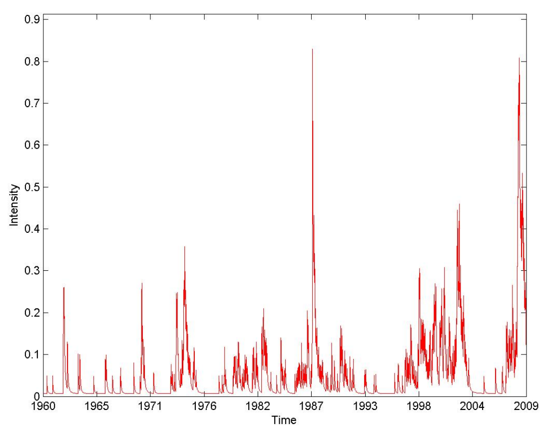
(a) Prices **(b)** Returns



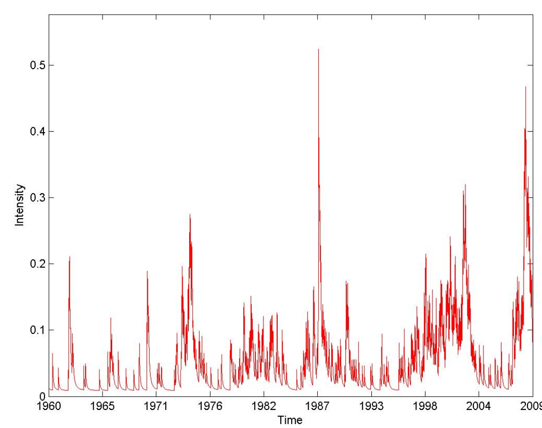
Evolution of the S&P 500 index prices and returns over the period January 2, 1957, until September 1, 2008

Figure 2: Conditional intensity model D

(a) Extremes



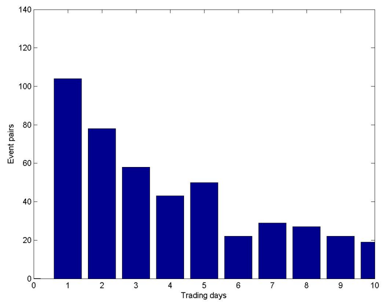
(b) Crashes



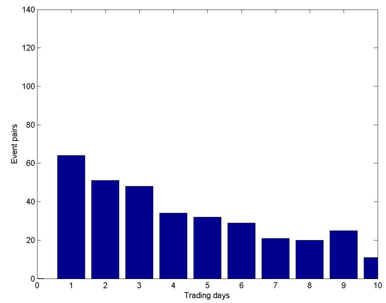
Estimated conditional intensity for the 95% quantile of daily extreme (negative) returns over the period January 2, 1957, until September 1, 2008, using model D

Figure 3: S&P 500 crashes and series simulated in discrete time

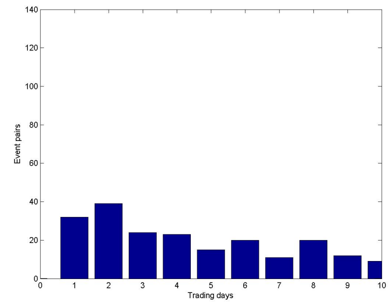
(a) S&P 500 crashes



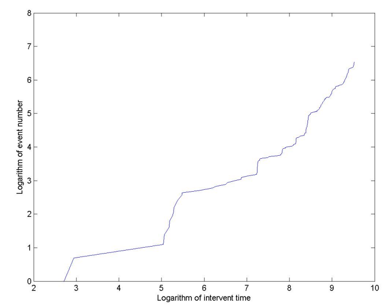
(b) Simulated events model D



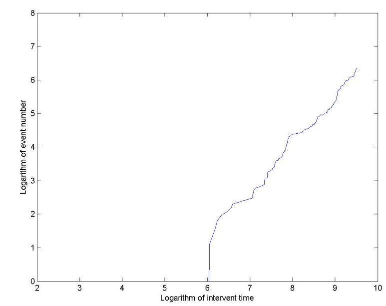
(c) Simulated events model H



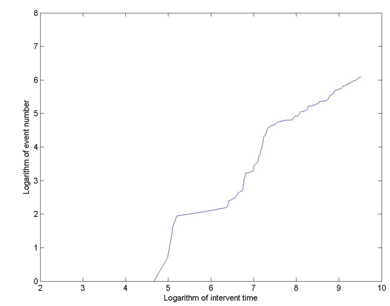
(d) S&P 500 crashes



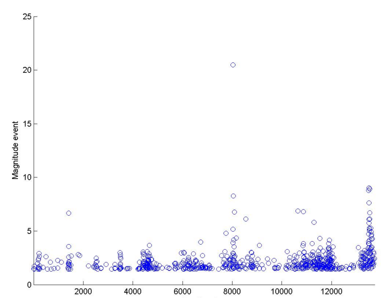
(e) Simulated events model D



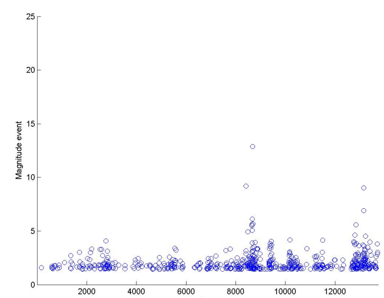
(f) Simulated events model H



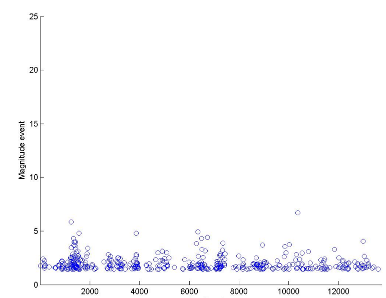
(g) S&P 500 crashes



(h) Simulated events model D



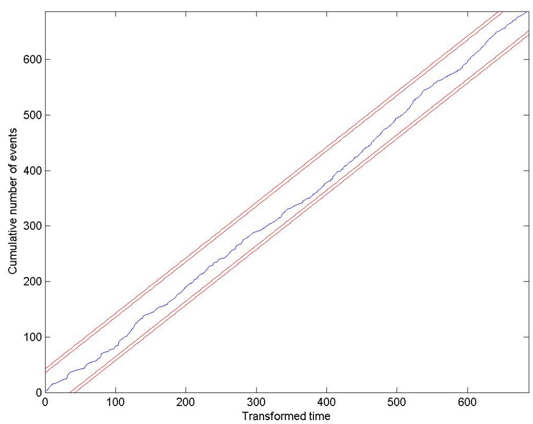
(i) Simulated events model H



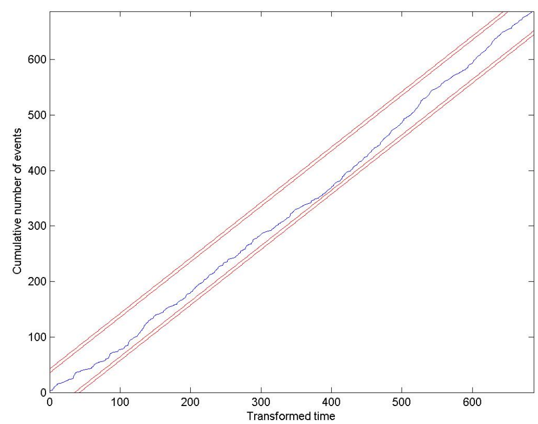
The S&P 500 crashes are shown together with a data series from model D and model H generated with the discrete time procedure. Histograms of times between events, plots of the logarithm of the cumulative number of events against the logarithm of time, and figures in which the magnitudes and times of the events are presented in this Figure.

Figure 4: Residual analysis for the S&P 500 crashes

(a) Model D



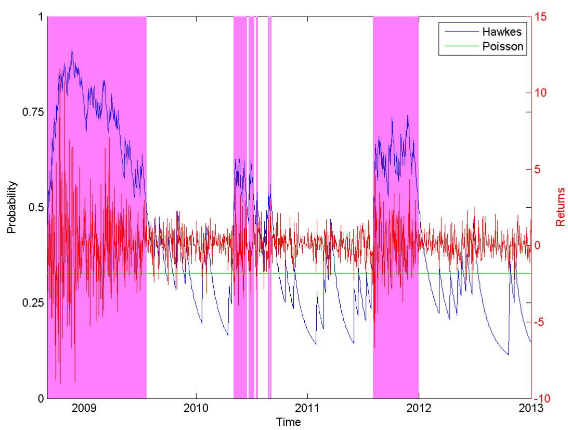
(b) Model H



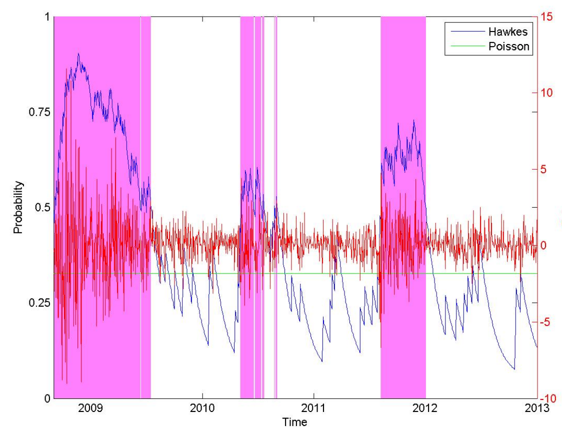
Cumulative number of events against the transformed time $\{\tau_i\}$. The red lines indicate the 95% and 99% error bounds of the Kolmogorov-Smirnov statistic.

Figure 5: Crash probability predictions

(a) Model C



(b) Model G



The probability of the occurrence of a crash within 5 days is predicted according to the models C and G (blue lines) from 1 September, 2008, to 1 January, 2013. When this probability is larger than 0.5 the background is shaded. The green line corresponds to the predicted probability of the occurrence of a crash within 5 days according to a homogeneous Poisson model with intensity equal to the number of crashes during the forecast period divided by the length of the forecast period.

List of Tables

Table 1: Specification ETAS models

Model	A	B	C	D	E	F	G	H
Triggering function	Power law				Exponential			
Influence magnitude events	$\alpha = 0$	$\alpha \neq 0$	$\alpha = 0$	$\alpha \neq 0$	$\alpha = 0$	$\alpha \neq 0$	$\alpha = 0$	$\alpha \neq 0$
Influence event history	$\eta = 0$	$\eta = 0$	$\eta \neq 0$	$\eta \neq 0$	$\eta = 0$	$\eta = 0$	$\eta \neq 0$	$\eta \neq 0$

Influence magnitude on triggering subsequent events is zero when α is restricted to 0. Influence event history on magnitude events is zero when η is restricted to 0.

Table 2: Estimation results extremes

Model	A	B	C	D	E	F	G	H
μ	0.0058	0.0060	0.0061	0.0063	0.0067	0.0071	0.0073	0.0078
	(0.0012)	(0.0012)	(0.0012)	(0.0013)	(0.0011)	(0.0011)	(0.0011)	(0.0012)
K_0	0.0441	0.0381	0.0522	0.0446	0.0380	0.0324	0.0440	0.0368
	(0.0057)	(0.0052)	(0.0065)	(0.0058)	(0.0041)	(0.0039)	(0.0045)	(0.0041)
γ	0.0185	0.0216	0.0269	0.0331				
	(0.0103)	(0.0111)	(0.0128)	(0.0144)				
ω	2.6687	2.2918	2.1865	1.7983				
	(1.3123)	(1.0314)	(0.9080)	(0.6726)				
α		0.1292		0.1435		0.1237		0.1388
		(0.0248)		(0.0181)		(0.0257)		(0.0186)
β					0.0436	0.0430	0.0513	0.0507
					(0.0047)	(0.0047)	(0.0053)	(0.0053)
ξ	0.2602	0.2602	0.1327	0.1157	0.2602	0.2602	0.1364	0.1202
	(0.0470)	(0.0470)	(0.0383)	(0.0366)	(0.0470)	(0.0470)	(0.0385)	(0.0369)
ϕ	0.6956	0.6956	0.3650	0.3870	0.6957	0.6957	0.3758	0.3974
	(0.0415)	(0.0415)	(0.0363)	(0.0364)	(0.0415)	(0.0415)	(0.0360)	(0.0362)
η			0.1478	0.1142			0.1207	0.0911
			(0.0249)	(0.0205)			(0.0181)	(0.0149)
$\log L(\theta)$	-2859.41	-2853.20	-2805.53	-2791.87	-2861.96	-2856.46	-2809.24	-2796.95
AIC	5730.83	5720.40	5623.07	5597.75	5737.92	5728.91	5630.48	5607.89

The models are applied to the 95% quantile of extreme returns on the S&P 500 index over the period January 2, 1957, until September 1, 2008. Model A, B, C and D correspond to an ETAS model with a power law triggering function. Model E, F, G and H correspond to an ETAS model with an exponential triggering function. In model A, C, E and G the magnitude of the events have no influence on the triggering of subsequent events, that is the parameter restriction $\alpha = 0$ is imposed. In model A, B, E and F the history of the events has no influence on the magnitude of subsequent events, that is the parameter restriction $\eta = 0$ is imposed. Standard deviations are shown in between parentheses.

Table 3: Estimation results crashes

Model	A	B	C	D	E	F	G	H
μ	0.0075	0.0079	0.0083	0.0085	0.0102	0.0106	0.0105	0.0111
	(0.0020)	(0.0020)	(0.0019)	(0.0020)	(0.0015)	(0.0016)	(0.0015)	(0.0015)
K_0	0.0338	0.0310	0.0359	0.0326	0.0279	0.0256	0.0297	0.0264
	(0.0052)	(0.0050)	(0.0052)	(0.0048)	(0.0036)	(0.0035)	(0.0034)	(0.0032)
γ	0.0265	0.0280	0.0253	0.0312				
	(0.0144)	(0.0148)	(0.0143)	(0.0163)				
ω	1.4766	1.4116	1.6781	1.4069				
	(0.7137)	(0.6642)	(0.8484)	(0.6546)				
α		0.0987		0.1304		0.0939		0.1248
		(0.0339)		(0.0205)		(0.0348)		(0.0211)
β					0.0349	0.0351	0.0375	0.0380
					(0.0047)	(0.0047)	(0.0044)	(0.0044)
ξ	0.2885	0.2885	0.1842	0.1640	0.2885	0.2885	0.1860	0.1680
	(0.0479)	(0.0479)	(0.0414)	(0.0402)	(0.0479)	(0.0479)	(0.0412)	(0.0401)
ϕ	0.5550	0.5550	0.2369	0.2477	0.5549	0.5550	0.2461	0.2583
	(0.0334)	(0.0334)	(0.0290)	(0.0293)	(0.0334)	(0.0334)	(0.0283)	(0.0286)
η			0.1461	0.1291			0.1210	0.1039
			(0.0246)	(0.0225)			(0.0171)	(0.0153)
$\log L(\theta)$	-2927.53	-2925.34	-2866.87	-2857.91	-2931.06	-2929.12	-2870.07	-2862.05
AIC	5867.05	5864.68	5745.74	5729.82	5876.13	5874.24	5752.14	5738.10

The models are applied to the 95% quantile of extreme negative returns on the S&P 500 index over the period January 2, 1957, until September 1, 2008. Model A, B, C and D correspond to an ETAS model with a power law triggering function. Model E, F, G and H correspond to an ETAS model with an exponential triggering function. In model A, C, E and G the magnitude of the events have no influence on the triggering of subsequent events, that is the parameter restriction $\alpha = 0$ is imposed. In model A, B, E and F the history of the events has no influence on the magnitude of subsequent events, that is the parameter restriction $\eta = 0$ is imposed. Standard deviations are shown in between parentheses.

Table 4: Estimation results simulated data

Model	A	B	C	D	E	F	G	H
μ	0,0083	0,0088	0,0090	0,0029	0,0104	0,0109	0,0107	0,0114
	(0,0021)	(0,0021)	(0,0020)	(0,0007)	(0,0016)	(0,0016)	(0,0016)	(0,0016)
K_0	0,0251	0,0227	0,0277	0,0202	0,0237	0,0214	0,0254	0,0230
	(0,0044)	(0,0044)	(0,0044)	(0,0050)	(0,0030)	(0,0031)	(0,0030)	(0,0032)
γ	0,0144	0,0147	0,0149	0,0193				
	(0,0116)	(0,0118)	(0,0116)	(0,0145)				
ω	3,7901	3,4995	4,0072	2,9153				
	(7,9141)	(6,0194)	(6,8997)	(4,7757)				
α		0,1039		0,1614		0,0990		0,0935
		(0,0625)		(0,1070)		(0,0605)		(0,0834)
β					0,0316	0,0314	0,0341	0,0346
					(0,0041)	(0,0041)	(0,0040)	(0,0043)
ξ	0,2834	0,2854	0,1803	0,1080	0,2881	0,2870	0,1772	0,1606
	(0,0548)	(0,0548)	(0,0505)	(0,0708)	(0,0544)	(0,0541)	(0,0496)	(0,0507)
ϕ	0,5589	0,5567	0,2357	0,4898	0,5566	0,5556	0,2454	0,2581
	(0,0381)	(0,0379)	(0,0319)	(0,0661)	(0,0376)	(0,0373)	(0,0301)	(0,0304)
η			0,1246	0,1174			0,1127	0,0995
			(0,0259)	(0,0491)			(0,0193)	(0,0200)
Convergence	854	865	884	887	1000	1000	1000	998

The discrete time procedure is used to generate 1000 data series from each model. Model A, B, C and D correspond to an ETAS model with a power law triggering function. Model E, F, G and H correspond to an ETAS model with an exponential triggering function. In model A, C, E and G the magnitude of the events have no influence on the triggering of subsequent events, that is the parameter restriction $\alpha = 0$ is imposed. In model A, B, E and F the history of the events has no influence on the magnitude of subsequent events, that is the parameter restriction $\eta = 0$ is imposed. Standard deviations are shown in between parentheses.

Table 5: Kolmogorov-Smirnov tests

Model	A	B	C	D	E	F	G	H
Extremes	0.2204	0.3906	0.1881	0.3277	0.1441	0.2594	0.0974	0.1697
Crashes	0.2003	0.3192	0.1488	0.2523	0.0613	0.1041	0.0501	0.0824

The tests are performed on the transformed times $\{\tau_i\}$ specified by the models. The null hypothesis of the test is transformed times $\{\tau_i\}$ are distributed according to a homogeneous Poisson process with intensity 1. In the Table the p-values of the Kolmogorov-Smirnov tests are reported.

Table 6: Quadratic and Log Probability Scores

		2008–2013				2008–2010			
	Model	A	C	E	G	A	C	E	G
Crashes	QPS	0.39	0.39	0.39	0.39	0.41	0.41	0.40	0.40
	LPS	0.58	0.58	0.58	0.58	0.60	0.60	0.59	0.59
Extremes	QPS	0.33	0.32	0.32	0.32	0.29	0.29	0.29	0.28
	LPS	0.51	0.50	0.51	0.50	0.46	0.45	0.45	0.45

QPS and LPS of probability predictions of the occurrence of events within 5 days from 1 September, 2008, to 1 January, 2013 and from 1 September, 2008, to 1 January, 2010.

Table 7: Diebold Mariano Statistics

		2008–2013		2008–2010	
	Models	E	G	E	G
Crashes	A	1.30	1.57	2.02	1.90
	C	0.81	1.20	1.97	1.89
Extremes	A	1.24	2.29	1.15	1.37
	C	0.16	1.53	0.56	0.98

DM statistics comparing predictions of non-nested models. Probability predictions of the occurrence of events within 5 days are made from 1 September, 2008, to 1 January, 2013 and from 1 September, 2008, to 1 January, 2010.

Table 8: Adjusted Mean Squared Prediction Errors

	2008–2013		2008–2010	
Models	A-C	E-G	A-C	E-G
Crashes	2.58	1.83	1.76	0.47
Extremes	2.29	2.94	1.88	1.50

Adjusted MSPE comparing predictions of nested models. Probability predictions of the occurrence of events within 5 days are made from 1 September, 2008, to 1 January, 2013 and from 1 September, 2008, to 1 January, 2010.

Table 9: Results for the Early Warning Systems

		2008–2013				2008–2010			
	Model	A	C	E	G	A	C	E	G
Crashes	Hits	248	249	243	242	161	161	158	156
	False alarms	140	137	136	134	68	66	64	62
	Hit rate	0.61	0.61	0.60	0.60	0.87	0.87	0.85	0.84
	False alarm rate	0.19	0.19	0.19	0.18	0.41	0.40	0.39	0.38
	KSS	0.42	0.43	0.41	0.41	0.45	0.46	0.46	0.46
Extremes	Hits	341	343	333	335	221	220	218	216
	False alarms	118	121	113	107	58	57	52	50
	Hit rate	0.74	0.74	0.72	0.72	0.94	0.93	0.92	0.92
	False alarm rate	0.18	0.18	0.17	0.16	0.51	0.50	0.46	0.44
	KSS	0.56	0.56	0.55	0.56	0.43	0.43	0.47	0.48

Hits, hit rates, false alarms, false alarm rates and the Hanssen-Kuiper Skill Scores (KSS) of the EWS predicting the occurrence of events within 5 days from 1 September, 2008, to 1 January, 2013 and from 1 September, 2008, to 1 January, 2010. An event is predicted when the probability according to the models given the history of the event process exceeds 0.5.

Table 10: Estimation results standardized residuals extremes

	GARCH				GJR			
Model	E	F	G	H	E	F	G	H
μ	0.0073	0.0073	0.0072	0.0072	0.0105	0.0105	0.0097	0.0097
	(0.0022)	(0.0022)	(0.0022)	(0.0022)	(0.0029)	(0.0029)	(0.0027)	(0.0027)
K_0	0.0096	0.0096	0.0094	0.0094	0.0128	0.0128	0.0117	0.0117
	(0.0016)	(0.0018)	(0.0015)	(0.0017)	(0.0025)	(0.0026)	(0.0022)	(0.0024)
β	0.0112	0.0112	0.0110	0.0110	0.0162	0.0162	0.0145	0.0146
	(0.0019)	(0.0019)	(0.0018)	(0.0018)	(0.0036)	(0.0036)	(0.0031)	(0.0031)
α		0.0010		0.0010		0.0010		0.0010
		(0.0499)		(0.0500)		(0.0445)		(0.0454)
ξ	0.1542	0.1541	0.1527	0.1526	0.1650	0.1650	0.1582	0.1582
	0.0412	0.0412	0.0410	0.0410	0.0406	0.0406	0.0401	0.0401
ϕ	1.2260	1.2261	0.9559	0.9558	1.2322	1.2321	0.9189	0.9190
	(0.0685)	(0.0685)	(0.1221)	(0.1222)	(0.0682)	(0.0682)	(0.1199)	(0.1198)
η			0.0435	0.0435			0.0702	0.0701
			(0.0186)	(0.0189)			(0.0251)	(0.0256)
$\log L(\theta)$	-3560.80	-3560.82	-3557.88	-3557.90	-3590.97	-3591.00	-3586.87	-3586.90
AIC	7131.59	7133.64	7127.75	7129.80	7191.94	7194.00	7185.74	7187.80

The models are applied to the 95% quantile of absolute standardized residuals from the GARCH-type models estimated on daily S&P 500 returns over the period January 2, 1957, until September 1, 2008. Model E, F, G and H correspond to an Hawkes model with an exponential triggering function. In model E and G the magnitude of the events have no influence on the triggering of subsequent events, that is the parameter restriction $\alpha = 0$ is imposed. In model E and F the history of the events has no influence on the magnitude of subsequent events, that is the parameter restriction $\eta = 0$ is imposed. Standard deviations are shown in between parentheses.

Table 11: Estimation results standardized residuals crashes

	GARCH				GJR			
Model	E	F	G	H	E	F	G	H
μ	0.0100	0.0100	0.0096	0.0096	0.0114	0.0114	0.0112	0.0112
	(0.0028)	(0.0028)	(0.0028)	(0.0028)	(0.0033)	(0.0034)	(0.0033)	(0.0033)
K_0	0.0090	0.0090	0.0086	0.0085	0.0063	0.0063	0.0060	0.0060
	(0.0019)	(0.0022)	(0.0019)	(0.0021)	(0.0012)	(0.0014)	(0.0011)	(0.0013)
β	0.0112	0.0112	0.0105	0.0105	0.0081	0.0081	0.0077	0.0077
	(0.0027)	(0.0027)	(0.0025)	(0.0025)	(0.0016)	(0.0016)	(0.0015)	(0.0015)
α		0.0010		0.0012		0.0010		0.0010
		(0.0614)		(0.0617)		(0.0575)		(0.0566)
ξ	0.1059	0.1059	0.1123	0.1123	0.1464	0.1464	0.1466	0.1466
	0.0374	0.0374	0.0372	0.0372	0.0388	0.0388	0.0380	0.0380
ϕ	1.4441	1.4440	1.1706	1.1705	1.3267	1.3266	0.9132	0.9133
	(0.0769)	(0.0769)	(0.1533)	(0.1533)	(0.0718)	(0.0718)	(0.1629)	(0.1628)
η			0.0422	0.0421			0.0544	0.0543
			(0.0221)	(0.0224)			(0.0221)	(0.0225)
$\log L(\theta)$	-3664.71	-3664.71	-3662.74	-3662.74	-3673.68	-3673.69	-3670.32	-3670.33
AIC	7339.42	7341.42	7337.48	7339.48	7357.37	7359.39	7352.64	7354.65

The models are applied to the 95% quantile of negative standardized residuals from the GARCH-type models estimated on daily S&P 500 returns over the period January 2, 1957, until September 1, 2008. Model E, F, G and H correspond to an ETAS model with an exponential triggering function. In model E and G the magnitude of the events have no influence on the triggering of subsequent events, that is the parameter restriction $\alpha = 0$ is imposed. In model E and F the history of the events has no influence on the magnitude of subsequent events, that is the parameter restriction $\eta = 0$ is imposed. Standard deviations are shown in between parentheses.

Table 12: Kolmogorov-Smirnov tests

		GARCH				GJR			
	Model	E	F	G	H	E	F	G	H
Extremes									
Quantile (%)	95	0.2099	0.2096	0.2106	0.2104	0.1589	0.1578	0.1773	0.1763
	96	0.1529	0.1523	0.1527	0.1522	0.2916	0.2903	0.2915	0.2902
	97	0.2410	0.2403	0.2395	0.2395	0.3825	0.3800	0.3803	0.3784
	98	0.1419	0.1418	0.1420	0.1418	0.4072	0.4062	0.4059	0.4055
	99	0.1246	0.1070	0.1245	0.1072	0.4243	0.4229	0.4240	0.4229
Crashes									
Quantile (%)	95	0.0378	0.0378	0.0428	0.0428	0.0789	0.0788	0.0798	0.0798
	96	0.0384	0.0384	0.0398	0.0400	0.0580	0.0577	0.0577	0.0573
	97	0.0959	0.0961	0.0961	0.0956	0.1234	0.1229	0.1230	0.1226
	98	0.1097	0.1095	0.1095	0.1094	0.1980	0.1977	0.1988	0.1981
	99	0.0725	0.0722	0.0725	0.0722	0.2531	0.2528	0.2570	0.2567

The tests are performed on the transformed times $\{\tau_i\}$ specified by the models. The null hypothesis of the test is transformed times $\{\tau_i\}$ are distributed according to a homogeneous Poisson process with intensity 1. In the Table the p-values of the Kolmogorov-Smirnov tests are reported.

Table 13: Comparison EWS for crashes

Quantile (%)	95		96		97		98		99	
Model	GARCH	GJR	GARCH	GJR	GARCH	GJR	GARCH	GJR	GARCH	GJR
Hits	245	246	219	221	187	187	127	136	62	80
False alarms	131	136	113	119	101	103	68	77	54	52
Hit rate	0.60	0.61	0.58	0.58	0.59	0.59	0.48	0.51	0.33	0.42
False alarm rate	0.18	0.19	0.15	0.16	0.12	0.13	0.08	0.09	0.06	0.06
KSS	0.42	0.42	0.43	0.43	0.46	0.46	0.40	0.42	0.27	0.37
Model	Pow	Exp	Pow	Exp	Pow	Exp	Pow	Exp	Pow	Exp
Hits	248	243	223	213	189	183	125	124	85	83
False alarms	139	135	134	132	112	106	90	84	49	48
Hit rate	0.61	0.60	0.59	0.56	0.59	0.57	0.47	0.46	0.45	0.44
False alarm rate	0.19	0.19	0.18	0.18	0.14	0.13	0.10	0.10	0.05	0.05
KSS	0.42	0.41	0.41	0.39	0.45	0.44	0.36	0.37	0.40	0.39
Model	ACD	log-ACD	ACD	log-ACD	ACD	log-ACD	ACD	log-ACD	ACD	log-ACD
Hits	325	9	303	14	260	18	219	14	157	12
False alarms	662	32	621	58	699	67	770	69	719	27
Hit rate	0.80	0.02	0.80	0.04	0.82	0.06	0.82	0.05	0.83	0.06
False alarm rate	0.91	0.04	0.82	0.08	0.86	0.08	0.89	0.08	0.76	0.03
KSS	-0.11	-0.02	-0.02	-0.04	-0.04	-0.03	-0.07	-0.03	0.07	0.03

Hits, hit rates, false alarms, false alarm rates and the Hanssen-Kuiper Skill Scores (KSS) of the EWS predicting the occurrence of a negative return exceeding the $0.95 - 0.99\%$ in sample quantile from 1 September, 2008, to 1 January, 2013. An alarm is set when the probability of an event within 5 days is above 0.5 according to the models. The models used are a GARCH(1,1) and a GJR(1,1) model combined with a Student-t distribution, a ACD(1,1) and a log-ACD(1,1) model combined with a Generalized Gamma distribution, and ETAS models with influence of the marks on the triggering probability and without influence of the history of the event process on the size of the marks combined with a power law triggering function and a exponential triggering function referred to as respectively Pow and Exp.

Table 14: Comparison EWS for extremes

Quantile (%)	95		96		97		98		99	
Model	GARCH	GJR	GARCH	GJR	GARCH	GJR	GARCH	GJR	GARCH	GJR
Hits	217	222	174	184	153	157	97	113	52	80
False alarms	23	21	18	21	17	17	14	15	10	52
Hit rate	0.47	0.48	0.42	0.45	0.43	0.44	0.34	0.40	0.29	0.29
False alarm rate	0.03	0.03	0.02	0.03	0.02	0.02	0.02	0.02	0.01	0.01
KSS	0.43	0.45	0.40	0.42	0.41	0.42	0.33	0.38	0.28	0.28
Model	Pow	Exp	Pow	Exp	Pow	Exp	Pow	Exp	Pow	Exp
Hits	341	330	306	303	268	266	190	190	125	124
False alarms	115	106	114	103	107	105	100	92	63	64
Hit rate	0.74	0.71	0.74	0.74	0.75	0.75	0.67	0.67	0.71	0.70
False alarm rate	0.17	0.16	0.16	0.14	0.14	0.14	0.12	0.11	0.07	0.06
KSS	0.56	0.55	0.59	0.59	0.62	0.61	0.55	0.56	0.64	0.64
Model	ACD	log-ACD	ACD	log-ACD	ACD	log-ACD	ACD	log-ACD	ACD	log-ACD
Hits	303	23	273	17	248	8	187	3	0	11
False alarms	453	41	464	51	708	54	363	19	0	3
Hit rate	0.65	0.05	0.66	0.04	0.70	0.02	0.66	0.01	0.00	0.06
False alarm rate	0.68	0.06	0.64	0.07	0.91	0.07	0.43	0.02	0.00	0.00
KSS	-0.02	-0.01	0.02	-0.03	-0.21	-0.05	0.23	-0.01	0.00	0.06

Hits, hit rates, false alarms, false alarm rates and the Hanssen-Kuiper Skill Scores (KSS) of the EWS predicting the occurrence of an absolute return exceeding the $0.95 - 0.99\%$ in sample quantile from 1 September, 2008, to 1 January, 2013. An alarm is set when the probability of an event within 5 days is above 0.5 according to the models. The models used are a GARCH(1,1) and a GJR(1,1) model combined with a Student-t distribution, a ACD(1,1) and a log-ACD(1,1) model combined with a Generalized Gamma distribution, and a ETAS models with influence of the marks on the triggering probability and without influence of the history of the event process on the size of the marks combined with a power law triggering function and a exponential triggering function referred to as respectively Pow and Exp.